International Journal of Statistical Sciences *ISSN 1683–5603*
Vol. 8, 2008, pp 103-110

Trend Analysis of the Production of Rice in Bangladesh

Provash Kumar Karmokar
Department of Statistics
University of Rajshahi,
Rajshahi-6205, Bangladesh
Eimail: sprovash@yahoo.com

A.H.M. Rahmatullah Imon
Department of Mathematical Sciences
Ball State University
Muncie, IN 47306, USA

[Received October 19, 2006; Revised August 10, 2008; Accepted November 1, 2008]

Abstract

The trend analysis of the production of rice is hugely important in a thickly populated country like Bangladesh, where rice is the main food and where, the entire economy hugely depends on the production of rice. But for the understanding of the trend in the production of rice, it is also very crucial to know which trend model is being used in the study, as a number of trend models are available in the literature for the same purpose. If we use a trend model, which is not appropriate, the subsequent analysis will be faulty and they will be no longer reliable. So the model that we use for the analysis must go through a tough theoretical justification. In this paper we have used a variety of recent diagnostic techniques to find a trend model which is appropriate in explaining the trend of rice production in Bangladesh.

Keywords and Phrases: Trend Analysis, Diagnostics, Normal probability plot, RM test for normality, Unusual observations.

AMS Classification: 62J05, 62J20, 62M10, 62G35.

1 Introduction

Rice is the most important food in the world. More than half of the world's population depends on rice for food calories and protein, especially in developing countries. In most of Asia, rice is not only the staple food, but also constitutes the major economic activity and a key source of employment and income for the rural population. More than 250 million farm families cultivate rice in Asia (Fischer and Cordova, 1998). Rice is a staple crop in Bangladesh and its production has to be enhanced to meet the food requirement of an over populated country where the size of the population is still going fast. That is why it is hugely important to investigate the trend of production of rice in Bangladesh.

A variety of trend models are now available in the literature [see Gupta et al. (1999), Sahu (2003)] for analyzing the production of major crops in a country. Among them the linear trend model, the quadratic trend model and the compound trend model are very commonly used. Some efforts have been made to analyze trends of production of crops over the years using these models. But we observe that results obtained from most of the studies are not much supported from diagnostic viewpoint. It is now evident that any statistical analysis that lacks diagnostics checking could produce much worse result than one can anticipate. All of the commonly used trend models largely depend on two basic assumptions: the normality of errors of the model and the non-existence of unusual observations (mainly outliers) in the data. In section 2, we introduce several trend models that have greater applications in explaining trends of crop productions. We also introduce several diagnostic methods in this section that are essential for examining the validity of assumptions the trend models require. In section 3, we investigate different trend models from diagnostic viewpoint using the data of the production of rice in Bangladesh for the last 43 years.

2 Trend Models and Diagnostics

In this section we introduce some commonly used methods available in the literature for analyzing trends of productions of crops. Here we consider (a) linear trend model, (b) quadratic trend model, (c) compound trend model as considered by Sahu (2003) in his study on production of four major crops of West Bengal in India.

If we denote the production of rice by Y and the year of production by t, the linear trend model becomes

$$Y = \alpha + \beta t + \in \tag{1}$$

The quadratic trend model can be defined as

$$Y = \alpha + \beta t + \gamma t^2 + \in \tag{2}$$

The compound trend model takes the form

$$Y = \alpha \beta^t \in \tag{3}$$

Apparently the compound model looks like a non-linear model, but taking logarithmic on both sides of (3) yields

$$Y^* = \alpha^* + \beta^* t + \in^* \tag{4}$$

where $Y^* = \ln Y$, $\alpha^* = \ln \alpha$, $\beta^* = \ln \beta$, $\in^* = \ln \in$. For all of the above models we have included the error term $\in$ and the basic assumption about the error term $\in$ is that it follows a Normal distribution with zero mean and constant variance.

It is a common practice over the years to judge the merit of a fitted model by the inspection of its R^2 values. This method is very simple, the higher value of R^2 indicates a better fit of the model. But the main limitation of R^2 is that it does not consider the number of explanatory variables used in the model and hence is not suitable for model comparisons. To overcome this problem we can consider the adjusted R^2 ($\bar{R}^2$). But it is now evident that [see Ryan (1997)] both R^2 and $\bar{R}^2$ can often produce misleading result if it is computed without the consideration of the assumptions it requires to show its merit. The basic assumption of having a satisfactory R^2 and/or $\bar{R}^2$is the normality of errors of the model. Other important assumption is the nonexistence of unusual observations in the data. The violation of these two basic assumptions could have drastic consequences on the fitting of the model and the subsequent analysis. Hence it is essential to check these assumptions before applying any statistical techniques to the data.

There is now a very large body of literature on tests for normality of errors. One very popular graphical method for testing normality of errors is the normal probability plot of residuals. Here the ordered residuals are plotted against their cumulative theoretical (normal) probabilities. Under normality, the resulting points should lie approximately on a straight line. Substantial departures from a straight line indicate non-normality of errors. In this paper we would use normal probability plots of residuals resulting from the different trend models we considered. Although the normal probability plots are very popular, we cannot always depend on these plots. The decision based on graphical methods may depend on the personal choice of the experimenter, so it is always desired to consider analytic formula for normality tests.

A variety of analytic tests are available in the literature for testing normality of errors, among them the Shapiro-Wilk normality test, the Anderson-Darling normality test and the Jarque-Bera normality tests have become very popular with the statisticians. Both the Shapiro-Wilk and the Anderson-Darling normality tests are computationally extensive and cannot be used without the help of special tables designed for these tests. On the other hand, the Jarque-Bera (JB) test is computationally very simple. This test requires the only coefficients of skewness and kurtosis of the observations and can be done without the help of any special tables. The JB test statistic is defined as

$$RM = \frac{b_1^2}{\lambda_1} + \frac{(b_2 - 3)^2}{\lambda_2} \tag{5}$$

where b_1 and b_2 are sample skewness and kurtosis of the residuals and $\lambda_1 = \frac{6}{n}$ and $\lambda_2 = \frac{24}{n}$ are the asymptotic variances of the sample skewness and kurtosis. Under normality the JB statistic follows a chi-square distribution with 2 degrees of freedom.

But the main disadvantage of the JB test is that it often suffers from possessing very poor power [see D'Agostino (1986)] when the errors are not normal. In recent years test based on Rescaled Moment (RM) of residuals suggested by Imon (2003) has become popular with the statisticians. Like the Bowman-Shenton test, this test requires only the coefficient of skewness and kurtosis of the residuals but in a rescaled form that have a very simple relationship with the original ones. This test can be performed without the help of any special tables, but the main advantage of this test is that it possesses very good power when the errors are not normal. The rescaled moments suggested in this test can be expressed in terms of the original moments as

$$m_2^*(\hat{\in}) = \tfrac{n}{n-p} m_2(\hat{\in}) = c m_2(\hat{\in})$$
$$m_3^*(\hat{\in}) \approx c^3 m_3(\hat{\in})$$
$$m_4^*(\hat{\in}) \approx c^4 \left[m_4(\hat{\in}) - 3\left(1 - c^{-2}\right) \{m_2(\hat{\in})\}^2 \right]$$

where $\hat{\in}$ is the set of OLS residuals, $m_k^*(\hat{\in}) = \frac{1}{n} \sum \hat{\in}^k$ and $c = \frac{n}{(n-p)}$

Using the above results we obtain the rescaled coefficient of skewness and kurtosis of residuals as

$$b_1^* = \frac{m_3^*}{m_2^{3/2}} = \frac{c^3 m_3}{c^{3/2} m_2^{3/2}} = c^{3/2} b_1 \tag{6}$$

$$b_2^* = \frac{m_4^*}{m_2^{*2}} = \frac{c^4 \left[m_4 - 3\left(1 - \frac{1}{c^2}\right) m_2^2 \right]}{c^2 m_2^2} = c^2 b_4 - 3\left(c^2 - 1\right) \tag{7}$$

Substituting the values of (6) and (7) in (5) we obtain the *RM* statistic for testing normality as

$$RM = c^3 \frac{n}{6} \left[b_1^2 + c \frac{(b_2 - 3)^2}{4} \right] \tag{8}$$

In this paper we would use the RM test for normality to see whether the trend models are satisfying the normality of errors or not.

We would also like to investigate the presence of autocorrelation and the presence of any unusual observations like outliers and influential observations. We employ the time series plot of residuals and the Durbin-Watson (DW) test to detect the presence of autocorrelation which affects the efficiency of the estimators. The existence of unusual observations does not necessarily mean that these observations are bad, it may indicate that the fitted models are inappropriate to accommodate the observations. So the problem of outliers can be viewed as the problem of inappropriateness of the model used to fit the data. There are a large number of diagnostic methods available in the literature to identify unusual observations. Here we consider the deleted Studentized residuals (often referred to as R-Student residuals) proposed by Cook and Weisberg (1982) to identify outliers and Cook's distance proposed by Cook (1977) to identify influential observations.

3 Trend Analysis of Production of Rice

In this section we offer a comparison between some commonly used methods available in the literature for analyzing trends of productions of crops from diagnostic point of view. We consider linear trend model, quadratic trend model and compound trend model as they are defined in the previous section. We use the data of FAOSTAT for the production of rice in Bangladesh from the year 1961 to 2003.

To have an idea about which of the above three models are appropriate we at first make a simple scatter plot of production of rice against the year of production as it in given in Figure 1(a). We observe from this plot that there is a nearly linear relationship between the production of rice and the year of production except few observations at the top right corner of the plot. This plot suggests that a compound trend model could be more appropriate for this data. This plot also suggests that a quadratic trend model could come in a little use for this data. This plot also suggests that a quadratic trend model could come in a little use for this data. Then we plot log production of rice against year as given in Figure 1(b) and it appears that, this plot is showing more linearity between the two variables than the former one.

Figure 1: Scatter plot of (a). Production of Rice (b). Log Production of Rice against Year

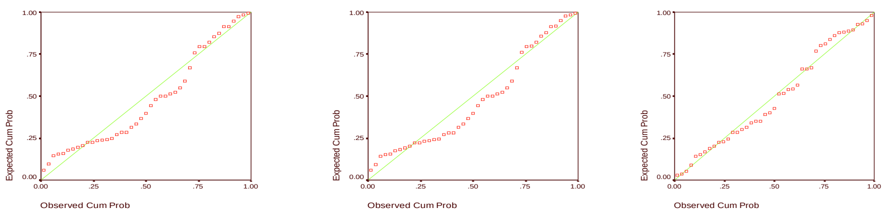

Figure 2: Normal probability plot of residuals obtained from (a) the linear trend model (b) the quadratic trend model (c) the compound trend model

We have fitted the three trend models using the data of rice production of Bangladesh from the year 1961 to 2003 by using the ordinary least squares (OLS) method. As we know that the OLS method heavily depends on the normality assumption of errors, we have presented the Normal probability plots of residuals [Figures 2(a)-2(c)] for each model. We observe from these plots that the linear and the quadratic trend models are not satisfying the normality assumption of the random errors. The linear trend model is slightly better than the quadratic model but its performance is not satisfactory. The wiggles of curves as shown in Figures 2(a) and 2(b) indicate that both the linear and quadratic trend models may be affected by outlying observations. However, the compound trend model perfectly obeys the normality assumption of the errors and performs best overall to explain the trend of production of rice in Bangladesh.

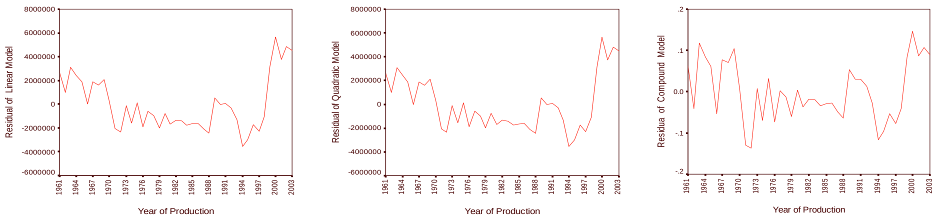

Figure 3: TS Plot of residuals for (a) linear trend model (b) the quadratic trend model (c) the compound trend model

Since the rice production data is a time series data, there is a strong possibility that the problem of autocorrelation might occur in this data. We have considered the time series (TS) plot of residuals for the three fitted trend models as shown in Figure 3. We observe from these plots that there is evidence of first order positive autocorrelation for all trend models considered for the rice production data of Bangladesh. But the autocorrelation problem is severe for linear and quadratic trend models but mild for the compound model.

Table 1 offers an analytic comparison of the three trend models when they are fitted by the OLS method. The results of the coefficients of the three fitted models together with their corresponding p values are presented in this table, which also presents the value of, adjusted, the values of JB and statistics for normality tests, the Durbin-Watson statistic for autocorrelation, maximum R-Student residual for the detection of outliers and maximum Cook's distance for the detection of influential observations.

Table 1: Diagnostics associated with the trend models for the production of rice in Bangladesh

Components	Linear	Quadratic	Compound
Estimating form of $\hat{Y}$	- 1004270884 + 518123.708 t (0.0001) (0.0001)	-491204420.9 + 518123.708 t + 130.802 t^2 (0.0001) (0.0001) (0.0001)	(1.00653E-12) 1.0227^t (0.0001) (0.0001)
R^2	0.891	0.892	0.941
Adjusted R^2	0.888	0.889	0.939
JB (4.61)	4.710	4.610	1.140
RM (4.61)	5.437	5.733	1.374
Max. R-Student residual (2.5)	2.737	2.741	2.250
Max. Cook's distance (1.0)	0.252	0.253	0.173
DW	0.420	0.424	1.003

We observe from the results presented in Table1 that the compound trend model is the best overall, as it possesses the highest R^2 and adjusted R^2 values among the three trend models considered in our study. The Jarque-Bera and RM test clearly show that the compound trend model obeys the normality assumption of errors. But both the linear and the quadratic trend models show nonnormality at the 10% level of significance and the RM test possesses better power than the JB test. Cook's distances for all the observations are less than the cut-off values presented in parentheses in Table 1. But we observe from the same table that both the linear and quadratic models are inappropriate to fit this data as they have at least one outlier each as the values of R-Student residuals indicate. However, the compound trend model is free from this problem. The DW statistic shows that severe autocorrelation problems exist for both linear and quadratic trend models, but this problem is only moderate for the compound trend model. So, we finally declare the compound trend model as the most suitable method from the diagnostic point of view for examining the trend of rice production in Bangladesh.

4 Conclusions

In this paper our main effort was to find out a suitable trend model for the production of rice in Bangladesh. In this paper we try to emphasize that the model under consideration must go through a tough theoretical justification, otherwise the inference may not be reliable. Here we compared the performances of the most commonly used linear, quadratic and compound trend models using the rice production data of Bangladesh for the last 43 years. Our study shows that the compound trend model outperforms the other two models in all respects. In this paper we also see that Bangladesh is maintaining a compound growth rate of 1.0227 per year in the production of rice, which is very encouraging.

Acknowledgements

The authors gratefully acknowledge all helpful comments and suggestions made by the reviewers to improve this paper.

References

[1] Cook, R.D. (1977). Detection of influential observations in linear regression, Technometrics, 19, 15-18.

[2] Cook, R.D. and Weisberg, S. (1982). Residuals and Influence in Regression, Chapman and Hall, London.

[3] D'Agostino, R.B. (1986). Tests for normal distribution, in Goodness-of-fit Techniques, R.B. D'Agostino and M. Stephens (Eds.), Marcell-Decker, New York.

[4] Fischer, K.S. and Cordova, V.G. (1998). Impact of IRRI on rice science and production, in Impact of Rice Research, P.L. Pingali and M. Hossain (Eds.), Thailand Development Research Institute, Bangkok, and International Rice Research Institute (IRRI), Philippines, 27-50.

[5] Gupta, D.S., Sahu, P.K., Sanyal, M. K. and Pal, S.R. (1999). Growth and trend analysis of potato production in West Bengal, J. Interacad, 3 (3&4), 345-350.

[6] Imon, A.H.M.R. (2003). Regression residuals, moments and their use in tests for normality, Commun. Statistist.-Theory and Meth., 32, 1021-1034.

[7] Sahu, P.K. (2003). Production model: In search of alternative food, Presented in Prof. K M Hossain International Seminar-Cum-Workshop, Department of Statistics, University of Rajshahi, Bangladesh, October 2003.

[8] FAOSTAT: http://www.faostat.fao.org/