Reported Murdered Cases in Pakistan: Application of Seasonal Box-Jenkins Methods

Muhammad Imran¹ and Jamal Abdul Nasir² and Najeeb Ul Rehman³

¹Visiting Faculty, Rahim Yar Khan Campus, The Islamia University of Bahawalpur, Pakistan.
²Director Campus Rahim Yar Khan (RYK), The Islamia University of Bahawalpur, Pakistan.
³MSc statistics, Department of Statistics, The Islamia University of Bahawalpur; Sub Campus (RYK)

Abstract Forecasting about crimes is often challengeable but possible by carefully implementation of forecasting models. Prediction about murdered incidents is an emerging for law enforcement agencies to adopt preventative policies and tactical operations. This study therefore aimed to uncover the pattern of reported homicide related deaths in Pakistan as well as development of mathematical model by incorporating the seasonal Box Jenkins methods. SARIMA (111) x (011)₁₂ is a correct model by taking into account the model selection criterion namely, AIC, HQ and SIC. The reported murdered cases in 2012 were 13846 while established model, SARIMA (111) x (011)₁₂ revealed that the increment 442 cases are expected in 2013.

Keywords: Homicide; Model Selection Criterion; Pakistan; SARIMA

Introduction
The history of homicide is as old as man itself. In worldwide almost half million people has lost their lives every year, mostly in Americas and Africa (UNODC 2013). The worldwide homicide related deaths are 437,000 in 2012 (UNODC 2013). By regionally 36% in Americas, 31% in Africa, 28% in Asia, 5% in Europe and 3% in Oceania The overall decline in homicide rate globally. The global mean homicide rate is 6.2 per 100,000 populations (UNODC 2013). The mean homicide rate from 2002 to 2012 in Pakistan per 100,000 populations is 6.75 (UNODC 2013).

Significant risk factors correlated with homicide rate
Variability exists in homicide pattern and magnitude region by region country by country. In some regions the structured crimes and the violent culture of the youth gang is highly responsible for homicide and others killing related to familiar partner and family related violence (UNODC 2013). Many researchers in cross-national study observed the risk factors that affect the homicide rates. Cole and Gramajo (2009) have made cross-national empirical study the number of socioeconomic variables that influence the level of lethal violence, like age stature, urbanization, income inequality and poverty, population density, ethno-linguistic, religious heterogeneity, education and governance indicators. They found the significant factors countries with high levels of cultural and ethnic heterogeneity tend to have relatively higher homicide rates (Cole and Gramajo 2009, Nivette 2011). On the other hand, countries with a high ranking on the World Bank’s quality of governance indicator tend to have lower rates, education was also found to be significantly associated with homicide rates (Cole and Gramajo 2009). LaFee (1999) studied the six risk factors and found that economic inequality was the consistent and significant factor that influences the homicide rate, while a negative relation hold between economic development and homicide rates. Balu (1982) found that socioeconomic discrimination raises the rate of criminal violence. Higher levels of homicide are linked with lower human and economic development (UNODC 2013). Nivette (2011) explored that positive correlation exists between income inequality divorce rate, population growth rate, female labor force participation, infinity mortality rate, ethnic heterogeneity and the homicide rate. Lack of good governance and crimes rate are linked. Good governance yet another significant factor. The governance is significantly and inversely connected to national homicide rate (Younker et al. 2013). The level of homicide can be reduced by improving the legitimacy in a nation (Nivette and Eisner 2012).

Causes of Homicide in Pakistan
Pakistan is a less developed country as well as 6th most populous country in the globe and rising trend in population, expected population approximately 230.7 million in 2027 (Zakria and Muhammad 2009).
Population density is yet another factor to create a more chance for crimes (Gillis 1974). Pakistan is facing many challenges, corruption and governance problem, growth and consumption inequality, and the effects of religious extremism in our society (Javaid 2010, Asad and Ahmad 2011, Javaid 2011). The major reasons for murder in Pakistan are family related violence, revenge, petty affair, old enmities honor killing and property dispute (Suhail and Javed 2004, Wasti 2010). Most of the murderers are illiterate unskilled young and poor (Suhail and Javed 2004). Different studies have been made to understand the homicide pattern in various districts of Pakistan like Haripur, Faisalabad, and Dera Ismail khan, and found that the ratio of male is higher as compared to female and the targeted age group is 21 to 30 years, while the most common weapons used for homicide are firearm (Bashir et al. 2004, Agha et al. 2012, Mujahid et al. 2006).

**Related work**

Effectively implementations of mathematical and statistical models allow us to better understand the trend and magnitude of crime as well as source allocation, preventative strategies and legislation through law enforcement agencies. For a univariate time series (Pankratz 1983), Box-Jenkins Methodology is considered as a more precise technique as compared to other traditional statistical models (Afzal et al. 2002). This methodology is capable to deal with the seasonality and has become a much popular and excessively used in various fields particularly in economics and social field. Shrivastav and Ekata (2012) in crime forecasting applied the ARIMA technique to forecast the crime in Gujarat State in India and present an adequate model for crime forecast. Osho and Daramola (2010) used Box-Jenkins approach to predict the crime rate in Harris County for the over specific period of time. Noor et al. (2013) used ARIMA model and fuzzy alpha-cut method to forecast the crime. Chen et al. (2008) used the ARIMA model for property crime in one city of china. A set of univariate, naïve methods were carried out to accurately forecast crime one month ahead Holt exponential smoothing considered as accurate method (Gorr et al. 2003).

**Methods and Material**

The relevant data for the present study has been taken from Pakistan Bureau of Statistics (Statistics House) government of Pakistan, Islamabad, Pakistan. The monthly data of the reported murdered from January 2005 to December 2012 with a total 96 observations. Let \( X_t \) is a monthly number of reported murdered cases in Pakistan. Minitab 15.0 and GRETL have been used for the execution of ARIMA models, while MS-Excel used for graphical representation.

**Box-Jenkins Methodology**

This methodology is a systematic iterative process until an adequate model is achieved. A procedure is enhanced step by step identification, estimation of parameters, diagnostic checks and finally forecast. A model consist the three parameters one autoregressive (p), second differencing order (d) and third moving average order(q) (Box-Jenkins 1976, Hoff 1983, Box-Jenkins and Reinsel 2008). The general equation of the auto regressive moving average model is given in equation (1)

\[
X_t = c + \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \ldots - \theta_q \epsilon_{t-q}
\]  

(1)

Where “\( \phi \)’’ and “\( \theta \)’’ are the autoregressive and moving average parameters to be estimated. “\( X's \)” and the “\( \epsilon's \)” are the original series and residuals respectively. The residuals assumed to follow a normal probability distribution. Backshift transformation is used on equation (1). By using this transformation we can get the ARMA (p, q) or ARIMA (p, 0, q) model, as in equation (3).

\[
(1-\phi_1 B - \ldots - \phi_p B^p) X_t = c + (1-\theta_1 B - \ldots - \theta_q B^q) \epsilon_t
\]  

(2)

\[
\phi \ (B)(1-B)^d X_t = \theta(B) \epsilon_t
\]  

(3)

Where \( \epsilon_t \) follow a white noise series with zero mean and variance \( \sigma^2 \).

SARIMA model given by Box – Jenkins

\[
\Phi(B^s) \phi(B) \nabla^d \nabla^D X_t = \Theta(B^s) \ \theta(B) \epsilon_t
\]  

(4)

Where \( \nabla^d = (1-B)^d \) and \( \nabla^D = (1-B^s)^D \)
\( \nabla^d \) = regular difference (non-seasonal) and \( \nabla^D \) = seasonal difference (s=12) while the non-seasonal and seasonal autoregressive and moving average components are denoted by polynomial \( \phi(B), \theta(B) \) and \( \phi(B^s), \theta(B^s) \) of order p, q and P, Q respectively.

Figure 1(a): Flow chart to identify an appropriate ARIMA model

Identification

It is an initial step of the Box-Jenkins process, in this step the given time series is checked whether it is mean reverting series or not. A plot of actual time series by taking reported murdered cases along y-axis showing the pattern of seasonality in Figure 1. There is a seasonal trend in homicide rate, higher number of incident occurs in summer season (Bashir et al. 2004). The actual series is not stationary, after the 1st differenced the stationary level is achieved. Stationarity is also verified by the help of statistical test Kwiatkowski-phillips-schmidt-shin (KPSS) test, the acceptance of null hypothesis tends to lead stationarity (Dhrymes 1998). Kwiatkowski-phillips-schmidt-shin test is performed including trend and seasonality and Lag truncation parameter is equal to 3. When \( d=0 \) the null hypothesis is rejected. When \( d=1 \) the actual series become stationary at 5% level of significance detail is given in Table 1. A plot of 1st differenced series is shown in figure 2.

Table 1: Kwiatkowski-phillips-schmidt-shin test

<table>
<thead>
<tr>
<th>Difference level</th>
<th>d=0</th>
<th>d=1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test statistic</td>
<td>0.16</td>
<td>0.036</td>
</tr>
<tr>
<td>Critical value at 5% LS</td>
<td>0.148</td>
<td>0.148</td>
</tr>
<tr>
<td>Conclusion</td>
<td>Reject H₀</td>
<td>Accept H₀</td>
</tr>
</tbody>
</table>

To identify an adequate ARIMA model, the graphical representations better allow us to choose appropriate initial value of parameter p and q for non-seasonal and P, D, Q for seasonal. The autocorrelation function (ACF) and the partial autocorrelation function (PACF) for actual and the 1st difference series is given in figure 3(i), 3(ii) and figure 4(i), 4(ii) respectively. With 5% significance limits for the autocorrelations and partial autocorrelations is (±1.96).
As the correlogram of the actual and 1st difference series clearly showing the pattern of seasonality and s=12 because the spikes at lag 12 is significantly different from zero in autocorrelation function from figure 4(i). By fixing the one regular difference (d=1) and one seasonal difference (D=1) we performed more than 25 combination of models. The tentative models along with model selection criterion are given in table 2.

The final model is selected by taking into the account the model selection criterion like Akaike information criteria (AIC), Hannan-Quinn criteria (HQ) and Schwarz information criteria (SIC). According to this the model with least AIC value (954.6354) will be selected, as detail summery of tentative models is given in table 2. SARIMA (111) X (011)12 is a final selected model. The estimation of the final selected model parameters is given in table 3 along with p-value standard error of coefficients and the t statistic.

Table 2: A tentative models along with model selection criterion

<table>
<thead>
<tr>
<th>SARIMA Model</th>
<th>AIC</th>
<th>HQ</th>
<th>SIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>(111)x(011)12*</td>
<td>954.6354*</td>
<td>959.4942*</td>
<td>966.7296*</td>
</tr>
<tr>
<td>(111)x(111)12</td>
<td>956.0535</td>
<td>961.8841</td>
<td>970.5666</td>
</tr>
<tr>
<td>(011)x(011)12</td>
<td>957.9431</td>
<td>961.8301</td>
<td>967.6185</td>
</tr>
<tr>
<td>(011)x(111)12</td>
<td>957.3812</td>
<td>962.24</td>
<td>969.4754</td>
</tr>
<tr>
<td>(011)x(110)12</td>
<td>959.1279</td>
<td>963.015</td>
<td>968.8033</td>
</tr>
<tr>
<td>(110)x(111)12</td>
<td>975.2879</td>
<td>980.1466</td>
<td>987.3821</td>
</tr>
<tr>
<td>(110)x(110)12</td>
<td>973.7561</td>
<td>977.6431</td>
<td>983.4315</td>
</tr>
<tr>
<td>(211)x(011)12</td>
<td>956.083</td>
<td>961.9136</td>
<td>970.5961</td>
</tr>
<tr>
<td>(211)x(111)12</td>
<td>957.0861</td>
<td>963.8883</td>
<td>974.0179</td>
</tr>
<tr>
<td>(211)x(110)12</td>
<td>959.2603</td>
<td>965.0909</td>
<td>973.7734</td>
</tr>
</tbody>
</table>

*Lowest value of AIC, SIC and HQ

By establishing the particular ARIMA model with estimated parameters under the sophisticated computational algorithms, we next see whether the chosen model is adequate or not. A model is considered as an adequate the residual of that model are distributed as white noise, i.e., they are independent and normality distributed with zero mean and constant variance and uncorrelated for all lags. Ljung-Box.Q statistic is used as a diagnostic purpose, the acceptance of null hypothesis leads toward that the fit is good, a detail statistics is given in table 4. The null hypothesis is accepted because the p-value at lags 12, 24, 36 and 48 is greater than the level of significance (5%) that indicating the best fit. P-value given in table 3 for the final estimates of parameters of SARIMA (111) x (011)12 are less than the level of significance 5%, indicating that the coefficients are significantly different from zero.

Table 3: Final Estimates of Parameters of ARIMA (111) x (011)12

<table>
<thead>
<tr>
<th>Type</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR 1</td>
<td>0.2678</td>
<td>0.1311</td>
<td>2.04</td>
<td>0.044</td>
</tr>
<tr>
<td>MA 1</td>
<td>0.9083</td>
<td>0.0662</td>
<td>13.71</td>
<td>0.00</td>
</tr>
<tr>
<td>SMA 12</td>
<td>0.7634</td>
<td>0.1068</td>
<td>7.15</td>
<td>0.00</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.3096</td>
<td>0.2983</td>
<td>-1.04</td>
<td>0.03</td>
</tr>
</tbody>
</table>
Reported Murdered Cases in Pakistan: Application of Seasonal Box-Jenkins Methods

Table 4: Modified Box-Pierce (Ljung-Box) Chi-Square statistic

<table>
<thead>
<tr>
<th>Lag</th>
<th>12</th>
<th>24</th>
<th>36</th>
<th>48</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-Square</td>
<td>3.8</td>
<td>18.6</td>
<td>23.4</td>
<td>44.1</td>
</tr>
<tr>
<td>DF</td>
<td>8</td>
<td>20</td>
<td>32</td>
<td>44</td>
</tr>
<tr>
<td>P-Value</td>
<td>0.874</td>
<td>0.548</td>
<td>0.865</td>
<td>0.466</td>
</tr>
</tbody>
</table>

Forecast

The final elected model for the reported murdered cases in Pakistan monthly is SARIMA (1,1,1) x (0,1,1)12. A graphical representation, actual versus fit along with 12 month prediction and 95% confidence limits is shown in figure 5.

The auto correlation function (ACF) and partial auto correlation function (PACF) at 5% limits are given in figure 6(i) and 6(ii). All the spikes are well in the 95% critical limits. Shapiro-Wilk normality test was also performed, \( W = 0.991655 \), with p-value 0.815871. This test shows that the p value is greater than the level of significance 5%, thus the residual follow a normal distribution.

Discussion and Conclusion

After the application of seasonal box-Jenkins methods on reported homicide related deaths in Pakistan from 2005 to 2012. SARIMA (1,1,1) x (0,1,1)12 is an established final efficient model for the prediction of monthly reported number of murdered cases in Pakistan. The total number of reported murdered cases in 2012 was 13846 while the predicted total number of murdered cases in the year of 2013 is 14288. The expected percent increment is 3.19% annually. A model predicted plot of reported murdered with actual and 12 months forecast along with 95% confidence limits is shown in figure 5, the forecast values agree well with the observe data. While the autocorrelation function plot and partial auto correlation function plot indicating that all spikes are well in within 95% critical limits are shown in figure 6 (i) and figure 6 (ii). While descriptively 95807 murdered cases were reported in Pakistan from 2005 to 2012, while the average reported murdered cases per month 998 with absolute measure of variability (standard deviation) 189.
Hence SARIMA (111) x (011)12 is an optimum established model for forecasting, monthly reported murdered cases in Pakistan. This statistical finding will serve as guide to the government and the law enforcement agencies to formulate smart, effective and proactive crime defensive policies, legislation, and timely intervention and mobilization of sources to tackle the homicide one of the worst crime.

References


