A longitudinal analysis of the crime rate in London: to what extent socioeconomic factors can explain violence in London.

1. Data source, application domain and analytical questions

Introduction

In the last 25 years there was a significant decrease in crime rate in most of the western countries and big cities including London. According to Mailley et al. (2010) such a major decrease in crimes still cannot be fully explained by the scientists, even though crime is a widely researched area.

However, in the last few years London crime and specifically violent crime have been back to the newspaper headlines. According to Dodd and Grierson (2018) violent death is on its peak since 2009. As a result, understanding what caused such a big increase in violence and how to deal with it became critically important.

Analytical questions and scope of work

The scientists name a lot of possible reasons for a violent crime and among all, socioeconomic group of factors such as poverty or immigration is one of the most heavily investigated areas on both local and global scales.

There are many socioeconomic indicators, however this paper would only concentrate on the following factors: poverty and income inequality (Coccia, 2017), unemployment (Lobont et al, 2017), immigration (Bianchi, Pinotti and Buonanno, 2012) and alcohol availability (Brempong, 2006).

This paper aims to address the following questions:

- What is going on with overall crime rate in London? Are there any trends?
- Is violent rate growing in London? How violence trend differs from overall crime trend?
- What are the relationship between socioeconomic factors and violent crime in London?

It is obvious that to some extent crime and violence are random factors and hard to predict but understanding what can influence it might help authorities to deal with violent crime in a more effective way.

Data source

The following datasets are used:

- Recorded Crime: Borough Rates (main dataset used in our analysis)
- Income Support Claimants (as indicator of poverty)
- Job Seekers Allowance (as indicator of unemployment)
- Personal Insolvency Statistics (as a possible socioeconomic indicator)
- Total affordable housing completions (as a possible socioeconomic indicator)
- National Insurance Number registrations to adult overseas nationals entering the UK (as indicator of immigration)
- House Price/Earnings ratio (as indicator of income inequality)
- Proportion of pubs/bars and clubs to all food and drinks businesses (as indicator of alcohol availability).

Source of data: London Datastore - https://data.london.gov.uk/

2. Analysis Strategy and Plans

In order to answer the questions raised in this paper the analysis would consist of two parts:

Part 1 – Investigating overall crime and violent crime patterns in London between 1999 and 2016.

Part 2 – Building multiple linear regression models (Ordinary Least-Squares Regression) to test the relationship between violence and socioeconomic factors.

Part 1:

- After loading London Crime Data in Python notebook all the required data wrangling activities will be taken including removing unnecessary rows and columns, handling missing values, indexing and data frame reshaping.
- Crime data will be split in to ten data frames according to the type of crime.
- First this paper will investigate the patterns of the overall crime (all offences) in different boroughs and how it changed between years 1999 and 2016.
- Then the analysis will concentrate on understanding the variations in types of crime as a proportion of total crime with a specific attention to violence against the person (violence).
- Also the paper will investigate the year to year violence trend in comparison to overall crime rate. A simple linear regression model will be build to test the change in all offences and violence.

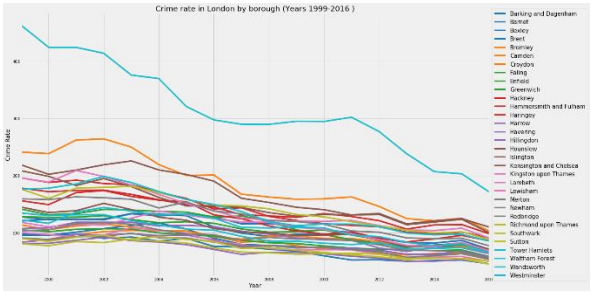
Part 2:

- Loading the necessary datasets to extract the required socioeconomic factors.
- Cleaning, melting and reshaping data to make it suitable for the required analysis.
- Once all the required data is merged all attributes will be checked for multicollinearity.
- Finally, preparation and building a various multiple regression models and evaluation of the results.

3. Findings and reflections

As can be seen in figure 1 the initial analysis has shown that for majority of boroughs there is a clear trend to decrease in overall crime rate (all offences). Westminster has shown the highest 'all offences' rate for the whole period and was an obvious outlier in comparison to all other boroughs. It can be explained by the fact that Westminster is more a tourist destination rather than a living area, so it attracts a lot of criminals. However, at the same time Westminster has demonstrated the highest drop in crime (63%) between years 1999 and 2016. One interesting observation is that predominantly the higher was the initial crime rate the more significant was the percentage drop (e.g. Camden).

Figure 1: overall crime rate by borough (year 1999-2016)

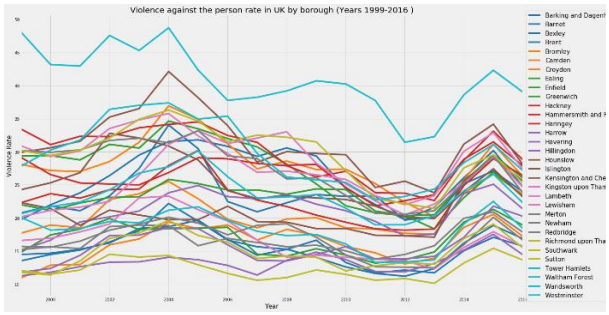


Looking at the different types of crime In London it is seen that in year 1999 Theft and Handling was the highest in comparison to all other crime types for most of boroughs. However by year 2016 we can see a big increase in violence against the person (violence) as a proportion of overall crime rate (it has gone up from 15 % in year 1999 to 31 % in year 2016). This explains such a big attention to the violent crime in media in recent years.

The time series analysis of the change in violence by borough (figure 2) did not show

any clear trends especially if you compare it to overall crime rate where a downtrend was obvious. The simple linear regression model for overall crime had majority of fitted coefficients clearly less than one. It proves a decrease in overall crime rate over the years. However, testing the violence rate has shown a mix of < 1 and > 1 coefficients, which implies that the violence rate is jumping up and down from year to year. The fact that average coefficient is 1.0022 means that there was a slight increase in violence over the whole period.

Figure 2: Violence rate by borough (Year 1999-2016)



In the second part of the analysis we concentrated on the violent crime and its relationship with chosen socioeconomic factors. No multicollinearity between the variables have been identified so all the factors were kept.

Two regression models have been built, one model included only some of the predictor factors (semi-saturated model) and the other model included all the variables (Saturated model). Semi-saturated model (figure 3) has shown (as per the p-values) that majority of variables are highly significant (< 0.05) in their impact on the violent crime. The variables 'Proportion of pubs/bars ' and 'Affordable Housing Supply' were found insignificant (>0.05) in that model. It is important to mention that it doesn't mean that those two socioeconomic factors are not important for determining violence rate as p-values presented in the model are calculated assuming that all other variables are already included in the model. So the same variables can be found significant in a smaller models with different set of features.

Figure 3: Semi-saturated regression model

OLS Regression Results

Dep. Variable:	Violence Rate	R-squared:	0.440
Model:	OLS	Adj. R-squared:	0.433
Method:	Least Squares	F-statistic:	61.95
Date:	Wed, 11 Dec 2019	Prob (F-statistic):	1.52e-56
Time:	22:19:27	Log-Likelihood:	-1487.3
No. Observations:	480	AIC:	2989.
Df Residuals:	473	BIC:	3018.
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	13.1544	1.807	7.279	0.000	9.603	16.706
Job Seekers Allowance Claimants (Rate)	-1.0032	0.273	-3.674	0.000	-1.540	-0.467
National Insurance Number Registrations (Immigration)(Rate)	0.1369	0.013	10.210	0.000	0.111	0.163
Income Support Claimants (Rate)	1.1834	0.121	9.784	0.000	0.946	1.421
Affordable Housing Supply	0.0006	0.001	0.622	0.534	-0.001	0.003
Proportion of pubs/bars	0.0959	0.068	1.420	0.156	-0.037	0.229
Proportion of clubs	-0.5352	0.188	-2.849	0.005	-0.904	-0.166

Omnibus:	46.776	Durbin-Watson:	1.457
Prob(Omnibus):	0.000	Jarque-Bera (JB):	64.163
Skew:	0.715	Prob(JB):	1.17e-14
Kurtosis:	4.077	Cond. No.	3.45e+03

In figure 3 we can observe R^2 to be higher than 40 %, which means that more than 40% of variation in violent crime rate can be explained by the model. This is impressive given that not all predictors were included.

In the saturated model (figure 4) we've observed that some of the factors that in semi-saturated model were shown to be not significant for violent crime in fully saturated model became significant and vice versa. This proves the point that significance of the variables sometimes depends on the presence of other variables in the model.

Figure 4: Saturated regression model.

OLS Regression Results

Dep. Variable:	Violence Rate	R-squared:	0.517
Model:	OLS	Adj. R-squared:	0.508
Method:	Least Squares	F-statistic:	55.89
Date:	Wed, 11 Dec 2019	Prob (F-statistic):	9.78e-69
Time:	22:36:28	Log-Likelihood:	-1451.8
No. Observations:	480	AIC:	2924.
Df Residuals:	470	BIC:	2965.
Df Model:	9		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-0.5534	2.464	-0.225	0.822	-5.395	4.289
Job Seekers Allowance Claimants (Rate)	-0.6342	0.290	-2.188	0.029	-1.204	-0.065
National Insurance Number Registrations (Immigration)(Rate)	0.0879	0.014	6.157	0.000	0.060	0.116
Income Support Claimants (Rate)	1.4943	0.134	11.161	0.000	1.231	1.757
Affordable Housing Supply	0.0022	0.001	2.194	0.029	0.000	0.004
Proportion of pubs/bars	0.1452	0.067	2.172	0.030	0.014	0.277
Proportion of clubs	0.0965	0.190	0.507	0.612	-0.277	0.470
Ratio of House Prices to Earnings	0.8931	0.109	8.223	0.000	0.680	1.106
Personal Insolvency (Rate)	0.1819	0.079	2.289	0.023	0.026	0.338
Personal Bankruptcy (Rate)	-0.3315	0.122	-2.709	0.007	-0.572	-0.091

Omnibus:	32.629	Durbin-Watson:	1.480
Prob(Omnibus):	0.000	Jarque-Bera (JB):	43.756
Skew:	0.545	Prob(JB):	3.15e-10
Kurtosis:	3.999	Cond. No.	5.04e+03

Since more independent variables are included in the above model it was expected for R^2 to be higher. In this case R^2 shows that we can explain over half the variance in violent crime rate using the listed factors above. However Adj. R^2 (> 50 %) has also gone up, which means

that additional factors improves the model explanation of violence.

Checking for normality of residuals we found that the residuals of both models a normally distributed which validates our models. The results show that the violence crime rate is inherently complicated and may not be able to be fully explained by the factors included. This is not surprising as violent crime is known to be unpredictable and in some cases random.

The removal of Westminster borough as outlier from our analysis has led to some further improvement in both models performance.

4. Limitations and future work

The data was used at the borough level which might be not detailed enough. It would be interesting to look at the crime patterns at a finer level than boroughs (e.g. wards).

Also due to the limited size of the analysis this paper does not fully examine the actual socioeconomic factors that were used in the analysis. It is important to investigate each factor in more details.

It is crucial to continue investigating how socioeconomic factors relate to a violent crime with a more detailed look at the regression models. Since there are many different types of data involved, we can do a log transformation of the features to see if it has any effect on the model performance.

5. Conclusion

It has been observed that overall crime rate has gone down over the past 15 years and London became a much safer city. However increase in violent crime rate is not just a media speculation, it is a reality as shown by our analysis. Violence has completely different patterns and trends than overall crime rate. Therefore it is crucial to continue investigating violent crime and its association with socioeconomic factors, which were shown to be important in explaining violence, so it is definitely an interesting area for the future exploration.

References:

- Barr, C., Kommenda, N., Voce, A. and Ibbetson, C. (2019). https://www.theguardian.com/uk-news/2018/apr/27/why-are-knife-and-gun-offences-on-the-rise-and-who-is-most-at-risk.
- Bianchi, M., Buonanno, B., Pinotti, P. (2012) 'Do immigrants cause crime?', Journal of the European Economic Association December, 2012 10(6):1318–1347
- Brempong, K.G. (2012) 'Neighborhood income, alcohol availability, and crime rates', The Review of Black Political Economy/Winter.
- Coccia, M. (2017) 'A Theory of general causes of violent crime: Homicides, income inequality and deficiencies of the heat hypothesis and of the model of CLASH', Aggression and Violent Behavior 37, 190–200
- Cohen, J., Cohen, P., West, S.G., and Aiken, L.S. (2002) 'Applied Multiple Regression/Correlation Analysis for the Behavioural Sciences', Routledge, Mahwah.
- Datastore – Greater London Authority, Available at https://data.london.gov.uk/
- Lobonţ, O.R., Nicolescu, A.C., Nicoleta-Claudia Moldovan, N.C. and Kuloğlu, A. (2017) 'The effect of socioeconomic factors on crime rat
- es in Romania: a macro-level analysis', Economic Research, 30:1, 91-111.
- The Met, Available at https://www.met.police.uk.
- Tseloni, A., Mailley, J., Farrell G. and Tilley, N. (2010) 'Exploring the international decline in crime rates', European Journal of Criminology, 7(5) 375–394