May 16th, 12:00 PM - 3:00 PM

Modeling and Forecasting Crime Patterns in Bellingham, Washington

Zachary Domingo  
*Western Washington University*

Eric Shoner  
*Western Washington University*

Follow this and additional works at: [https://cedar.wwu.edu/scholwk](https://cedar.wwu.edu/scholwk)

Part of the [Higher Education Commons](https://cedar.wwu.edu/scholwk)


This Event is brought to you for free and open access by the Conferences and Events at Western CEDAR. It has been accepted for inclusion in Scholars Week by an authorized administrator of Western CEDAR. For more information, please contact [westerncedar@wwu.edu](mailto:westerncedar@wwu.edu).
Abstract:
Our purpose is to use time series analysis to model and forecast the underlying dynamics behind crime in Bellingham, Washington. Using recent monthly data from the Bellingham Police Department, we considered singular spectrum analysis (SSA) and autoregressive integrated moving average (ARIMA) modeling techniques to estimate significant deterministic patterns in the data. We created two models for alcohol offenses and domestic violence datasets then compared them. The better performing model was used to forecast the number of crime incidents for the next ten months and identify trends and seasonality.

Background and Significance:
The federal and state governments spend billions of dollars on maintaining law enforcement agencies. If we had the ability to predict trends or periodicities in crime patterns, these agencies could then properly allocate their resources to reduce the incidence. This would allow the local law enforcement to take more meaningful action or make better informed decisions in measures to reduce crime. This research takes a somewhat narrow scope, considering only crime data from Bellingham, Washington (Fig. 1, Fig. 2) [1].

Research Questions:
- How well can we model crime trends/seasonalities in Bellingham, specifically, alcohol offenses and domestic violence?
- Which of two models (SSA-based or ARIMA) performs better in terms of forecasting?

Methods:
- Compared an ARIMA model with a combination of SSA and ARIMA model (ARIMA without any differencing) for each data set.
- Performed a logarithmic transformation on all data sets to ensure appropriate comparisons.
- Removed one outlier in alcohol data set.
- Used the first 48 observations (first 4 years) to create our models so that we could forecast the next 10 (if an outlier was removed) months and compared them to observations from months 49 to 58 (with an outlier removed).
- Used the Dickey-Fuller-Marriott test (D-M test) and root mean square error (RMSE) values to see if one model performed better than the other per dataset.

ARIMA Modeling and Forecasting for Domestic Violence:
- Using the first 48 observations as training data, the AIC criterion selected an ARIMA(2,1,0) model [3].
- Forecasted values using the ARIMA(2,1,0) model for the trained data with 80% and 95% prediction intervals (Fig. 3).
- Examined the autoregression function (ACF) and partial autocorrelation function (PACF) plot, and applied a Portmanteau test on the residuals of the model to ensure the residuals were uncorrelated (Fig. 4).
- Prediction intervals from model capture the data.

Summary of Statistical Models:
- SSA is a relatively new method that can model the deterministic component (trend and seasonality) simultaneously without any restrictive assumptions [2]. If the residuals of the SSA are correlated, then they may be modeled by the traditional autoregressive integrated moving average (ARIMA) model.
- ARIMA is a typical time series model that handles non-stationary time series data by differencing [3].

Analysis on Forecasting Performance:
- Domestic Violence:
  - The RMSE values for the SSA model are lower, suggesting its superior forecast performance (Fig. 11).
  - On the other hand, the D-M test shows no significant difference in performance between the two models for all forecast values except one, implying a weak advantage of the SSA model (Fig. 12).
- Alcohol Offenses:
  - The RMSE values for the SSA model are lower, suggesting its superior forecast performance (Fig. 13).
  - On the other hand, the D-M test does not indicate any statistically significant advantage of the SSA compared to the AR(1) model (Fig. 14).

Conclusions:
The SSA-based models tend to produce much more accurate forecasts than the ARIMA models for the two data sets being analyzed. These results may be attributed to the fact that the SSA can be used even for relatively small data sets. The SSA-based models for both data sets indicate yearly periodicity for both datasets. Both data sets show increases in frequency during winter and peaks in summer months, and decreases for the rest of the year.

The above finding could be attributed to the fact that children are not in school and thus are more likely to become the victim of domestic violence. Similarly, the increase in alcohol crimes could be attributed to minors seeking out alcohol when school is not in session. Further investigation is required to substantiate these claims and it is likely that unaccounted for factors contribute to these periodicities.

To further this research, a multivariate analysis of correlation between domestic violence crimes and alcohol crimes over time may be considered.

References:

Zak Domingo, Eric Shoner, Phillip Thomas, Justin Irvine, advised by Kimihiro Noguchi
Department of Mathematics, Western Washington University, Bellingham, WA 98225