CHAPTER 2

GENERATION OF FORECASTS

2.1 Introduction

In this chapter, we describe briefly the methodology of each of the four methods of forecasting, their limitations and then indicate possible improvement in methods by suitably transforming the series. The entire exercise of forecasting undertaken in the present study is carried out to assess empirically the relative performance of the four methods and not to undertake any routine forecasting exercise as such. In doing so we, first of all, have generated forecasts of the series upto lead period 1, 2, ..., 12 and then tried to improve them, if possible, within the framework of the respective method and finally considered these improved forecasts to evaluate the comparative performance of the methods. While the forecasting and generation of its improved version will be discussed in this chapter, the other part namely the assessment of the relative performance of the methods will be carried out in Chapter 3. In Section 2.2, the Box-Jenkins method is discussed along with the Akaike's Information Criterion. The methodology of the Holt-Winters method and the necessary modification in the series to improve forecast
are described in Section 2.3. The forecasts of the series generated by the Brown Exponential Smoothing method and Stepwise Autoregression method are presented in Section 2.4 and Section 2.5 respectively.

2.2 Box-Jenkins method

2.2.1 Model description

If \( \{X_t, t = 0, 1, 2, \ldots\} \) is a stationary stochastic process having an absolutely continuous spectrum, then Priestley (1980) has shown that \( X_t \) can be expressed as

\[
X_t = \sum_{u=0}^{\infty} \alpha_u a_{t-u},
\]

(1)

where \( \alpha_0, \alpha_1, \alpha_2, \ldots \) are parameters and \( \{a_t\} \) is a sequence of uncorrelated random variables. If we define \( B \) as the back-shift operator i.e. \( B a_t = a_{t-1}, B^2 a_t = a_{t-2} \) etc., \( X_t \) can be written as

\[
X_t = (\sum_{u=0}^{\infty} \alpha_u B^u) a_t = F(B) a_t
\]

(2)

where

\[
F(B) = \sum_{u=0}^{\infty} \alpha_u B^u.
\]

(3)

If it is possible to express \( F(B) \) as the ratio of two polynomials \( \phi(B) \) and \( \theta(B) \) of finite order, at least
approximately, then we can write (2) as

\[ X_t = \frac{\Phi(B)}{\Theta(B)} \alpha_t \]

or \[ \phi(B) X_t = \theta(B) \alpha_t, \quad t = 0, \pm 1, \pm 2, \ldots \] \hspace{1cm} (4)

where \( \phi(B) \) and \( \theta(B) \) are the polynomials of degree \( p \) and \( q \) respectively. If \( \{\alpha_t\} \) is a sequence of independently and identically distributed random variables, the model defined in (4) is known as Auto-Regressive Moving Average (ARMA) model of order \( p \) and \( q \). In the particular cases when \( q = 0 \) or \( p = 0 \), the models are known as Auto-Regressive (AR) and Moving Average (MA) models respectively. Their mathematical expressions are

\[ \phi(B) X_t = \alpha_t, \quad t = 0, \pm 1, \pm 2, \ldots \] \hspace{1cm} (5)

and \[ X_t = \theta(B) \alpha_t, \quad t = 0, \pm 1, \pm 2, \ldots \] \hspace{1cm} (6)

respectively.

Now, it is a common experience that an economic time series is generally not stationary. This arises due to the presence of trend in mean or variance or may be in both. In these circumstances, it is not possible to express the series directly in terms of an AR, MA or ARMA model. But Box and Jenkins (1976) argue that such a nonstationary series can be transformed to a stationary or to an almost
stationary series, if it is differenced an appropriate number of times. Thus, if we have a stochastic process 
\{X_t, t = 0, \pm 1, \pm 2, \ldots\} which is nonstationary and has a trend, we can find a positive integer d such that the transformed series \( W_t = \nabla^d X_t \) becomes stationary, \( \nabla \) being the difference operator viz., \( \nabla X_t = X_t - X_{t-1} \), \( \nabla^2 X_t = X_t - 2X_{t-1} + X_{t-2} \) and so on. Hence, in general case, \( X_t \) can be approximated as

\[
\phi(B) \nabla^d X_t = \Theta(B) a_t, \quad t = 0, \pm 1, \pm 2, \ldots
\]

The above representation is known as Auto-Regressive Integrated Moving Average (ARIMA) model.

2.2.2 Problem of Uniqueness

The complete procedure of model building and forecasting are fully described by Box and Jenkins (1976). In short, they have suggested four basic steps namely

(i) Identification of the model
(ii) Estimation of parameters of the model
(iii) Diagnostic checking of the model
(iv) Forecasting.

First of all, tentative values of \( p, d \) and \( q \) of the model (7) are obtained with the help of autocorrelation and partial autocorrelation functions. Having obtained them, the parameters \( \phi = (\phi_1, \phi_2, \ldots, \phi_p)' \) and \( \theta = (\theta_1, \theta_2, \ldots, \theta_q)' \)
are estimated by minimising

\[ S(\phi, \theta) = \sum_{t=-n'}^{n} \hat{a}_t^2 (\phi, \theta), \]  

where (i) \( \hat{a}_t (\phi, \theta) \) is the estimated value of \( a_t \) for given \( \phi \) and \( \theta \) and (ii) \( n' \) is an integer such that \( \hat{a}_t (\phi, \theta) \) is very near to zero for \( t < -n' \). \( S(\phi, \theta) \) is minimised by Non-linear Least Square method as suggested by Box and Jenkins.

Having estimated the parameters in this fashion, we test the adequacy of the model by the statistic

\[ Q = n \sum_{k=1}^{L} \frac{n_{-k}}{n-k} \gamma_k^2(a), \]

where (i) \( \gamma_k(a) \) is the \( k \)-th order autocorrelation of the estimated \( \hat{a}_t \) and (ii) \( L \) is some arbitrary integer. Under the hypothesis that the model is adequate, \( Q \) is distributed as \( \chi^2 \) with \((L-p-q)\) degrees of freedom. Once the model is found to be adequate, it is used for forecasting.

Apparently, one does not visualise any difficulty in building a suitable model based on the above procedure. But sometimes, the above procedure identifies a number of models which apparently seem to fit the data equally well. In such a situation, one naturally faces the problem of choosing the best model out of many. This problem has been discussed by Chatfield and Prothero (1977) and with the help of a case
study, they have shown that although a Box-Jenkins seasonal model fits the data very well, it fails to generate good forecasts. Kendall (1971) also pointed out a similar difficulty.

2.2.3 Akaike's Information Criterion

In the above context, Akaike (1973) has suggested a criterion, now known as Akaike's Information Criterion (AIC) which is used to identify a model uniquely. This identified model will generate optimum forecasts in mean square sense. The AIC has been defined as:

$$AIC = -2 \log_e (\text{Maximum of the Likelihood}) + 2 (\text{Number of parameters in the model})$$ (10)

and we select a model for which AIC is minimum. In the case of a Gaussian process, it is well known that the maximisation of likelihood is equivalent to the minimisation of the error sum of squares. Hence, AIC strikes a balance between the reduction of mean square error and the increase in the number of parameters. Ozaki (1977) has shown that this criterion has worked quite well in identifying the best parsimonious models for all series discussed by Box and Jenkins (1976). For forecasting the total deposits of all scheduled commercial banks, AIC has also worked reasonably well (Ray, 1980). In Ray (1980), we have studied the capability of AIC in identifying ARIMA models by considering...
the time series giving the aggregate deposits of all scheduled commercial banks for the period January 1970 to December 1978. For this purpose, we had identified five models and with the help of each model, forecasts were generated up to the lead period twelve. A model was considered best if total absolute forecast error of all twelve lead periods combined is least. In this context, we had noted that the model identified by AIC was in fact ranked second out of five models considered by us. The work has been extended in the present study by considering a large number of series as well as by extending the period of coverage for the above mentioned series.

2.2.4 AIC of ARIMA model

It may be recalled from Section 1.6 that our twelve series are all seasonal in character. Accordingly, we have built seasonal ARIMA models the description of which is as follows:

$$
\phi_1(B) \phi_2(B^s) \psi_\delta \psi_\delta \delta X_t = \psi_1(B) \psi_2(B^s) a_t,
$$

where (i) $\phi_1(B)$, $\phi_2(B^s)$, $\psi_1(B)$ and $\psi_2(B^s)$ are polynomials in $B$, $B^s$ of order $p$, $P$, $q$ and $Q$, respectively, (ii) $\psi$ and $\psi_s$ are difference operators such that $\psi X_t = X_t - X_{t-1}$ and $\psi_s X_t = X_t - X_{t-s}$, (iii) $a_t$ are independently and identically distributed random variables with $E(a_t) = 0$, $V(a_t) = \sigma^2$, 

(iv) s is the period of seasonality of the series.

Sometimes, before the building of a seasonal ARIMA model, it is necessary to transform the series suitably. In this connection, Box and Cox (1964) have suggested a family of transformation of the form:

\[ Y = \frac{X^\lambda - 1}{\lambda}, \text{ if } \lambda \neq 0, \]

\[ = \log_e X, \text{ if } \lambda = 0. \]  \hspace{1cm} (12)

In our study, except for (i) Index Numbers of Industrial Production, (ii) Index Numbers of Wholesale Prices and (iii) Index Numbers of Consumer Prices for Industrial Workers, we have transformed the nine series logarithmically before the Box-Jenkins method is applied.

Now we evaluate the values of AIC of seasonal ARIMA model defined in (11). If we assume that \( a_t \)'s are normally distributed in addition to the assumptions stated in (iii), we can write the log-likelihood function as:

\[ L = -\frac{n}{2} \log \sigma^2 - \frac{n}{2} \log 2\pi - \frac{1}{2\sigma^2} S(\phi, \theta) \]

\[ + h(\phi, \theta), \]  \hspace{1cm} (13)

where (i) \( n \) is the length of the series on which the model is built,

(ii) \( \phi \) and \( \theta \) are parameters of the model,
(iii) \( S(\phi, \theta) = \sum_{t=-\infty}^{n} [a_t(\phi, \theta)]^2, [a_t(\phi, \theta)] \) being the expected value of \( a_t \) for given \( \phi \) and \( \theta \),

(iv) \( h(\phi, \theta) \) is a function of \( \phi \) and \( \theta \) whose value is sufficiently small compared to \( S(\phi, \theta) \) when \( n \) is large.

Thus, when \( n \) is sufficiently large, we can write approximately

\[
L = -\frac{n}{2} \log \sigma^2 - \frac{n}{2} \log 2\pi - \frac{1}{2\sigma^2} S(\phi, \theta). \tag{14}
\]

Now, in the Box-Jenkins procedure, we estimate \( \phi \) and \( \theta \) by minimising \( S(\phi, \theta) \). Hence, in this case, AIC takes the value

\[
AIC = n \log \sigma^2 + n \log 2\pi + n + 2M, \tag{15}
\]

where \( M \) is the number of parameters estimated in the model.

As we are interested in identifying the best seasonal ARIMA models on the basis of AIC, we have considered only the relevant forms of AIC which may be written as

\[
AIC = n \log \sigma^2 + 2M. \tag{16}
\]

Actually, this is the definition of AIC which has been followed in our study. Since \( \sigma^2 \) is unknown, for the evaluation of AIC we have used the following estimate of \( \sigma^2 \):

\[
\hat{\sigma}^2 = \frac{1}{n'} \sum_{t=-n'}^{n} \hat{a}_t^2(\phi, \theta), \tag{17}
\]

where \( n' \) is an integer such that the value of \( \hat{a}_t(\phi, \theta) \), \( t < -n' \) is very close to zero. Here \( \hat{a}_t(\phi, \theta) \) is the estimate of \( [a_t(\phi, \theta)] \).
2.2.5 Empirical assessment of AIC

In this section, we shall try to find out the capability of AIC in identifying the best ARIMA model from a set of models identified by the Box-Jenkins procedure. For this purpose, the same set of twelve series given in Table A.1 of Appendix A has been taken and a family of five models have been identified for each of the twelve series. The five models for a series are designated as A, B, C, D and E respectively. Of course, it is kept in mind that the models are parsimonious or at least close to it. The complete list of the sixty models so obtained together with other relevant particulars are shown in Table B.1 of Appendix B.

Now the complete assessment of AIC is carried out on the basis of mean percentage square errors vis-a-vis the rank values on the same line as described in Chapter 1. Further, the evaluation of performance will be attempted at two levels; firstly, at lead period 1 and secondly, at all lead periods combined. Now, it is argued by Akaike that AIC is capable of identifying the best ARIMA model and if it is so, then we expect that a model identified by AIC will have the minimum rank value for mean percentage square error. If for any reason whatsoever, it does not attain minimum rank, it
should take at least the rank value 2 or atmost 3 as the rank can take the values 1 to 5. The basic data for the assessment of AIC, namely mean percentage square errors and the corresponding rank values are shown in Table B.2 and Table B.3 of Appendix B.

Let $R_{ijk}$ ($i, j = 1, 2, \ldots, 12; k = 1, 2, \ldots, 5$) be the rank value of the $k$-th model for lead period $j$ when the model is fitted to the $i$-th series. The superiority of the model can be judged by looking at its rank values. If Akaike's claim is justified, $R_{ijk}$ should be 1 or 2 for fixed $i$ and $j$. Table 2.1 represents the rank scores of AIC models for the lead period 1 - 12 based on all twelve series.

**TABLE 2.1 TOTAL RANK SCORE OF AIC-MODELS FOR THE LEAD PERIOD 1 - 12**

<table>
<thead>
<tr>
<th>LEAD PERIOD</th>
<th>TOTAL RANK SCORE</th>
<th>LEAD PERIOD</th>
<th>TOTAL RANK SCORE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>38</td>
<td>7</td>
<td>36</td>
</tr>
<tr>
<td>2</td>
<td>39</td>
<td>8</td>
<td>36</td>
</tr>
<tr>
<td>3</td>
<td>39</td>
<td>9</td>
<td>38</td>
</tr>
<tr>
<td>4</td>
<td>43</td>
<td>10</td>
<td>36</td>
</tr>
<tr>
<td>5</td>
<td>40</td>
<td>11</td>
<td>37</td>
</tr>
<tr>
<td>6</td>
<td>35</td>
<td>12</td>
<td>34</td>
</tr>
</tbody>
</table>
From the Table 2.1, it may be noted that the total rank scores are scattered ranging from 34 to 43 over the various lead periods. Specially, at lead period 1, the total score is 38 which is higher than mid rank value which is 36. Further, if the performance of AIC-models are considered for all lead periods combined, the total score comes out to be 451 which is again higher than mid rank value of 432. Hence, it can be concluded that Akaike's Information Criterion has not identified the best ARIMA model which leads to good forecasts in the sense of minimum mean percentage square errors at least in the case of series considered in our study. While, in our present study, we are not attempting to investigate the reasons for failure of AIC-models, nevertheless, we have noted the fact that AIC may not be always good in identifying the best ARIMA model. The above empirical investigation throws some light on the capability of AIC in tackling the problem of identification.

2.2.6 Parsimony and conclusions

We have also considered the criterion of parsimony in identifying the best ARIMA model as suggested by Box and Jenkins. It is argued by them that the model should be parsimonious. The models so selected are described in Table B.4 of Appendix B. In case, where two or more models are having minimum but equal number of parameters, the model which
is having minimum estimated error variance is selected. Having selected the models on the basis of parsimony, the mean percentage square errors and rank scores are evaluated and they are shown respectively in Table B.5 and Table B.6 of Appendix B. It is quite surprising that the total rank scores of all twelve models combined for the lead period 1 is as high as 39. Once again it may be concluded that the models selected on the basis of parsimony are not good. The same inference can be made if the performance of the models of all lead periods are combined. Thus, we see that neither the Box-Jenkins criterion of parsimony nor AIC is capable of identifying the best ARIMA model. In this context, it may be suggested that the best way to select a model is to identify a family of models and select one whose past performance is best. Accordingly, from the Table B.1 of Appendix B, we have selected those models whose total rank scores of all lead periods combined are minimum. Such models are shown in Table 2.2 and the corresponding mean percentage square errors are shown in Table 2.3. It may be stated that on the basis of past performance, the model is identified uniquely. The discussion in Subsections 2.2.2, 2.2.5 and the above discussion bring out that the existing criteria for model identification do not work satisfactorily for the twelve Indian economic time series considered by us.
<table>
<thead>
<tr>
<th>SERIES/MODEL</th>
<th>$W_t$</th>
<th>MODEL DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1B</td>
<td>$\nabla \nabla_{12} \log X_t$</td>
<td>$W_t = (1+0.17B)(1 - 0.86B^{12})a_t$</td>
</tr>
<tr>
<td>2D</td>
<td>$\nabla \nabla_{12} \log X_t$</td>
<td>$W_t = (1 - 0.12B - 0.19B^2)(1 - 0.85B^{12})a_t$</td>
</tr>
<tr>
<td>3B</td>
<td>$\nabla \log X_t$</td>
<td>$(1 - 0.30B) W_t = (1 - 0.30B^3 + 0.25B^6)(1 + 0.20B^{12})a_t$</td>
</tr>
<tr>
<td>4A</td>
<td>$\nabla \log X_t$</td>
<td>$(1 - 0.46B^{12}) W_t = (1 - 0.22B)a_t$</td>
</tr>
<tr>
<td>5D</td>
<td>$\nabla \log X_t$</td>
<td>$(1 - 0.18B - 0.15B^2 - 0.12B^5)(1 - 0.77B^{12}) W_t = a_t$</td>
</tr>
<tr>
<td>6D</td>
<td>$\nabla \log X_t$</td>
<td>$(1 - 0.19B^6)(1 - 0.58B^{12}) W_t = (1 - 0.63B)a_t$</td>
</tr>
<tr>
<td>7B</td>
<td>$\nabla \log X_t$</td>
<td>$W_t = (1 + 0.06B)(1 + 0.46B^{12}) a_t$</td>
</tr>
<tr>
<td>8B</td>
<td>$\nabla \log X_t$</td>
<td>$(1 - 0.76B^{12}) W_t = a_t$</td>
</tr>
<tr>
<td>9E</td>
<td>$\nabla \nabla_{12}X_t$</td>
<td>$(1 + 0.36B + 0.07B^2 + 0.17B^4)(1 + 0.16B^{12}) W_t = (1-0.77B^{12})a_t$</td>
</tr>
<tr>
<td>10A</td>
<td>$\nabla \nabla_{12}X_t$</td>
<td>$(1 - 0.46B)(1 + 0.92B^{12}) W_t = a_t$</td>
</tr>
<tr>
<td>11A</td>
<td>$\nabla \nabla_{12}X_t$</td>
<td>$(1 - 0.90B)(1 + 0.32B^{12}) W_t = (1 - 0.39B)(1 - 0.8CB^{12})a_t$</td>
</tr>
<tr>
<td>12C</td>
<td>$\nabla \log X_t$</td>
<td>$(1 + 0.66B)(1 - 0.53B^{12}) W_t = (1 - 0.13B - 0.52B^2 - 0.16B^3)a_t$</td>
</tr>
<tr>
<td>LEAD PERIOD</td>
<td>1B</td>
<td>2D</td>
</tr>
<tr>
<td>------------</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>1</td>
<td>2.53</td>
<td>16.07</td>
</tr>
<tr>
<td>2</td>
<td>5.73</td>
<td>32.87</td>
</tr>
<tr>
<td>3</td>
<td>6.36</td>
<td>40.27</td>
</tr>
<tr>
<td>4</td>
<td>7.72</td>
<td>38.03</td>
</tr>
<tr>
<td>5</td>
<td>7.08</td>
<td>28.37</td>
</tr>
<tr>
<td>6</td>
<td>7.32</td>
<td>20.81</td>
</tr>
<tr>
<td>7</td>
<td>11.40</td>
<td>13.55</td>
</tr>
<tr>
<td>8</td>
<td>11.09</td>
<td>14.67</td>
</tr>
<tr>
<td>9</td>
<td>17.25</td>
<td>18.55</td>
</tr>
<tr>
<td>10</td>
<td>21.20</td>
<td>16.65</td>
</tr>
<tr>
<td>11</td>
<td>25.21</td>
<td>22.11</td>
</tr>
<tr>
<td>12</td>
<td>30.67</td>
<td>19.87</td>
</tr>
</tbody>
</table>
2.3 Holt-Winters Method

2.3.1 Model description

This is the second forecasting technique discussed in our study. From our earlier discussion in Section 2.2, it has become clear that the Box-Jenkins procedure is capable of describing adequately the behaviour of time series belonging to the ARIMA class. As most of the time series which occur in practice either belong to the ARIMA class or can be reasonably approximated by a time series belonging to the ARIMA class, one would expect that the Box-Jenkins procedure will generate sufficiently good forecasts for a given time series. However, the Box-Jenkins suffers from the defect that it requires heavy computation. Further, this procedure, though computer based, is not fully automatic in the sense that intermediate manual intervention is necessary to identify the proper ARIMA model. Naturally, one has to search for slightly less sophisticated but fully automatic procedure particularly when very accurate forecasts are not required. This is often the case in an industry where future sales of many products are required for inventory control or for planning production. Generally, forecasts are needed for a short period. A number of methods under the heading "Exponential Smoothing" have been developed to handle routine forecasts of this kind. The primary objective
of such a method is to generate forecasts of sufficient accuracy quickly and inexpensively. Ideally, the method should be such that it can be carried out by a comparatively unskilled worker. Such a method has to be fully automatic. All exponential smoothing techniques satisfy this requirement and the Holt-Winters method comes under this category.

This is an exponential adaptive smoothing technique developed by Holt (1957) and Winters (1960). In this method, the series $X_t$ is assumed to be composed of a linear trend, a seasonal component (assuming the series is seasonal) and an error component. If we further assume that the model is multiplicative in the sense that the seasonal effect is multiplicative with the trend, then we can write

$$X_t = L_t \cdot F_t \cdot E_t,$$

where
(i) $L_t$ is the linear trend at time $t$,
(ii) $F_t$ is the seasonal component at time $t$,
(iii) $E_t$ is the error component at time $t$.

The two components of the linear trend $L_t = \mu_t + t \cdot T_t$ at time $t$ namely $\mu_t$ (level) and $T_t$ (slope) and the seasonal component $F_t$ are successively smoothed with the arrival of a new observation. The set of equations used to smooth and update the above three parameters is as follows:
\[ \mu_t = A(X_t/F_{t-s}) + (1-A) (\mu_{t-1} + T_{t-1}), \]

\[ T_t = C(\mu_t - \mu_{t-1}) + (1-C) T_{t-1}, \quad (19) \]

\[ F_t = D(X_t/\mu_t) + (1 - D) F_{t-s}, \]

where \( s \) is the length of periodicity; \( A, C \) and \( D \) are smoothing constants lying between 0 and 1. Now it may be noted that the lower the values of these three constants, the more steady will be the final forecasts, since the use of low values implies that more weight is given to the past observations and consequently any random fluctuations in the near past will have less influence on the forecasts. As the seasonal factors are multiplicative, the forecast equation may be written as

\[ \hat{X}_{N+h} = (\mu_N + h T_N) F_{N+h'-s}, \quad h = 1, 2, \ldots, \quad (20) \]

where (i) \( \hat{X}_{N+h} \) is the forecast value of \( X_{N+h} \) at time point \( (N+h) \) and (ii) \( h' \) is the modular value of \( h \) with respect to \( s \) i.e. \( h' \) is the remainder when \( h \) is divided by \( s \).

Now the main difficulty is to obtain the optimum values of \( A, C \) and \( D \). Though the criterion of the optimality value is subjective, these values are generally determined on the
basis of forecast performance over a period of time. Assuming a quadratic loss function, the optimum combination of (A,C,D) is obtained from a set of feasible grid of values over the region \(0 < A, C, D < 1\) for which the sum of square errors

\[
U = \sum_{t=M}^{M+H-1} (\hat{X}_{t+1} - X_{t+1})^2
\]

becomes minimum. Here, \((\hat{X}_{t+1} - X_{t+1})\) is the one step ahead forecast error based on a model which is derived on the basis of observations \(X_1, X_2, \ldots, X_t\); \(M\) being an integer less than \(N\). In our exercise, we have taken \(H\) as 12 uniformly for all series. As the object of our study is to study the comparative performance of the various methods, we evaluate the mean percentage square errors up to the lead period 12 for each of the twelve series and the values are shown under case I of Table B.7 of the Appendix B.

2.3.2 Problem of nonlinearity and logarithmic transformation

From the earlier section, we know that the Holt-Winters method presupposes the linearity of trend on which the forecast equation (20) is formulated. So long as the linearity assumption holds good, the technique does perform well. But once this assumption breaks down, the forecast may not be as good as it is expected to be. From the Section 1.6, we know that
only one series namely Index Numbers of Wholesale Prices is linear while the remaining series are nonlinear. As the Holt-Winters method performs best for a linear series, it is desirable to ensure that a non-linear series is transformed so that the transformed series is linear or very near to linear. Thus, if the trend component is quadratic, then the first difference will be linear. Actually, based on this idea, we have taken the first difference of all twelve series before the Holt-Winters method is applied. But to our surprise, we find that the forecasts are still not satisfactory. To justify our claim, we present mean percentage square errors of forecasts for the first two series namely Currency Notes in Circulation and Rupee Securities in Table 2.4. We also present in the same table the mean percentage square errors when the method is applied directly on the original series. These two sets of errors help us to assess the relative performance of the two approaches.
TABLE 2.4 MEAN PERCENTAGE SQUARE ERRORS WHEN THE HOLT-WINTERS METHOD IS APPLIED ON $X_t$ AND $VX_t$

<table>
<thead>
<tr>
<th>LEAD PERIOD</th>
<th>MEAN PERCENTAGE SQUARE ERRORS</th>
<th>CURRENCY NOTES IN CIRCULATION</th>
<th>RUPEE SECURITIES</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$X_t$</td>
<td>$VX_t$</td>
<td>$X_t$</td>
</tr>
<tr>
<td>1</td>
<td>2.00</td>
<td>41.41</td>
<td>11.07</td>
</tr>
<tr>
<td>2</td>
<td>3.12</td>
<td>89.34</td>
<td>18.77</td>
</tr>
<tr>
<td>3</td>
<td>2.93</td>
<td>170.56</td>
<td>19.61</td>
</tr>
<tr>
<td>4</td>
<td>3.77</td>
<td>240.31</td>
<td>20.38</td>
</tr>
<tr>
<td>5</td>
<td>4.27</td>
<td>281.90</td>
<td>16.66</td>
</tr>
<tr>
<td>6</td>
<td>4.68</td>
<td>244.53</td>
<td>12.31</td>
</tr>
<tr>
<td>7</td>
<td>7.70</td>
<td>218.81</td>
<td>17.20</td>
</tr>
<tr>
<td>8</td>
<td>7.16</td>
<td>143.56</td>
<td>22.76</td>
</tr>
<tr>
<td>9</td>
<td>8.17</td>
<td>177.66</td>
<td>31.22</td>
</tr>
<tr>
<td>10</td>
<td>10.34</td>
<td>206.20</td>
<td>26.89</td>
</tr>
<tr>
<td>11</td>
<td>11.77</td>
<td>189.81</td>
<td>24.53</td>
</tr>
<tr>
<td>12</td>
<td>16.49</td>
<td>148.59</td>
<td>26.80</td>
</tr>
</tbody>
</table>

Similar results are obtained for other series also. Mean percentage square errors increase further if the series is difference twice. Perhaps, this has happened due to the fact that when a series is differenced and subsequently the Holt-Winters method is applied, both trend and seasonal components
fluctuate over a wide range of values thus making the forecast equation very unstable. This is more so if the differenced series is having some negative values.

Alternately, to achieve linearity for a non-linear series, we make a simple logarithmic transformation because it is believed that if a series is quadratic, cubic or exponential, the log transformed series will be more linear than the original series. This is more so if a series obeys multiplicative model. However, to see the impact of logarithmic transformation on a series, we apply the Holt-Winters technique on all series after making logarithmic transformation. The results so obtained are presented in Table B.7 of Appendix B, under the heading case II.

2.3.3 Performance of logarithmic transformation and conclusions

Now, to examine the superiority or otherwise of logarithmic transformation of the Holt-Winters technique when the series is non-linear, we count the number of times one type of transformation outperforms the other with respect to mean percentage square error. The result is shown in Table 2.5. The analysis is carried out between two types of series namely linear and nonlinear and within type of series, 'without' or 'logarithmic' transformation.
TABLE 2.5 NUMBER OF TIMES ONE OUTPERFORMS OTHER

<table>
<thead>
<tr>
<th>TYPE OF SERIES</th>
<th>WITHOUT LOGARITHMIC TOTAL TRANSFORMATION</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>LINEAR</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>NON-LINEAR</td>
<td>52 80</td>
<td>132</td>
</tr>
</tbody>
</table>

From the above table, we find that when the series is linear, 'without transformation' outperforms 'logarithmic transformation' in all the twelve cases. Thus, we may conclude that when a series is linear, no logarithmic transformation is required.

On the other hand, when a series is nonlinear, we want to test the hypothesis that the Holt-Winters method does better if a logarithmic transformation is made. If \( \pi \) denotes the population proportion that the Holt-Winters does better in such cases, we like to test \( H_0 : (\pi = 1/2) \) against the alternative hypothesis \( H_1 : (\pi > 1/2) \). Let \( p \) be the sample estimate of \( \pi \) based on \( n \) observations. Now it can be shown that under \( H_0 \), the statistic \( U = 2\sqrt{n} (p - 1/2) \) is asymptotically normal with mean zero and variance unity. From the Table 2.5, we get the estimate of \( p \) as \( p = 0.61 \). Substituting the values of \( n = 132 \) and \( p = 0.61 \) in the
expression for $U$, we get $U = 2.53$ which is significant at 5% level of significance when a one sided test is used. Hence we conclude that if a series is nonlinear, the Holt-Winters method performs better with a logarithmic transformation than without a transformation. Thus, from our analysis, we may summarise our conclusion as follows:

When one is interested in forecasting with the help of the Holt-Winters method, it is better to make logarithmic transformation if the series is nonlinear. If the series is linear, no transformation is necessary.

However, for the purpose of comparative assessment of four methods, we have considered those set of mean percentage square errors which are minimum of Case I and Case II. The complete list of mean percentage square errors to be considered for the comparative assessment is shown in Table 2.6 and the actual comparative assessment will be carried out in the next chapter.

2.4 Brown exponential smoothing method

2.4.1 Model description

The Brown Exponential Smoothing method is an adaptive and generalised version of simple exponential smoothing. If we consider a realisation $X_1, X_2, \ldots, X_N$ of a stochastic process $\{X_t, t = 0, 1, 2, \ldots\}$ we can develop a forecasting model based on the following assumptions:
<table>
<thead>
<tr>
<th>LEAD PERIOD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.00</td>
<td>11.07</td>
<td>22.94</td>
<td>1.02</td>
<td>1.42</td>
<td>2.14</td>
<td>2.28</td>
<td>1.45</td>
<td>8.83</td>
<td>4.10</td>
<td>1.58</td>
<td>412.03</td>
</tr>
<tr>
<td>2</td>
<td>3.12</td>
<td>18.77</td>
<td>56.97</td>
<td>1.28</td>
<td>2.26</td>
<td>3.11</td>
<td>5.60</td>
<td>2.11</td>
<td>12.13</td>
<td>10.99</td>
<td>2.10</td>
<td>228.55</td>
</tr>
<tr>
<td>3</td>
<td>2.93</td>
<td>19.61</td>
<td>63.23</td>
<td>1.82</td>
<td>4.18</td>
<td>3.83</td>
<td>6.58</td>
<td>3.27</td>
<td>11.81</td>
<td>21.69</td>
<td>3.68</td>
<td>507.98</td>
</tr>
<tr>
<td>4</td>
<td>3.77</td>
<td>20.38</td>
<td>114.07</td>
<td>1.78</td>
<td>6.22</td>
<td>3.57</td>
<td>11.09</td>
<td>4.29</td>
<td>21.86</td>
<td>35.82</td>
<td>5.25</td>
<td>547.16</td>
</tr>
<tr>
<td>5</td>
<td>4.27</td>
<td>16.66</td>
<td>141.45</td>
<td>1.98</td>
<td>7.86</td>
<td>3.87</td>
<td>12.65</td>
<td>4.92</td>
<td>23.37</td>
<td>47.02</td>
<td>6.87</td>
<td>552.37</td>
</tr>
<tr>
<td>6</td>
<td>4.68</td>
<td>12.31</td>
<td>153.62</td>
<td>1.59</td>
<td>9.51</td>
<td>4.86</td>
<td>14.50</td>
<td>4.85</td>
<td>29.54</td>
<td>52.55</td>
<td>8.20</td>
<td>600.20</td>
</tr>
<tr>
<td>7</td>
<td>7.70</td>
<td>17.20</td>
<td>164.18</td>
<td>1.62</td>
<td>11.26</td>
<td>5.84</td>
<td>17.45</td>
<td>3.97</td>
<td>27.57</td>
<td>57.87</td>
<td>11.02</td>
<td>530.44</td>
</tr>
<tr>
<td>8</td>
<td>7.16</td>
<td>22.76</td>
<td>153.59</td>
<td>1.72</td>
<td>12.94</td>
<td>6.69</td>
<td>18.71</td>
<td>3.50</td>
<td>34.58</td>
<td>69.02</td>
<td>11.62</td>
<td>486.13</td>
</tr>
<tr>
<td>9</td>
<td>8.17</td>
<td>31.22</td>
<td>152.25</td>
<td>1.73</td>
<td>14.67</td>
<td>6.40</td>
<td>21.40</td>
<td>3.02</td>
<td>29.92</td>
<td>78.18</td>
<td>13.71</td>
<td>220.09</td>
</tr>
<tr>
<td>10</td>
<td>10.34</td>
<td>26.89</td>
<td>153.46</td>
<td>3.12</td>
<td>16.29</td>
<td>6.49</td>
<td>20.63</td>
<td>3.20</td>
<td>20.51</td>
<td>74.76</td>
<td>16.06</td>
<td>344.55</td>
</tr>
<tr>
<td>11</td>
<td>11.77</td>
<td>24.53</td>
<td>149.43</td>
<td>3.57</td>
<td>19.70</td>
<td>6.34</td>
<td>17.71</td>
<td>2.65</td>
<td>15.61</td>
<td>61.55</td>
<td>16.19</td>
<td>238.39</td>
</tr>
<tr>
<td>12</td>
<td>16.49</td>
<td>26.80</td>
<td>153.94</td>
<td>6.26</td>
<td>22.08</td>
<td>7.34</td>
<td>22.82</td>
<td>3.92</td>
<td>10.77</td>
<td>58.41</td>
<td>17.49</td>
<td>238.55</td>
</tr>
</tbody>
</table>
(i) The random variable $X_t$ at time point $t$ consists of two additive components namely $\xi_t$ and $\varepsilon_t$ where $\xi_t$ is completely non-random i.e. deterministic, $\varepsilon_t$ s are uncorrelated random components with mean zero and constant variance $\sigma^2$.

(ii) The deterministic part $\xi_t$ may be reasonably represented, at least locally, by a linear combination of $p$ fitting functions which are deterministic functions of time. It is further assumed that these fitting functions are such that they obey the relationship,

$$f(t+1) = L f(t), \quad (22)$$

where $f(t) = (f_1(t), f_2(t), \ldots, f_p(t))'$ is a column vector of $p$ fitting functions namely $f_1(t), f_2(t), \ldots, f_p(t)$ and $L$ is a nonsingular $p \times p$ matrix. This matrix $L$, generally known as the transition matrix, can be determined when the fitting functions are polynomials, exponentials or sinusoidals.

(iii) The forecast model is built on the assumption that the influence of the observation on the forecast value goes on decreasing exponentially as one moves from the present to the past. In other words, while estimating the parameters of the model, we minimise $\sum_{t=1}^{N} \beta^{N-t} \varepsilon_t^2$ $(0 < \beta < 1)$ rather than minimising $\sum_{t=1}^{N} \varepsilon_t^2$. This is known as Discounted Least Square Technique.
(iv) The forecast model is developed by taking the current period as the point of origin. Now, because of the above assumptions, \( X_t \) can be expressed as

\[
X_t = a_1(N) f_1(t-N) + a_2(N) f_2(t-N) \\
+ \ldots + a_p(N) f_p(t-N) + \epsilon_t \\
= a'(N) f(t-N) + \epsilon_t ,
\]

(23)

where \( a'(N) \) is the column vector of regression coefficients estimated from the \( N \) realisations namely \( X_1, X_2, \ldots, X_N \).

\( f(t) \) is the vector of \( p \) fitting functions and \( \epsilon_t \) s are uncorrelated and identically distributed random variables with \( E(\epsilon_t) = 0 \) and \( V(\epsilon_t) = \sigma^2 \). The estimation procedure and the updating formula of \( a(N) \) are fully discussed by Brown (1962) and following him we get:

(i) \( a(N) = F^{-1} G(N) \),

(ii) \( a(N) = L' a(N-1) + F^{-1} f(0) (X_N - \hat{X}_N) \),

(24)

where (a) \( F = \sum_{j=0}^{\infty} \beta^j f(-j) f'(-j) \),

(b) \( L \) is the transition matrix,

(c) \( \hat{X}_N \) is the one step ahead forecast,

(d) \( \beta \) is the discount factor lying between 0 and 1.
Once the estimate of $a(N)$ is available we can write the forecast function as

$$\hat{X}_{N+h} = \sum_{i=1}^{P} a_i(N) f_i(h)$$

$$= \hat{a}(N) f(h), \quad (25)$$

where $\hat{a}(N)$ is the estimate of $a(N)$ and $\hat{X}_{N+h}$ is the forecast of lead period $h$.

### 2.4.2 Choice of Smoothing Constant

The most crucial factor in the development of a good model for forecasting is the proper choice of smoothing constant $\beta$. While the recommended range of $\beta$ is 0.70 to 0.99, it is usually not known apriori. At this place we may mention that the value of $\beta$ indirectly influences the forecast values. Since the model involves powers of $\beta$, large values of $\beta$ result in inclusion of a larger number of past observations in the model as compared to the situation where smaller values of $\beta$ include less number of past observations. In fact, for large values of $\beta$ one may expect a slow response of forecasting system due to any structural changes in the system. On the other hand if $\beta$ is small, one would require very few observations in the recent past to build the forecast equation, as such, any structural change observed in the system will have immediate impact on the forecasting system. In the extreme case when $\beta = 0$, only
the most recent observation is used to build the forecast model and when $\beta = 1$, all observations are considered equally important and all of them are taken into account for generating forecasts.

However, given a set of observations, we may determine the optimum value of $\beta$ in a way similar to the one adopted to determine the smoothing constants of the Holt-Winters method. If $N$ observations are given, we may choose an integer $M (M < N)$ arbitrarily and on the basis of first $M$ observations, we generate a series of forecasts over a grid of possible values of $\beta$ and compare them with the actuals. We choose that value of $\beta$ for which some measure of effectiveness of forecasts attains optimum. In our exercise, we have taken the measure of effectiveness as

$$ U = \sum_{h=1}^{H} (\hat{X}_{M+h} - X_{M+h})^2, \quad (26) $$

where $\hat{X}_{M+h}$ is the forecast value of $X_{M+h}$ obtained from the Brown model which is based on $M$ observations.

2.4.3 Empirical results.

We know that all of our series are seasonal. Further, they are having trend. These two components namely trend and seasonality may be considered as the deterministic part
(\xi_t) of the series and the residual as the random component
($c_t$). To describe trend, we have taken the linear or quadra-
tic function of time as the fitting functions while the
seasonal parts are represented by Sine and Cosine functions.
Now, to capture the seasonal pattern of the series fully,
we have considered adequate number of Sine and Cosine terms
and in general, \( \xi(t) \) is

\[
\xi_t = \phi(t) + \sum_{i=1}^{5} b_i \sin \frac{2\pi i}{12} t + \sum_{i=1}^{6} c_i \cos \frac{2\pi i}{12} t,
\]

(27)

where \( \phi(t) \) is a linear or quadratic function of \( t \) and
\( b_1, b_2, \ldots, b_5, c_1, c_2, \ldots, c_6 \) are unknown coefficients.
Initially, we estimate an appropriate value of \( \beta \) along the
lines discussed in the earlier paragraph and estimate all
coefficients given in (27). Then keeping this value of \( \beta \)
fixed, we update these coefficients eleven times, each
time adding one observation. In all, we generate forecasts
twelve times up to the lead period 12 and from there, we
calculate the mean percentage square errors which are shown
in Table 2.7.
<table>
<thead>
<tr>
<th>PERIOD</th>
<th>SERIES 1</th>
<th>SERIES 2</th>
<th>SERIES 3</th>
<th>SERIES 4</th>
<th>SERIES 5</th>
<th>SERIES 6</th>
<th>SERIES 7</th>
<th>SERIES 8</th>
<th>SERIES 9</th>
<th>SERIES 10</th>
<th>SERIES 11</th>
<th>SERIES 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.31</td>
<td>22.77</td>
<td>51.58</td>
<td>1.07</td>
<td>0.33</td>
<td>1.95</td>
<td>6.77</td>
<td>0.66</td>
<td>11.12</td>
<td>5.61</td>
<td>2.65</td>
<td>11.02</td>
</tr>
<tr>
<td>2</td>
<td>7.07</td>
<td>24.15</td>
<td>63.60</td>
<td>1.01</td>
<td>1.09</td>
<td>2.28</td>
<td>7.01</td>
<td>1.16</td>
<td>13.32</td>
<td>5.21</td>
<td>5.50</td>
<td>3.00</td>
</tr>
<tr>
<td>3</td>
<td>8.76</td>
<td>24.14</td>
<td>91.93</td>
<td>1.76</td>
<td>2.26</td>
<td>2.99</td>
<td>15.80</td>
<td>3.05</td>
<td>14.65</td>
<td>1.75</td>
<td>4.38</td>
<td>3.31</td>
</tr>
<tr>
<td>4</td>
<td>7.74</td>
<td>27.98</td>
<td>14.89</td>
<td>1.66</td>
<td>2.90</td>
<td>3.42</td>
<td>4.81</td>
<td>3.24</td>
<td>14.38</td>
<td>27.11</td>
<td>4.49</td>
<td>361.73</td>
</tr>
<tr>
<td>5</td>
<td>7.03</td>
<td>33.24</td>
<td>14.47</td>
<td>2.22</td>
<td>3.64</td>
<td>4.35</td>
<td>22.74</td>
<td>5.44</td>
<td>14.73</td>
<td>28.14</td>
<td>3.28</td>
<td>379.73</td>
</tr>
<tr>
<td>7</td>
<td>11.42</td>
<td>47.80</td>
<td>30.01</td>
<td>2.62</td>
<td>7.54</td>
<td>6.76</td>
<td>37.04</td>
<td>5.90</td>
<td>13.81</td>
<td>35.46</td>
<td>2.75</td>
<td>421.81</td>
</tr>
<tr>
<td>8</td>
<td>11.70</td>
<td>56.65</td>
<td>30.47</td>
<td>2.71</td>
<td>10.68</td>
<td>7.60</td>
<td>36.14</td>
<td>6.95</td>
<td>15.96</td>
<td>35.46</td>
<td>2.75</td>
<td>315.40</td>
</tr>
<tr>
<td>9</td>
<td>26.05</td>
<td>67.44</td>
<td>30.41</td>
<td>3.45</td>
<td>15.40</td>
<td>8.83</td>
<td>56.18</td>
<td>7.30</td>
<td>11.91</td>
<td>42.17</td>
<td>2.27</td>
<td>330.50</td>
</tr>
<tr>
<td>10</td>
<td>32.71</td>
<td>58.21</td>
<td>41.45</td>
<td>3.45</td>
<td>20.83</td>
<td>5.02</td>
<td>59.80</td>
<td>6.94</td>
<td>11.76</td>
<td>42.17</td>
<td>2.27</td>
<td>241.48</td>
</tr>
<tr>
<td>11</td>
<td>47.31</td>
<td>67.21</td>
<td>56.01</td>
<td>3.71</td>
<td>27.84</td>
<td>10.64</td>
<td>77.96</td>
<td>8.21</td>
<td>7.67</td>
<td>41.58</td>
<td>2.02</td>
<td>252.42</td>
</tr>
<tr>
<td>12</td>
<td>51.99</td>
<td>80.55</td>
<td>59.14</td>
<td>4.23</td>
<td>36.45</td>
<td>9.79</td>
<td>91.50</td>
<td>10.38</td>
<td>7.19</td>
<td>41.58</td>
<td>2.02</td>
<td></td>
</tr>
</tbody>
</table>
2.5 Stepwise autoregression

2.5.1 Model description and estimation

It may be recalled from our earlier discussion in Subsection 2.3.1 that the Box-Jenkins procedure is not fully automatic and that at the identification stage, some manual intervention is required. Generally, one likes to have a forecasting procedure which automatically generates forecasts. While the Box-Jenkins method comes under the category of non-automatic group, the Stepwise Autoregression method is an automatic one. This procedure tries to construct a series of autoregressive models to describe the behaviour of a given time series and select one which fits the data best in the least square sense.

Let, \( X_1, X_2, ..., X_N \) be the realisation of a stochastic process \( \{ X_t, t = 0, \pm 1, \pm 2, ... \} \) at time points \( t = 1, 2, ..., N \). If we are interested in fitting an autoregressive model to this data, we may consider a general autoregressive model of order \( k \) of the following form

\[
X_t = \alpha + \sum_{i=1}^{k} \beta_i X_{t-i} + \varepsilon_t, \tag{28}
\]

where \( \varepsilon_t \) is the uncorrelated but identically distributed random variables with mean zero and variance \( \sigma^2 \).
(ii) $\alpha, \beta_1, \ldots, \beta_k$ are parameters to be estimated from the data.

Now, it may happen that some of the parameters of (28) are identically zero but unless one knows them a priori, it may be quite difficult to obtain its correct form of (28). Further, the selection of $k$ also may be equally difficult. If a small value of $k$ is chosen, some important lagged variables may be omitted while if $k$ is large, the model may be overparametrized. One way to overcome these difficulties is to employ the technique of Stepwise Autoregression which is described in the following steps:

(i) **Step 1**: To start with, we decide the maximum lagged variable $X_{t-p}$ ($p \geq 1$) which is to be included in the regression (28). Having decided the value of $p$, we select that lagged variable $X_{t-j}$ for which the absolute value of the correlation between $X_t$ and $X_{t-r}$, $r = 1, \ldots, p$ is maximum. Now the regression equation takes the form

$$X_t = \alpha + \beta_1 X_{t-j} + \epsilon_t. \quad (29)$$

Under the assumption of normality of $\epsilon_t$'s, we can estimate and test the significance of $\alpha$ and $\beta_1$ with our usual $t/F$ tests. In case, the result is significant, we go to Step 2. Otherwise, we stop computation and accept the model

$$X_t = \alpha + \epsilon_t. \quad (30)$$
(ii) **Step 2**: For convenience, we partition the complete set \( (X_{t-1}, \ldots, X_{t-p}) \) of lagged variables into two sets namely \( S_1 \) and \( S_2 \). The set \( S_1 \) comprises of the lagged variables already included in the regression and the set \( S_2 \) represents remaining lagged variables. Of course, initially \( S_1 \) will contain only the lagged variable \( X_{t-1} \) selected at Step 1 for the first time. At this very step, our main task is to select a new variable from \( S_2 \) for transfer to \( S_1 \). This selection is carried out on the basis of partial autocorrelation between \( X_t \) and \( X_{t-r} \) \( (X_{t-r} \in S_2) \). A variable which has maximum (in absolute sense) partial autocorrelation is selected. For the selection, no test procedure is adopted at this stage.

(iii) **Step 3**: We regress \( X_t \) on all variables included in \( S_1 \) and test for the significance of the regression coefficients. If all the regression coefficients are significant, we go back to Step 2 for the inclusion of another lagged variable in the regression. Otherwise we delete an insignificant variable from \( S_1 \) and put it in \( S_2 \) and go back to Step 2 for the inclusion of a new variable.

(iv) **Step 4**: If at any stage of Step 3, we find that a variable which is included in just previous stage is getting omitted, we stop further search for the addition or deletion of a variable. The set of variables thus obtained
in $S_1$ is final and $X_t$ is regressed on all the variables in $S_1$ taken together to get the regression line.

For fitting an autoregressive model by the Stepwise Autoregression method, generally three decisions are taken.

(i) A value of $k$ for the maximum permissible lag in (28) must be chosen. There is no method by which the value of $k$ can be determined mathematically or statistically. Nevertheless, it depends on the nature of series at hand and amount of data available. It may be possible to have some thumb rule for the value of $k$. In this connection Granger and Newbold (1977) remarked that "for shorter monthly seasonal series, a value of $k = 13$ is generally adequate. For longer monthly seasonal series, a value of $k = 25$ is preferable ". As in our study, most of the series are having observations in the neighbourhood of 100 and they are monthly seasonal, we have taken $k$ as 25 for all the series.

(ii) Though in the case of autoregression analysis, the explanatory variables are stochastic, the usual regression analysis for estimation and testing is valid, at least asymptotically, due to the results of Mann and Wald (1943). In this connection, Newbold and Granger (1974) have considered 4.0 as a critical value of Fisher's $F$ ratio in their study to determine whether a variable is to be included or excluded.
As in our analysis also, we have fitted autoregressive model given by (28) in the same way as Newbold and Granger did, we have also taken 4.0 as the critical F-ratio to test whether a variable is to be included or excluded from the regression.

(iii) A suitable rule has to be framed to stop the computation of the Stepwise Autoregression. In our analysis, we have adopted the following rule: "If at any stage, a variable is included in the regression but the same variable is excluded from the regression in the very next step, stop further search". This is because when a variable is included in the regression on the basis of partial autocorrelation, it may be regarded as most important variable to explain $X_t$ among all the variables in $S_2$. But, when the same variable is excluded from the regression because of the test, it appears that the variable cannot explain $X_t$ significantly. When the most important variable among $S_2$ cannot explain $X_t$ significantly, it is not worth searching further a variable which will explain $X_t$ significantly.

2.5.2 Empirical results and conclusions

Now once the set of lagged variables which are included in the regression equation (28) is determined by the Stepwise Autoregression method, the same equation can be conveniently used for forecasting. But, we are not merely interested in forecast values. Rather, we have to assess the comparative
performance of this method in comparison with the other methods. For this purpose, we have computed the mean percentage square errors upto the lead period 12 for all the twelve series and the results are shown in the table B.8 of the appendix B. It is thought that the Stepwise Autoregression may generate better forecasts if regression is fitted on VX_t rather than X_t. We have, first of all, estimated autoregressive model on the first difference of X_t and then mean percentage square errors are estimated. These are presented in the table B.9 of the appendix B.

Now we like to test whether the Stepwise Autoregression performs better if autoregression is fitted on VX_t rather than on original series X_t. For this purpose, we count the number of times the first procedure (autoregression on VX_t) outperforms the other one on the basis of mean percentage square errors for all series and lead periods combined. From the tables B.8 and B.9, we count the number of times (U) the first procedure outperforms the other and we find the value of U as 91. Under the null hypothesis that both the procedures perform equally well, the distribution of T = (U - 72)/6 is asymptotically normal with mean zero and variance one. Substituting the value of U = 91, we get T = 3.17 which is significant at 5% level. Thus we conclude that the autoregression fitted on the first difference of X_t by the Stepwise
Autoregression performs better than that of autoregression if regression is fitted on original series \( X_t \). However, we find that for some of the series, the autoregression on \( X_t \) has performed better. As we are interested in assessing the performance of the Stepwise Autoregression technique with the other techniques, we have considered, for the assessment purpose, the minimum mean percentage square errors between the two procedures and they are shown in Table 2.8.

At this stage, one may also examine how best a Stepwise Autoregressive model generates forecasts if it is fitted on \( \log X_t \). Accordingly, we have transformed all the twelve series logarithmically and then Stepwise Autoregressive models are fitted. The mean percentage square errors so obtained are shown in Table B.10 of Appendix B. If we compare the mean percentage square errors of Table B.10 with that of Table 2.8, we find that Stepwise Autoregressive model on \( \log X_t \) has not worked well. In the case of only two series namely (i) Rupee Securities and (ii) Index Numbers of Consumer Prices for Industrial Workers, there is an improvement. But this improvement in terms of the reduction of mean percentage square errors is too marginal and it seems that it is not worth the trouble. Hence, we have not considered the logarithmic transformation for any further analysis.
TABLE 2.8 MEAN PERCENTAGE SQUARE ERRORS OF STEPWISE AUTOREGRESSION

<table>
<thead>
<tr>
<th>LEAD PERIOD</th>
<th>SERIES 1</th>
<th>SERIES 2</th>
<th>SERIES 3</th>
<th>SERIES 4</th>
<th>SERIES 5</th>
<th>SERIES 6</th>
<th>SERIES 7</th>
<th>SERIES 8</th>
<th>SERIES 9</th>
<th>SERIES 10</th>
<th>SERIES 11</th>
<th>SERIES 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.00</td>
<td>17.09</td>
<td>9.89</td>
<td>0.71</td>
<td>1.07</td>
<td>1.53</td>
<td>5.20</td>
<td>0.69</td>
<td>12.43</td>
<td>2.12</td>
<td>1.39</td>
<td>281.35</td>
</tr>
<tr>
<td>2</td>
<td>7.41</td>
<td>20.53</td>
<td>30.29</td>
<td>0.91</td>
<td>2.08</td>
<td>1.88</td>
<td>7.72</td>
<td>1.54</td>
<td>10.87</td>
<td>5.47</td>
<td>1.99</td>
<td>207.89</td>
</tr>
<tr>
<td>3</td>
<td>10.45</td>
<td>26.70</td>
<td>55.99</td>
<td>1.38</td>
<td>4.32</td>
<td>2.07</td>
<td>7.26</td>
<td>2.72</td>
<td>10.96</td>
<td>6.22</td>
<td>3.53</td>
<td>319.69</td>
</tr>
<tr>
<td>4</td>
<td>13.91</td>
<td>26.80</td>
<td>88.60</td>
<td>1.85</td>
<td>6.64</td>
<td>3.16</td>
<td>12.80</td>
<td>4.18</td>
<td>21.45</td>
<td>8.25</td>
<td>6.81</td>
<td>293.73</td>
</tr>
<tr>
<td>5</td>
<td>15.90</td>
<td>22.63</td>
<td>121.12</td>
<td>1.30</td>
<td>9.04</td>
<td>4.55</td>
<td>7.16</td>
<td>6.35</td>
<td>21.20</td>
<td>7.22</td>
<td>9.31</td>
<td>252.92</td>
</tr>
<tr>
<td>6</td>
<td>19.04</td>
<td>20.37</td>
<td>170.40</td>
<td>1.40</td>
<td>11.89</td>
<td>5.83</td>
<td>12.64</td>
<td>7.31</td>
<td>23.97</td>
<td>8.51</td>
<td>12.64</td>
<td>235.77</td>
</tr>
<tr>
<td>7</td>
<td>26.00</td>
<td>24.99</td>
<td>224.30</td>
<td>1.46</td>
<td>14.77</td>
<td>7.60</td>
<td>22.79</td>
<td>8.88</td>
<td>19.75</td>
<td>11.76</td>
<td>16.80</td>
<td>269.60</td>
</tr>
<tr>
<td>8</td>
<td>39.74</td>
<td>27.89</td>
<td>258.95</td>
<td>1.74</td>
<td>17.65</td>
<td>10.16</td>
<td>24.25</td>
<td>11.12</td>
<td>30.62</td>
<td>15.27</td>
<td>17.64</td>
<td>313.53</td>
</tr>
<tr>
<td>9</td>
<td>54.62</td>
<td>42.25</td>
<td>259.79</td>
<td>2.10</td>
<td>21.91</td>
<td>11.68</td>
<td>30.91</td>
<td>13.28</td>
<td>23.17</td>
<td>21.14</td>
<td>16.04</td>
<td>296.38</td>
</tr>
<tr>
<td>10</td>
<td>72.45</td>
<td>30.84</td>
<td>311.17</td>
<td>2.98</td>
<td>27.23</td>
<td>14.76</td>
<td>27.83</td>
<td>16.71</td>
<td>32.93</td>
<td>33.21</td>
<td>15.69</td>
<td>431.57</td>
</tr>
<tr>
<td>11</td>
<td>86.94</td>
<td>33.39</td>
<td>405.14</td>
<td>3.65</td>
<td>32.51</td>
<td>17.67</td>
<td>29.16</td>
<td>21.29</td>
<td>22.68</td>
<td>39.56</td>
<td>13.24</td>
<td>300.56</td>
</tr>
<tr>
<td>12</td>
<td>103.11</td>
<td>46.47</td>
<td>500.48</td>
<td>5.02</td>
<td>39.87</td>
<td>18.95</td>
<td>52.18</td>
<td>35.81</td>
<td>24.90</td>
<td>45.86</td>
<td>14.34</td>
<td>208.38</td>
</tr>
</tbody>
</table>