Forecasting the future values of rubber yield and Cost of Products by fitting the best TIME SERIES models

P. K. B. N. M. Pallawala[1] and D. D. M. Jayasundara[2]

1. Department of Statistics and Computer Science,
University of Kelaniya,
Kelaniya, Sri Lanka.
Email: madhupallawala@kln.ac.lk
+94 (0) 722309945

2. Department of Statistics and Computer Science,
University of Kelaniya,
Kelaniya, Sri Lanka.
Email: jayasund@kln.ac.lk
+94 (0) 112 903 377

Corresponding Author: - P. K. B. N. M. Pallawala

ABSTRACT
Forecasting future values would be helpful to make the future plans and to take the require actions to safeguard the situation which can happen in the future. More frequently the future values are influenced by the past data and therefore most of the forecasting methods are mainly based on observing the past data. Time Series Models are one of the best models which can be used to model the data to forecast the future values using the past data. The focus of this paper is to come up with time-series models to forecast the future values of rubber yield and to forecast the future values of Cost of Products (CoP) values and to identify the behavior of rubber yield within the one year period of time in a selected rubber estate in Sri Lanka.

Keywords: forecast, CoP values, yield, Time Series Models.

1. INTRODUCTION

Rubber plays an important role in the Sri Lankan economy as one of the main export agricultural crops in the country. Rubber yield is the economically important part of rubber tree as well as the main outcome of the plantations. Identifying the variations of amount of rubber yield within a year would be an advantage for the future plans and it would be useful to control the effective situations in future. In rubber cultivation, use of rain guards is a common factor. Analyzing the effectiveness of the rain guard would be an advantage in the rubber cultivation. Forecasting future values using regression models is common approach. However, time series models are more advance than regression models and more suitable to predict the future values. Many research have been done to forecast the future values using Time Series models.

In 2003 Javier Contreras et al carried out a research to predict the next day electricity prices in Spain and California using ARIMA models. They have found that the ARIMA models are reliable and they have used the fitted models to forecast the electricity prices with high accuracy.
In Zhan-jun Qiao et al research they predicted the mid – long term regional power load with lower error percentage, based on the SARIMA model and using the CensusX12-SARIMA seasonal adjustment method.
In 2011 Alnaa, S., & Ahiakpor, F. had published papers on ARIMA approach to predicting inflation in Ghana mainly based on the AIC value and finally came up with a efficient model. Same approach was found in 2010 by Aidoo using SARIMA models and models were identified mainly by lowest AIC and BIC values. Many research based on forecasting tourist demand for several countries could be found in literature review. Chaitip in 2008 for Thailand and Nanthakumar in 2008 for Malaysia and Songs' research are some of the related research which used the Time Series models to forecast the tourist demand. Chaitip has used many models in two concepts as methods establishing one variable and methods establishing more variables including SARIMA models. Padhanhas carried out a research to forecast the international tourist's footfalls in India in 2011. His research provides evidence that the ARIMA models perform better with comparing to other competing models using the MAPE, MAD and other decisive. Time Series models can use for the purpose of forecasting in many areas. Literature can be found almost in all the areas like research by Faganel to forecast primary demand for a beer brand in 2008 and Earnest et al in 2005 ARIMA models to predict and monitor the number of beds occupied during a SARS outbreak in a tertiary hospital in Singapore.
Some research papers and works have done to review the past research done on Time Series forecasting and to compare the Time Series models with the other models. Jan et al in 2006 and Zhang in 2008 and J.S. Armstrong in 1983 had similar approaches. Many other research were done to forecast and analyze the data using Time Series models specially ARIMA and SARIMA models in past. Oduro-Gyimah et al in 2012, T. Nochai and R Nochai in 2006, Zhang et al in 2007 and Subsorn are also have work out with the Time Series models to forecast and analyze the data in various fields.
In this research paper will be discussing importance of the stationarity of the data before fitting a suitable forecasting model. Ways of removing the trend, seasonal and regular variations of the series. Special focus will be paid to the ACF and PACF of the series and they will be used to identify the lags of the suitable models. Distribution of the residuals and the p-values of the fitted models will be discussed using results and charts. The best model will be proposed to forecast the future values of the series and the forecasts will be compared with the actual values.

2. METHODOLOGY

2.1. Objectives of time series analysis

There are several possible objectives in analyzing a time series. They can be classified as

- Description
- Explanation
- Prediction
- Control

2.1.1. Components of a Time Series

- Trend
- Seasonal effects
- Cyclic components
- Irregular variations

2.1.1.1. Trend

Trend may be defined as “long term change in the mean value”. The trend measures the average change in the variable per unit time.

2.2.1.2. Seasonal effects

Seasonal variations are cycles that occur during the one year period of time and tend to repeat themselves each year.

2.1.1.3. Cyclic components (business cyclic movements)

This is refers to ups and downs around the trend lines observed over long period of time (usually more than one year).

2.1.1.4. Irregular variations

Irregular fluctuations in the time series cannot be predicted and these are occurring as a result of unexpected events such as strikes, droughts or political situations.

Much statistical theories concerned with random samples of independent observations. The special feature of time series analysis is the fact that successive observations are usually not independent and that the analysis must take in to account the time order of the observations. When successive observations are dependent future values may be predicted from past observations. To forecast there are several statistical procedures to determine the independencies and fitting suitable models. A common assumption in many time series techniques is that the data are stationary. A Stationary process has the property that the mean, variance and autocorrelation structure do not change over time. If the original time series is not stationary, it can be converted to a stationary series by differencing the original series with appropriate lag (trend removal) or transforming to an appropriate scale (stabilize the variance).

2.1.2. Stationary time series

A time series is stationary if

- There is no systematic change in the mean (i.e. no trend).
- There is no systematic change in variance.
- Periodic variances have been strictly removed.

2.1.2.1. Weak stationary

A time series weakly stationary or 2nd order stationary time series if

- Its mean is constant.
  i.e. $\epsilon[X_t] = \mu_t = \mu <$ infinity, that is the expectation of X_t is finite and does not depend on time t.
- Auto – covariance function does not depend on time. (i.e. depends only on lag)
  i.e.

$\gamma(X_{t1}, X_{t2}) = \gamma(X_{t1+h}, X_{t2+h})$ for all h, where h is the lag

2.1.2.2. Strictly stationary

A Time series $\{X_t\}$ is strictly stationary if $(X_1, X_2 \ldots, X_n)$ and $(X_{1+h}, X_{2+h}, X_{3+h}, \ldots X_{n+h})$ have the same joint distribution for all integers h, n >0.

$\therefore$ the joint distribution of (X_t) depends only on the gaps between $(t_1\ t_2\ t_3 \cdots t_n)$

2.1.3. White Noise (WN)

Purely random processes is sometimes called “white noise”, is a discrete time series and consist of a sequence of random variables $\{Z_t\}$, which are mutually independent and identically distributed. That is $\{z_t\}$ is a sequence of uncorrelated random variables, each with zero mean and variance σ^2.

$\{z_t\} \sim N(0, \sigma^2)$

Since this is mutually independent $\gamma_{(h)} = \mathrm{COV}(Z_t, Z_{t+h}) = 0$ for $h = \pm 1, \pm 2, \cdots \pm n$.

2.1.4.1. Autoregressive Process –AR(p)

Let $\{Z_t\}$ be a purely random process with mean zero and variance σ^2. Then $\{X_t\}$ is an autoregressive process of order p if

$X_t = \alpha_1 X_{t-1} + \ldots\ldots\ldots\ldots + \alpha_p X_{t-p} + z_t$, where $\{z_t\}$~WN(0, σ^2) and $\alpha_1, \alpha_2, \ldots, \alpha_p$ are constants.

2.1.4.2. Seasonal AR-models (SAR(p,P))

$$y_t = \delta + \phi_1 \cdot y_{t-1} + \ldots + \phi_p \cdot y_{t-p} + \phi_{1,L} \cdot y_{t-L} + \ldots + \phi_{P,L} \cdot y_{t-P \cdot L} + a_t$$

where *L* is the number of seasons (during a year) and
p – order of the non-seasonal level ,
P- order of the seasonal level.

2.1.5.1. Moving Average Process- MA(q)

Let $\{Z_t\}$ be a purely random process with mean zero and variance σ^2. Then
$\{X_t\}$ is a moving average process of order q if

$X_t = \beta_0 z_t + \beta_1 z_{t-1} + \ldots\ldots\ldots\ldots + \beta_q z_{t-q}$, where $\{z_t\}$~WN(0, σ^2) and $\beta_0, \beta_1, \beta_2, \ldots, \beta_q$ are constants usually $\beta_0 = 1$.

2.1.5.2. Seasonal MA-models (SMA(q,Q))

$$y_t = \delta + a_t - \theta_1 \cdot a_{t-1} - \ldots - \theta_q \cdot a_{t-q} - \theta_{1,L} \cdot a_{t-L} - \ldots - \theta_{Q,L} \cdot a_{t-Q}$$

where *L* is the number of seasons (during a year) and
q – order of the non-seasonal level, Q – order of the seasonal level

2.1.6. Mixed ARMA models

A useful class of models for time series is formed by combining MA and AR processes. A mixed autoregressive/moving-average process containing p AR terms and q MA terms is said to be an ARMA process of order (p,q). It is given by

$$X_t = \propto_1 X_{t-1} + \propto_2 X_{t-2} + \cdots + \propto_p X_{t-p} + Z_t + \beta_1 Z_{t-1} + \cdots + \beta_q Z_{t-q}$$

2.1.6.1. Integrated ARMA or ARIMA models.

In practices most time series are non – stationary. In order to fit a stationary model it is necessary to remove non-stationary sources of variation. If the observed time series is non-stationary in the mean, then we can difference the series.

The general ARIMA (p,d,q) process has the form

$$\phi(B)(1-B)^d X_t = \theta(B) Z_t$$

Where $\phi(B)$ is the polynomial of AR process and $\theta(B)$ is the polynomial of MA process.'d' is the degree of differencing and B is the backshift operator.

2.1.6.2. Seasonal ARIMA Models (ARIMA(p,d,q,P,D,Q)$_L$)

In practice, many time series contain a seasonal periodic component, which repeats every s observations. Box and Jenkin's generalized the ARIMA model to deal with seasonality, and defined a general multiplicative seasonal ARIMA (SARIMA) model as

$\phi_p(B)\Phi_P(B^s)W_t = \theta_q(B)\Theta_Q(B^s)Z_t$, where B denotes the Backward shift operator,

$\phi_p, \Phi_P, \theta_q, \Theta_Q$ are polynomials of order p,P,q,Q respectively and Z_t denotes a purely random process and

$W_t = \nabla_d \nabla_s^d X_t$

2.1.7.1. Autocorrelation Function (ACF)

Let $X_1, X_2, \ldots\ldots X_n$ be observations of a time series. The sample auto covariance function is defined as

$$\gamma(h) = \frac{1}{n}\sum_{t=1}^{n-h}(x_t - \bar{x})(x_{t+h} - \bar{x}) \quad , \text{ where } \bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t$$

The sample autocorrelation function is defined as

$$\rho(h) = \frac{\gamma(h)}{\hat{\gamma}(0)}, \qquad -n < h < n$$

2.1.7.2. Partial Autocorrelation Function (PACF)

Partial autocorrelations are used to measure the degree of association between X_t and X_{t-k} when the effect of other time lags 1,2,…..k-1 are removed.

Yule- Walker equation can be used to calculate the partial autocorrelations $\phi_1, \phi_2, \ldots \phi_\kappa$

$$\rho_1 = \phi_1 + \phi_2\rho_1 + \phi_3\rho_2 + \ldots + \phi_\kappa\rho_{\kappa-1}$$
$$\rho_2 = \phi_1\rho_1 + \phi_2 + \phi_3\rho_2 + \ldots + \phi_\kappa\rho_{\kappa-2}$$
....
....
$$\rho_\kappa = \phi_1\rho_{\kappa-1} + \phi_2\rho_{\kappa-2} + \phi_3\rho_{\kappa-3} + \ldots + \phi_\kappa$$

2.1.8.1. Parameter Estimation

Having made suitable model identification, the AR and MA parameters have to be determined in the best possible manner.

2.1.8.2. Method of Maximum likelihood

Under this method parameters are estimated by maximizing the likelihood function,

$$L(.) = f_{\underline{x}}(\underline{X})$$

2.1.8.3. Diagnostic Checking

The graphs for the ACF and PACF of the ARIMA residuals include lines representing two standard errors to either side of zero. Values that extend beyond two standard errors are statistically significant at approximately $\alpha = 0.05$, and show evidence that the model has not explained all autocorrelation in the data. For ACF the distance between the lines and zero for the i^{th} autocorrelation are determined by the following formula:

$$\sqrt[2]{1 + 2\sum_{k=1}^{i-1}\frac{\tau_k^2}{\sqrt{n}}}$$

where n = the number of observations in the series, and Γ_k = the k_i autocorrelation.

2.1.8.4. Residual Analysis

Although the selected model may appear to be best among those models considered, it is also necessary to do diagnostic checking to verify that the model is adequate. As with most statistical models, this is usually done by studying the residuals to see if any pattern remains unaccounted. For a good forecasting model, the residuals left over after fitting the model should be white noise. Therefore ACF and PACF of the residuals should be non-significant or they should be within the limits $\pm 1.96/\sqrt{n}$.

2.1.8.4.1. Assessing the normality of residuals

Histogram and normal probability plot of the residuals can be used for detecting the departures from the assumption that residuals are normally distributed. The assumption of normally distributed residuals is important when using prediction intervals based on normal distribution.

2.1.8.4.2. Ljung-Box Test

The ljung-box approximate chi-squared statistic provides a test to assess the reasonableness of the k autocorrelations as a set rather than individually.

Null Hypothesis: $H_0: \rho_1 = \rho_2 = ... = \rho_k = 0$

Testy statistics: $Q = n(n+2)\sum_{k=1}^{K}(n-k)^{-1}\hat{\rho}_k^2 \approx \chi^2(k-\rho)$

Where n - number of observations used to estimate the model

ρ - number of parameters estimated in the ARIMA model.

If the value of test statistics is larger than $\chi^2(1-\alpha, k-\rho)$, the 1-α quantile of the chi-square distribution of k-ρ degrees of freedom, there is sufficient evidence that the model is inadequate, under the α level of significance.

2.1.8.5. Forecasting

Forecasting the future values of an observed time series is important to make decisions.

2.1.8.5.1. Method of Minimum Mean Square Error (MSE)

Let $x_{1,} x_2, ..., x_t$ be observed time series. Then the forecast of x_{t+1} made at time t for l step ahead will be denoted by $\hat{x}_t(l)$.

The MSE forecast of x_{t+1} at time t is the conditional expectation of x_{t+1} at time t.

$\hat{x}_t(l) = E[x_{t+1} / x_t, x_{t-1}, \ldots.., x_1]$

2.1.8.5.2. Forecast Error

The Forecast error is defined as the difference between the actual and the forecast value.

$e_t(l) = x_{t+l} - \hat{x}_t(l)$

2.1.85.3. Prediction Limits

The 95% confidence interval for the forecasts is approximately equals to

$\hat{x}_t(l) \pm Z_{0.025}\sqrt{\text{var}(e_t(l))}$

The confidence interval typically widen as the forecast horizon increases, due to the expected build up error. The rate at which the confidence intervals widen will depend on the type of forecasting model.

2.2. Regression Analysis

2.2.1. Simple Linear Regression

Consider the case where the relationship between a dependent variable Y and a single independent variable X is studied. Then the model is

$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$

Where ε_i= Random Error = Errors of Y due to factors other than X

2.2.1.1. Assumptions of Regression (LINE Assumptions)

1. Linearity
   The response Y and the predictors(X) are linearly related.
2. Independence of error
   Error values are statistically independent.
3. Normality of Error
   Error values (ε) are normally distributed for any given value of X,
   $\varepsilon_i \sim N(0, \sigma^2)$
4. Equal variance (Homoscedasticity)
   The probability distribution of the errors has constant variance.

2.2.2. The Durbin-Watson Statistic

The Durbin-Watson statistic is used to test for auto-correlation.

H_0: Residuals are not correlated
H_1: Autocorrelation is present
The Durbin-Watson statistic is

$$D = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2}$$

- The possible range of the D is, $0 \leq D \geq 4$
- D should be closed to 2 if H_0 is true.
- D less than 2 may signal positive auto-correlation
- D greater than 2 may signal negative auto-correlation

2.2.3. Inference about the slope

Application of t-test for a population slope, is whether there is a linear relationship between X and Y.
$H_0: \beta_1 = 0$; no linear relationship
$H_1: \beta_1 \neq 0$; linear relationship does exists
Test –statistic

$$t = \frac{b_1 - \beta_1}{S_{b_1}}$$

Where b_1 – regression slope coefficient, β_1 – hypothesized slope, S_{b_1}- Standard error of the slope

2.2.4. F-test for significance

$H_0: \beta_1 = 0$; no linear relationship
$H_1: \beta_1 \neq 0$; linear relationship does exists
F-statistic:

$$F = \frac{\text{Regression Mean Square}}{\text{Residual Mean Square}}$$

Where F- follows an F- distribution with k numerator and (n-k-1) denominator degrees of freedom.
k- # of independent variables in the regression model.

2.2.5.1. R^2 (R-sq)

Coefficient of determination; indicates how much variation in the response is explained by the model. The higher the R^2, the better the model fits your data. The formula is:

$$1 - \frac{SS_{error}}{SS_{total}}$$

2.2.5.2. Adjusted R^2 (R-sq adj)

Adjusted R^2 accounts for the number of factors in your model. The formula is:

$$1 - \frac{MS_{error}}{SS_{total} / DF_{total}}$$

2.2.6. p value

The p value of a test of hypothesis is defined to be the smallest level of significance at which the null hypothesis is rejected. In hypothesis testing at 95% confidence level, the hypothesis is rejected if p-value obtained from the mini-tab output is less than α-value. If the p-value is greater than the α-value, then we don't have enough evidence to reject the null hypothesis of the test. Typically we take α-value as the 0.05.
Where $\alpha = P(\text{type I error}) = \text{P}(\text{reject } H_0 \mid H_0 \; is \; true)$

3. RESULTS AND DISCUSSION

3.1 Times series analysis on Monthly Crop(kg).

The data for the crop is taken from January in 1998 to December in 2009 for each month and these data were analyzed to fit a suitable model. To identify the existence of the trend and seasonal components of the series time plot was analyzed.

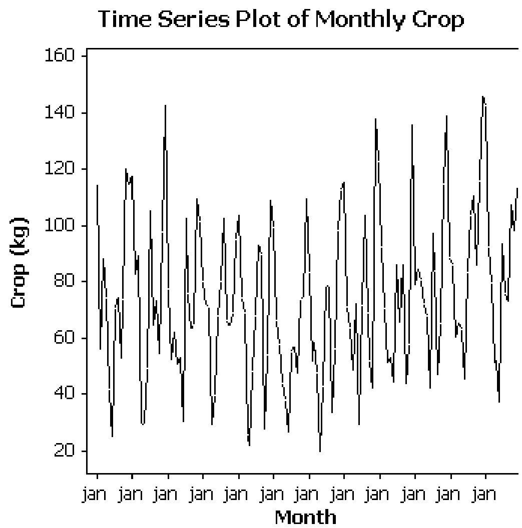


Figure 3.1.1 : Time series plot of crop vs. month.

Initially existing components of the time series of crop was identified by the plot. ACF of the series was used to identify the seasonality and other facts.

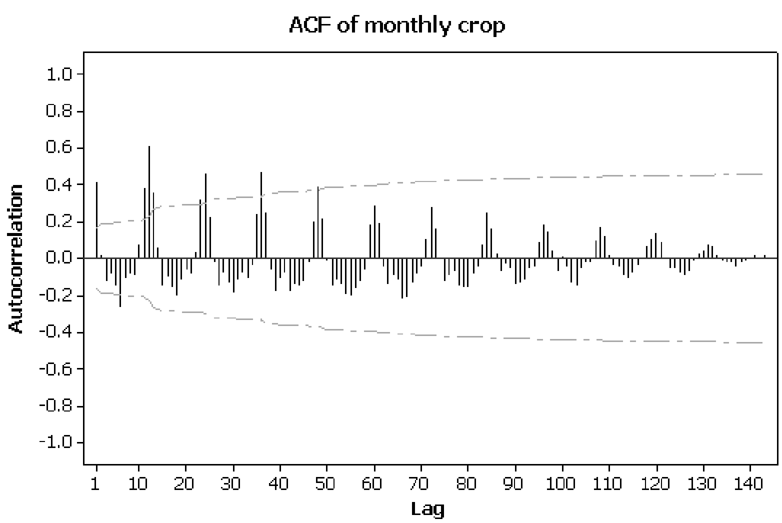


Figure 3.1.2: ACF of monthly crop vs. lag.

Seasonal model was identified to model the data by examining the ACF.

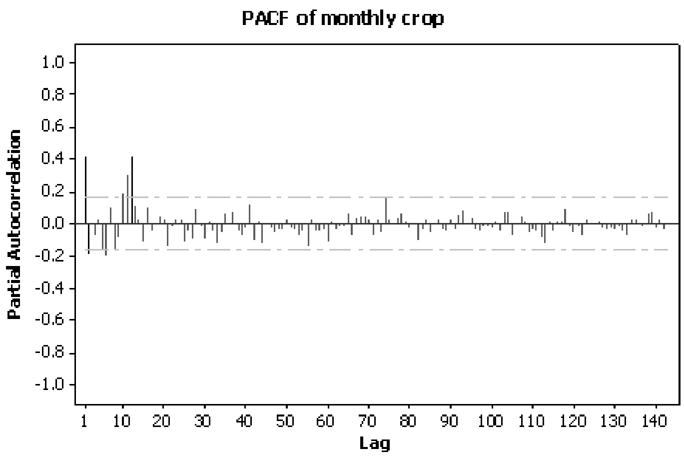


Figure 3.1.3 : PACF of monthly crop vs. lag

Before move to the SARIMA( ) models two SAR( ) models were identified. Identified models were fitted and analyzed the results and errors. Since there is a trend and seasonal in the series followings were done in order to make the series stationary. Trend was removed by taking the first non seasonal difference and using time series plot, ACF and PACF plots two possible time series models were identified using lags and since it shows strong seasonality first seasonal difference of the data set was take. Time plot, ACF and PACF of the series were analyzed. Corresponding plots were attached in the appendix. All possible SARIMA( ) models were identified and fitted. Comparing the results of all fitted models including the previously fitted models SARIMA(0,0,1) $(0,1,1)_{12}$ was identified as the best model. Due to lack of the space comparison of last two SARIMA ( ) model is included here. After comparing all the models SARIMA(0,0,1) $(0,1,1)_{12}$ and SARIMA(0,0,2) $(0,1,1)_{12}$ were identified as the best models. The results of the above two models were compared as follows:

SARIMA(0,0,1) $(0,1,1)_{12}$	SARIMA(0,0,2) $(0,1,1)_{12}$
Ljung-box statistic is ok Residual MS is low Parameter estimates are ok Residuals are distributed normally and residual fittings are ok	Ljung-box statistic is ok Residual MS is high Parameter estimates are not exact. Residuals are distributed normally and residual fittings are ok

Table 3.1.1: Comparison of the SARIMA(0,0,1) $(0,1,1)_{12}$ models.

SARIMA(0,0,1)$(0,1,1)_{12}$ is identified as the best model to forecast the monthly crop.
Using the fitted model forecast from January to August of 2010 was done in 95 percent limit.

Month	Forecast	Lower limits	Upper limits	Actual	Error	Error%
January	**105.321**	70.475	140.167	**110.619**	5.29779	4.78924
February	**82.498**	46.987	118.008	**76.512**	-5.98599	7.82363
March	**76.073**	40.562	111.583	**76.619**	0.54607	0.71271
April	**57.561**	22.051	93.072	**59.349**	1.7878	3.0124
May	**55.366**	19.856	90.876	**49.529**	-5.8404	11.79
June	**46.664**	11.153	82.174	**50.837**	4.1732	8.2090
July	**89.485**	53.975	124.995	**78.070**	-11.4152	4.6218
August	**89.563**	54.053	125.073	**86.047**	-3.5165	4.0867

Table 3.1.2.: Comparing the forecast and actual values

3.2.Time series analysis of monthly COP(Rs/kg)

The data for the COP is taken from January in 2000 to December in 2009 for each month and these data will be analyzed to fit a suitable model. Time plot for the COP values was analyzed.

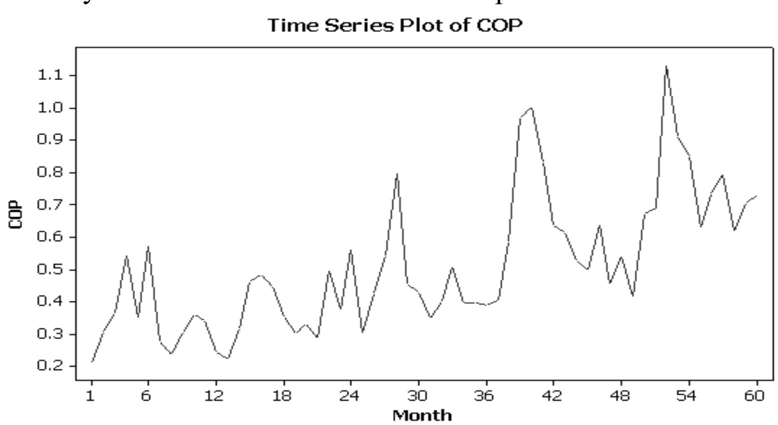


Figure 3.2.1: Time plot of CoP values

Using the same procedure used for monthly crop SARIMA$(1,1,1)(1,1,0)_{12}$ was identified as the most suitable model to forecast the CoP values. Using the residual plots it is shown that the residuals are uncorrelated and normally distributing with this model.

Forecasting from January to June in 2010 using the fitted model in 95 percent limits.

Month	Forecast	Lower limits	Upper limits	Actual	Error	Error%
January	**152.289**	111.983	192.595	**141.87**	-10.4188	7.34391
February	**193.618**	148.135	239.099	**214.37**	20.753	9.68034
March	**262.403**	215.527	309.282	**246.61**	15.793	6.40412
April	**286.804**	239.484	334.121	**257.06**	29.744	11.5
May	**248.927**	201.459	296.429	**264.31**	15.3830	5.8201
June	**213.463**	165.903	261.025	**222.66**	9.197	4.130

Table 3.2.1: Comparing forecast and actual values.

3.3. Analysis of monthly crop vs. the rainfall

For the analyzing the effectiveness of the rain guards to the rubber crop of the estate regression analysis was carried out for monthly crop vs. monthly rainfall.

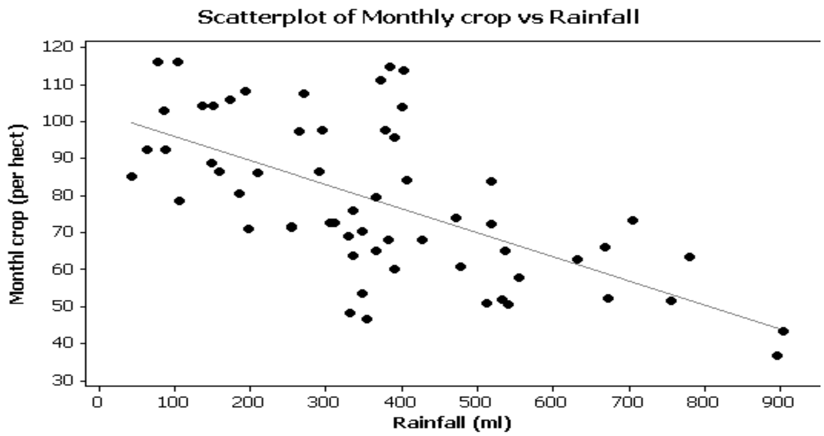


Figure3.3.1: Scatter plot of monthly crop vs. monthly rainfall

Following regression line was identified to model the data:

Monthly crop = 102 - 0.0649 Rainfall

Predictor	Coef	SE Coef	T	**P**
Constant	102.183	4.375	23.35	**0.000**
Rainfall	-0.06493	0.01041	-6.24	**0.000**

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	10383	10383	38.93	**0.000**
Residual Error	58	15468	**267**		
Total	59	25851			

S = 16.3306 R-Sq = 40.2% R-Sq(adj) = 39.1%

According to the results there exists a negative relationship between monthly crop and the monthly rainfall and therefore finally conclude the use of rain guards is affected the crop.

Finally the SARIMA (0,0,1) (0,1,1)12 was fitted to forecast the future values of the rubber yield. After fitting the proposed model, predictions was done for the next year and compare those predicted values with the actual values. Then error percentage to the actual value was done with the outcomes. Same procedure was carried out to identify and to predict the CoP values as for the monthly crop and SARIMA(1,1,1)(1,1,0)$_{12}$ was fitted as the best model. There was a negative relationship between monthly crop and the rainfall. Therefore the belief rain guards affect the rubber yield is not rejected. However according to the R^2 and R^2(adj) rain fall explains about 40% of the crop.

4. APPENDIX

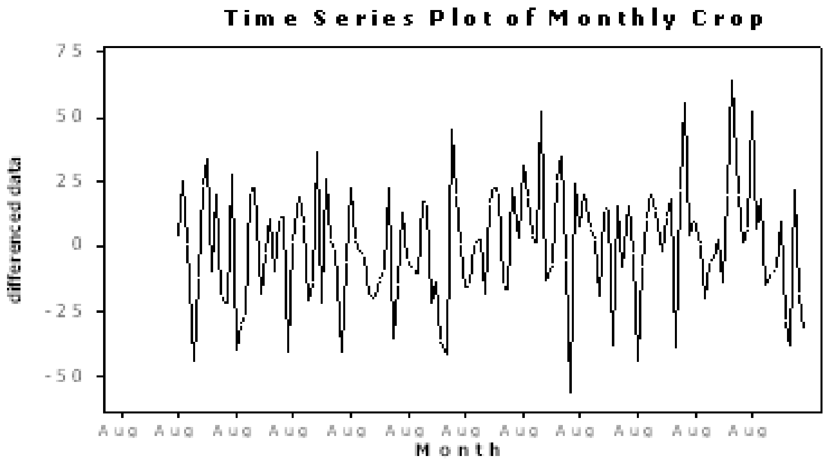


Figure 4.1.4: Time series plot of the first seasonal and non seasonal differenced monthly crop.

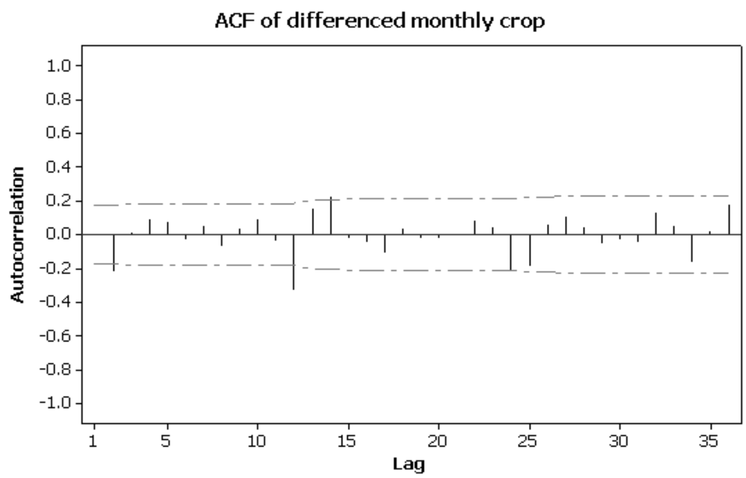


Figure 4.1.5: ACF of the first seasonal and non seasonal differenced monthly crop.

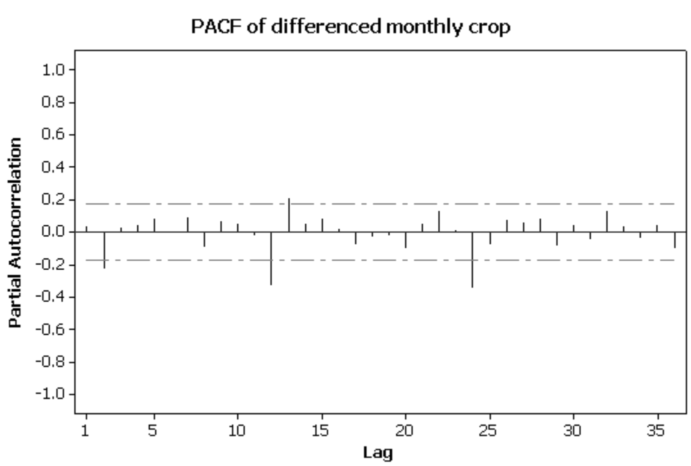
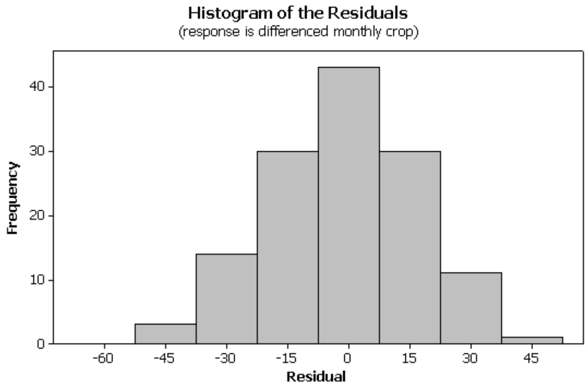
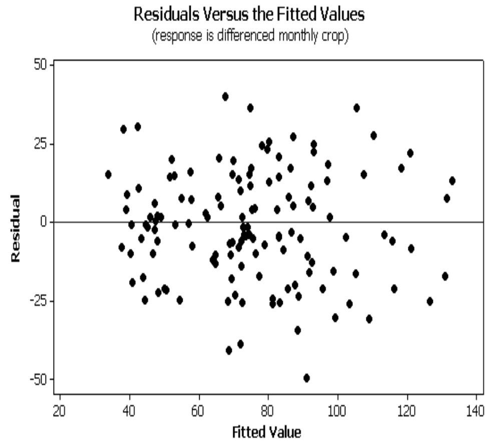


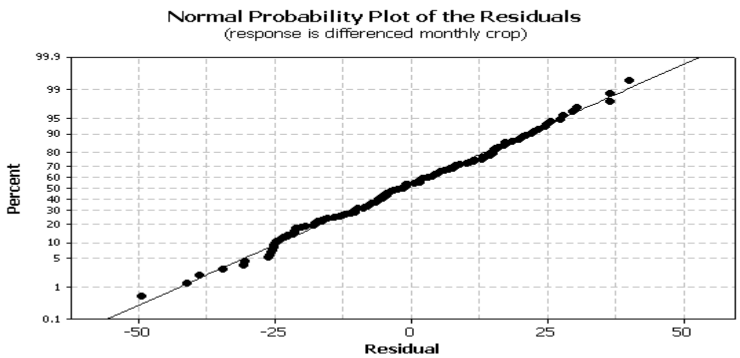
Figure 4.1.6: PACF of the first seasonal and non seasonal differenced monthly crop



(a) Histogram of the residuals

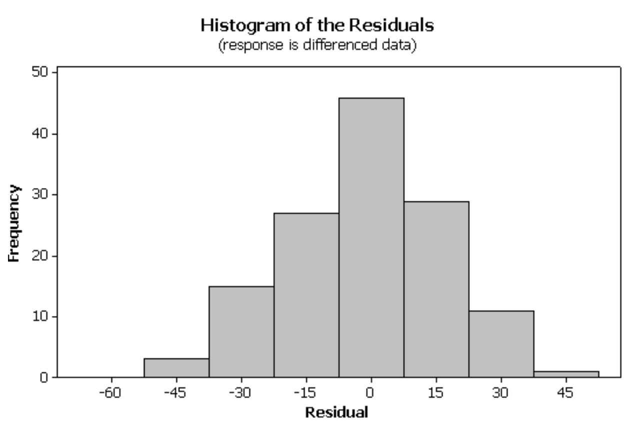


(b) residuals vs. fitted values

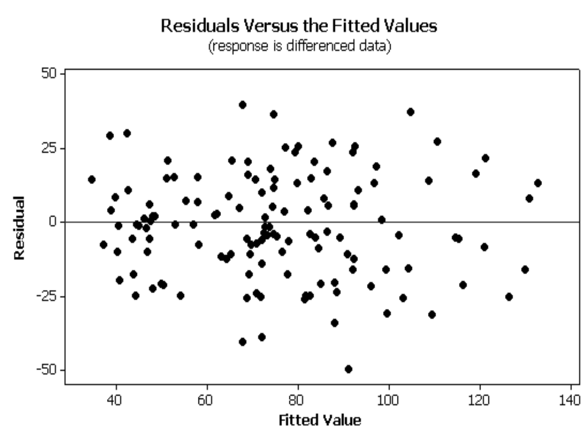


(c). Normal probability plot of the residuals

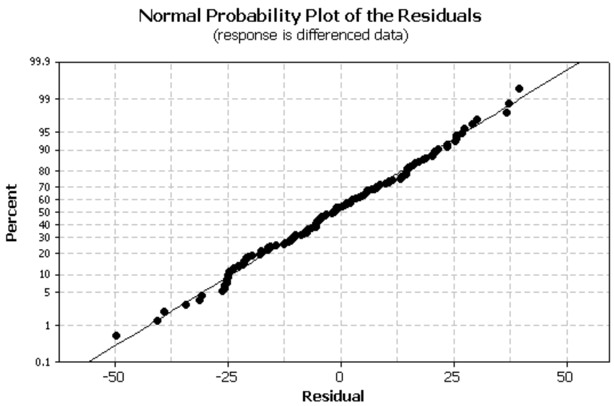
Figure 4.1.7: Residual plots of SARIMA (0,0,1) $(0,1,1)_1$



(a) Histogram of the residuals



(b) Residuals vs. fitted values



(c) Normal probability plot of residuals

Figure 4.1.8: Residual plots of SARIMA(0,0,2) $(0,1,1)_{12}$.

REFERENCES

Aidoo, E. (2010). Modelling and Forecasting Inflation Rates in Ghana: An Application of SARIMA Models (Doctoral dissertation, Dalarna University).

Alnaa, S., & Ahiakpor, F. (2011). ARIMA (autoregressive integrated moving average) approach to predicting inflation in Ghana. Journal of economics and international finance,3(5),(228-336).

Alnaa, S., & Ahiakpor, F. (2011). ARIMA (autoregressive integrated moving average) approach to predicting inflation in Ghana. Journal of economics and international finance, 4(3),83-87.

Armand Faganel, (2008). Forecasting primary demand for a beer brand using time series analysis. Organizacija, 41, 116-123.

Armstrong, J. S. (1983). Relative accuracy of judgemental and extrapolative methods in forecasting annual earnings. Journal of Forecasting, *2*(4), 437-447.

Chaitip, P., Chaiboonsri, C., & Mukhjang, R. (2008). Time Series Models for Forecasting International Visitor Arrivals to Thailand. International Conference on Applied Economics–ICOAE (159-13).

Contreras, J. ,Espínola, R., Nogales, F.J., &Antonio J. Conejo, (2003). ARIMA Models to Predict Next-Day Electricity Prices. IEEE transactions on power systems, 18, 1014-1020.

Earnest, A., Chen, M. I., Ng, D., & Sin, L. Y. (2005). Using autoregressive integrated moving average (ARIMA) models to predict and monitor the number of beds occupied during a SARS outbreak in a tertiary hospital in Singapore. BMC Health Services Research, 5(1), 36.

Jan G. De Gooijer , Rob J. Hyndman , (2006). 25 years of time series forecasting. International Journal of Forecasting, 22, 443–473.

Nanthakumar, L., Subramaniam, T., & Kogid, M. (2012). Is 'malaysia truly asia'? Forecasting tourism demand from ASEAN using SARIMA approach. Tourismos: an International Multidisciplinary Journal of Tourism, 7(1), 367-381.

Nochai, R., & Nochai, T. (2006). ARIMA model for forecasting oil palm price. In 2nd IMT-GT Reg. Conf, on Mathematics, Statistic and Apps., Univ. Sains Malaysia, Penang, June.

Oduro-Gyimah, F. K., Harris, E., & Darkwah, K. F. (2012). SARIMA Time Series Model Application to Microwave Transmission of Yeji-Salaga (Ghana) Line-Of-Sight Link. International Journal of Applied, 2(9),40-51.

Padhan, P. C. (2011). Forecasting International Tourists Footfalls in India: An Assortment of Competing Models. International Journal of Business and Management, 6(5),190-202.

Qiao, Z. J., Li, F. L., & Li, Y. Mid-long-Term Regional Load Forecasting based on Census X12-SARIMA Model.

Song, H., & Li, G. (2008). Tourism demand modelling and forecasting—A review of recent research. Tourism Management, 29(2), 203-220.

Subsorn, P., Xiao, J., & Clayden, J. Forecasting Rubber Production using Intelligent Time Series Analysis to Support Decision Makers.

Zhang, Y., Bi, P., Hiller, J. E., Sun, Y., & Ryan, P. (2007). Climate variations and bacillary dysentery in northern and southern cities of China. Journal of Infection, 55(2), 194-200.

Zhang, Y., Bi, P., & Hiller, J. (2008). Climate variations and salmonellosis transmission in Adelaide, South Australia: a comparison between regression models. International journal of biometeorology, 52(3), 179-187.