UCLA ASA DataFest – 2020 Challenge
Appendix - Team Ocean's Four

Introduction

Los Angeles County has issued stay-at-home orders starting from March 19, 2020. Almost two months into the lockdown, we have received four warning emails from the UCLA Police Department regarding crimes near campus. This has brought our attention to the impact of lockdown on crimes within LA County. After following news on the recent downward trend of LA County crime, we hypothesized that the total crime cases would go up when the order is loosened or the lockdown ends. In this project, we explored crime change over the past ten years and projected the potential growth of LA County crime in the following year. We believe that researching crime trends can shed light on COVID-19's effects from an integrated perspective. Furthermore, this project could help the public, the government and police officers to be more prepared as we enter the next stage of the pandemic.

Dataset

We collected 2010-2020 crime statistics from the Los Angeles County Sheriff's Department (LASD). To project the short-term effect of lockdown on the total crime cases as LA County issued another stay-at-home order for the next three months, we selected the crime data from 2020/03/19 to 2020/05/03 to better capture the effects of the lockdown order. We removed the most recent days of crime data from forecasting because there could be problems with report lag and record delay. After plotting total crime cases over the past ten years, we found the unusually high number of cases (over 1000) reported on January 1 of each year whereas a normal daily reporting was around 500. We researched one reason for such a spike was the delay in crime summary, so we decided to subtract the annual average from January 1 and add the excess part evenly to every day of the year. We also used the past ten years' crime data to model the long-term behavior of criminality in LA County. In addition, we consulted past research related to crime change and included factors such as law enforcement and unemployment rate in our analysis. As the spread of the virus may directly affect crimes in LA County, dataset on confirmed cases of COVID-19 was collected to analyze the trend.

Methods

We first focus on using total crime cases per day since lockdown to project future criminality in the short-term future. After plotting the graph, we found that the time series might be stationary. In other words, we can use the ARIMA model for forecasting. ARIMA models are one of the most widely used approaches to time series forecasting and they aim to describe the autocorrelations in the data. Therefore, we perform the unit-root test to see if the time series is stationary or not. Since the p-value is greater than 0.05, we fail to reject the null hypothesis that

the time series is non-stationary. Then we plot the ACF and PACF graph to determine the possible candidate models. From the ACF plot we can see that there is a significant spike at lag 2, followed by the spike at lag 1. Hence, we expect an AR(1) model to be appropriate. The auto.arima() function in R that uses a combination of unit root tests and minimisation of the AICc and MLE also verifies our expectation. We check the residuals from the ARIMA(1,0,0) model by plotting the ACF of the residuals and doing a portmanteau test of the residuals. The residuals look like white noise and the p-value of Ljung-Box test is also greater than 0.05. Therefore, the model is appropriate and we can use it for forecasting.

Likewise, we follow the same approach for long-term forecasting. In order to predict the total crime in LA County after the lockdown is dismissed, we use crime cases of the entire LA County from 2010 to 2019 and we take the weekly average as our sample because daily crime cases generated noisy data. After plotting the time series, we find the graph exhibited seasonality so that it might not be stationary. We then decomposed the time series to check if there is a seasonal or cyclical pattern. We observed a 1-year seasonal cycle and a possible 5-year cycle. However, since we only care about 1-year prediction, we decide to only consider the 1-year seasonality. For a stationary time series, the ACF will drop to zero relatively quickly, while the ACF of non-stationary data decreases slowly. We first take a seasonal difference, but the ACF also appears to be non-stationary, so we take an additional first difference. However, the time series after differencing fail both ACF and the unit-root test. Therefore, we decided to look for other modelling. The next model we consider is to use STL decomposition. STL is a versatile and robust method for decomposing time series that exhibit seasonality. However, after we fit the STL model, the residuals fail the Ljung-Box test with p-value much smaller than 0.05, which means that the model does not fit our observations. An alternative approach we consider is to use a dynamic harmonic regression model. Since seasonal versions of ARIMA models are designed for shorter periods such as months or quarters, seasonal differencing of high order does not make a lot of sense. A harmonic regression approach where the seasonal pattern is modelled using Fourier terms with short-term time series dynamics handled by an ARMA error is more appropriate in our case. Using a line search to minimize AICs values, we find that an ARIMA(1,1,2) process with 12 pairs of Fourier sin and cos is the best fit. The residuals of this model pass the Ljung-Box with p-value greater than 0.05 and the residuals ACF plot show that they are random noises.

As the time series since lockdown appears to be a stationary process with constant mean and the observations are comparatively limited with respect to the projection window we are interested, the predicted values soon converge to the mean of the observations. Hence, we decided to use the drift method to allow the short-term forecasts to increase over time and eventually reach historical mean and then evolve with the long-term forecasts.

Discussion

As forecasted by our model, total crime cases in LA County would experience a sharp increase after reopening and surpass the historical average of 576 cases per day in August 2020. By the end of 2020, total crime is projected to become relatively stabilized at the average. As the unemployment rate in the County has soared in March and is likely to continue the upward trend, this supports our hypothesis that unemployment is positively related to total crime. Moreover, since law enforcement is pivotal in controlling crime, the current high rates of infection among police officers suggest the potential lack of policing practices in the future.

From the maps showing total incidents in the same month of 2019 and 2020, there is an obvious decrease in crimes of LA County. But some cities like Los Angeles and Compton still face higher risks than others. Looking closely at changes in specific crimes, we noticed that child abuse, sex, drink drug, and traffic experienced a drop in cases. However, these crime categories would probably increase as children returning to school and people's lives back to normal.

Insights

Our results indicate the need for further economic support, including another round of stimulus check from the Federal Government and unemployment benefits. The issue that law enforcement agencies might be understaffed in the following months should be addressed by providing sufficient protective equipment to the police officers and allocating available funding to encourage recruitment in police departments. More broadly, the general public should continue to take precautionary measures in order to support the efforts by people who cannot work from home. We believe the joint action from the public, the government and the police will smooth the transition to the future stages of the pandemic.

References

2010-2020 Crime and Law Enforcement in LA County. Los Angeles County Sheriff's Department. Retrieved May 9, 2020, from
http://shq.lasdnews.net/CrimeStats/CAASS/desc.html

2010-2020 Unemployment rate and population by city in LA County. Bureau of Labor Statistics. Retrieved May 9, 2020, from
https://www.bls.gov/web/metro/laucntycur14.txt

2020 COVID-19 dataset by city in LA County. County of Los Angeles Public Health. Retrieved May 9, 2020, from
http://dashboard.publichealth.lacounty.gov/covid19_surveillance_dashboard/

Brooks, Rosa. 2020, April 24. Police officers nationwide need to consider going hands-off during this crisis. *The Washington Post*. https://www.washingtonpost.com/opinions/2020/04/24/this-pandemic-standard-police-practices-risk-spreading-infection/

Identifying the numbers of AR or MA terms in an ARIMA model. Accessed May 12, 2020, from http://people.duke.edu/~rnau/411arim3.htm

Hyndman, Rob J and George Athanasopoulos. 2018. *Forecasting: Principles and Practice*. https://otexts.com/fpp2/

Supplement Information

Our research categories of crime were adapted from the LA Sheriff Department. The detailed keywords for each category are listed below:

Category	Description
Property	Burglary, Larceny Theft, Grand Theft Auto, Arson, Federal Offenses with Money, Fraud and NSF Checks
Traffic	Vehicle / Boating Laws, Accidents Traffic / Vehicle / Boat, Drunk Driving Vehicle / Boat
Violence*	Aggravated Assaults, Non-aggravated Assaults, Weapon Laws, Robbery, Criminal Homicide, Federal Offenses without Money
Vandalism	Vandalism
Domestic Violence	Offenses against Family, Suicide and Attempt
Drink Drug	Narcotics, Drunk / Alcohol / Drugs, Liquor Laws
Sex	Sex Offenses Felonies, Sex Offenses Misdemeanors, Forcible Rape
Misdemeanor	Disorderly Conduct, Vagrancy, Gambling, Misdemeanors Miscellaneous, Mentally Ill, Felonies Miscellaneous
Other	Receiving Stolen Property, Juvenile Non-criminal, Persons Missing, Forgery, Persons Dead, Accidents Miscellaneous, Commitments, Miscellaneous Non-criminal, Warrants
* Domestic violence and child abuse are subtracted from violence.	