Tanzania Journal of Science 47(2): 495-506, 2021
ISSN 0856-1761, e-ISSN 2507-7961

Count Time Series Models for Road Traffic Accidents in Tanzania Mainland

Amina S Msengwa* and Florence D Ngari
Department of Statistics, University of Dar es Salaam, P. O. Box 35047, Dar es Salaam, Tanzania.
**Corresponding author e-mails: amina.msengwa@udsm.ac.tz; msengwaa@yahoo.com*
**Corresponding author ORCID: orcid.org/0000-0001-9418-9008*
Co-author e-mail: ngariflorence@yahoo.com
Received 11 Dec 2020, Revised 11 Mar 2021, Accepted 21 Apr 2021, Published May 2021
DOI: https://dx.doi.org/10.4314/tjs.v47i2.8

Abstract

A pairwise analysis was conducted to assess the trends and factors associated with road traffic accidents in Tanzania. The Poisson and Negative Binomial Autoregressive Models were used to extend log linear functions by accounting time-varying components. A total of 85,514 road traffic accidents in Tanzania mainland that occurred from 2012 to 2017 were extracted from Tanzania Police Office records. Eleven factors were grouped into a human, vehicle, physical/environmental and pedestrian-related factors. The Likelihood ratio test, Akaike Information Criterion, Bayesian Information Criterion and residual ACF plots were used to evaluate the performance of the models in Dar es Salaam and other combined regions. The trend analysis indicated a declining pattern in all factors and human-related factors appeared higher than the other three factors. The highest number of road traffic accidents was observed in Dar es Salaam Region compared to other combined regions. The models, including its past values and time-varying factors, were in favour-of other models. In both, Dar es Salaam and other combined regions, non-linear pattern and Negative Binomial Autoregressive Models fitted the data well. The implementation of collective actions in recent years seems positive on road traffic accidents. Nevertheless, more emphasis is needed to monitor trends on the number of accidents and related fatalities.

Keywords: Road Traffic Accidents, Poisson, Negative binomial, Autoregressive Models, Tanzania.

Introduction

Road transport is the most important and common means of transportation. It is essential and convenient for transporting both humans and goods (WHO 2015). However, road traffic accidents are a global problem affecting all sectors of society and threatening everyday life. Estimates show that approximately 1.35 million people die each year worldwide due to road traffic accidents, and roughly half of those dying are pedestrians, cyclists, and motorcyclists (WHO 2018). Furthermore, about 20 to 50 million people are injured, the majority are people aged 5 to 29 years from low-and middle-income countries.

In developed countries, the magnitude of the problem (road traffic accidents) has been recognized since 1970s. Consequently, these countries have been working towards promoting road safety conditions and raising awareness among road users (Petridou and Moustaki 2000). The developed countries had the ability to regulate road traffic accidents through a combination of interventions, including improved roads, implementation of strict traffic laws and other road safety measures targeting important factors that contribute to the problem. However, some countries have effectively minimized the number of road fatalities, while for others, the number is increasing (WHO 2018).

Studies conducted in developing countries indicated that more than 90% of deaths and disabilities were related to road traffic accidents (Walugembe et al. 2020). The risks are still high, though Delavary et al. (2019) projected a global decline of 27% by 2020 on road traffic accidents. In Tanzania, road transport is the key mode of transport, playing an important role in carrying over 75% of freight traffic and over 90% of passengers (JICA 2014, Walugembe et al. 2020), and is characterized by a low level of safety (Tanzania Police Force 2016, Fell et al. 2017). WHO (2018) estimated 29.2 per 100,000 population and road traffic fatalities at around 3256 (79% males, 21% females). The estimates demand data-driven models to predict the number of road traffic accidents in Tanzania to monitor Sustainable Development Goals (SDG) target 3.6 that aimed at reducing the number of deaths and injuries from road traffic accidents by a half.

The observational-driven approach is specified by Cox et al. (1981) to extend Generalized Linear Models for predicting counts of time series outcomes. Contrary to some other traditional time series models, such as Auto-regression (AR) and Moving-Average (MA), the approach provides the best estimates for over-dispersed data, non-linear responses and autocorrelation functions (Fokianos 2012, Ravishanker 2014, Christou and Fokianos 2015, Chen et al. 2016). This study considers a class of Generalized Autoregressive models to assess time-varying covariates and autocorrelation functions to shed-light on the possible future direction of the number of road traffic accidents in Tanzania.

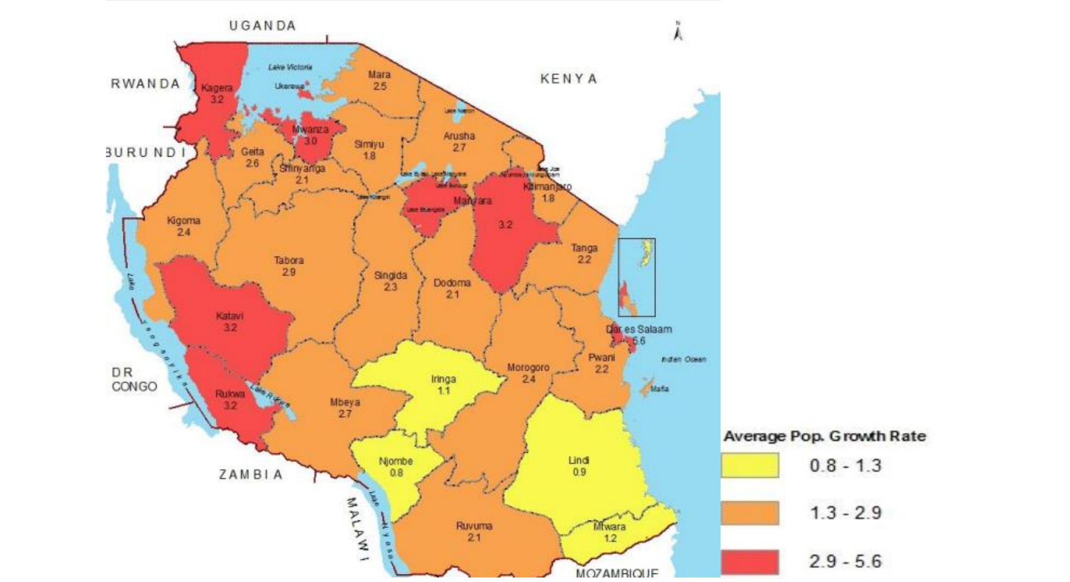
Material and Methods

Data type and source

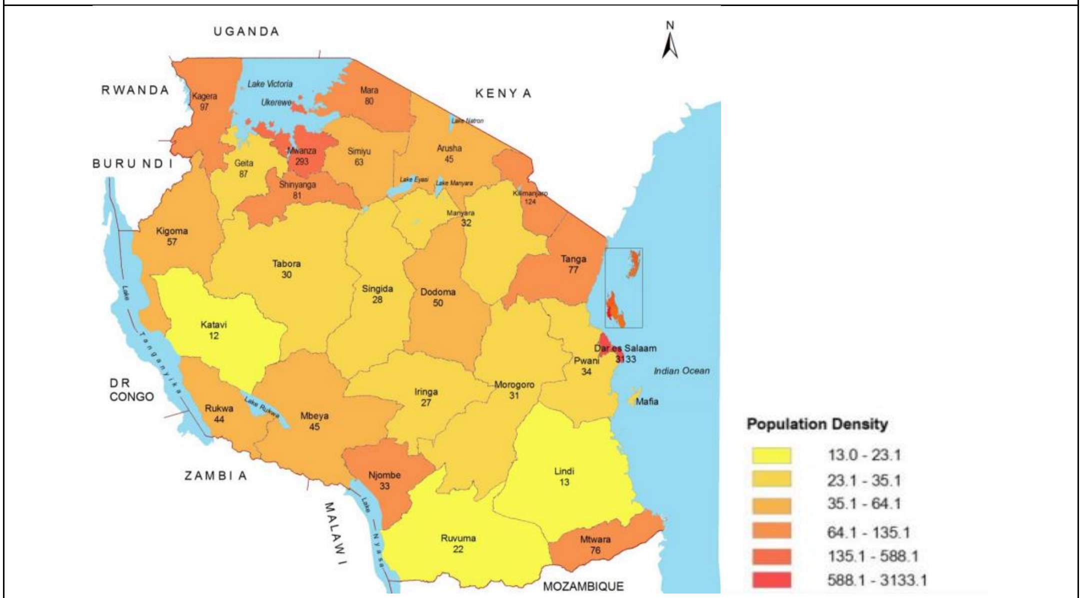
In this study, retrospective monthly time series data was used to analyze road traffic accidents in Tanzania mainland. Road traffic accident data was requested and extracted from the traffic police office for all regions from January 2012 to December 2017 (72 months). The regional classification was based on the 2012 census report, whereby the regions were purposely divided into two sub-regional groups, Dar es Salaam Region and other combined regions. Figure 1 illustrates the population parameters of Tanzania regions. With the exception of Dar es Salaam Region, the population is sparsely distributed across regions. Dar es Salaam is the largest city in terms of people and a regional economic centre. Its average annual population growth rate is high, estimated at 5.6 between 2002-2012 censuses and the population density is approximated to 3,313 persons per square kilometer. These differences are very high, guarantee pairwise analysis across sub-regional accidents.

Study variables

Count Time series data is used for an ordered sequence of observations in time, referred as (Y_{it}, X_{ijt}). In this sequence, y_t denotes dependent variables, and x_t denotes independent variables at a specific time t (where t represents month, $t = 1, 2, ...72$ months), predictor j (where $j = 1,2,3,4$ represents factor associated with RTA) and region i (where i represents sub-regional group $i = 1,2$ ($1 =$ Dar es Salaam Region and $2 =$ Other combined regions whereby the number of road traffic accident were added together). The total number of road traffic accidents was considered as a dependent variable and the number of accidents was comprised of eleven independent variables grouped into four factors (human, vehicle, physical/environmental and pedestrian), namely reckless/dangerous driving, careless driving, over speeding, overtaking, vehicle defects, poor vehicle lighting, fire, road barriers, poor road infrastructure, railway crossing and pedestrians. The independent variables are considered as human X_{i1t} (reckless/dangerous driving, careless driving, over speeding and careless overtaking), vehicle X_{i2t} (vehicle defects and poor vehicle lighting), physical/environmental X_{i3t} (fire, road barriers, poor road infrastructure and railway crossing) and pedestrian X_{i4t} were considered as independent variables.



(a) Population growth rate 2002-2012



(b) Population density 2012

Figure 1: Map showing population growth rate 2002-2012 and population density of 2012 (Source: NBS 2012).

Modeling

In modeling time series of counts, one often assumes that, given the history Y_t , variable Y_{t+1} follows a Poisson distribution. When the conditional variance of the counts grow un-proportionally with their means, this assumption may be violated and the Negative Binomial distribution becomes appropriate. To better understand the pattern of road traffic accidents in Tanzania, the Poisson Autoregressive and Negative Binomial Autoregressive models were performed and compared.

Before fitting count time series models, the recorded dataset was assessed as to whether linear and non-linear trends fit well to the data-points. The observations were examined through plotting the data points, fitted linear trend and Loess curve. Estimating the regression effects and the serial-correlations of the number of road traffic accidents for each time point, Y_t are assumed independent Poisson distributed as $Y_t \sim \text{Poisson}(\mu_t)$ and Negative Binomial distributed as $Y_t \sim \text{NB}(\mu_t)$. Based on the Poisson primary model: $f(\text{y}_t;\mu_t) = \dfrac{\text{e}^{\mu_t}\mu_t^{-y_t}}{y_t!}$ and for a case of over-dispersion, the NB model: $f(y_t;\mu_t,\theta) = \dfrac{\Gamma(y_t+\theta)}{\Gamma(\theta).y_t!}.\dfrac{\mu_t^{y_t}\theta^{\theta}}{(\mu_t+\theta)^{y_t+\theta}}$

with mean μ_t and shape parameter θ.

The expected count is represented by $\mu_t = \exp(X_t\beta + V_t)$ and the stochastic process V_t. Under observation-driven approach (as presented by Zeileis et al. (2008), and Ahmad and Francq (2016)) Y_t / μ_t is assumed Poisson (μ_t), $\log(\mu_t) = X_t\beta + V_t$ and V_t defines a function of past observations Y_s, s < t such that $V_t = \gamma_1 Y_{t-1} + ... + \gamma_p Y_{t-p}$. Note that if $V_t = 0$, the model becomes $\log(\mu_t) = X_t\beta_t$ indicating a standard Poisson or Negative Binomial regression model, and if $\beta = 0$, the model depends on past observations commonly known as Autoregressive Conditional Poisson (ACP) model. In this case, four (4) models were formulated to explain these scenarios as shown in Table 1.

Table 1: Specification of count time-series models

<table>
<tr><td>Poisson Regression Model
$Y_t \sim \mathit{Poisson}(\mu_t)$</td><td>Negative Binomial Regression Model
$Y_t \sim \mathit{NB}(\mu_t)$</td></tr>
<tr><td colspan="2">$\mu_t = \exp(X_t\beta + V_t)$, where $X_t = [x_{i1t}, x_{i2t}, x_{i3t}, x_{i4t}]$ and $V_t = \gamma_1 Y_{t-1} + ... + \gamma_p Y_{t-p}$
time $t = 1,2,......72$ months. β and γ are coefficients associated with regressed covariate and time covariate factors</td></tr>
<tr><td>Model 1: $\mu_t = \mathit{exp}(V_t)$</td><td>Model 3: $\mu_t = \exp(V_t)$</td></tr>
<tr><td>Model 2: $\mu_t = \exp(X_t\beta + V_t)$</td><td>Model 4: $\mu_t = \exp(X_t\beta + V_t)$</td></tr>
</table>

In order to diagnose the best-fitted model, the identified models were compared by Likelihood ratio (LR), Akaike information Criterion (AIC), Bayesian Information Criterion (BIC) and residual ACF plots as a goodness of fit criteria. The lower the value of LR, AIC and BIC that represents the best fit among the identified models (Jung et al. 2006). ACF and PACF plots examine the residuals of the model; points should be within the 95% confidence intervals of the ACF functions. The "tscount" package in R was used to analyze road traffic accidents in Tanzania mainland using the Maximum Likelihood approach.

Results

Number of road traffic accidents in Tanzania Mainland

A total of 85,514 road traffic accidents in Tanzania were recorded from 2012 to 2017 as presented in Table 2. Dar es Salaam Region bear the highest burden with more than fifty percent of all accidents in Tanzania. These findings are in line with Boniface et al. (2016) and Tanzania Police Force (2016). Both reported that Dar es Salaam is the leading city in road traffic accidents compared to other regions in Tanzania.

Table 2: Number of road traffic accidents in Tanzania Mainland 2012-2017

Year	**Region n (%)**		**Tanzania Mainland**
	Dar es Salaam	**Other**	
2012	11,410 (48)	12,135 (52)	23,545
2013	12,983 (54)	10,859 (46)	23,842
2014	7,811 (54)	6,549 (46)	14,360
2015	3,574 (43)	4,763 (57)	8,337
2016	5,719 (58)	4,137 (42)	9,856
2017	2,687 (48)	2,887 (52)	5,574
Total	**44,184 (52)**	**41,330 (48)**	**85,514**

Source: Traffic Police records from 2012 to 2017.

The trend analysis is presented in Figure 2. It illustrates a downward pattern in both regions; Dar es Salaam as well as other combined regions. From 2015 to 2016, in Dar es Salaam Region the number of accident has mainly increased, suggesting the application of a more flexible approximation. Looking into the actual values, it is clearly demonstrated that a non-linear polynomial (Loess function) gives a much better fit. The difference is seen by comparing the actual values (in black dot), fitted linear function (in blue line) and Loess curve (in red curve), see Figure 2.

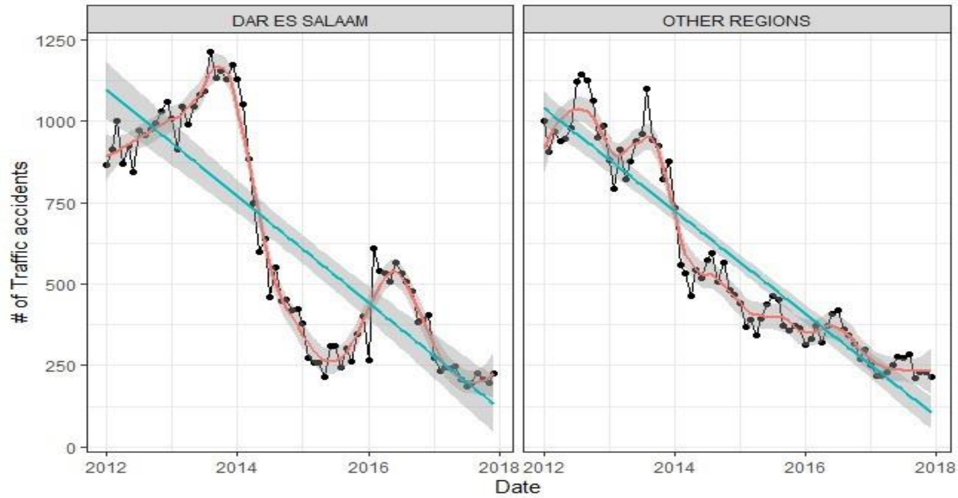


Figure 2: Trend of road traffic accidents during 2012-2017 (black dots), fitted linear function (blue) and fitted Loess curve (red) in Dar es Salaam and Other regions.

Figure 3 shows a similar trend of the number of road traffic accidents caused by human (reckless/dangerous driving, careless driving, over speeding and overtaking), vehicle (vehicle defects and poor vehicle lighting), physical/environmental (fire, road barriers, poor road infrastructure and railway crossing) and others (pedestrians). Since 2014, the plot seems to have a similar magnitude, and most observations are smaller than three hundred per month.

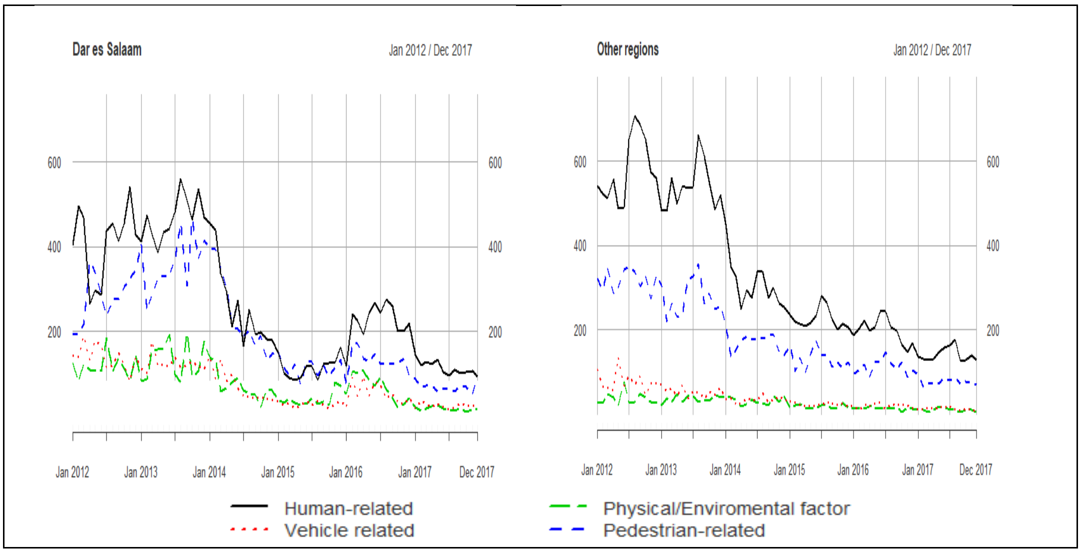


Figure 3: Recorded number of road traffic accidents by cause and sub-regional category.

Poisson and Negative Binomial Autoregressive Models

The estimates obtained for Dar es Salaam Region and other combined regions are presented in Tables 3-5 and Figures 4-6. The assessments look similar in terms of estimated parameters, with small differences in AIC and BIC values. In Dar es Salaam Region, Negative Binomial Models out-perform, while in other combined regions, the models including predictor component and time-covariate factors, are in favour-off of other models.

Table 3: AIC and BIC results for Dar es Salaam and other combined regions models

Variables	Poisson		Negative Binomial	
	Model 1	**Model 2**	**Model 3**	**Model 4**
Dar es Salaam Region				
AIC	1476.52	939.47	857.5	803.79
BIC	1485.63	959.96	868.88	826.55
Other Combined Regions				
AIC	1180.27	751.31	805.80	732.26
BIC	1189.38	771.80	817.18	755.03

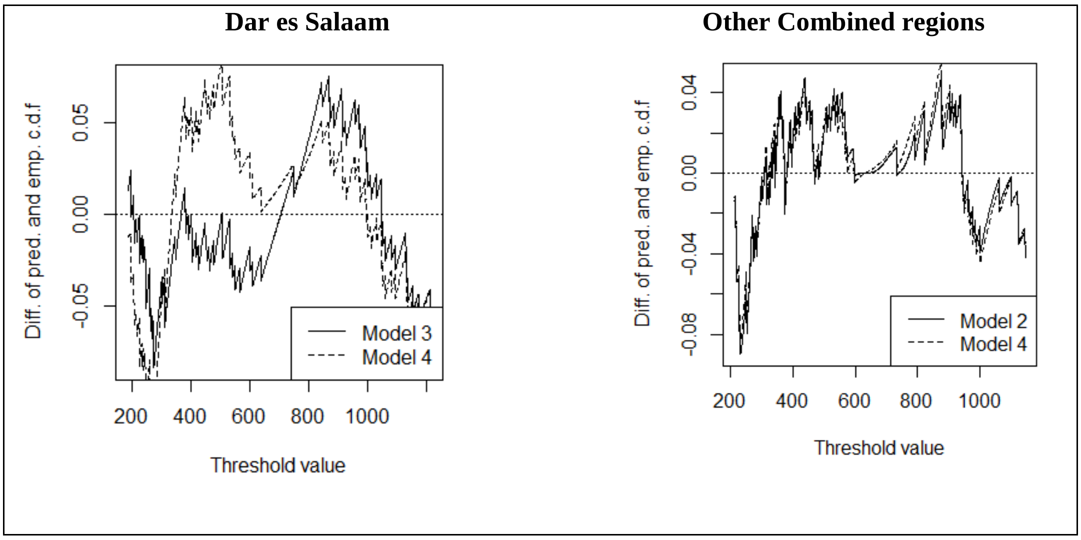
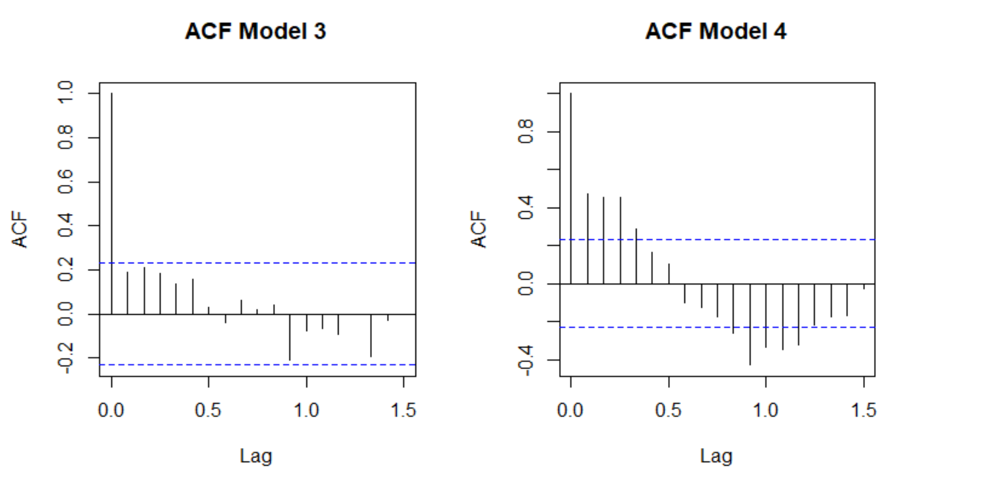


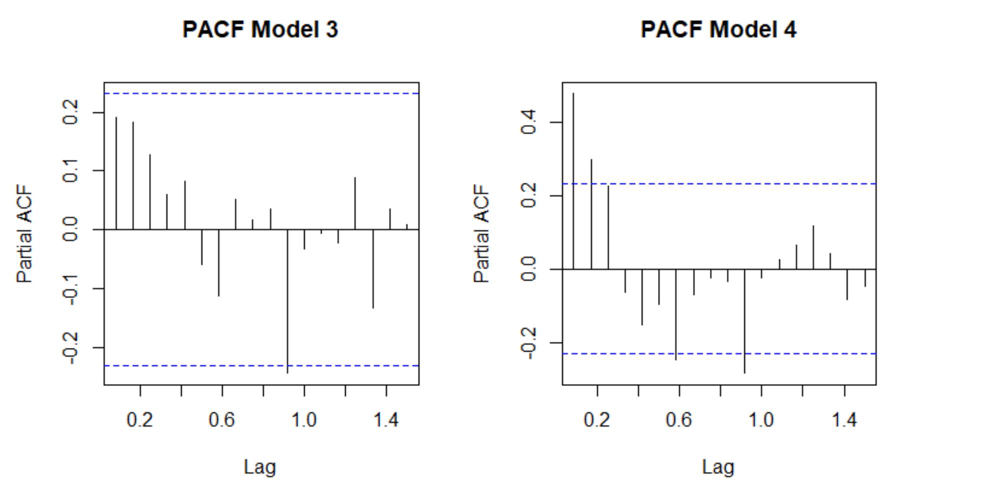
Figure 4: Marginal calibration of best fitted models for Dar es Salaam and other regions.

Table 4: Regression results for road traffic accident in Dar es Salaam

	Poisson		Negative Binomial	
Variables	**Model 1**	**Model 2**	**Model 3**	**Model 4**
(Intercept)	0.264 [0.12,4.26]	4.5003 [4.261,4.739]	0.2636 [-0.224,0.751]	4.5003 [3.835,5.166]
y_{t-1}	0.645 [0.03,0.05]	0.103 [0.05,0.16]	0.645 [0.37,0.92]	0.103 [-0.05,0.26]
y_{t-2}	0.362 [0.03,-0.05]	0 [-0.05,0.05]	0.362 [0.05,0.68]	0 [-0.17,0.17]
y_{t-3}	-0.046 [0.02,0.02]	0.068 [0.02,0.11]	-0.046 [-0.32,0.23]	0.068 [-0.07,0.21]
Human-related		0.001 [0.00121,0.0015]		0.001 [0.0009,0.0018]
Vehicles-related		0.002 [0.00154,0.0024]		0.002 [0.00063,0.0033]
Physical/ environmental factors		0.001 [0.00118,0.0017]		0.001 [0.00047,0.0024]
Pedestrian-related		0.001 [0.00061,0.001]		0.001 [0.00016,0.0014]
Parameters for				
Linear Trend		-0.028 [-0.04,-0.02]		-0.028 [-0.06,0.01]
Sigma square			0.0292	.010821



(a)

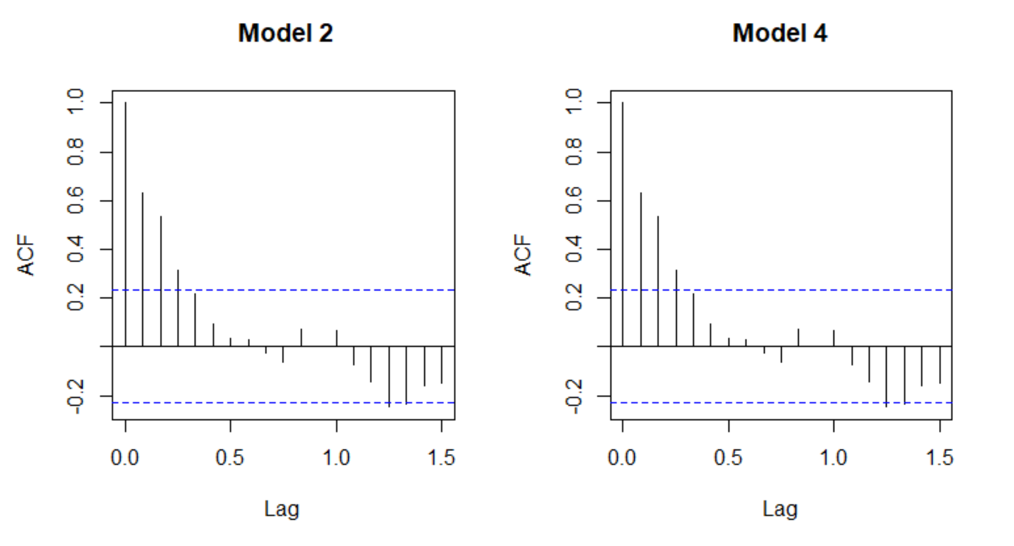


(b)

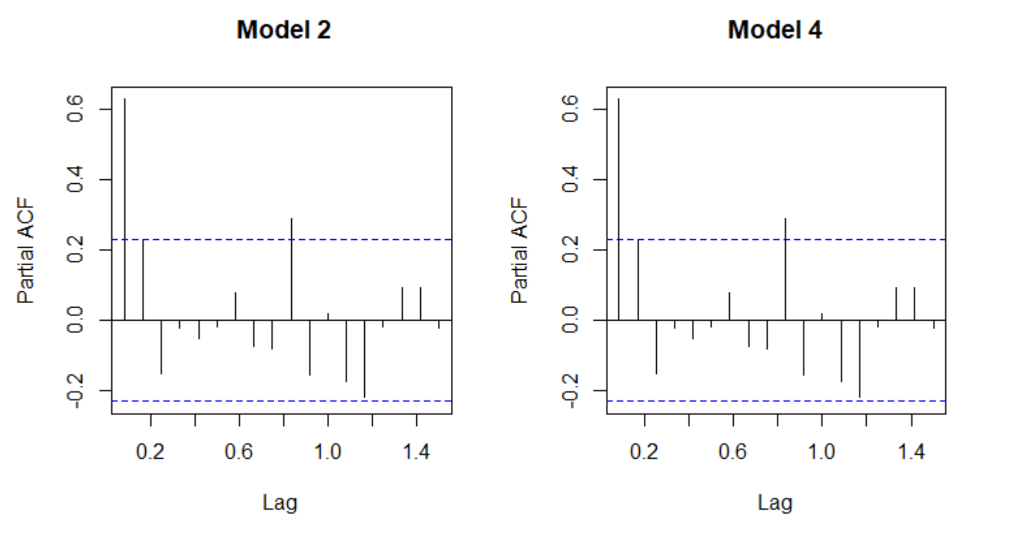
Figure 5: (a) ACF and (b) PACF of response residuals of best fitted model for Negative Binomial AR Models in Dar es Salaam.

Table 5: Regression results for road traffic accidents in other combined regions

	Poisson		**Negative Binomial**	
Variables	**Model 1**	**Model 2**	**Model 3**	**Model 4**
(Intercept)	0.105 [-0.03,0.24]	5.034 [4.69,5.37]	0.105 [-0.28,0.49]	5.034 [4.4,5.67]
y_{t-1}	0.85 [0.77,0.93]	0.075 [0.00,0.15]	0.85 [0.59,1.11]	0.075 [-0.06,0.21]
y_{t-2}	0.239 [0.13,0.34]	0.016 [-0.05,0.08]	0.239 [-0.11,0.59]	0.016 [-0.11,0.15]
y_{t-3}	-0.105 [-0.19,-0.02]	0.041 [-0.01,0.09]	-0.105 [-0.37,0.16]	0.041 [-0.06,0.14]
Human-related		0.001 [0.00084,0.0012]		0.001 [0.00067,0.0013]
Vehicles-related		0.001 [-0.00015,0.0012]		0.001 [-0.00078,0.0018]
Physical/ environmental factors		0.002 [0.00057,0.0026]		0.002 [-0.00037,0.0036]
Pedestrian		0.001 [0.00081,0.0015]		0.001 [0.0005,0.0018]
Parameters for				
Linear Trend		-0.08 [-0.1,-0.06]		-0.08 [-0.11,-0.05]
Sigma square			0.0136	.002986



(a)



(b)

Figure 6: (a) ACF and (b) PACF of response residuals of best fitted model for Poisson and Negative Binomial AR Models in other regions.

Discussion

All factors associated with RTA indicate a declining pattern. Human-related factors were higher in both regions than in the other three factors throughout the whole time frame. These findings are consistent with the results of Mcharo (2012), WHO (2018) and Wangdi et al. (2018), who reported that human factors were the most typical causes of road traffic accidents.

Having fitted two count time series models to assess trends and factors associated with the number of road traffic accidents in Tanzania, it is essential to compare them using model performance criteria (i.e. AIC and BIC). The values of AIC and BIC varied from 732.26 to 1476. 52 and 755.03 to 1485.63, respectively. Based on these results, the Negative Binomial Autoregressive models had smaller values of AIC and BIC than the Poisson Autoregressive models. This implies that the Negative Binomial Autoregressive models are the best-fitted models for these data (see Table 3). The marginal calibration is presented in Figure 4. In Dar es Salaam Region, the gap between the predicted and observed values for the best-fitted models was substantially large, while in other regions, the graphs are closely related. For the best fitted model, Figure 5 and 6 indicated a significant spike of one-month before in Dar es Salaam Region as well as in other combined regions.

The coefficients of the predictors are presented in Tables 4 and 5. All the estimates were positive, indicating an increase in the rate of reckless/dangerous driving, careless driving, over speeding and overtaking, vehicle defects and poor vehicle lighting, fire, road barriers, poor road infrastructure and railway crossing and pedestrians will increase the number of RTAs. In Dar es Salaam Regions, all factors appears to be significant (see results in Table 4), while in other combined regions (Table 5), human related and pedestrian were statistically significant at 5% level.

Conclusion

The results portrayed that the numbers of road traffic accidents in Tanzania were decreasing. This information is crucial for road safety professions to emphasize collective measures that reduce the number of accidents. Therefore, much focus should be put on people during road safety programs to increase road traffic awareness and encourage good driving behavior. The Poisson and Negative Binomial estimates give almost similar results for each region, though the Negative Binomial Autoregressive models seem superior. This justified that the data were over-dispersed and time-varying components are essential for analyzing the number of road traffic accidents in Tanzania. In order to evaluate the SDG target 3.6, further study should be conducted to relate number of accidents against fatalities based on Negative Binomial Autoregressive model. However, unavailability of timely data is a big challenge to promote evidence-based decisions.

Acknowledgments
We are grateful to the Tanzania Traffic Police for giving us access to the raw data.

Conflict of Interest: The authors declare that there is no conflict of interest.

Research Permit: Research permit was obtained from the Vice Chancellor, University of Dar es Salaam.

References

Ahmad A and Francq C 2016 Poisson QMLE of count time series models. *J. Time Ser. Anal.* 37(3): 291–314.

Boniface R, Museru L, Kiloloma O and Munthali V 2016 Factors associated with road traffic injuries in Tanzania. *Pan Afr. Med. J.* 23(1). https://doi.org/10.11604/pamj.2016.23.46.7487

Chen CWS, So MKP, Li JC and Sriboonchitta S 2016 Autoregressive conditional negative binomial model applied to over-dispersed time series of counts. *Stat. Methodol.* 31: 73–90.

Christou V and Fokianos K 2015 On count time series prediction. *J. Stat. Comput. Simul.* 85(2): 357–373.

Cox DR, Gudmundsson G, Lindgren G, Bondesson L, Harsaae E, Laake P, Juselius K and Lauritzen SL 1981 Statistical analysis of time series: Some recent developments. *Scand. J. Stat.* 8: 93-115.

Delavary FM, Mohammadzadeh MA and Fakoor V 2019 Time trends in gender-specific incidence rates of road traffic injuries in Iran. *Plos One* 14(5): e0216462.

Fell M, Jones A, Mallow P, Lema E and Mertner J 2017 Improvement of road safety in Tanzania mainland. *Final Report COWI for SUMATRA*. Tanzania.

Fokianos K 2012 Count time series models I. *Handbook of Statistics*. 30: 315–347. Elsevier.

JICA 2014 *Comprehensive Transport and Trade System Development Master Plan in the United Republic of Tanzania*. Ministry of Transport, The United Republic of Tanzania. Japan International Cooperation Agency (JICA) https://openjicareport.jica.go.jp/pdf/12150512_01.pdf

Jung R, Kukuk M and Liesenfeld R 2006 Time series of count data: Modeling, estimation and diagnostics. *Comput. Stat. Data Anal.* 51: 2350–2364.

Mcharo B 2012 *Motorcycle crash: Injuries pattern and associated factors among patients treated at Muhimbili Orthopaedic Institute (MOI)* [PhD Thesis]. Muhimbili University of Health and Allied Sciences.

NBS 2012 *Population and Housing Census*. National Bureau of Statistics (NBS), Ministry of Finances of Tanzania Dar es Salaam, Tanzania.

Petridou E and Moustaki M 2000 Human factors in the causation of road traffic crashes. *Eur. J. Epidemiol.* 16(9): 819–826.

Ravishanker N, Serhiyenko V and Willig M 2014 Hierarchical dynamic models for multivariate time series of counts. *Statist. Interface* 7: 559-570.

Tanzania Police Force 2016 *Crime and traffic incidents statistics report*. Tanzania.

Walugembe F, Levira F, Ganesh B and Lwetoijera DW 2020 A retrospective study on the epidemiology and trends of road traffic accidents, fatalities and injuries in three Municipalities of Dar es Salaam region, Tanzania between 2014-2018. *Pan Afr. Med. J.* 36(24).

Wangdi C, Gurung MS, Duba T, Wilkinson E, Tun ZM and Tripathy JP 2018 Burden, pattern and causes of road traffic accidents in Bhutan, 2013–2014: A police record review. *Int. J. Inj. Contr. Saf. Promot.* 25(1): 65–69.

WHO 2015 Global status report on road safety 2015. Geneva.

WHO 2018 Global status report on road safety 2018. Geneva.

Zeileis A, Kleiber C and Jackman S 2008 Regression models for count data in R. *J. Stat. Softw.* 27(8): 1-25.