

WHEAT PRODUCTION TRENDS IN PAKISTAN- A STATISTICAL ANALYSIS

Mukhtar Ahmed , Muhammad Salim Khan** and Nasir Iqbal**

ABSTRACT

The present study was carried out at statistical Section, Ayub Agricultural Research Institute, Faisalabad, Pakistan during the year 2013 to examine the production and productivity trends of wheat in Pakistan for the period 1947-48 to 2012-13. Several nonlinear models are applied to fit the trends. Most suitable model selected on the basis of maximum Adjusted R square, relatively minimum values of mean absolute error, root mean squared error and normally, independently, identically distributed error term. None of the parametric models is selected to fit trend of area sown under wheat in Pakistan, nonparametric regression model with kernel smoothing was employed. Sinusoidal model was selected as a suitable model for the both production and productivity of wheat in Pakistan as it fits the trends with 99 percent predictability in case of production and 98 percent for productivity alongwith all the satisfactory assumptions regarding error term. The root mean squared error of Sinusoidal model for production (0.779) and productivity (99.58) are relatively lower than all other models compared in the study to fit the trends. Mean absolute error also observed minimum in case of Sinusoidal model for both the production and productivity with values 0.623 and 84.778 respectively. The rate of yearly change due to suggested model is 2.95 percent in case of production and 1.89 percent for productivity.

KEYWORDS: Nonlinear models, Adjusted R square, Mean Absolute Error, Root Mean Squared Error, Sinusoidal model.

INTRODUCTION

Agriculture is a major sector in Pakistan's economy which not only contributes 21.4 percent in GDP, and 45% to employment but also plays a key role in development of other sectors concluding its significance (Economic Survey of Pakistan, 2012-13). Wheat, being an imperative grain and staple food, was sown on 220.39 million hectares in 2011 and 704.80 million tons with an average production of 3005 kg per hectare around the

*Statistical Section, Ayub Agricultural Research Institute, Faisalabad, **Department of Statistics, Government Post Graduate College, Jhang, Pakistan.

world. Among wheat producers worldwide, China has emerged as topper followed by India and Russian Federation. Even with production of 117.41 million tons, China is forced to import wheat to feed its 1.2 billion population. India and Pakistan producing 80.80 & 25.21 million tonnes of wheat respectively ranked second and seventh amongst top wheat producers globally. Pakistan exported 413.7 thousand tons of wheat in 2011-12. Wheat requirement in future for such a rapid growth of population needs special attention and forecasting to be able for effective planning regarding production, consumption, marketing, storage, domestic and foreign trade, etc.(17). Careful selection of statistical models is of great importance for the purpose of forecasting and growth rates.

Parametric models, assuming linear or exponential forms, are mostly used to estimate growth rates for different crops. A number of researchers (3, 9, 10, 11, 12, 13, 14, 20) have used parametric nonlinear models. Sinusoidal model found as suitable in an earlier study (13) to fit the trends in production and productivity in India alongwith non parametric regression as best fit for area under wheat.

The objective of the present study was to develop appropriate statistical model through evaluating the trends and calculation rate of change in production and productivity for wheat crop grown in Pakistan.

MATERIALS AND METHODS

This study was conducted at Statistical Section, Ayub Agricultural Research Institute, Faisalabad, Pakistan out on the basis of time-series data pertaining to the period 1947-48 to 2012-13 of Punjab province and Pakistan. The data was collected from various annual publications named Agricultural Statistics of Pakistan by Ministry of National Food Security and Research (Economic Wing), Government of Pakistan.

Non-linear parametric models (1, 2, 4, 8, 15, 18) were employed using R (16). The following models (Table1) were analyzed for the final solution by trying several initial estimates of parameters obtained using different methods in literature. Some other models were also considered but not discussed in this study because of lower predictability.

Nonparametric regression model with additive error is of the form i.e.

$$Y_i = m(x_i) + \varepsilon_i$$

Here Y_i is the observation (area, production and productivity) of the i^{th} time point, m is the trend function, which is assumed to be smooth, and ε_i is

random error with mean zero and finite variance, $\sigma^2 < \infty$. The kernel weighted linear regression smoother is used to estimate the trend function (8, 13).

Table 1. Models compared in the study.

Model name	Short name	Model expression
Sinusoidal	SND	$Y_t = P_1 + P_2 \text{Cos} (P_3 t + P_4) + \varepsilon_t$
Morgan Mercerer	MM	$Y_t = \frac{(P_1 P_2 + P_3 t P_4)}{(P_2 + t P_4)} + \varepsilon_t$
Gompertz	GMTZ	$Y_t = P_1 \text{EXP} (-\text{EXP} (P_2 - P_3 t)) + \varepsilon_t$
Guassian	GUAS	$Y_t = P_1 \text{EXP} (-0.5 (t - P_2/P_3)^2) + \varepsilon_t$
Logistic	LGS	$Y_t = \frac{P_1}{1 + P_2 \text{EXP} (-P_3 t)} + \varepsilon_t$
Yield Density	YD	$Y_t = (P_1 + P_2 t + P_3 t^2)^{-1} + \varepsilon_t$

P_1, P_2, P_3, P_4 are parameters and ε_t is error due to the model.

The statistically most suited models were selected on the basis of coefficient of determination (R-square), adjusted R-square, significant regression coefficients, Root Mean Square Error (RMSE), Mean Absolute Error (MAE) values and assumptions of residuals normality (19), independence (Durbin Watson Statistic) and randomness (Run Test) (4, 6). (D'Agostino & Stephens, 1986; Garson, 2012). Once identification of the best model for a data set is made then next step is to compute the compound growth rate (11), which is given as:

$$\frac{D[\ln y]}{dt} = Y^{-1} \frac{dy}{dt} \dots\dots\dots (1)$$

RESULTS AND DISCUSSION

Several non-linear models are studied to fit the trends in area sown under wheat, production and productivity in Pakistan. Suitable models are suggested on the basis of several verification of assumptions.

Area trends

Morgan-Mercer-Flodin and Sinoidal models, employed maximum adjusted R squared (98%) and comparatively lower values of RMSE (MM = 217.878, SND = 221.92) and MAE (MM = 177.910, SND = 175.392) to that of other

non-linear models in study (Table2). Shapiro wilk test of normality found statistically non-significant and randomness due to run test found significant which assures that assumption about randomness of error term is violated by both of these models. Thus none of parametric non-linear models are suitable to fit trends in area sown under wheat in Pakistan. To accomplish the stipulated objective nonparametric regression model with kernel smoothing is used. The optimum band width using cross validation method is computed as 1.738. Residual due to nonparametric regression model showed that the assumptions of independence of errors is not violated. The RMSE and MAE values were found to be 137.649 and 109.327 respectively. Hence the nonparametric regression model was selected to fit the trends in area under the wheat crop. The graph of the fitted trend for area under the wheat crop using the nonparametric regression is depicted in the Fig. Study of Rajarathinum and Vinoth, (13) depicted that nonparametric model with optimum bandwidth of 0.50 was employed to fit the trends in area of wheat.

Table 2. Characteristics of fitted models for Area sown under wheat in Pakistan.

Model	Parameters				Goodness of Fit							
	P1	P2	P3	P4	R ²	R ² _{adj}	S.W.T	R. T	D.W.S	RMSE	MAE	
SND (S.E)	6156.756* 146.8785	2517.256* 185.76	0.041* 0.0041	3.618* 0.1685	0.980	0.979	0.715	0.002	0.974	221.920	175.392	
EXP (S.E)	4331.699* 82.4312	0.0121* 0.0004	—	—								
GMTZ (S.E)	11288.424* 600.3759	0.135* 0.0351	0.025* 0.0026	—	0.975	0.974	0.172	0.000	0.775	250.134	214.003	
GUAS (S.E)	8899.153* 159.4196	72.962* 3.3907	54.704* 2.4996	—	0.978	0.978	0.599	0.000	0.877	234.357	196.205	
LGS (S.E)	10160.287* 322.9	1.793* 0.0678	0.039* 0.0027	—	0.977	0.976	0.371	0.000	0.837	240.224	202.985	
NM (S.E)	4118.739* 95.5187	2048.1083 1368.0069	10130.735* 466.9171	2.131* 0.2235	0.981	0.980	0.911	0.009	1.016	217.878	177.910	

* Significance at 5%, S.E= Standard Error, S.W.T= Shapiro Wilk Test, R.T= Run Test, D.W= Durbin Watson Statistic, RMSE= Root Mean Squared Error, MAE= Mean Absolute Error.

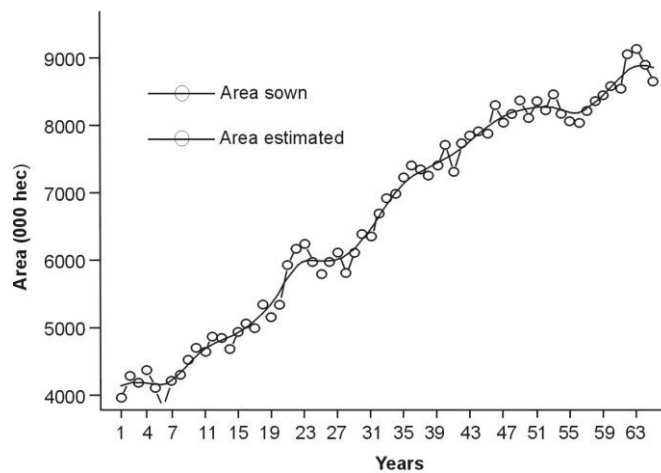


Fig. 1. Area sown under wheat in Pakistan (1947-2013)

Production trends

Sinusoidal model, with maximum predictability (Pakistan = 99%, Punjab = 98%) and all parameters significant at 1%, is found suitable for production trends in Pakistan as well as in Punjab because of lower values of RMSE (Pakistan = 0.779, Punjab = 0.652) and MAE (Pakistan = 0.623, Punjab = 0.501) (Table 3). Assumptions of normality, independence and no autocorrelation of residual are also satisfied by this model. Rajarathinum and Vinoth, (13) has also reported sinusoidal model as a suitable fit for the production trends of wheat in India.

In the starting period, rate of yearly change is negative indicating a slight downward trend, and Afterward increasing trend in both cases of Punjab and Pakistan with overall average rate of changes was calculated from sinusoidal model as 3.11 and 2.95 percent, respectively.

The graph of fitted trend for wheat crop production using the Sinusoidal Model is depicted in the Fig. 2.

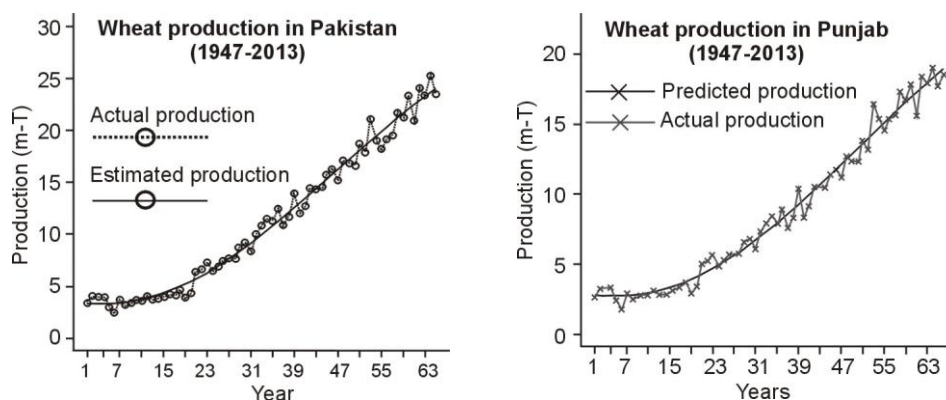


Fig. 2. Wheat production in Punjab and Pakistan.

The predictions due sinusoidal model for production in Punjab and Pakistan are given in Table 4.

Productivity trends

The results (Table 4) for productivity of wheat crop in Pakistan revealed that among the fitted non-linear models, maximum adjusted R squared of 98% is observed in case of Sinusoidal model with comparatively lower values of RMSE (99.577) and MAE (84.778) than other models. (Table 5) Sinusoidal

model for the case of wheat productivity in Punjab depicted comparatively lower values of root mean squared error (117.252), mean absolute error (96.890) and higher value of adjusted R squared 97%. Its parameters are highly significant at 1% level of significance and also satisfied all the assumptions as well regarding residuals of model for both Pakistan and Punjab. In comparison with the study of India, the same (Sinusoidal) model with different parameters is suggested suitable to fit the trends of wheat productivity (13). In period 1947-53 rate of yearly change is negative indicating a slight downward trend and afterward increasing trend in both cases of Pakistan and Punjab with overall average rate of change i.e. 1.89, and 1.76 percent, respectively.

Table 3. Characteristics of fitted models for production of wheat.

Model		PAKISTAN				
		SND	MM	GMZ	GUAS	LGS
Parameters	P1	16.002**	3.126**	77.733**	31.194**	34.924**
	(S.E)	(1.232)	(2.609)	(18.756)	(3.094)	(2.609)
	P2	12.725**	19900.1	1.306**	93.238**	14.834**
	(S.E)	(1.319)	(13016)	(0.046)	(6.730)	(0.820)
	P3	0.037**	47.321**	0.018**	39.996**	0.005**
	(S.E)	(0.004)	(9.990)	(0.002)	(2.678)	(0.003)
Goodness of Fit	P4	3.002**	2.346**	—	—	—
	(S.E)	(0.088)	(0.245)			
	R ²	0.99	0.99	0.98	0.98	0.99
	R ² _{adj.}	0.99	0.99	0.98	0.98	0.98
	S.W.T	0.931	0.922	0.694	0.591	0.509
	R.T	0.899	0.899	0.170	0.382	0.382
	D.W.S	2.081	2.080	1.647	1.716	1.768
RMSE	0.779	0.780	0.878	0.860	0.846	
MAE	0.623	0.629	0.709	0.693	0.688	
		PUNJAB				
Parameters		SND	MM	GMZ	GUAS	LGS
	P1	13.623**	2.621**	100.405**	29.137**	31.091**
	(S.E)	(1.491)	(0.218)	(41.806)	(4.734)	(3.237)
	P2	10.905**	35002.9	1.414**	106.409**	16.461**
	(S.E)	(1.542)	(25265)	(0.082)	(10.708)	(1.179)
	P3	0.034**	40.583**	0.014**	44.489**	0.050**
Goodness of Fit	(S.E)	(0.004)	(10.824)	(0.003)	(3.800)	(0.003)
	P4	2.966**	2.432**	—	—	—
	(S.E)	(0.091)	(0.276)			
	R ²	0.98	0.98	0.98	0.98	0.98
	R ² _{adj.}	0.98	0.98	0.98	0.98	0.98
	S.W.T	0.535	0.531	0.539	0.715	0.788
	R.T	0.804	0.321	0.620	0.321	1.000
D.W.S	2.191	2.150	1.730	1.809	1.897	
RMSE	0.652	0.658	0.735	0.718	0.701	
MAE	0.501	0.508	0.566	0.554	0.544	

**Significance at 1%. S.E = Standard Error, S.W.T = Shapiro Wilk Test, R.T = Run Test, D.W = Durbin Watson Statistic, RMSE = Root Mean Squared Error, MAE = Mean Absolute Error.

Table 4. Predictions for next 3 years of wheat production (million tons).

Years	2013-14	2014-15	2015-16
Punjab	19.28	19.60	19.90
Pakistan	24.67	25.00	25.33

Table 5. Fitted models with characteristics of productivity.

Model		PAKISTAN					
		SND	MM	GMZ	GUAS	LGS	YD
Parameters	P1	1839.43**	765.90**	37121.3	6637.74	6002.77**	0.0015**
	(S.E)	(81.556)	(34.379)	(48959)	(3435.6)	(1524.33)	(5.1e ⁻⁵)
	P2	1048.27**	22262.18	1.41**	162.86**	8.74**	-3.2e ⁻⁵ **
	(S.E)	(87.012)	(21257)	(0.311)	(47.355)	(2.132)	(2.2e ⁻⁶)
	P3	0.043**	4217.74**	0.007**	74.60**	0.031**	2.2e ⁻⁷ **
	(S.E)	((0.004)	(747.523)	(0.003)	(14.309)	(0.003)	(2e ⁻⁰⁸)
Goodness of Fit	P4	2.913**	2.461**				
	(S.E)	(0.112)	(0.329)				
	R ²	0.98	0.98	0.97	0.97	0.97	0.97
	R ² _{adj.}	0.98	0.98	0.97	0.97	0.97	0.97
	S.W.T	0.192	0.309	0.411	0.343	0.336	0.368
	R. T	0.104	0.104	0.709	0.899	0.899	0.261
		PUNJAB					
Parameters	P1	1924.76**	841.19**	10756.3	7400.35**	1.4e ⁻³ **	
	(S.E)	(91.579)	(39.260)	(11278)	(2886.525)	(5.2e ⁻⁰⁵)	
	P2	1060.66**	35125.5	209.79**	9.96**	-2.9e ⁻⁵ **	
	(S.E)	(94.267)	40859	(98.873)	(3.849)	(2.3e ⁻⁰⁶)	
	P3	0.04**	4336.03**	89.15**	0.028**	2e ⁻⁷ **	
	(S.E)	(0.005)	(862.736)	(25.934)	(0.004)	(2e ⁻⁰⁸)	
Goodness of Fit	P4	2.85**	2.57**				
	(S.E)	(0.122)	(0.393)				
	R ²	0.97	0.97	0.96	0.96	0.96	
	R ² _{adj.}	0.97	0.97	0.96	0.96	0.96	
	S.W.T	0.639	0.672	0.265	0.212	0.249	
	R. T	0.457	0.457	0.137	0.137	0.804	
	D.W.S	1.891	1.818	1.381	1.405	1.564	
	RMSE	117.252	119.576	137.517	136.307	129.082	
	MAE	96.890	98.910	106.902	105.872	102.756	

**Significance at 1%. S.E = Standard Error, S.W.T = Shapiro Wilk Test, R.T = Run Test, D.W = Durbin Watson Statistic, RMSE = Root Mean Squared Error, MAE = Mean Absolute Error.

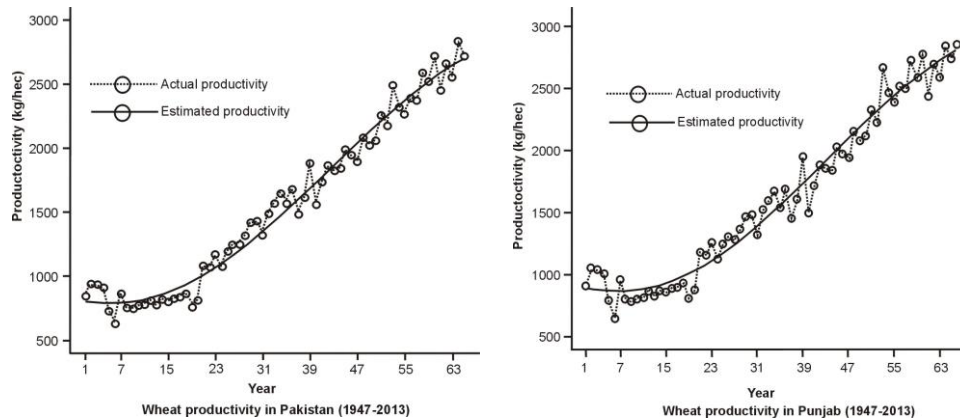


Fig. 3. Wheat productivity in Punjab/Pakistan.

CONCLUSION

Non-linear models are compared in the study for the production, productivity trends and best model is selected on the basis of satisfactory diagnostics. None of the non linear models is found suitable to fit trends in area sown under wheat in Pakistan due to violation of assumptions of error term. It is concluded that sinusoidal model with all significant parameters served our purpose to detect the trends in production and productivity of wheat grown in Punjab and Pakistan while nonparametric regression with kernel smoothing parameter (1.738) by method of cross validation is fitted to area sown under wheat in Pakistan. Wheat production in Pakistan had increased at an average rate of 2.95% which was due to combined effect of increase in area under cultivation and 1.89% average annual rate of change in productivity (kg per hectare). Area under cultivation plays a key role in production enhancement, effective strategies may be planned to add new area under cultivation of wheat. It can be more permissive in the sense to forecast or predict future production of wheat if more information regarding management, inputs or relief to farmer given by government be included.

REFERENCES

1. Bard, Y. 1974. Nonlinear Parameter Estimation. New York: Academic Press.
2. Bates, D. M., and D. G. Watts, 1988. Nonlinear Regression Analysis and Its Applications. USA: John Wiley & Sons.

3. Borthakur, S., and B. Bhattacharya, 1998. Trend analysis of area, production and productivity of potato in Assam: 1951-1993. *Economic Affairs* 43:221-226.
4. D'Agostino, R. B., and M. A. Stephens. 1986. *Goodness-of-Fit Techniques*. New York and Basel: Marcel Dekker, Inc.
5. Drapper, A. & H. Smith. 1998. *Applied Regression Analysis*. USA: John and Wiley & Sons.
6. Garson, G. D. 2012. *Testing Statistical Assumptions*. Asheboro, USA: Statistical Associates Publishing.
7. Hardle, W. 1990. *Applied Nonparametric Regression (1st Ed.)*. New York, USA: Cambridge University Press.
8. Huet. S., A. Bouvier, M.A. Poursat and E. Jolivet. 2004. *Statistical Tools for Nonlinear Regression: A Practical Guide With S-PLUS and R Examples (2nd Ed.)*. New York: Springer Series in Statistics.
9. Joshi, P., and R. Saxena. 2002. A profile of pulses production in India. Facts, trends and opportunities. *Indian J. Agric. Econ.* 57 (3):326-339.
10. Panse, V. 1964. Yield trends of rice and wheat in first two five-year plans in India. *J. Ind. Soc. Agric. Stat* , 16:1-50.
11. Prajneshu, and K. Chandran. 2005. Computation of compound growth rates in agriculture: Revisited*. *Agric. Econ. Res. Rev.* 18 (July-December), 317-324.
12. Rajarathinam, A., and B. Vinoth. 2013. Computations of jump points in tobacco (*Nicotiana tabacum*) crop production. *Int. J. Prob. and Stat.* 2 (2):13-20.
13. Rajarathinam, A., and B. Vinoth, 2012. Statistical modeling for wheat (*Triticum aestivum*) crop Production. *Int. J. Stat. Appli.*, 2(4):40-46.
14. Rajarathinam, A., Parmar, R., & Vaishnav, P. 2010. Estimating models for area, production and productivity trends of tobacco (*Nicotiana tabacum*) crop for Anand Region Of Gujarat State, India. *J. Appli. Sci.* 10 (20):2419-2425.
15. Ratkowsky, D. 1990. *Handbook of Non-linear Regression Models*. New York: Marcel Dekker.
16. Ritz, C. and J. C. Streibig. 2008. *Nonlinear Regression with R*. New York: Springer Science+Business Media.
17. Sami-Ullah, B. Din, and G. Haider. 2010. Modeling and Forecasting Wheat Yield of Pakistan. *Int. J. Agric. Appl. Sci.* 2(1):15-19.
18. Seber, G. A., and C. J. Wild. 2003. *Nonlinear Regression*. Hoboken, New Jersey: John Wiley & Sons.

124 *M. Ahmed et al.*

19. Shapiro, S. S. and Wilk, M. B. 1965. An analysis of variance test for normality (complete samples). *Biometrika* . 52 (3/4):591-611.
20. Singh, A. and R. Srivastava.2003. Growth and instability in sugarcane production Uttar Pradesh: A regional study. *Ind. J.Agric. Econo.* 58 (2):279-282.

Received: July 07, 2014 Accepted: April 27, 2016

CONTRIBUTION OF AUTHORS:

- | | | |
|----------------------------|----------|---|
| Mukhtar Ahmed | : | Collected and analyzed research data, Interpretation of results and discussion, prepared write up |
| Muhammad Salim Khan | : | Prepared introduction and review of literature, Tabulation of results, analysed data, discussed results. |
| Nasir Iqbal | : | Collected data from described sources. |

Table 2. Characteristics of fitted models for Area sown under wheat in Pakistan.

Model	Parameters				Goodness of Fit						
	P1	P2	P3	P4	R ²	R ² _{adj.}	S.W.T	R. T	D.W.S	RMSE	MAE
SND (S.E)	6156.756*	2517.256*	0.041*	3.618*	0.980	0.979	0.715	0.002	0.974	221.920	175.392
	146.8785	185.76	0.0041	0.1685							
EXP (S.E)	4331.699*	0.0121*	—	—	0.935	0.934	0.018	0.000	0.304	404.969	355.399
	82.4312	0.0004									
GMTZ (S.E)	11288.424*	0.135*	0.025*	—	0.975	0.974	0.172	0.000	0.775	250.134	214.003
	600.3759	0.0351	0.0026								
GUAS (S.E)	8899.153*	72.962*	54.704*	—	0.978	0.978	0.599	0.000	0.877	234.357	196.205
	159.4196	3.3907	2.4996								
LGS (S.E)	10160.287*	1.793*	0.039*	—	0.977	0.976	0.371	0.000	0.837	240.224	202.985
	322.9	0.0678	0.0027								
MM (S.E)	4118.739*	2048.1083	10130.735*	2.131*	0.981	0.980	0.911	0.009	1.016	217.878	177.910
	95.5187	1368.0069	466.9171	0.2235							

*Significance at 5%, S.E= Standard Error, S.W.T= Shapiro Wilk Test, R.T= Run Test, D.W= Durbin Watson Statistic, RMSE= Root Mean Squared Error, MAE= Mean Absolute Error.