Chapter 6

Conclusion

Forecasting future production is a challenge for the decision-maker and knowledge of this has always promised advantages and opportunities of many kinds. Production can be affected by many factors, which in turn creates uncertainty in production and makes decision-making and planning for the future a difficult task. The forecaster's goal is to find a useful way to express a time-structured relationship so that forecasts of future production can be incorporated into the decision-making process.

This study developed models to generate forecasts for tea production. Forecasting plays an important role in tea production especially in the areas of production planning, marketing, labour recruitment and scheduling of work plans. Forecasts of yield can also help in determining the amount of tea that can be supplied by estates in Peninsular Malaysia to meet local demand, and hence the amount that has to be imported in the future.

The research is based on data for quarterly tea production and average hectareage in production in Peninsular Malaysia, using tea statistics compiled by the Department of Statistics, Malaysia. The first objective of this study was to examine some trends in the tea industry and understand the factors causing the fluctuations and patterns in the tea yield series. Next, the basic patterns in the yield series were examined by using Classical Decomposition and then the need for transformation of the data before modelling was carried. Thereafter forecasting models for quarterly tea production per hectare were developed using Exponential Smoothing and ARIMA models. The fit of these models was evaluated on sample forecast range (SFR) and the forecast performance of these models on the consecutive sample of observations that was not included in the estimation process, that is, post sample forecast range (PSFR).

6.1 Summary of Findings

The following are the main findings that the study has identified in the yield series of made tea in Peninsular Malaysia:

6.1.1 Trends In Tea Production And Factors Causing The Observed Variations

Some of the noted trends in the tea industry in Peninsular Malaysia are:

1. Over the years more highland tea has been produced. In the early sixties, about forty percent of the total production was lowland tea. By mid-nineties, production of lowland tea decreased to about eighteen percent of the total production.
2. The conversion ratio from green leaves to made tea remained constant and stable during the period 1960 to 1975, although in more recent years there has been slight changes due to machine plucking. On the average the conversion rate is 22.5 per cent.
3. The average hectareage in production from 1960 to 1996 has decreased. However, the production of made tea has increased during the same period.

4. Productivity per hectare has increased. Annual production of made tea per hectare (yield) in 1960 was 384 Kilograms per hectare and by the mid-nineties had increased to 2823 Kilograms per hectare
5. For the period 1960 to 1996, the statistics show that only small scale replanting and new planting has taken place.

Some of the identified factors that can be linked to the regular patterns in the yield series are given below. It is assumed that these following factors will continue to cause the regular patterns in the yield series in the future:

1. The upward trend will continue if replanting and new planting is carried out by planting new clonal varieties, which have high yield. Majority of the tea plants, which exist now, were all planted more than thirty years ago. If these tea plants are gradually replaced by high yielding varieties then this will bring an upward trend in the yield. Effective and efficient method in the application of fertilisers has also continued to increase yield. For, example Boh Tea Plantations, which is using aerial fertilisation, has seen an increase in yield and it has been able to overcome labour shortage problems. Improved control over weeds, pests and diseases in the coming years will also continue to give an upward trend.
2. The seasonal pattern in yield series in Peninsular Malaysia is caused by climatic conditions in the highlands, that is, a combined effect of rainfall, humidity and sunshine hours. This is because over the years the increasing proportion of tea is coming from Cameron Highlands. The seasonality pattern and the fluctuations in the yield have

changed over the years. In the early years of the sample, the fluctuations and seasonality were low as the yield is a combination of lowland and highland tea and these areas contributed an equal proportion towards total tea produced. After 1975, more highland tea was produced, and the proportion of highland tea production increased to about 80 to 90 percent in mid-nineties. Highland tea production is affected severely by climatic conditions and other factors such as pests and diseases, which result in high fluctuations and a seasonality effect, as seen in the yield series. This explains the high fluctuations and strong seasonality effect in the yield series after 1975.

6.1.2 Basic Patterns in the Yield Series

The regular patterns identified in the yield series of tea in Peninsular Malaysia are noted below:

1. There is increasing variation in yield series. This could be due to increased mean level.
2. There is a significant upward trend in yield over the sample period. This is linked to the improvements in the agronomic practices.
3. The yield series has a significant seasonal pattern. The peak occurs at second quarter and the trough during the first quarter. This is linked to climatic factors and improved agronomic practices.
4. Graphical plot of the residuals and the Quarterly Cyclical Dominance ratio spans shows that noise dominates the cycle.

6.1.3 Data Transformation

In view of the observed increasing variance in the yield series, an appropriate transformation was identified. The exact value for variance stabilising parameter of the Box-Cox power transformation to stabilise the increasing variance was λ = -0.67, but λ = -0.5 was taken for easier interpretation. Two other values of λ, λ = 0 and λ = -1 were tried, as the visual plots of these power transformation indicated that the variance was reasonably constant. Moreover, the use of alternative values permitted comparison with λ = -0.5 for tracking and forecasting performance. The transformed series based on these variance stabilising parameters of power transformation are LY_t (λ = 0), $Y1_t$ (λ = -0.5) and $Y2_t$ (λ = -1).

Abraham and Ledolter, (1986) provided some guidance on the interpretation of models. According to them, a logarithmic transformation is easy to explain, since difference of log transformed data approximate percentage changes. It is much more difficult to explain a transformed series for a variance stabilising parameter (λ) between 0 and 1 of power transformation. They provided one possible explanation for a λ between 0 and 1 is that the data may come from the model

$$Y_{t+l} = (\beta_o + \beta_1 l)S_l + a_{t+l}$$

Where, trend and the seasonal component are multiplicative, but the error is additive. If the variability in the first part on the right-hand side of the above equation dominates the error, one should consider a log transformation. If the error dominates, the parameter λ in Y_t^{λ} should be close to 1. If both components contribute to the variability, one can:

(a) consider a transformation Y_t^{λ} where λ is somewhere between 0 and 1;

(b) consider the original data and use Winters' multiplicative smoothing; or

(c) consider a seasonal ARIMA model with a simplifying operator $(1 - B^{12})^2$.

In this study the parameter λ in Y_t^{λ} is between 0 and 1 indicating the possibility that the yield data follows the above-said model and the study in fact considers the option (a) and (b) in the modelling process.

6.1.4 Forecasting

The Holt-Winters Method and Box-Jenkins Methodology were used to develop forecasting models and generate quarterly forecasts for the yield series.

For Holt-Winters method, the fitted models were:

1. Holt-Winters Multiplicative Seasonality (HWMS) with the three optimal smoothing constants fitted to the original yield series, Y_t
2. Holt-Winters Additive Seasonality with the three optimal smoothing constants fitted to the transformed series, LY_t, $Y1_t$ and $Y2_t$. Between the two methods, Holt-Winters Additive Seasonality on $Y1_t$ gives the best forecasting performance.

For Box-Jenkins methodology, the fitted ARIMA models are:

1. ARIMA$(0,1,1)(0,1,1)_4$ on LY_t,

2. ARIMA(0,1,1)(0,1,1)$_4$ on $Y1_t$

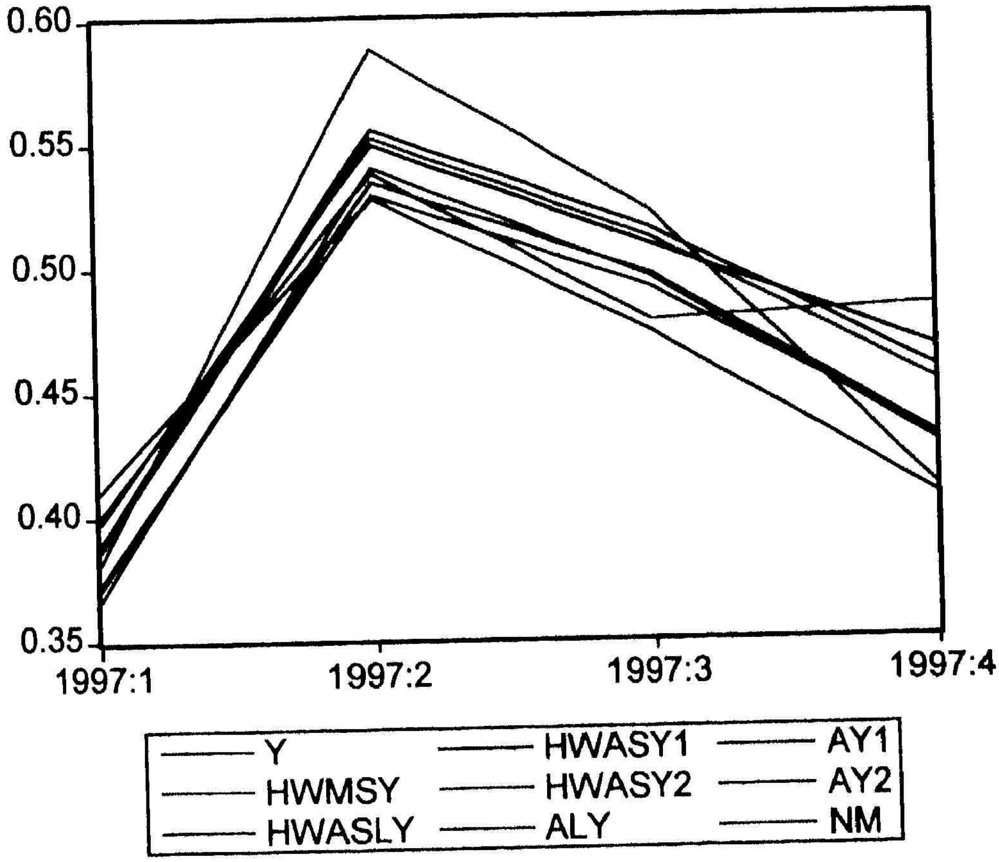
3. ARIMA(0,1,1)(0,1,1)$_4$ on $Y2_t$.

The forecasts generated by the above models are brought back to original scale before it is compared with actual quarterly yield of 1997. Table 6.1 shows the forecasted yield generated by Holt-Winters method and Box-Jenkins methodology in the post sample forecast range. Figure 6.1 shows the of the plot actual and the forecasted yield for all the fitted models in this study. Table 6.1 also includes the naïve model, where the forecasts for naïve model (NM) is done by taking the value of the previous quarterly yield.

Table 6.1: Forecast yield Generated By Holt-Winters method And ARIMA Models

PSFR -1997	Q1	Q2	Q3	Q4
Actual	0.410390	0.527636	0.473456	0.407115
HWMS (Y_t) (α= 0.46, γ = 0, δ =0.39	0.386691	0.552312	0.511290	0.453109
HWAS (LY_t) α= 0.45, γ = 0, δ =0.37	0.389160	0.555639	0.515292	0.458569
HWAS ($Y1_t$) α= 0.44, γ = 0, δ =0.23	0.398120	0.549524	0.508633	0.466886
HWAS ($Y2_t$) α= 0.47, γ = 0, δ =0	0.400725	0.537963	0.478024	0.483526
ARIMA(0, 1, 1)(0, 1, 1)$_4$ LY_t	0.366843	0.540249	0.495416	0.428295
ARIMA(0, 1, 1)(0, 1, 1)$_4$ $Y1_t$	0.371454	0.534506	0.493345	0.431126
ARIMA(0, 1, 1)(0, 1, 1)$_4$ $Y2_t$	0.373273	0.529193	0.491431	0.430839
Naïve Model(NM)	0.381618	0.588038	0.522234	0.410550

Figure 6.1: Plot Of Actual Yield And Forecast Yield Generated By Holt-Winters Method And ARIMA Models



6.1.5 Evaluation of models

Table 6.2 gives the forecast performance of forecast yield in the post sample forecast range. Holt-Winters on $Y1_t$ shows the best performance, followed by ARIMA(0,1,1)(0,1,1)$_4$ on $Y1_t$ and $Y2_t$. The forecasting performance of other models do differ very greatly from naïve model performance. Even for the best models, forecast performance differ from naïve model in the range of 4 percent. Therefore, the suitable model to forecast yield in Peninsular Malaysia is Holt-Winter Additive

Seasonality method fitted to an appropriate transformed yield series so that the seasonal fluctuation does not vary with the level of the series.

Table 6.2: Forecasting Performance In Post Sample Forecast Range

Fitted Model	RMSE	MAPE
HWMS (Y_t)	0.0343	7.4
HWAS (LY_t)	0.0819	9.7
HWAS ($Y1_t$)	0.0607	3.4
HWAS ($Y2_t$)	0.1974	5.3
ARIMA(0, 1, 1)(0, 1, 1)$_4$ LY_t	0.0273	5.7
ARIMA(0, 1, 1)(0, 1, 1)$_4$ $Y1_t$	0.0252	5.2
ARIMA(0, 1, 1)(0, 1, 1)$_4$ $Y2_t$	0.0238	4.7
Naïve Model(NM)	0.0414	7.4

6.2 Discussion

6.2.1 Factors Affecting Yield

The factors that are strongly linked to the upward trend in the yield series for the sample period are the use of fertilisers as well as an effective method of application, effective control of pests and diseases, machine-plucking and change in the method of pruning, i.e. table form. The impact of clonal varieties which are high yielders on trend, is less, as majority of tea the plants are plants which were planted thirty to forty years ago. These plants have lower yielding effect compared with clonal varieties. Moreover, during the sample period not much replanting was carried out. In the coming years, if more intensive replanting with clonal varieties is carried out, this will have a greater impact on the upward trend of the yield series.

Tea produced in Peninsular Malaysia mainly comes from the highlands, especially, from Cameron Highlands. The high fluctuations and seasonality is strongly linked to climatic conditions. That is, a combination of factors like temperature, rainfall, humidity and sunshine hours. Rapid development is taking place in Cameron Highlands as of late and this has an impact on climatic conditions. If development is unchecked and Cameron Highlands continues to be a major contributor of tea, one can expect to see a change in the seasonality pattern and fluctuations in the yield series of tea in Peninsular Malaysia for the coming years.

6.2.2 *Forecasting Models and Interpretation*

Holt-Winters method with three smoothing parameters on the yield series shows that the trend smoothing parameter is zero. This implies that the updating trend is equal to the previous trend. The implication is that, it could take the forecasting system a long time to overcome the influence of such a large downward shift when the overall trend is upward.

The ARIMA model developed for all three transformed series (Y^*_t) takes the following representation.

$$Y^*_t = Y^*_{t-1} + Y^*_{t-4} - Y^*_{t-5} + \varepsilon_t + 0.473520\varepsilon_{t-1} + 0.784360\varepsilon_{t-4} + 0.371410\varepsilon_{t-5}$$

The model has exponential weighted moving average (EWMA) components with no deterministic time trend component. The forecast yield of Y^*_t at time t is related to the immediate previous value of Y^*_t, that is, Y^*_{t-1}, the successive changes of Y^*_{t-4} - Y^*_{t-5}, that is, change in yield experienced by the same quarter for the previous year and past random shocks ε_{t-1}, ε_{t-4} , ε_{t-5} through the coefficients of 0.473520, 0.784360 and 0.371410 respectively.

Long term forecast for the fitted ARIMA models is equal to yield for the previous quarter plus change in yield (Y^*_{t-4} - Y^*_{t-5}) experienced by the same quarter for the previous year.

6.2.3 Does Transformation Matter

For the Holt-Winters method, Holt-Winter Multiplicative Seasonality is fitted to the original series, Y_t as this series displays increasing seasonal fluctuation. But if one decides to fitt Holt-Winters Additive Seasonality to a series then the series must display seasonal fluctuation which does not vary with the level of the series, otherwise, the series must be transformed to achieve this characteristic. In this study an appropriate transformation is carried out on the yield series to achieve this characteristics and then Holt-Winters Additive seasonality is fitted to the transformed series. The outcome of the two fitted models is that Holt-Winter on the transformed series gave an improved forecasting performance. However, the improvement is only small. The issue to be considered at this point is, does transformation matters? The answer would be that if the procedure of transformation is tedious and the improvement to forecasting performance is small then it would be advisable to use Holt-Winters Multiplicative Seasonality on the original series

Does data transformation matter in Box-Jenkins methodology? The answer is definitely yes, as this methodology requires the series to have a constant variance, if the series shows non-stationarity in variance, then an appropriate transformation must be determined to make the series have constant variance before modelling is carried out.

6.2.4 Transformation Bias

Another aspect that need to be noted here is that the mean absolute percentage error (MAPE) and root mean sum square error (RMSE) for the fitted and forecast values of the transformed series before and after bringing back to original scale differs. In this study, it differed in the range of 3 percent. This is because, the optimal forecasts obtained for a power-transformation series does not retain the optimal properties when brought back to the original scale, Granger and Newbold (1976). This is known as transformation bias and at times, the bias can be very high and makes an efficient forecasting method to produce less reliable forecasts.

6.2.5 Comparing Methodologies

In this study Holt-Winters method has outperformed Box-Jenkins Methodology. Some guidance on the relative accuracy of the different procedures is provided by the large-scale empirical comparisons of Newbold and Granger (1974) and Reid (1975) using, in each case, over 100 series of data. The study concluded that Box-Jenkins method gave more accurate forecasts than the Holt-Winter procedure for about two third of the series analysed. Nevertheless, the Holt-Winter procedure does actually outperform the Box-Jenkins procedure for about one-third of the 100 series of data. Therefore, it is not surprising to note that Holt-Winters Method can outperform Box-Jenkins Methodology for some time series, as has happened in the study.

Gupta (1993) also used Box-Jenkins methods to develop forecasting models for tea. Gupta came to the conclusion that ARIMA model offers a good technique for predicting the magnitude of any variable as the strength lies on the method which is suitable for any time series with any pattern of change. Moreover, ARIMA model does not require the forecasters to choose prior value of any parameter. Gupta also noted some limitations including the requirement for long time series and the use of rather sophisticated techniques.

6.3 Limitations

The Box-Jenkins methodology has been applied to the transformed series. However, optimal forecasts obtained for the power-transformed series do not retain their optimal properties when brought back to the original scale. Thus, the comparison of the models based on the non-optimal forecasts in the original scale may not be fully valid.

This study only focused on univariate time series model and this method is only able generate forecasts. This method does not quantify the effects of factors affecting the yield series. If explanatory modelling is considered, then the factors affecting the yield can act as explanatory third of the series analysed. Nevertheless, the Holt-Winter procedure does actually outperform the Box-Jenkins procedure for about one-third of the 100 series of data. Therefore, it is not surprising to note that Holt-Winters Method can outperform Box-Jenkins Methodology for some time series, as has happened in the study.

Gupta (1993) also used Box-Jenkins methods to develop forecasting models for tea. Gupta came to the conclusion that ARIMA model offers a good technique for predicting the magnitude of any variable as the strength lies on the method which is suitable for any time series with any pattern of change. Moreover, ARIMA model does not require the forecasters to choose prior value of any parameter. Gupta also noted some limitations including the requirement for long time series and the use of rather sophisticated techniques.

6.3 Limitations

The Box-Jenkins methodology has been applied to the transformed series. However, optimal forecasts obtained for the power-transformed series do not retain their optimal properties when brought back to the original scale. Thus, the comparison of the models based on the non-optimal forecasts in the original scale may not be fully valid.

This study only focused on univariate time series model and this method is only able generate forecasts. This method does not quantify the effects of factors affecting the yield series. If explanatory modelling is considered, then the factors affecting the yield can act as explanatory variables and will be able to measure the effects of these factors on yield variable.

6.4 Further Scope of This Study

There are some papers and studies that propose certain methods for correcting the bias arising from transforming forecasts back to the original scale, namely that of Neyman and Scott (1960), Miller (1984) and Taylor (1986). Although these authors did not consider explicitly time-series models, but it is possible to adapt their solutions to these models. On the other hand, Granger and Newbold (1976), Pankratz and Dudley (1987) as well as Victor (1993) addressed specifically the time series case and devised procedures for obtaining de-biasing factors that can be applied to correct transformation bias. It would be useful to evaluate whether these more complex techniques arrive at a better forecasting model.

Further research could also be done using Dynamic Regression (DR) models. That is, output variable (Y_t), yield is linearly related to current and past values of one or more input variables ($X_{1,t}$, $X_{2,t}$, $X_{3,t}$, . . .). The word "Dynamic" means that it pays special attention to the time structure of input-output relationship and the time structure of the disturbance series, Pankratz (1991).