CHAPTER 2

RESEARCH METHODOLOGY

The derivation of reliable forecasts is an important task in many areas. There are many different types of forecasting procedures, but no simple answer as to which is "best". Different forecasting problems require different treatments and thus the choice depends on a variety of considerations including the objectives of the study and the properties of the data. This chapter discusses the methods used to forecast tea yield.

2.1 The Data

The Study is based on tea statistics compiled by the Department of Statistics, Malaysia. Since tea is one of the earliest industrial crops in Peninsular Malaysia, data collection on the production of tea has been given importance. Tea statistics date back to the early thirties and time series data on tea is from 1947 onwards[1]. The data is collected during the Annual and Quarterly Census from tea estates. Information is available for hectareage planted, average hectareage in production, production of green leaves and made tea, yield, sales, stocks, imports and exports.

The study uses the series on the quarterly production of tea measured in tonnes and average hectareage in production for Peninsular Malaysia from

[1] Agricultural Statistics-Time Series (1988), Department of Statistics, Malaysia.

1960 to 1996. The quarterly production of tea is the total production of tea from both the Highlands and the Lowlands. The Highland tea comes from Cameron Highlands and Lowland tea comes from Perak, Southern Pahang, Selangor and Johor.

2.1.1 Source of Data

Statistics on Quarterly Tea production is published in the Statistical Monthly Bulletin. The quarterly production of tea from 1960 to 1970 is given in the unit of pounds (Lbs.). These figures were converted to tonnes using the following conversion: 1 tonne is equal to 2066 Lbs. Thereafter the production is quoted in Kilograms and then by Tonnes. The Department of Statistics, Malaysia, publishes the data on average hectareage in production in the Cocoa, Coconut and Tea Statistics Handbook.

Since 1990, the Department of Statistics started publishing the quarterly production of Green Leaves in Peninsular Malaysia instead of Made Tea. However, the study is on quarterly production of Tea. On request for quarterly production of Tea from the Department of Statistics, the department provided the figures of quarterly production of tea from 1992 to 1997 from their ledgers but data for the year 1990 and 1991 was unavailable. An expert opinion suggested that conversion rate can be used to convert the data from green leaves to tea. The department showed some statistics on conversion and suggested that the conversion rate is in the range of 22 per cent to 24 per cent. Some of the managers of the major tea plantations in Cameron Highlands who were interviewed agreed with this conversion rate. Based on the above evidence, the study used the following conversion rates: 22 percent for quarter 1, quarter 3, and quarter 4 and 23 percent for quarter 2, because these rates corresponded closely to the data set for the years where the actual amount of made tea and green leaves are known.

2.1.2 Measure of Yield

The production of tea per hectare, yield, measured as tonne per hectare, is calculated by taking the ratio of quarterly production of tea to average hectareage in production. Since only annual quarterly hectareage in production is available, the following is assumed in the calculation of production per hectare (yield) of tea:

(i) Quarterly average hectareage in production is assumed to be constant for every quarter of a given year.

(ii) Quarterly average hectareage in production is assumed to be equal to the annual average hectareage in production of that given year.

2.2 Identification Of Factors Affecting Tea Yield

The identification process begins by examining the trends in tea production in Peninsular Malaysia. It reviews the literature on factors that generally affect tea production or yield and then narrows down to factors affecting the production or yield of tea in Peninsular Malaysia. The identification process covers both uncontrollable and controllable factors affecting the yield.

2.3 Evaluating Patterns In Data

The analysis begins by providing descriptive statistics of yield series and then the plot of the yield series is examined. Decomposition analysis and Statistical Tests are used to check for patterns of trend, seasonality and cycles.

2.3.1 Examination of Plots

Generally, the examination of plots is to provide visual evidence of the existence and the behaviour of the series or patterns making up the series. The examination of the tea yield series plot is to provide visual evidence of the patterns making up the yield series. Examination of individual plots of the patterns making up the series is to show the behaviour of patterns separately and provide evidence of their existence.

2.3.2 Decomposition Method

Developed in the 1920s, the classical time series decomposition procedure provides information on the patterns in the data. Graphical insights into the behaviour of a time series from the decomposition process can help in identifying the structure of a series and hence in the selection of appropriate models for forecasting. The general mathematical representation of the decomposition approach is (see, for example, Makridakis, Wheelwright and Hyndman, 1998, pp 80-126) :-

$Y_t = F(S_t\,, T_t\,, E_t)$,

where

Y_t is the time series value (yield) at period t,

S_t is the seasonal component (or index) at period t,

T_t is the trend-cycle component at period t, and

E_t is the irregular (or remainder) component at period t.

The components can be combined in one of the two ways:

(i) Additive Model: $Y_t = S_t + T_t + E_t$

(ii) Multiplicative Model: $Y_t = S_t .\ T_t .\ E_t$

The additive model is appropriate if the magnitude of the seasonal fluctuations does not vary with the level of the series. If the seasonal fluctuations increases or decreases proportionally with the increase or decrease in the level of the series, then the multiplicative model is appropriate.

A classical decomposition is carried out using the following four steps.

Step 1: The trend-cycle (T_t) is computed using a centered moving average.

$$T_t = \frac{1}{2}[(\frac{Y_{t-2}+Y_{t-1}+Y_t+Y_{t+1}}{4}) + (\frac{Y_{t-1}+Y_t+Y_{t+1}+Y_{t+2}}{4})]$$

Step 2: The de-trended series is computed by removing the trend-cycle component from the data, leaving the seasonal and irregular terms. That is,

Additive model: $Y_t - T_t = S_t + E_t$

Multiplicative model: $S_t \,.\, E_t = \frac{Y_t}{T_t}$

Step3: Once the trend-cycle component has been removed, the seasonal component is relatively easy to estimate. In classical decomposition, it is assumed that the seasonal component is constant from year to year. So only one seasonal value (index) is calculated for each season. This is the adjusted average of the detrended values for each season. The values are adjusted so that the sum of the seasonal indices is zero (additive model) or one (multiplicative model).

Step4: The irregular series E_t is computed by subtracting the estimated seasonality and trend-cycle from the original data series.

Additive model: $E_t = Y_t - T_t - S_t$

Multiplicative model: $E_t = Y_t /(T_t.S_t)$

A time series plot of these components gives the best-visualised effect of these components.

2.3.3 Statistical Tests

The significance of the components can be determined by statistical tests (see, Farnum and Stanton, 1989). The ones used in this study are non-parametric tests and are described below. The reason for using non-parametric test is that, it is free from distribution assumptions, Bradley (1968).

2.3.3.1 Run Test for Trend

The run test (see, for example, Farnum and Stanton, 1989, pg. 57-60) is based on the premise that any observations from a horizontal time series with independent error terms is equally likely to be above or below the median of the series. To run the test, we first compute the median of the series, then assign a plus to observations above the median and a minus sign to observation below it. Finally we list the pluses and minuses in the chronological order and count the number of runs, or blocks of pluses and minuses.

The number of pluses and minuses depends on whether the series has an even or an odd number of observations. If it is odd, the median, which is itself an observation, is ignored and we get $\frac{(n-1)}{2}$ pluses and $\frac{(n-1)}{2}$ minuses. Otherwise we get $\frac{n}{2}$ of each. Suppose we let:

m= the numbers of +'s = the number of -'s

= $\frac{(n-1)}{2}$, if n is odd

or

= $\frac{n}{2}$, if n is even

Then a random no-trend series should produce a random string of pluses and minuses, that is, a string with neither too few nor too many runs. A series with trend, or one with large autocorrelation, will tend to have fewer runs; a series with negative autocorrelation will tend to have many more runs.

The statistic R = the numbers of runs in a random sequence of m pluses and m minuses have the following mean and standard deviation (see, Hogg and Craig, 1995, pg.519-520).

μ_R = expected number of runs = m+1

σ_R = standard deviation of the number of runs

$= \sqrt{\frac{m(m-1)}{(2m-1)}}$

Decision rule: $z = \frac{|R-\mu_R|}{\sigma_R}$, Reject H_o if $|z| > z_{\alpha/2}$

Decision rule: Reject H_o if $R > R_U$ or $R < R_L$

2.3.3.2 Kruskal-Wallis for Seasonality

The test (see, for example, Farnum and Stantom, 1989) is that if the specific seasonals are purely random with no seasonality, their distribution should be the same in all L seasons. Then, if ranks are assigned to specific seasonals, a given rank should be likely to fall in one season as in another. There should not be a preponderance of, say, low ranks in one season, so the average ranks in all seasons should be about the same, i.e., within sampling variability of one another. If we let

R_i = the sum of the ranks of the Y_i 's in the j^{th} season

n_i = the number of specific seasonals in the j^{th} season

n = the total number of specific seasonals = $n_1 + n_2 + \ldots + n_L$

then the H statistic used to test the randomness hypothesis is the weighted sum of squared differences between the average ranks within seasons and average rank.

$$H = \frac{12}{n(n+1)} \sum \frac{R_j^2}{n_j} - 3(n+1)$$

If the specific seasonals are random, H will likely to be fairly small; so if H is too large we would reject randomness in favor of seasonality. When the null hypothesis is true and the n_l 's are even moderately large, H is known to follow approximately a chi-square distribution, Kruskal and Wallis (1952), with L -1 degrees of freedom, so the critical values for test may be found in statistical tables.

2.3.3.3 Quarterly Cyclical Dominance(QCD)

Using the cyclical-irregular term in forecasting requires knowledge of the relative importance of its two components, C_t and ε_t. If the cycle dominates the noise, then the cyclical-irregular term will be useful for forecasting. If ε_t dominates, then the cyclical-irregular term, being primarily noise, has zero forecast value.

One way of evaluating the relative strength of the cyclical component is by comparing the trend-cycle to the noise in the series. The procedure for calculating the quarters for cyclical dominance(QCD) is given here (see, for example, Farnum and Stantom, 1989).

Given: A sample of n observations, Y_t; t = 1, 2, . . .,n

A set of normalised seasonal indexes, S_1, S_2, S_3, ..., S_L

The set of centered moving averages, MA_t, of the Y_t's.

Calculation of percentage changes for Trend-Cycle and Irregular term :

Multiplicative Model: $T_t C_t = MA_t$

Calculate percentage changes in MA

$\varepsilon_t = Y_t /(S_t Ma_t)$

Calculate percentage changes in ε_t.

Calculate the average absolute percentage change in the trend-cycle and irregular terms for 1, 2, 3,.... period time spans.

$$QCD = \frac{\text{Average absolute \% change in irregular term}}{\text{Average absolute \% change in irregular term in trend-cycle}}$$

Interpretation: QCD of 1 indicates that the trend-cycle is strong: A QCD beyond 6 indicates very weak or nonexistent cyclical, Skiskin (1984).

2.4 Data Transformation

An examination of the plot of the data provides information on whether data transformation is necessary. Increasing or decreasing variations in the fluctuations of the series in the plot indicates the need of data transformation. There are number of reasons why transformation may be necessary. One reason is that many statistical techniques require the assumption of normality of data. Data transformation may also be used to achieve stationarity in variance. That is, the variance of the error terms remains essentially constant or stable over time.

To stabilise the variance the common approach taken is to try out the common transformations based on some characteristics of the series and then view the plots to decide which transformations is best able to result in a stable variance. The commonly used transformations are the logarithmic, square, cube, square root and inverse. Sometimes, a certain transformation may be preferred for the reason that it is more interpretable. Chatfield and Prothero (1973a) used Log transformation but did not obtain a statisfactory model using Box-Jenkins methods. Several discussants of the Chatfield and Prothero (1973a) paper, Box and Jenkins (1973), Harrison (1973) and Wilson (1973) suggested that a more flexible parametric family of transformations, introduced by Box and Cox (1964) for stabilising the variance should be considered to improve applicability of statistical models rather than making a decision based on visual of plots of the commonly used transformation.

Box and Cox (1964), indicated that the choice of this preliminary transformation is of critical importance, particularly in seasonal time series. For a given value of the parameter λ, the transformation is defined by :

$$Y^*_t = \begin{cases} (Y_t^{\lambda} - 1)/\lambda, & \lambda \neq 0 \\ \mathrm{Log}(Y_t), & \lambda = 0 \end{cases}$$

Where for our purposes Y_t, assumed to be positive, denotes a non-stationary time series and λ is the transformation parameter. The parameter λ is chosen by the user, is used to achieve normality and to stabilise the variance. The values of $\lambda < 1$ are useful for positively skewed data $\lambda > 1$ for negatively skewed data. A modified form of this transformation is usually employed to preserve the original order in the data, is given by Makridakis, Wheelwright and Hyndman (1998) :

$$Y^*_t = \begin{cases} -Y_t^{\lambda} & \lambda < 0 \\ \operatorname{Log}(Y_t), & \lambda = 0 \\ Y_t^{\lambda} & \lambda > 0 \end{cases}$$

In this study, we use the method proposed by Victor (1993) for selecting a value for the parameter λ. This procedure is carried out by grouping N observations of the series into H subseries, so that a local estimate of mean and variance within subseries can be obtained. Let $z_{h,r}$ be the r th observation of subseries h. We then compute

$$\bar{z}_h = \frac{\sum_{r=1}^{R} z_{h,r}}{R}, \qquad S_h = [\frac{\sum_{r=1}^{R} (z_{h,r} - \bar{z}_h)^2}{R-1}]^{1/2}$$

and λ should be chosen in such a way that $\frac{S_h}{\bar{z}_h^{(1-\lambda)}} = a$ where h =1,2...H

holds for some constant a >0. This procedure uses the following approaches to decide the optimal value for variance stabilising parameter of power transformation.

Approach 1 : Minimising relative variation

An empirical interpretation of the equation $\frac{S_h}{\bar{z}_h^{(1-\lambda)}} = a$ leads us to look

for a λ value such that the ratios $\frac{S_h}{\bar{z}_h^{(1-\lambda)}}$ show minimum variation across the h subseries. We select the power by looking for the smallest coefficient of variation(CV) of $\frac{S_h}{\bar{Z}_h^{(1-\lambda)}}$ as a function of λ.

Approach 2 : Estimating Linear regression in logarithms

Another empirical interpretation of condition $\frac{S_h}{\bar{Z}_h^{(1-\lambda)}}$ = a can be put in the form of the following simple regression model:

$$\text{Log}(S_h) = \text{Log}(a) + (1-\lambda)\text{Log}(\bar{Z}_h) + \varepsilon_h, \qquad h = 1, 2, 3, \ldots, H$$

with the ε_h's being a random sample of errors uncorrelated with Log($\bar{Z}_h$).It is clear that from this form λ can be estimated by ordinary least squares method.

The study would use the second approach to find the best variance stabilising parameter of power transformation.

2.5 Forecasting Methods

Forecasting methods can be grouped into: quantitative and qualitative methods. Table 2.1 summaries this categorisation. Quantitative methods require numerical past data. In both approaches methods vary by whether the variable under study is seen as a function of its past values or whether other variables play a role as well. The methods used in this study are quantitative forecasting procedures as there is sufficient quantified data available about the yield of tea in Peninsular Malaysia.

Table 2.1
CLASSIFICATION OF FORECASTING TECHNIQUES

Category	Condition of Applicability	Process	Forecasting Technique	Assumption
Quantitative Forecasting	Information of past. Quantified in the form of numerical data	History repeats itself	**Time Series Methods:** Decomposition ; Exponential Smoothing; ARIMA; Filters; Leading Indicators	Some aspects of past pattern will continue into the future.
		External and Internal Factors determine events	**Explanatory Methods:** Regression; Econometric models; Multivariate ARMA; Input/Output	
Qualitative Forecasting	Little or no quantitative information. Sufficient qualitative knowledge required.	History repeats itself	**Exploratory methods:** Anticipatory surveys; Catastrophe theory; Delphi; Historical analogies; Life cycle analysis	Some aspects of past pattern will continue into the future.
		External and Internal Factors determine events	**Normative methods:** Cross-impact matrices; Relevance trees; Delphi; System dynamics; market research	

Source : Markridakis and Wheelwright (1979), pg. 4

Quantitative methods can be further grouped into time series and explanatory models. There has been considerable debate about the comparative performance of these models in forecasting. Studies done by Stekler (1968) suggest that economic models have not been entirely successful in forecasting economic activity, especially during the sixties. Labys and Pollak (1984) also expressed similar views on the use of econometric models for modelling commodities. The same period also saw the development of certain time series methods. There was an edge to look for an alternative to econometric models to improve forecast values. This has led to more attention being given to Exponential Smoothing also known as the Holt-Winter (1960) method and the parametric modelling of time series which developed by Box and Jenkins (1970). Table 2.2 outlines the basic differences between the two approaches.

Table 2.2
Characteristics Of Explanatory And Time Series Univariate Models

Explanatory Models	Time Series Models
Explanatory models assume that the variable to be forecasted exhibits an explanatory relationship with other variables, including its own past.	Time series forecasting treats the system as a black box and makes no attempt to discover the factors affecting its behavior. Therefore, prediction of the future is based on past values of a variable.
The objective of the explanatory model is to discover the form of the relationship and use it to forecast future values of the forecast variable.	The objective of time series models is to discover the pattern in the historical data series and extrapolate that pattern into future.
The models assume a specific relationship between dependent and independent variables.	Time series forecasting treats the system as a black box, and makes no attempt to discover the factors affecting its behaviour.

Source: Makridakis, Wheelwright and Hyndman (1998), pg. 11.

The focus in this Study is on the use of time series models, for the following reasons:

(i) The objective of the study is to generate forecasts and understand some of the basic time structured patterns in the yield of tea.

(ii) A time series model can be easily implemented by organisations since only yield information needs to be maintained.

(iii) Explanatory models for tea yield would require information on independent variables that is not easily available.

Exponential smoothing models and ARIMA models are then used to generate forecasts for tea yield.

2.5.1 Exponential Smoothing Method

This study is concerned with a variant of exponential smoothing, which is often known as the Holt-Winters method (see, for example, Makridakis, Wheelwright and Hyndman, 1998 pg.160-168). This method allows data to be modelled by a local mean, a local trend and a local seasonal factor, which are all updated by exponential smoothing. The seasonal effect may be additive or multiplicative.

The Holt-Winters Method

The study denotes the local mean level, trend and seasonal index at t by L_t , T_t and S_t respectively. Let α, γ and δ denote the smoothing parameters for updating the mean level, trend and seasonal index respectively and let p denote the number of observations per seasonal cycle. The formulae for updating L_t , T_t and S_t when a new observation Y_t becomes available, is given in Table 2.3

Table 2.3: Holt-Winters' Updating Equations

	Additive Seasonality	Multiplicative Seasonality
Level	$L_t = \alpha(Y_t - S_{t-p}) + (1-\alpha)(L_{t-1} + T_{t-1})$	$L_t = \alpha(Y_t / S_{t-p}) + (1-\alpha)(L_{t-1} + T_{t-1})$
Trend	$T_t = \gamma(L_t - L_{t-1}) + (1-\gamma)T_{t-1}$	$T_t = \gamma(L_t - L_{t-1}) + (1-\gamma)T_{t-1}$
Seasonal	$S_t = \delta(Y_t - L_t) + (1-\delta)S_{t-p}$	$S_t = \delta(Y_t / L_t) + (1-\delta)S_{t-p}$

Source: Makridakis, Wheelwright and Hyndman (1998), pg.164-168. .

Then the new forecast, $\hat{Y}_t(k)$ made at time t of the values k periods ahead is given in Table 2.4.For k=1,2 .

Table 2.4: Holt-Winters' Forecast Equations

	Additive Seasonality	Multiplicative Seasonality
Forecast	$\hat{Y}_t(k) = L_t + kT_t + S_{t-p+k}$	$\hat{Y}_t(k) = (L_t + kT_t)S_{t-p+k}$

Source: Makridakis, Wheelwright and Hyndman (1998), pg.164-168.

In order to implement this method, the user must

(a) decide whether to use Holt-Winters additive or multiplicative method,

(b) obtain starting or intial values for L_t , T_t and S_t at the beginning of the series, (see, Bowerman and O'Connell, 1987, pg. 273-276).

(c) the values of α, γ and δ are chosen such it gives the minimum MSE or MAPE,

(d) updating and forecasting is performed using the values in part (a) and (b) in sample range,

(e) forecasts are generated by further updating into the post-sample range using updating equations and forecast equation.

2.5.2 Box-Jenkins approach to univariate model building

(a) The general form of the integrated autoregressive-moving average (ARIMA) structure (see, for example, Bowerman and O'Connell, 1987, pg. 100) is

$$\phi_p(B)\ \phi_P(B^L)(1-B)^d(1-B^L)^D\ Y_t = \delta + \theta_q(B)\ \theta_Q(B^L)a_t,$$

where

B is the backshift operator ($BY_t = Y_{t-1}$);

$\phi_p(B) = (1- \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)$ are the non seasonal regular autoregressive parameters;

$\phi_P(B^L) = (1 - \phi_{1,L}B^L - \phi_{2,L}B^{2L} - \ldots - \phi_{P.L}B^{PL})$ are the seasonal autoregressive parameters;

$(1 - B)^d$ is the difference term of order d;

$(1 - B^L)^D$ is the seasonal difference term of order D;

$\delta = \mu \phi_p(B)\ \phi_P(B^L)$ is the constant term or deterministic trend constant and

μ is the true mean of the stationary time series being modelled;

$\theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q)$ are the non seasonal regular moving average terms;

$\theta_Q(B^L) = (1 - \theta_{1,L}B^L - \theta_{2,L}B^{2L} - \ldots - \theta_{Q.L}B^{QL})$ are the seasonal moving average terms;

a_t are random shocks that are assumed to be statistically independent of each other; each is assumed to have been randomly selected from a normal distribution that has mean zero and a variance that is the same for each and every time period t;

Y_t is the value of the series or transformed series at time t.

Considering the general Box-Jenkins model, it can be shown (Box and Jenkins, 1976) that constant term δ and the autoregressive and differencing operators $\phi_p(B)$ $\phi_P(B^L)(1-B)^d(1-B^L)^D$ determine the basic nature of the forecasts provided by this model, and it can also be shown (Box and Jenkins, 1976) that the moving average operators $\theta_q(B)$ $\theta_Q(B^L)$ determines how previous random shocks (or residuals, which are predictions of previous random shocks) modify the basic nature of the forecast

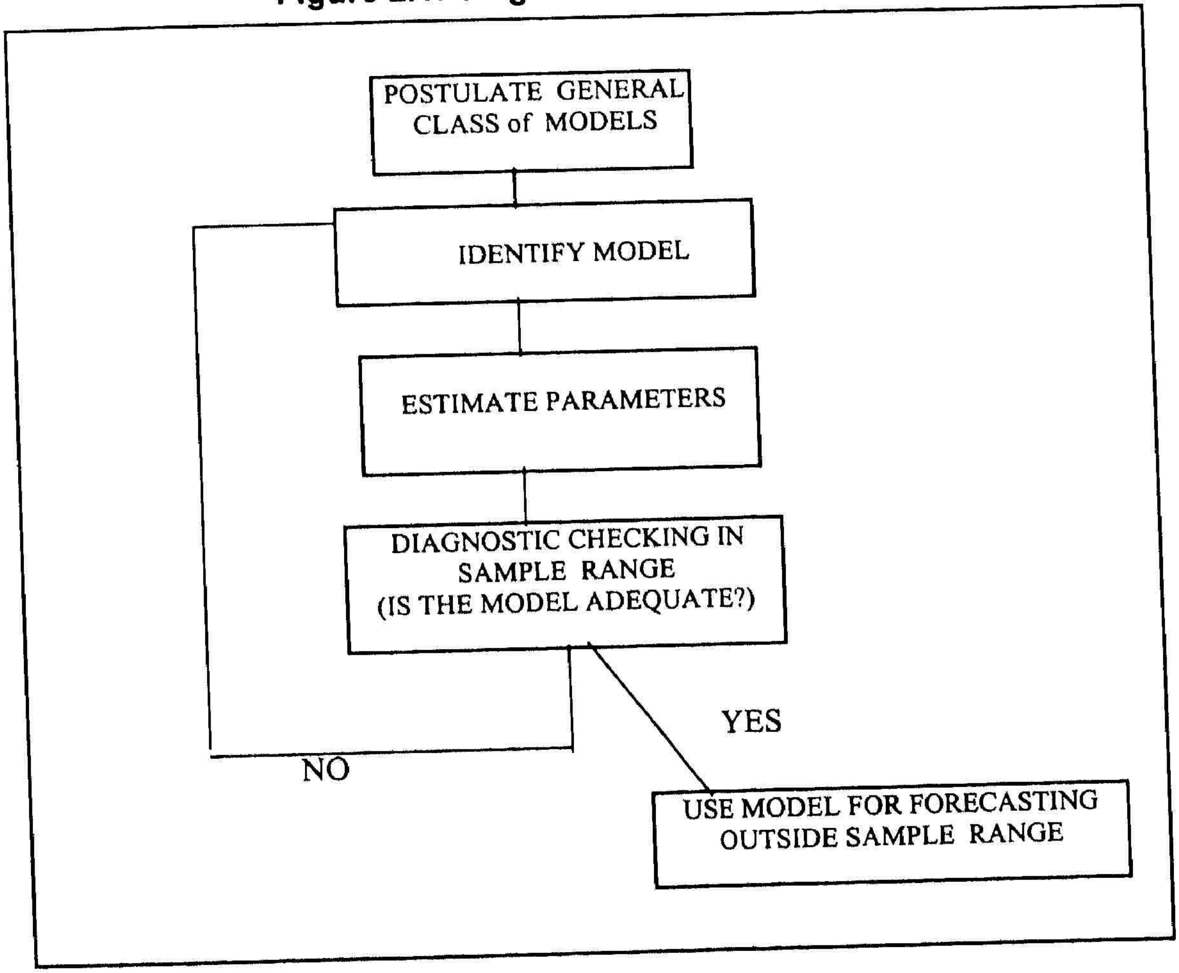
The Box-Jenkins methodology requires that the model to be used in describing and forecasting a time series be both stationary and invertible. The stationarity and invertibility conditions on the parameters forms of the operators $\phi_p(B)$, $\phi_P(B^L)$, $\theta_q(B)$ and $\theta_Q(B^L)$ are complicated (Hamilton,1994) and will not be given here. However, it can be said that

(a) A necessary but not sufficient stationarity condition on parameters of any form of each of the operator $\phi_p(B)$ and $\phi_P(B^L)$ is that the sum of the values of the parameter in the operator is less than 1

(b) A necessary but not sufficient invertibility condition on parameters of any form of each of the operator $\theta_q(B)$ and $\theta_Q(B^L)$ is that the sum of the values of the parameter in the operator is less than 1

Developing an accurate but parsimonious ARIMA model of this general form requires a three-stage iterative process. Figure 2.1 shows the stages in the iterative approach to model building Box and Jenkins (1976):

Figure 2.1: Stages In Model Building



POSTULATE GENERAL CLASS of MODELS
IDENTIFY MODEL
ESTIMATE PARAMETERS
DIAGNOSTIC CHECKING IN SAMPLE RANGE (IS THE MODEL ADEQUATE?)
YES
NO
USE MODEL FOR FORECASTING OUTSIDE SAMPLE RANGE

Source: Box and Jenkin (1976), pg 19.

(a) Model Identification: Examination of the data to see which model in the class of ARIMA processes appears to be most appropriate. The principal tools used in identification are sample autocorrelation function and the partial autocorrelation function. Since these functions exist only for stationary series, it is necessary to manipulate the original time series until it can assume to be stationary. In other words the data will fluctuate around a constant mean, independent of time. One way of removing non-stationarity in mean is through the method of differencing.

$$Z_t = \nabla_L^D \nabla^d Y_t ,$$

or

$$Z_t = \nabla_L^D \nabla^d Y^*_t ,$$

where ∇ operator for differencing,

∇_L seasonal operator for seasoanl differencing,

d is the degree of non-seasonal differencing used and D is the degree of seasonal differencing used,

Y_t is the original series where the series is stationary in variance,

Y^*_t represents the transformed series to stabilise variance.

In general to determine a particular differencing as given above, one can rely on the visual plot of a time series. This is often enough to convince a forecaster that the data is stationary or non-stationary in mean. But, this does not give the exact degree of differencing required for stationarity. The autocorrelation function (ACF) and partial autocorrelation function (PACF) plot can suggest non-stationarity in the mean. If the time plot shows the data scattered horizontally around a constant mean, the autocorrelations of data drop to zero relatively quickly. If the time plot is not horizontal, or the autocorrelations do not drop to zero, non-stationarity in mean is implied. The degree of differencing for non-stationary data is determined by carrying out the appropriate non-seasaonal and seasonal differencing till the autocorrelations of data drop to zero relatively quickly. When stationarity has been achieved, examine the autocorrelations to see for dominant spikes in ACF and PACF plot. The pattern of these dominant spikes is then used to decide the tentative model.

(b) Estimation: Having made tentative model identification, maximum likelihood estimation is carried to estimate the parameters. The method of maximum likelihood finds the values of the parameters which maximise the likelihood function, L. These estimates are found iteratively. EViews program on estimation of parameters uses this method of estimation. This method is usually favoured because it has some desirable statistical properties (see Box, Jenkins and Reinsell, p. 225).

(c) Diagnostic Checking: Examine the estimated residuals from the fitted model to see if it is adequate. Ljung-Box Test (Ljung and Box, 1978) is applied to residuals to test for white noise. If the Ljung-Box Test (Q-Statistic) is insignificant, then the residuals can be considered as white noise and the model is adequate.

Test Procedure: Ljung-Box Test

H_0: $\rho_k = 0$ for all $k \leq m$

H_0: $\rho_k \neq 0$ for some value of $k \leq m$

Test Statistic: $Q_m = n(n+2)\sum_{k=1}^{m}\frac{r_k^2}{n-k}$

m = the number of coefficients being tested

n = the number of observations in the series

Decision Rule: (Level α, n, m)

Reject H_0 if $Q_m > \chi^2_\alpha(m)$

Otherwise do not reject H_0

If there may be more than one model identified with white noise then there is a need to determine which one of them is to be preferred. A plausible criterion for choosing the best ARIMA model might appear to be to choose the model which gives the largest values for the log likelihood function or smallest Akaike's Information Criterion (AIC), (see, Makridakis, Wheelwright and Hyndman, 1998 pg. 360).

Log Likelihood Function (In L): The method of maximum likelihood finds the values of the parameters, which maximises the likelihood function L and its logarithm function (In L) at the same time (see Hogg and Craig, 1995, p.260-261). For ARIMA, the likelihood function is penalised for each additional term in the model. If extra term does not improve the likelihood function more than the penalty amount, it is not worth adding.

Akaike's Information Criterion (AIC): It is the penalised likelihood procedure. Let m = p + q + P +Q be the number of terms estimated in the model. Then we choose the values of p, q, P and Q by minimising AIC:

$$AIC = -2\log L + 2m$$

where, L denotes the likelihood.

If the first model appears to be inadequate for some reason, then other ARIMA models may be considered by repeating the above procedure until a satisfactory model is found.

2.6 Assessment Of Model Fit And Forecast Evaluation

Figure 2.2 shows the forecasting scenario of any forecasting methodology. Assessments of model fit are carried out for the sample range (1960 - 1996). Firstly, by examining the plot of Y_t (actual) versus F_t(forecast) to see for tracking ability of the model and then followed by examination of correlogram of forecast errors to determine whether it follows a white noise model using Ljung-Box Test (see, Diagnostic testing, Section 2.5.2).

The models are evaluated for forecast performance. The measures that will be used to carry out the evaluation are:

MAE (mean absolute error) = $\frac{1}{n}\sum_{t=1}^{n}| e_t |$

RMSE (root mean sum squared error) = $\sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^2}$

MAPE (mean absolute percentage error) = $\frac{1}{n}\sum_{t=1}^{n}| PE_t |$,

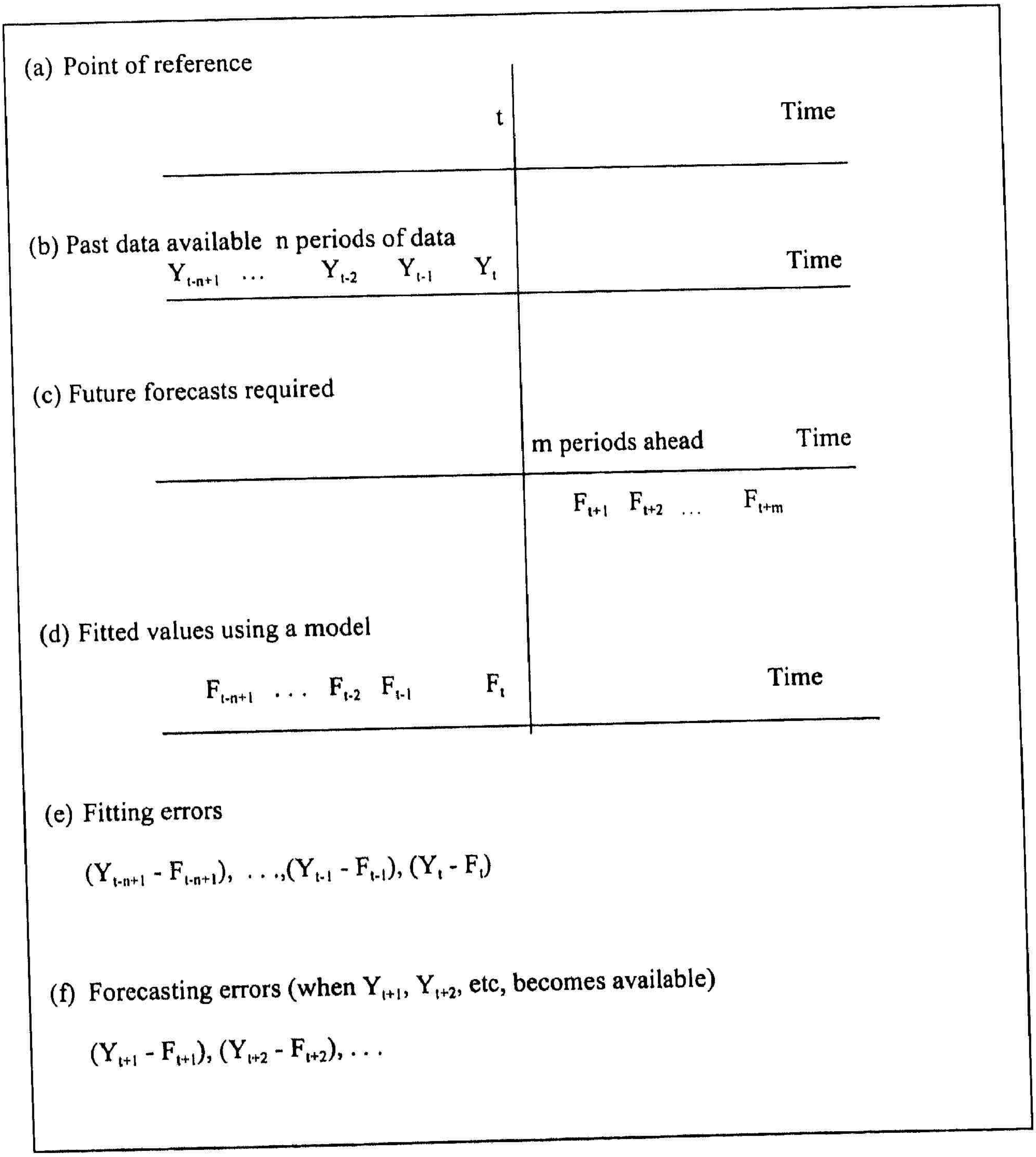
$$\text{where } PE_e = (\frac{Y_t - F_t}{Y_t}) \times 100$$

Y_t is the actual observation for time period t and F_t is the forecast for the same period, the error is defined as e_t = Y_t - F_t.

2.7 Forecasting Software

EViews is used for developing forecasting models. It is utilised for estimating, forecasting and assessing of models in this study. It is also used for plotting time series graphs. Excel is used to carry out ranking for specific seasonals in the Kruskal-Wallis test for Seasonality (see, Kruskal-Wallis test for seasonality, Section 2.3.3 (b)). It is also used to calculate H-Statistic in the Kruskal-Wallis test.

Figure 2.2: The Forecasting Scenario


(a) Point of reference
t
Time
(b) Past data available n periods of data
Y_{t-n+1} ... Y_{t-2} Y_{t-1} Y_t
Time
(c) Future forecasts required
m periods ahead
Time
F_{t+1} F_{t+2} ... F_{t+m}
(d) Fitted values using a model
F_{t-n+1} ... F_{t-2} F_{t-1} F_t
Time
(e) Fitting errors
$(Y_{t-n+1} - F_{t-n+1})$, . . .,$(Y_{t-1} - F_{t-1})$, $(Y_t - F_t)$
(f) Forecasting errors (when Y_{t+1}, Y_{t+2}, etc, becomes available)
$(Y_{t+1} - F_{t+1})$, $(Y_{t+2} - F_{t+2})$, . . .

Note: F_{t+1} F_{t+2} etc., refers to forecasted values of Y_{t+1}, Y_{t+2}
A fitted value, such as F_{t-1} could be represented as Y_{t-1} (Estimated value of Y_{t-1})
Source: Makridakis, Wheelwright and Hyndman (1998), pg. 139.