Zanco Journal of Human Sciences (ZJHS)
ISSN (Print):2218-0222
ISSN (Online): 2412-396X

Vol.28 Issue 1 2024
ID No.1442 (PP 278 -289)

https://doi.org/10.21271/zjhs.28.1.18

Research Article

Building a Statistical Model to Forecast Traffic Accidents for Death and Injuries by Using Bivariate Time Series Analysis

Nozad Hussen Mahmood*
Dler Hussen Kadir **
Obaid Mahmud Mohsin Alzawbaee*

* Department of Business Administration, Cihan University Sulaimaniya, 46001, Kurdistan region, Iraq
** Department of Statistics and Informatics, College of Administration and Economics, Salahaddin University-Erbil
** Kurdistan Region, Iraq
Department of Business Administration, Cihan University-Erbil, Kurdistan Region, Iraq
nozad.mahmood@sulicihan.edu.krd
dler.kadir@su.edu.krd
Obed.muhsin@sulicihan.edu.krd

Reiceved ‖ 26/07/2023
Accepted ‖ 10/09/2023
Published ‖ 15/02/2024

Keywords:
Multivariate time series,
VARMA (p, q),
Forecasting,
Traffic Accident.

Abstract

In our study, multivariate time series were used that included two variables, namely, the death and injury rates from car accidents in Erbil City Iraq. The data for the two series were collected monthly from January 2015 to December 2020, so there are 72 units in each series. The most important finding is that the time series is stationary, and the appropriate model to represent the phenomenon studied is VARMA (1,0). A statistical model was adopted to forecast accidents resulting in death and injuries for 2024, and it was found to be appropriate. Furthermore, we use R-programing and STATA version 17 to analyze our data. As a result, the study suggested that the Iraqi Kurdistan Traffic Department could use the model developed to forecast the phenomenon's future trends.

About the Journal
ZANCO Journal of Humanity Sciences (ZJHS) is an international, multi-disciplinary, peer-reviewed, double-blind and open-access journal that enhances research in all fields of basic and applied sciences through the publication of high-quality articles that describe significant and novel works; and advance knowledge in a diversity of scientific fields.
https://zancojournal.su.edu.krd/index.php/JAHS/about

1. Introduction

According to the work of Farag, Hashim, and others, road traffic accidents are chance occurrences that result in injury, death, and damage (Farag et al., 2014). The World Health Organization (WHO) has found that car accidents cause the most deaths among people aged 5 to 29 worldwide, and in 2020, 10726 people died in car accidents in Iraq, which is 7.32 percent of all deaths. The average death rate is 34.41 per 100000 population, and Iraq is 37th in the world (WHO, 2020; Getahun, 2021).

Accidents on the road can happen anywhere in the world, but the specific things that cause them to happen vary a lot from place to place. There are some factors for car accidents, such as driver or passenger mistakes, fast driving, and road and highway conditions, which are just a few of the causes of car accidents. Human factors contribute to road traffic accidents, including driver attitude, road etiquette, driving under the influence of substances, especially alcohol, driver gender, seat belt use, and driver age (especially in teenage and elderly drivers) (Smart & Mann, 2002; Bjerre et al., 2008). To the tune of 20–50 million people, car accident-related injuries and disabilities are a global problem. It predicts that by 2020, traffic accidents will rank third among all causes of death (Peden, 2005).

Multivariate time series is a set of measurements taken from several variables arranged according to the time of their occurrence. In most cases, these periods are regular. The time series analysis consists of sequential steps that begin with the Identification step of the model, followed by the step of estimation of the model's parameters, and then the step of Diagnostic Checking for the model to come to the last step, which is forecasting (ALN, 2019 and Doornik & Hansen, 2008).

A multivariate time-series study described and predicted Kurdistan Province vehicle accident injuries. From March 2009 to February 2015, the Box-Jenkins time-series analysis used an autoregressive integrated moving average (ARIMA) and a seasonal ARIMA to model injury observations and predict accidents up to 24 months in the future. ARIMA (1, 0, 0) and SARIMA (1, 1, 1) (0, 0, 1) were the best models for car occupant and motorcyclist injuries (Parvareh et al., 2018).

A multivariate time series were used to examine how road safety standards have changed. A model has been made to evaluate the effect of research methodology on the number of injuries and deaths caused by car accidents. The results show that lowering blood alcohol concentration levels indirectly affects the number of people who die in car accidents. In contrast, seatbelt legislation and the number of miles driven have direct effects (Chamlin & Sanders, 2018).

The multivariate time series techniques were used to analyze contemporaneous relationships and dynamic interactions among police enforcement, traffic violations, and traffic crashes. A vector autoregressive (VAR) model was applied, and the results showed that traffic accidents and violations changed weekly and were significantly affected by holidays and the weather (Feng et al., 2020).

A study was conducted based on the multivariate time series to examine the numerous accident situations, which demonstrate interdependencies with each other and their impact on the occurrence of accidents. Multivariate forecasting was also utilized to show how accidents might get worse in different areas in the future. An ARIMA method extension was used to calculate forecast values (Meibner & Rieck, 2021).

The primary goals of this study are:

1. Explain model selection for time series analysis by explaining and comparing approaches for creating VARMA models in a stationary situation.
2. Summarize recent methodological developments for simplifying VARMA model identification and estimation in the literature.

3. To evaluate the model performance and the difference between the predicted and true values of every car accident.

The study's significance lies in the VARMA (p, q) model's ability to predict traffic accident deaths and injuries. Furthermore, using multivariate time series analysis to construct a model for forecasting traffic accidents facilitates the authorities in developing plans to achieve traffic safety.

2. Multivariate time series analysis

A multivariate time series is a list of measurements from different variables ordered by when they happened. The length of these periods depends on what is being studied. The identification stage is the first step that makes up multivariate time series analysis. Next comes the estimation of model parameters; the model's health is checked through diagnostics, and predictions are made (Engle, 1982, Mohammed et al. 2020).

The goal of multivariate time series analysis is to accurately describe how the process that creates time series works and to explain how time series behave by building a statistical model that fits the phenomenon being studied and forecasting how time series will behave in the future based on how they behaved in the past. The series is said to be stationary if the probabilistic properties do not change over time. If the probabilistic properties change over time, the series is considered non-stationary (Elliott et al., 1992, Al-Zawbaee and Mahmood, 2023).

That is, the time series is stationary if the arithmetic mean and variance of the time series are constant, and the failure of either of the previous two conditions leads to the impossibility of analyzing the time series, and therefore it must be addressed first. Among the methods used to establish the stationarity of time series are the following:

The differences between the nonstationary series are used to turn a mean nonstationary time series into a stationary series. Critical transformations are used to deal with the nonstationary variance and make it stationary (Alzawbaee et al., 2020)

The Dickey-Fuller test (ADF) and the Phelps-Perron test (PP) are used to determine whether the series is stationary. The following hypothesis tests for each of the tests (ADF) and (PP) (Dickey & Fuller, 1981, Perone, 2020, Smith et al., 2003).

$$H_0: Nonstationary\ time\ series$$
$$H_1: Stationary\ time\ series$$

If the null hypothesis H_0 is accepted, then the time series under investigation is not stationary, and more testing is required after appropriate adjustments have been made. If H0 is not accepted, then the time series is stationary.

The Augmented Dickey-Fuller Test (ADF) Formula:

$$\nabla Zt = b_1 z_{t-1} + \sum_{j=1}^{k} \alpha_j \Delta z_{t-j+} a_t \tag{1}$$

where:

ΔZ_t: represents the first difference of the variable Z at time t.

b_1: is the coefficient associated with the lagged level of Zt-1.

Zt-1: represents the level of the variable Z at the previous time period, t-1

k: is the number of lags.

$\Delta Z_{t-1} = Z_{t-1} - Z_{t-2}$, $\Delta Z_{t-2} = Z_{t-2} - Z_{t-3}$, and so on.

$\sum_{j=1}^{k} \alpha_j \Delta z_{t-j+} a_t$ are the additional autoregressive terms.

a_t: Random error term

The Philips-Perron Test. It is an extension of the Dickey-Fuller test, which considers the heterogeneity of error variance for the residuals. We find that Phillips and Perron developed a generalization of the Dickey Fuller method, and its mathematical formula:

$$\Delta Z_t = \emptyset + (\rho - 1)Z_{t-1} + \gamma\left(t - \frac{T}{2}\right) + \psi\Delta Z_{t-1} + \varepsilon_t \qquad (2)$$

Where:

ΔZ_t: represents the first difference of the variable Z at time t.

$\emptyset$: represents the constant term or intercept.

ρ: is the autoregressive coefficient, which indicates the relationship between Zt and Zt-1.

γ: represents the coefficient of the deterministic trend.

t: represents the time index or period.

T: Sample Size

ψ: is the coefficient associated with the lagged first difference of Z.

ΔZ_{t-1}: represents the lagged first difference of the variable Z.

ε_t: random error term.

3. Multivariate Time Series Models

In general, multivariate time series models include Vector Autoregressive models VAR(p), Vector Moving Average Models VAM(q), and Mixed Models (Vector Autoregressive-Moving Average Models) VARMA (p, q). The multivariate time series analysis process goes through the steps of Identification, Estimation, Diagnostic Checking, and Forecasting.

The identification step is the most important step for the analysis of time series and is represented by knowing the type of the model and determining the rank by applying the model order criteria. The type of the model determines through the behavior of ACF (autocorrelation function) and PACF (partial autocorrelation function), we determine initially whether the model is VAR(p), VMA(q), or VARMA (p, q).

But for some cases, the model cannot be accurately determined by drawing (ACF) and (PACF). We can take all possible arrangements of the model by giving (q, p) on the order of values (0,1,2) for the stationary time series, so we have 8 Forms: (Dickey & Fuller, 1981, Alzawbaee et al., 2020 and Kadir, 2020).

VARMA (p, q): All state, take p=3 and q=3

In the case that the series is nonstationary, we take the differences (d) and the number of differences required d = 1,2, and 3 to convert the nonstationary time series into stationary series.

Where (p) is the autoregressive rank, (d) is the number of differences (integration), and (q) is the moving average rank. And every time we test each one of the models, by applying the hierarchies of the eight possible models and applying the model order criteria of the models, and finding the (MSE) mean squares error values, the correct and efficient model is selected among the models, which is adopted for forecasting purposes.

In a recent study, researchers provided a proposed approach for identifying effective models for time series, and they evaluated the method and process for doing so using a criterion called mean squares error, which works well for both univariate and multivariate time series (Mahmood & Al-Takriti, 2013; Ding et al., 2020). There are several criteria for choosing the rank of the model, including:

Akaike Information criterion. It is denoted by the symbol AIC and is given in the following form (ALN, 2019 and Khidir et al., 2023)

$$\text{AIC} = -2\ \log\ (L\ )\ +2\ k$$

Where:
L: is the maximized value of the likelihood function of the model.
K: is the number of parameters estimated.

Then the formula for the Akaike Information Criterion (AIC) that is used to correct for small sample sizes is:

$$AIC_c = AIC + \frac{2*k*(k+1)}{n-k-1} \qquad (3)$$

Where:
k: is the number of parameters in the model.
n: is the number of observations.

1. Schwartz Bayesian Information criterion:

It is denoted by the symbol (BIC) and its formula is as follows:

$$\text{BIC} = -2\ \log\ (L\ )\ +k \ln\ (n)$$

Where:
L: is the maximized value of the likelihood function of the model.
k: is the number of parameters estimated.
n: is the number of observations.

Then with gaussian special case

$$BIC = \frac{\log|\hat{\Sigma}|+r\ \log(T)}{T} \qquad (4)$$

Where:
$\hat{\Sigma}$: Covariance matrix of the estimated residuals.
r: is the number of parameters in the model.
T: is the number of observations.

Then, after the identification step, we go to the estimation step. The least squares method is one way to estimate model parameters and build multivariate time series models. After the estimation step, we test the model (the diagnostic checking step). In this step, the model is tested to see how well it fits the data of the studied phenomenon and how well it can be used to make predictions. Many tests can be used for this purpose. One of these is the Box-Pierce test, symbolized by the symbol (Q), as it relies on the equality of the autocorrelation of the estimated residuals to zero. It indicates testing the null hypothesis based on the autocorrelation of the estimated residuals according to the following formula (ALN, 2019 and Mahmood & Al-Takriti, 2013).

$$Q = n\sum_{k=1}^{L} r_k^2(a) \sim \ \chi^2{}_{((L-m),\alpha)} \qquad (5)$$

Where:
L: Number of offsets, and m: number of estimated parameters.

If the value of (Q) is smaller than χ^2 tabular, it accepts the null hypothesis and concludes that the autocorrelations of the estimated errors are not significant, which indicates that the residual estimations of the estimated correlations are random and distributed independently, which confirms that the model is good and appropriate. Then the last step is the forecasting step, after the model identification, its parameters are estimated and diagnostic checking, it is used to forecast the future values of the series, where the behavior of the studied phenomenon is known in the future (Fanelli & Piazza, 2020, He & Tao, 2018 and Ljung & Box, 1978).

4. Methodology

The essential goal of this research is to build an efficient multivariate time series model for traffic accidents in the Kurdistan region of Iraq. For that purpose, the data needed for this study was obtained from the Erbil General Directorate of Traffic, the capital of the Kurdistan Region, which has been keeping track of the number of people killed or injured in traffic accidents each month for six years, starting from January 2015 to December 2020. In other words, the number of observations or sample size of the study is 72 months. Thus, we have two different variables (series):

$\{Z_1\}$ = Death is a monthly time series of people killed in traffic accidents from 2015 to 2020.
$\{Z_2\}$ = Injure is a monthly time series of people injured in traffic accidents from 2015 to 2020

Furthermore, for the analysis of our data, we use R-programing and STATA version 17.

5. Results

Compute the correlation between two variables in the statistical analysis of time series to test the stationary of the time series in question. This is done through drawing the series, as well as the ADF and Phillips-Peron test: The following figure shows the behavior of the time series of traffic accidents (death, injuries) combined.

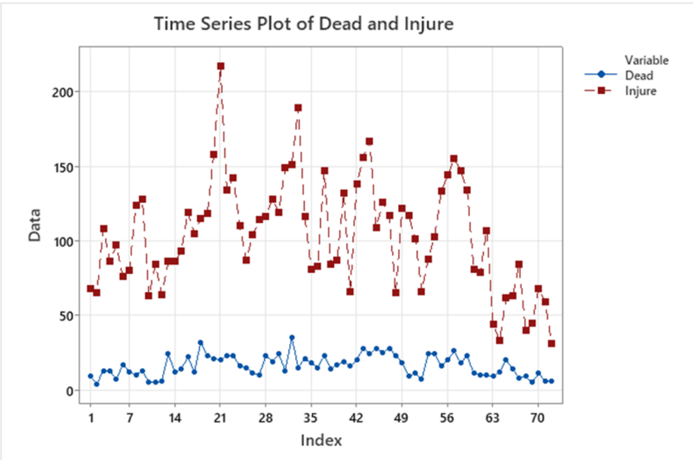


Figure 1. The number of deaths and injuries over time

Since it is not possible to definitively determine the extent of the stationary of the series, we apply the Dickey-Feller test, and the Phillips-Peron test to test the stationary of the series, and the results are shown in the following table 1, bearing in mind that the hypothesis is:

H_0: *Serieas are non − stationary*

H_1: *Serieas are non − stationary*

Table 1. ADF and the Phillips-Peron tests for stationary of the series

Stationary Tests	Significance level			Test Statistics	P-value
	0.01	0.05	0.1		
Dickey-Fuller	-4.56221	-3.92817	-3.61198	-5.13212	0.001
Phillips-Perron	-19.278	-13.468	-10.826	-43.082	0.000

It is clear from Table 1 that the absolute values of the test (DF&PP) are greater than the critical values at the significance level (0.01), so we reject the null hypothesis (H_0) which states that the time series are nonstationary, and we accept the alternative hypothesis (H_1), which indicates that the time series in question are stationary. After ensuring the stationary of the time series, the steps of time series analysis are applied, and we start with the identification step, where all possible possibilities of the expected models were taken, and as explained in the theoretical side, where the models were taken, VARMA(p, q), (p, q) = 0, 1,2, and the model order criteria for each model were found, as well as finding (Weights, MSE), as shown in the following table 2:

Table 2. Possible models for the time series in question

Models VARMA (p, q)	RMSE	AIC	SBC	MAPE	MAE	Weights
(1,0)	19.061*	630.708*	635.262*	16.789*	15.451*	0.413*
(2,0)	19.129	632.261	639.091	16.827	15.458	0.19
(1, 1)	19.137	632.314	639.144	16.809	15.457	0.185
(1, 2)	19.234	633.969	643.076	16.971	15.458	0.081
(2, 1)	19.24	633.979	643.085	16.832	15.557	0.081
(2, 2)	19.267	634.959	646.343	17.072	15.573	0.049
(0,2)	20.998	645.506	652.336	19.012	16.881	0
(0,1)	22.163	652.422	656.975	20.729	18.204	0

The results of Table 2 illustrate that the VARMA (1,0) model has the lowest values based on the criteria of RMSE = 19.061, AIC = 630.708, SBC = 635.262, MAPE = 16.789, and MAE = 15. 45. Conversely, VARMA (1,0) has the most significant effect because it shows the highest weights = 0.413 compared to other models. Thus, it is the most efficient model to analyze our data.

After identification the models and choosing the appropriate model, the second step of the time series analysis was applied, which is the estimation step, where the model parameters were estimated, and the following table (3) shows the estimations of the model parameters:

Table 3. Estimation of the parameters of the VARMA (1,0) model.

Dependent Variable	Constant	Death	Injure
Z1t	5.744	0.259	0.06
Z2t	34.993	1.561	0.417

Where the estimated model is:

$$\begin{bmatrix} Z_{1t} \\ Z_{2t} \end{bmatrix} = \begin{bmatrix} 5.744 \\ 34.993 \end{bmatrix} + \begin{bmatrix} 0.259 & 0.06 \\ 1.561 & 0.417 \end{bmatrix} \begin{bmatrix} Z_{1t-1} \\ Z_{2t-1} \end{bmatrix} + \begin{bmatrix} \hat{a}_{1t} \\ \hat{a}_{2t} \end{bmatrix} \ldots \quad (6)$$

The model for the death series rate due to the traffic accident series is below.

$$Z_{1t} = 5.744 + 0.259\, Z_{1,t-1} + 0.06\, Z_{2,t-1} + a_{1t} \ldots \quad (7)$$

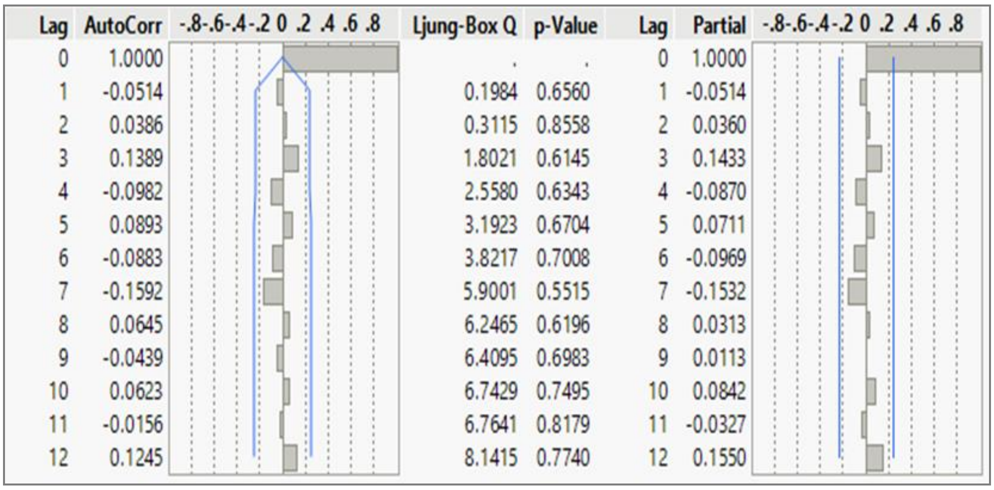
The model for the injures series rate due to the traffic accident series is below.

$$Z_{1t} = 34.993 + 1.561\, Z_{1,t-1} + 0.417\, Z_{2,t-1} + a_{2t} \quad (8)$$

After selecting the model and estimating the parameters, the diagnostic checking of the model was tested. We look at the diagnostic tests to see if the idea that residuals are just random noise is actually true. For that purpose, we use the Ljung-Box (Q) Test for checking within each lag and the Portmanteau test for testing the whole model with the following hypothesis.

H_0: The correlations between the estimated residuals are independently distributed.

H_1: The correlations between the estimated residuals are not independently distributed.



Lag	AutoCorr	Ljung-Box Q	p-Value	Lag	Partial
0	1.0000	.	.	0	1.0000
1	-0.0514	0.1984	0.6560	1	-0.0514
2	0.0386	0.3115	0.8558	2	0.0360
3	0.1389	1.8021	0.6145	3	0.1433
4	-0.0982	2.5580	0.6343	4	-0.0870
5	0.0893	3.1923	0.6704	5	0.0711
6	-0.0883	3.8217	0.7008	6	-0.0969
7	-0.1592	5.9001	0.5515	7	-0.1532
8	0.0645	6.2465	0.6196	8	0.0313
9	-0.0439	6.4095	0.6983	9	0.0113
10	0.0623	6.7429	0.7495	10	0.0842
11	-0.0156	6.7641	0.8179	11	-0.0327
12	0.1245	8.1415	0.7740	12	0.1550

Figure 2. Ljung-Box Q test for ACF and PCF according to each lag.

Table 4. Portmanteau Q test for white noise

Portmanteau (Q) statistic	p-Value
44.5524	0.1064

Figure 2 shows that for all the lags, all the p-values for the Ljung-Box Q test are more than 0.05, meaning there is no reason to reject the null hypothesis. In other words, the estimated residuals are not correlated with each other over time. It shows that the residuals in our time series model are independent, which is one of the assumptions for the fitting model. On the other hand, the portmanteau Q test in Table 4 illustrated that the p-value is insignificant (0.1064), so we cannot reject the null hypothesis.

The proposed model for forecasting was adopted, as it was forecasting for the four years of (2021, 2022, 2023, 2024), as follows:

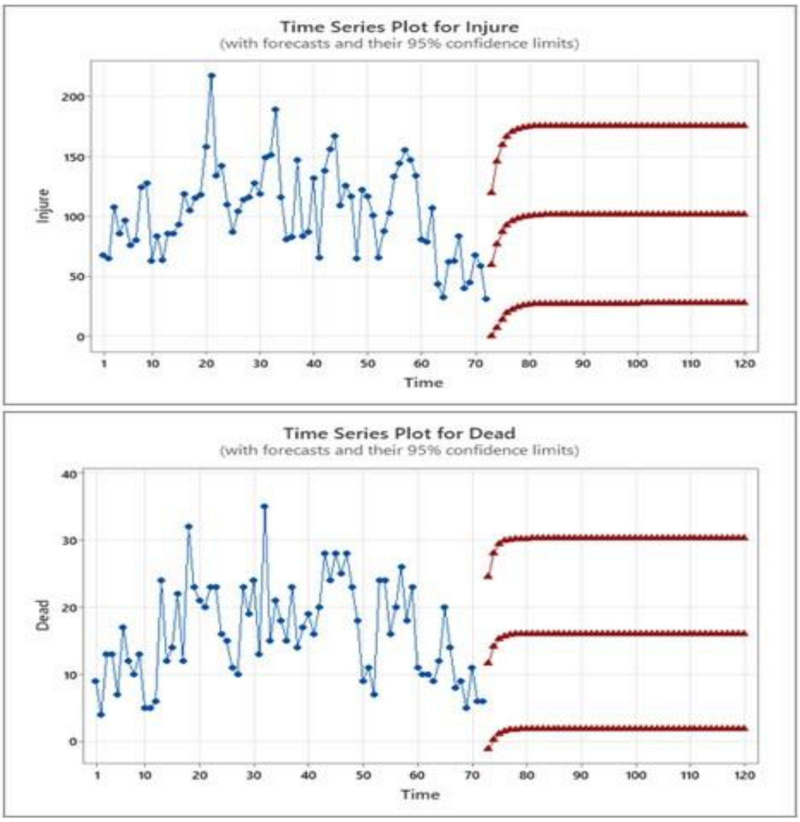


Figure 3. Forecasting for dead and injure for the next twelve months.

Table 5. Forecasting traffic accidents for the next four years

Time Period	Injure	Death	Time Period	Injure	Death	Time Period	Injure	Death	Time Period	Injure	Death
Jan-21	60.275	11.736	Jan-22	101.542	16.031	Jan-23	101.609	16.031	Jan-24	101.609	16.031
Feb-21	77.412	14.192	Feb-22	101.570	16.031	Feb-23	101.609	16.031	Feb-24	101.609	16.031
Mar-21	87.445	15.244	Mar-22	101.586	16.031	Mar-23	101.609	16.031	Mar-24	101.609	16.031
Apr-21	93.317	15.694	Apr-22	101.596	16.031	Apr-23	101.609	16.031	Apr-24	101.609	16.031
May-21	96.755	15.887	May-22	101.601	16.031	May-23	101.609	16.031	May-24	101.609	16.031
Jun-21	98.768	15.969	Jun-22	101.604	16.031	Jun-23	101.609	16.031	Jun-24	101.609	16.031
Jul-21	99.946	16.004	Jul-22	101.606	16.031	Jul-23	101.609	16.031	Jul-24	101.609	16.031
Aug-21	100.635	16.020	Aug-22	101.607	16.031	Aug-23	101.609	16.031	Aug-24	101.609	16.031

Sep-21	101.039	16.026	Sep-22	101.608	16.031	Sep-23	101.609	16.031	Sep-24	101.609	16.031
Oct-21	101.275	16.029	Oct-22	101.609	16.031	Oct-23	101.609	16.031	Oct-24	101.609	16.031
Nov-21	101.414	16.030	Nov-22	101.609	16.031	Nov-23	101.609	16.031	Nov-24	101.609	16.031
Dec-21	101.495	16.031	Dec-22	101.609	16.031	Dec-23	101.609	16.031	Dec-24	101.609	16.031

Table 5 and Figure 3 forecast the number of deaths and injuries for the next four years (2021, 2022, 2023, and 2024) due to traffic accidents, and the results show that the number of deaths and injuries will increase over time. In such a way that for the next year (2024), there will be around 102 injured and 16 dead people in Erbil (the capital of the Kurdistan region of Iraq) every month caused by car accidents.

6. Conclusions and Recommendations:

The most important conclusions and recommendations reached by the research are the following:

1- The appropriate and efficient model to represent the bivariate time series of traffic accidents (deaths and injuries) is VARMA (1, 0).
2- The model that was built to forecast traffic accidents (deaths and injuries) has been adopted for the four years (2021, 2022, 2023, and 2024). The findings show a remarkable convergence between the estimated and actual behavior of the series, which supports the model's efficiency and reliability in forecasting future values.
3- This approach can help determine the phenomenon's future trends and develop the necessary plans for the General Directorate of Traffic in the Kurdistan Region.

7. References:

- ALN. A. 2019. Using multivariate time series analysis to study highway traffic accidents in Sudan.. PhD Thesis, University of Science and Technology, Sudan.
- Al-Zawbaee, O.M. and Mahmood, N.H., 2023. Building an Efficient Model by Using Panel Data. Cihan University-Erbil Journal of Humanities and Social Sciences, 7(1), pp.82-87.
- Alzawbaee, O. M., Hamad, A. P. D. A. A., & Mahmood, A. L. A. O. 2020. Using Time Series Analysis to Forecast the prevalence of the COVID-19 Pandemic (Iraq and Gulf Cooperation Council countries (GCC)).
- BJERRE, J., KIRKEBJERG, P. G. & LARSEN, L. B. 2006. Prevention of traffic deaths in accidents involving motor vehicles. Ugeskrift for Læger, 168, 1764-1768.
- CHAMLIN, M. B. & SANDERS, B. A. 2018. Social policy and crash fatalities: a multivariate time series analysis. Journal of crime and justice, 41, 322-333.
- DICKEY, D. A. & FULLER, W. A. 1981. Likelihood ratio statistics for autoregressive time series with a unit root. Econometrica: journal of the Econometric Society, 1057-1072.
- DING, G., LI, X., JIAO, F. & SHEN, Y. 2020. Brief Analysis of the ARIMA model on the COVID-19 in Italy. medRxiv.
- DOORNIK, J. A. & HANSEN, H. 2008. An omnibus test for univariate and multivariate normality. Oxford bulletin of economics and statistics, 70, 927-939.
- ELLIOTT, G., ROTHENBERG, T. J. & STOCK, J. H. 1992. Efficient tests for an autoregressive unit root. National Bureau of Economic Research Cambridge, Mass., USA.
- ENGLE, R. F. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica: Journal of the econometric society, 987-1007.
- FANELLI, D. & PIAZZA, F. 2020. Analysis and forecast of COVID-19 spreading in China, Italy and France. Chaos, Solitons & Fractals, 134, 109761.
- FARAG, S. G., HASHIM, I. H. & EL-HAMRAWY, S. A. 2014. Analysis and assessment of accident characteristics: Case study of Dhofar Governorate, Sultanate of Oman. Int J Traffic Trans Eng, 3, 189-98.
- FENG, M., WANG, X. & QUDDUS, M. 2020. Developing multivariate time series models to examine the interrelations between police enforcement, traffic violations, and traffic crashes. Analytic methods in accident research, 28, 100139.
- GETAHUN, K. A. 2021. Time series modeling of road traffic accidents in Amhara Region. Journal of big data, 8, 1-15.

- Global status report on road safety. Accident: WHO; 2020.

- https://www.worldlifeexpectancy.com/iraq-road-traffic-accidents

- HE, Z. & TAO, H. 2018. Epidemiology and ARIMA model of positive-rate of influenza viruses among children in Wuhan, China: A nine-year retrospective study. International Journal of Infectious Diseases, 74, 61-70.

- KADIR, D. 2020. Time Series Modeling to Forecast on Consuming Electricity: A case study Analysis of electrical consumption in Erbil City from 2014 to 2018. Journal of Al-Rafidain University College For Sciences (Print ISSN: 1681-6870, Online ISSN: 2790-2293), 472-485.

- Khidir, H.A., Etikan, İ., Kadir, D.H., Mahmood, N.H. and Sabetvand, R., 2023. Bayesian machine learning analysis with Markov Chain Monte Carlo techniques for assessing characteristics and risk factors of Covid-19 in Erbil City-Iraq 2020–2021. Alexandria Engineering Journal, 78, pp.162-174.

- LJUNG, G. M. & BOX, G. E. 1978. On a measure of lack of fit in time series models. Biometrika, 65, 297-303.

- Mahmood, A. L. A. O & AL-Takriti, M. S. 2013. A New Suggested Technique to Determine the Best Models of Time Series. Damascus University Journal for Basic Sciences, V (29).

- MEIßNER, K. & RIECK, J. 2021. Multivariate Forecasting of Road Accidents Based on Geographically Separated Data. Vietnam Journal of Computer Science, 8, 433-454.

- MOHAMMED, P. A., OBED, S. A., ALI, I. M. & KADIR, D. H. 2022. US Dollar/IQ Dinar Currency Exchange Rates Time Series Forecasting Using ARIMA Model. Cihan University-Erbil Scientific Journal, 6, 12-19.

- PARVAREH, M., KARIMI, A., REZAEI, S., WOLDEMICHAEL, A., NILI, S., NOURI, B. & NASAB, N. 2018. Assessment and prediction of road accident injuries trend using time-series models in Kurdistan', Burns & Trauma 6 (1), 9.PEDEN, M. 2005. Global collaboration on road traffic injury prevention. International journal of injury control and safety promotion, 12, 85-91.

- PERONE, G. 2020. An ARIMA model to forecast the spread and the final size of COVID-2019 epidemic in Italy. MedRxiv.

- SMART, R. G. & MANN, R. E. 2002. Deaths and injuries from road rage: cases in Canadian newspapers. CMAJ, 167, 761-762.

- SMITH, R. L., WIGLEY, T. M. & SANTER, B. D. 2003. A bivariate time series approach to anthropogenic trend detection in hemispheric mean temperatures. Journal of Climate, 16, 1228-1240.

دروستکردنی مۆدێلێکی ئاماری بۆ پێشبینیکردنی ڕووداوەکانی هاتوچۆ بۆ مردن و برینداربوون بە بەکارهێنانی زنجیرە کاتیەکانی دوانەی گۆڕاو
Bivariate Time Series Analysis

نەوزاد حسین محمود
بەشی کارگێڕی کار، زانکۆی جیهان-سلێمانی
nozad.mahmood@sulicihan.edu.krd

دلێر حسین قادر
بەشی ئامارو زانیاری، کۆلێژی بەڕێوەبردن و ئابوری، زانکۆی سەلاحەدین-هەولیر
بەشی کارگێڕی کار، زانکۆی جیهان هەولێر
dler.kadir@su.edu.krd

عوبید محمود محسن
بەشی کارگێڕی کار، زانکۆی جیهان-سلێمانی
obed.muhsin@sulicihan.edu.krd

پوختە

داتاکانی ئەم توێژینەوەیە بریتیبوو لە دیاردەی مردن و برینداربوون بە هۆی ڕووداوەکانی ئۆتۆمبێلەوە لە مانگی یەکی ساڵی ٢٠١٥ تا کانوونی دووەمی ٢٠٢٠، لە لێکۆڵینەوەکەماندا دەرکەوت کە زنجیرە کاتیەکانی دوانەی گۆڕاو جێگرن کە ئەمەش یەکێکە لە مەرجە بنەڕەتیەکانی بەکارهێنانی ئەم جۆرە مۆدێلە، هەربۆیە ڕێگەمان پێدەدات کە مۆدێلەکانی VARMA (1,0) بەکاربهێنین. پاشان دوای بەکارهێنانی هەمان مۆدێل بۆ پێشبینیکردنی (Forecasting) دیاردەی مردن و برینداربوون بۆ ساڵی ٢٠٢٤، دەرکەوت کە ئەو مۆدێلەی دروستمان کردوە زۆر سەرکەوتووەو دەتوانرێت پشتی پێببەسترێ. بۆ ئەم مەبەستەش پرۆگرامی (ڕ) وستاتا-وەشانی ١٧، بەکارهێنراوە بۆ شیکردنەوەی داتاکانمان. هەربۆیە پێشنیار دەکەین کە بەڕێوەبەرایەتی هاتووچۆی کوردستان سود لەم توێژینەوەیە وەربگرێت بۆ کەمکردنەوە و کۆنترۆڵکردنی ڕووداوەکانی داهاتوو لە ساڵانی داهاتودا.

وشە سەرەکییەکان: زنجیرە کاتیەکانی دوانەی گۆڕاو، VARMA (p, q)، پێشبینیکردن، ڕوداوی هاوتوچۆ.

بناء نموذج إحصائي للتنبؤ بحوادث المرور للوفيات والإصابات باستخدام تحليل السلاسل الزمنية ثنائية المتغير

نوزاد حسين محمود
قسم ادارة الاعمال- جامعة جيهان-السليمانية
nozad.mahmood@sulicihan.edu.krd

دلير حسين قادر
قسم الاحصاء والمعلومات، كلية الإدارة والإقتصاد،جامعة صلاح الدين-اربيل
قسم ادارة الاعمال- جامعة جيهان-اربيل
Dler.Kadir@su.edu.krd

عوبيد محمود محسن
قسم ادارة الاعمال- جامعة جيهان-السليمانية
obed.muhsin@sulicihan.edu.krd

ملخص

في دراستنا ، تم استخدام سلاسل زمنية متعددة المتغيرات تضمنت متغيرين ، وهما معدلات الوفيات والإصابات الناجمة عن حوادث السيارات في مدينة أربيل العراقية. تم جمع بيانات السلسلتين شهريا من يناير 2015 إلى ديسمبر 2020 ، لذلك هناك 72 وحدة في كل سلسلة. أهم نتيجة هي أن السلسلة الزمنية ثابتة ، والنموذج المناسب لتمثيل الظاهرة المدروسة هو فارما (1 ، 0). تم اعتماد نموذج إحصائي للتنبؤ بالحوادث التي أدت إلى الوفاة والإصابات لعام 4202 ، ووجد أنه مناسب. وعلاوة على ذلك ، لتحليل البيانات لدينا ، ونحن نستخدم برمجة (ر) وستاتا الإصدار 17. ونتيجة لذلك ، أشارت الدراسة إلى أن إدارة المرور في كردستان العراق يمكن أن تستخدم النموذج الذي تم تطويره للتنبؤ بالاتجاهات المستقبلية للظاهرة.

الكلمات المفتاحية: سلاسل زمنية ثنائية المتغير ، VARMA (p, q) ، التنبؤ، حوادث المرور.