

## Research Article

# Practical Minimum Sample Size for Road Crash Time-Series Prediction Models

Fady M. A. Hassouna  and Khaled Al-Sahili 

*Civil Engineering Department, An-Najah National University, Nablus, P.O. Box 7, State of Palestine*

Correspondence should be addressed to Fady M. A. Hassouna; [fady.h@najah.edu](mailto:fady.h@najah.edu)

Received 24 October 2020; Revised 9 December 2020; Accepted 18 December 2020; Published 29 December 2020

Academic Editor: Valeria Vignali

Copyright © 2020 Fady M. A. Hassouna and Khaled Al-Sahili. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Road crashes are problems facing the transportation sector. Crash data in many countries are available only for the past 10 to 20 years, which makes it difficult to determine whether the data are sufficient to establish reasonable and accurate prediction rates. In this study, the effect of sample size (number of years used to develop a prediction model) on the crash prediction accuracy using Autoregressive integrated moving average (ARIMA) method was investigated using crash data for years 1971–2015. Based on the availability of annual crash records, road crash data for four selected countries (Denmark, Turkey, Germany, and Israel) were used to develop the crash prediction models based on different sample sizes (45, 35, 25, and 15 years). Then, crash data for 2016 and 2017 were used to verify the accuracy of the developed models. Furthermore, crash data for Palestine were used to test the validity of the results. The used data included fatality, injury, and property damage crashes. The results showed similar trends in the models' prediction accuracy for all four countries when predicting road crashes for year 2016. Decreasing the sample sizes led to less prediction accuracy up to a sample size of 25; then, the accuracy increased for the 15-year sample size. Whereas there was no specific trend in the prediction accuracy for year 2017, a higher range of prediction error was also obtained. It is concluded that the prediction accuracy would vary based on the varying socioeconomic, traffic safety programs and development conditions of the country over the study years. For countries with steady and stable conditions, modeling using larger sample sizes would yield higher accuracy models with higher prediction capabilities. As for countries with less steady and stable conditions, modeling using smaller sample sizes (15 years, for example) would lead to high accuracy models with good prediction capabilities. Therefore, it is recommended that the socioeconomic and traffic safety program status of the country is considered before selecting the practical minimum sample size that would give an acceptable prediction accuracy, therefore saving efforts and time spent in collecting data (more is not always better). Moreover, based on the data analysis results, long-term ARIMA prediction models should be used with caution.

## 1. Introduction

Time-series analysis is a common technique used by numerous research studies to analyze trends of certain phenomena and to predict future conditions. This technique has been applied in various fields: engineering, scientific, social, medical, etc. Typical time-series analysis includes historical data over an extended period of time. There are also several statistical models used for representing the time-series and trends and for forecasting.

A quick scan of the literature shows that different studies used different sample sizes for modeling, depending on the

phenomenon under investigation, the type of model used, and the constraints of data availability. The majority of the published developed models yielded reasonable level of goodness-of-fit and significance, which were represented by various statistical indicators. At the same time, a scan through the web showed many researchers asking, "What is the reasonable minimum sample size for appropriate modeling?"

It is a common rule-of-thumb in statistics that more is better; for time-series analysis. However, some statisticians revealed that "more may not always be better." The data-generating process may have changed strongly over time, so

that the data from before the change may not reflect current and future dynamics of your series any more [1]. In this case, one should look for the “change-point,” if it exists.

In transportation and traffic engineering, many time-series studies were also conducted. One of the common applications is modeling road traffic crashes for a region or a country and predicting future crash trends. The Box and Jenkins ARIMA models were commonly used in these studies. The Box and Jenkins method typically recommends a minimum of 50 observations for an ARIMA model. This is recommended to cover seasonal variations and effects. At the same time, the method recommends that a “good” prediction is the one that produces appropriate (narrow) confidence limits (which is typically in the vicinity of 95%) and useful results.

In the context of road traffic crashes, the question remains, “What is the ‘reasonable’ and ‘practical’ value for the minimum sample size?” The occurrence of road traffic crashes might depend on many socioeconomic and political characteristics of the region, which are drastically changing in some countries over a period of 50 years. On the other hand, a good model will allow for changing trend and seasonal patterns. However, the issue is whether the way things change can be properly modeled [1].

Changes in some countries may have gone steady, for example, several developed countries. However, changes in developing countries are very dynamic and subject to numerous social and economic, as well as political, conditions and traffic safety programs. Over the past fifty years, which is the recommended minimum sample size on a yearly basis for ARIMA models, several countries moved up from being “underdeveloped” to the level of “developed” country. Certainly, these changes greatly affect the road crash patterns, in a way that ARIMA models, or other time-series models, might not be able to properly work.

In the Palestinian case, for example, road traffic crashes were collected and reported by the Israeli authority starting from 1970 following certain procedures for crash recording and archiving system. In 1993, the Oslo Peace Accord was signed between the Palestinian Liberation Organization and Israel. As a result, the Palestinian National Authority was created and took partial administrative control over parts of Palestine, including crash data collection and achieving [2]. The socioeconomic characteristics of the region have dramatically changed since then: higher income level, mobility, auto ownership, etc. In year 2000, there was a Palestinian uprising against the Israeli authorities, which lasted for few years and involved road closures and blockage of road mobility. As a result, road traffic crash patterns dramatically changed during these two major events. Therefore, using road crashes over that extended period of time will include those “not-normal” patterns.

The issue at hand in this paper is not to question the statistical requirements of the minimum sample size for modeling time-series and forecasting using ARIMA models; this is left for statisticians. It is rather, and for road crash analysis, what the “reasonable” and “practical” sample size would be for modeling time-series and forecasting using the Box and Jenkins ARIMA models with “appropriate” confidence limits.

The paper answers this question by modeling road crashes from four countries, testing different sample sizes for each country, and assessing their prediction capabilities and the associated significance levels for each country, using the ARIMA models. The results were also verified using road crash for Palestine.

## 2. Literature Review

Numerous studies of road crashes time-series and trends analysis were conducted in several parts of the world. Models and forecasts were also developed based on the established databases and their analyses. The sample sizes in these studies varied with differences among the researchers.

In Nigeria, a time-series approach was followed to investigate the consequences (casualties) of road crashes in the country [3]. The authors established that road crashes were increasing at an alarming rate in Nigeria. A time-series model was developed using ARIMA model for the yearly crash data based on crash records for years (1960–2013). A sample size of 51 years (1960–2011) was used for developing a model, and 2 years (2012–2013) were used for validation purposes. Two ARIMA models were obtained for the injury and total casualty consequences and for fatal consequences with high significance levels. The models forecasted that the road crashes will continue to increase in the following 7 years. The authors also recommended that the models should be used with caution for predicting future conditions beyond the forecasted period because long time forecast may give arbitrary large forecast, and enforcement conditions in the country may change in the future.

In another study in Nigeria, based on an annual database of 32 years (1970–2001), time-series analysis showed that road traffic accidents were on the decrease with the exception of Lagos Island local government area [4]. Another study in Lagos, Nigeria, used time-series to analyze road crashes based on a 20-year database (1989–2008) and to forecast crashes for the following 4 years (2009–2012).

Mutangi [5] analyzed road traffic crashes in Zimbabwe using time-series based on a database for years 1997 to 2013 (17 years); the study also predicted the annual number of crashes in the future years. The author used ARIMA model, where three models were suggested and the most appropriate one with the smallest corrected Akaike Information Criteria (AICc) and Bayesian Information Criteria (BIC) was chosen as the best model. The crash data included numbers of crashes, injuries, and fatalities. The results showed that forecasting of the number of road traffic crashes using a white noise process is difficult because the values at different times are statistically independent.

In Ghana, Avuglah et al. [6] applied ARIMA models to study the trends and patterns of road crashes and to make a five-year prediction. Twenty years (1991 to 2011) of annual data were used for modeling purposes. The study showed an increasing trend for the coming five years.

Several studies of time-series on road crashes were conducted in Iran. Parvareh et al. [7] assessed and predicted road crash injuries trends using ARIMA time-series models in Kurdistan Province, Iran. The authors established that

fatality rates, which resulted from road crashes in Iran, were twice the global average in 2002, continued at an increasing rate, and remained among the highest in the world. The study used a database of crash records for the period of March 2009 to February 2015 (72 months) and predicted crashes up to February 2017. Different ARIMA models were developed for car occupant, motorcyclist, and pedestrian injuries, all with a significance level of 5% or higher. The analysis of motorcyclists and pedestrian injuries showed a seasonal pattern with a peak during August. The study also showed an increasing trend in the frequency of nonfatal injuries. The decline in fatalities was attributed to the national extensive efforts and interventions to improve safety.

Another study forecasted the trend of traffic crash mortality in Kermanshah Province in west Iran [8]. The study assessed fatalities of road crashes using time-series forecast based on crash records from January 2013 to December 2015 (36 months). The mean square error was used to determine the model with the best goodness-of-fit. The study found that the SARIMA model, which is an expanded version of the ARIMA model that is influenced by seasonality factors, was the best-fit model for data with Pearson correlation coefficient of 0.75 ( $P=0$ ). Based on the model's prediction, traffic fatalities showed a decreasing trend in the following years, with a confidence level of 95%. Similar work was done by Yousefzadeh-Chabok [9] for Zanjan Province, Iran, using ARIMA model based on a monthly database for years 2007–2013.

Similarly, predicted results from ARIMA time-series model, which was developed based on data from 1951 to 2011 (60 years), revealed that road traffic crashes would be increasing in year 2012 in China [10]. Furthermore, based on fatality data from 1972 to 2010 (40 years), ARIMA model was used in predicting Malaysian road fatalities for the years 2015 and 2020, and the results showed the predicted fatalities for the following five years [11]. The model was significant at 95% confidence level.

Al-Ghamdi [12] developed three ARIMA forecasting models for the traffic crash frequencies, injuries, and fatalities in Saudi Arabia based on the annual crash records for years 1980–1991 (11 years). The developed models achieved a 95% significance level. The study revealed that, based on the trend analysis, there would be an increase in the studied types of crashes. Therefore, the study highlighted the importance of improving the existing safety programs in the country.

Al-Zyood [13] developed the most appropriate ARIMA model, out of nine tested ARIMA modes, to forecast car crashes in Saudi Arabia based on crash database for years 1998 to 2016 (18 years). Based on the developed model, the author forecasted car crashes for the following 7 years with 95% prediction intervals. The prediction showed that traffic crashes, injuries, and fatalities will be increasing.

In Jordan, Al-Omari et al. [14] analyzed traffic crash trends and characteristics in the country for years 1998–2010 (13 years). The authors established that traffic crashes were continuously increasing over the study period as a result of increasing population and motorization level. The authors modeled the traffic crash frequencies and fatalities in terms

of motorization level. Multiple-linear regression and logarithmic models were developed, and the most appropriate models were used based on  $R^2$  and standards error criteria.

Jadaan et al. [15] also analyzed the traffic safety in Jordan based on traffic crashes database for 10 years (2000–2009). The authors used multiple-linear regression models to correlate fatalities with population, number of registered vehicles, and total length of paved roads. The authors established that crash cost was \$475 million in 2009 and is expected to reach \$786 million in 2030, which is a burden on the national economy.

Ljubič et al. [16] analyzed road crashes in the UK for the years of 1979–1999 (21 years) using simple statistical visualization and graphic presentation. The authors also used a clustering approach to the analysis of time-series data to find trends of road crashes in the country through different time segments.

Another study in the UK analyzed traffic crashes trends for 1.6 million traffic crashes from 2005 to 2014 [17]. The author analyzed the time-series based on the number of weekly crashes, casualties, and vehicles involved across the UK. Exponential smoothing and ARIMA models were used for the analysis. The author established that both models are useful in making future predictions.

In the USA, Tang et al. [18] evaluated the predictive power of SPF for crash frequencies on two-lane rural highways in three districts in the State of Pennsylvania. The authors used the negative binomial (NB) and random parameters negative binomial (RPNB) modeling frameworks. The study found that random parameters negative binomial (RPNB) model provides better predictions than the negative binomial (NB) model when applied to within-sample observations, while the NB model provides better predictions than RPNB model when applied to out-of-sample observations. The study concluded that when sample sizes were more than 50, the estimates were fairly insensitive.

Liu and Sharma [19] analyzed the trends of spatial-temporal effects in road crashes based on crash data for 10 years (2006–2015) in all Iowa, USA, counties. The study showed that both spatial and temporal correlations are critical in crash frequency modeling; the vehicle-miles-traveled was the only significant variable. The study also used the integrated nested Laplace approximation (INLA) to estimate Bayesian spatio-temporal models. The study found that fatal crashes showed decreasing trends in all counties with various rates among them.

Furthermore, as related to sample size, Freeman and Leith [20] estimated the number of traffic crash-related cervical spine injuries in the United States. The authors concluded that because of a limited sample size and restrictive criteria for both crash and injury inclusion, they could not make such estimates.

Russo et al. [21] studied crash severity on more than 2000 km of undivided rural roads in Italy for a period of 5 years (2006–2010). The study found that annual average daily traffic, lane width, curvature change rate, length, and vertical grade are important variables in explaining the severity of crashes. The negative binomial regression model was used to predict injury frequency and fatality per year on

homogeneous road segments. The research confirmed the effectiveness of the Highway Safety Manual (HSM), while it showed that the HSM overestimates the crash frequency for rural undivided roads.

The summary of the reviewed literature is presented in Table 1. In summary, numerous studies have been conducted on the subject of time-series analysis and trends of road traffic crashes in several parts of the world. However, other studies have followed different approaches focusing on speed and safety issues, such as Russo et al. [22, 23]. ARIMA models were among the most commonly used models for analysis and forecasting. There was clear variation among the research works in the sample sizes used for ARIMA models, which ranged from as low as 5 to as high as 84. The majority of the models provided a reasonable level of significance and prediction capabilities. None of them questioned the requirement for minimum sample size. Therefore, the question remains, "What is practically the minimum sample size required to provide a reasonable significance level of ARIMA models based on data availability and varying socioeconomic constraints?"

### 3. Data and Methodology

The data used in this study were acquired from the Organisation for Economic Co-operation and Development [24]. Road crash data for four countries (Denmark, Turkey, Germany, and Israel) for the period 1971–2015 were used in this study to determine the effect of sample size on the accuracy and validity of road crash prediction models (see Figure 1). Crash data for years 2016 and 2017 were used to test the prediction capabilities of the developed models.

Four different sample sizes (number of years) were used to develop crash prediction models for each country in order to assess the effect of sample size on the accuracy of the models developed separately for each country. Based on the available data, the sample sizes were 45 years (1971–2015), 35 years (1981–2015), 25 years (1991–2015), and 15 years (2001–2015).

The developed road crash prediction models in this study are based on the Box–Jenkins methodology for Autoregressive Integrated Moving Average (ARIMA) modeling. The ARIMA model has three parts: (1) the autoregressive part is a linear regression that relates past values of data series to future values; (2) the integrated part indicates how many times the data series has to be differenced to produce a stationary series; and (3) the moving average part relates past forecast errors to future values of the data series [25–27].

In the ARIMA model, the parameters  $p$ ,  $d$ , and  $q$  represent the number of ordinary autoregressive, differencing, and moving average parameters, respectively. In other words, the  $p$  and  $q$  are the number of significant lags of the autocorrelation function (ACF) and the partial autocorrelation function (PACF) plots, respectively, and  $d$  is the different order needed to remove the ordinary non-stationarity in the mean of error terms [28].

The method applied to develop the ARIMA forecast model was developed by Box and Jenkins [29] as follows:

- (i) Examine the stationarity of the time-series. If the time-series is nonstationary, then the difference and power transformation are applied to it. Thus, the time-series converts into the stationary time-series with uniform variance.
- (ii) Estimate the model parameters  $p$ ,  $d$ , and  $q$  for the ARIMA ( $p$ ,  $d$ ,  $q$ ).
- (iii) Check the validity of the model; the goodness-of-fit tests are performed using statistical information.

In order to apply Box and Jenkins methodology, several statistical tests were conducted. The Augmented Dickey–Fuller test was used to examine the stationarity of the time-series. The Mean Absolute Error (MAE) was used to select the best-fit model. The Ljung–Box, quantile–quantile (q–q) plot of the residuals and the Kolmogorov–Smirnov test of the residuals were used to test the validity of the developed model, as recommended by Makridakis et al. [30].

Four crash prediction models of different sample sizes (45, 35, 25, and 15 years) were developed for each of the four subject countries using the annual number of road crashes for the period 1971–2015, as shown in Table 2. Finally, the crash data from 2016 and 2017 were used to verify the accuracy of the prediction for each model in order to determine the effect of sample size on the accuracy of prediction models.

### 4. Data Analysis and Discussion

In this study, the ARIMA method was used to develop 16 crash prediction models, four models for each of the four countries. Following the ARIMA procedure for modeling, all possible prediction models were developed for each sample size of the subject countries. A total of 56 possible models were tested for each country (grand total of 224 tested models). Table 3 shows all possible crash prediction models for Germany as an example; similarly, all possible crash prediction models for the remaining countries were also developed.

The best-fit model was selected for each sample size, based on the minimum MAE. Ljung–Box and Kolmogorov–Smirnov tests of the residuals were then used to test the validity of the selected model. Finally, the best-fit models for each sample size of all subject countries were summarized in Table 4.

After selecting the best-fit model for each sample size of each country, the validity of each model was verified by testing the normal distribution of the residuals using Ljung–Box and Shapiro–Wilk tests, and the parameters of all models were determined. As an example, (1)–(4) show the prediction models of Germany at the sample sizes of 45, 35, 25, and 15, respectively. Similarly, the equations of models for all remaining countries were determined as well.



TABLE 1: Summary of literature related to sample size, models, and results.

Reference	Sample size/ period	Method/location	Findings
Adebola et al. [3]	51 years (1960–2013)	ARIMA model, Nigeria	Road injury and fatal crashes will increase in the following 7 years; however, the models should be used with caution for predicting future conditions beyond the forecasted period
Atubi, 2013 [4]	32 years (1970–2001)	Time-series analysis, Nigeria	Road traffic accidents were generally on the decrease
Mutangi, 2015 [5]	17 years (1997–2013)	ARIMA model, Zimbabwe	Forecasting of the number of road traffic crashes using a white noise process is difficult because the values at different times are statistically independent
Avuglah et al. [6]	20 years (1991–2011)	ARIMA model, Ghana	The trends and patterns of road crashes were studied, and a five-year prediction was made. The study showed an increasing trend for the coming five years. Injury crashes for the following 24 months were predicted. Motorcyclists and pedestrian injuries showed a seasonal pattern. Results showed an increasing trend in the frequency of nonfatal injuries and a decline in fatalities.
Parvareh et al. [7]	72 months (2009–2015)	ARIMA model, Kurdistan Province, Iran	Traffic fatalities showed a decreasing trend in the following years
Zolala et al. [8]	36 months (2013–2015)	SARIMA model, Kermanshah Province, Iran	They found similar results to Zolala et al., 2016 [8]
Yousefzadeh-Chabok et al. [9]	84 months (2007–2013)	ARIMA model, Zanjan Province, Iran	Road traffic crashes would be increasing in year 2012
Yuan et al., 2013 [10]	60 years (1951–2011)	ARIMA model, China	They predicted road fatalities for the following five years (2015 and 2020)
Rohayu et al. [11]	40 years (1972–2010)	ARIMA model, Malaysia	An increase in the studied types of crashes was revealed
Al-Ghamdi, 1995 [12]	11 years (1980–1991)	ARIMA model, Saudi Arabia	Car crashes, injuries, and fatalities were forecasted for the following 7 years; all will be increasing
Al-Zyood, 2017 [13]	18 years (1998–2016)	ARIMA model, Saudi Arabia	Traffic crash frequencies and fatalities in terms of motorization level were modeled; traffic crashes were continuously increasing
Al-Omari et al. [14]	13 years (1998–2010)	Multiple-linear regression and logarithmic models, Jordan	Fatalities were correlated with population, number of registered vehicles, and total length of paved roads; the crash cost will be increasing
Jadaan et al. [15]	10 years (2000–2009)	Multiple-linear regression, Jordan	A clustering approach was used to find trends of road crashes through different time segments
Ljubič et al. [16]	21 years (1979–1999)	Simple statistical visualization and graphic presentation, UK	Crashes, casualties, and vehicles were modeled. The models are useful in making future predictions.
Colum, 2018 [17]	Weekly data (2005–2014)	ARIMA model, UK	The RPNB model provides better predictions when applied to within-sample observations, while the NB model provides better predictions when applied to out-of-sample observations
Tang et al. [18]	8 years (2005–2012)	Negative binomial (NB) and random parameters negative binomial (RPNB), USA	Spatial and temporal correlations are critical in crash frequency modeling; the vehicle-miles-traveled was the only significant variable. Fatal crashes showed decreasing trends
Liu and Sharma, 2017 [19]	10 years (2006–2015)	Integrated nested Laplace approximation (INLA) to estimate Bayesian spatiotemporal models, USA	The AADT, lane width, curvature change rate, length, and vertical grade are important variables in explaining the severity of crashes. The HSM overestimates the crash frequency for rural undivided roads.
Russo et al. [21]	5 years (2006–2010)	Negative binomial regression, Italy	

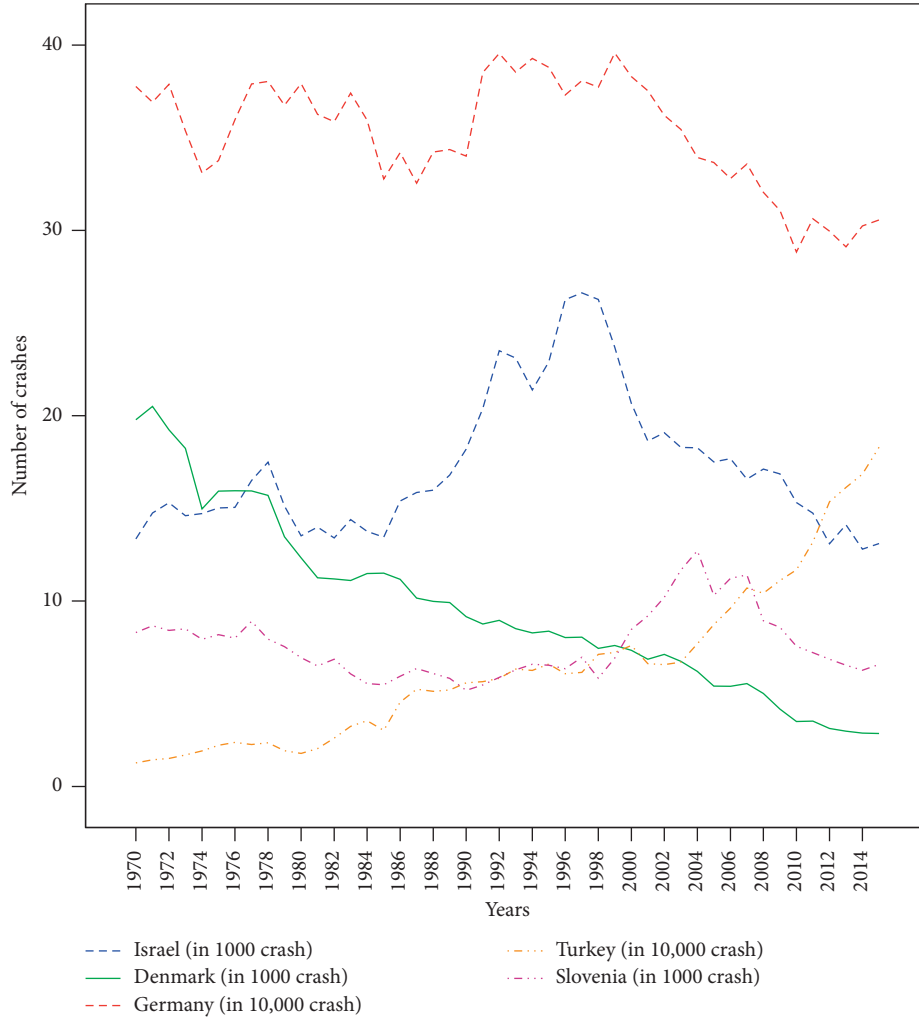


FIGURE 1: Trends of road crashes for the studied countries (1971–2015).

TABLE 2: Prediction models based on different sample sizes (number of years).

	Sample size (years)			
Country	45 (1971–2015)	35 (1981–2015)	25 (1991–2015)	15 (2001–2015)
Denmark	Model 1	Model 2	Model 3	Model 4
Turkey				
Germany				
Israel				

$$\Delta(Y_t, 2) = -54.942 - 1.227Y_{t-1} - 0.993Y_{t-2} + 0.186\varepsilon_{t-1} - 0.197\varepsilon_{t-2} - 0.988\varepsilon_{t-3}, \quad (1)$$

$$\Delta(Y_t, 2) = -74.934 - 1.186Y_{t-1} - 0.974Y_{t-2} + 0.940\varepsilon_{t-1} - 0.136\varepsilon_{t-2} - 0.949\varepsilon_{t-3}, \quad (2)$$

$$\Delta(Y_t, 2) = -1001.693 - 1.478Y_{t-1} - 0.791Y_{t-2} - 0.313Y_{t-3} + 0.984\varepsilon_{t-3}, \quad (3)$$

$$\Delta(Y_t, 2) = -917.691 - 0.945Y_{t-1} - 0.526Y_{t-2} - 0.344Y_{t-3} - 0.532\varepsilon_{t-1} - 0.467\varepsilon_{t-2}, \quad (4)$$

$Y_t$  is the road crash forecast at time (year)  $t$ .  $Y_{t-1}$ ,  $Y_{t-2}$ , and  $Y_{t-3}$  are the road crashes at time (years) lag  $t-1$ , lag  $t-2$ ,

and lag  $t-3$ .  $\varepsilon_{t-1}$ ,  $\varepsilon_{t-2}$ , and  $\varepsilon_{t-3}$  are the error terms at time (years)  $t-1$ ,  $t-2$ , and  $t-3$ .

TABLE 3: All possible crash prediction models for Germany.

Sample size	Model	Model fit statistics	
		Stationary R-squared	MAE
45	(0, 2, 1)	0.476	13208.783
	(1, 2, 0)	0.249	15495.564
	(1, 2, 1)	0.476	13143.372
	(2, 2, 0)	0.375	14227.440
	(0, 2, 2)	0.476	13136.482
	(2, 2, 2)	0.480	13020.020
	(1, 2, 2)	0.476	13159.988
	(2, 2, 1)	0.477	13143.070
	(0, 2, 3)	0.475	13089.132
	(3, 2, 0)	0.389	13912.752
	(1, 2, 3)	0.479	13160.837
	(3, 2, 1)	0.481	12830.792
	(2, 2, 3)	0.519	12101.438
	(3, 2, 2)	0.482	12830.982
35	(0, 2, 1)	0.490	12691.374
	(1, 2, 0)	0.309	14341.679
	(1, 2, 1)	0.496	12658.162
	(2, 2, 0)	0.453	12935.550
	(0, 2, 2)	0.495	12614.442
	(2, 2, 2)	0.496	12689.400
	(1, 2, 2)	0.493	12647.033
	(2, 2, 1)	0.496	12668.552
	(0, 2, 3)	0.497	12526.896
	(3, 2, 0)	0.468	12926.740
	(1, 2, 3)	0.496	12701.953
	(3, 2, 1)	0.516	12127.803
	(2, 2, 3)	0.538	12001.359
	(3, 2, 2)	0.516	12098.514
25	(0, 2, 1)	0.490	12691.374
	(1, 2, 0)	0.226	12442.211
	(1, 2, 1)	0.257	12424.256
	(2, 2, 0)	0.245	12544.195
	(0, 2, 2)	0.256	12420.705
	(2, 2, 2)	0.276	11818.040
	(1, 2, 2)	0.296	11666.041
	(2, 2, 1)	0.263	12649.132
	(0, 2, 3)	0.256	12423.869
	(3, 2, 0)	0.259	12368.022
	(1, 2, 3)	0.292	11609.888
	(3, 2, 1)	0.308	11446.555
	(2, 2, 3)	0.292	11848.622
	(3, 2, 2)	0.297	11773.639
15	(0, 2, 1)	0.595	8823.144
	(1, 2, 0)	0.429	9775.784
	(1, 2, 1)	0.661	8052.656
	(2, 2, 0)	0.495	9523.153
	(0, 2, 2)	0.702	7022.490
	(2, 2, 2)	0.665	8067.804
	(1, 2, 2)	0.649	8227.577
	(2, 2, 1)	0.672	7703.740
	(0, 2, 3)	0.715	7109.722
	(3, 2, 0)	0.625	8523.649
	(1, 2, 3)	0.702	7431.874
	(3, 2, 1)	0.709	7159.754
	(2, 2, 3)	0.728	6992.708
	(3, 2, 2)	0.712	6934.599

Finally, the annual crash data for 2016 and 2017 were used to verify the accuracy of each model. The difference between the observed and the forecasted crash data

(forecast error) was determined, and the percentage of errors was used to compare the results of all models for the different sample sizes in order to identify the forecast error trends, as shown in Tables 5 and 6, for years 2016 and 2017, respectively.

The results of forecast error for 2016 showed that the percentage of forecast error varied from as low as 1.30% for Germany for the model with 35 sample size to as high as 15.89% for Denmark for the model of 25 sample size, as shown in Table 5

As for the forecast errors for 2017, the results showed that the percentage of forecast error varied from 0.02% for Germany with the model of sample size 25 to 30.06% for Denmark with the model of sample size 25, as shown in Table 6.

For Denmark, the results in Table 5 show that the forecast errors for 2016 for models of sample sizes 45, 35, 25, and 15 were 4.86%, 13.22%, 15.89%, and 9.54%, respectively. In other words, there was an increase in the percentage of forecast error by decreasing the sample size to 25, which was followed by a decrease in the percentage of forecast error at sample size 15. On the other hand, for 2017, the forecast errors for models of sample sizes of 45, 35, 25, and 15 were 7.86%, 26.99%, 30.06%, and 26.01%, respectively, as presented in Table 6.

For Turkey, the results show that the forecast errors for 2016 for models of sample sizes of 45, 35, 25, and 15 were  $-4.14\%$ ,  $-4.63\%$ ,  $-6.19\%$ , and  $5.64\%$ , respectively. There was a slight increase in the percentage of forecast error by decreasing the sample size to 25. This was followed by a slight decrease in the percentage of forecast error at a sample size of 15. On the other hand, for 2017, the forecast errors for models of sample sizes 45, 35, 25, and 15 were  $-10.36\%$ ,  $-11.73\%$ ,  $-17.61\%$ , and  $-19.41\%$ , respectively, all above 10% difference.

As for Germany, the results for 2016 models of designated sample sizes were 1.3%, 2.31%, 3.5%, and 1.79%, respectively. There was an increase in the percentage of forecast error when decreasing the sample size to 25. As for 2017, the forecast errors were  $-1.15\%$ ,  $-0.26\%$ ,  $-0.02\%$ , and  $-0.69\%$ , respectively; all forecast errors for 2016 and 2017 were relatively low.

For Israel, the results show that the forecast errors for 2016 for models of the different sample sizes were 3.19%, 5.60%, 9.92%, and 9.00%, respectively. There was a general increase in the percentage of forecast error by decreasing the sample size. As for 2017, the same trend was generally observed (12.10%, 16.87%, 12.23%, and 26.01%, respectively), except at sample size of 35.

Overall, the results of the forecast errors for 2016 showed that the model of 45-year sample size had the lowest percentage of forecast error among all sample sizes for all subject countries, as shown in Figure 2. In other words, the best prediction accuracy was obtained using the 45-year sample size.

On the other hand, the results for year 2016 showed that the percentages of forecast errors increased by decreasing the sample size from 45 to 25 years, and then the forecast errors decreased at a sample size of 15 for all subject countries. The

TABLE 4: The selected best-fit models for each sample size of all subject countries.

Country	Sample size	Best-fit ARIMA model	Model fit statistics		
			Stationary <i>R</i> -squared	<i>R</i> -squared	MAE
Denmark	45	(2, 2, 2)	0.361	0.966	523.517
	35	(1, 2, 3)	0.411	0.979	292.596
	25	(1, 2, 3)	0.545	0.975	271.426
	15	(2, 2, 3)	0.488	0.947	276.796
Turkey	45	(2, 2, 3)	0.389	0.984	3851.541
	35	(2, 2, 3)	0.393	0.977	4515.151
	25	(2, 2, 3)	0.505	0.981	4189.615
	15	(1, 2, 2)	0.193	0.966	5514.014
Germany	45	(2, 2, 3)	0.519	0.742	12101.438
	35	(2, 2, 3)	0.538	0.780	12001.359
	25	(3, 2, 1)	0.308	0.821	11446.555
	15	(3, 2, 2)	0.712	0.834	6934.599
Israel	45	(0, 2, 1)	0.246	0.863	1095.791
	35	(0, 2, 1)	0.255	0.874	1132.513
	25	(2, 2, 3)	0.347	0.881	1181.735
	15	(2, 2, 3)	0.662	0.813	665.480

TABLE 5: The forecast error of the best-fit models for year 2016.

Country	Sample size	Observed	Forecast	Error	Error (%)
Denmark	45	2882	2742	140	4.86
	35	2882	2501	381	13.22
	25	2882	2424	458	15.89
	15	2882	2607	275	9.54
Turkey	45	185128	192793	-7665	-4.14
	35	185128	193693	-8565	-4.63
	25	185128	196584	-11456	-6.19
	15	185128	195566	-10438	-5.64
Germany	45	308145	304128	4017	1.30
	35	308145	301026	7119	2.31
	25	308145	297365	10780	3.50
	15	308145	302629	5516	1.79
Israel	45	12966	12553	413	3.19
	35	12966	12240	726	5.60
	25	12966	11680	1286	9.92
	15	12966	11799	1167	9.00

TABLE 6: The forecast error of the best-fit models for year 2017.

Country	Sample size	Observed	Forecast	Error	Error (%)
Denmark	45	2864	2639	225	7.86
	35	2864	2091	773	26.99
	25	2864	2003	861	30.06
	15	2864	2119	745	26.01
Turkey	45	182669	201597	-18928	-10.36
	35	182669	204089	-21420	-11.73
	25	182669	214840	-32171	-17.61
	15	182669	218131	-35462	-19.41
Germany	45	302656	306144	-3488	-1.15
	35	302656	303456	-800	-0.26
	25	302656	302712	-56	-0.02
	15	302656	304741	-2085	-0.69
Israel	45	13628	11979	1649	12.10
	35	13628	11329	2299	16.87
	25	13628	11096	2532	18.58
	15	13628	11961	1667	12.23

reason for this reduction is that these four countries had noticeable socioeconomic developments and safety programs during the last 20–25 years. Therefore, using the prediction models to forecast future road crashes based on the past short period (15 years) led to higher prediction accuracy than models with a longer period (25 years); however, the highest prediction accuracy was achieved using the 45-year models.

The results of forecast errors for 2017 showed that they followed a somewhat different trend, as shown in Figure 3. The prediction error generally decreased for models with 45-year sample size as compared to other sample sizes. The trend at sample sizes of 15–35 years was slightly different among the four countries. As for Germany, despite the different behavior at different sample sizes, the prediction accuracy for all sample sizes was very high (errors range: 0.02–1.1%). Furthermore, models at 15-year sample size generally had a relatively low prediction error.

As a general observation, the percentages of forecast errors increased for predicting crashes in year 2017 as compared to 2016 for almost all the sample sizes. Therefore, using ARIMA models for predicting crashes for a longer period of time (farther than the following year) would lead to a higher range of prediction errors as compared to the prediction error of the following year. Therefore, using ARIMA models for predicting crashes for a long period in the future could lead to lower prediction accuracy and thus should be used with caution.

## 5. Case of Palestine

In order to test the validity of the previous analysis, time-series modeling and verification were conducted for a country with high political and economic instability, Palestine. Its political and socioeconomic conditions were briefly presented before. Road crash time-series are presented in Figure 4. The road network in Palestine has been developing over the years. It ranged from a total length of



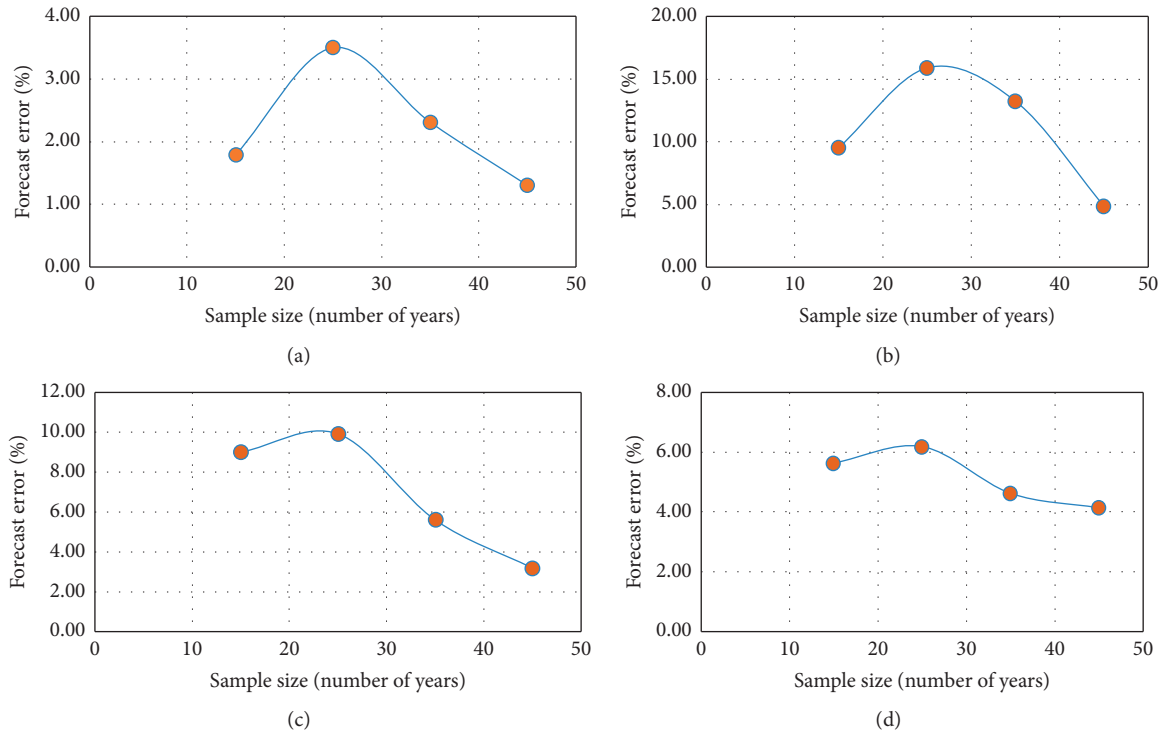


FIGURE 2: Forecast error for the subject countries, 2016. (a) Forecast error for Germany, 2016. (b) Forecast error for Denmark, 2016. (c) Forecast error for Israel, 2016. (d) Forecast error for Turkey, 2016.

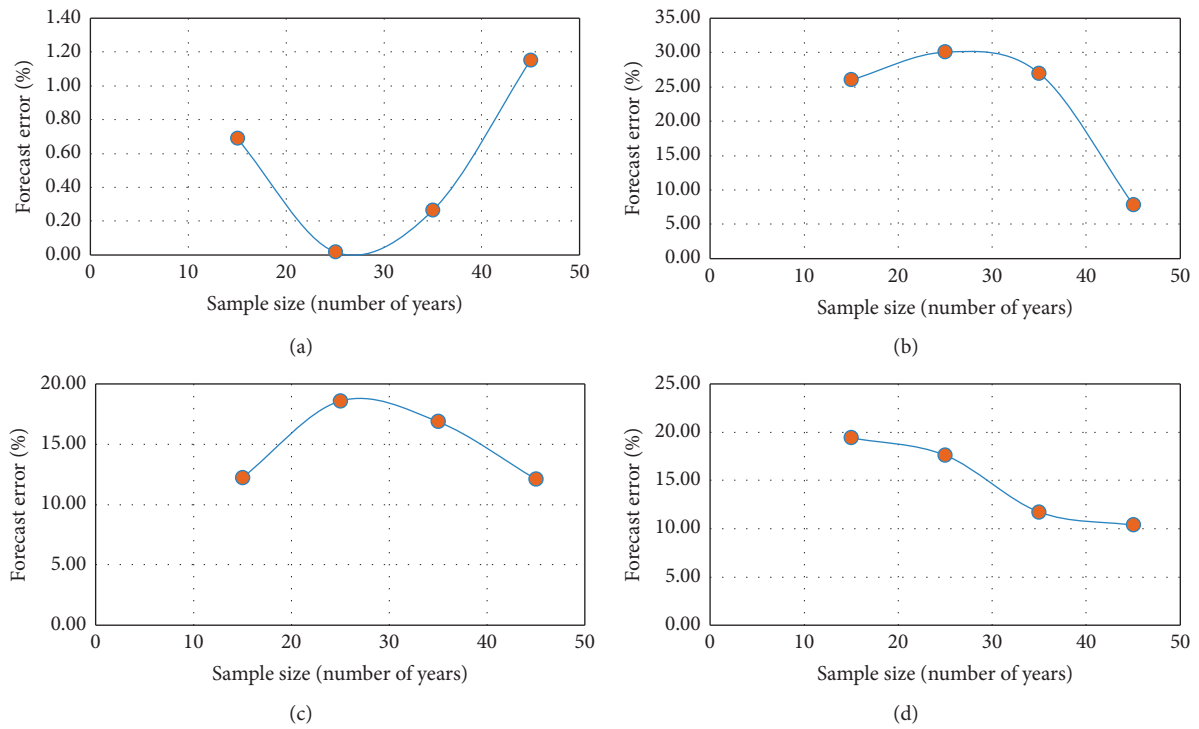


FIGURE 3: Forecast error for the subject countries, 2017. (a) Forecast error for Germany, 2017. (b) Forecast error for Denmark, 2017. (c) Forecast error for Israel, 2017. (d) Forecast error for Turkey, 2017.

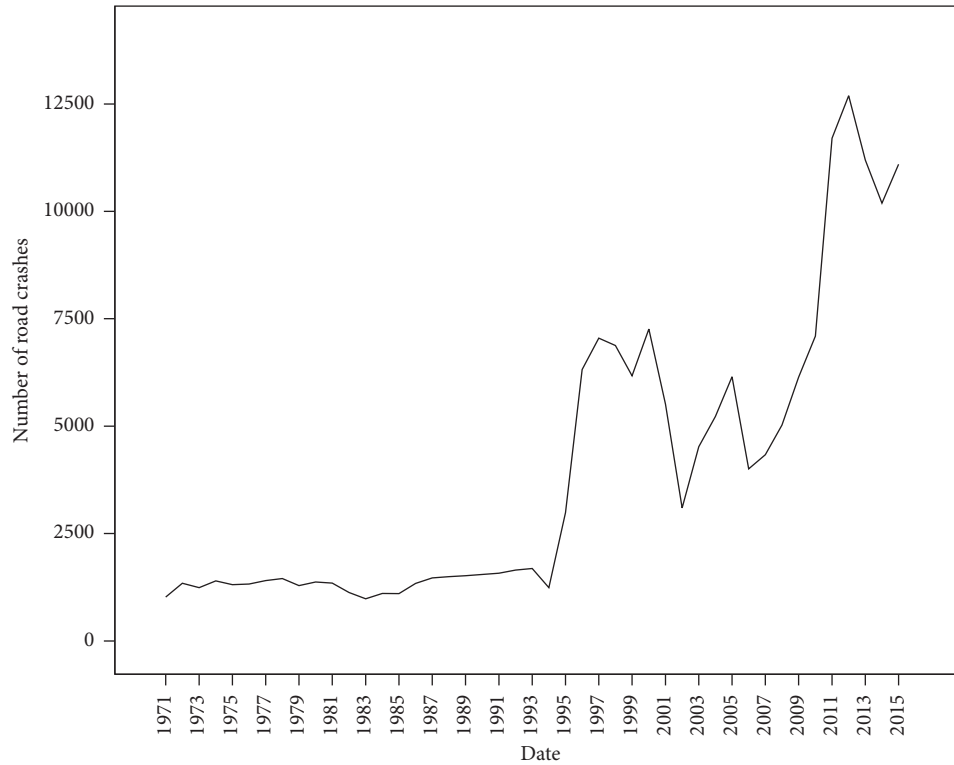


FIGURE 4: Road crash time-series for Palestine (1971–2015).

1576 km in year 1967 to 1716 km in 1997 [31]. The paved roadway widths ranged from 3.0 m to 6.0 m, with the vast majority being two-lane undivided highways. The length increased to 3501 km in 2010 [32] and up to approximately 5200 km in 2018 [33]. The paved roadway width increased up to 20 m of multilane divided highways.

The results of ARIMA models for road crashes in Palestine are presented in Table 7. It should be noted that the ARIMA method was not applicable at the sample sizes of 45 and 35 (for years before 1991) since the Shapiro–Wilk test showed that the residuals do not follow a normal distribution, which is one of the main conditions for applying ARIMA method. Therefore, results at 35- and 45-year sample size are not presented. As mentioned before, a new era in Palestine started in 1993 as a result of the Oslo Peace Accord, which significantly affected road crash patterns, as shown in Figure 4. Therefore, the ARIMA method could not model these abrupt changes and their consequences for road crashes in the period before 1991 [34]. The results of forecast errors for years 2016 and 2017 are presented in Table 8.

The results showed that predicting road crashes for models of sample sizes of 25 and 15 years produced somewhat close results for year 2016 (5.4% and 5.9% errors). More specifically, the prediction of road crashes for the 15-year sample size model produced slightly higher accuracy as compared to the prediction of road crashes for year 2017. The 15-year sample size model produced a higher accuracy with less forecast errors than a sample size of 25 years (3.7% compared to 6.0%), as shown in Figure 5.

TABLE 7: The selected best-fit models for each sample size for Palestine.

Country	Sample size	Best-fit ARIMA model	Model fit statistics		
			Stationary <i>R</i> -squared	<i>R</i> -squared	MAE
Palestine	—	—	—	—	—
	25	(3, 2, 3)	0.530	0.821	1092.853
	15	(3, 2, 3)	0.520	0.742	1229.727

TABLE 8: The forecast error of the best-fit models for Palestine for 2016 and 2017.

Year	Crashes	25 years (1991–2015)	15 years (2001–2015)
2016	Observed	11856	11856
	Forecast	12506	11152
	Error (%)	5.4	5.9
2017	Observed	12767	12767
	Forecast	12009	12299
	Error (%)	6.0	3.7

As a result, the prediction accuracies would vary based on the varying socioeconomic and development conditions of the country over the study years. For countries with steady and stable conditions, modeling using larger sample sizes (45 years or more) would yield higher accuracy models with higher prediction capabilities. As for countries with less steady and stable conditions, modeling using smaller sample

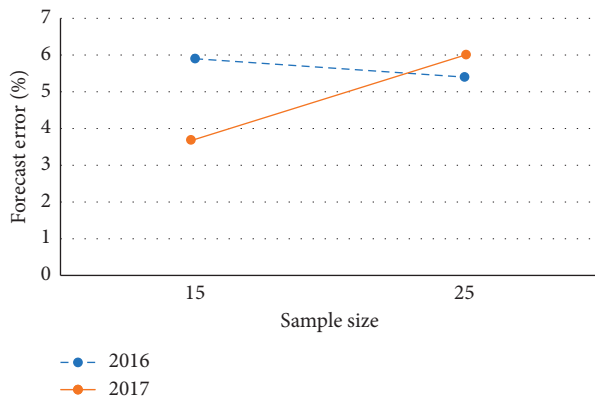


FIGURE 5: Forecast error for the subject countries.

sizes (15 years, for example) would lead to high accuracy models with good prediction capabilities.

## 6. Conclusions

During the last three decades, an important question has arisen, “What is the reasonable minimum sample size for appropriate time-series modeling?” However, there is a common rule-of-thumb in statistics that more is better. Practically this rule may not be ultimately accurate for all aspects of engineering. In traffic engineering, more specifically in road crash prediction, several interrelated factors may affect the prediction accuracy and the minimum sample size requirements. In this study, road crash prediction models were developed for four countries using different sample sizes, using ARIMA methodology. The analysis was further verified for a country with drastically changing socioeconomic and political conditions. Based on the results of this study, the following findings are offered.

- (i) It was verified that using larger sample size of 45 years, for road crash prediction, would generally lead to higher prediction accuracy as compared to 35, 25, and 15 years. On the other hand, the results generally showed that the prediction accuracy using a sample size of 15 years was higher than the accuracy for sample sizes of 25 and 35 years.
- (ii) In general, using ARIMA models to predict future road crashes for a longer period of time (farther than the following year) would lead to an increase in the prediction error, as compared to the prediction error of the following year; therefore, this should be done with caution.
- (iii) The socioeconomic and political factors as well as implemented safety programs for the study area are elements that affect the selection of the minimum acceptable sample size for road crash prediction. More specifically, in countries that have well-established and steady socioeconomic developments and an effective traffic safety program during the past few decades, using a sample size of 15 years could be acceptable, although a larger sample size of 45 is better. On the other hand, in countries with

nonsteady socioeconomic development or effective safety programs over the past few decades, using a smaller sample size, as small as 15 years, would be practical and would yield reasonably accurate results. Furthermore, crash data for an extended period of time might not always be available in those countries. However, the prediction capability of such models is also limited for only a few future years.

- (iv) The offered results are related to road crash prediction and might not be valid for the prediction in other fields. Therefore, different studies should be conducted separately for each field.
- (v) Using other prediction methods, other than ARIMA, requires specific studies in order to select the minimum acceptable sample size since each method of prediction depends on different statistical techniques and parameters.
- (vi) It is recommended for future work that crash data are tested for other countries with varying socioeconomic, political, and safety program conditions and that the conclusions of this research are confirmed. It is also recommended that smaller sample size intervals are used for modeling and testing of models’ prediction accuracies.

## Data Availability

All relevant data are included in the manuscript.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] R. J. Hyndman and A. V. Kostenko, “Minimum sample size requirements for seasonal forecasting models,” *Foresight: The International Journal of Applied Forecasting*, vol. 6, pp. 12–15, 2007.
- [2] F. M. A. Hassouna and M. Assad, “Towards a sustainable public transportation: replacing the conventional taxis by a hybrid taxi fleet in the West Bank, Palestine,” *International Journal of Environmental Research and Public Health*, vol. 17, no. 23, p. 8940, 2020.
- [3] F. B. Adebola, N. A. Adegoke, and R. A. Sanusi, “Consequences of road traffic accident in Nigeria: time series approach,” *International Journal of Computer Applications Technology and Research*, vol. 4, no. 4, pp. 262–273, 2015.
- [4] A. O. Atubi, “Time series and trend analysis of fatalities from road traffic accident in Lagos State, Nigeria,” *Mediterranean Journal of Social Sciences*, vol. 4, no. 1, 2013.
- [5] K. Mutangi, “Time series analysis of road traffic accidents in Zimbabwe,” *International Journal of Statistics and Applications*, vol. 5, no. 4, pp. 141–149, 2015.
- [6] R. K. Avuglah, K. A. Adu-Poku, and E. Harris, “Application of ARIMA models to road traffic accident cases in Ghana,” *International Journal of Statistics and Applications*, vol. 4, no. 5, pp. 233–239, 2014.

- [7] M. Parvareh, A. Karimi, S. Rezaei et al., "Assessment and prediction of road accident injuries trend using time-series models in Kurdistan," *Burns & Trauma*, vol. 6, no. 9, 2018.
- [8] F. Zolala, A. A. Haghdoost, T. Ahmadijouybari et al., "Forecasting the trend of traffic accident mortality in West Iran," *Health Scope*, vol. 5, no. 3, Article ID e31336, 2016.
- [9] S. Yousefzadeh-Chabok, F. Ranjbar-Taklimie, R. Malekpouri, and A. Razzaghi, "A time series model for assessing the trend and forecasting the road traffic accident mortality," *Archives of Trauma Research*, vol. 5, no. 3, Article ID e36570, 2016.
- [10] Y. P. Yuan, Z. Xu-jun, T. Zhi-bin, C. Meng-jing, and G. Yue, "Autoregressive integrated moving average model in predicting road traffic injury in China," *Zhonghua Liu Xing Bing Xue Za Zhi*, vol. 34, no. 7, pp. 736–739, 2013.
- [11] S. Rohayu, S. M. R. Sharifah Allyana, M. M. Jamilah, and S. V. Wong, "Predicting Malaysian road fatalities for year 2020," in *Proceedings of the Road Safety and Simulation International Conference (RSS2013)*, Rome, Italy, October 2013.
- [12] A. S. Al-Ghamdi, "Time series forecasts for traffic accidents, injuries, and fatalities in Saudi Arabia," *Journal of King Saud University-Engineering Sciences*, vol. 7, no. 2, pp. 199–217, 1995.
- [13] M. Al-Zyood, "Forecast car accident in Saudi Arabia with ARIMA models," *International Journal of Soft Computing and Engineering (IJSCE)*, vol. 7, no. 3, pp. 30–33, 2017.
- [14] B. Al-Omari, K. Ghuzlan, and H. Hasan, "Traffic accidents trends and characteristics in Jordan," *International Journal of Civil & Environmental Engineering IJCEE-IJENS*, vol. 13, no. 05, pp. 9–16, 2013.
- [15] K. Jadaan, I. Al-Hyari, H. Naghawi, R. Ammourah, and Z. A. Nabulsi, "Traffic safety in Jordan: magnitude, cost and potential countermeasures," *Journal of Traffic and Logistics Engineering*, vol. 1, no. 1, pp. 54–57, 2013.
- [16] P. Ljubič, L. Todorovski, and N. Lavrač, "Time-series analysis of UK traffic accident data," in *Proceedings of the Fifth International Multi-Conference Information Society*, pp. 131–134, Ljubljana, Slovenia, 2002.
- [17] K. Colum, "Time series analysis of UK traffic accident data using R," 2018, <http://columk.dbsdataproyects.com/2018/05/28/time-series-analysis-of-uk-traffic-accident-data-using-r/>.
- [18] H. Tang, V. V. Gayah, and E. T. Donnell, "Evaluating the predictive power of an SPF for two-lane rural roads with random parameters on out-of-sample observations," *Accident Analysis & Prevention*, vol. 132, p. 105275, 2019.
- [19] C. Liu and A. Sharma, "Exploring spatio-temporal effects in traffic crash trend analysis," *Analytic Methods in Accident Research*, vol. 16, pp. 104–116, 2017.
- [20] M. Freeman and W. Leith, "Estimating the number of traffic crash-related cervical spine injuries in the United States; an analysis and comparison of national crash and hospital data," *Accident Analysis and Prevention*, vol. 142, Article ID 105571, 2020.
- [21] F. Russo, M. Busiello, and G. Dell Acqua, "Safety performance functions for crash severity on undivided rural roads," *Accident Analysis & Prevention*, vol. 93, pp. 75–91, 2016.
- [22] F. Russo, S. Fric, S. A. Biancardo, and D. Gavran, "Driver speed behavior on circular curves of undivided two-lane rural roads," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2472, no. 1, pp. 117–128, 2015.
- [23] F. Russo, S. A. Biancardo, and M. Busiello, "Operating speed as a key factor in studying the driver behaviour in a rural context," *Transport*, vol. 31, no. 2, pp. 260–270, 2016.
- [24] OCED Road, *Accidents*, OCED Publication, Paris, France, 2019, <https://data.oecd.org/transport/road-accidents.htm>.
- [25] M. A. Din, "ARIMA by Box Jenkins methodology for estimation and forecasting models in higher education," in *Proceedings of the ATINER's Conference Paper Series EMS015-1846*, Athens, Greece, January 2015.
- [26] F. M. A. Hassouna and I. Pringle, "Analysis and prediction of crash fatalities in Australia," *The Open Transportation Journal*, vol. 13, no. 1, pp. 134–140, 2019.
- [27] F. M. A. Hassouna and K. Al-Sahili, "Environmental impact assessment of the transportation sector and hybrid vehicle implications in Palestine," *Sustainability*, vol. 12, no. 19, p. 7878, 2020.
- [28] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, OTexts, Heathmont, Australia, 2nd edition, 2018.
- [29] G. Box and G. Jenkins, *Series Analysis: Forecasting and Control*, Wiley & Sons, San Francisco, CA, USA, 4th edition, 1976.
- [30] S. G. Makridakis, S. C. Wheelwright, and R. J. Hyndman, *Forecasting: Methods and Applications*, J. Wiley, New York, NY, USA, 1998.
- [31] PCBS, *Transportation and Communications Statistics in the West Bank and Gaza, 1998*, The Palestinian Central Bureau of Statistics, Ramallah, Palestine, 1999.
- [32] PCBS, *Transportation and Communications Statistics in the West Bank and Gaza, Annual Report 2010*, The Palestinian Central Bureau of Statistics, Ramallah, Palestine, 2011.
- [33] PCBS, *Transportation and Communications Statistics in the West Bank and Gaza, Annual Report 2018*, The Palestinian Central Bureau of Statistics, Ramallah, Palestine, 2019.
- [34] F. M. A. Hassouna, S. Abu-Eisheh, and K. Al-Sahili, "Analysis and modeling of road crash trends in Palestine," *Arabian Journal for Science and Engineering*, vol. 45, no. 10, pp. 8515–8527, 2020.