CHAPTER-VI
FORECASTING OF
PUBLIC EXPENDITURE IN MANIPUR
Chapter : VI

Forecasting of Public Expenditure in Manipur

The following approaches of economic forecasting based on time series data are considered for the purpose of forecasting of public expenditure.

1. Box-Jenkin’s Methodology,
2. Almond’s PDL model,
3. Trend model and
4. Vector Autoregressions (VARs) models.

6.1 Box-Jenkin’s Methodology

ARIMA Theory[1]: A univariate time series model described the behaviour of a variable in terms of its own past values and some disturbances.

The general expression is:

\[ Y_t = f(Y_{t-1}, Y_{t-2}, \ldots, u_t) \]  \hspace{1cm} 6.1.1

To make equation 6.1 operational, it must be specified three things:

1. functional form,
2. number of lags and
3. structure for disturbance term.

ARIMA models are generalisation of the simple AR(Autoregressive) model that uses three tools for modelling the serial correlation in the disturbance.

1. The first tool is the autoregressive(AR) term. The AR(1) uses only the first order term but in general it may have higher order AR terms. Each AR term corresponds to the use of the lagged values of the residuals in the forecasting equation for the conditional residual. An autoregressive model of order \( p \), AR(\( p \)) has the form:

\[ Y_t = \mu + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \ldots + \alpha_p Y_{t-p} + \varepsilon_t \]  \hspace{1cm} 6.1.2

2. The second tool is the integration order term. Each integration order corresponds to differencing the series being forecast. A first order integrated component means that the forecasting model is designed for the first difference of the original series. A second order component corresponds to using second order difference and so on.

3. The third tool is the Moving average (MA) term. A moving average (MA) forecasting model uses lagged values of the forecast error to improve the current forecast. A first order moving average MA(1) term uses the most recent forecast error, a second order moving average MA(2) uses the forecast error from the two most recent periods and so on. MA of the order q, MA(q) has the form:

\[ u_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} \]  

The autoregressive and moving average of the dth order difference variable can be combined to form an ARIMA(p, d, q) specification.

\[ y_t = \mu + \gamma_1 y_{t-1} + \gamma_2 y_{t-2} + \ldots + \gamma_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} \]

where \[ y_t = (1 - L^d) Y_t \]

This methodology is known as Box-Jenkin's methodology.

In this specification:

p = order of autoregressive term

d = order of difference to get an stationary time series

The inclusion of non-stochastic part -

\[ \beta^T X_t = \beta_1 X_{rt} + \beta_2 X_{rt} + \ldots + \beta_k X_{kt} \]

to the equation (6.4) yields the ARMAX(p, d, q) have the following specification:

\[ y_t = \mu + \hat{\beta}^T X_t + \gamma_1 y_{t-1} + \gamma_2 y_{t-2} + \ldots + \gamma_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} \]

where \[ y_t = (1 - L^d) Y_t \] and \( L \) is the lag operator that is \( Y_t \) is the original series differenced d times, if this is desired. More compactly, the specification (6.1.5) can be written as:
Chapter: VI Forecasting of Public Expenditures

\[ \Phi(L)\delta(L)^dY_t = \mu + \beta \hat{\chi}_t + \Theta(L) \varepsilon_t \]

where \( \Phi \) and \( \Theta \) are polynomials in the lag operator \( L \) and \( \delta(L) = (1 - L) \). The non-stochastic part \( \beta \hat{\chi}_t \) is optional. Without this term the ARMAX(p, d, q) model becomes ARIMA(p, d, q) model.

This Box-Jenkin’s methodology of time series modelling and forecasting consists of four steps:

1. Identification: The first step of this Box-Jenkin’s methodology is to find the appropriate value of \( p, d \) and \( q \). For this purpose, Correlogram and Partial correlogram are used as tools of identification of value of \( p, d \), and \( q \). LM serial correlation test is also used to identify this values.

2. Estimation: Having identified the appropriate value of \( p, d \) and \( q \), the next step is to estimate the parameters of the autoregressive and moving average terms included in the model.

3. Diagnostic Checking: Having chosen the particular ARIMA model and having estimated its parameters, the next step is to test whether the chosen model fits the data reasonably well or not. So, this methodology is an iterative process.

The testing of resultant equation can be done in two ways;

i. One set of test can be applied to the estimated coefficients of the model by using the concept of significance of the coefficients of the included variables. It can also test the effect of adding one or more variables to the specification.

ii. Another set of test can be applied to the estimated residual of the estimated model, which also provide an important information for testing. If the residual is white noise then the model selected fits the data very well.
4. Forecasting: One of the most important reasons for the popularity of the ARIMA methodology is its success in forecasting. In many cases, the forecast obtained by this method are more reliable than those obtained from the traditional econometric modelling.

Through the Box-Jenkins methodology, it was found that TEM series is non-stationary, whereas the first difference of this series is stationary. The transformed variable $d(\text{TEM})$ as defined below, turned out to be stationary.

$$d(\text{TEM})_t = \text{TEM}_t - \text{TEM}_{t-1}$$

6.1.6

The autocorrelation and partial correlation up to lag 16 are given in table 6.1. For $d(\text{TEM})$ series, from autocorrelation function (ACF) and partial autocorrelation function (PACF) from table 6.1, it is observed that autocorrelation and partial autocorrelation of $d(\text{TEM})$ are found to be spiked at lag 1 and 3. Thus, AR of the order three and MA of the order one were expected to be significant.

<table>
<thead>
<tr>
<th>Autocorrelation</th>
<th>Partial Correlation</th>
<th>lag</th>
<th>AC</th>
<th>PAC</th>
<th>Q-Stat</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;***&quot;</td>
<td>&quot;***&quot;</td>
<td>1</td>
<td>-0.357</td>
<td>-0.357</td>
<td>4.3342</td>
<td>0.037</td>
</tr>
<tr>
<td>.</td>
<td>&quot;*&quot;</td>
<td>2</td>
<td>0.052</td>
<td>-0.086</td>
<td>4.4297</td>
<td>0.109</td>
</tr>
<tr>
<td>.</td>
<td>&quot;*&quot;</td>
<td>3</td>
<td>0.231</td>
<td>0.255</td>
<td>6.3835</td>
<td>0.094</td>
</tr>
<tr>
<td>.</td>
<td>&quot;*&quot;</td>
<td>4</td>
<td>-0.129</td>
<td>0.056</td>
<td>7.0149</td>
<td>0.135</td>
</tr>
<tr>
<td>.</td>
<td>&quot;*&quot;</td>
<td>5</td>
<td>0.111</td>
<td>0.081</td>
<td>7.5030</td>
<td>0.186</td>
</tr>
<tr>
<td>.</td>
<td>&quot;*&quot;</td>
<td>6</td>
<td>0.040</td>
<td>0.053</td>
<td>7.5689</td>
<td>0.271</td>
</tr>
<tr>
<td>.</td>
<td>&quot;*&quot;</td>
<td>7</td>
<td>-0.043</td>
<td>0.001</td>
<td>7.6495</td>
<td>0.365</td>
</tr>
<tr>
<td>.</td>
<td>&quot;*&quot;</td>
<td>8</td>
<td>-0.086</td>
<td>-0.182</td>
<td>7.9798</td>
<td>0.435</td>
</tr>
<tr>
<td>.</td>
<td>&quot;*&quot;</td>
<td>9</td>
<td>-0.025</td>
<td>-0.175</td>
<td>8.0087</td>
<td>0.533</td>
</tr>
<tr>
<td>.</td>
<td>&quot;*&quot;</td>
<td>10</td>
<td>-0.054</td>
<td>-0.130</td>
<td>8.1514</td>
<td>0.614</td>
</tr>
<tr>
<td>.</td>
<td>&quot;*&quot;</td>
<td>11</td>
<td>0.121</td>
<td>0.170</td>
<td>8.9036</td>
<td>0.631</td>
</tr>
<tr>
<td>.</td>
<td>&quot;*&quot;</td>
<td>12</td>
<td>-0.281</td>
<td>-0.172</td>
<td>13.142</td>
<td>0.359</td>
</tr>
<tr>
<td>.</td>
<td>&quot;*&quot;</td>
<td>13</td>
<td>0.185</td>
<td>0.102</td>
<td>15.039</td>
<td>0.302</td>
</tr>
<tr>
<td>.</td>
<td>&quot;*&quot;</td>
<td>14</td>
<td>-0.097</td>
<td>-0.020</td>
<td>15.665</td>
<td>0.334</td>
</tr>
<tr>
<td>.</td>
<td>&quot;*&quot;</td>
<td>15</td>
<td>-0.121</td>
<td>-0.055</td>
<td>16.694</td>
<td>0.344</td>
</tr>
<tr>
<td>.</td>
<td>&quot;*&quot;</td>
<td>16</td>
<td>0.171</td>
<td>-0.015</td>
<td>18.586</td>
<td>0.290</td>
</tr>
</tbody>
</table>

Note: Area between the two dots in the above table indicates the 95% confidence interval.
But some competing models viz, ARIMA(1,1,1), ARIMA(1,1,3) and ARIMA(3,1,1) were estimated. The model selection statistic such as R-bar square(Adj R²), Akaike Information Criterion(AIC) and Schwarz Criterion(SC) are given for these competing models are reported in table 6.2. The best model is chosen on the basis of these statistics.

Table: 6.2 Values of Adj-R², AIC and SC of the competing ARIMA models.

<table>
<thead>
<tr>
<th>Models</th>
<th>adj-R²</th>
<th>AIC</th>
<th>SC</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1,1,1)</td>
<td>0.07</td>
<td>24.01</td>
<td>24.16</td>
</tr>
<tr>
<td>(1,1,3)</td>
<td>0.15</td>
<td>23.94</td>
<td>24.07</td>
</tr>
<tr>
<td>(3,1,1)*</td>
<td>0.15</td>
<td>24.00</td>
<td>24.15</td>
</tr>
</tbody>
</table>

Comparing adj-R², AIC and SC, the **ARIMA(1,1,3)** model is selected for the purpose of forecasting of TEM series. Besides, the residual series of the model **ARIMA(1,1,3)** is also found to be stationary on the basis of correlogram. The estimated form of equation for the model **ARIMA(1,1,3)** is presented below:

\[
d(TEM)_t = 16563 + U_t \\
(2.25) \\
(1 - 0.385 L) = (1 - 0.397 L) e_t \\
(1.72) \quad (-1.87) \\
\text{Adj } R^2 = 0.153 \quad \text{DW} = 2.09 \quad \text{F-stat} = 3.43 \quad \text{level of significance} = 5% \\
\]

which is equivalent to

\[
d(TEM)_t = 10269 + 0.385 d(TEM)_{t-1} - et + 0.397 e_{t-1} \\
(6.1.8)
\]

where, \(e_t\) = actual TEM - fitted TEM.

* indicates the best model out of the competing models
In the above equations, all variables are denominated in lakhs of rupees, numbers in parenthesis are corresponding t-values, \( \text{adj-R}^2 \) is the coefficient of determination and DW is the Durbin - Watson statistic. All \textbf{t-ratios} are significant at 5 percent and less level of significance. DW rules out the serial correlation problem and \( \text{adj-R}^2 \) value is low particularly for the first difference variable. The level of significance is the level of significance for rejecting the overall significance of the estimated parameters. From the estimated levels of significance, we see that all the estimated parameters are simultaneously zero is rejected at 5% level of significance.

In order to verify the appropriateness of the model, ACF and PACF coefficients of residual for various lags were obtained and the Q-statistic was applied to the group of first 16 lags' autocorrelation coefficients. None of this individual coefficient as well as group of coefficients were found to be significant. Another test for goodness of fit of the model would be to examine the accuracy of within sample period forecasts. This was attempted on the basis of four quantitative statistics viz; Root mean square error(RMSE), Mean absolute error(MAE), Mean absolute percentage error(MAPE) and Theil’s Inequality statistic(Theil \( U_2 \)). RMSE is the measure of deviation of the forecasted variable from its actual time path. This statistic depends on the scale of the data. MAE measures the mean error in absolute value. This statistic is also depends on scale of the data. MAPE measures the percentage of mean absolute error, which is a scale free statistic. Theil \( U_2 \) measures the proportion of inequality. It is a relative measure, which compares the forecasting accuracy of the method in relation to that of forecasts of variable(TEM) in period \( t+1 \). It is also a scale free statistic. If Theil \( U_2 =0 \), then it means actual time series is exactly equal to fitted time series. All these errors[2] are rather small and this establishes suitability of ARIMA model for forecasting public expenditure.

---

2. The term 'error' here refers to RMSE, MAE, MAPE and Theil \( U_2 \). For calculation, see reference on page-249
Chapter: VI  Forecasting of Public Expenditures

The forecasts of the public expenditure (TEMF) were obtained through using
the equation (6.1.1) presented above and estimated the forecast values as follows:

$$\text{TEMF}_t = \text{fitted}(d(\text{TEM}))_t + \text{TEM}_{t-1}$$  \hspace{1cm} 6.1.8

The fitted values of the first difference of TEM were obtained through equation(6.1.8) and the actual values of TEM\(_{t-1}\) was used for the periods until which they
were available and the other periods the extrapolation[3] method was resorted to,
under which the estimated values of the lagged variables instead of their true values
were used.

A discussion on this forecast is presented later when all alternative forecasts
are estimated and compared. The forecasts is presented in table 6.8.

6.2 Almond's Polynomial Distributed lag (PDL) Methodology

Under PDL method[4], TEM is taken a dependent variable and TRM as
explanatory variable. TRM is considered as explanatory variable because it is the
only significant variable which determine TEM. The model considered is given
below:

$$\text{TEM}_t = \sum \beta_i \text{TRM}_{t-i} + U_t$$  \hspace{1cm} 6.2.1

in which $\beta_i = a_0 + a_1i + a_2i^2 + a_3i^3$

We assume a polynomial of degree three (3) because the degree of the
polynomial must less than or equal to number of parameters. To get the appropriate
values of number of parameters and degree of polynomial, we considered various
possible models. Their corresponding adj-$R^2$, AIC and SC are shown in table 6.3.

---

3. Extrapolation is a technique of inserting values of variable that lies outside the extreme points.
4. details of PDL model was explained in section 5.6.1.3.
Table: 6.3 Estimated values of Adj $R^2$, AIC and SC of different PDL models.

<table>
<thead>
<tr>
<th>Models</th>
<th>Adj $R^2$</th>
<th>AIC</th>
<th>SC</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDL(3,2)</td>
<td>0.986</td>
<td>22.15</td>
<td>22.34</td>
</tr>
<tr>
<td>PDL(3,3)</td>
<td>0.991</td>
<td>21.66</td>
<td>21.90</td>
</tr>
<tr>
<td>PDL(4,2)</td>
<td>0.982</td>
<td>22.39</td>
<td>22.58</td>
</tr>
<tr>
<td>PDL(4,3)</td>
<td>0.987</td>
<td>22.03</td>
<td>22.27</td>
</tr>
</tbody>
</table>

From the above table 6.3, it is seen that after comparing the values of Adj $R^2$, AIC and SC, PDL(3,3) is the best model. The estimated version of the model PDL(3,3) is given below:

$$\text{TEM} = -4141 - 0.096\text{PDL}01 - 0.05\text{PDL}02 + 0.69\text{PDL}03 - 0.32\text{PDL}04$$

|   | (-1.19) | (-1.24) | (-0.42) | (5.92) | (-4.22) |

This estimated version is used to forecasts TEM. The forecasted values are reported in table: 6.8.

6.3 Trend Model

Most of the time series data that we have encountered are not continuous in time, instead they consist of discrete observations made at regular intervals of time. One of the basic characteristics of a time series that can be described is its long run growth pattern. Despite the short run up and down movements, it is possible that the time series might exhibit a clear cut upward trend. If we believe that this upward trend exists and will continue, we can construct a simple model that describes that trend which can be used to forecast[5].

Chapter: VI  Forecasting of Public Expenditures

Under this method, the following specifications are considered for the purpose.

1. **Linear Trend Model**: The simple specification for linear trend model is as follows:

   \[ Y_t = c_1 + c_2 \ t \]  \hspace{1cm} 6.3.1

   In this model, it is assumed that series \( Y_t \) will increase in constant absolute amounts in each time period.

2. **Exponential Growth Curve**: The mathematical form of the exponential growth curve is given below:

   \[ Y_t = Ae^{rt} \]

   \[ \Rightarrow \log(Y_t) = c_1 + c_2 \ t \]  \hspace{1cm} 6.3.2

   where, \( c_1 = \log A, \ c_2 = r \)

   In this model, it is assumed that \( Y_t \) will increase exponentially at the growth rate \( r \).

3. **Autoregressive Trend Model**: The mathematical form of this autoregressive trend model is as follows:

   \[ Y_t = c_1 + c_2 \ Y_{t-1} \]  \hspace{1cm} 6.3.3

   In this model, if \( c_1 = 0 \), in which \( c_2 \) represents the rate of change of series \( Y_t \).

4. **Logarithmic Autoregressive Trend Model**: This model is specified as follows:

   \[ \log(Y_t) = c_1 + c_2 \log(Y_{t-1}) \]  \hspace{1cm} 6.3.4

   In this model, if \( c_1 \) is fixed to be 0, then the value of \( c_2 \) is the compound rate of growth of the \( Y_t \) series.

5. **Quadratic Trend Model**: The mathematical form of the quadratic trend model is given below:

   \[ Y_t = c_1 + c_2 \ t + c_3 \ t^2 \]  \hspace{1cm} 6.3.5

   This quadratic trend model is a simple extension of the linear model and involves adding a term in \( t^2 \). If \( c_2 \) and \( c_3 \) are both positive, \( Y_t \) will always be increasing but
even more rapidly as time goes on. If $c_2$ is negative and $c_3$ is positive, $Y_t$ will at first decrease and later increase. If both $c_2$ and $c_3$ are negative, $Y_t$ will always decrease.

The above specified models are estimated using the time series data of total expenditure (TEM) from 1972-73 to 2003-04. The models are estimated using OLS and their selected statistics are reported in table 6.4.

Table 6.4 The estimated values of Adj-$R^2$, AIC, SC and nature of estimated residuals.

<table>
<thead>
<tr>
<th>Models</th>
<th>R-bar square</th>
<th>AIC</th>
<th>SC</th>
<th>Nature of estimated residuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear Trend</td>
<td>0.80</td>
<td>24.80</td>
<td>24.86</td>
<td>non-stationary</td>
</tr>
<tr>
<td>Exponential trend</td>
<td>0.86</td>
<td>0.94</td>
<td>1.03</td>
<td>non-stationary</td>
</tr>
<tr>
<td>Autoregressive Trend</td>
<td>0.90</td>
<td>24.06</td>
<td>24.16</td>
<td>non-stationary</td>
</tr>
<tr>
<td>Logarithmic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Autoregressive Trend*</td>
<td>0.92</td>
<td>0.18</td>
<td>0.29</td>
<td>stationary</td>
</tr>
<tr>
<td>Quadratic Trend</td>
<td>0.88</td>
<td>24.26</td>
<td>24.40</td>
<td>non-stationary</td>
</tr>
</tbody>
</table>

From table 6.4, comparing the model selection statistics, logarithmic autoregressive trend model given by the specification (6.3.4) is the best model in the sense that it has got maximum Adj-$R^2$ value, minimum values of AIC and SC and further it is noticed that the estimated residuals series is also found to be stationary. The estimated version of the equation of model (6.3.4) is reported below:

\[
\text{Log}(\text{TEM}_t) = 1.415 + 0.89 \text{ Log}(\text{TEM}_{t-1})
\]

(6.3.6)

\[
R^2 = 0.924 \quad \text{Adj-}R^2 = 0.921
\]

\[
\text{DW} = 2.34 \quad \text{AIC} = 0.177 \quad \text{SC} = 0.27
\]

The forecasts of this model is reported in table 6.8.
6.4 Vector Autoregressions (VARs) Model

The Vector Autoregressions (VARs) is commonly used for forecasting system of interrelated time series and for analysis of the dynamic impact of random disturbance on the system of variables. The VAR approach model every endogenous variables in the system is taken as a function of the lagged endogenous variables and exogenous variables in the system.

The econometric form of the VAR[6] is

\[ Y_t = A_1 Y_{t-1} + A_2 Y_{t-2} + \ldots + A_p Y_{t-p} + B X_t + \varepsilon_t \]  

where \( Y_t \) is a \( k \) vector of endogenous variables, \( X_t \) is a \( d \) vector of exogenous variables, \( A_1, A_2, \ldots, A_p \) and \( B \) are matrices of coefficients to be estimated and \( \varepsilon_t \) is a vector of innovations that may be contemporaneously correlated with each other but are uncorrelated with all of the right hand side of variables. Because most of the public expenditure time series data are non-stationary as they contain unit roots in the series, it has spurred the development of the theory of non-stationary time series analysis. Engle and Granger (1987) pointed out that a linear combination of two or more non-stationary time series may be stationary[6]. If such stationary or \( I(0) \) linear combination exists, then the non-stationary time series (with unit root) are said to be Cointegrated. The stationary linear combination is called cointegrating equation and which may be interpreted as a long run equilibrium relationship between the variables. A Vector Error Correction (VEC) model is a restricted VAR model that has cointegrated restriction built into the specification, so that it is designed for use with non-stationary time series that are known to be cointegrated.

In order to forecast the total expenditure (TEM), of Manipur, the total revenue of Manipur (TRM) is considered as explanatory variable because it is a significant determinant of public expenditure (TEM).

To adopt the VAR methodology of estimation, the pre requisites are

1. the variables must be stationary or
2. the variables must be cointegrated

The test of stationarity of TEM and TRM series using ADF test and P-P test are given in table 6.5. From the table, we noticed that TEM and TRM are difference stationary.

**Table 6.5 Stationarity test for the TEM and TRM series.**

<table>
<thead>
<tr>
<th>Variables</th>
<th>level</th>
<th>first difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ADF</td>
<td>P-P</td>
</tr>
<tr>
<td>TEM</td>
<td>2.17</td>
<td>2.74</td>
</tr>
<tr>
<td>TRM</td>
<td>-0.76</td>
<td>-1.38</td>
</tr>
</tbody>
</table>

* indicates the rejection of the hypothesis of non-stationarity at 1 percent level of significance.

The next step is to test the existence of cointegration between TEM and TRM. For this purpose, Johansen Cointegration test is used. The test statistics are given in table 6.6.

**Table 6.6 Johansen Cointegration Test for TEM and TRM.**

<table>
<thead>
<tr>
<th>Eigenvalues</th>
<th>LR statistic</th>
<th>Hypothesised no of CEs</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.344283</td>
<td>16.18385</td>
<td>none*</td>
</tr>
<tr>
<td>0.144414</td>
<td>4.367130</td>
<td>at most one*</td>
</tr>
</tbody>
</table>

* indicates the rejection of the hypothesis of no cointegration at 1 percent level of significance. 

note: CEs indicates the cointegration equations.
The normalised cointegration equation reported by the econometric software is given below:

\[ \text{TEM}_t = 9286 + 0.945 \text{TRM}_t \]

From the above table 6.6, it is confirmed that TEM and TRM series are cointegrated in the first difference. This confirms the suitability of the application of VAR model. Various VAR models such as VAR(1,0), VAR(1,1), VAR(1,2), VAR(1,3), VAR(1,4), VAR(1,5), and VAR(1,6) are estimated using TEM and TRM as endogenous variables. Based on the significance of the estimated coefficients the VAR(1,3) is chosen for forecasting purpose. The estimated VAR(1,3) is presented below:

\[
\begin{align*}
\text{d(TEM)}_t &= 32178.23 - 6.6 \text{d(TEM)}_{t-1} - 7.5 \text{d(TEM)}_{t-2} - 4.1 \text{d(TEM)}_{t-3} \\
&\quad + 6.1 \text{d(TRM)}_{t-1} + 7.3 \text{d(TRM)}_{t-2} + 4.2 \text{d(TRM)}_{t-3} \\
&\quad - 6.4.1 \\
&\quad (2.603) \quad (-2.6) \quad (-2.41) \quad (-1.66) \\
&\quad (2.47) \quad (2.43) \quad (1.72) \\
\text{Adj-R}^2 &= 0.23 \quad \text{AIC} = 24.04 \quad \text{SC} = 24.43
\end{align*}
\]

The above estimated equation (6.4.1) is used to forecast the TEM series. The forecasts of TEM series using this method is reported in table 6.8.

For obtaining the ex-ante forecasts through PDL and VAR models, the forecast is based on the explanatory variable TRM for the post sample period. These were obtained by the trend method. Taking the period 1972-73 to 2003-04 as sample period the models are estimated and using the estimated models we are trying to forecast the TEM series for the post sample period 2004-05 to 2008-09 that is a period of 5 years.

In order to verify the appropriateness of the forecasts models, an attempt is made on the basis of four quantitative statistics (RMSE, MAE, MAPE and Theil $U_2$) whose values are reported in table 6.7.
Chapter: VI  Forecasting of Public Expenditures

On the basis of accuracy measures, PDL has smaller RMSE and Theil U₂; on the other hand VAR has smaller MAE and MAPE which are of the same origin.

So, PDL is selected as best model.

Table: 6.7: The estimated RMSE, MAE, MAPE and Theil U₂ of different models[7]

<table>
<thead>
<tr>
<th>Models</th>
<th>RMSE</th>
<th>MAE</th>
<th>MAPE</th>
<th>Theil U₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIMA</td>
<td>55469.25</td>
<td>35445.96</td>
<td>18.09</td>
<td>0.079</td>
</tr>
<tr>
<td>PDL@</td>
<td>11190.52</td>
<td>6693.81</td>
<td>3.027</td>
<td>0.016</td>
</tr>
<tr>
<td>TREND</td>
<td>55671.06</td>
<td>31580.70</td>
<td>16.06</td>
<td>0.084</td>
</tr>
<tr>
<td>VAR</td>
<td>15753.4</td>
<td>568.62</td>
<td>0.006</td>
<td>0.07</td>
</tr>
</tbody>
</table>

@ indicates the best model on the basis of the above given statistics

7. The formula for estimating such evaluating statistics are given by:
   \[
   \text{RMSE} = \left( \frac{1}{n} \sum (y - \hat{y})^2 \right)^{\frac{1}{2}}
   \]
   \[
   \text{MAE} = \frac{1}{n} \sum |y - \hat{y}|
   \]
   \[
   \text{MAPE} = \frac{1}{n} \sum \frac{|y - \hat{y}|}{y}
   \]
   and \[
   \text{Theil U₂} = \left[ \frac{1}{n} \sum (y - \hat{y})^2 \right] \left[ \frac{1}{n} \sum y^2 \right] \left[ \frac{1}{n} \sum (y - \hat{y})^2 \right] + \left[ \frac{1}{n} \sum y \hat{y} \right]
   \]
   where \( n \) = the number of periods to be forecasted.
   \( y \hat{y} = \) the predicted value of \( y \).

Post sample period forecasts for the various alternative models are reported in table 6.8
Chapter: VI  Forecasting of Public Expenditures

Table: 6.8  Forecasts on Total public expenditure of Manipur(TEM) series through alternative methods.

(Rs. in lakhs)

<table>
<thead>
<tr>
<th>Year</th>
<th>ARIMA</th>
<th>PDL</th>
<th>Trend</th>
<th>VAR</th>
<th>Actuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>471238.4</td>
<td>454889.9</td>
<td>442662.9</td>
<td>655006.2</td>
<td>678189.6</td>
</tr>
<tr>
<td>2004-05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-06</td>
<td>681444.5</td>
<td>686817.8</td>
<td>683818.1</td>
<td>735007.2</td>
<td>737892.7</td>
</tr>
<tr>
<td>2006-07</td>
<td>728494.0</td>
<td>710921.9</td>
<td>739505.5</td>
<td>763419.9</td>
<td>761850(BE)</td>
</tr>
<tr>
<td>2007-08</td>
<td>867735.7</td>
<td>881352.7</td>
<td>761758.8</td>
<td>808177.7</td>
<td>NA</td>
</tr>
<tr>
<td>2008-09</td>
<td>870418.5</td>
<td>858470.8</td>
<td>815592.7</td>
<td>880486.3</td>
<td>NA</td>
</tr>
</tbody>
</table>

Figure 6.1  Plot of the different forecasts of total public expenditure by different methods.
6.5 Evaluation of the Alternative Forecasts

All the forecasts are subject to errors and the alternative forecasts on total public expenditure (TEM) of Manipur contained in table 6.8 are also no exception. Further, since the data on actual total expenditure of the forecasts periods are not yet published, there is no way to assess their accuracy for the whole forecasts periods and to support the one set of forecasts over the other. However, certain comparisons and observations can still be made using the actual available values, which are as follows:

1. As expected, all forecasts reveal an increasing size of public expenditure (TEM) in Manipur.

2. Theoretically speaking, no forecasting model has sound theoretical background.

3. On the basis of accuracy measures of ex-post forecasts, Almond’s PDL is the best.

4. On the basis of the line graph, presented in figure 6.1, it is noticed that VAR model forecasting provides the forecasting model in the sense that it is nearest to the actual values.

It is impossible to recommend one set of forecasts over the other on the basis of theory and past records. If this were possible, only one method of forecasting would have been used instead of several models. For the forecasts of VAR and PDL models the data points of the explanatory variables for the period to be forecasted are required. But for the alternative methods ARIMA and logarithmic autoregressive trend models, there is no explanatory variable except the lag values of the dependent variables. Thus, the findings are as follows:

1. If one were to provide a single forecast, then those contained in column 3 of the table 6.8 may be taken as our forecasts for the series TEM in Manipur during 2004-05 to 2008-09.
2. If alternative forecasts would be given, the most conservative forecasts are those contained in column (2) of table 6.8 and which is the most liberal one. Thus the most likely forecasts could be taken as column (3) of the table 6.8 given by the Almond’s PDL model.

From figure 6.1, it is observed that forecasts obtained from the VARs model is closest to the actual lines. It indicates that VARs model gives the best forecast but it is against the theoretical conclusion.

6.6 Conclusion

This chapter tries to forecast the total expenditure in Manipur. The alternative methods used for forecasting public expenditure will have varying results. No method is perfect. As per the results, the total expenditure in Manipur would be as small as Rs. 815592.70 lakhs by trend model and as large as Rs. 880486.30 lakhs by PDL model in 2008-09. On the basis of accuracy measures of ex-post forecasts, the PDL forecasts are the best. But, on the other hand comparing the forecasts with actuals, VAR model is found to be the best. So, our best estimate should be around Rs. 880000.00 lakhs as yielded by the VAR method. It is needless to say that the estimates are derived on the basis of an analysis of historical data and so, if far reaching economic reforms continue or become more vigorous, we may witness the total expenditure figure different from the forecasts. Nevertheless, the magnitude of total expenditure whatever it assumes, would have a significant impact on the level of NSDP, inflation and among other economic variables.