Trends in Agricultural Economics 3 (4): 207-215, 2010
ISSN 1994-7933

A Comparison of Univariate Time Series Methods for Forecasting Cocoa Bean Prices

[1]K. Assis, [2]A. Amran, [3]Y. Remali and [4]H. Affendy
[1]School of Sustainable Agriculture,
[2]School of Science and Technology,
[3]School of Business and Economics,
[4]School of International Tropical Forestry, Universiti Malaysia Sabah,
Locked Bag 2073, 88999 Kota Kinabalu, Sabah, Malaysia

Abstract: The purpose of this study was to compare the forecasting performances of different time series methods for forecasting cocoa bean prices. The monthly average data of Tawau cocoa bean prices graded SMC 1B for the period of January 1992-December 2006 was used. Tawau is one of the top cocoa producers in the world along with the Ivory Coast, Ghana and Indonesia. Four different types of univariate time series methods or models were compared, namely the exponential smoothing, autoregressive integrated moving average (ARIMA), generalized autoregressive conditional heteroskedasticity (GARCH) and the mixed ARIMA/GARCH models. Root mean squared error (RMSE), mean absolute percentage error (MAPE), mean absolute error (MAE) and Theil's inequality coefficient (U-STATISTICS) were used as the selection criteria to determine the best forecasting model. This study revealed that the time series data were influenced by a positive linear trend factor while a regression test result showed the non-existence of seasonal factors. Moreover, the Autocorrelation function (ACF) and the Augmented Dickey-Fuller (ADF) tests have shown that the time series data was not stationary but became stationary after the first order of the differentiating process was carried out. Based on the results of the ex-post forecasting (starting from January until December 2006), the mixed ARIMA/GARCH model outperformed the exponential smoothing, ARIMA and GARCH models.

Key words: Univariate time series models, exponential smoothing, ARIMA, GARCH, ARIMA/GARCH, model selection criteria

INTRODUCTION

Cocoa, scientifically known as *Theobroma cacao* L. is the third-largest agricultural commodity in Malaysia after oil palms and rubber. Malaysia now exports cocoa products to sixty-six countries (Ministry of Plantation Industries and Commodities, 2006). Tawau is one of the top cocoa producers in Malaysia and even in the world along with the Ivory Coast, Ghana and Indonesia (Shanti, 2006). Domestic cocoa bean prices are changing from time to time and very volatile (Yusoff and Salleh, 1987; Arshad and Zainalabidin, 1994). Instability of cocoa prices creates significant risks to producers, suppliers, consumers and other parties that are involved in the marketing and production of cocoa beans, particularly

Corresponding Author: K. Assis, School of Sustainable Agriculture, Universiti Malaysia Sabah, Locked Bag 2073, 88999 Kota Kinabalu, Sabah, Malaysia

in Malaysia. In risky conditions and amidst price instability, forecasting is very important in helping to make decisions. Accurate price forecasts are particularly important to facilitate efficient decision making as there is time lag intervenes between making decisions and the actual output of the commodity in the market.

Modelling or forecasting of agricultural price series, like that of other economic time series, has traditionally been carried out either by building an econometric model or by applying techniques developed for analyzing stationary time series. Time series forecasting is a major challenge in many real world applications such as stock price analysis, palm oil prices, natural rubber prices, electricity prices and flood forecasting. This type of forecasting is to predict the values of a continuous variable (called as response variable or output variable) with a forecasting model based on historical data. There are two types of time series forecasting modeling methods; univariate and multivariate. Univariate modeling methods generally used time only as an input variable with no other outside explanatory variables (Celia *et al*., 2003). This forecasting method is often called univariate time series modeling. A few commonly employed methods in univariate time series models are exponential smoothing, autoregressive-integrated-moving average (ARIMA) and Autoregressive Conditional Heteroscedastic (ARCH) (Kahforoushan *et al*., 2010).

The last few decades have witnessed significant advances in the topic of exponential smoothing. It has established itself as one of the leading forecasting strategies (Robert and Amir, 2009). Fatimah and Roslan (1986) confirmed the suitability of univariate ARIMA models in agricultural prices forecasting. Shamsudin *et al*. (1992) has noted that ARIMA models have the advantage of relatively low research costs when compared with econometric models, as well as efficiency in short term forecasting. One of the earliest time series models allowing for heteroscedasticity is the Autoregressive Conditional Heteroscedastic (ARCH) model introduced by Engel (1982). Bollerslev (1986) extended this idea into Generalized Autoregressive Conditional Heteroscedastic (GARCH) models which give more parsimonious results than ARCH models, similar to the situation where ARMA models are preferred over AR models. Kamil and Noor (2006) have developed a time series model of Malaysian palm oil prices by using ARCH models. Zhou *et al*. (2006) have proposed a new network traffic prediction model based on non-linear time series ARIMA/GARCH. They found that the proposed ARIMA/GARCH outperformed the existing Fractional Autoregressive Integrated Moving Average (FARIMA) model in terms of prediction accuracy. Therefore, the objective of this research was to compare the forecasting performances of four different univariate time series methods or models for forecasting cocoa bean prices (i.e., Tawau cocoa bean prices), namely exponential smoothing, ARIMA, GARCH and the mixed ARIMA/GARCH models.

MATERIALS AND METHODS

The monthly Tawau cocoa bean prices graded SMC 1B was used for this study which was collected from the official website of The Malaysian Cocoa Board (http://www.koko.gov.my/lkmbm/loader.cfm?page=statisticsFrm.cfm). The time series data was measured in Ringgit Malaysia per tonne (RM/tonne). The time series data ranged from January 1992 until December 2006. The coefficient of variation (V) was used to measure the index of instability of the time series data. The coefficient of variation (V) is defined as:

$$V = \frac{\sigma}{\bar{Y}}$$

where σ is the standard deviation and

$$\bar{Y} = \frac{\sum_{t=1}^{n} Y_t}{n}$$

is the mean of Tawau cocoa bean prices changes.

A completely stable data has V = 1, but unstable data are characterized by a V>1 (Telesca *et al*., 2008).

Regression analysis was used to test whether trends and seasonal factors exist in the time series data. The existence of linear trend factors was tested through this regression equation

$$Y = \beta_0 + \beta_1 \text{Trend} + \varepsilon \qquad \varepsilon \sim WN(0, \sigma^2)$$

with Y is the time series data of the study, Trend is the linear trend factor, β_0 and β_1 are parameters and ε is the error of the model with an assumption of White Noise (WN). The hypothesis of the model was

- **H_0:** β_1 = (Non-existence of linear trend factor)
- **H_1:** $\beta_1 \neq$ (Linear trend factor exists)

With the month of January as the base month, the existence of seasonal factor was detected by using regression as shown below:

$$Y = \beta_0 + \beta_1 \text{Trend} + \beta_2 \text{Feb} + \beta_3 \text{Mar} + \beta_4 \text{Apr} + \beta_5 \text{May} + \beta_6 \text{Jun} + \beta_7 \text{Jul} + \beta_8 \text{Aug} + \beta_9 \text{Sep} + \beta_{10} \text{Oct} + \beta_{11} \text{Nov} + \beta_{12} \text{Dec} + \varepsilon$$

and hypothesis was defined as:

- **H_0:** $\beta_2 = \beta_3 = \beta_4 = \ldots \beta_{12} = 0$ (Non-existence of seasonal factor)
- **H_1:** At least one of $\beta_2, \beta_3, \ldots, \beta_{12} \neq 0$ (Seasonal factor exists)

The correlogram and Augmented Dickey-Fuller (ADF) test were chosen to test the stationary of the time series data.

Exponential Smoothing

The h-periods-ahead forecast is given by:

$$\hat{Y}_{t+h} = a + bh$$

with a and b are permanent components. Both of these parameters are counted by the following equations

$$a_t = \alpha Y_t + (1-\alpha)(a_{t-1} + b_{t-1})$$

$$b_t = \beta(a_t - a_{t-1}) + (1-\beta) b_{t-1}$$

with $0<\alpha, \beta<1$.

ARIMA

This study followed the Box-Jenkins methodology which involves four steps. These are identification, estimation, model checking and forecasting. ARMA (p, q) processes can be simply expressed as the following two Eq.

$$Y_t = x_t \gamma + e_t \tag{1}$$

$$e_t = \sum_{i=1}^{p} \phi_i \mu_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t \tag{2}$$

where, x_t is the explanatory variables, e_t is the disturbance term, ε_t is the innovation in the disturbance, p is the order of AR term, q: the order of MA term. In Eq. 2, the disturbance term (μ_t) again consists of three parts. The first part is AR terms and the second part is MA terms. The last one is just a white-noise innovation term. If we replace the data (Y) with the difference data ($\Delta y_t = Y_t - Y_{t-1}$), then the ARMA models become ARIMA(p, d, q) models.

GARCH

The standard form of GARCH(p,q) models can be specified as following three equations:

$$Y_t = x_t \gamma + \varepsilon_t \tag{3}$$

$$\varepsilon_t = v_t \sqrt{\sigma_t^2} \tag{4}$$

$$\sigma_t^2 = \delta + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \tag{5}$$

where, p is the order of GARCH term, q is the order of ARCH term and σ^2_v. Equation 3 and 5 are called mean equation and conditional variance equation, respectively. The mean equation is written as a function of exogenous variables (x_t) with an error term (μ_t). The variance equation is a function of mean (δ), ARCH (μ^2_{t-i}) and GARCH term (μ^2_{t-t}).

ARIMA/GARCH

Combination of ARIMA(p,d,q) and GARCH(p,q) are written as below:

$$(\Delta Y_t)^d = \sum_{i=1}^{p} \phi_i (\Delta Y_{t-i})^d + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \qquad \varepsilon_t \sim WN(0, \sigma_t^2) \tag{6}$$

$$\sigma_t^2 = \delta + \sum_{j=1}^{q} \beta_j \varepsilon_{t-j}^2 + \sum_{i=1}^{p} \alpha_i \sigma_{t-i}^2 \tag{7}$$

Eight model selection criteria as suggested by Ramanathan (2002) were used to chose the best forecasting models among ARIMA and GARCH models (Table 1). While, the best time series methods for forecasting Tawau cocoa bean prices was chosen based on the values of four criteria, namely RMSE, MAE, MAPE and U-statistics (Table 2). Finally, the selected model was used to perform short-term forecasting for the next twelve months for Tawau cocoa bean prices starting from January 2007 until December 2007.

Table 1: Model Selection Criteria (Ramanathan, 2002)

Criteria	Formula
AIC	$\left(\frac{ESS}{n}\right)e^{2f/n}$
FPE	$\left(\frac{ESS}{n}\right)\frac{n+f}{n-f}$
GCV	$\left(\frac{ESS}{n}\right)\left[1-\left(\frac{f}{n}\right)\right]^{-2}$
HQ	$\left(\frac{ESS}{n}\right)(\ln n)^{2f/n}$
RICE	$\left(\frac{ESS}{n}\right)\left[1-\left(\frac{2f}{n}\right)\right]^{-1}$
SCHWARZ	$\left(\frac{ESS}{n}\right)n^{f/n}$
SGMASQ	$\left(\frac{ESS}{n}\right)\left[1-\left(\frac{f}{n}\right)\right]^{-1}$
SHIBATA	$\left(\frac{ESS}{n}\right)\frac{n+2f}{n}$

n: Number of observations, f: Number of parameters, ESS: Error sum of square

Table 2: Forecast accuracy criteria

Criteria	Formula
RMSE	$\sqrt{\frac{ESS}{n}}$
MAE	$\frac{\sum_{t=1}^{n}\left\|Y_t-\hat{Y}_t\right\|}{n}$
MAPE	$\frac{\sum_{t=1}^{n}\left\|\frac{Y_t-\hat{Y}_t}{Y_t}\right\|}{n}\times 100\%$
U-statistics	$\frac{RMSE}{\sqrt{\sum_{t=1}^{n}\hat{Y}^2/n}+\sqrt{\sum_{t=1}^{n}Y^2/n}}$

Y_t: The actual value at time t, $\hat{Y}_t$: The forecast value at time t, n: The number of observations; ESS: The error sum of square

RESULTS

The results showed that the coefficient of variation (V) of the time series data was 1.012 (V>1). Because of the V value was closed to 1, so this study was concluded that the time series data was stable (Telesca *et al*., 2008). The results of the regression analysis have shown that positive linear trend factor exists in the time series data but seasonal factor was not. With referring to the correlogram and the Augmented Dickey-Fuller tests results, the time series data of the study was not stationary. But after the first order of differencing was carried out, the time series data became stationary (Fig. 1).

Exponential Smoothing

The double exponential smoothing method was used as the regression result has showed the positive linear trend factor exists in the time series data. Double exponential smoothing models consist with two parameters which symbolized as α for mean and β for trend. The best model of the double exponential smoothing has been selected based on the lowest value of MSE (Mean Square Error) from combination of α and β with condition $0<\alpha$,

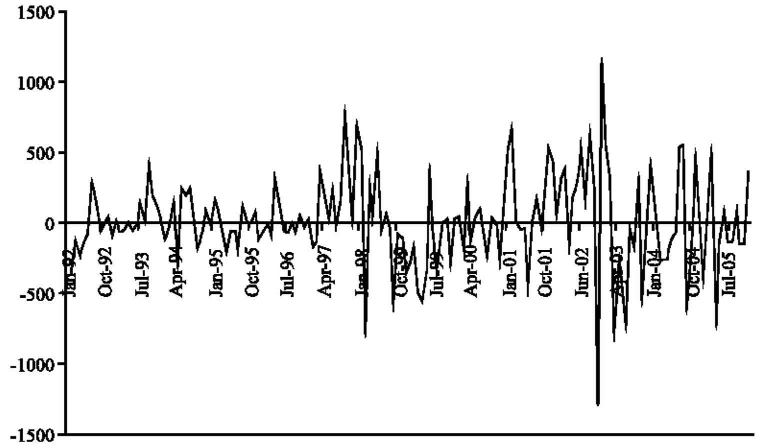

Fig. 1: Time series data (after first order of differencing)

Table 3: Error Sum of Square (ESS) according to α and β values

	α								
β	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	173185627	70362670	41488913	30011334	24384697	21244412	19366607	18231686	17589723
0.2	171444565	61358397	36273998	27053389	22754641	20391822	19006661	18223235	17866015
0.3	138700812	53824840	32336246	25412232	22160031	20307532	19226853	18673826	18528354
0.4	130053979	46589861	30068608	24960672	22290468	20684210	19769690	19377921	19415785
0.5	130417274	41204810	29317287	25242334	22771650	21264717	20458721	20212951	20443538

Table 4: EViews output of the double exponential smoothing model

Sample: 1992M01 2005M12
Included observations: 168
Method: Holt-Winters No Seasonal
Original Series: TAWAU
Forecast Series: TAWAUSM

Analysis	Values
Parameters	
Alpha	0.9000
Beta	0.1000
Sum of squared residuals	17589723
Root mean squared error	323.5749
End of period levels	
Mean	4852.068
Trend	-23.43109

$\beta<1$. The result showed that combination $\alpha = 0.9$ and $\beta = 0.1$ was the best forecasting model of double exponential smoothing method (Table 3). The double exponential smoothing model was written in equation form as (Table 4).

$$F_{T+h} = a + bh = 4852.068 + (h) * (-23.43109)$$

ARIMA

All models which fulfilled the criteria of $p+q \leq 5$ have been considered and compared in this study and there were twenty ARIMA (p, d, q) models which fulfilled the criteria. Parameters of the models were estimated with the least square method. Parameters which were not significant at 5% confidence level were dropped from the model. Using the eight model selection criteria suggested by Ramanathan (2002), the ARIMA (3, 1, 2) model was selected as the best model among the other ARIMA models. However, the parameters of AR

Table 5: Estimation of ARIMA (3, 1, 2)

Variables	Coefficient	Standard error	Z-statistic	p-value
Constant	18.130660	26.185420	0.692395	0.4897
AR(2)	-0.895200	0.051784	-17.287200	<0.0001*
AR(3)	0.118147	0.039606	2.983055	0.0033*
MA(2)	0.936905	0.045159	20.746770	<0.0001*

*p<0.05

Table 6: Estimation of GARCH (1, 1)

	Mean equation			
Variables	Coefficient	Standard error	Z-statistic	p-value
Constant	2.2945	17.5522	0.1307	0.8960
Conditional variance equation				
Constant	5896.95	2841.67	2.0752	0.0380*
ε^2_{t-1}	0.2753	0.1009	2.7286	0.0064*
σ^2_{t-1}	0.6887	0.0961	7.1633	<0.0001*

*p<0.05

Table 7: Estimation of ARIMA (3, 1, 2)/GARCH (1, 1)

Variables	Coefficient	Standard error	Z-statistic	p-value
ARIMA (3, 1, 2)				
Constant	10.42818	19.2344	0.542163	0.5877
AR (2)	-0.86716	0.050758	-17.084200	<0.0001*
AR (3)	0.087526	0.036952	2.368676	0.0179*
MA (2)	0.944859	0.030012	31.483150	<0.0001*
GARCH (1, 1)				
C	4606.362	2778.705	1.657737	0.0974
ε^2_{t-1}	0.237459	0.099897	2.377032	0.0175*
σ^2_{t-1}	0.736343	0.106755	6.897485	<0.0001*

*p<0.05

(1) and MA (1) were found not significant and thus dropped from the model. The ARIMA (3, 1, 2) model was written in equation form as (Table 5).

GARCH

Identification and estimation of GARCH (p, q) models in this study were done by following the four steps that were ARCH effect checking, estimation, model checking and forecasting. Four GARCH (p,q) models were selected and compared, namely GARCH (1, 1), GARCH (1, 2), GARCH (2, 1) and GARCH (2, 2). Using the eight model selection criteria suggested by Ramanathan (2002), the GARCH (1, 1) model has been selected as the best model among the other three GARCH models.

The GARCH(1,1) model was written in equation form as (Table 6):

$$\hat{z}_t = 2.2945 \qquad \text{(Mean equation)}$$

$$\hat{\sigma}_t^2 = 5896.95 + 0.2753\varepsilon_{t-1}^2 + 0.6887\sigma_{t-1}^2 \qquad \text{(Conditional variance equation)}$$

ARIMA/GARCH

ARCH effect which was tested by using a regression analysis exists in the ARIMA (3, 1, 2) model. That means the ARIMA (3, 1, 2) model could be mixed with the best GARCH model (i.e., GARCH(1, 1)).

The ARIMA(3, 1, 2)/GARCH(1, 1) model was written in equation form as (Table 7):

$$\hat{z}_t = 10.42818 - 0.86716z_{t-2} + 0.087526z_{t-3} + 0.944859\varepsilon_{t-2}$$

$$\hat{\sigma}_t^2 = 4606.362 + 0.237459\varepsilon_{t-1}^2 + 0.736343\sigma_{t-1}^2$$

Table 8: Four model selection criteria

Criteria	Double exponential Smoothing	ARIMA (3, 1, 2)	GARCH (1, 1)	ARIMA (3, 1, 2)/ GARCH (1, 1)
RMSE	324.45006	183.3087257	158.8800832	155.5006431
MAE	292.65075	144.9835	122.80825	126.74
MAPE	5.805836015	2.936916969	2.417800798	2.552741574
U-Statistics	0.033466951	0.018211057	0.01604509	0.015509933

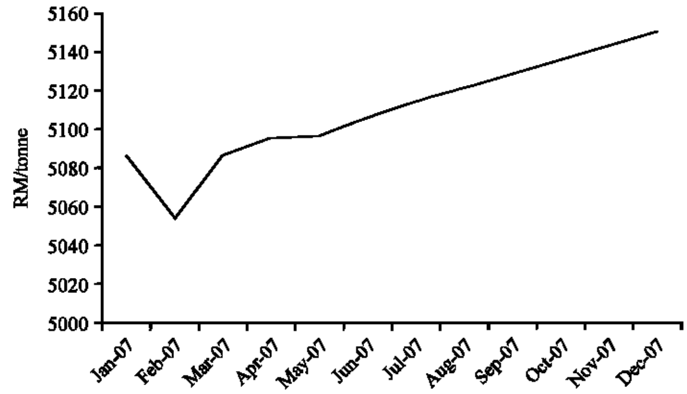

Fig. 2: Short-term forecasting of tawau cocoa bean prices

Model Selection

Four model selection criteria were used to select the best forecasting model from the four different types of time series methods. Based on the results of the ex-post forecasting (starting from January until December 2006), the ARIMA (3, 1, 2)/GARCH (1, 1) model was the best short-term forecasting model of Tawau cocoa bean price graded SMC 1B (Table 8).

Ex-Ante Forecasting

Based on the ex-ante forecasting by using the mixed ARIMA/GARCH model, Fig. 2 shows that the short-term forecasting indicated an upward trend of Tawau cocoa bean prices for the period January-December 2007.

DISCUSSION

The result showed that the time series data (starting January 1992 until December 2006) was stable. This is contradict with the previous researches (Yusoff and Salleh, 1987; Arshad and Zainalabidin, 1994) which stated that domestic cocoa bean prices are changing from time to time and very volatile. The results of the regression analysis have shown that positive linear trend factor exists in the time series data but seasonal factor was not. That means the cocoa bean prices of Tawau have increased in the period of 1992-2006 but seasonal factor which is usually related to climate change has not given any significant influence on the monthly changes of cocoa bean prices. The mixed ARIMA/GARCH model outperformed the exponential smoothing, ARIMA and GARCH for the case of forecasting monthly Tawau cocoa bean prices. This is in agreement with the findings in the literature (Zhou *et al*., 2006). Some of previous research have found that ARIMA models (Fatimah and Roslan, 1986; Shamsudin *et al*., 1992; Kahforoushan *et al*., 2010) and also GARCH-type models (Kamil and Noor, 2006) were the best or suitable price forecasting models in terms of prediction accuracy, but the accuracy of the mixed ARIMA/GARCH should also be considered in price forecasting for the future researches.

CONCLUSION

This study investigates four different types of univariate time series methods, namely exponential smoothing, ARIMA, GARCH and the mixed ARIMA/GARCH. The results showed that the mixed ARIMA/GARCH model outperformed the exponential smoothing, ARIMA and GARCH for forecasting Tawau cocoa bean prices. Forecasting the future prices of cocoa bean through the most accurate univariate time series model can help the Malaysian government as well as the buyers (e.g., exporters and millers) and sellers (e.g., farmers and dealers) in cocoa bean industry to perform better strategic planning and also to help them in maximizing revenue and minimizing the cost of price.

REFERENCES

Arshad, F.A. and M. Zainalabidin, 1994. Price discovery through crude palm oil futures market: An economic evaluation. Proceedings of the 3rd Annual Congress on Capitalising the Potentials of Globalisation-Strategies and Dynamics of Business, (CPGSDB'94), IMDA, Malaysia, pp: 73-92.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroscedasticity. J. Econ., 31: 307-327.

Celia, F., G. Ashish, R. Amar and S. Les, 2003. Forecasting women's apparel sales using mathematical modeling. Int. J. Clothing Sci. Technol., 15: 107-125.

Engel, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica, 50: 987-1007.

Fatimah, M.A. and A.G. Roslan, 1986. Univariate approach towards cocoa price forecasting. Malaysian Agric. Econ., 3: 1-11.

Kahforoushan, E., M. Zarif and E.B. Mashahir, 2010. Prediction of added value of agricultural subsections using artificial neural networks: Box-Jenkins and holt-winters methods. J. Dev. Agric. Econ., 2: 115-121.

Kamil, A.A. and A. Md. Noor, 2006. Time series modeling of Malaysian raw palm oil price: Autoregressive conditional heteroskedasticity (ARCH) model approach. Discov. Math., 28: 19-32.

Ministry of Plantation Industries and Commodities, 2006. Report on Malaysia plantation industries and commodities (2001-2005). Putrajaya.

Ramanathan, R., 2002. Introductory Econometric with Applications. 5th Edn., Thomson Learning, South Carolina, pp: 688.

Robert, R.A. and F.A. Amir, 2009. A new bayesian formulation for holt's exponential smoothing. J. Forecast., 28: 218-234.

Shamsudin, M.N., M.L. Rosdi and T.C. Ann, 1992. An econometric analysis of cocoa prices: A structural approach. J. Ekonomi Malaysia, 25: 3-17.

Shanti, G., 2006. Destination cocoa town. COPAL Cocoa Info, A Weekly Newsletter of Cocoa Producers Alliance, Cocoa Producers Alliance, Issue No. 180. May 22-26. http://www.copal-cpa.org/newsletters/Newsletter%20No%20180.pdf.

Telesca, L., M. Bernardi and C. Rovelli, 2008. Time-scaling analysis of lightning in Italy. Commun. Nonlinear Sci. Numer. Simulation, 13: 1384-1396.

Yusoff, M. and M. Salleh, 1987. The elasticities of supply and demand for malaysian primary commodity exports. Malaysian J. Agric. Econ., 4: 59-72.

Zhou, B., D. He and Z. Sun, 2006. Modeling and Simulation Tools for Emerging Telecommunication Networks. Springer, USA.