CHAPTER-VII

FORECASTING TOURISM DEMAND AND MARKET TRENDS

7.1 INTRODUCTION

Two important concepts for tourism analysis are forecasting and demand. Forecasting refers to the task of making predictions. In practice, most of the predictions made by tourism analysis concern the demand for tourism commodities. An understanding of the concept of demand and of its various connotations can help in understanding the practice and problems of forecasting tourism trends. Conversely, an appreciation of the nature of the forecasting can shed light on how demand is studied and measured. The nature of forecasting and some of the general issues associated with selecting a forecasting model are described below. The nature of demand is then considered, with special attention given to both the definitions of demand and the forces that cause demand to change. Finally, three different forecasting models are described which illustrate various approaches which can be used in forecasting tourism trends.
7.2. Nature of Forecasting

Virtually all policy analysis and planning problems in tourism require forecasts of future conditions. Estimates of future levels of demand for different commodities, travel volumes, the market share of various destinations or business, household incomes, interest rates on loans, changes in consumer tastes, and many other economic and social variables are vital to managing and planning tourism development. Forecasting can give us an idea of what future conditions may be like if we fail to take corrective action, and it can provide us with an assessment of the possible outcomes of alternative courses of action.

The challenges of successful forecasting are more than just the technical difficulties of developing an accurate model. Forecasting models must be developed with a clear understanding of both the nature of the problem for which forecasts are desired and of the resources available to the analyst charged with making the forecast. Stynes identified four factors that should be considered when developing a forecasting model: (1) the organizational environment, (2) the decision making situation, (3) existing knowledge; and (4) the nature of the phenomenon being studied.
7.2.1 The Organisational Environment

Each organisation has resources, ways of operating, and objectives specific to it. These characteristics influence the goals and types of forecasting the decision-makers in that organisation want. An agency that prides itself on being politically neutral and on producing objective, high-quality forecasts will have very different standards for forecasting than an organisation devoted to the lobbying of a predetermined political position. The availability of resources such as data banks, computers, software, statisticians, and other technical experts will also influence the type of forecast that can be developed. We need to be aware of all these aspects of the organisational environment in order to design a model that will function effectively within that environment.

7.2.2 The Decision-making Situation

This is related to the organisational environment. Some organisations need to make decisions quickly for their immediate future, others work with a more distant planning horizon and have a longer time period available for developing their model. The level of precision required for a decision is also important in selecting the appropriate forecasting technology. Generally, the greater the precision required, the more complex the model and the longer the lead time
required. Another aspect of the decision-making situation is the level of accuracy required. Whereas precision refers to the amount of detail, accuracy is a measure of the correctness of the forecast. For example, a forecast that the demand for international air travel between the India and the USA will increase next year by 63.25 per cent is precise, but probably not accurate. On the other hand, simply saying that the demand will increase may well be accurate, but it is not precise. As with precision, greater accuracy usually requires more resources and a longer lead time for model development.

7.2.3 Existing Knowledge

Scientific forecasts are based on information about past and current conditions. Some types of forecasting models, such as trend extrapolation or systems simulation models require significant amounts of historical data. The issue of existing knowledge also refers to our understanding of theoretical issues associated with the phenomenon being forecast and our familiarity with the forecasting technology. A match has to be made between the theoretical and technical requirements of the problem.

7.2.4 Nature of the Phenomenon being Forecast

Certain phenomena show a high degree of stability. The percentage of Indians taking vacations has remained virtually unchanged
since the early 1990s. Other phenomena exhibit dramatic changes from year to year in response to fads, local crises or other forces. The former, by their very nature, are much easier to predict than the latter. We will also need to consider whether the phenomenon being modelled is best studied with a stochastic model (which predicts percentage or probabilities) or a deterministic model (which predicts absolute numbers). A choice between a linear and a non-linear model will also depend on the nature of the phenomenon. More general knowledge of the forces that have affected the past behaviour of the phenomenon can assist in the selection of the most useful variables and perhaps even the best model structure.

7.3 Forecasting Models

Forecasting models in tourism may be classified as belonging to three categories: (1) trend extrapolation models; (2) structural models; and (3) simulation models.

Trend extrapolation models, as their name suggests, rely on the extrapolation of a historical series of data into the future. One of the simplest way is a manual plot of data on a graph. The vertical axis on the graph is some measure of tourism demand or market activity, while the horizontal axis contains units of time, such as years. A line is visually fitted to the data and then extended beyond the observed data
to a desired point in the future. More sophisticated models are available to accomplish the same task. These include simple regression models, exponential models, logistic models, quadratic equations, and harmonic analysis. Despite the differences in the statistical complexity of each and the shape of the extrapolation curve, we assume in each model that the observed trend will continue for some reasonable period of time into the future.

Structural models depend on the identification of the relationship between some measure of tourism demand and a series of casual variables, such as price, income, distance, or competition. These relationships are usually identified using multiple regression or analysis of variance and cross-sectional data. Once the model has been calibrated, estimates of future values of the causal variables are used in the model to make a forecast of future tourism demand.

Simulation models are a complex set of equations that typically combine both trend extrapolation and structural models into a more comprehensive systems simulation. Relationships between many variables, including feedback, synergistic, and dampening effects, are specified through a series of interrelated equations. These models also rely on historical data for model calibration. Forecasts are made by specifying expected values for the causal variables, and then solving the
system of equations to arrive at predicted values of the dependent variables.

The choice of the most appropriate model involves consideration of the four factors described earlier. The ultimate choice often requires trade-off between a model that will provide the ideals of the greatest accuracy and precision possible and the constraints imposed by time, budget, and other resources. No single model is best on all criteria. All models are capable of producing good-quality forecasts if they are properly developed and applied, if adequate data are available, and if the problem being studied conforms closely to the assumption implicit in the specific model. The degree to which the development and application of a forecasting model departs from these conditions ultimately determines the quality of the forecast.

7.4 NATURE OF DEMAND

Demand is an ambiguous word with at least four definitions used by tourism analysts. The most traditional definition is that of neo-classified economics; demand is the schedule of quantities of some goods or service that will be consumed at various specified prices. Higher consumption is usually associated with lower prices; lower
consumption with higher prices. Demand in this sense, can be described graphically as:

The downward sloping line, DD', reflects the inverse relationship between price and consumption. Consumption, in the context of tourism, refers to the purchase of some goods or service, such as a hotel room; participation in some activity, such as a pleasure drive in a private automobile; or attendance at an attraction, such as visiting a historical site.

Demand is also used to refer to actual consumption. This definition of demand would be represented as a single point on DD' in the above figure. Such a point, labelled X, is the pairing of a specific price, P₁, and observed consumption, Q₁. This is arguably the most
price. $P_1$, and observed consumption, $Q_1$. This is arguably the most common use of the word ‘demand’ but it is of limited usefulness to tourism analysts because it tells us nothing about trends of levels of unmet demand. It is not, therefore, a useful definition for forecasting.

A third definition is that of unmet demand, which is also referred to as latent demand. Latent demand is a measure of the difference between the potential level of consumption and the observed level. The difference may be due to shortage of supply, excessively high prices, scheduling problems, or other barriers. Latent demand is of special interest to tourism planners because it represents the potential for market expansion.

Finally, demand is used to refer directly to a forecast of future consumption. This conception of demand is closely related to the neo-classical definition, but there are important distinctions. Demand in the sense of future participation is seen as a function of many variables, not just price. It also refers to the anticipated mix and values of those variables that affect demand. Demand as future consumption is thus both broader and narrower than the neo-classical definition. It includes more variables, but the focus is on estimating one single value, not a schedule of values.
7.4.1 Demand Shifters

The variables implicit in demand as future consumption are known collectively as demand shifters. These include consumer characteristics such as age, education, previous experience with similar products, and tastes as well as the effects of promotional efforts, product innovation, and new technology. Consider the demand for rooms at a given hotel. This demand, in the neo-classical sense would be represented by curve DD' in our previous figure. If the hotel adds a new recreational complex, shuttle services to a nearby airport, or expands its conference and meeting facilities, the demand is likely to grow. This would be represented by shifting DD' to the right, to D_1D_1'. Consumers are willing to pay more, P_2 for the same level of consumption. They are also willing to consume more, Q_2, if the price remains at P_1.

If the hotel begins to deteriorate through poorer maintenance, a lessened quality of service, or unfavourable publicity about hotel problems, the demand can be expected to drop. This is reflected by shifting DD' to the left, to D_2D_2'. Consumers would be willing to purchase the original number of rooms only if the price drops to P_3. If prices do not change, total consumption will fall to Q_3.
Demand analysis, including forecasting, may focus on either an individual or a group. The patterns for individuals tend to be more complex and have a higher degree of variance, and thus are harder to predict accurately than demand patterns for groups. The main reason for this is the fact that large numbers of individuals tend to average out the idiosyncratic behaviour of single individuals. Large groups tend to display more stable patterns that cluster around a mean value. Young and Smith described effects of the level of aggregation on demand forecasting. Their work confirmed the experience of many others who have noted that the reliability and accuracy of models increases as the level of aggregation increases. While this is desirable to a point, the most accurate models are often obtained at the most general levels of analysis – analysis so generalized and based on such highly aggregated data that the results have little value for policy and planning problems. As with the other issues surrounding the selection of a forecasting model, the analyst must make a trade-off between a highly accurate, highly aggregated, but less useful model and one that has a lower level of aggregation and thus potentially greater usefulness but with lower accuracy and reliability.
7.4.2 Elasticity

Elasticity is a concept closely tied to the neo-classical definition of demand. In the previous figure, the slope of DD' indicates the degree to which consumption changes given a change in the price of the commodity. A steep line indicates that a large change in price has relatively little effect on consumption; a flatter line reflects large changes in consumption associated with modest changes in price. Quantitatively, elasticity may be defined as the ratio between the observed percentage change in consumption and 1 per cent change in price. A commodity with unitary elasticity is one whose consumption changes at the same rate as price; a 1 per cent drop in price causes a 1 per cent rise in consumption, and vice versa. If the consumption changes at a percentage rate lower than that of a price change (a steep line), the commodity is described as being inelastic. Conversely, if the consumption changes at a rate faster than changes in price (a flatter line), the commodity is elastic.

Two major characteristics of commodities influence their degree of elasticity. Those that are necessities tend to be inelastic. Food staples such as bread and salt, modern necessities such as petrol or telephone service, and life-supporting goods such as some prescription drugs show relatively little short-term variation in consumption due to
changes in price. In contrast, the purchase of luxury goods, which include many tourism commodities, tend to be elastic.

Elasticity may also be examined from the perspective of income. This shift in perspective is of special value to tourism analysts because of the close relationship between the ability to pay for tourism experiences (measured by income) and the willingness to pay for them (measured by demand). If we replace price on the vertical axis of a demand curve by income, the slope of the demand curve becomes positive. Higher incomes are usually associated with higher levels of consumption. The degree of association, reflected by the slope of the line, is the income elasticity. Commodities that are purchased at only slightly elevated levels as incomes rise have low elasticity. Expenditure on other goods, however, rises at a rate faster than income rises. These high elasticity goods are known as superior goods. Commodities whose rate of consumption rises at the same rate as income increases have unitary elasticity. Finally, the consumption of other commodities may actually drop as income rise. These are known as inferior goods. These observations were first formally made by a German statistician, Ernst Engle, in the middle of the nineteenth century. Engle predicted that as incomes rise: (1) the percentage spent on food would fall; (2) the
percentage spent on lodging and clothing would remain about the same;
and (3) the percentage spent on all other goods would rise.

7.5 TREND EXTRAPOLATION

7.5.1 Simple Regression Analysis

One of the simplest but most useful methods of trend extrapolation is simple regression analysis. Simple regression is a method for correlating two variables against each other. Both variables must be measured on an interval scale. The form of a simple regression model is:

\[ Y = a + bX \]

Where:
- \( Y \) = dependent variable;
- \( a, b \) = coefficients to be estimated;
- \( X \) = independent variable.

The dependent variable is some measure of tourism demand or consumption, such as a count of visitors to an attraction, total receipts, or the number of scheduled airline flights. The independent variable may be any of a wide variety of demand shifters such as income, or other aggregate variables such as total market size. One especially useful independent variable is time. Data related to levels of demand are collected for a number of specified units of time, such as years. If the change in business levels is fairly stable over time, a reasonably
accurate forecasting model may be developed by correlating level of demand against time.

Regardless of the independent variable selected, the process of making a forecast is the same. The coefficients in the above equation are estimated using least-squares estimation method using historical data. Once the model has been calibrated, we substitute an expected future value of the independent variable (obtained from an independent forecast) into the equation and solve for $Y$. This new value of $Y$ is the forecast value of future business levels.

The central problem in simple regression is the statistical definition of a linear function that best summarizes a set of data. The procedure of forecasting using simple regression analysis is given below:

**Procedure**

1. Select the appropriate dependent variable and an independent variable, usually some unit of time. Designate the dependent variable as $Y$ and the independent variable as $X$. Select appropriate units of analysis and collect data. A minimum of 10 to 15 observations is normally desirable.
2. Prepare a table. The first column, $X$, will contain values of the independent variable; the second column, $Y$, lists values of the dependent variable. The third column is the product of $X$ and $Y$. The fourth and fifth columns are $X^2$ and $Y^2$ respectively. Also obtain the sums of each column.

3. Calculate $b$ with the equation:

$$b = \frac{n \left( \sum XY \right) - \left( \sum X \right) \left( \sum Y \right)}{n \left( \sum X^2 \right) - \left( \sum X \right)^2}$$

4. Once we have a value for $b$, the value of $a$ is obtained from:

$$a = \frac{\sum Y - b \left( \sum X \right)}{n}$$

Where: $n =$ number of observations.

5. The coefficient of correlation, $r$, is the measure of the goodness of fit between the estimated regression line and the data. It indicates the degree to which there is a linear relationship between $X$ and $Y$. It is calculated with the equation:

$$r = \frac{n \left( \sum XY \right) - \left( \sum X \right) \left( \sum Y \right)}{\sqrt{n(\sum X^2) - (\sum X)^2} \sqrt{n(\sum Y^2) - (\sum Y)^2}}$$

The sign of $r$ will be the same as that of $b$, the slope of the regression line. A positive sign reflects a direct correlation between $X$ and $Y$; a negative sign reflects an inverse relationship. A value close to zero indicates a horizontal line, or no correlation between $X$ and $Y$. The values of $r$ range from 1.00 to $-1.00$. These extreme values as well as the mid-point of 0.00 are easy to interpret, but intermediate values
are more frequently obtained. One method of interpreting this is to square the value, obtaining $r^2$. This statistic may be interpreted as a measure of the explained variance attributable to the independent variable. The extreme value of $r^2$ are 0.00 and 1.00. A value of 0.80 indicates that the independent variable explains 80 per cent of the variation in the dependent variable; a model with an $r^2$ of 0.80 would be considered four times as strong as one that produced an $r^2$ of 0.20.

The above procedure was applied on tourism data to forecast the tourist arrivals for next 10 years. Since the values of $X$ are years, a short cut was used to simplify the calculations of $a$ and $b$. The values of $X$ as years was replaced with values that will cancel out when summed. If the number of years is even, the years are replaced with the value: -3, -2, -1, 1, 2, 3

....If the number is odd, we use....-3, -2, -1, 0, 1, 2, 3.

....The equations for $a$, $b$, and $r$ then become:

$$a = \frac{\sum Y}{n}$$

$$b = \frac{\sum XY}{\sum X^2}$$

$$r = \frac{n (\sum XY)}{\left(\sqrt{n(\sum X^2)}\right) \left(\sqrt{n(\sum Y^2)} - (\sum Y)^2\right)}$$
A forecast is then made by extending the series of codes to the future years. This procedure was applied to forecast tourist arrivals to India in next 10 years. The calculations are given below:

<table>
<thead>
<tr>
<th>Year</th>
<th>X</th>
<th>Tourist arrival (in lacs) (Y)</th>
<th>XY</th>
<th>X²</th>
<th>Y²</th>
</tr>
</thead>
<tbody>
<tr>
<td>1989</td>
<td>-5</td>
<td>17.36</td>
<td>-86.80</td>
<td>25</td>
<td>301.36</td>
</tr>
<tr>
<td>1990</td>
<td>-4</td>
<td>17.07</td>
<td>-68.28</td>
<td>16</td>
<td>291.38</td>
</tr>
<tr>
<td>1991</td>
<td>-3</td>
<td>16.77</td>
<td>-50.31</td>
<td>9</td>
<td>281.23</td>
</tr>
<tr>
<td>1992</td>
<td>-2</td>
<td>18.68</td>
<td>-37.36</td>
<td>4</td>
<td>348.94</td>
</tr>
<tr>
<td>1993</td>
<td>-1</td>
<td>17.65</td>
<td>-17.65</td>
<td>1</td>
<td>311.52</td>
</tr>
<tr>
<td>1994</td>
<td>0</td>
<td>18.86</td>
<td>0</td>
<td>0</td>
<td>355.69</td>
</tr>
<tr>
<td>1996</td>
<td>2</td>
<td>22.88</td>
<td>45.76</td>
<td>4</td>
<td>523.49</td>
</tr>
<tr>
<td>1997</td>
<td>3</td>
<td>23.74</td>
<td>71.22</td>
<td>9</td>
<td>563.58</td>
</tr>
<tr>
<td>1998</td>
<td>4</td>
<td>23.59</td>
<td>94.36</td>
<td>16</td>
<td>556.48</td>
</tr>
<tr>
<td>1999</td>
<td>5</td>
<td>24.82</td>
<td>121.10</td>
<td>25</td>
<td>616.03</td>
</tr>
<tr>
<td>TOTAL</td>
<td>0</td>
<td>222.66</td>
<td>96.28</td>
<td>110</td>
<td>4600.83</td>
</tr>
</tbody>
</table>

\[ a = \frac{\Sigma Y}{n} = \frac{222.66}{11} = 20.24 \]

\[ b = \frac{\Sigma XY}{\Sigma X^2} = \frac{96.28}{110} = 0.88 \]

\[ r = \frac{n(\Sigma XY)}{\left(\sqrt{n(\Sigma X^2)}\right) \left(\sqrt{n(\Sigma Y^2)} - (\Sigma Y)^2\right)} \]

\[ = \frac{11(96.28)}{\left(\sqrt{11(110)} \right) \left(\sqrt{11(4600.83)} - (222.66)^2\right)} = \frac{1059.08}{1117.11} = 0.95 \]
Using the above values of a & b, the tourism arrivals were forecasted for the next 10 years. The following estimated figures were obtained.

<table>
<thead>
<tr>
<th>Year</th>
<th>Estimated tourist arrival (in lacs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>25.52</td>
</tr>
<tr>
<td>2001</td>
<td>26.40</td>
</tr>
<tr>
<td>2002</td>
<td>27.28</td>
</tr>
<tr>
<td>2003</td>
<td>28.16</td>
</tr>
<tr>
<td>2004</td>
<td>29.04</td>
</tr>
<tr>
<td>2005</td>
<td>29.22</td>
</tr>
<tr>
<td>2006</td>
<td>30.80</td>
</tr>
<tr>
<td>2007</td>
<td>31.68</td>
</tr>
<tr>
<td>2008</td>
<td>32.56</td>
</tr>
<tr>
<td>2009</td>
<td>33.44</td>
</tr>
<tr>
<td>2010</td>
<td>34.32</td>
</tr>
</tbody>
</table>

The value of r is 0.95, meaning thereby that $r^2 = 0.90$, which means a very strong model. Therefore, simple regression analysis gives a very reliable estimate of tourism arrival forecast in the country. In fact the data of tourist arrival for the year 2000 was available, which permitted me to check the accuracy of the forecast. The actual value was 26.24 lacs. This represents an error of only 2%, which is negligible for any marketing forecast, depicting the strongness of simple regression analysis model for forecasting tourist arrival. We will now fit exponential model to forecast tourist arrival.
7.5.2 Fitting of exponential model

The exponential model or the logarithmic straight line is used as an expression of the secular movement, when the series is increasing or decreasing by a constant percentage rather than a constant absolute amount. In this case, the data plotted on a semi-logarithmic scale will give a straight line graph. The form of a exponential model is:

\[ Y = ab^x \]

Where \( Y \) : dependent variable (tourist arrival)

\( a, b \): Coefficients to be estimated

\( x \) : independent variable (time)

Taking log on both sides, we get:

\[ \log Y = \log a + X \log b \]

Or \( V = A + BX \)

Where \( V = \log Y \)

\( A = \log a \)

\( B = \log b \)

The curve fitting is then done as was done for simple regression analysis. As done in case of simple regression analysis, we may, to
make the calculations easy, take \( u = X - 1994 \). The calculations are
given in the following table:-

<table>
<thead>
<tr>
<th>Year X</th>
<th>Tourist Arrival Y</th>
<th>( U = X - 1994 )</th>
<th>( V = \log Y )</th>
<th>( U^2 )</th>
<th>UV</th>
<th>( V^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1989</td>
<td>17.36</td>
<td>-5</td>
<td>1.239</td>
<td>25</td>
<td>-6.195</td>
<td>1.535</td>
</tr>
<tr>
<td>1990</td>
<td>17.07</td>
<td>-4</td>
<td>1.232</td>
<td>16</td>
<td>-4.928</td>
<td>1.518</td>
</tr>
<tr>
<td>1991</td>
<td>16.77</td>
<td>-3</td>
<td>1.224</td>
<td>9</td>
<td>-3.672</td>
<td>1.498</td>
</tr>
<tr>
<td>1992</td>
<td>18.68</td>
<td>-2</td>
<td>1.271</td>
<td>4</td>
<td>-2.542</td>
<td>1.615</td>
</tr>
<tr>
<td>1993</td>
<td>17.65</td>
<td>-1</td>
<td>1.246</td>
<td>1</td>
<td>-1.246</td>
<td>1.55</td>
</tr>
<tr>
<td>1994</td>
<td>18.86</td>
<td>0</td>
<td>1.275</td>
<td>0</td>
<td>0</td>
<td>1.626</td>
</tr>
<tr>
<td>1995</td>
<td>21.24</td>
<td>1</td>
<td>1.327</td>
<td>1</td>
<td>1.327</td>
<td>1.761</td>
</tr>
<tr>
<td>1996</td>
<td>22.88</td>
<td>2</td>
<td>1.359</td>
<td>4</td>
<td>2.718</td>
<td>1.847</td>
</tr>
<tr>
<td>1997</td>
<td>23.74</td>
<td>3</td>
<td>1.375</td>
<td>9</td>
<td>4.125</td>
<td>1.891</td>
</tr>
<tr>
<td>1998</td>
<td>23.59</td>
<td>4</td>
<td>1.372</td>
<td>16</td>
<td>5.488</td>
<td>1.882</td>
</tr>
<tr>
<td>1999</td>
<td>24.82</td>
<td>5</td>
<td>1.394</td>
<td>25</td>
<td>6.970</td>
<td>1.943</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td>0</td>
<td>14.314</td>
<td>110</td>
<td>2.045</td>
<td>18.666</td>
</tr>
</tbody>
</table>

We can thus calculate \( A \) & \( B \) as follows :-

\[
A = \frac{\sum V}{N} = \frac{14.314}{11} = 1.301
\]

\( a = \text{Antilog} \ (A) = 20.01 \)

\[
B = \frac{\sum UV}{\sum U^2} = \frac{2.045}{110} = 0.0185
\]

\( b = \text{Antilog} \ (B) = 1.0437 \)

The actual curve fitted would then be :

\[
Y = (20.01)(1.0437)\ u
\]

Or \( V = A + BU \)

\( = (1.301) + (0.0185)U \)
The expected value of V can then be calculated by substituting subsequent years for U. The actual tourist arrival forecast would be antilog (V). The estimated values of tourist arrivals for the next ten years using exponential model is given in the following table:

<table>
<thead>
<tr>
<th>YEAR</th>
<th>U</th>
<th>( V = 1.301 + (0.0185)^u )</th>
<th>Trend value Y = antilog (V)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>6</td>
<td>1.4120</td>
<td>25.82</td>
</tr>
<tr>
<td>2001</td>
<td>7</td>
<td>1.4305</td>
<td>26.95</td>
</tr>
<tr>
<td>2002</td>
<td>8</td>
<td>1.4490</td>
<td>28.12</td>
</tr>
<tr>
<td>2003</td>
<td>9</td>
<td>1.4675</td>
<td>29.34</td>
</tr>
<tr>
<td>2004</td>
<td>10</td>
<td>1.4860</td>
<td>30.62</td>
</tr>
<tr>
<td>2005</td>
<td>11</td>
<td>1.5045</td>
<td>31.95</td>
</tr>
<tr>
<td>2006</td>
<td>12</td>
<td>1.5230</td>
<td>33.34</td>
</tr>
<tr>
<td>2007</td>
<td>13</td>
<td>1.5415</td>
<td>34.79</td>
</tr>
<tr>
<td>2008</td>
<td>14</td>
<td>1.5600</td>
<td>36.31</td>
</tr>
<tr>
<td>2009</td>
<td>15</td>
<td>1.5780</td>
<td>37.89</td>
</tr>
<tr>
<td>2010</td>
<td>16</td>
<td>1.5970</td>
<td>39.54</td>
</tr>
</tbody>
</table>

The co-relation coefficient was calculated using the formula:

\[
r = \frac{n \sum UV}{(\sqrt{n \sum U^2})(\sqrt{n \sum V^2} - (\sum V)^2)}
\]

\[
= \frac{22.495}{(34.78)(0.6603)} = 0.979
\]

The value of r was 0.979 meaning thereby that \( r^2 = 0.959 \), which means a very strong model, stronger than simple regression analysis. The predicted value of tourist arrival was 25.82 lacs as compared with the actual value of 26.24 lacs which represented an error of only 1.6%.
This is very negligible. Exponential model thus approved to be stronger than simple regression analysis.

We will now explain rest of the two methods, viz. Gravity Model and Probabilistic Model. However, when these were applied to forecast tourist arrivals in Indian situation, they did not prove to be good models. The estimated figures were too far from real figures. Without going into details, these methods are given below in brief for sake of completeness.

7.6 GRAVITY MODEL

The gravity model is a well-known structural forecasting model. As its name suggests, the gravity model is based on an analogy to Newton’s law of gravitation. Newton’s law states that the gravitational attraction between any two bodies is directly proportional to the masses of the two bodies and inversely proportional to the square of the distance between them:

\[ I_{ij} = \frac{GM_i M_j}{D_{ij}^2} \]

Where: \( I_{ij} \) = gravitation attraction between two bodies, i and j;

\[ G \] = gravitational constant;

\[ M_{ij} \] = masses of i and j;

\[ D_{ij} \] = distance between the centers of i and j.
This rather simple formulation has been the inspiration for a growing body of travel and interaction models in the social sciences. Interaction here refers potentially to any form of exchange between two social groups. This may be financial flows, telephone calls, mail volumes, marriages, trips, and literally hundreds of other variables. The masses of the social groups may be expressed in terms of population, relative wealth, retail floor space, destination attractiveness and many other variables. Distance is usually measured in terms of physical separation, but measures of travel time or social distance can also be used.

Crampon was the first to demonstrate explicitly the usefulness of the gravity model to tourism research. Crampon’s basic model, as well as that of most other researchers who have used the gravity model is:

\[ T_{ij} = \frac{GP_iA_j}{D_{ij}^a} \]

Where:
- \( T_{ij} \) = some measure of tourist travel between origin \( i \) and destination \( j \);
- \( G, a \) = coefficients to be estimated;
- \( P_i \) = measure of the population size of origin \( i \);
- \( A_j \) = attractiveness or capacity of destination \( j \);
- \( D_{ij} \) = distance between \( i \) and \( j \).
As with other structural models, the above equation must be calibrated with historical data before it can be used for forecasting. If this calibration is done successfully, not only do we then have a forecasting model, but the coefficients 'G' and 'a' may have some intrinsic interest. The value of 'a' for example, reflects the relative strength of distance as a deterrent to travel. The larger the estimated value of 'a', the greater the effect of distance on reducing the number of trips. The value of 'G' is less easily interpreted. In Newton's model, 'G' is a universal constant – one of the four universal constants that shape the structure of the universe. For tourism, 'G' is a proportionality constant that adjusts the magnitude of the other variables so that they explain as closely as possible the observed level of tourism activity, $T_{ij}$. The relative values of 'G' in different modeling situations might contain some meaning of use to tourism research, but the subject has not yet been fully addressed.

The most important reason for developing a gravity model is not to replicate observed travel patterns or to examine the magnitudes of $a$ and $G$, but to provide a forecasting methodology. Given estimates of future values of $P_i$ and $A_j$, and assuming $a$, $G$, and $D_{ij}$ are constant, we can predict future levels of tourism demand.
The gravity model is used to forecast trips between a single origin and a single destination within a specified time period. If we want to make a forecast for a system with multiple origins and/or multiple destinations, we can simply calibrate a gravity model for each origin-destination pair. We shall now discuss some of the major limitations in gravity models. Some of the modifications have then been proposed to overcome those limitations.

One weakness in the basic gravity model, as expressed in the form of above equation is that it is unconstrained. In other words, there is no upper limit on the number of trips that the model may forecast. If, for example, we are calibrating a model to forecast travel by residents of some state to a given hill station and if we were to use the number of waterfall as the attraction component, our model would predict a doubling of visitors with a doubling of waterfall numbers. If the capacity were to increase tenfold, our forecast would increase tenfold. Unlimited use increases are not realistic. There is some upper bound to the number of trips a given population can make in one year; yet the basic gravity model can not reflect this fact. The solution to the problem is to develop a constrained gravity model in which a realistic upper limit is identified. This is usually accomplished by developing a two-stage model. The first stage estimates the total number of trips that
can be generated under specified conditions; the second part allocates those trips to competing destinations. A common form of the trip-generation component of a constrained gravity model is a simple regression equation relating income, population size, or mobility to the total number of trips expected. Multiple regression equations might also be used to combine several independent variables in a more accurate and comprehensive model. The trip-distribution component then allocates the total number of trips to available destinations.

Another criticism of gravity models, whether in the unconstrained or constrained format, is that they have no theoretical basis. This criticism was historically correct but is irrelevant, and is no longer true. Stewart and Zipf who independently developed the concept of the gravity model based their formulations explicitly on an analogy to Newton’s law of gravitation. Although their models had no theoretical basis, it has been shown empirically that their models and various modifications that developed were as successful or more successful in forecasting travel patterns than models developed directly from theory. Further, Niedercorn and Bechdoldt have derived the gravity model from existing economic theory. They demonstrated that the gravity model is a logical and theoretically sound solution for the
Another reason that the gravity model continues to be a popular structural forecasting tool is that it allows for substantial refinement and modification.

**Procedure**

1. Specify the origin-destination pairs and a relevant time period for data collection. Collect data on (i) the total number of trips from the origin to each destination; (ii) the population of the origin; (iii) the capacity or attractiveness of the destination; and (4) some measure of the distance between origin and destination.

2. Define the per capita trip rate for the origin’s population:

   \[ K = \frac{\sum T_{ij}}{P_i} \]

   Where: \( k \) = per capita trip rate;

   \[ \sum T_{ij} = \text{number of trips to all destinations by residents of i;} \]

   \[ P_i = \text{population of i;} \]

3. Calculate the total attractiveness of all destinations by summing the attractiveness measure of each destination:

   \[ A = \sum A_j \]

   Where: \( A = \text{aggregate attractiveness of all destinations;} \)
4. Calculate the expected number of trips for all travelers from the origin to each destination under the assumption that distance has no effect:

\[ V_{ij} = \frac{kP_iA_i}{A} \]

Where: \( V_{ij} \) = expected number of trips; other variables are as defined previously.

5. Calculate the effect of distance on the expected number of trips by dividing the actual number of trips by the expected number:

\[ \frac{T_{ij}}{V_{ij}} \]

6. Obtain a measure of the distance between the origin and each destination and carry out the following regression:

\[ \log \frac{T_{ij}}{V_{ij}} = \log a + b (\log D_{ij}) \]

7. Remove logs, substitute using above equation and let \( G = \propto k/A \)

(where \( \propto = \text{antilog } a \)) to obtain:

\[ T_{ij} = \frac{GP_iA_i}{D_{ij}^b} \]

8. To make a forecast, substitute predicted values for \( P_i \) and \( A_i \) for the values used in calibrating the model. Solve for the predicted \( T_{ij} \).
This procedure may be used either for future travel patterns between an existing origin and destination or the model may be calibrated using an existing origin and destination pair and then applied to a different but similar pair.

7.7 PROBABILISTIC TRAVEL MODEL

The probabilistic travel model is another example of a structural forecasting model in which a prediction about travel volumes is made on the basis of a hypothesized structure relating several travel variables. This particular model differs from the gravity model in that the forecasts are expressed in terms of probabilities or percentages of total trips rather than as numbers of actual trips. This model can be combined with a trip-generation model such as a trend extrapolation model to develop a constrained gravity model.

The model is based on the argument that the probability a consumer will select a particular product such as a tourism destination is directly proportional to the ‘utility’ of that product with respect to all alternative products. An important advantage of this model for tourism is that it allows the analyst to avoid the unrealistic assumption that a tourist will always go to the most desirable destination and that all other destinations will be totally ignored. The model accommodates the fact that the same traveler may go to different destinations and that many
travelers will go to all available destinations in varying numbers. Specifically, the model assigns a probability estimate to each destination expressing the odds that the average traveler will select that destination. Since the probabilities total to 1.00 for all destinations in a set of competing destinations, the probabilities may also be interpreted as the expected market share of each tourism product or destination.

The central issue associated with the use of this model is the definition of utility. Utility reflects more than just attractiveness; it also includes the effects of cost or access limitations. Although many destination characteristics affect utility, we are limited to those that can be measured on an interval scale.

Before we examine the procedures to be followed in developing a probabilistic travel model, it will be helpful to identify several assumptions implicit in such models. First, the model is based on the assumption that travelers from any origin are homogeneous in tastes, their willingness to travel, and in their perceptions of utility. Or to put that differently, it is assumed that the average travelers in any region is an adequate indicator of the behaviour of the population of all travelers. This assumption may be relaxed by developing separate models for different types of travelers – such as different income strata. The model
form, though, remains the same; we simply calculate a larger number of probabilities.

The model is also based on the assumption of equal (not necessarily perfect) knowledge. All destinations in a set are assumed to be equally familiar to the potential travelers. This, too, may be unrealistic and it can be relaxed if we are able to define weights that can be used to adjust probabilities to reflect knowledge levels.

The importance of finding a valid measure of utility is obvious in this procedure. The challenge involves not only the identification of the relevant variables but also the proper specification of how they are to be combined.

Finally, it will be helpful to describe a test of the explanatory power of this model. The accuracy of the calibrated model may be estimated using the formula:

\[ r^2 = 1 - \frac{\sum(P_{ij} - \overline{P}_{ij})^2}{\sum(P_{ij} - \overline{P}_{ij})^2} \]

where: \( r^2 \) = the coefficient of determination;

\( P_{ij} = \) actual percentage of trips made by travelers from i to j;

\( \overline{P}_{ij} = \) average percentage of trips made by travelers from i to all destinations;
\[ P_{ij} = \text{predicted percentage of trips made by travelers from } \]
\[ i \text{ to } j. \]

The interpretation of \( r^2 \) is the same as \( r^2 \) calculated for simple regression.

**Procedure**

1. Develop a quantitative measure of destination utility, incorporating both positive and negative qualities relevant to the travel system. The definition of utility must include not just the variables that affect utility, but also how these variables are to be combined.

2. After deciding on an appropriate measure of utility, collect the data necessary to calculate utility for every destination. Record the data and calculate total utility by summing the individual utility measures; \( \sum U_i \).

3. Determine the probability a traveler will choose any particular destination by dividing the utility of that destination by the total utility of all destinations:

\[ P_{ij} = \frac{U_i}{\sum U_j} \]

where: \( P_{ij} = \text{probability a traveler will select } j \).
If desired, we can also develop a multiple regression model to predict the total number of trips that travelers from origin ii are likely to generate to all the resorts, thus producing a two-component model that would be equivalent to a constrained gravity model.

7.8 CONCLUSION

It can thus be concluded that forecasting the tourist demand is an important task in tourism planning. There are a range of tools available to the forecaster, including both qualitative and quantitative methods. The trend extrapolation method which is based directly on statistical analysis of past behaviour proved to be the best in Indian conditions. In this method, one of the inherent assumption was that historical pattern will continue. Such an assumption is not unreasonable for short term and middle term forecasts. If applied cautiously, forecasting can be successfully used for planning and management of Indian Tourism Industry. The Government can prepare for the expected number of tourists in a rational manner rather than doing an adhoc planning based on subjective judgements.