OPENING MINDS: RESEARCH FOR SUSTAINABLE DEVELOPMENT

Temporal Patterns Analysis of Paddy Production in Sri Lanka

N.B.W.I. Udeshika* and T.M.J.A. Cooray

University of Moratuwa, Katubedda, Sri Lanka

**Corresponding author: Email:imaliudeshika88@gmail.com*

1 INTRODUCTION

Rice is the main crop cultivated by the majority of farmers in rural areas and it is the staple food of approximately 20 million inhabitants in Sri Lanka. Hence, rice sector makes a significant contribution to the economy of Sri Lanka. About 30% of the crop sector contribution to the agricultural GDP (Gross Domestic Product) is from the rice sector. Approximately 800,000 farm families, which are about 20% of the population, depend on paddy cultivation for their livelihood (Statistics. (2017).

In Sri Lanka paddy cultivation is mainly divided into two seasons (time periods) known as "Maha" and "Yala" which are associated with the two monsoons. Maha season is the main season in paddy cultivation associated with the north-east monsoon during the period of September to March. Yala is the secondary season which is associated with south-west monsoon during the period of May to August (Statistics. (2017). However, the whole area devoted for paddy is not being cultivated due to number of reasons such as shortage of water during the seasons, the prevailing unsettling conditions on the ground, etc.

2 METHODOLOGY

2.1 Seasonal Autoregressive Integrated Moving Average (SARIMA) Model

Log transformation is applied to the data series to remove the non-constant variance of the series. Four SARIMA models are developed for log transformed series. Numerous statistics such as AIC, SBC, $R^{2,}$,DW are used to identify the most adequate model among formed SARIMA models. Validity of the assumptions of the fitted model are checked by considering results of the hypothesis tests specifically; Box –Pierce test, Serial Correlation LM Test and Histogram Normality Test**.**

2.2 Vector Error correction (VEC) model

Integration order of the secondary data series is investigated to build up a multivariate approach for the paddy production. Consequently Vector Error correction model (VEC) is fitted including disequilibrium term. VEC Lag exclusion wald test, Portmanteau Test for Autocorrelations and Lagranch Multiplier test are used to examine the goodness of fit of the formed VEC model.

3 RESULTS AND DISCUSSION

Table 1: Parameter Estimates of SARIMA Model

Model	AIC	R^2	DW	Q Statistic*	Test1**	Test 2***
SARIMA $(011)(010)_2$	-0.3798	0.4747	1.8767	Sig	Not sig	Sig.
SARIMA $(010)(011)_2$	-0.4210	0.4959	2.8174	Not sig	Not Sig	Sig
SARIMA $(011)(011)_2$	-0.7958	0.6592	1.8811	Not Sig	Not Sig	Not Sig
SARIMA $(111)(011)_2$	-0.7108	0.6249	2.0179	Not Sig	Not Sig	Not Sig

* Ho: No serial correlation up to lag 12 of residual series, H1: There exists serial correlation up to lag 12 of residuals series
** Normality Test (H_0: Residuals are normally distributed, H_1: Residuals are not-normally distributed)
*** Correlation Test (H_0 : Residuals are uncorrelated, H_1: Residuals are correlated)

When comparing four models, SARIMA (011) $(011)_2$ model has taken smallest AIC value, largest R^2 value. Further DW statistic value of that model is much closer to 2. Q statistic, Normality Test and correlation Test have given insignificant results. Hence SARIMA (011) $(011)_2$ model can be selected as the most adequate model to capture the trends and seasonal patterns of the paddy production series.

3.1 Diagnostic checking of SARIMA $(011)(011)_2$Model

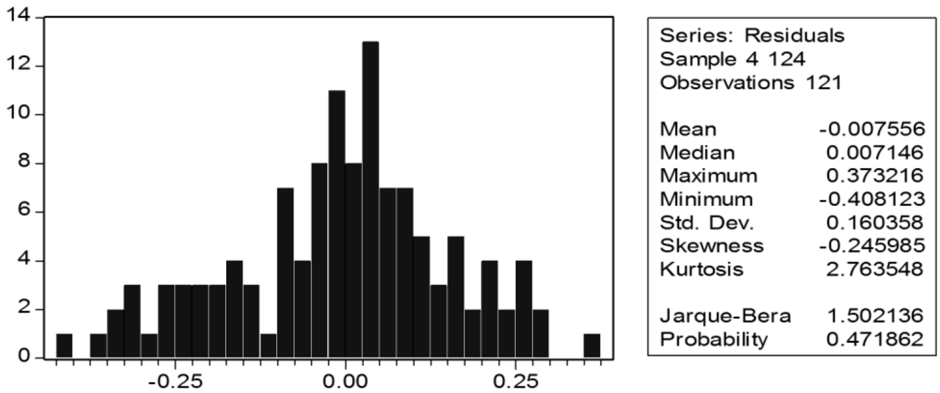

Figure 1: Histogram of Normality Test

When considering the model adequacy of the fitted model, residuals are fairly normally distributed when the few large and small values are not considered. Results of the Correlation LM test is given that no correlations among residuals. Q-

ISSN 2012-9916

statistics of the correlogrm of standardized residuals are not significant. Hence there is no serial autocorrelation among residuals of the fitted model.

3.2 Model Equation

$$(\mathbf{1}-\boldsymbol{B})(\mathbf{1}-\boldsymbol{B}^{\mathbf{2}})\boldsymbol{lny_t} = (\mathbf{1}-\mathbf{0.784903}.\boldsymbol{B})(\mathbf{1}-\mathbf{0.967740}.\boldsymbol{B}^{\mathbf{2}})\boldsymbol{e_t} \ldots\ldots\ldots\ldots(1)$$

yt = Paddy production in season t
B = Backshift Operator

3.3 Validating the requirements for Multivariate Model

Table 2: Results of unit root test for logarithmic variables (1953-2013)

Variable	Probability of log series	Significant	Order
Paddy production (y)	0.4516	Not Sig.	Cannot identified
Harvested Area(y_1)	0.2587	Not Sig.	
Rain fall (y_2)	0.002	Sig.	I(0)
	Probability of 1st difference of log series		
Paddy production (y)	0.0000	Sig.	I(1)
Harvested Area(y_1)	0.0000	Sig.	I(1)

The Table 2 illustrates that paddy production(y) and harvested area(y1) logarithmic series have taken equal orders(I(1)) while rain fall (y2) logarithmic series has taken order 0. Therefore only y and y1 variables can be used to build Multivariate time series model.

3.4 Test for Cointegration

According to the cointegration rank test at most one cointegration equation exists at 5% level of significance and the cointegration by maximum Eigen value indicates same results. Thus there exists a cointegration equations and it implies that variables are cointegrated. A Vector Error correction model has to be fitted including the disequilibrium term. So disequilibrium terms are added to the model as explanatory variables.

3.5 The vector Error correction estimates

Table 3: Results of ECM for LNY and LNY1 (1952-2013)

Variables of Error Correction Model	D(LNY)	D(LNY1)
CointEq1	0.450530*	0.479370*
D(LNY(-1))	-1.398502*	-0.807413*
D(LNY(-2))	-0.519834*	-0.483595*
D(LNY1(-1))	1.082698*	0.535566*
D(LNY1(-2))	0.796267*	0.760925
C	0.037602*	0.021614

*Significant at 5%, LNY= log(y), LNY1= log (y1)

3.6 Diagnostic checking of VEC Model

Residual Portmanteau Tests for Autocorrelations is indicated that residuals of the model are uncorrelated. Residual Portmanteau Tests for Autocorrelations is also concluded the adequacy of the model. Residual plots are indicated that residuals are randomly distributed (not white noise). Apart from the very few data points most of the data points of the correlograms are inside the bandwidth, showed that the auto correlation function support the stationary of the model.

3.7 Model Equation

d(lny) = 0.4505297098*(lny(-1) - 2.723984106*lny1(-1) + 9.17796832) (2)

1.398502266*d(lny(-1)) - 0.5198337118*d(lny(-2)) + 1.082697997*d(lny1(-1)) +

0.7962672213*d(lny1(-2)) + 0.0376020718

d(lny1) = 0.4793701563*(lny(-1) - 2.723984106*lny1(-1) + 9.17796832) - (3)

0.8074134735*d(lny(-1)) - 0.4835952083*d(lny(-2)) + 0.5355655228*d(lny1(-1))

In above equations 2 and 3, the cointegrating variables are estimated together, alternating dependent variable and independent variable. The model is fit to the first differences of the non-stationary variables, but a lagged error-correction term is added to the relationship. Since two variables exist here, error correlation term is the lagged residual from the cointegrating regression, of one of the series on the other in levels. It expresses the prior disequilibrium from the long-run relationship, in which that residual would be zero.

ISSN 2012-9916 © The Open University of Sri Lanka

3.8 Forecasting and Model Comparison

Table 4: Actual versus forecasted values

Year	Season	Actual Value (Ha.000)	Forecast Value of SARIMA Model (Ha.000)	Forecast Value of VEC Model (Ha.000)
2014	Yala	1145	1147.086	1412
2014	Maha	2877	2683.481	2456
2015	Yala	1942	1616.341	1764
2015	Maha	2902	2756.867	2433
MAPE			**7.16%**	**15.81%**

SARMA model has given smallest MAPE value.

4 CONCLUSIONS AND RECOMMENDATIONS

SARIMA $(011)(011)_2$ model is the best model which can be used to forecast the paddy production in Sri Lanka based on MAPE value. The best model which was developed for paddy production data captures 65.92% of variation (R^2=0.6592) of the original log series.

Since order of the series is equal, paddy production data and harvested area data can be used to develop a multivariate model; however, rain fall data cannot be included in the model due to uneven order.

The data set consist of only 124 data points, if it can increase to a larger value it will could support the development of a long-term forecasting model. The VEC model was fitted using Eviews software and it does not have an option to forecast just for next few data points exclusive of all data points of the fitted model. As future works VEC model can be improved with new software

REFERENCE

Statistics. (n.d.). Retrieved March 2017, from Ministry of Agriculture: http://www.agrimin.gov.lk/

Paddy Statistics. (n.d.). Retrieved March 2017, from Department of Cencus and Statistics: *http://www.statistics.gov.lk/*

Statistics. (2017, March). Retrieved from The Central Bank of Sri Lanka: *http://www.cbsl.gov.lk/*