Crime Data Mapping and Spatial Regression

Mudabbira Mushtary
University of Bonn

March 31, 2023

Point Patterns of Burglary Incidents

- **Kernel density estimates:** Counting the number of points in a continuous way.
- Burglaries are centered around the city center.

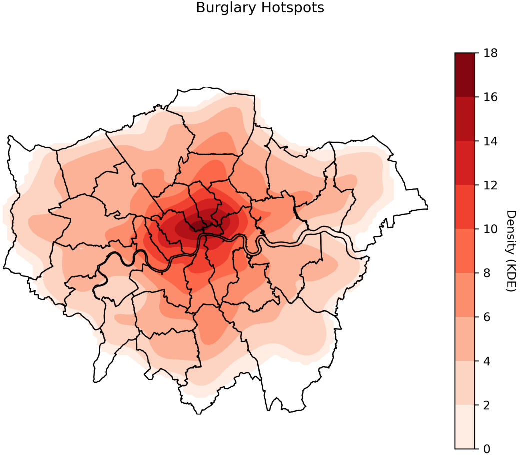


Figure 1: Burglary hotspots within greater London Area.

Spatial Clustering of Burglary Incidents

- **DBSCAN:** density-based clustering
- **Assumption**: clusters are dense regions in space separated by regions of lower density
- Burglaries are centered around the city center.

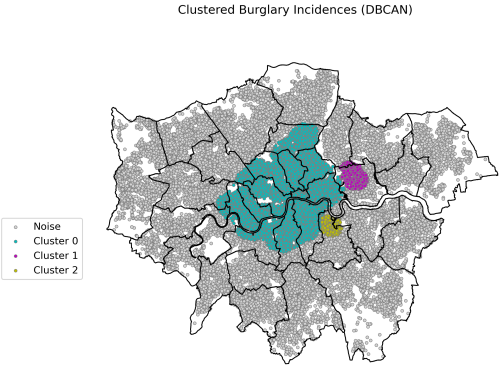


Figure 2: Burglary clusters within greater London Area.

Spatial Weights Matrix

- A spatial weights matrix is the way geographical space is formally encoded into a numerical form so it is easy for a computer (or a statistical method) to understand.
- We can define a spatial weights matrix, such as contiguity, distance-based, or block. The following analysis has used a Queen contiguity matrix.
- **Queen Contiguity Matrix:** For a pair of local authorities in the dataset to be considered neighbours under this W, they will need to be sharing border or, in other words, "touching" each other to some degree.

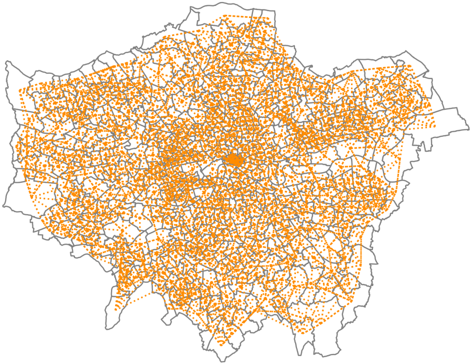
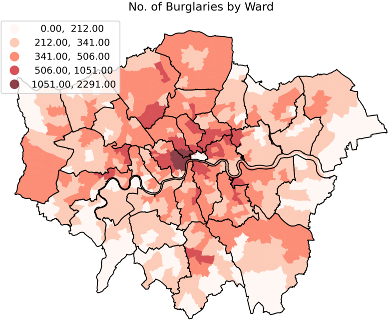


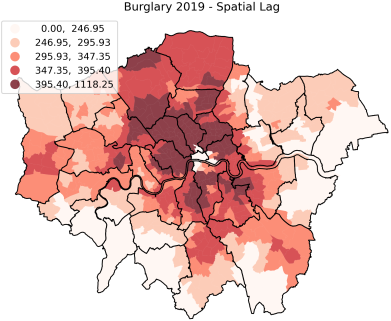
Figure 3: Burglary clusters within greater London Area.

Spatial Lag of Burglary Incidents

Spatial lag is a spatially weighted average of the values of a variable in the neighbourhood of each observation.Once we have weights matrix, we can look at the spatial lag of the burglary in the Greater London area



(a) No. of Burglary Incidents (2019)

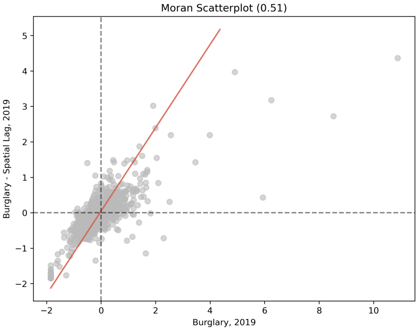


(b) Burglary Lag (2019)

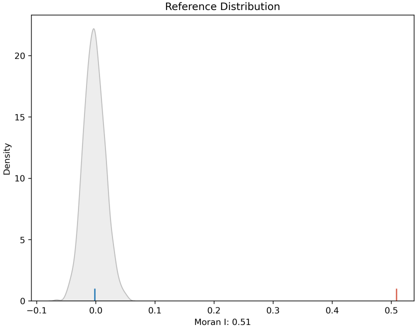
Figure 4: Burglary incidents and the corresponding lag on a ward level.

Moran Plot

- The plot displays a positive relationship between the standardized rate of burglary and its spatial lag.



(a) Moran's Plot



(b) Distribution after random Sampling

Figure 5: Moran's I and Global Spatial Autocorrelation

Moran Plot

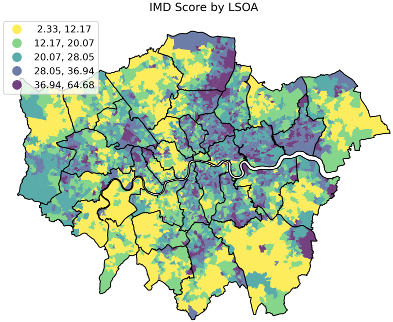
- Moran plot is a way of visualizing a spatial dataset to explore the nature and strength of spatial autocorrelation.
- This is associated with the presence of **positive spatial autocorrelation:** similar values tend to be located close to each other. This means that the **overall trend** is for high values to be close to other high values, and for low values to be surrounded by other low values.

Moran's I and Spatial Autocorrelation

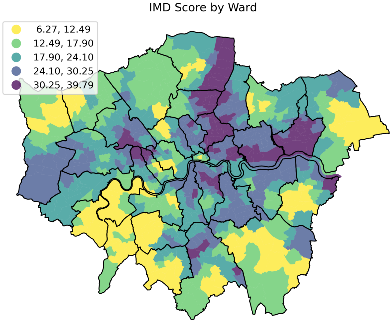
- **Moran's I** is a measure of **spatial autocorrelation**.
- From the figure 5a, we can see that the value of the Moran's I is 0.5, which is a positive value.
- With the small p-value associated with the Moran's I, we can conclude that the map displays more spatial pattern that we would expect if the values had been randomly allocated to a particular location.
- **Global Spatial Autocorrelation** relates to the overall geographical pattern present in the data.
- Global autocorrelation analysis shows us that observations do seem positively correlated over space.

Resource Deprivation in Greater London Area

Resource deprivation scores on a LSOA level as shown in Figure 6a have been aggregated to a ward level for subsequent spatial regression analysis as shown in Figure 7b.



(a) Indices of Multiple Deprivation LSOA level

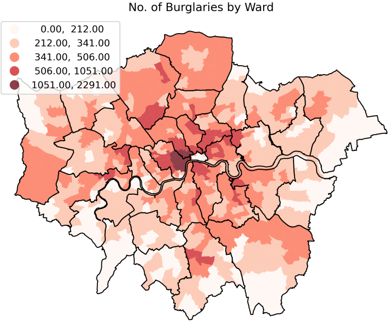


(b) Indices of Multiple Deprivation on ward level

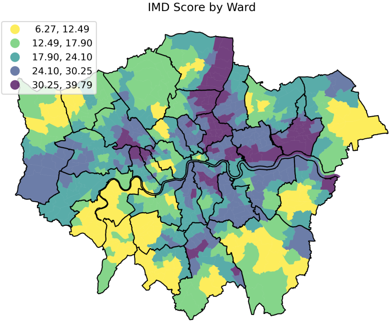
Figure 6: English Indices of Multiple Deprivation (IMD)

Spatial Regression: Burglaries and Resource Deprivation

Number of burglaries a shown in in Figure 7a and resource deprivation as shown in Figure 7b are used for spatial regression analysis on a ward level.



(a) Indices of Multiple Deprivation LSOA level



(b) Indices of Multiple Deprivation on ward level

Figure 7: English Indices of Multiple Deprivation (IMD)

Spatial Regression

	Independent Variable	Coefficient	t-Statistic	Probabilty	Dependent Variable
0	CONSTANT	7.913793e-17	2.336944e-15	1.000000e+00	burglaryRate2019
1	IncScore	7.152022e-02	4.229583e-01	6.724671e-01	burglaryRate2019
2	EmpScore	1.516256e-01	9.474071e-01	3.437873e-01	burglaryRate2019
3	EnvScore	4.559383e-01	1.115281e+01	1.575439e-26	burglaryRate2019
4	BHSScore	-3.098920e-01	-7.037613e+00	5.027381e-12	burglaryRate2019
5	EduScore	-3.807009e-02	-6.756828e-01	4.994850e-01	burglaryRate2019

Table 1: Estimation results of regression.

- **Independent variables:** Resource deprivation scores.
- **Dependent Variable:** Burglary rate in GreaterLondon Area .
- **Null hypothesis:** No Spatial autocorrelation in the residuals.
- **Alternative hypothesis:** There is spatial autocorrelation in the residuals.
- **Conclusion:** The null hypothesis is rejected. There is spatial autocorrelation in the residuals.

Spatial Regression

Table 2: Spatial Diagnostic Tests

	Lagrange Multiplier (error)	Lagrange Multiplier (lag)	Robust LM (error)	Robust LM (lag)	Moran's I
Value	5.710069e+02	4.870404e+02	8.698442e+01	3.017904	4.503319e-01
p-value	3.390663e-126	6.277545e-108	1.093795e-20	0.082350	2.490353e-137

- **Lagrange Multiplier Tests** suggests that a spatial error or spatial lag model would improve the fit of our model.
- We have got significant results for both LM error and LM lag. This shows that we should consider the spatial correlation.
- However, according to the Robust LM and Robust LM(error) values, we should only consider the Spatial Error model as Robust LM is not significant anymore.
- We will continue with the spatial error and spatial lag models even though the error model is more than likely the correct path.

Spatial Regression: ML_Lag

$$y_i = X_i\beta + \lambda w_i \epsilon_i + u_i \qquad (1)$$

- **Spatial Lag Model** is a global model where the dependent variable among where neighbors influence the dependent variable.
- For Example, If burglary incidence is higher in the neighborhood of the City Centre of London, its highly likely to have high burglary incidence in the city center of London. Therefore, there is a feedback loop that occurs where affects on our neighbor(s) y affects our y and our neighbor(s) y variable.
- **Null hypothesis:** Spatial dependencies are not captured by the spatial lag Wy in the endogenous variable Y
- **Alternative hypothesis:** Spatial dependencies are captured by the spatial lag Wy in the endogenous variable Y

Spatial Regression: ML_Lag

	Independent Variable	Coefficient	z-Statistic	Probabilty	Dependent Variable
0	CONSTANT	0.029678	1.146564	2.515619e-01	burglaryRate2019
1	IncScore	-0.136874	-1.123234	2.613379e-01	burglaryRate2019
2	EmpScore	0.068506	0.593948	5.525471e-01	burglaryRate2019
3	EnvScore	0.241367	6.460562	1.043147e-10	burglaryRate2019
4	BHSScore	-0.080167	-2.452719	1.417812e-02	burglaryRate2019
5	EduScore	0.080989	1.997165	4.580727e-02	burglaryRate2019
6	W_burglaryRate2019	0.847833	35.295624	6.854114e-273	burglaryRate2019

Table 3: Estimation results of ML_Lag regression.

- **W_burglaryRate2019** is a spatially lagged y multiplier.
- Coefficient is 0.84 (with a significant p-value)
- Burglary incidence tends to higher in an area when it is higher in the neighborhood with some spillover effects.

Spatial Regression: ML_Error

$$y_i = X_i\beta + \lambda w_i \epsilon_i + u_i \tag{2}$$

- **Spatial Error Model** does not include lagged dependent or independent variables, but instead includes a function of our unexplained error and that of our neighbors.
- Higher than expected residual values suggest a missing explanatory variable that is spatially correlated. In this analysis, Lambda is the error multiplier.
- **Null hypothesis:** There is no dependency of error values of an area with errors in other areas associated with it
- **Alternative hypothesis:** There is dependency of error values of an area with errors in other areas associated with it

Spatial Regression: ML_Error

	Independent Variable	Coefficient	z-Statistic	Probabilty	Dependent Variable
0	CONSTANT	0.358446	2.006287	4.482566e-02	burglaryRate2019
1	IncScore	0.150740	0.711830	4.765700e-01	burglaryRate2019
2	EmpScore	0.026335	0.137548	8.905978e-01	burglaryRate2019
3	EnvScore	0.688048	9.886850	4.747296e-23	burglaryRate2019
4	BHSScore	-0.391575	-6.094279	1.099321e-09	burglaryRate2019
5	EduScore	-0.041581	-0.559195	5.760286e-01	burglaryRate2019
6	lambda	0.867445	35.846630	2.075406e-281	burglaryRate2019

Table 4: Estimation results of ML_Error regression.

- Lambda coefficient is 0.86 with a significant p-value.
- Thus, we reject the null hypothesis and conclude that there is dependency of error values of an area with errors in other areas associated with it.

Model Compasiron and Conclusion

	MAXIMUM LIKELIHOOD SPATIAL LAG (METHOD = FULL)
Pseudo R-squared	0.626991
Spatial Pseudo R-squared	0.626991
Log likelihood	-652.866938
Schwarz criterion	1351.061904

Table 5: Statistics of ML_Lag model.

	MAXIMUM LIKELIHOOD SPATIAL ERROR (METHOD = FULL)
Pseudo R-squared	0.253184
Spatial Pseudo R-squared	0.253184
Log likelihood	-632.844586
Schwarz criterion	1304.541768

Table 6: Statistics of ML_Error model.

Model Compasiron and Conclusion

- We got lower log-likelihood for the Error model.In addition, Schwarz criterion of the error model is also lower than the lag model.
- So, we can conclude that this research is missing spatially correlated independent variable.
- Further work on this area is necessary to find the correct explanatory variables that are more appropriate for explaining burglary incidence in Greater London Area.

Resources

Raw data was gathered from the following sources:

- Crime data is taken from `https://data.police.uk/data/`
- Deprivation data is taken from `https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019`

Methods were selected from the following sources:

- Geospatial analysis: Lawhead (2019)
- Spatial Econometrics: Anselin (1988) and Kopczewska (2020)

LaTeX template was taken from von Gaudecker (2019) and modified.

References I

Anselin, Luc (1988). *Spatial econometrics: methods and models*. Vol. 4. Springer Science & Business Media.

Gaudecker, Hans-Martin von (2019). "Templates for Reproducible Research Projects in Economics". `https://doi.org/10.5281/zenodo.2533241`.

Kopczewska, Katarzyna (2020). *Applied spatial statistics and econometrics: data analysis in R*. Routledge.

Lawhead, Joel (2019). *Learning Geospatial Analysis with Python: Understand GIS fundamentals and perform remote sensing data analysis using Python 3.7*. Packt Publishing Ltd.