Modeling and Forecasting of Saudi Arabia Revenue and Expenditure Data

A THESIS
SUBMITTED TO THE GRADUATE EDUCATIONAL POLICIES COUNCIL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
for the degree

MASTER OF SCIENCE

By
Hleil Aed A Alrweili

Advisor - Dr. Rahmatullah Imon

Ball State University

Muncie, Indiana

May 2015
Modeling and Forecasting of Saudi Arabia Revenue and Expenditure Data
A THESIS
SUBMITTED TO THE GRADUATE EDUCATIONAL POLICIES COUNCIL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
for the degree
MASTER OF SCIENCE
By
Hleil Aed A Alrweili

Committee Approval:

Committee Chairman

Committee Member

Committee Member

Department Head Approval:

Head of Department

Graduate office Check:

Dean of Graduate School

Ball State University
Muncie, Indiana
May 2015
I would like to express the deepest appreciation to my committee chair Professor Rahmatullah Imon who has shown the attitude and the substance of a genius: he continually and persuasively conveyed a spirit of adventure in regard to research and scholarship, and an excitement in regard to teaching. Without his supervision and constant help this thesis would not have been possible.

I would like to thank my committee member and the statistic graduate adviser Dr. Munni Begum who's always available to answer my questions and show her supporting for my thesis. I would like to thank Dr. Yayuan Xiao who have been helpful regarding the mathematical part.

A special thanks to the dean of Science College in Northern Border University Professor Saad Alanazi.

Hleil Aed A Alrweili
Table of Contents

CHAPTER 1  
1 INTRODUCTION  
1.1 Saudi Arabia Revenue and Expenditure Data  
1.2 Outline of the Study

CHAPTER 2  
12 MODELING AND FITTING OF DATA USING CLASSICAL, ROBUST AND NONPARAMETRIC REGRESSION METHODS  
2.1 Classical Regression Analysis  
2.2 Robust Regression  
2.3 Nonparametric Regression  
2.4 Fitting of Saudi Arabia Expenditure Revenue Data Using Regression Models

CHAPTER 3  
34 FORECASTING WITH ARIMA MODELS  
3.1 The Box-Jenkins Methodology  
3.2 Stationary and Nonstationary Time Series  
3.3 Test for Significance of a White Noise Autocorrelation Function  
3.4 ARIMA Models  
3.5 Estimation and Specification of ARIMA Models  
3.6 Diagnostic Checking  
3.7 Computing a Forecast  
3.8 Fitting of the ARIMA Model to Saudi Arabia Expenditure Data

CHAPTER 4  
59 EVALUATION OF FORECASTS BY REGRESSION AND TIME SERIES METHODS USING CROSS VALIDATION  
4.1 Cross Validation in Regression and Time Series Models  
4.2 Generation of Forecasts for Saudi Arabia Expenditure Data Using Cross Validation

CHAPTER 5  
66 CONCLUSIONS AND AREAS OF FUTURE RESEARCH  
5.1 Conclusions  
5.2 Areas of Future Research

REFERENCES  
68
List of Tables

Chapter 1

Table 1.1: Revenue and Expenditure Data of Saudi Arabia 2

Chapter 2

Table 2.1: ANOVA Table for a Two Variable Regression Model 15
Table 2.2: Summary Statistics of Nine Different Classical Regression Models 27
Table 2.3: MAPE for Nine Different Regression Fits Using OLS, LOWESS and LMS 33

Chapter 3

Table 3.1: Specification of ARIMA Models 47
Table 3.2: ACF and PACF of Current Expenditure of Saudi Arabia 56
Table 3.3: ACF and PACF of Capital Expenditure of Saudi Arabia 56
Table 3.4: ACF and PACF of Current Expenditure of Saudi Arabia 57
Table 3.5: AR (1) Fits for Current, Capital and Total Expenditure of Saudi Arabia 58

Chapter 4

Table 4.1: Original and Forecast Values for Current Expenditure of Saudi Arabia 62
Table 4.2: Original and Forecast Values for Capital Expenditure of Saudi Arabia 62
Table 4.3: Original and Forecast Values for Total Expenditure of Saudi Arabia 63
Table 4.4: MSPE Values for Current, Capital and Total Expenditure of Saudi Arabia 63
Table 4.5: MSPE Values for Current, Capital and Total Expenditure of Saudi Arabia
List of Figures

Chapter 1

Figure 1.1: Time Series Plot of Oil Revenue in Saudi Arabia 4
Figure 1.2: Time Series Plot of Other Revenue in Saudi Arabia 4
Figure 1.3: Time Series Plot of Total Revenue in Saudi Arabia 5
Figure 1.4: Time Series Plot of Current Expenditure in Saudi Arabia 5
Figure 1.5: Time Series Plot of Capital Expenditure in Saudi Arabia 6
Figure 1.6: Time Series Plot Total Expenditure in Saudi Arabia 6
Figure 1.7: Regression Plot of Current Expenditure versus Oil Revenue for Saudi Arabia 7
Figure 1.8: Regression Plot of Current Expenditure versus Other Revenue for Saudi Arabia 7
Figure 1.9: Regression Plot of Current Expenditure versus Total Revenue for Saudi Arabia 8
Figure 1.10: Regression Plot of Capital Expenditure versus Oil Revenue for Saudi Arabia 8
Figure 1.11: Regression Plot of Capital Expenditure versus Other Revenue for Saudi Arabia 9
Figure 1.12: Regression Plot of Capital Expenditure versus Total Revenue for Saudi Arabia 9
Figure 1.13: Regression Plot of Total Expenditure versus Oil Revenue for Saudi Arabia 10
Figure 1.14: Regression Plot of Total Expenditure versus Other Revenue for Saudi Arabia 10
Figure 1.15: Regression Plot of Total Expenditure versus Total Revenue for Saudi Arabia 11

Chapter 2
Chapter 3
Figure 3.1: ACF and PACF of Current Expenditure Data 55
Figure 3.2: ACF and PACF of Capital Expenditure Data 55
Figure 3.3: ACF and PACF of Total Expenditure Data 55
CHAPTER 1

INTRODUCTION

Revenue and expenditure determine the economic condition of a country. Saudi Arabia has an oil-based economy with strong government controls over major economic activities. Since the oil prices have been high, Saudi Arabia economy is in very good condition. According to recent World Bank report the GDP of Saudi Arabia is 745.27 billion US dollars. Its rank is 19th among world countries. The oil prices control the economy of Saudi Arabia. For example, if the oil prices are high, economy of Saudi Arabia will be in very good condition. According to the Ministry of Finance of Saudi Arabia, Saudi Arabia government spends most of its money on education, health services, municipal services, transport, communication, and agriculture. Especially the education and health services, Saudi Arabia government in last 10 years tends to spend money on education and health services. For example, in 2013, Saudi Arabia spent 45% of its budget on education and 30% on health services.

1.1 Saudi Arabia Revenue and Expenditure Data

In our study, we would like to consider revenue and expenditure data of Saudi Arabia from year 1969 to 2012. This data set is taken from the official website of Saudi Arabian Ministry of Finance: www.mof.gov.sa measured in million Saudi Riyals (1 US Dollar = 3.75 Saudi Riyal). We observe six sets of results: Oil revenue, other revenue, total revenue, current expenditure, capital expenditure, total expenditure. This data is presented in Table 1.1.
Table 1.1: Revenue and Expenditure Data of Saudi Arabia

<table>
<thead>
<tr>
<th>Year</th>
<th>Oil Revenue</th>
<th>Other Revenue</th>
<th>Total Revenue</th>
<th>Current Expenditure</th>
<th>Capital Expenditure</th>
<th>Total Expenditure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1969</td>
<td>5119</td>
<td>549</td>
<td>5668</td>
<td>3853</td>
<td>2175</td>
<td>6028</td>
</tr>
<tr>
<td>1970</td>
<td>7122</td>
<td>818</td>
<td>7940</td>
<td>3989</td>
<td>2304</td>
<td>6293</td>
</tr>
<tr>
<td>1971</td>
<td>9685</td>
<td>1435</td>
<td>11120</td>
<td>4756</td>
<td>3374</td>
<td>8130</td>
</tr>
<tr>
<td>1972</td>
<td>13480</td>
<td>1888</td>
<td>15368</td>
<td>5654</td>
<td>4504</td>
<td>10158</td>
</tr>
<tr>
<td>1973</td>
<td>39285</td>
<td>2420</td>
<td>41705</td>
<td>8470</td>
<td>10125</td>
<td>18595</td>
</tr>
<tr>
<td>1974</td>
<td>94190</td>
<td>5913</td>
<td>100103</td>
<td>15207</td>
<td>19832</td>
<td>35039</td>
</tr>
<tr>
<td>1975</td>
<td>93481</td>
<td>9903</td>
<td>103384</td>
<td>37931</td>
<td>43304</td>
<td>81235</td>
</tr>
<tr>
<td>1976</td>
<td>121191</td>
<td>14766</td>
<td>135957</td>
<td>73621</td>
<td>54652</td>
<td>128273</td>
</tr>
<tr>
<td>1977</td>
<td>114042</td>
<td>16617</td>
<td>130659</td>
<td>71417</td>
<td>66631</td>
<td>138048</td>
</tr>
<tr>
<td>1978</td>
<td>115078</td>
<td>16427</td>
<td>131505</td>
<td>83488</td>
<td>64484</td>
<td>147972</td>
</tr>
<tr>
<td>1979</td>
<td>189295</td>
<td>21901</td>
<td>211196</td>
<td>102447</td>
<td>83277</td>
<td>185724</td>
</tr>
<tr>
<td>1980</td>
<td>319305</td>
<td>28795</td>
<td>348100</td>
<td>132661</td>
<td>104094</td>
<td>236755</td>
</tr>
<tr>
<td>1981</td>
<td>328594</td>
<td>39412</td>
<td>368006</td>
<td>113636</td>
<td>171014</td>
<td>284650</td>
</tr>
<tr>
<td>1982</td>
<td>186006</td>
<td>60176</td>
<td>246182</td>
<td>102248</td>
<td>142664</td>
<td>244912</td>
</tr>
<tr>
<td>1983</td>
<td>145123</td>
<td>61296</td>
<td>206419</td>
<td>124052</td>
<td>106134</td>
<td>230186</td>
</tr>
<tr>
<td>1984</td>
<td>121348</td>
<td>50161</td>
<td>171509</td>
<td>121696</td>
<td>94667</td>
<td>216363</td>
</tr>
<tr>
<td>1985</td>
<td>88425</td>
<td>45140</td>
<td>133565</td>
<td>119865</td>
<td>64139</td>
<td>184004</td>
</tr>
<tr>
<td>1986</td>
<td>42464</td>
<td>34034</td>
<td>76498</td>
<td>98894</td>
<td>38528</td>
<td>137422</td>
</tr>
</tbody>
</table>
At first we consider time series plot of all the six variables we consider in our study.

<table>
<thead>
<tr>
<th>Year</th>
<th>Value1</th>
<th>Value2</th>
<th>Value3</th>
<th>Value4</th>
<th>Value5</th>
<th>Value6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1987</td>
<td>67405</td>
<td>36406</td>
<td>103811</td>
<td>134419</td>
<td>50500</td>
<td>184919</td>
</tr>
<tr>
<td>1988</td>
<td>48400</td>
<td>36200</td>
<td>84600</td>
<td>116283</td>
<td>24573</td>
<td>140856</td>
</tr>
<tr>
<td>1989</td>
<td>75900</td>
<td>38700</td>
<td>114600</td>
<td>118303</td>
<td>36567</td>
<td>154870</td>
</tr>
<tr>
<td>1990</td>
<td>96800</td>
<td>39600</td>
<td>136400</td>
<td>114300</td>
<td>45215</td>
<td>143515</td>
</tr>
<tr>
<td>1991</td>
<td>149497</td>
<td>30742</td>
<td>180239</td>
<td>262905</td>
<td>65005</td>
<td>327910</td>
</tr>
<tr>
<td>1992</td>
<td>128790</td>
<td>40857</td>
<td>169647</td>
<td>162350</td>
<td>76637</td>
<td>238987</td>
</tr>
<tr>
<td>1993</td>
<td>105976</td>
<td>35469</td>
<td>141445</td>
<td>184878</td>
<td>3012</td>
<td>187890</td>
</tr>
<tr>
<td>1994</td>
<td>95505</td>
<td>33486</td>
<td>128991</td>
<td>161380</td>
<td>2396</td>
<td>163776</td>
</tr>
<tr>
<td>1995</td>
<td>105728</td>
<td>40772</td>
<td>146500</td>
<td>148776</td>
<td>25167</td>
<td>173943</td>
</tr>
<tr>
<td>1996</td>
<td>135982</td>
<td>43103</td>
<td>179085</td>
<td>171258</td>
<td>26859</td>
<td>198117</td>
</tr>
<tr>
<td>1997</td>
<td>159985</td>
<td>45515</td>
<td>205500</td>
<td>218880</td>
<td>2392</td>
<td>221272</td>
</tr>
<tr>
<td>1998</td>
<td>79998</td>
<td>61610</td>
<td>141608</td>
<td>171163</td>
<td>18897</td>
<td>190060</td>
</tr>
<tr>
<td>1999</td>
<td>104447</td>
<td>43007</td>
<td>147454</td>
<td>167195</td>
<td>16646</td>
<td>183841</td>
</tr>
<tr>
<td>2000</td>
<td>214424</td>
<td>43641</td>
<td>258065</td>
<td>216958</td>
<td>18364</td>
<td>235322</td>
</tr>
<tr>
<td>2001</td>
<td>183915</td>
<td>44244</td>
<td>228159</td>
<td>223508</td>
<td>31632</td>
<td>255140</td>
</tr>
<tr>
<td>2002</td>
<td>166100</td>
<td>46900</td>
<td>213000</td>
<td>203500</td>
<td>30000</td>
<td>233500</td>
</tr>
<tr>
<td>2003</td>
<td>231000</td>
<td>62000</td>
<td>293000</td>
<td>223530</td>
<td>33470</td>
<td>257000</td>
</tr>
<tr>
<td>2004</td>
<td>330000</td>
<td>62291</td>
<td>392291</td>
<td>247649</td>
<td>37551</td>
<td>285200</td>
</tr>
</tbody>
</table>
Figure 1.1: Time Series Plot of Oil Revenue in Saudi Arabia

Figure 1.2: Time Series Plot of Other Revenue in Saudi Arabia
Figure 1.3: Time Series Plot of Total Revenue in Saudi Arabia

Figure 1.4: Time Series Plot of Current Expenditure in Saudi Arabia
The above plots show an increasing trend of revenue and expenditure of Saudi Arabia over the years. The growth have been steady for each variable except capital expenditure. It showed decreasing pattern from 1980-2000, but since then it shows increasing pattern. For all variables we observe a huge growth in recent years.

Now we present several scatterplots to see the relationship between expenditure and revenue.
Figure 1.7: Regression Plot of Current Expenditure versus Oil Revenue for Saudi Arabia

Figure 1.8: Regression Plot of Current Expenditure versus Other Revenue for Saudi Arabia
Figure 1.9: Regression Plot of Current Expenditure versus Total Revenue for Saudi Arabia

Figure 1.10: Regression Plot of Capital Expenditure versus Oil Revenue for Saudi Arabia
Figure 1.11: Regression Plot of Capital Expenditure versus Other Revenue for Saudi Arabia

Figure 1.12: Regression Plot of Capital Expenditure versus Total Revenue for Saudi Arabia
Figure 1.13: Regression Plot of Total Expenditure versus Oil Revenue for Saudi Arabia

Figure 1.14: Regression Plot of Total Expenditure versus Other Revenue for Saudi Arabia
Regression plots shown in Figures 1.7 – 1.15 show that linear relationship exist between expenditure and revenue data however few of them (e.g., Figures 1.10 – 1.12) need further investigation.

1.2 Outline of the Study

We organize this thesis in the following way. In chapter 2, we introduce different classical, robust and nonparametric regression methodologies for fitting the data. We employ these methods to Saudi Arabia revenue and expenditure data. Fitting and forecasting of data using ARIMA models is discussed in chapter 3. In chapter 4, we offer a comparison between regression and time series methods to investigate which methods can forecast future observations in a better way. The whole experiment is done using the cross validation technique.
CHAPTER 2
MODELING AND FITTING OF DATA USING CLASSICAL, ROBUST AND NONPARAMETRIC REGRESSION METHODS

In this chapter at first we discuss different classical, robust and nonparametric methods commonly used in regression and later use them for the revenue and expenditure data of Saudi Arabia.

2.1 Classical Regression Analysis

Regression is probably the most popular and commonly used statistical method in all branches of knowledge. It is a conceptually simple method for investigating functional relationships among variables. The user of regression analysis attempts to discern the relationship between a dependent (response) variable and one or more independent (explanatory/predictor/regressor) variables. Regression can be used to predict the value of a response variable from knowledge of the values of one or more explanatory variables.

To describe this situation formally, we define a simple linear regression model

\[ Y_i = \alpha + \beta X_i + u_i \]  

(2.1)
where $Y$ is a random variable, $X$ is a fixed (nonstochastic) variable and $u$ is a random error term whose value is based on an underlying probability distribution (usually normal). For every value $X$ there exists a probability distribution of $u$ and therefore a probability distribution of the $Y$'s.

We can now fully specify the two-variable linear regression model as given in (2.1) by listing its important assumptions.

1. The relationship between $Y$ and $X$ is linear.

2. The $X$’s are nonstochastic variables whose values are fixed.

3. The error has zero expected values: $E(u) = 0$

4. The error term has constant variance for all observations, i.e.,

$$E(u_i^2) = \sigma^2, \quad i = 1, 2, \ldots, n.$$  

5. The random variables $u_i$ are statistically independent. Thus,

$$E(u_i u_j) = 0, \quad \text{for all } i \neq j.$$  

6. The error term is normally distributed.

### 2.1.1 Tests of Regression Coefficients, Analysis of Variance and Goodness of Fit

We often like to establish that the explanatory variable $X$ has a significant effect on $Y$, that the coefficient of $X$ (which is $\beta$) is significant. In this situation the null hypothesis is constructed in a way that makes its rejection possible. We begin with a null hypothesis, which usually states that a certain effect is not present, i.e., $\beta = 0$. We estimate $\hat{\beta}$ and its standard error from the data and compute the statistic
Residuals can provide a useful measure of the fit between the estimated regression line and the data. A good regression equation is one, which helps explain a large proportion of the variance of $Y$. Large residuals imply a poor fit, while small residuals imply a good fit. The problem with using the residuals as a measure of goodness of fit is that its value depends on the units of the dependent variable. For this we require a unit-free measure for the goodness of fit. Thus the total variation of $Y$ (usually known as total sum of squares TSS) can be decomposed into two parts: the residual variation of $Y$ (error sum of squares ESS) and the explained variation of $Y$ (regression sum of squares RSS). To standardize, we divide both sides of the above equation by the TSS to get

$$1 = \frac{\text{ESS}}{\text{TSS}} + \frac{\text{RSS}}{\text{TSS}}$$

We define the $R^2$ (or $R$-squared) of the regression equation as

$$R^2 = \frac{\text{RSS}}{\text{TSS}} = 1 - \frac{\text{ESS}}{\text{TSS}}$$

$R^2$ is the proportion of the total variation in $Y$ explained by the regression of $Y$ on $X$. It is easy to show that $R^2$ ranges in value between 0 and 1. But it is only a descriptive statistics. Roughly speaking, we associate a high value of $R^2$ (close to 1) with a good fit of the model by the regression line and associate a low value of $R^2$ (close to 0) with a poor fit. How large must $R^2$ be for the regression equation to be useful? That depends upon the area of application. If we could develop a regression equation to predict the stock market, we would be ecstatic if $R^2 = 0.50$. On the other hand, if we were predicting death in road accident, we would want the prediction equation to have strong predictive ability, since the consequences of poor prediction could be quite serious.
It is often useful to summarize the decomposition of the variation in \( Y \) in terms of an analysis of variance (ANOVA). In such a case the total explained and unexplained variations in \( Y \) are converted into variances by dividing by the appropriate degrees of freedom. This helps us to develop a formal procedure to test the goodness of fit by the regression line. Initially we set the null hypothesis that the fit is not good. In other words, our hypothesis is that the overall regression is not significant in a sense that the explanatory variable is not able to explain the response variable in a satisfactory way.

**ANOVA Table for a Two Variable Regression Model**

<table>
<thead>
<tr>
<th>Components</th>
<th>Sum of Squares</th>
<th>Degrees of freedom</th>
<th>Mean SS</th>
<th>( F ) statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>RSS</td>
<td>1</td>
<td>( \text{RSS}/1 = \text{RMS} )</td>
<td>( \text{RMS}/\text{EMS} \sim F_{1,n-2} )</td>
</tr>
<tr>
<td>Error</td>
<td>ESS</td>
<td>( n - 2 )</td>
<td>( \text{ESS}/(n - 2) = \text{EMS} )</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>TSS</td>
<td>( n - 1 )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

We compute the ratio RMS / EMS which follows an \( F \) distribution with 1 and \( n - 2 \) degrees of freedom (which is also a square of a \( t \) distribution with \( n - 2 \) d. f.). If the calculated value of this ratio is greater than \( F_{1,n-2,0.05} \), we reject the null hypothesis and conclude that the overall regression is significant at the 5% level of significance.

**2.1.2. Regression Diagnostics and Tests for Normality**

Diagnostics are designed to find problems with the assumptions of any statistical procedure. In diagnostic approach we estimate the parameters (in regression fit the model) by the classical method (the OLS) and then see whether there is any violation of assumptions and/or irregularity.
in the results regarding the six standard assumptions mentioned at the beginning of this section. But among them the assumption of normality is the most important assumption.

The normality assumption means the errors are distributed as normal. The simplest graphical display for checking normality in regression analysis is the normal probability plot. This method is based in the fact that if the ordered residuals are plotted against their cumulative probabilities on normal probability paper, the resulting points should lie approximately on a straight line. An excellent review of different analytical tests for normality is available in Imon (2003). A test based on the correlation of true observations and the expectation of normalized order statistics is known as the Shapiro – Wilk test. A test based on empirical distribution function is known as Anderson – Darling test. It is often very useful to test whether a given data set approximates a normal distribution. This can be evaluated informally by checking to see whether the mean and the median are nearly equal, whether the skewness is approximately zero, and whether the kurtosis is close to 3. A more formal test for normality is given by the Jarque – Bera statistic:

\[
JB = \frac{n}{6} \left[ S^2 + (K - 3)^2 / 4 \right]
\] (2.4)

Imon (2003) suggests a slight adjustment to the JB statistic to make it more suitable for the regression problems. His proposed statistic based on rescaled moments (RM) of ordinary least squares residuals is defined as

\[
RM = \frac{n c^3}{6} \left[ S^2 + c(K - 3)^2 / 4 \right]
\] (2.5)

where \(c = n/(n - k)\), \(k\) is the number of independent variables in a regression model. Both the JB and the RM statistic follow a chi square distribution with 2 degrees of freedom. If the values of these statistics are greater than the critical value of the chi square, we reject the null hypothesis of normality.
2.2 Robust Regression

Robustness is now playing a key role in time series. According to Kadane (1984) ‘Robustness is a fundamental issue for all statistical analyses; in fact it might be argued that robustness is the subject of statistics.’ The term robustness signifies insensitivity to small deviations from the assumption. That means a robust procedure is nearly as efficient as the classical procedure when classical assumptions hold strictly but is considerably more efficient over all when there is a small departure from them. The main application of robust techniques in a time series problem is to try to devise estimators that are not strongly affected by outliers or departures from the assumed model. In time series, robust techniques grew up in parallel to diagnostics (Hampel et al., 1986) and initially they were used to estimate parameters and to construct confidence intervals in such a way that outliers or departures from the assumptions do not affect them.

A large body of literature is now available [Rousseeuw and Leroy (1987), Maronna, Martin, and Yohai (2006), Hadi, Imon and Werner (2009)] for robust techniques that are readily applicable in linear regression or in time series.

2.2.1. L – estimator

A first step toward a more robust time series estimator was the consideration of least absolute values estimator (often referred to as L – estimator). In the OLS method, outliers may have a very large influence since the estimated parameters are estimated by minimizing the sum of squared residuals

\[ \sum_{i=1}^{n} u_i^2 \]
L estimates are then considered to be less sensitive since they are determined by minimizing the sum of absolute residuals

\[ \sum_{t=1}^{n} |u_t| \]

The L estimator was first introduced by Edgeworth in 1887 who argued that the OLS method is over influenced by outliers, but because of computational difficulties it was not popular and not much used until quite recently. Sometimes we consider the \( L \)–estimator as a special case of \( L_p \)-norm estimator in the literature where the estimators are obtained by minimizing

\[ \sum_{t=1}^{n} |u_t|^p \]

The \( L_1 \)-norm estimator is the OLS, while the \( L_2 \) - norm estimator is the \( L \) – estimator. But unfortunately a single erroneous observation (high leverage point) can still totally offset the \( L \)-estimator.

### 2.2.2. Least Median of Squares

Rousseeuw (1984) proposed Least Median of Squares (LMS) method which is a fitting technique less sensitive to outliers than the OLS. In OLS, we estimate parameters by

minimizing the sum of squared residuals \( \sum_{t=1}^{n} u_t^2 \)

which is obviously the same if we

minimize the mean of squared residuals \( \frac{1}{n} \sum_{t=1}^{n} u_t^2 \).

Sample means are sensitive to outliers, but medians are not. Hence to make it less sensitive we can replace the mean by a median to obtain median sum of squared residuals

[18]
MSR (\(\hat{\beta}\)) = \text{Median}\{\hat{u}_i^2\} \hspace{1cm} (2.6)

Then the LMS estimate of \(\beta\) is the value that minimizes MSR (\(\hat{\beta}\)). Rousseeuw and Leroy (1987) have shown that LMS estimates are very robust with respect to outliers and have the highest possible 50% breakdown point.

### 2.2.3. Least Trimmed Squares

The least trimmed (sum of) squares (LTS) estimator is proposed by Rousseeuw (1984). In this method we try to estimate \(\beta\) in such a way that

\[
\text{LTS (}\hat{\beta}\text{) = minimize } \sum_{i=1}^{h} \hat{u}_{(i)}^2 \hspace{1cm} (2.7)
\]

Here \(\hat{u}_{(i)}\) is the \(i\)-th ordered residual. For a trimming percentage of \(\alpha\), Rousseeuw and Leroy (1987) suggested choosing the number of observations \(h\) based on which the model is fitted as \(h = [n (1 - \alpha)] + 1\). The advantage of using LTS over LMS is that, in the LMS we always fit the regression line based on roughly 50% of the data, but in the LTS we can control the level of trimming. When we suspect that the data contains nearly 10% outliers, the LTS with 10% trimming will certainly produce better result than the LMS. We can increase the level of trimming if we suspect there are more outliers in the data.

### 2.2.4. M – estimator

Huber (1973) generalized the estimation of parameters by considering a class of estimators, which chooses \(\rho(\bullet)\) to

\[
\text{Minimize } \sum_{i=1}^{n} \rho(\varepsilon_i) = \text{Minimize } \sum_{i=1}^{n} \rho(y_i - x_i^T\beta) \hspace{1cm} (2.8)
\]

where \(\rho(\bullet)\) is a symmetric function less sensitive to outliers than squares. An estimator of this type is called an \(M – \) estimator, where \(M\) stands for maximum likelihood. It is easy to see from (2.8) that the
function $\rho$ is related to the likelihood function for an appropriate choice of an error distribution. For example if the error distribution is normal, then $\rho(z) = z^2 / 2, -\infty < z < \infty$, which also yields the OLS estimator. The $M$ – estimator obtained from (2.8) is not scale invariant. To obtain a scale invariant version of this estimator we solve

$$
(2.9)
$$

In most of the practical applications, the value of $\sigma$ is unknown and it is usually estimated before solving the equation (2.9). A popular choice of $\sigma$ is

$$
\tilde{\sigma} = \text{MAD (normalized)}. \tag{2.10}
$$

To minimize (2.8), we have to equate the first partial derivatives of $\rho$ w.r.t. $\beta$ for $j = 0, 1, \ldots, p$ to zero, yielding a necessary condition for a minimum. This gives a system of $k = p + 1$ equations

$$
\sum_{i=1}^{n} x_{ij} \psi \left( \frac{x_{ij} - \beta_j}{\tilde{\sigma}} \right), 1, \ldots, p \tag{2.11}
$$

where $\psi = \rho'$. In general the $\psi$ function is nonlinear and (2.11) must be solved by iterative methods.

### 2.2.5. $S$ – estimator

Rousseeuw and Yohai (1984) suggested another class of high breakdown estimator based on the minimization of the dispersion of the residuals:

$$
s(\hat{\varepsilon}_1 (\beta), \hat{\varepsilon}_2 (\beta), \ldots, \hat{\varepsilon}_n (\beta))
$$

The above dispersion is defined as the solution of

$$
\frac{1}{n} \sum_{i=1}^{n} \rho(\hat{\varepsilon}_i / s) = K \tag{2.12}
$$

$K$ is often put equal to $E_{\Phi}(\rho)$ where $\Phi$ is the standard normal. The function $\rho$ must satisfy the following conditions: $\rho$ is symmetric and continuously differentiable and $\rho(0) = 0$. The estimator thus obtained is called an $S$ – estimator because it is derived from a scale statistic in an implicit way. In fact $s$ given in the above estimating equation is an $M$ – estimator of scale.
2.2.6. MM – estimator

The MM – estimator was originally proposed by Yohai (1987). The objective was to produce a high breakdown point estimator that maintained good efficiency. The MM – estimator has three stages:

- The initial estimate is an S – estimate, so it has a high breakdown point.
- The second stage computed an M – estimate of the error standard deviation using the residuals from the initial S – estimate.
- The last step is an M – estimate of the parameters using a hard redescending weight function to put a very small (often zero) weight to sufficiently large residuals.

In an extensive performance evaluation of several robust regression estimators, Simpson and Montgomery (1998) report that MM – estimators have high efficiency and work well in most outlier scenarios.

2.3 Nonparametric Regression

In regression we start with a tentative model and then apply diagnostics to see if the model should be modified. Sometimes it may be extremely useful to employ a strictly adaptive approach where the regression equation is determined from the data. To quote Hastie and Tibshirani (1987), ‘... residual ... plots are used to detect departure from linearity and often suggest parametric fixes. An attractive alternative to this indirect approach is to model the regression function nonparametrically and let the data decide on the functional form.’ According to Silverman (1985) ‘... an initial nonparametric estimate may well suggest a suitable parametric model (such as linear regression), but nevertheless will give the data more of a chance to speak for themselves in choosing the model to be fitted.’ We will follow Altman
(1992) to define nonparametric regression as ‘Nonparametric regression is a collection of techniques for fitting a curve when there is a little a priori knowledge about its shape.’

In linear regression we start with a tentative model and then apply diagnostics to see whether this model should be modified, while in non-parametric regression the regression equation is determined from the data. In linear regression we have four basic assumptions. At first we assume that a linear model is appropriate. Then we assume that errors are independent, have a common normal distribution with a zero mean, and have a constant variance. What if we relax the second assumption? If we knew that the errors are nonnormal and they violate other conditions as well, we would not want to use the OLS method and may not use the linear regression techniques.

Now consider a case where the error term is assumed to have a nonnormal distribution, the single regressor is assumed to be fixed, and the relationship between $X$ and $Y$ is linear. What if we used the OLS method, recognizing that it is inappropriate? By the virtue of Gauss-Markov theorem, the OLS estimators are still ‘best’ in the sense of having the smallest variance of any unbiased estimator that are linear functions of $Y$’s. Since the normal theory inference procedures cannot be used, one possible approach would be to have the subsequent inferences based on nonparametric procedures. For example, we could use a nonparametric test for testing $H_0 : \beta = 0$. The test is to compute the Spearman rank correlation coefficient for the paired observations and compare it with the appropriate tabular value. A confidence interval for $\beta$ could also be using nonparametric method [see Iman and Conover (1980)]. We may also obtain confidence intervals for regression parameters by using the bootstrap method. Another possible alternative is to use robust regression when the error term is not normal.
2.3.1. Smoothers

Smoothers are very commonly used tools in nonparametric regression. A smoother is a tool for describing the trend of $Y$ as a function of one or more regressors. Smoothers are useful because the amount of horizontal scatter in data will often make it difficult to see the trend in a data set when there is a trend. Another advantage is that here we need not to specify a model. For a single regressor, the smoothing is called scatter plot smoothing, which entails estimating $E(Y \mid x_0)$, where $x_0$ denotes the regressor value for which the estimate is desired. The set of such estimates for the sample $x$ values is called a smooth. If $Y_j$ can be written as a linear combination of the $Y_j$, we call it a linear smoother. In general, we try to estimate $\mu(x_j)$, where $x_j$ is an arbitrary value of the regressor and the ‘model’ is assumed to be

$$Y_i = \mu(x_i) + u_i, \ i = 1, 2, \ldots, n$$

for which we would like to be able to assume that $\mu(x_j)$ is a smooth function. If so, then we would expect to be able to use $X$-values that are close to $x_j$ in estimating $\mu(x_j)$.

2.3.2. Monotone Regression

Assume that we have a regression data with a single regressor where the relationship between $X$ and $Y$ is nonlinear, but $Y$ may be considered as a monotonic function of $X$. In this situation one might attempt to use polynomial and/or other nonlinear terms in the model, or perhaps use a nonlinear regression model. A simpler approach, however, would be to use some form of monotone regression.

The simplest monotone regression model is a rank regression model suggested by Iman and Conover (1981). For a single regressor, the method entails converting $X$ and $Y$ to ranks and
then applying simple linear regression to the ranks. For a rank regression model we obtain the 
estimates of the regression parameters as

\[ \hat{\beta} = \frac{\sum R_x R_y - n(n+1)^2 / 4}{\sum R_x^2 - n(n+1)^2 / 4}, \quad \hat{\alpha} = \frac{(1 - \hat{\beta})(n+1)}{2} \]  

(2.14)

where \( R_x \) and \( R_y \) denote the ranks of \( X \) and \( Y \) respectively.

Monotone regression is analogous to the least squares with no assumptions in that we are 
simply fitting a line through the ranks of the points. The main disadvantage of this method is that 
since no assumptions are being made, there are essentially no confidence intervals, hypothesis 
tests, prediction intervals, and so forth. The use of this method is limited to predicting \( Y \). 
Obviously, there would also not be any regression plots, residual plots and there would be no 
checking of the model since there is no model. Monotone regression may also be used in 
multiple regression [Iman and Conover (1981)].

### 2.3.3. Running Line

One of the simplest type of smoothers for regression data is a running line smoother (Cleveland, 
1979). This is similar to a moving average; for the moving average an average is computed each 
time a point is deleted and another point is added, whereas with the running line smoother we get 
moving windows by deleting a point and adding another point and a simple linear regression line 
is computed for each moving window of data points. It is often useful to consider the updating 
formulae for the running line smoother that would take advantage of previous computations.

For a two variable regression, we get

\[ \hat{\beta} = \frac{S_{xy}}{S_{xx}}, \quad \hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X} \]

Therefore we need the updating formula for \( S_{xx}, \ S_{xy}, \ \bar{X} \) and \( \bar{Y} \).
Let \( n^* \) denote the size of the neighbourhood with \((x^+, y^+)\) representing the point that is to be added and \((x^-, y^-)\) the point is to be deleted.

Then the updating formulae for the required statistics are

\[
\bar{X}_{n^*} = \frac{(n^* \bar{X}_n + x^+ - x^-)}{n^*}
\]

\[
\bar{Y}_{n^*} = \frac{(n^* \bar{Y}_n + y^+ - y^-)}{n^*}
\]

\[
S_{xx}^* = S_{xx} + \frac{(x^+ - \bar{X}_{n^*})^2 - (x^- - \bar{X}_{n^*})^2 - (x^+ - x^-)^2}{n^*}
\]

\[
S_{xy}^* = S_{xy} + \frac{(x^+ - \bar{X}_{n^*})(y^+ - \bar{Y}_{n^*}) - (x^- - \bar{X}_{n^*})(y^- - \bar{Y}_{n^*}) - (x^+ - x^-)(y^+ - y^-)}{n^*}
\]

Hastie and Tibshirani (1987) suggested considering 10-50\% of the data to be the window size (size of the neighbourhood), \( n^* \). A nearest neighbourhood is defined as the set of \( n^* \) points that are closest to \( x_0 \). A symmetric nearest neighbourhood has \( n^*/2 \) points on each side of \( x_0 \). When it is not possible to have \( n^*/2 \) points on a given side, as many points as possible are used. Hastie and Tibshirani (1987) argue for the use of symmetric nearest neighbourhood for a running line smoother indicating that a nearest neighbourhood would give too much weight to points far away from \( X \), since the neighbourhood size at the end points would be the same as the neighbourhood size in the center of the data. Regardless of which approach we use, selecting \( n^* \) too small will cause the data not to be too smoothed at all, rather the running line will have a considerable variability.

**2.3.4. Local Regression**

Local regression was introduced by Cleveland (1979). It is more popularly known as locally weighted smoothing scatterplots (LOWESS). LOWESS uses the data from a neighbourhood around the specific location. Typically the neighbourhood is defined as the span, which is the fraction of the total data points used to form neighbourhood. A span of 0.5 (which is a very
popular choice) indicates that the closest half of the data points are used as the neighbourhood. The LOWESS procedure then uses the points in the neighbourhood to generate a weighted least squares estimate of the specific response. The weighted least squares procedure uses a low-order polynomial, usually a simple linear regression or a quadratic regression model. The weights for WLS portion of the estimation are based on the distance of the points used in the estimation from a specific location of interest. Here the fit in each neighbourhood is actually a weighted least squares fit. Local regression is similar to a running line smoother in that overlapping neighbourhoods are used. The fraction of observation (span) to be used in each neighbourhood must be specified. As we have already mentioned that a very popular choice of span is 0.5, but some other considerations are also available in the literature (Ryan, 1997).

2.4 Fitting of Saudi Arabia Expenditure Revenue Data Using Regression Models

In this section we employ the regression methods discussed above to fit expenditure on revenue data for Saudi Arabia. Initially we fit nine different regression models.

(i) Current expenditure (CE) on Oil revenue (OR)

(ii) Current expenditure on Other revenue (OrR)

(iii) Current expenditure on Total revenue (TR)

(iv) Capital expenditure (CpE) on Oil revenue

(v) Capital expenditure on Other revenue

(vi) Capital expenditure on Total revenue

(vii) Total expenditure (TE) on Oil revenue
(viii) Total expenditure on Other revenue
(ix) Total expenditure on Total revenue

Table 2.1 presents summary statistics of nine different regression models considered here. We employ the classical ordinary least squares (OLS) methods to fit each of the model and Table 2.1 presents the goodness of fit statistic $R^2$ in percentages together with the original and p-value of t used for testing the significance of regression, p-value of the F statistic from ANOVA and two normality test statistic (p-value of the Anderson-Darling test and the rescaled moments test).

<table>
<thead>
<tr>
<th>Model</th>
<th>$R^2$ (%)</th>
<th>$t$ (p)</th>
<th>$F$ (p)</th>
<th>AD (p)</th>
<th>RM (p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CE on OR</td>
<td>80.0</td>
<td>12.95 (0.000)</td>
<td>0.000</td>
<td>0.254</td>
<td>0.39 (0.823)</td>
</tr>
<tr>
<td>CE on OrR</td>
<td>75.1</td>
<td>11.26 (0.000)</td>
<td>0.000</td>
<td><strong>0.005</strong></td>
<td><strong>7.99 (0.018)</strong></td>
</tr>
<tr>
<td>CE on TR</td>
<td>82.1</td>
<td>13.90 (0.000)</td>
<td>0.000</td>
<td>0.555</td>
<td>0.01 (0.998)</td>
</tr>
<tr>
<td>CpE on OR</td>
<td>65.9</td>
<td>9.01 (0.000)</td>
<td>0.000</td>
<td>0.322</td>
<td>4.85 (0.088)</td>
</tr>
<tr>
<td>CpE on OrR</td>
<td>42.5</td>
<td>5.57 (0.000)</td>
<td>0.000</td>
<td>0.132</td>
<td><strong>11.36 (0.003)</strong></td>
</tr>
<tr>
<td>CpE on TR</td>
<td>65.7</td>
<td>8.96 (0.000)</td>
<td>0.000</td>
<td>0.576</td>
<td>3.71 (0.196)</td>
</tr>
<tr>
<td>TE on OR</td>
<td>85.5</td>
<td>15.74 (0.000)</td>
<td>0.000</td>
<td><strong>0.013</strong></td>
<td><strong>6.83 (0.032)</strong></td>
</tr>
<tr>
<td>TE on OrR</td>
<td>75.3</td>
<td>11.33 (0.000)</td>
<td>0.000</td>
<td><strong>0.005</strong></td>
<td><strong>6.03 (0.049)</strong></td>
</tr>
<tr>
<td>TE on TR</td>
<td>87.0</td>
<td>16.79 (0.000)</td>
<td>0.000</td>
<td><strong>0.005</strong></td>
<td><strong>18.56 (0.000)</strong></td>
</tr>
</tbody>
</table>
The above table shows that all nine regressions are significant. However on few occasions their $R^2$ values are not very strong. But the more alarming feature of this table is the fact that on five out of nine occasions the hypothesis of normality of errors are rejected which clearly indicates that we should not apply the OLS method to fit these five models. Similar remark may apply with the normal probability plot of errors as shown in Figure 2.1.

Figure 2.1: Normal Probability Plot of Nine Different Regression Models
Next we employ the nonparametric and robust fit for all these nine models. We have followed the best non-parametric method, the local regression (LOWESS) and the best robust method, the least median of squares (LMS) as suggested by Ryan (1997).

Figure 2.2: LMS and LOWESS Fit of Current Expenditure versus Oil Revenue for Saudi Arabia

Figure 2.3: LMS and LOWESS Fit of Current Expenditure versus Other Revenue for Saudi Arabia
Figure 2.4: LMS and LOWESS Fit of Current Expenditure versus Total Revenue for Saudi Arabia

Figure 2.5: LMS and LOWESS Fit of Capital Expenditure versus Oil Revenue for Saudi Arabia

Figure 2.6: LMS and LOWESS Fit of Capital Expenditure versus Other Revenue for Saudi Arabia
Figure 2.7: LMS and LOWESS Fit of Capital Expenditure versus Total Revenue for Saudi Arabia

Figure 2.8: LMS and LOWESS Fit of Total Expenditure versus Oil Revenue for Saudi Arabia

Figure 2.9: LMS and LOWESS Fit of Total Expenditure versus Other Revenue for Saudi Arabia
Figure 2.10: LMS and LOWESS Fit of Total Expenditure versus Total Revenue for Saudi Arabia

We really observe some interesting features of fitting regression model by different methods. The OLS indeed produces the minimum mean square error fits, but we often show that the LOWESS fits are getting closer to the most of the observations. The robust LMS fits the majority of the data very well, but most of the times it produces very poor fit for the other observations. From Figure 2.2 – 2.10 we get an impression that the LOESS produces the better fit followed by the OLS. The LMS fit is the worst overall.

Since graphical displays are always subjective, here we introduce some numerical summaries to assess the goodness of fit of the different regression methods. For a regular data three measures of accuracy of the fitted model: MAPE, MAD, and MSD for each of the simple fitting and smoothing methods. For all three measures, the smaller the value, the better the fit of the model. Use these statistics to compare the fits of the different methods.

**MAPE**, or Mean Absolute Percentage Error, measures the accuracy of fitted values. It expresses accuracy as a percentage.

\[
MAPE = \frac{\sum |(y_i - \hat{y}_i)/y_i|}{T} \times 100
\]  

(2.16)
where $y_i$ equals the actual value, $\hat{y}_i$ equals the fitted value, and $T$ equals the number of observations.

**MAD** (Mean), which stands for Mean Absolute Deviation, measures the accuracy of fitted values. It expresses accuracy in the same units as the data, which helps conceptualize the amount of error.

$$\text{MAD (Mean)} = \frac{\sum |y_i - \hat{y}_i|}{T}$$

(2.17)

where $y_i$ equals the actual value, $\hat{y}_i$ equals the fitted value, and $T$ equals the number of observations.

**MSD** stands for Mean Squared Deviation. MSD is always computed using the same denominator, $T$, regardless of the model, so you can compare MSD values across models. MSD is a more sensitive measure of an unusually large forecast error than MAD.

$$\text{MSD} = \frac{\sum(y_i - \hat{y}_i)^2}{T}$$

(2.18)

where $y_i$ equals the actual value, $\hat{y}_i$ equals the fitted value, and $T$ equals the number of observations.

Table 2.2: MAPE for Nine Different Regression Fits Using OLS, LOWESS and LMS

<table>
<thead>
<tr>
<th>Model</th>
<th>OLS</th>
<th>LOWESS</th>
<th>LMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>CE on OR</td>
<td>182.10</td>
<td><strong>154.40</strong></td>
<td>203.3</td>
</tr>
<tr>
<td>CE on OrR</td>
<td>55.80</td>
<td><strong>39.35</strong></td>
<td>77.60</td>
</tr>
<tr>
<td>CE on TR</td>
<td>138.90</td>
<td><strong>116.40</strong></td>
<td>173.60</td>
</tr>
<tr>
<td>CpE on OR</td>
<td>194.90</td>
<td><strong>108.30</strong></td>
<td>215.5</td>
</tr>
</tbody>
</table>
Table 2.2 offers a comparison of OLS, LOWESS and LMS employed to all nine different regression models in terms of MAPE and the best results (the lowest) are presented in boldfaces. This result shows LOWESS performs the best. The OLS is the next choice. The worst result is produced by the LMS,

### CHAPTER 3

**FORECASTING WITH ARIMA MODELS**

In this chapter we discuss different aspects of data analysis techniques useful in time series analysis. Here the prime topic of our discussion will be ARIMA models. We will talk about fitting of a model and generating forecasts using ARIMA models. An excellent review of different aspects of stochastic time series modelling is available in Pyndick and Rubenfield (1998), Bowerman *et al.* (2005) and Imon (2015). We assume that the time series models have been generated by a stochastic process. In other words, we assume that each value $y_1, y_2, \ldots, y_T$ in the series is randomly drawn from a probability distribution. We could assume that the observed series $y_1, y_2, \ldots, y_T$ is drawn from a set of jointly distributed random variables. If we could specify the probability distribution function of our series, we could determine the probability of one or another future outcome. Unfortunately, the complete specification of the
probability distribution function for a time series is usually impossible. However, it usually is possible to construct a simplified model of the time series which explains its randomness in a manner that is useful for forecasting purpose.

3.1 The Box-Jenkins Methodology

The Box-Jenkins methodology consists of a four-step iterative procedure.

Step 1: Tentative Identification: Historical data are used to tentatively identify an appropriate Box-Jenkins model.

Step 2: Estimation: Historical data are used to estimate the parameters of tentatively identified model.

Step 3: Diagnostic Checking: Various diagnostics are used to check the adequacy of the tentatively identified model and, if need be, to suggest an improved model, which is then regarded as a new tentatively identified model.

Step 4: Forecasting: Once a final model is obtained, it is used to forecast future time series values.

3.2 Stationary and Nonstationary Time Series

It is important to know whether the stochastic process that generates the series can be assumed to be invariant with respect to time. If the characteristic of the stochastic process changes over time we call the process nonstationary. If the process is nonstationary, it will often be difficult to represent the time series over past and future intervals of time by a simple algebraic model. By contrast, if the process is stationary, one can model the process via an equation with fixed coefficients that can be estimated from past data.
Properties of Stationary Process

We have said that any stochastic time series \( y_1, y_2, \ldots, y_T \) can be thought of as having been generated by a set of jointly distributed random variables; i.e., the set of data points \( y_1, y_2, \ldots, y_T \) represents a particular outcome (also known as a realization) of the joint probability distribution function \( p(y_1, y_2, \ldots, y_T) \). Similarly, a future observation \( y_{T+1} \) can be thought of as being generated by a conditional probability distribution function

\[
p(y_{T+1} \mid y_1, y_2, \ldots, y_T)
\]

that is, a probability distribution for \( y_{T+1} \) given the past observations \( y_1, y_2, \ldots, y_T \). We define a stationary process, then, as one whose joint distribution and conditional distribution both are invariant with respect to displacement in time. In other words, if the series is stationary, then

\[
p(y_t, y_{t+1}, \ldots, y_{t+k}) = p(y_{t+m}, y_{t+m+1}, \ldots, y_{t+m+k})
\]

and

\[
p(y_t) = p(y_{t+m})
\]

for any \( t, k, \) and \( m \).

If the series \( y_t \) is stationary, the mean of the series, which is defined as

\[
\mu_y = E(y_t)
\]

must also be stationary, so that \( E(y_t) = E(y_{t+m}) \), for any \( t \) and \( m \). Furthermore, the variance of the series

\[
\sigma_y^2 = E[(y_t - \mu_y)^2]
\]

must be stationary, so that

\[
E[(y_t - \mu_y)^2] = E[(y_{t+m} - \mu_y)^2].
\]
Finally, for any lag $k$, the covariance of the series

$$
\gamma_k = \text{Cov} (y_t, y_{t+k}) = E[(y_t - \mu_y)(y_{t+k} - \mu_y)]
$$

must be stationary, so that $\text{Cov} (y_t, y_{t+k}) = \text{Cov} (y_{t+m}, y_{t+m+k})$.

If a stochastic process is stationary, the probability distribution $p(y_t)$ is the same for all time $t$ and its shape can be inferred by looking at the histogram of the observations $y_1, y_2, \ldots, y_T$. An estimate of the mean $\mu_y$ can be obtained from the sample mean

$$
\bar{y} = \frac{\sum_{t=1}^{T} y_t}{T}
$$

and an estimate of the variance $\sigma_y^2$ can be obtained from the sample variance

$$
\hat{\sigma}_y^2 = \frac{\sum_{t=1}^{T} (y_t - \bar{y})^2}{T}.
$$

Usually it is very difficult to get a complete description of a stochastic process. The autocorrelation function could be extremely useful because it provides a partial description of the process for modeling purposes. The autocorrelation function tells us how much correlation there is between neighboring data points in the series $y_t$. We define the autocorrelation with lag $k$ as

$$
\rho_k = \frac{\text{Cov}(y_t, y_{t+k})}{\sqrt{\text{V}(y_t)\text{V}(y_{t+k})}}
$$

For a stationary time series the variance at time $t$ is the same as the variance at time $t + k$; thus, the autocorrelation becomes

$$
\rho_k = \frac{\gamma_k}{\sigma_y^2} = \frac{\gamma_k}{\gamma_0}
$$

and thus $\rho_0 = 1$ for any stochastic process.

Suppose the stochastic process is simply
\[ y_t = \epsilon_t \]

where \( \epsilon_t \) is an independently distributed random variable with zero mean. Then it is easy to show that for this process, \( \rho_0 = 1 \) and \( \rho_k = 0 \) for \( k > 0 \). This particular process is known as white noise, and there is no model that can provide a forecast any better than \( \hat{y}_{T+l} = 0 \) for all \( l \).

Thus, if the autocorrelation function is zero (or close to zero) for all \( k > 0 \), there is little or no value in using a model to forecast the series.

### 3.3 Test for Significance of a White Noise Autocorrelation Function

In practice, we use an estimate of the autocorrelation function, called the sample autocorrelation (SAC) function

\[
 r_k = \frac{\sum_{i=1}^{T-k} (y_i - \bar{y})(y_{i+k} - \bar{y})}{\sum_{i=1}^{T} (y_i - \bar{y})^2} \quad (3.11)
\]

It is easy to see from their definitions that both the theoretical and the estimated autocorrelation functions are symmetrical; i.e., \( \rho_k = \rho_{-k} \) and \( r_k = r_{-k} \).

#### 3.3.1. Bartlett’s test

Here the null hypothesis is \( H_0: \rho_k = 0 \) for \( k > 0 \). Bartlett shows that when the time series is generated by a white noise process, the sample autocorrelation function is distributed approximately as a normal with mean 0 and variance \( 1/T \). Hence, the test statistic is
\[ |z| = \sqrt{T} |r_k| \]  

and we reject the null hypothesis at the 95% level of significance, if \(|z|\) is greater than 1.96.

3.3.2. The t-test based on SAC

The standard error of \( r_k \) is given by

\[
SE(r_k) = \begin{cases} 
\frac{1}{\sqrt{T}} & \text{if } k = 1 \\
\sqrt{ \frac{1 + 2 \sum_{i=1}^{k-1} r_i^2}{T} } & \text{if } k > 1 
\end{cases}
\]

The t-statistic for testing the hypothesis \( H_0: \rho_k = 0 \) for \( k > 0 \) is defined as

\[
T = \frac{r_k}{SE(r_k)}
\]

and this test is significant when \(|T| > 2\).

3.3.3. Box and Pierce Test and Ljung and Box Test

To test the joint hypothesis that all the autocorrelation coefficients are zero we use a test statistic introduced by Box and Pierce. Here the null hypothesis is

\[
H_0: \rho_1 = \rho_2 = \ldots = \rho_k = 0.
\]

Box and Pierce show that the appropriate statistic for testing this null hypothesis is

\[
Q = T \sum_{i=1}^{k} r_i^2
\]

is distributed as chi-square with \( k \) degrees of freedom.

A slight modification of the Box-Pierce test was suggested by Ljung and Box, which is known as the Ljung-Box Q (LBQ) test defined as

\[
Q = T(T + 2) \sum_{i=1}^{k} (T - k)^{-1} r_i^2
\]
Thus, if the calculated value of $Q$ is greater than, say, the critical 5% level, we can be 95% sure that the true autocorrelation coefficients are not all zero.

### 3.3.4. Stationarity and the Autocorrelation Function

How can we decide that whether a series is stationary or determine the appropriate number of times a homogenous nonstationary series should be differenced to arrive at a stationary series?

The correlogram (a plot of autocorrelation coefficients against the number of lag periods) could be a useful indicator of it. For a stationary series, the autocorrelation function drops off as $k$ becomes large, but this usually is not the case for a nonstationary series. In order to employ the Box-Jenkins methodology, we must examine the behavior of the SAC. The SAC for a nonseasonal time series can display a variety of behaviors. First, the SAC for a nonseasonal time series can cut off at lag $k$. We say that a spike at lag $k$ exists in the SAC if the SAC at lag $k$ is statistically significant. Second, we say that the SAC dies down if this function does not cut-off but rather decreases in a steady fashion. In general, it can be said that

1. If the SAC of the time series values either cuts off fairly quickly or dies down fairly quickly, then the time series values should be considered stationary.
2. If the SAC of the time series values dies down extremely slowly, then the time series values should be considered nonstationary.

### 3.4 ARIMA Models

In this section we introduce some commonly used ARIMA models.

#### 3.4.1. Moving Average Models
In the moving average process of order $q$ each observation $y_t$ is generated by a weighted average of random disturbances going back to $q$ periods. We denote this process as MA($q$) and write its equation as

$$y_t = \mu + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \ldots - \theta_q \epsilon_{t-q}$$

(3.17)

In the moving average model the random disturbances are assumed to be independently distributed across time, i.e., generated by a white noise process. In particular, each $\epsilon_t$ is assumed to be normal random variable with mean 0 and variance $\sigma_\epsilon^2$, and covariance $\gamma_k = 0$ for $k \neq 0$. We have $E(y_t) = \mu$ which shows that a moving average process is independent of time.

The process MA($q$) is described by exactly $q + 2$ parameters, the mean $\mu$, the disturbance $\sigma_\epsilon^2$ and the parameters $\theta_1, \theta_2, \ldots, \theta_q$.

Let us now look at the variance of the moving average process of order $q$.

$$\gamma_0 = E[(y_t - \mu)^2]$$

(3.18)

$$= E(\epsilon_t^2 + \theta_1^2 \epsilon_{t-1}^2 + \theta_2^2 \epsilon_{t-2}^2 + \ldots + \theta_q^2 \epsilon_{t-q}^2 - 2\theta_1 \epsilon_t \epsilon_{t-1} y_{t-1} - \ldots)$$

$$= \sigma_\epsilon^2 (1 + \theta_1^2 + \theta_2^2 + \ldots + \theta_q^2)$$

The moving average process of order $q$ has autocorrelation function

$$\rho_k = \begin{cases} 
- \theta_k + \theta_1 \theta_{k+1} + \theta_2 \theta_{k+2} + \ldots + \theta_{q-k} \theta_q & \text{for } k = 1, 2, \ldots, q \\
1 + \theta_1^2 + \theta_2^2 + \ldots + \theta_q^2 & \text{for } k > q 
\end{cases}$$

(3.19)

### 3.4.2. Autoregressive Models
In the autoregressive process of order $p$ the current observation $y_t$ is generated by a weighted average of past observations going back $p$ periods, together with a random disturbance in the current period. We denote this process as AR($p$) and write its equation as

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \epsilon_t$$  

(3.20)

If the autoregressive process is stationary, then its mean, which we denote by $\mu$, must be invariant with respect to time; i.e., $E(y_t) = E(y_{t-1}) = E(y_{t-p}) = \mu$. The mean $\mu$ is thus given by

$$\mu = \delta + \phi_1 \mu + \phi_2 \mu + \ldots + \phi_p \mu$$

or

$$\mu = \frac{\delta}{1-\phi_1-\phi_2-\ldots-\phi_p}$$

If the process is stationary, then $\mu$ must be finite. For this, it is necessary that $\phi_1 + \phi_2 + \ldots + \phi_p < 1$.

For the first-order process AR(1): $y_t = \delta + \phi_1 y_{t-1} + \epsilon_t$,

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \phi_1^k$$  

(3.21)

In general, for $k > 1$,

$$\gamma_k = E[y_{t-k}(\phi_1 y_{t-k-1} + \phi_2 y_{t-k-2} + \epsilon_t)] = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2}$$  

(3.22)

Thus the autocorrelation function is given by

$$\rho_1 = \frac{\phi_1}{1-\phi_2}$$  

(3.23)

$$\rho_k = \rho_1 \rho_{k-1} + \rho_2 \rho_{k-2} \quad \text{for } k > 1$$

### 3.4.3. The Partial Autocorrelation Function

The partial autocorrelation function is used to determine the order of an autoregressive process.

For an autoregressive process of order $p$, the covariance with displacement $k$ is determined from
\[ \gamma_k = E \left[ y_{t-k} (\phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \epsilon_t) \right] \]  

(3.24)

which gives

\[ \gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \ldots + \phi_p \gamma_p + \sigma_e^2 \]

\[ \gamma_1 = \phi_1 \gamma_0 + \phi_2 \gamma_1 + \ldots + \phi_p \gamma_{p-1} \]

…………………

\[ \gamma_p = \phi_1 \gamma_{p-1} + \phi_2 \gamma_{p-2} + \ldots + \phi_p \gamma_0 \]  

(3.25)

The above equations also give a set of \( p \) equations, known as Yule-Walker equations, to determine the first \( p \) values of the autocorrelation functions:

\[ \rho_1 = \phi_1 + \phi_2 \rho_1 + \phi_p \rho_{p-1} \]

…………………

\[ \rho_p = \phi_1 \rho_{p-1} + \phi_2 \rho_{p-2} + \ldots + \phi_p \]  

(3.26)

The solution of the Yule-Walker equations requires the knowledge of \( p \). Therefore we solve these equations for successive values of \( p \). We begin by hypothesizing that \( p = 1 \). We compute the sample autocorrelation \( \hat{\rho}_1 \) as an estimate of \( \rho_1 \). If this value is significantly different from 0, we know that the autoregressive process is at least order 1. Next we consider the hypothesis that \( p = 2 \). We solve the Yule-Walker equations for \( p = 2 \) and obtain a new set of estimates for \( \phi_1 \) and \( \phi_2 \). If \( \phi_2 \) is significantly different from 0, we may conclude that the process is at least order 2. Otherwise we conclude that the process is order 1. We repeat this process for successive values of \( p \). We call the series \( \phi_1, \phi_2, \ldots \), partial autocorrelation function. If the true order of the process is \( p \), we should observe that \( \hat{\phi}_j \approx 0 \) for \( j > p \). To test whether a particular \( \phi_j \) is zero, we can use the fact that it is approximately normally distributed with mean 0 and variance \( 1 / T \).
Hence we can check whether it is statistically significant at, say, the 5% level by determining whether it exceeds $2 / \sqrt{T}$ in magnitude.

### 3.4.4. Mixed Autoregressive – Moving Average (ARMA) Models

Many stationary random processes cannot be modeled as purely moving average or as purely autoregressive, since they have quality of both types of processes. In this situation we can use the mixed autoregressive – moving average (ARMA) process. An ARMA process of order $(p, q)$ is defined as

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \ldots - \theta_q \epsilon_{t-q}$$  \hspace{1cm} (3.27)

We assume that the process is stationary, then its mean, which we denote by $\mu$, must be invariant with respect to time and is given by $\mu = \delta + \phi_1 \mu + \phi_2 \mu + \ldots + \phi_p \mu$ or

$$\mu = \frac{\delta}{1 - \phi_1 - \phi_2 - \ldots - \phi_p}.$$  

For the general ARMA$(p, q)$ process it is not easy to obtain the variances, covariances and autocorrelations by solving equations. It can be shown easily, however, that

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \ldots + \phi_p \rho_{k-p}, \quad k > q$$  \hspace{1cm} (3.28)

It is interesting to note that $q$ is the memory of the moving average part of the process, so that for $k > q$, the autocorrelation function exhibits the properties of a purely autoregressive process.

The above equations also give a set of $p$ equations, known as Yule-Walker equations, to determine the first $p$ values of the autocorrelation functions:

$$\rho_1 = \phi_1 + \phi_2 \rho_1 + \phi_p \rho_{p-1}$$

…………………………
\[ \rho_p = \phi_1 \rho_{p-1} + \phi_2 \rho_{p-2} + \ldots + \phi_p \]  

(3.29)

The solution of the Yule-Walker equations requires the knowledge of \( p \). Therefore we solve these equations for successive values of \( p \). We begin by hypothesizing that \( p = 1 \). We compute the sample autocorrelation \( \hat{\phi}_1 \) as an estimate of \( \rho_1 \). If this value is significantly different from 0, we know that the autoregressive process is at least order 1. Next we consider the hypothesis that \( p = 2 \). We solve the Yule-Walker equations for \( p = 2 \) and obtain a new set of estimates for \( \phi_1 \) and \( \phi_2 \). If \( \phi_2 \) is significantly different from 0, we may conclude that the process is at least order 2. Otherwise we conclude that the process is order 1. We repeat this process for successive values of \( p \). We call the series \( \phi_1, \phi_2, \ldots \), partial autocorrelation function. If the true order of the process is \( p \), we should observe that \( \hat{\phi}_j \approx 0 \) for \( j > p \).

To test whether a particular \( \phi_j \) is zero, we can use the fact that it is approximately normally distributed with mean 0 and variance 1 / \( T \). Hence we can check whether it is statistically significant at, say, the 5% level by determining whether it exceeds \( 2 / \sqrt{T} \) in magnitude.

### 3.4.5. Homogenous Nonstationary Processes: ARIMA Models

Probably very few of the time series one meets in practice are stationary. Fortunately, however, many of the nonstationary time series that are encountered have the desirable property that they are differentiated one or more times, the resulting series will be stationary. Such a nonstationary series is termed homogenous. The number of times the original series must be differenced before a stationary series results in is called the order of homogeneity. Thus, if \( y_t \) is first order homogenous nonstationary, the series \( w_t = y_t - y_{t-1} = \Delta y_t \) is stationary. Here we should
construct models for those nonstationary series, which can be transformed into stationary series by differencing them one or more times. We say that $y_t$ is homogenous nonstationary of order $d$ if

$$w_t = \Delta^d y_t,$$  \hspace{1cm} (3.30)

is a stationary series. If $w_t = \Delta^d y_t$ and $w_t$ is an ARMA($p$, $q$) process, then we say that $y_t$ is an integrated autoregressive moving average process of order $p$, $d$ and $q$, or simply ARIMA($p$, $d$, $q$).

### 3.5 Estimation and Specification of ARIMA Models

It is often convenient to describe time lags by using the backward shift operator $B$. The operator $B$ imposes a one-period time lag each time it is applied to a variable. Thus

$$B \epsilon_t = \epsilon_{t-1}, B^2 \epsilon_t = \epsilon_{t-2}, \ldots, B^n \epsilon_t = \epsilon_{t-n}.$$  \hspace{1cm} (3.31)

Using this operator, we can write an MA($q$) process as:

$$y_t = \mu + \epsilon_t \left(1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q\right) = \mu + \epsilon_t \theta(B)$$  \hspace{1cm} (3.32)

In a similar way, the AR($p$) process can be rewritten as

$$y_t \left(1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p\right) = \delta + \epsilon_t$$  \hspace{1cm} (3.33)

Finally, an ARMA($p$, $q$) process can be reexpressed as

$$y_t \left(1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p\right) = \delta + \epsilon_t \left(1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q\right)$$  \hspace{1cm} (3.34)

It is easy to show that any homogenous nonstationary process can be modeled as an ARIMA process. We can write the equation for an ARIMA($p$, $d$, $q$) as:

$$\phi(B) \Delta^d y_t = \delta + \theta(B) \epsilon_t$$  \hspace{1cm} (3.35)

If $d = 0$, we obtain
\[
\phi(B)y_t = \delta + \theta(B)\varepsilon_t
\]  
(3.36)

which is an ARMA(\(p, q\)). When \(q = 0\), i.e., when \(w_t = \Delta^d y_t \) is just AR(\(p\)), we call \(y_t\) an integrated autoregressive process of order \((p, d)\) and denote it as ARIMA(\(p, d, 0\)) or ARI(\(p, d, 0\)). When \(p = 0\), i.e., when \(w_t\) is just MA(\(q\)), we call \(y_t\) an integrated moving average process of order \((d, q)\) and denote it as ARIMA(\(0, d, q\)) or IMA(\(0, d, q\)). In practice, it is crucial to specify the ARIMA model, i.e., to choose the most appropriate values for \(p\), \(d\) and \(q\).

Given a series \(y_t\), the first problem is to determine the degree of homogeneity \(d\). To do this one first examines the autocorrelation function of the original series \(y_t\) and determine whether it is stationary. If it is not, difference the series and examine the autocorrelation function for \(\Delta^d y_t\). Repeat this process until a value of \(d\) is reached such that \(\Delta^d y_t\) is stationary, i.e., the autocorrelation function goes to 0 as \(k\) becomes large. After \(d\) is determined, one can work with the stationary series \(w_t = \Delta^d y_t\) and examine both its autocorrelation and partial autocorrelation function to determine possible specifications for \(p\) and \(q\). The lower order processes like AR(1), AR(2), MA(1), MA(2), ARMA(1, 1) etc are easy to recognize.

<table>
<thead>
<tr>
<th>Model</th>
<th>ACF</th>
<th>PACF</th>
</tr>
</thead>
<tbody>
<tr>
<td>White noise</td>
<td>All zero</td>
<td>All zero</td>
</tr>
<tr>
<td>MA(1)</td>
<td>Zero after 1 lag</td>
<td>Declining from 1st lag</td>
</tr>
<tr>
<td>MA(2)</td>
<td>Zero after 2 lags</td>
<td>Declining from 2nd lag</td>
</tr>
<tr>
<td>MA((q))</td>
<td>Zero after (q) lags</td>
<td>Declining from (q)th lag</td>
</tr>
</tbody>
</table>

Table 3.1: Specification of ARIMA Models
<table>
<thead>
<tr>
<th>Model</th>
<th>Description</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(1)</td>
<td>Geometric decline from 1 lag</td>
<td>Zero after 1 lag</td>
</tr>
<tr>
<td>AR(2)</td>
<td>Geometric decline from 2 lags</td>
<td>Zero after 2 lags</td>
</tr>
<tr>
<td>AR(p)</td>
<td>Geometric decline from pth lag</td>
<td>Zero after p-lags</td>
</tr>
<tr>
<td>ARMA(1,1)</td>
<td>Geometric decline from 1 lag</td>
<td>Declining from first lag</td>
</tr>
<tr>
<td>ARMA(p,q)</td>
<td>Geometric decline from pth lag</td>
<td>Declining from qth lag</td>
</tr>
</tbody>
</table>

The autoregressive and moving average of order \( p \) and \( q \) respectively are determined by using the partial autocorrelation function (PACF) and ACF respectively. Table 3.1 summarizes the characteristics of the ARIMA(\( p,0,q \)) or ARMA(\( p,q \)) model.

### 3.6 Diagnostic Checking

After a time series model has been estimated, one must test whether the specification was correct. We assume that the random errors \( \varepsilon_i \) in the actual process are normally distributed and independent. Then, if the model is specified correctly, the residuals \( \hat{\varepsilon}_i \) should resemble a white noise process. Consequently a sample autocorrelation function of the residuals

\[
\hat{r}_k = \frac{\sum_{t=1}^{\infty} \hat{\varepsilon}_t \hat{\varepsilon}_{t-k}}{\sum_{t=1}^{\infty} \hat{\varepsilon}_t^2}
\]

would be close to 0 for \( k > 0 \). We can use the Box and Pierce Test for this purpose. Consider the statistic \( Q \) composed of the first \( K \) residual autocorrelations.
\[ Q = T \sum_{k=1}^{K} \hat{\rho}_k^2 \tag{3.38} \]

which is distributed as chi-square with \( K - p - q \) degrees of freedom.

### 3.7 Computing a Forecast

In ARIMA models we generate forecasts by the minimum mean-square error method. Here our objective is to predict future values of a time series. For this reason we consider the optimum forecast to be that forecast which has the minimum mean square forecast error. Since the forecast error is a random variable, we minimize the expected value. Thus, we wish to choose our forecast \( \hat{y}_T(l) \) so that

\[ E \left[ e^2(l) \right] = E \left[ \{ y_{T+l} - \hat{y}_T(l) \}^2 \right] \tag{3.39} \]

is minimized. It is easy to show that this forecast is given by the conditional expectation of \( y_{T+l} \), that is, by

\[ \hat{y}_T(l) = E(y_{T+l} | y_T, y_{T-1}, \ldots, y_1) \]

The computation of the forecast \( \hat{y}_T(l) \) can be done recursively by using the estimated ARIMA model. This involves first computing a forecast one period ahead, then using this forecast to compute a forecast two periods ahead, and continuing until the \( l \)-period forecast has been reached.

Let us write the ARIMA\((p, d, q)\) model as

\[ w_t = \phi_1 w_{t-1} + \phi_2 w_{t-2} + \ldots + \phi_p w_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \ldots - \theta_q \epsilon_{t-q} + \delta \tag{3.40} \]

To compute the forecast \( \hat{y}_T(l) \), we begin by computing the one-period forecast of \( w_t, \hat{w}_T(1) \). To do so, we write
\[ w_{T+1} = \phi_1 w_T + \phi_2 w_{T-1} + \ldots + \phi_p w_{T-p+1} + \epsilon_{T+1} - \theta_1 \hat{\epsilon}_T - \ldots - \theta_q \hat{\epsilon}_{q+1} + \delta \]  \hspace{1cm} (3.41)

We obtain

\[
\hat{w}_T(1) = E(w_{T+1} | w_T, \ldots) = \phi_1 w_T + \phi_2 w_{T-1} + \ldots + \phi_p w_{T-p+1} - \theta_1 \hat{\epsilon}_T - \ldots - \theta_q \hat{\epsilon}_{q+1} + \delta
\]

Now, using the one-period forecast \( \hat{w}_T(1) \), we can obtain the second-period forecast \( \hat{w}_T(2) \):

\[
\hat{w}_T(2) = \phi_1 \hat{w}_T(1) + \phi_2 w_T + \ldots + \phi_{p-1} \hat{w}_T(p-1) + \epsilon_{T+2} - \theta_1 \hat{\epsilon}_T - \ldots - \theta_q \hat{\epsilon}_{q+2} + \delta \]  \hspace{1cm} (3.43)

Thus the \( l \)-period forecast is given by

\[
\hat{w}_T(l) = \phi_1 \hat{w}_T(l-1) + \ldots + \phi_1 w_T + \ldots + \phi_p \hat{w}_T(p-1) - \theta_1 \hat{\epsilon}_T - \ldots - \theta_q \hat{\epsilon}_{q+2} + \delta \]  \hspace{1cm} (3.44)

If \( l > p \) and \( l > q \), this forecast will be

\[
\hat{w}_T(l) = \phi_1 \hat{w}_T(l-1) + \ldots + \phi_p \hat{w}_T(p-1) w_{T-p+1}
\]

When \( d = 1 \), our \( l \)-period forecast \( \hat{y}_T(l) \) is given by

\[
\hat{y}_T(l) = \hat{w}_T(1) + \hat{w}_T(2) + \ldots + \hat{w}_T(l)
\]

### 3.7.1. The Forecast Error

We can write the ARIMA \((p, d, q)\) model as

\[
\phi(B)(1-B)^d y_{t} = \theta(B) \epsilon_t
\]

where \( \delta = 0 \) and \( \Delta = 1 - B \). Therefore,

\[
y_{t} = \phi^{-1}(B)(1-B)^{-d} \theta(B) \epsilon_{t} = \psi(B) \epsilon_{t} = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}
\]

Thus an ARIMA model can be expressed as a purely moving average process of infinite order.

Then

\[
y_{T+t} = \psi_0 \epsilon_{T+t} + \psi_1 \epsilon_{T+t-1} + \ldots + \psi_{T-1} \epsilon_{T+1} + \sum_{j=0}^{\infty} \psi_{T+j} \epsilon_{T-j}
\]
The desired forecast $\hat{y}_T(l)$ can be based only on information available up to time $T$ that gives

$$\hat{y}_T(l) = \sum_{j=0}^{\alpha} \psi_{l+j} \in_{T-j}$$

We define the forecast error $e_T(l)$ as

$$e_T(l) = y_{T+l} - \hat{y}_T(l) = \psi_0 \in_{T+l} + \psi_1 \in_{T+l-1} + \ldots + \psi_{l-1} \in_{T+l}$$

(3.45)

The variance of the forecast error is given by

$$E(e^2_T(l)) = (\psi_0^2 + \psi_1^2 + \ldots + \psi_{l-1}^2) \sigma_e^2$$

It is easy to show that

$$\psi_0 = 1$$

Therefore, for any ARIMA specification, the forecast error one period ahead is just

$$e_T(l) = \epsilon_{T+l}$$

3.7.2. Forecast Confidence Interval

The estimate of $\sigma_e^2$ is given by

$$\hat{\sigma_e^2} = \sum_{i=1}^{T} \epsilon_i^2 / (T - p - q)$$

(3.46)

Hence the 100 $(1 - \alpha)\%$ confidence interval around a forecast $l$ periods ahead would be given by

$$\hat{y}_T(l) \pm z_{(1-\alpha/2)} \sqrt{1 + \sum_{j=1}^{l-1} \psi_j^2} \sigma_e$$

(3.47)

3.7.3. The AR(1) Process
Let us consider the AR(1) process:

\[ y_t = \phi_1 y_{t-1} + \delta + \epsilon_t \]

The one-period forecast is

\[ \hat{y}_T(1) = \phi_1 y_T + \delta \]

The two-period forecast is

\[ \hat{y}_T(2) = \phi_1 \hat{y}_T(1) = \phi_1^2 y_T + (\phi_1 + 1) \delta \]

Thus the \( l \)-period forecast is

\[ \hat{y}_T(l) = \phi_1^l y_T + (\phi_1^{l-1} + \phi_1^{l-2} + \ldots + \phi_1 + 1) \delta \] (3.48)

As \( l \) becomes large, the forecast converges to the value

\[ \lim_{l \to \infty} \hat{y}_T(l) = \delta \sum_{j=0}^{\infty} \phi_1^j = \delta / (1 - \phi_1) \]

which is the mean of the process. The forecast error is given by

\[ e_T(l) = \epsilon_{T+l} + \phi_1 \epsilon_{T+l-1} + \ldots + \phi_1^{l-1} \epsilon_{T+1} \]

which has a variance

\[ E(e_T^2(l)) = (1 + \phi_1^2 + \ldots + \phi_1^{2l-2}) \sigma_\epsilon^2 \]

### 3.7.4. The MA(1) Process

Let us consider the MA(1) process:

\[ y_t = \delta + \epsilon_t - \theta_1 \epsilon_{t-1} \]

The one-period forecast is

\[ \hat{y}_T(1) = \delta - \theta_1 \hat{\epsilon}_T \]

The \( l \)-period forecast is

\[ \hat{y}_T(l) = \delta - \theta_1 \hat{\epsilon}_T + \phi_1 \hat{y}_T(l-1) \]
\[ \hat{y}_T(l) = E(\delta + \epsilon_{T+l} - \theta_1 \epsilon_{T+l-1}) = \delta \]  

(3.49)

for \( l > 1 \). The forecast error is given by

\[ e_T(l) = \epsilon_{T+l} - \theta_1 \epsilon_{T+l-1} \]

which has a variance

\[ E(e^2_T(l)) = (1 + \theta_1^2) \sigma_e^2 \]

### 3.7.5. The ARMA(1, 1) Process

An ARMA (1, 1) process is:

\[ y_t = \delta + \phi_1 y_{t-1} + \epsilon_t - \theta_1 \epsilon_{t-1} \]

The one-period forecast is

\[ \hat{y}_T(1) = E(\phi_1 y_T + \delta + \epsilon_{T+l} - \theta_1 \epsilon_T) = \phi_1 y_{T-1} + \delta + \theta_1 \hat{\epsilon}_T \]

The two-period forecast is

\[ \hat{y}_T(2) = \phi_1 \hat{y}_T(1) + \delta = \phi_1^2 y_T + (\phi_1 + 1) \delta - \phi_1 \theta_1 \hat{\epsilon}_T \]

Finally the \( l \)-period forecast is

\[ \hat{y}_T(l) = \phi_1^l y_T + (\phi_1^{l-1} + \phi_1^{l-2} + \ldots + \phi_1 + 1) \delta - \phi_1^{l-1} \theta_1 \hat{\epsilon}_T \]  

(3.51)

### 3.7.6. The ARI(1, 1, 0) Process

Now we examine a simple nonstationary process, the integrated autoregressive process ARI(1, 1, 0):

\[ w_t = \phi_1 w_{t-1} + \delta + \epsilon_t \]

with \( w_t = y_t - y_{t-1} \). Since \( w_t \) is AR(1), the \( l \)-period forecast is

\[ \hat{w}_T(l) = \phi_1^l w_T + (\phi_1^{l-1} + \phi_1^{l-2} + \ldots + \phi_1 + 1) \delta \]
We also have
\[ \hat{y}_T(l) = y_T + \hat{w}_T(1) + \hat{w}_T(2) + \ldots + \hat{w}_T(l) \]

The one-period forecast is
\[ \hat{y}_T(1) = y_T + \phi_1(y_T - y_{T-1}) + \delta = (1 + \phi_1) y_T - \phi_1 y_{T-1} \]

The two-period forecast is
\[ \hat{y}_T(2) = y_T + \hat{w}_T(1) + \hat{w}_T(2) = \hat{y}_T(1) + \hat{w}_T(2) \]
\[ = (1 + \phi_1) y_T - \phi_1 y_{T-1} + \phi_1^2 (y_T - y_{T-1}) + (\phi_1 + 1) \delta \]
\[ = (1 + \phi_1 + \phi_1^2) y_T - (\phi_1 + \phi_1^2) y_{T-1} + (\phi_1 + 1) \delta + \delta \]

Since \( \hat{w}_T(2) = \phi_1^2 w_T + (\phi_1 + 1) \delta = \phi_1 (\phi_1 w_T + \delta) + \delta = \phi_1 \hat{w}_T(1) + \delta \), hence
\[ \hat{y}_T(2) = \hat{y}_T(1) + \phi_1 \hat{w}_T(1) + \delta \]

Similarly,
\[ \hat{y}_T(l) = \hat{y}_T(l-1) + \phi_1 \hat{w}_T(l-1) + \delta \quad (3.52) \]

The forecast error for one-period is given by
\[ e_T(1) = y_{T+1} - \hat{y}_T(1) = y_T + w_{T+1} - y_T - \hat{w}_T(1) = \epsilon_{T+1} \]

The two-period forecast error is given by
\[ e_T(2) = y_{T+2} - \hat{y}_T(2) = y_T + w_{T+1} + w_{T+2} - y_T - \hat{w}_T(1) - \hat{w}_T(2) \]
\[ = [w_{T+1} - \hat{w}_T(1)] + [w_{T+2} - \hat{w}_T(2)] = (1 + \phi_1) \epsilon_{T+1} + \epsilon_{T+2} \]

Finally
\[ e_T(l) = (1 + \phi_1 + \phi_1^2 + \ldots + \phi_1^{l-1}) \epsilon_{T+1} + (1 + \phi_1 + \phi_1^2 + \ldots + \phi_1^{l-2}) \epsilon_{T+2} + \ldots \]
\[ + (1 + \phi_1) \epsilon_{T+l-1} + \epsilon_{T+l} \]

This has a variance
\[ E(e_r^2(l)) = \left( \sum_{i=1}^{l} \sum_{j=0}^{l-j} \phi_i \right) \sigma^2 \]

### 3.8 Fitting of the ARIMA Model to Saudi Arabia Expenditure Data

In our study we have considered three variables, Current expenditure, Capital expenditure, and Total expenditure as response variables. In this section we would like to employ ARIMA models to fit these three variables. In order to determine the order of ARIMA we compute the ACF, the PACF, and the corresponding \( t \) value for each of the three variables and the results are given below.

Figures 3.1 – 3.3 present the ACF and the PACF of each of the three variables. It is clear from the plots that each of the variable should perfectly match with AR(1) since the ACF shows a geometric declining pattern from the first lag and the PACF dies down to zero right after the first lag. Similar remarks may apply with the ACF and PACF values with the associated \( t \) and Ljung-Box tests as shown in Tables 3.2- 3.4.

![Autocorrelation Function for Current Expenditure](image1.png)

![Partial Autocorrelation Function for Current Expenditure](image2.png)

Figure 3.1: ACF and PACF of Current Expenditure Data
Figure 3.2: ACF and PACF of Capital Expenditure Data

Figure 3.3: ACF and PACF of Total Expenditure Data

Table 3.2: ACF and PACF of Current Expenditure of Saudi Arabia

<table>
<thead>
<tr>
<th>ACF</th>
<th>TSTAT</th>
<th>LBQ</th>
<th>PACF</th>
<th>TSTAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.833165</td>
<td>5.52659</td>
<td>32.674</td>
<td>0.833165</td>
<td>5.52659</td>
</tr>
<tr>
<td>0.701360</td>
<td>3.01037</td>
<td>56.379</td>
<td>0.023532</td>
<td>0.15609</td>
</tr>
<tr>
<td>0.604037</td>
<td>2.18191</td>
<td>74.391</td>
<td>0.045258</td>
<td>0.30021</td>
</tr>
<tr>
<td>0.516495</td>
<td>1.69162</td>
<td>87.889</td>
<td>-0.008749</td>
<td>-0.05803</td>
</tr>
<tr>
<td>0.419928</td>
<td>1.29377</td>
<td>97.041</td>
<td>-0.069723</td>
<td>-0.46249</td>
</tr>
</tbody>
</table>
Table 3.3: ACF and PACF of Capital Expenditure of Saudi Arabia

<table>
<thead>
<tr>
<th>ACF</th>
<th>TSTAT</th>
<th>LBQ</th>
<th>PACF</th>
<th>TSTAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.348457</td>
<td>1.03492</td>
<td>103.508</td>
<td>0.019122</td>
<td>0.12684</td>
</tr>
<tr>
<td>0.271412</td>
<td>0.78717</td>
<td>107.538</td>
<td>-0.062198</td>
<td>-0.41258</td>
</tr>
<tr>
<td>0.220287</td>
<td>0.63008</td>
<td>110.266</td>
<td>0.035099</td>
<td>0.23282</td>
</tr>
<tr>
<td>0.182959</td>
<td>0.51865</td>
<td>112.202</td>
<td>0.016148</td>
<td>0.10711</td>
</tr>
<tr>
<td>0.155286</td>
<td>0.43754</td>
<td>113.637</td>
<td>0.013669</td>
<td>0.09067</td>
</tr>
</tbody>
</table>

Table 3.4: ACF and PACF of Current Expenditure of Saudi Arabia

<table>
<thead>
<tr>
<th>ACF</th>
<th>TSTAT</th>
<th>LBQ</th>
<th>PACF</th>
<th>TSTAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.811515</td>
<td>5.38298</td>
<td>30.9981</td>
<td>0.811515</td>
<td>5.38298</td>
</tr>
<tr>
<td>0.584720</td>
<td>2.54801</td>
<td>47.4743</td>
<td>-0.216245</td>
<td>-1.43441</td>
</tr>
<tr>
<td>0.409950</td>
<td>1.56975</td>
<td>55.7707</td>
<td>0.025554</td>
<td>0.16950</td>
</tr>
<tr>
<td>0.241911</td>
<td>0.87842</td>
<td>58.7318</td>
<td>-0.136878</td>
<td>-0.90795</td>
</tr>
<tr>
<td>0.098896</td>
<td>0.35297</td>
<td>59.2394</td>
<td>-0.040152</td>
<td>-0.26634</td>
</tr>
<tr>
<td>-0.028611</td>
<td>-0.10183</td>
<td>59.2830</td>
<td>-0.107154</td>
<td>-0.71078</td>
</tr>
<tr>
<td>-0.111853</td>
<td>-0.39800</td>
<td>59.9674</td>
<td>0.011067</td>
<td>0.07341</td>
</tr>
<tr>
<td>-0.173032</td>
<td>-0.61348</td>
<td>61.6507</td>
<td>-0.078123</td>
<td>-0.51821</td>
</tr>
<tr>
<td>-0.204928</td>
<td>-0.72043</td>
<td>64.0792</td>
<td>0.005271</td>
<td>0.03497</td>
</tr>
<tr>
<td>-0.234778</td>
<td>-0.81580</td>
<td>67.3605</td>
<td>-0.105745</td>
<td>-0.70143</td>
</tr>
</tbody>
</table>
Now we fit AR (1) model for current, capital and total expenditure of Saudi Arabia data and the results are presented in Table 3.5.

Table 3.5: AR (1) Fits for Current, Capital and Total Expenditure of Saudi Arabia

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model</th>
<th>Coefficient</th>
<th>SE</th>
<th>T</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current Expenditure</td>
<td>AR (1)</td>
<td>1.008</td>
<td>0.0450</td>
<td>22.39</td>
<td>0.000</td>
</tr>
<tr>
<td>Capital Expenditure</td>
<td>AR (1)</td>
<td>0.9332</td>
<td>0.0633</td>
<td>14.75</td>
<td>0.000</td>
</tr>
<tr>
<td>Total Expenditure</td>
<td>AR (1)</td>
<td>1.0032</td>
<td>0.0416</td>
<td>24.09</td>
<td>0.000</td>
</tr>
</tbody>
</table>
Results presented in Table 3.5 clearly show that all three fits are very satisfactory as their corresponding p-values are 0.000 each. We will use AR (1) for these three variables to generate forecasts in the next chapter.

CHAPTER 4

EVALUATION OF FORECASTS BY REGRESSION AND TIME SERIES METHODS USING CROSS VALIDATION

In this section our main objective is to evaluate forecasts made by different regression and time series methods. We would employ the cross validation technique for doing this.

4.1 Cross Validation in Regression and Time Series Models
Cross-validation is a technique for assessing how the results of a statistical analysis will 
generalize to an independent data set. It is mainly used in settings where the goal is prediction, 
and one wants to estimate how accurately a predictive model will perform in practice. One round 
of cross-validation involves partitioning a sample of data into complementary subsets, 
performing the analysis on one subset (called the training set), and validating the analysis on the 
other subset (called the validation set or testing set). An excellent review of different type of 
cross validation techniques is available in Izenman (2008). Picard and Cook (1984) developed all 
basic fundamentals of applying cross validation technique in regression and time series.

According to Montgomery et al. (2008), three types of procedures are useful for 
validating a regression or time series model.

(i) Analysis of the model coefficients and predicted values including comparisons with prior 
experience, physical theory, and other analytical models or simulation results,

(ii) Collection of new data with which to investigate the model’s predictive performance,

(iii) Data splitting, that is, setting aside some of the original data and using these observations to 
investigate the model’s predictive performance. Since we have a large number of data set, we 
prefer the data splitting technique for cross-validation of the fitted model.

In randomly missing data, generally three performance indicators; say, mean absolute error 
(MAE), root mean square error (RMSE) and estimated bias (EB) are considered to examine the 
accuracy of theses imputation methods. In order to select the best method for estimation missing 
values, the predicted and observed data were compared. The mean absolute error is the average 
difference between predicted and actual data values, and is given by

\[ MAE = \frac{1}{N} \sum_{i=1}^{N} |P_i - O_i| \] 

(4.1)
where $N$ is the number of imputations, $P_i$ and $O_i$ are the imputed and observed data points, respectively. MAE varies from 0 to infinity and perfect fit is obtained when MAE = 0.

The root mean squared error is one of the most commonly used measure and it is computed by

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (P_i - O_i)^2}$$  \hspace{1cm} (4.2)

The smaller is the RMSE value, the better is the performance of the model.

The estimated bias is the absolute difference between the observed and the estimated value of the respective parameters and defined as

$$EB = |O_i - E_i|$$  \hspace{1cm} (4.3)

where $E_i$ is the estimated value of the parameter that obtained from the imputation methods.

In order to find out the best prediction model we usually leave out say, $l$ observations aside as holdback period. The size of $l$ is usually 10% to 20% of the original data. Suppose that we tentatively select two models namely, $A$ and $B$. We fit both the models using $(T - l)$ set of observations. Then we compute

$$MSPE_A = \frac{1}{l} \sum_{i=1}^{l} e_{Ai}^2$$  \hspace{1cm} (4.4)

for model $A$ and

$$MSPE_B = \frac{1}{l} \sum_{i=1}^{l} e_{Bi}^2$$  \hspace{1cm} (4.5)

for model $B$. Several methods have been devised to determine whether one MSPE is statistically different from the other. One such popular method of testing is the $F$-test approach, where $F$-statistic is constructed as a ratio between the two MSPEs keeping the larger MSPE in the numerator of the $F$-statistic. If the MSPE for model $A$ is larger, this statistic takes the form:
\[ F = \frac{MSPE_A}{MSPE_B} \] (4.6)

This statistic follows an \( F \) distribution with \((l, l)\) degrees of freedom under the null hypothesis of equal forecasting performance. If the \( F \)-test is significant we will choose model \( B \) for this data otherwise, we would conclude that there is a little bit difference in choosing between these two models.

### 4.2 Generation of Forecasts for Saudi Arabia Expenditure Data

Using Cross Validation

To evaluate forecasts we carry out an experiment using cross validation. Here we compare four different methods: the classical OLS, the robust LMS, the nonparametric LOWESS and the ARIMA \((1, 0, 0)\) or the AR \((1)\) model. Since we have 44 observations, we use 36 of them (roughly 80\%) as the training set and left the last 8 observations for cross validation. We have produced ten sets of results for each of the three variables. Three of them are generated from the OLS, three from the LMS, three from the LOWESS and one from AR \((1)\) and the results are presented in Tables 4.1 – 4.3.

#### Table 4.1: Original and Forecast Values for Current Expenditure of Saudi Arabia

<table>
<thead>
<tr>
<th>Original</th>
<th>OLS</th>
<th>LMS</th>
<th>LOWESS</th>
<th>AR(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR</td>
<td>OrR</td>
<td>TR</td>
<td>OR</td>
</tr>
<tr>
<td>284173</td>
<td>322180</td>
<td>205037</td>
<td>337241</td>
<td>663722</td>
</tr>
<tr>
<td>322411</td>
<td>374343</td>
<td>233664</td>
<td>394648</td>
<td>791433</td>
</tr>
<tr>
<td>347199</td>
<td>352271</td>
<td>268327</td>
<td>378435</td>
<td>737394</td>
</tr>
<tr>
<td></td>
<td>Original</td>
<td>OLS</td>
<td>LMS</td>
<td>LOWESS</td>
</tr>
<tr>
<td>--------</td>
<td>----------</td>
<td>---------</td>
<td>---------</td>
<td>----------</td>
</tr>
<tr>
<td></td>
<td>OR</td>
<td>OrR</td>
<td>TR</td>
<td>OR</td>
</tr>
<tr>
<td>62301</td>
<td>150090</td>
<td>63127</td>
<td>142841</td>
<td>36017.5</td>
</tr>
<tr>
<td>70911</td>
<td>177770</td>
<td>69550</td>
<td>169193</td>
<td>41222.8</td>
</tr>
<tr>
<td>119049</td>
<td>166058</td>
<td>77326</td>
<td>161751</td>
<td>39020.3</td>
</tr>
<tr>
<td>131230</td>
<td>282725</td>
<td>102567</td>
<td>272175</td>
<td>60959.7</td>
</tr>
<tr>
<td>179840</td>
<td>130666</td>
<td>73760</td>
<td>129699</td>
<td>32364.9</td>
</tr>
<tr>
<td>198842</td>
<td>195995</td>
<td>71008</td>
<td>185565</td>
<td>44650.1</td>
</tr>
<tr>
<td>276200</td>
<td>296850</td>
<td>79248</td>
<td>276224</td>
<td>63615.8</td>
</tr>
<tr>
<td>261679</td>
<td>327447</td>
<td>92307</td>
<td>307459</td>
<td>69369.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Original</th>
<th>OLS</th>
<th>LMS</th>
<th>LOWESS</th>
<th>AR(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR</td>
<td>OrR</td>
<td>TR</td>
<td>OR</td>
<td>OrR</td>
</tr>
<tr>
<td>346474</td>
<td>472094</td>
<td>267265</td>
<td>480024</td>
<td>388296</td>
<td>257475</td>
</tr>
<tr>
<td>393322</td>
<td>552038</td>
<td>302203</td>
<td>563893</td>
<td>438841</td>
<td>297168</td>
</tr>
<tr>
<td>466248</td>
<td>518211</td>
<td>344504</td>
<td>540207</td>
<td>417454</td>
<td>345228</td>
</tr>
<tr>
<td>520069</td>
<td>855157</td>
<td>481811</td>
<td>891641</td>
<td>630488</td>
<td>501225</td>
</tr>
<tr>
<td>596434</td>
<td>415998</td>
<td>325104</td>
<td>438199</td>
<td>352830</td>
<td>323187</td>
</tr>
</tbody>
</table>
Now we employ some methods to evaluate forecast. At first we compute the MSPE values as defined in (4.4) and the results are presented in Table 4.4.

Table 4.4: MSPE Values for Current, Capital and Total Expenditure of Saudi Arabia

<table>
<thead>
<tr>
<th>Model</th>
<th>Current Expenditure</th>
<th>Capital Expenditure</th>
<th>Total Expenditure</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS (OR)</td>
<td>7.67616E+09</td>
<td>6.43303E+09</td>
<td>2.60067E+10</td>
</tr>
<tr>
<td>OLS (OrR)</td>
<td>3.06088E+10</td>
<td>1.22046E+10</td>
<td>8.06584E+10</td>
</tr>
<tr>
<td>OLS (TR)</td>
<td>1.09149E+10</td>
<td>5.32766E+09</td>
<td>3.00074E+10</td>
</tr>
<tr>
<td>LMS (OR)</td>
<td>3.60817E+11</td>
<td>1.75767E+10</td>
<td>2.07238E+10</td>
</tr>
<tr>
<td>LMS (OrR)</td>
<td>1.77716E+10</td>
<td>2.27570E+10</td>
<td>7.97373E+10</td>
</tr>
<tr>
<td>LMS (TR)</td>
<td>1.87280E+11</td>
<td>4.96611E+10</td>
<td>1.78768E+10</td>
</tr>
<tr>
<td>LOWESS (OR)</td>
<td>5.87433E+09</td>
<td>3.19672E+09</td>
<td>1.62306E+10</td>
</tr>
<tr>
<td>LOWESS (OrR)</td>
<td>3.06088E+10</td>
<td>5.83689E+09</td>
<td>3.54981E+10</td>
</tr>
<tr>
<td>LOWESS (TR)</td>
<td>5.93560E+09</td>
<td>3.26517E+09</td>
<td>1.65486E+10</td>
</tr>
<tr>
<td>AR (1)</td>
<td>3.72173E+10</td>
<td>2.45063E+10</td>
<td>1.18662E+11</td>
</tr>
</tbody>
</table>

Table 4.5: MSPE Values for Current, Capital and Total Expenditure of Saudi Arabia

<table>
<thead>
<tr>
<th>Model</th>
<th>Current Expenditure</th>
<th>Capital Expenditure</th>
<th>Total Expenditure</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS (OR)</td>
<td>1.307 (0.704)</td>
<td>2.012 (0.314)</td>
<td>1.602 (0.520)</td>
</tr>
</tbody>
</table>
If we carefully investigate the results given in Table 4.4 we will see that the LOWESS forecast based on the oil revenue produces the best set of result as it has the minimum MSPE values for each of the three response variables. But these results are difficult to compare and we cannot use them for any further statistical analysis. Hence we compute the F-statistic based on the MSPE’s as defined in (4.6) and the results together with their associated p-values are presented in Table 4.5.

Results presented here is the key finding of our research. We observe that for all three response variables LOWESS based on oil revenue performs the best, The performance of LOWESS based on total revenue is equally as good. Out of 30 F results presented here 12 of them (which are underlined) are not acceptable at the 5% level. Likewise the fitting, the LMS performs very poor in forecasting as 6 out of 9 LMS results are unacceptable. But the stunning result that we observe here is the performance of the ARIMA models. In time series it is

<table>
<thead>
<tr>
<th>Method</th>
<th>MSPE Oil Revenue</th>
<th>MSPE Total Revenue</th>
<th>MSPE LMS (OR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS (OrR)</td>
<td>5.211 (0.031)</td>
<td>3.818 (0.075)</td>
<td>4.970 (0.035)</td>
</tr>
<tr>
<td>OLS (TR)</td>
<td>1.858 (0.398)</td>
<td>1.667 (0.485)</td>
<td>1.849 (0.402)</td>
</tr>
<tr>
<td>LMS (OR)</td>
<td>61.423 (0.000)</td>
<td>5.498 (0.026)</td>
<td>1.277 (0.737)</td>
</tr>
<tr>
<td>LMS (OrR)</td>
<td>3.025 (0.138)</td>
<td>7.119 (0.012)</td>
<td>4.913 (0.037)</td>
</tr>
<tr>
<td>LMS (TR)</td>
<td>31.881 (0.000)</td>
<td>15.535 (0.000)</td>
<td>1.101 (0.895)</td>
</tr>
<tr>
<td>LOWESS (OR)</td>
<td>1.000 (1.000)</td>
<td>1.000 (1.000)</td>
<td>1.000 (1.000)</td>
</tr>
<tr>
<td>LOESS (OrR)</td>
<td>5.211 (0.031)</td>
<td>1.826 (0.412)</td>
<td>2.187 (0.289)</td>
</tr>
<tr>
<td>LOWESS (TR)</td>
<td>1.010 (0.989)</td>
<td>1.021 (0.977)</td>
<td>1.020 (0.978)</td>
</tr>
<tr>
<td>AR (1)</td>
<td>6.336 (0.017)</td>
<td>7.666 (0.009)</td>
<td>7.311 (0.011)</td>
</tr>
</tbody>
</table>
conventional to apply ARIMA models for forecasting but we observe in our study that all 3 AR(1) models are unacceptable in terms of generating forecasts.

CHAPTER 5
CONCLUSIONS AND AREAS OF FUTURE RESEARCH

In this chapter we will summarize the findings of our research to draw some conclusions and outline ideas for our future research.

5.1 Conclusions
My prime objective was to find the most appropriate models for fitting of regression and/or time series models and for generating forecasts. Although the OLS produces the line of best fit in terms of producing the minimum MSE, but there are some other aspects of fitting where the OLS
may not be that good. The OLS fit heavily relies on the normality assumption, but we observe in our study that this assumption is often violated for the OLS model. It is generally believed that robust regression fits the model in a better way since they are resistant to outliers. But our experience here is that not always the case. Of course they fit the best model for the majority of the data but may not be good overall. We are quite impressed with the performance of the local regression (LOWESS) method. Since it is a nonparametric method it does not rely on the normality assumption at all and it can produce very good fit and forecasts. In time series its almost conventional to fit the data by the ARIMA models, but we observe in our study that the ARIMA forecast may be much inferior to regression forecasts when it is possible to apply regression methods to generate forecasts.

5.2 Areas of Future Research
In our study we were not able to consider a variety of regression and time series methods due to time constraints. We only consider those methods which are reported as the best by other researchers. In future we would like to extend our research by considering a large number of alternative methods. Since our data is time series data volatility could be an essential part of it. We could not consider ARCH/GARCH or ARFIMA/GARFIMA models in our present study. In future we would like to extend our research in those areas as well.
REFERENCES


