African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 3, Issue 5, 2020 (pp. 63-74)
www.abjournals.or

MATHEMATICAL MODELING OF ROAD TRAFFIC FATALITIES

Matthew Femi Olayiwola

Mathematics Department, African Leadership Academy South Africa.

Corresponding Author: femmat08@gmail.com

ABSTRACT: *The development of a mathematical model for predicting the future values of road fatality in South Africa is of great concern as road accidents become a public health concern and a threat to the economy. This study aims to develop an ARIMA model for the prediction of quarterly road fatality in all the provinces of South Africa. Nineteen years of quarterly data from 2001 to 2019 were used to build the model. The best two models were selected using the Akaike information criterion (AIC); volatility and adjusted R^2 values. Thus, ARIMA $(4, 1, 1)$ and ARIMA $(4, 1, 4)$ were found to be the best models. The ARIMA (4, 1, 4) model forecast produced an excellent match with the actual series. Hence, predicting the future values of road fatality in South Africa using these models will help the policymakers and all stakeholders make well-informed decisions. In the next study, the economic impact of road crashes and fatalities on the South African economy will be considered.*

KEYWORDS: Mathematical Modeling, Road Traffic Fatality, Box-Jerkins, ARIMA, Time Series

INTRODUCTION

It is no longer news that road traffic injuries are a serious threat to every nation's public health and economy. According to the World Medical Association (2006), serious injuries and mortality road collisions are a public health problem with consequences similar to those of major diseases such as cancer and cardiovascular disease. World Health Organisation (2004) reported that; over 1.2 million people lost their lives on the road while close to 35 million people sustained injuries ranging from minor to severe. The report further mentioned that 2.1% of all deaths globally were due to road traffic accidents, which makes it become the 11th leading cause of death across the globe (World Health Organization, 2004). Sadly, 90% of the global road traffic deaths occurred in low income and middle-income countries, mostly in the continent of Africa. The global mortality was predicted to be 67% by 2020 if proper attention is not paid to this threat (World Health Organisation, 2004). Thus, this brings every nation of the world under pressure since 2004.

Moira Winslow, Chairman of Drive Alive, in South Africa correctly argues that; "The human suffering for victims and their families of road traffic-related injuries is incalculable. Their endless repercussions; families break up; high counseling costs for a family if a breadwinner is lost; and thousands of Rands to care for injured and paralyzed people." (World Health Organization, 2004).

In total, 3 280 931 deaths were recorded in South Africa between 2001 and 2006 of which 9.5% were due to non-natural causes (Statistics South Africa, 2006). Road traffic accident deaths comprised 9.3% of non-natural deaths. Data from the National Injury Mortality Surveillance System (NIMSS) show that in 2005 transport-related injuries accounted for 74.3% of all accidental (or unintentional) deaths (Medical Research Council South Africa, 2007). Analysis of the injury burden in South Africa showed that the age-standardized road traffic injury mortality rates for South Africa were about double the global rate for both males and females (Norman, Leslie George, and World Health Organization, 1962).

A road traffic crash results from a combination of factors related to the components of the system comprising roads, the environment, vehicles and road users, and the way they interact. Normaan and WHO (1962) attributed the high burden of traffic injury mortality in South Africa to unsafe road environments, poor enforcement of existing traffic laws, road rage and aggressive driving as well as alcohol misuse.

The leading cause of death among adolescent and younger people in their prime age has been traced down to road traffic crashes (Atunbi, 2009). The experience of the surge of road traffic crashes varies from nation to nation. Emenike and Ogbole (2008) submitted that nations with more industrial activities experience a decrease in road traffic crashes by 20% compared to their counterparts. In research carried out by the duo, the economic giant of Africa is experiencing road traffic crashes at an alarming rate.

Nations of the world pay much attention to improving their economies. To improve the country's economy, South Africa is no exemption in paying much attention to economic researches. South African economy has been described to be energy-dependent as its petrol consumption (majorly used by the automobile industry) shows a yearly upward trend (Olayiwola and Seeletse, 2020). In another research on the cointegration relationship between petrol price and consumption, the duo opined that a hike in petrol price shifts the commuters' preference from taxis to trains within the same form of transportation, road (Olayiwola and Seeletse, 2020). This implies that the road as a means of transportation plays a vital role in the growth of the South African economy. Olayiwola and Seeletse (2020) opined that a model developed for forecasting should be able to provide short term forecasts that give indications of the possible performance on medium (and maybe long term) forecasting.

Poisson and negative binomial regression analyses have been suggested by researchers to be suitable for modeling road traffic crashes due to the nature of its data. These two methods are often used in modeling the occurrence of events that are in the discrete form (Cameroon and Trivedi, 1998). Oppong and Assuah (2015) as an alternative to the Cox model for survival analysis further buttressed this. ARIMA (3,1,1) and MA (0,1,2) are concluded to be the best models for road accidents (Balogun et al, 2015). Contrary to this, in 2016 seasonal-ARIMA was argued to be the best model for the traffic accidents data by Iwok (2016). The Seasonal-ARIMA model was found to be the best fit for monthly road traffic accidents in Port Harcourt, Nigeria (Iwok, 2016). Lack of concentration due to the use of mobile phones while driving among many other factors was claimed to be responsible for the alarming rate of road traffic crashes in Nigeria (Emenike and Kanu, 2017). In a fatal road accident in Nigeria, an upward trend was noticed as the simple exponential smoothing model fitted the data (Afere et al, 2015). The incidence of the accident in Ibadan shows an upward trend as revealed in research conducted by Abdulkabir et al (2015).

A lot of researches have been done on which model best fits the road traffic crashes around the world. However, a suitable model for the fatalities resulting from road traffic accidents is still a cause for concern. Hence, in this study, mathematical models that could be used to predict road traffic accidents in South Africa as a tool for decision-making by all the stakeholders in both the short and medium terms are developed.

Time Series Procedures

Augmented Dickey-Fuller (ADF) test is considered in testing for the null hypothesis that the series is nonstationary against the alternative hypothesis of staitionarity. This test is based on simple regression of actual series on its one-period lagged value.

$$Y_t = \theta Y_{t-1} + e_t \quad (1)$$

Where: Y_t = the actual series

Y_{t-1}= one period lagged of the actual series

e_t= a serially uncorrected white noise error term with a mean of zero and a constant variance.

The parametrised form of (1) yields (2):

$$\Delta Y_t = \delta Y_{t-1} + e_t \quad (2)$$

where $\delta = (\theta - 1)$, and Δ is the first difference operator.

The null hypothesis of $\delta = 0$ from (2) has an estimated t-value

The t-value of the estimated coefficient of δ in (2) does not follow the *t*-distribution even if the sample size is large, that is, it does not have an asymptotic normal distribution (Erdogdu, 2007). The Dickey-Fuller (DF) test assumed that the residuals are uncorrelated. But in practice, the error terms of the DF test are found to be serially correlated. So, this assumption is proven to be untrue and was corrected by DF in another test known as the Augmented Dickey-Fuller test. In the ADF test, the lags of the first difference are included in the regression equation to make the error terms white noise. This is represented below:

$$\Delta Y_t = \delta Y_{t-1} + \alpha_i \sum_{i=1}^{m} \Delta Y_{t-i} + e_t \quad (3)$$

Including intercept and time, t, in equation 3 now givess:

$$\Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + \alpha_i \sum_{i=1}^{m} \Delta Y_{t-i} + e_t \quad (4)$$

The ADF unit root testing procedure uses this model:

$$\Delta y_t = \alpha + \beta t + \gamma Y y_{t-1} + \sum_{j=1}^{\rho} \delta_j \Delta y_{t-j} + e_{it} \quad (5)$$

where α= the intercept (a constant),

β= the coefficient of time, t

γ = the coefficient of y_{t-1},

P= the lag order of the autoregressive process

Δy_t = the first difference of y_t

y_{t-1} = the one time period lag values of y_t

Δy_{t-j} = the changes in lagged values, and

e_{it} = the white noise.

γ is the parameter of interest in (5), γ = 0 implies that the series is non-stationary.

The generalized AR(p) model is represented in (6):

$$\text{AR}(p)\text{: } fatality_t = \alpha + \sum_{i=1}^{p} \vartheta_i fatality_{t-i} + u_t \quad (6)$$

$p = the\ number\ of\ lagged\ values\ of\ the\ regressand\ included\ in\ the\ model.$

(6) is the regression equation of the actual series on its past values, that is, the series is explained by its past values.

Also, the MA(q) is represented in (7):

$$\text{MA}(q)\text{: } fatality_t = \pi + \sum_{j=1}^{q} d_j u_{t-i} + d_0 u_t \quad (7)$$

$q = the\ number\ of\ lagged\ values\ of\ the\ error\ term\ in\ the\ model.$

(7) is the regression equation of the actual series on its error past values, that is, the series is explained by its error past values.

Combining (6) and (7) produces the ARMA (p, q) model as represented in (8) below.

$$\text{ARMA}(p,q)\text{: } fatality_t = \varphi + \sum_{i=1}^{p} \vartheta_i fatality_{t-i} + \sum_{j=1}^{q} d_j u_{t-i} + d_0 u_t + \mu_t \quad (8)$$

(8) contains p lags of the dependent variable, fatality, and q lags of the error term.

Hence when the series has a unit root, then (8) becomes ARIMA (p, d, q) with d been the order of integration.

Materials and Methods

Quarterly road fatality data were collected for the period of 19 years, specifically from years 2001-2019 from the Road Transport Management Corporation South Africa. A basic time series plot is used to investigate the presence of any time series components such as trend, season, etc. The correlogram is used to identify the appropriate model and order of the lag-variables to be included in the model based on the behavior of the ACF and PACF. The series is then pretested to determine the order of integration using the Augmented Dickey-Fuller (ADF) test. The significant lags from the ACF and PACF are used in estimating tentative models. The best models are chosen from the estimated models using each model's volatility, adjusted R^2, the significance of coefficients and information criterion values. The chosen model is diagnosed as required by the theory. The forecast is done using the best model and its adequacy In predicting the fatality is considered. Eviews 11 is used for all the analyses, recommendations and conclusions are based on the discussions of results.

RESULTS AND DISCUSSIONS

The time series plot of original road fatality data shown in Fig. 1 shows that there is trend variation in the series as the series mean changes systematically slowly with time. Though this is an indication of nonstationarity in the series but we cannot conclude based on the time series plot. Hence, the Augmented Dickey-Fuller test is required to check for the presence of the unit root in the series.

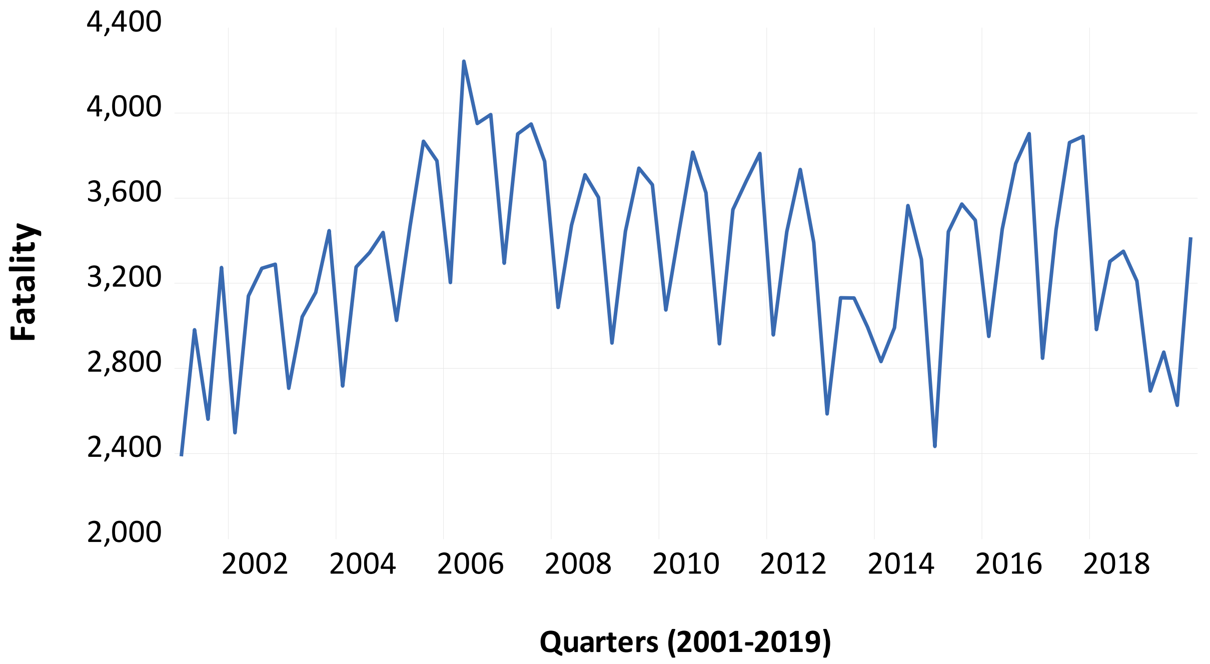


Fig 1: Students Achievement in Mathematics and Physical Sciences

The results of the ADF unit root tests for both level and differenced series of fatality are shown in table 1. The results show that the level series has a unit root while the first differenced series of fatality has no unit root. Hence, fatality is integrated of the first order, $I(1)$. Do for the crash as well.

Table 1: Augmented Dickey-Fuller results of road traffic fatality

$\boldsymbol{H_0}$: Series has a unit root

		Fatality
Series	P-value	Decision
Level series	0.5893	H0 is not rejected
First differenced series	0.0299	H0 is rejected

Source: Appendices I and II

Table 1: Augmented Dickey-Fuller results of road traffic fatality and crash

H_0: Series has a unit root

	Fatality		Crash	
Series	P-value	Decision	P-value	Decision
Level	0.5893	H0 is not rejected	0.5893	H0 is not rejected
First differenced	0.0299	H0 is rejected	0.0299	H0 is rejected

Source: Appendices I and II

The correlogram plot of the autocorrelation function (ACF) and partial autocorrelation function (PACF) for the first differenced fatality series is plotted. This plot shows that the fatality series can be modeled by an autoregressive integrated moving average (ARIMA). Few lags are found to be significant for both the ACF and PACF, thus various model are estimated. Table 2 shows the results of tentative estimated ARIMA models.

Table 2: Results of tentative ARIMA (p, d, q) models

	ARIMA (2, 1, 1)	**ARIMA (4, 1, 1)**	**ARIMA (4, 1, 3)**	**ARIMA (4, 1, 4)**	**ARIMA (8, 1, 1)**
Intercept	0.7653	0.9676	0.9065	0.9670	0.9138
Autoregresive	0.0355	0.0000	0.0000	0.0000	0.0000
Moving Average	0.0000	0.0000	0.0000	0.0000	0.0000
Sigma (volatility)	141984.20	67740.94	82099.80	66810.50	67927.86
Adjusted R (in %)	41.36	70.84	66.09	72.40	71.94
AIC	14.82	14.12	14.37	14.18	14.18
SC	14.94	14.24	14.49	14.30	14.30

Source: Authors' computation from Eviews 11

The p-values for the intercept, autoregressive, and moving average are reported in table 2. None of the models has a significant intercept/constant while each of the tentative models has significant AR and MA. ARIMA (2, 1, 1) has the volatility and information criterion with the lowest adjusted R^2 value. AIC and SC. This is followed by ARIMA (4, 1, 3). ARIMA (4, 1, 4) and ARIMA (8, 1, 1) have the same AIC and SC values with the later having the highest volatility and lowest adjusted R^2 compared to the former. Thus, ARIMA (4, 1, 4) is preferred to ARIMA (8, 1, 1). ARIMA (4, 1, 1) has the lowest AIC and SC, but slight high volatility compared to ARIMA (4, 1, 4); and a slightly lower adjusted R^2 value compared to ARIMA (4, 1, 4). Therefore, ARIMA (4, 1, 1) and ARIMA (4, 1, 4) are considered to be the best models in table 2.

The two selected models are diagnosed through the correlogram of their residuals. Each correlogram indicates that all information has been captured, hence the forecast will be based on the two selected models, ARIMA (4, 1, 1) and ARIMA (4, 1, 4).

Table 3: Forecast Model Adequacy Results

	ARIMA (4, 1, 1)	ARIMA (4, 1, 4)
Root Mean Squared Error	331.0893	341.1166
Mean Absolute Error	290.3863	291.9272
Mean Absolute Percentage Error	9.9172	10.4078
Symmetric MAPE	9.7495	9.9506

Source: Appendices III and IV

Table 3 shows the results of the forecast models. ARIMA (4, 1, 1) and ARIMA (4, 1, 4) are used to forecast for 2019 quarter 1 to 2019 quarter 4. ARIMA (4, 1, 1) forecast shows a slight large deviation from the actual fatality series from 2019q1 to 2019q2 compared to ARIMA (4, 1, 4) forecast (see appendices I and II). The forecast by ARIMA (4, 1, 4) is exactly equal to the actual fatality at two points while ARIMA (4, 1, 4) forecast intersects the actual fatality at only one point (see appendices I and II). From table 3, ARIMA (4, 1, 1) is seen to have an accuracy value of 90.08 while ARIMA (4, 1, 4) has an accuracy of 89.59. Thus, the two models have almost the same accuracy value.

CONCLUSIONS

The 19 years fatality series in South Africa has a trend component, thus the series is differenced to remove the trend and found to be integrated of order 1, $I(1)$. Many lags of the first differenced are found to be significant from the correlogram of both the ACF and PACF, however, the parsimonious models are estimated and checked for the best models based on the number of the significant coefficients; the lowest volatility; highest adjusted R^2 and the lowest information criterion. ARIMA (4, 1, 1) and ARIMA (4, 1, 4) are found to be the best predictive models for road fatality in South Africa. The residuals correlogram plots of the two models are found to be flat, meaning that all information is captured. ARIMA (4, 1, 4) which has a forecast accuracy of 89.95% predicts the series better as its forecast values intersect the actual values at two points and with a little deviation from the actual. The policymakers and all road traffic stakeholders to predict the future values of life that will be lost on South African roads in the coming years can therefore use these models. This is necessary since transportation by road plays a vital role in the automobile industry that contributes significantly to the economy of South Africa. Also, planning with these models will help reduce the public health challenge and economic threat the road fatality and accident placed on the South African economy. The impact of the cost of accidents on the South African economy will be considered in the next study.

Acknowledgments

The author acknowledge the support of Statistics South Africa and Road Traffic Management Corporation of South Africa for providing the data used in this study.

Author's contributions

MF Olayiwola proposed the study topic, reviewed the relevant pieces of literature, collected and analyzed the data and discussed the results, and conducted the study. Also, he prepared this manuscript and finally identified the relevant journal.

REFERECENS

Abdulkabir, M., Raji, S. T., & Udokang, A. E. (2015). Trend analysis on road traffic accident in Nigeria. Science, 3(5), 52-57.

Afere, B. A., Oyewole, S. A., & Haruna, I. (2015). On the Time Series Forecasting of Road Traffic Accidents in Ondo State of Nigeria. Journal of Statistical Science and Application, 3, 153-152.

Alive Arrive. (2007). Report Back to the Nation on State of the Road. South Africa: Arrive Alive.

Atunbi, A. O. (2009). Urban Transportation. An Appraisal of Features and Problems in Nigerin Sociaty. International Journal of Geography and Regional Planning, 1.

Balogun, O., Oguntunde, P., Akinrefon, A., & Modibbo, U. (2015). The Comparison of the Performance of ARIMA and MA model selection on Road Accident Data in Nigeria. European Journal of Academi, 2(3), 13-31.

Cameron, A. C., & Trivedi, P. K. (2013). Regression analysis of count data. Cambridge university press.

Emenike, G. C., & Ogbole, A. (2008). Accidents and the Road Transport Industry in Nigeria. Journal of the International Centre for Constructive Research (ICCR).

Emenike, G., & Kanu, C. (2017). Driver's distraction and road traffic crashes in Port Harcourt Metropolis, Rivers State, Nigeria. International Journal of New Technology and Research, 3(4), 01-07.

Erdogdu, E. (2007). Electricity demand analysis using cointegration and ARIMA modelling. Energy policy, 35(2), 1129-1146.

Iwok, I. A. (2016). Seasonal Modelling of Road Traffic Accident. Mathematical Theory and Modeling, 6(4), 83-94.

Matzopoulos, R., & Mohamed, S. (2008). National Injury Mortality Surveillance System-a call for city-level injury prevention. SOUTH AFRICAN MEDICAL JOURNAL-CAPE TOWN-MEDICAL ASSOCIATION OF SOUTH AFRICA, 705.

Medical Reserach Council South Africa. (2007). A profile of fatal injuries in South Africa. Pretoria.

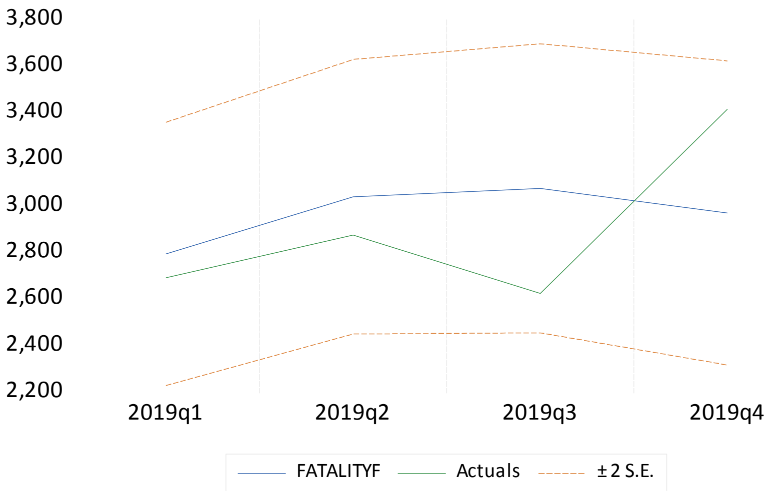
Norman, Leslie George, and World Health Organization. (1962). Road traffic accidents: epidemiology, control, and prevention.

Olayiwola, M. F., & Seeletse, S. M. (2020). South African Petrol Price and Consumption Cointegration Analysis. Journal of Engineering and Applied Sciences, 15(6), 1579-1588. doi:http://dx.doi.org/10.36478/jeasci.2020.1579.1588

Olayiwola, M. F., & Seeletse, S. M. (2020). Statistical Forecasting of Petrol Price in South Africa. Journal of Engineering and Applied Sciences, 15(2), 602-606.
Oppong, R. A. (2015). Comparative Assessment of Poisson and Negative Binomial Regressions As Best Models For Road Count Data. International Journal of Scientific Research and Engineering Studies, 11(2), 28-32.
Statistics South Africa. (2006). Mortality and causes of death in South Africa, 2003 and 2004: findings from death notification. Pretoria: Statistics South Africa.
World Health Organization. (2004). Road safety is no accident: A brochure for World Health. Geneva: Department of Injuries and Violence Prevention.
World Medical Association. (2006). Statement on Traffic Injury". Adopted by the 42nd World Medical Assembly. CA-USA: Rancho Mirage.

APPENDIX

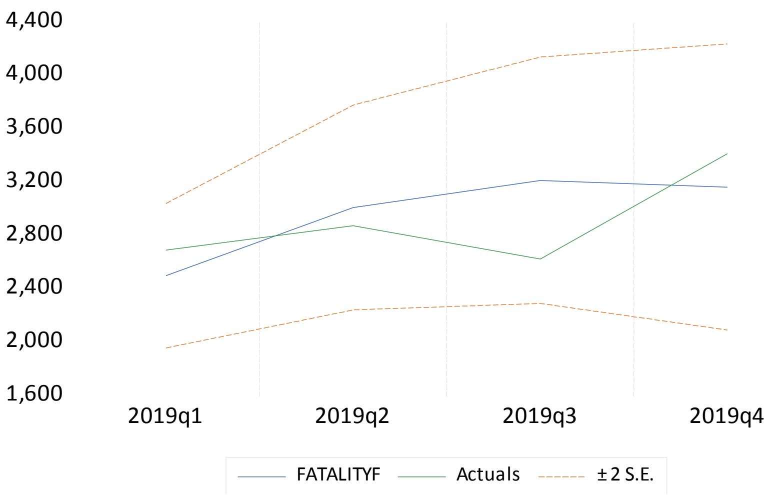
I: Graphs of ARIMA (4, 1, 1) forecasted and actual fatality series



Forecast: FATALITYF	
Actual: FATALITY	
Forecast sample: 2019Q1 2019Q4	
Included observations: 4	
Root Mean Squared Error	331.0893
Mean Absolute Error	290.3863
Mean Abs. Percent Error	9.917198
Theil Inequality Coef.	0.056197
Bias Proportion	0.041632
Variance Proportion	0.372060
Covariance Proportion	0.586309
Theil U2 Coefficient	0.745091
Symmetric MAPE	9.749475

Source: Authors' computation from Eviews 11

II: Graphs of ARIMA (4, 1, 4) forecasted and actual fatality series



Forecast: FATALITYF	
Actual: FATALITY	
Forecast sample: 2019Q1 2019Q4	
Included observations: 4	
Root Mean Squared Error	341.1166
Mean Absolute Error	291.9272
Mean Abs. Percent Error	10.40782
Theil Inequality Coef.	0.057758
Bias Proportion	0.042667
Variance Proportion	0.006279
Covariance Proportion	0.951054
Theil U2 Coefficient	0.722605
Symmetric MAPE	9.950638

Source: Authors' computation from Eviews 11

III: Graphs of ARIMA (4, 1, 4) forecasted and actual fatality series

Dependent Variable: D(FATALITY)
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 11/26/20 Time: 09:49
Sample: 2001Q2 2019Q4
Included observations: 75
Convergence achieved after 16 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.712093	42.00656	0.040758	0.9676
AR(4)	0.764199	0.079446	9.619075	0.0000
MA(1)	-0.652789	0.080483	-8.110911	0.0000
SIGMASQ	67740.94	12234.11	5.537055	0.0000
R-squared	0.720205	Mean dependent var		13.73333
Adjusted R-squared	0.708382	S.D. dependent var		495.3592
S.E. of regression	267.5020	Akaike info criterion		14.11819
Sum squared resid	5080570.	Schwarz criterion		14.24179
Log likelihood	-525.4320	Hannan-Quinn criter.		14.16754
F-statistic	60.91900	Durbin-Watson stat		1.994068
Prob(F-statistic)	0.000000			
Inverted AR Roots	.93	.00-.93i	-.00+.93i	-.93
Inverted MA Roots	.65			

Source: Authors' computation from Eviews 11

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 3, Issue 5, 2020 (pp. 63-74)
www.abjournals.or

IV: Graphs of ARIMA (4, 1, 4) forecasted and actual fatality series

Dependent Variable: D(FATALITY)
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 11/26/20 Time: 09:50
Sample: 2001Q2 2019Q4
Included observations: 75
Convergence achieved after 25 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.310184	175.9295	0.041552	0.9670
AR(4)	0.996215	0.007934	125.5704	0.0000
MA(4)	-0.854920	0.133914	-6.384095	0.0000
SIGMASQ	66810.48	12023.27	5.556766	0.0000

R-squared	0.724048	Mean dependent var	13.73333
Adjusted R-squared	0.712388	S.D. dependent var	495.3592
S.E. of regression	265.6585	Akaike info criterion	14.18102
Sum squared resid	5010786.	Schwarz criterion	14.30462
Log likelihood	-527.7881	Hannan-Quinn criter.	14.23037
F-statistic	62.09701	Durbin-Watson stat	2.630678
Prob(F-statistic)	0.000000		

Inverted AR Roots	1.00	.00+1.00i	-.00-1.00i	-1.00
Inverted MA Roots	.96	.00-.96i	.00+.96i	-.96

Source: Authors' computation from Eviews 11