Forecasting crimes in Northern Mindanao using autoregressive integrated moving average (ARIMA) model

Exequiel R. Gono, Jr.[1] **John Vianne B. Murcia**[2]
[1]Professional Schools, University of Mindanao, Philippines, exequielgono@umindanao.edu.ph
[2]Department of Marketing Management, University of Mindanao, Philippines, jv_murcia@umindanao.edu.ph

ABSTRACT

The study was conducted to forecast the crime rates in Northern Mindanao. The ARMA model was applied to describe the crime rates series and adopted the Box-Jenkins approach to ascertain the most appropriate ARMA model. There is an erratic movement of the crime rates from 2009 to 2013. A decrease in crime rates from 2009 to 2012 and increase in crime rates in 2013. The ADF revealed that the data is stationary at first differencing. The ARIMA (2, 1, 2) model is the best approach to predict the crimes rate in Northern Mindanao, denoted as $Y_t = 0.495859 + \text{-}0.04041Y_{t-1} - 0.746630_{t-2} + .077274_{e-1} + 0.967364_{e-2}$. Furthermore, there is no ARCH/GARCH effect on this model.

Key words: crimes, forecast model, autoregressive integrated moving average, Northern Mindanao, Philippines.

1. INTRODUCTION

Crime forecasting is generally not exercised by the police and security sector, more so by the Philippine police. Although there are several econometric crime literatures or the inclusion of crime in research studies, it is difficult to find police departments or other police agencies that make frequent utilization of econometric forecasting or extrapolative implementation of constrained resources. Several reasons exist including the thought that that crime forecasting is actually not considered useful or possible until recently [1] [2] [3].

The main target of police operations are people engaged to any criminal activities – for instance, investigating the *modus operandi* of serial offenders and detaining them [4]. Certainly, traditional approaches are barely not useful in forecasting the characteristics of each serial offenders [5][6]. Crime forecasting, by means of the approaches studied in this paper, would not become significant until two incidents happened: first, the criminality of locations was developed on the basis of theories like repetitive activities [7], the ecology of crime [8], and hotspots (spatial clusters of crime) [9], and second, police have started systematically mapping incidents through geographic information systems (GIS) over the last 5 to 10 years. The target of the forecast are the criminals; however, the objective of crime reduction and law implementation focuses on areas where criminality appears to occur [10]. Police data could include incidence report of crimes over time and geographical location, and these data could be forecastable.But is precise forecasting of crime *feasible*? There is a significant amount of crime-related distinctiveness and randomness [11]. Nonetheless, forecasts produced by different statistical and algorithm outputs could provide sense on the regularity of behavior [11], crimes included. This fact was notably established through the influence of the recent concepts of opportunity of crime commission, which view crime as a routine activity.

Dealing with uncertainties is very crucial to the institution whose responsibility is to protect and secure the lives of the people. It is very important to come up with information that will happen in the future to avoid worry among the citizens. However, with the drastic change in the agenda of the new president of the Republic of the Philippines number crimes committed are yet a question. In fact, a 2016 report from the national police suggests a 13% dropped on the number of crimes perpetrated in the country, manifesting a continuous downtrend in the past 4 years, amid President Duterte's war on drugs which commenced a year ago. The information implies that the overall number of incidents throughout the country dropped from 675,816 reported crimes in 2015 to 584,809 in 2016. Almost one-fifth of the reported incident was characterized as "index crimes", which referred to as those incidents which are "serious in nature and which occur with sufficient frequency and regularity such that they can serve as an index to the crime situation". Crimes in this classification – which involves murder, homicide, physical injury, rape and burglary – numbered 139,462 for 2016.

In Northern Mindanao, there is a significant decline on the number of crimes in July in comparison with the same month for the past two years. Police data reported a 44% drop on the overall crime rate for July 2016, a month after President

Rodrigo Duterte seated as the country's president. The report is notably significant as data suggests a drastic drop compared to records from the last two years. Our evaluation refers to the direct influence of the reduced prevalence of illegal drugs on the streets, because of the strengthened anti-illegal drug initiative. In sum, these scenarios prompted the researcher to forecast crimes committed in Northern Mindanao. Furthermore, it will test the accuracy and reliability of the model in forecasting crimes.

2. METHOD

This paper utilized time-series analysis, such that variations of past behavior was estimated to provide a prediction of future behavior. Moreover, this approach denotes understanding of series of time-ordered observations as is essential in creating forecast based preceding events and their patterns [13][14]. This research employed the ARIMA model in order to forecast the crime rate of Northern Mindanao.

The autoregressive moving average (ARMA) model was first described by Whittle [15], describing it using statistical inference and both Laurent series and Fourier analysis. However, ARMA gained attention when Box and Jenkins [9] provided a comprehensive discussion of its assumptions and applications. Accordingly, ARMA is particularly useful for polynomials in the lower order, i.e. the third-degree polynomials or lower, and is iterative in the estimation and selection procedure of time-series analysis. The notation of ARMA(p,q) is expanded as:

$$X_t = c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-1} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-1} \qquad (1)$$

ARMA is useful in understanding time-series data, denoted as X_t, and to understand the information of the series essential in forecast estimation [16]. The ARMA model is denoted with boy p, which represents the autoregressive terms (AR) of the series, and q, which denotes the moving average (MA) terms of the series. Estimation of either AR or MA in time-series analysis could become a problem due to possible inaccuracy of the forecasts, given that a high-order analysis could provide an accurate explanation of the complexity of the data and the representation of the correct parameters in them. To address this concern, ARMA models are applicable as they estimate both analyses at the same time.

Box and Jenkins [17] established methods in creating ARMA models which influences time-series analysis as well as forecasting. Box, Jenkins and MacGregor [18] endorsed the application of both autocorrelation (ACF) and partial autocorrelation function (PACF) into the sample data since

these are key instruments in determining the direction of ARMA model. Within classification phase, the conversion of data is required to construct the time series stationary [19].

Stationarityis considered as an important phase in building ARMA models and the resulting forecasting procedure thereafter. In addition, a stationary time-series is characterized having to exhibit certain statistical properties, such as having both stable mean and autocorrelation structure [20]. While heteroscedasticity and sequence both exist in the experimental time series, power transformation and differentiation are employed to remove the pattern and render the variance stable before an ARMA model can be used [21].Finally, when the white noise error term has an equal variance and when the series' ARMA (p, q)model's order is established, the assumption of constant variance will be removed and presence of autoregressive conditional heteroscedasticity (ARCH) effect could now be estimated [22].

3. RESULTS

The study was conducted to forecast the crime rates in Northern Mindanao. The ARMA model was applied and intended to pattern the crime rates series. Further, the adapted Box-Jenkins method was employed to ascertain the most appropriate ARMA model. Patterns in the crimes rates were shown in figure 1. We can observe an erratic pattern of crime rates from 2009 to 2013. A decrease in crime rates from 2009 to 2012 and increase in crime rates in 2013. Furthermore, we can observe that the data is not stationary.
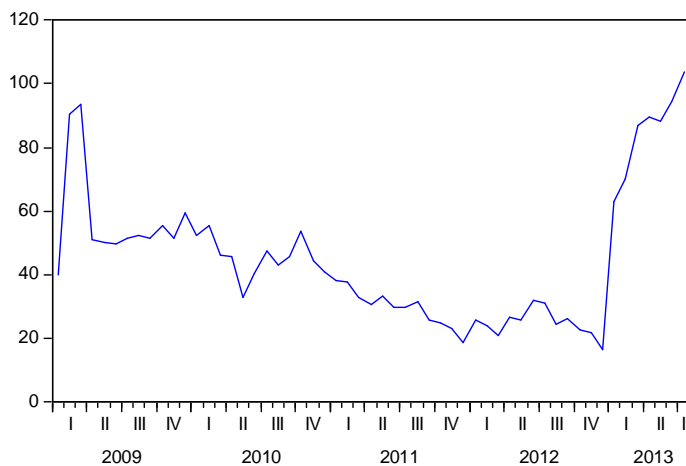

Figure 1: Trend of crime rates in Northern Mindanao

The crime rates are plotted using both autocorrelation (ACF) and partial autocorrelation function (PACF), both of which can distinguish the property of time series [18]. Figure 2 shows the ACF and PACF plots of crime rates in Mindanao are shown in Figure 2. A visual inspection of the series can

preliminarily confirm its stationarity as characterized by a correlogram that decays rapidly. However, visual checking of the correlograms alone does not suffice to conclusively confirm its stationarity [23]; in such case, we utilized Augmented Dickey Fuller (ADF) test to check the time-series' property.

Shown in table 1 is the test for stationarity of the crime rates in Northern Mindanao. Using ADF test, the time-series data of the crime rates was found to be non-stationary ($p=0.6469$). With this, we subjected the data to first-level differencing. The core principle of differentiation is to smoothen or stabilize the time-series data [24]. After differencing, an ADF test confirmed that the data is stationary; hence, the d of the series is at order 1.

Table 1: Augmented Dickey-Fuller (ADF) test of stationarity of crime rates in Northern Mindanao

Degree of the series	t-statistics	Probability
At level	-1.248446	0.6469
First difference	-9.30891	0.000*

Running the correlograms of the first differencing of the crime rates is performed to give an idea of number of lags that will fitted in the ARMA model. With regards to the autocorrelation and partial correlation of the correlogram, the decay was observed on lag 2.

Moreover, shown in table 2 is the ARMA model to represent the represent the crime rates in Northern Mindanao. In the tables on ARMA model identification, both Akaike (AIC) and Schwarz Information Criterion (SIC) for every model was presented on the second and third columns. Based on the table, the desired model appears to have the highest coefficient of determination or the R^2, with the least non-significant parameter estimates and non-significant $Q12$ statistic. In contrast, desired model reduces AIC and SIC values. Further, if the p-value correlated in $Q12$ is superior than 10% significance level, the hypothesis indicated on the first twelve (12) autocorrelations were not significantly different from zero would be "accepted" and hence, implicitly "accepting" that the model is accurate [12] [18].

Table 2: Different ARMA estimations to represent the crimes in Northern Mindanao

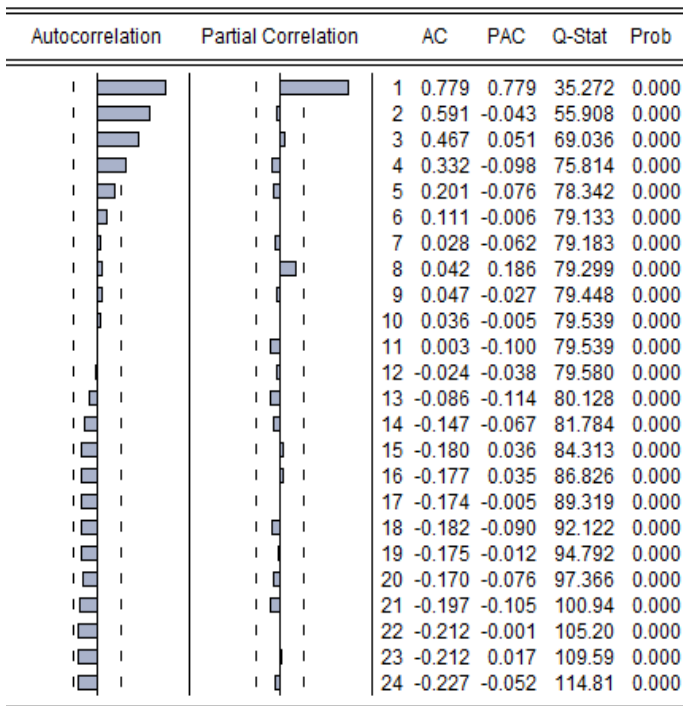| Model	AIC	SIC	Not Sig	R^2
AR1	7.56	7.63	*B1*	0.00
AR2	7.534	7.609	*B1, B2*	0.04
AR3	7.18	7.26	*B1, B2, B3*	0.003
ARMA(1,1)	7.59	7.7	*B1, e1*	0.00
ARMA(2,1)	7.68	7.62	*B1, B2, e1*	0.04
ARMA(1,2)	7.62	7.56	*B1, B2*	0.07
ARMA(2,2)	7.37	7.3	none	0.30

Figure 2: Autocorrelation and partial autocorrelation functions of crime rates in Northern Mindanao at level
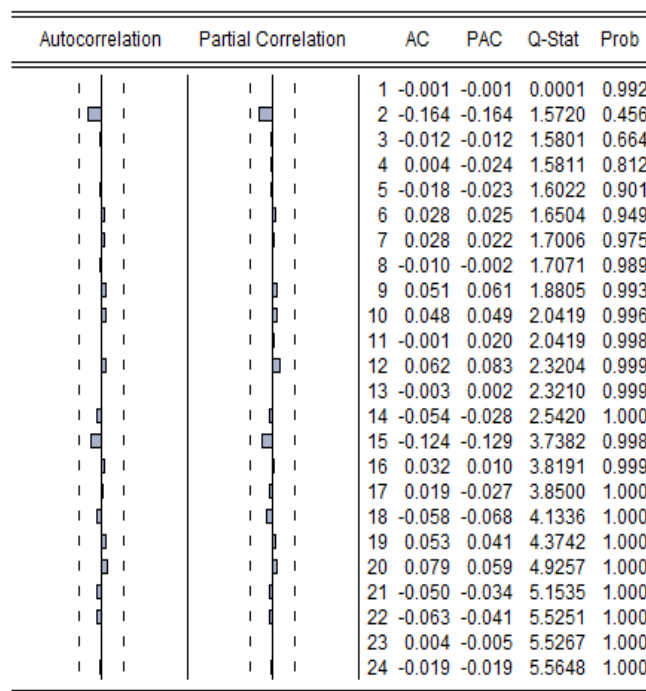

Figure 3: Autocorrelation and partial autocorrelation functions of crime rates in Northern Mindanao after first difference

Shown in table 3 is the ARIMA (2,1,2) model to forecast the crimes rate in Northern Mindanao. The results shows that the crime rates can be predicted in the model

$$Y_t = 0.495859 + -0.04041 Y_{t-1} - 0.746630_{t-2} \quad (1)$$
$$+ 0.077274_{e-1} + 0.967364_{e-2}$$

Table 3: ARIMA(2,1,2) model to forecast the crimes in Northern Mindanao

Variable	Coefficient	S.E.	*t*-statistic	Prob.
Intercept	0.495859	1.511418	0.328075	0.7443
AR(1)	-0.04041	0.087005	-0.46449	0.6444
AR(2)	-0.74663	0.078231	-9.54386	0.0000
MA(1)	0.077274	0.053126	1.454542	0.1524
MA(2)	0.967364	0.029257	33.06399	0.0000
F-statistic	0.033462	Prob. F(1,51)		0.8556
Obs*R^2	0.034751	Prob. χ^2(1)		0.8521

Finally, a test to determine volatility of the series revealed that both F test and χ^2 test did not yield significant results; thus, the model has no ARCH or GARCH effect.

4. CONCLUSION

The study was conducted to forecast the crime rates in Northern Mindanao. The ARMA model was used to model the crime rates series, as well as the adaptation of Box-Jenkins method was employed to further ascertain the most appropriate ARMA model. The ADF revealed that the data is stationary at first differencing. We can observe an erratic pattern of the crime rates from 2009 to 2013. A decrease in crime rates from 2009 to 2012 and increase in crime rates in 2013. The data is not stationary and necessary tests to stabilize the series were done before proceeding to the forecasting estimation procedure. Finally, the best model derived is the ARIMA (2,1,2) model to forecast the crimes rate in Northern Mindanao. Crime rates in this region, however, do not exhibit volatility.

ACKNOWLEDGMENT

REFERENCES

1. D. H. Klepinger, and J. G. Weis. **Projecting crime rates: An age, period, and cohort model using ARIMA techniques**, *Journal of Quantitative Criminology* Vol. 1, no. 4, pp. 387-416,1985.
2. P. Chen, H. Yuan, and X. Shu. **Forecasting crime using the ARIMA model**, in *2008 Fifth International Conference on Fuzzy Systems and Knowledge Discovery*, IEEE, Vol. 5, pp. 627-630, 2008.
3. M. Zakria, and F. Muhammad. **Forecasting the population of Pakistan using ARIMA models**, *Pakistan Journal of Agricultural Sciences* Vol. 46, no. 3,pp. 214-223, 2009.
4. L. W. Sherman. **Attacking crime: Police and crime control**, *Crime and Justice* Vol. 15, pp. 159-230, 1992.
5. W. Gorr, and R. Harries. **Introduction to crime forecasting**, *International Journal of Forecasting* Vol. 19, no. 4, pp. 551-555, 2003.
6. E. R. Groff, and N. G. La Vigne. **Forecasting the future of predictive crime mapping**, *Crime Prevention Studies* Vol. 13, pp. 29-58, 2002.
7. L. E. Cohen, and M. Felson. **Social change and crime rate trends: A routine activity approach**, *American Sociological Review*, pp. 588-608, 1979.
8. M. Felson, and L. E. Cohen. **Human ecology and crime: A routine activity approach**, *Human Ecology* Vol. 8, no. 4, pp. 389-406, 1986.
9. L. W. Sherman, P. R. Gartin, and M. E. Buerger. **Hot spots of predatory crime: Routine activities and the criminology of place**, *Criminology* Vol. 27, no. 1, pp. 27-56, 1989.
10. E. Groff, D. Weisburd, and N. A. Morris. **Where the action is at places: Examining spatio-temporal patterns of juvenile crime at places using trajectory analysis and GIS**, in *Putting Crime in Its Place*, Springer, New York, pp. 61-86, 2009.
11. J. Ackerman. *Quantitative Methods in the Study of Deviance and Crime*, The Handbook of Deviance, 2015, p. 191.
12. M. Tawarish andK. Satyanarayana. **A review on pricing prediction on stock market by different techniques in the field of data mining and genetic algorithm**,*International Journal of Advanced Trends in Computer Science and Engineering* Vol. 8, no. 1, pp. 23-26, 2019.
13. M. K. Singla, J. Gupta, P. Nijhawan andA. S. Oberoi. **Electrical load forecasting using machine learning.** *International Journal of Advanced Trends in Computer Science and Engineering* Vol. 8, no. 3, pp. 615-619, 2019.
14. A. S. Weigend. *Time Series Prediction: Forecasting the Future and Understanding the Past*, New York, U.S.: Routledge, 2018, ch. 1, p. 11.
15. P. Whittle. **The analysis of multiple stationary time series**, *Journal of the Royal Statistical Society: Series B (Methodological)* Vol. 15, no. 1, pp. 125-139, 1953.

16. M. Y. Anwar, J. A. Lewnard, S. Parikh, and V. E. Pitzer. **Time series analysis of malaria in Afghanistan: using ARIMA models to predict future trends in incidence**, *Malaria Journal* Vol. 15, no. 1, pp. 1-10, 2016.
17. G. E. Box, and G. M. Jenkins. **Time Series Analysis: Forecasting and Control**, Revised Ed., California, U.S.: Holden-Day, 1992.
18. G. E. Box, G. M. Jenkins, and J. F. MacGregor. **Some recent advances in forecasting and control**, *Journal of the Royal Statistical Society: Series C (Applied Statistics)* Vol. 23, no. 2, pp. 158-179, 1974.
19. M. Khashei, and M. Bijari. **An artificial neural network (p, d, q) model for timeseries forecasting**, *Expert Systems with Applications* Vol. 37, no. 1, pp. 479-489, 2010.
20. F. J. Breidt, N. Crato, and P. De Lima. **The detection and estimation of long memory in stochastic volatility**, *Journal of Econometrics* Vol. 83, no. 1-2, pp. 325-348, 1998.
21. K. W. Hipel, A. I. McLeod, and W. C. Lennox. **Advances in Box Jenkins modeling: 1. Model construction**, *Water Resources Research* Vol. 13, no. 3, pp. 567-575, 1977.
22. S. Degiannakis, and E. Xekalaki. **Autoregressive conditional heteroscedasticity (ARCH) models: A review**, *Quality Technology & Quantitative Management* Vol. 1, no. 2, pp. 271-324, 2004.
23. C. W. J. Granger, and P. Newbold. **The time series approach to econometric model building**, *Econometric Society Monographs* Vol. 32, pp. 302-316, 2001.
24. S. Leybourne, P. Newbold, and D. Vougas. **Unit roots and smooth transitions**, *Journal of Time Series Analysis* Vol. 19, no. 1, pp. 83-97, 1998.