Bangladesh J. Sci. Res. **26**(1&2): 47-56, 2013 (December)

MARKET MODEL ANALYSIS AND FORECASTING BEHAVIOR OF WATERMELON PRODUCTION IN BANGLADESH

Murshida Khanam* and Umme Hafsa

Department of Statistics, Biostatistics & Informatics, University of Dhaka, Dhaka-1000, Bangladesh

Abstract

An attempt has been made to study various models regarding watermelon production in Bangladesh and to identify the best model that may be used for forecasting purposes. Here, supply, log linear, ARIMA, MARMA models have been used to do a statistical analysis and forecasting behavior of production of watermelon in Bangladesh by using time series data covering whole Bangladesh. It has been found that, between the supply and log linear models; log linear is the best model. Comparing ARIMA and MARMA models it has been concluded that ARIMA model is the best for forecasting purposes.

Key words: ARIMA, MARMA, supply model, log linear model

Introduction

Bangladesh is a developing country and facing numerous economic problems. The whole economy of Bangladesh is primarily dependent on agriculture. So the overall economic development depends on the proper development of our agricultural sector to a great extent. Watermelon is a very common delicious food with vitamin A and C which is also a good source of carbohydrate. Nowadays, it is cultivated commercially in our country and we can earn a lot of foreign currency by exporting this. So, watermelon production can play an important role for our economic development.

Salauddin (2011) has conducted a study on current guava production by considering the time series data for 23 districts of Bangladesh over a period of 27 years. The study has considered three models: CLRM (Classical Linear Regression Model), ANCOVA model and log linear model) for whole Bangladesh. Comparing all models, he found that the log linear model is the best, because the goodness of fit obtained by this model is higher than other models.

A study has been conducted regarding forecasting purpose of wheat production in Bangladesh (Karim *et al.* 2010). They found that wheat production in Bangladesh is increasing.

Rahman (2010) considered a study regarding forecasting of boro rice production in Bangladesh using data from 2008-2009 to 2012-2013. It has been observed from this analysis that short term forecast is more efficient for ARIMA models.

Shukla and Jharkharia (2011) studied the applicability of ARIMA models in wholesale vegetable market. In this study it has been found that ARIMA model can be applied to forecast the demand and price which are highly unstable.

*Corresponding author: <murshidakhanam@yahoo.com>.

So, by using the concept of previous literature, the present study has been tried to fit different types of models. This study was conducted to fit the supply and the log linear models and to identify the best one. Regarding forecasting purpose, the present study also considered the ARIMA and MARMA models and to identify the best model that may be used for forecasting purposes.

Data: This study has considered the annual time series data for the period 1990-1991 to 2009-2010 covering whole Bangladesh from various publications of BBS (Bangladesh Bureaus of Statistics) and the price of watermelon has been collected from BADC (Bangladesh Agricultural Development Council), Khamar Bari, Farmgate, Dhaka.

Variables of the study: Realizing the limitations of data availability, three variables have been considered. Production of water melon is taken as the dependent variable for all models where price has been considered as independent variable for supply, log linear and ARIMA models. In MARMA model price, distribution of improves seeds and natural calamities have been considered as independent variable.

Theoretical background: The present study has considered supply, log linear, ARIMA, MARMA models.

Supply model: A supply model shows a relationship between price and quantity of a product that a producer is willing and able to supply onto the market at a given price in a given time period. In this study the following supply model has been considered:

$$Y_{it} = a_1 + a_2X_t + U_{it} \quad (1)$$

where, Y_{it} =total production of watermelon for ith administrative division at the tth year in Bangladesh; X_t = Price of water melon; U_{it} = Disturbance terms; a_1 and a_2 are known as the regression coefficient. They also called as intercept and slope coefficients, respectively.

Log linear model: Let us consider the following regression model:

$$Y_i = \beta_1 + X_i^{\beta 2} e^{ui} \quad (2)$$

This may be expressed as,

$$\log Y_i = \log \beta_1 + \beta_2 \log X_i + u_i \quad (3)$$

If we write as

$$\log Y_i = a + \beta_2 \log X_i + u_i \quad (4)$$

where $a = \log \beta_1$.

This model is linear in the parameters a and β_2, linear in the logarithms of the variables Y and X, and can be estimated by OLS. Because of this linearity, such models are called log-log, double-log or log-linear models (Gujarati 2008)

The log-linear model, considered in this study is:

$$\log Y_{it} = a_1 + a_2 \log X_t + U_{it} \quad (5)$$

where, $\log Y_{it}$ = log of total production of watermelon for ith administrative division at the tth year in Bangladesh; $\log X_t$ = log of price of water melon; U_{it} = disturbance terms. a_1 and a_2 are known as the coefficient. They are also called as intercept and slope coefficients, respectively (Gujarati 2008).

ARIMA model: ARIMA (Auto-regressive Integrated Moving Average) model popularly known as the Box-Jenkins (BJ) methodology. Lags of the differenced series appearing in the forecasting equation are called "auto-regressive" terms, lags of the forecast errors are called "moving average" terms, and a time series which needs to be differenced to make stationary is said to be an "integrated" version of a stationary series. If we take difference a time series "d" times to make it stationary and then apply the ARMA (p, q) model to it, we can say that the original time series is ARIMA (p, d, q). ARIMA (p,d,q) is an auto-regressive integrated moving average time series, where p denotes the number of auto-regressive terms, d be the number of times the series has to be differenced before it becomes stationary, and q be the number of moving average terms (Gujarati 2008).

MARMA model: Suppose that, we would like to forecast the variable y_t using a regression model. Suppose that our regression model contains two variables, x_1 and x_2 which are not themselves collinear, as follows:

$$y_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + u_t \quad \ldots \qquad (1)$$

This equation has an additive error term that accounts for unexpected variance in y_t . The ARIMA model provides some information as to what future values of u_t are likely to be; i.e., it helps "explain" the unexpected variance in the regression equation. The combined regression time series model is $y_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + \Phi^{-1} (B)\, \theta (B)\, \psi_t$ (2)

where, ψ_t is a normally distributed error term which may have a different variance from u_t. This model is likely to provide better forecasts than regression equation (1) alone or a time series model alone. Equation (2) is an example of what is sometimes referred to as a transfer function model or, alternatively, a multivariate auto-regressive moving average model (MARMA). A MARMA model relates a dependent variable to lagged values of itself, current and lagged values of one or more independent variables, and an error term which is partially "explained" by a time series model.

The technique of MARMA modeling involves examination of partial and total auto-correlation functions for the independent variables as well as the dependent variable in an effort to specify the lag polynomials. (Pindyck and Rubinfeld 1991)

Econometric validation: For econometric validation multicollinearity, heteroscedasticity and stationarity have been checked. For multicollinearity variance inflation factor (VIF) has been calculated for each variable. And it has been found that VIF for all variables are less than 10. So, there is no multicollinearity. To detect heteroscedasticity Breusch-Pagan-Godfrey test has been used. The test indicated that there exists heteroscedasticity. To remove this problem the Prais and Houthakker approach has been used.

In this paper Box and Pierce and Ljung-Box tests have been performed to test the stationarity for all the models. The results of the above tests have been given in Table 1. In this table Q stands for the Box-Pierce Q statistic value which has been calculated. And LB stands for Ljung-Box (LB) statistic value which has been calculated. It has been found from the results of Box and

Pierce and Ljung-Box tests that the overall supply model are non stationary. So to achieve stationarity difference method has been used.

Table 1.Test results of Box and Pierce and Ljung-Box tests.

Test results of Box and Pierce test				
Model	Lag	Calculated Q	$x^2{}_{tab}$	Decision
Overall supply model	7	15.338	14.067	Non stationary
Test results of Ljung-Box test				
Model	Lag	Calculated LB	$x^2{}_{tab}$	Decision
Overall supply model	7	15.338	14.067	Non-stationary

The similar procedure has been followed for Log linear, ARIMA and MARMA models to check stationarity and to remove stationarity.

Results and Discussion

Referring Table 2, for the supply model it has been found that the value of R^2 is 0.623 which indicates that 62% of the total variation of the dependent variable production has been explained by the independent variable price. For supply model it has been found that, the coefficient of the parameter estimate of price is 12.596. The corresponding p value is 0.009 which is highly significant at 5% level of significance. That means 1 unit increase in price will lead to the increase of production with the rate of 12.596. Referring back to Table 1, for the log linear model it has been found that the value of R^2 is 0.714. That means 71% of the total variation of the dependent variable production has been explained by the independent variable. It has also been found that, the coefficient of the parameter estimate price is 0.658. The corresponding p value is 0.003 which is highly significant at 5% level of significance. That means 1 unit increase of price will lead to the increase of production with the rate of 0.658.

Table 2. Parameter estimates of the supply and the log linear models.

Name of the model	Value of R^2	Value of adjusted R^2	Parameter estimates	p values
Supply model	0.623	0.297	12.596	0.009
Log linear model	0.714	0.327	0.658	0.003

Here the value of adjusted R^2 is reduced significantly because the adjusted R^2 can be calculated by considering the potential variables in the model.

The diagnostic criteria of ARIMA model has been given in Table 3 for the overall supply model.

Table 3. The values of diagnostic criteria of ARIMA model for overall supply model data series).

Criteria	Period	Value of diagnostic criteria for the model					
		ARIMA (1,1,0)	ARIMA (2,1,0)	ARIMA (0,1,1)	ARIMA (0,1,2)	ARIMA (1,1,1)	ARIMA (2,1,2)
AIC	Estimation	454.289	456.469	**453.514**	455.634	456.461	458.182
AIC_c	Estimation	454.989	457.169	**454.214**	456.334	457.161	458.882
BIC	Estimation	457.276	460.452	**456.501**	459.617	460.443	464.157
AME	Estimation	–1148.73	–1059.48	–1516.35	–1115.51	**–1516.638**	–0.640019
	Validation	2799.80	21049.00	26909.54	21418.27	26017.071	**19942.89**
	Total	4863.54	**4467.650**	5590.116	4517.928	5366.78	4985.249
RMSE	Estimation	6028.01	5977.265	5930.756	5921.991	5909.1967	**5844.2731**
	Validation	47443.1	46513.06	48477.61	46463.66	473337.62	**44125.038**
	Total	24870.5	23825.66	24759.20	23791.18	24215.731	**22635.622**
MAPE	Estimation	7.3430	–1.238	–1.7258	–1.2994	–1.7282	–.09444
	Validation	14.6605	11.1943	14.0921	11.4459	13.804	**10.1919**
		9.1724	**1.869909**	2.2286	1.8868	2.1548	2.4771
No. of lowest values		0	2	3	0	2	5

Note: Bold values indicate the lowest value in each row.

Using this table, ARIMA (2, 1, 2) model has been selected as the best model for forecasting purpose regarding the overall supply model. The structure of the ARIMA (2,1,2) model selected for our data series is:

ARIMA (2, 1, 2): $(1+1.1166*B + 0.3749*B^2)(1 - B)\ Y_t = (1 + 1.3197*B - 0.8267*B^2)\ e_t$

For Overall supply model:

ARIMA (x = Y, order = c(2, 1, 2))

Coefficients:

	ar 1	ar 2	ma 1	ma 2
	–1.1166	–0.3749	–1.3197	0.8267
s.e.	–1.0155	–0.3515	–1.2548	0.8014

Sigma^2 estimated as 223605597: Log likelihood = –171.645, aic = 458.1

From Fig. 1, it has been observed that although the production of watermelon in Bangladesh is increasing but there are some fluctuations due to natural calamities. On the other hand, it can be said that our sample represents such type of result.

The overall supply model data series has been forecasted for the next 10 years of the selected model by using computer software R-version 2.15.1 for windows. The predicted value of year 2010-2011 to 2019-2020 of production of watermelon in whole Bangladesh has been shown in Table 4.

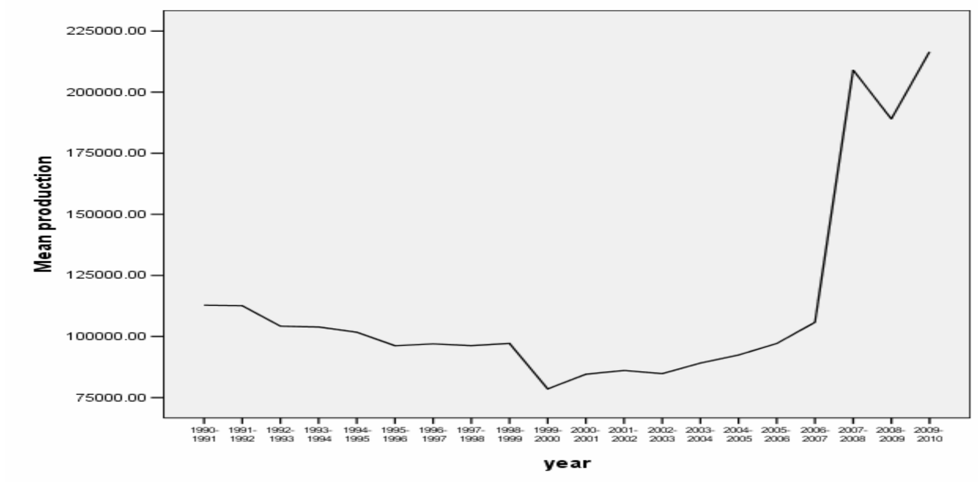

Fig. 1. Year-wise comparison of production of watermelon.

Table 4. Predicted value of year 2010-2011 to 2019-2020.

Year	Predicted value
2010-2011	170594.74
2011-2012	116861.20
2012-2013	93792.76
2013-2014	96090.87
2014-2015	107278.89
2015-2016	115709.69
2016-2017	118264.16
2017-2018	117013.20
2018-2019	114941.17
2019-2020	113711.33

The prediction of watermelon production has been shown graphically in Fig. 2.

From the Fig. 2 and Table 4 of predicted values, it has been observed that predicted values decreases first but then it increase gradually. The above figure can be explained in another way. In the Fig. 1 we have seen there is a sudden jump. The observed values have been increased and then the predicted values have been decreased suddenly. After that it can be said that there is a sign to increase slowly and gradually. For the reason behind sudden jump we cannot say that it is only the consequences of fluctuation due to natural calamities. It may be due to the fact that our sample represents such type of result.

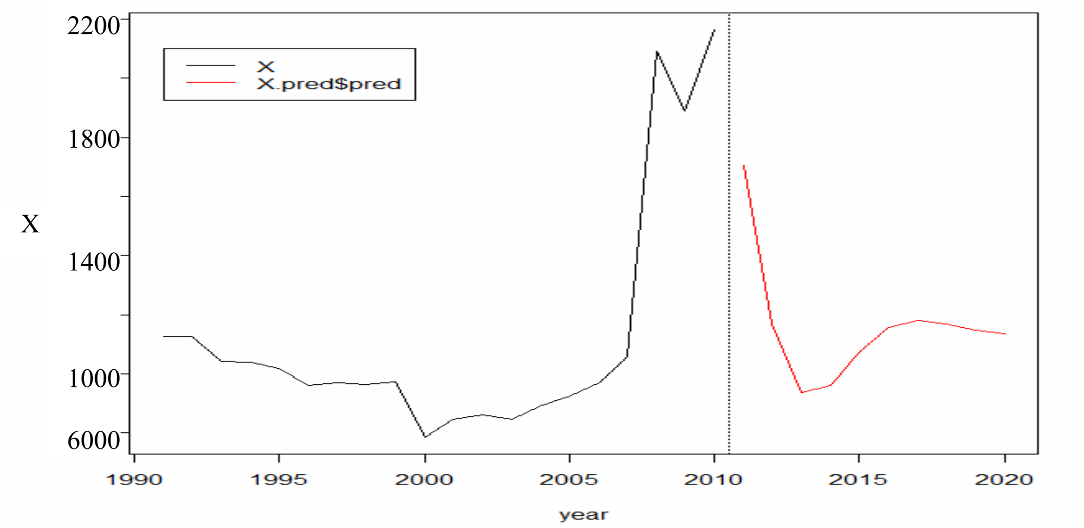

Fig. 2. Graph of predicted value of production of watermelon for overall supply model is given below.

The results of the MARMA model have been discussed by using the parameter estimates and the value of R^2. p values are compared with .05 as 5% level of significance has been considered. The estimated values for the model are given in the Tables 5 and 6:

Table 5. Parameter estimates of the MARMA model.

Variable	γ	Parameter	t	sig
Intercept	γ_o	8270.083	0.727	0.028
Price	γ_1	24.606	0.656	0.021
Distribution of improved seeds, D	γ_2	1.101	5.191	0.000
Fluctuation due to natural calamities, F	γ_3	–6.004	–0.279	0.004

Table 6. The value of R^2 and adjusted R^2 of MARMA model.

Model	R square	Adjusted R square
MARMA model for overall data series	0.721	0.669

From Table 5, it has been found that the production is positively related to price of watermelon and distribution of improved seeds but negatively related to fluctuation due to natural calamities and they are individually statistically significant. This result is very much logical in the context of general sense. Because as price increases then the production of watermelon also increases. Even then using improved seeds also responsible for a good production of watermelon.

On the other hand production of watermelon decreases as fluctuation due to natural calamities increases. From Table 6, it has been found that the value of R^2 is 0.721 and the value of adjusted R^2 is 0.669 which indicated the good fit of the model. Here the value of adjusted R^2 is reduced significantly. This is because the value of R^2 can be calculated by considering all of the variables in the model. But the adjusted R^2 can be calculated by considering those variables which are potential and significantly contributing in the model.

We have to forecast observation for next year by using "eye ball analysis" technique in MARMA model. We have compared the year-wise production of watermelon. For this comparison Fig. 1 has also been used.

Referring back to Table 7, it has been found that the predicted value of production of watermelon for the next three years in three conditions. For forecasting purpose we compared the watermelon production in the year of 2009-2010. If we consider the average of our 20 years data of price of watermelon, distribution of improved seeds, and assuming flood does not occur or if flood occurs it affects less than one third of the total area then the watermelon production may be increased 216497.0 m.ton to 227897 m.ton in year 2010-2011, 216497.0 m.ton to 248678 m.ton in year 2011-2012, 216497.0 m.ton to 268759 m.ton in year 2012-1013. If we increase price of watermelon, distribution of improved seeds by 3% and assuming flood does not occur or if flood occurs it affects less than one third of the total area (compared to 2009-2010) then the watermelon production may be increased 216497.0 m.ton to 241598 m.ton in year 2010- 2011, 216497.0 m.ton to 264857m.ton in year 2010-2011, 216497.0 m.ton to 289546 m.ton in year 2012-1013. If we increase price of watermelon, distribution of improved seeds by 4% and assuming flood does not occur or if flood occurs it affects less than one third of the total area (compared to 2009-2010) then the watermelon production may be increase 216497.0 m.ton to 267532 m.ton in year 2010-2011, 216497.0 m.ton to 284967 m.ton in year 2011-2012, 216497.0 m.ton to 298574 m.ton in year 2012-1013.

The above discussion has been placed in Table 7 considering three assumptions in the following ways:

A: Considering average of our 20 years data of price of watermelon, distribution of seeds, fertilizer, area irrigated by different methods and assuming flood does not occur or if flood occurs it affects less than one third of the total area.

B: Assume we increase price of watermelon, distribution of improved seeds, fertilizer, area irrigated by different methods by 3% and assuming flood does not occur or if flood occurs it affects less than one third of the total area.

C: Assume we increase price of watermelon, distribution of improved seeds, fertilizer, area irrigated by different methods by 4% and assuming flood does not occur or if flood occurs it affects less than one third of the total area.

Table 7. The predicted value of production of watermelon for the next three years in three conditions.

Year	Production of watermelon (in metric ton) (Compared to the year 2008-2009)		
	Assumption A	Assumption B	Assumption C
2010-2011	227897	241598	267532
2011-2012	248678	264857	284967
2012-2013	268759	289546	298574

Conclusion

According to the process of estimation of a model we have estimated our parameters of the models and calculated their respective R^2 and adjusted R^2 to compare the models. It has been observed that, for supply model R^2 which is 0.623 and is good enough that means the model is well fitted. On the other hand, log linear model also fitted well because of higher value of R^2 which is 0.714. For both the cases we have got significant result of parameters. Comparing the results of R^2 and adjusted R^2 of supply model and log linear model it can be concluded that log linear model is better than supply model because the value of R^2 and adjusted R^2 of log linear model is greater than supply model. To compare the models we have used R^2 and adjusted R^2 as they are the measures of goodness of fit.

In this study, for statistical forecasting we have studied the sample data with ARIMA and MARMA models. After obtaining the result, comparing these two forecasting models we observed that, the ARIMA model is better than MARMA model. We have used R-version 2.15.1and SPSS to identify the best ARIMA and MARMA models. We estimated the parameters and forecasting has been made for the future value of next years with the similar sample data. Moreover, we have drawn graphs to represent the current and future value in a same plot to notice the future trends of production of watermelon for both models (Fig. 2).

In this comparison, in MARMA model we have used “eye ball analysis” technique in forecasting the future values, which can be related to the chance of error. In this technique, the forecasting values might be varied based on research as well as on the researcher over similar sample data. In this situation, forecasting has been made regarding future value only for three years. This result is a little bit subjective and based on personal judgment.

On the other hand, in ARIMA model, for forecasting purpose regarding future value we have obtained some exact and accurate figure as the method is fully mathematical and the method has been performed by using R-version 2.15.1. And this method is not like as the forecasting procedure of MARMA model. Because of the method related to ARIMA model there is no chance to get different results using similar data.

In this case we can forecast for five or ten or 20 years and can get exact figure. In this research forecasting has been carried out for the next ten years. We have obtained an exact figure of forecasting value which will be acceptable and suitable comparing with the MARMA model.

So, as a final point it can be said that the ARIMA model is better than MARMA model regarding forecasting purpose.

References

BBS 1992. Yearbook of Agricultural Statistics of Bangladesh 1991, Bangladesh Bureau of Statistics, Dhaka.

BBS 1996. Yearbook of Agricultural Statistics of Bangladesh 1995, Bangladesh Bureau of Statistics, Dhaka.

BBS 2000. Yearbook of Agricultural Statistics of Bangladesh 1999, Bangladesh Bureau of Statistics, Dhaka.

BBS 2004. Yearbook of Agricultural Statistics of Bangladesh 2003, Bangladesh Bureau of Statistics, Dhaka.

BBS 2008. Yearbook of Agricultural Statistics of Bangladesh 2007, Bangladesh Bureau of Statistics, Dhaka.

BBS 2011. Yearbook of Agricultural Statistics of Bangladesh 2010, Bangladesh Bureau of Statistics, Dhaka.

Gujarati, N. 2008. Basic Econometrics, Fourth edition, Tata McGraw Hill Publishing Company Limited, New Delhi.

Karim, R., Awal and Akter. 2010. *Forecasting of Wheat Production in Bangladesh. Bangladesh J. Agril. Res. ,* 2010, **35(1)**: 17-28. Retrieved October 16 2012 http://www.banglajol.info/index.php/BJAR/article/download/5863/4598

Pindyck, R. S. and Rubinfeld, D. L.1991. Econometric Models and Econometric Forecasts. McGRAW-HILL, INC.

Rahman, N.M.F. 2010. *Forecasting of Boro Rice Production in Bangladesh - An ARIMA approach. J. Bangladesh Agril. Univ., 2010*, **8(1):** 103-112. Retrieved October 16 2012 from http://www.banglajol.info/index.php/JBAU/article/download/6406/4901

Salahuddin, G. 2011. A statistical analysis of current production of guava and its prospects in Bangladesh, Unpublished MS thesis in the Department of Statistics, Biostatistics and Informatics, University of Dhaka.

Shukla, M. and Jharkharia, S. 2011. Applicability of ARIMA Models in Wholesale Vegetable Market : An Investigation. *Proceedings of the 2011 International Conference on Industrial Engineering and Operations Management.* Retrieved October 16 2012 from http://www.iieom.org/ieom2011/pdfs/IEOM172.pdf

(Manuscript received on 8 January, 2013; revised on 9 June, 2013)