WORD FILE I

Alexander Konkin

BPS212

Block 1. Introduction

1. In this work I would like to explore the possible connections between number of crimes committed in the area and election results in the respective area. In order to do it I will investigate the case of the UK and its local authorities. I expect that the authorities where more crimes are committed tend to vote more conservatively. As the Conservative Party stance on police funding and crime are harsher rather than Labour – their main competitors. In order to understand whether it is really so I will explore the data for year 2021 collected by the Office for National Statistics.
2. The data comes from three datasets that I have called `crime\_statistics`, `elections\_results` and `econ\_stat`. In `crime\_statistics` variables stand for number of registered offences (types are in the names of variables). For my research I will mostly use such variables as `Drug offences`, `Possession of weapons offences` and `Theft offences` which all stand for number of respective crimes in the local authority. I use these variables as they are on one level of analysis – they stand for groups of crimes and not exact ones as well as might be influential on elections results from theoretical point of view. In `elections\_results` the results of local elections per ward are listed. In order to make this dataset comparable to previous I had to make a sum of all votes per ward so that I can get a summary of votes for parties per local authority. From `econ\_stat` I use the only variable, namely GDP per capita in pounds for year 2021.

Block 2. Table with descriptive statistics

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Descriptive statistics table** | | | | |
|  | | | | |
| Statistic | St. Dev. | Min | Max | Median |
|  | | | | |
| Conservative vote share (%) | 11.48 | 4.87 | 68.72 | 43.72 |
| Labour vote share (%) | 13.69 | 2.88 | 65.38 | 26.82 |
| Theft offences | 3,641.67 | 648 | 26,578 | 2,421.5 |
| Drug offences | 606.96 | 44 | 6,141 | 289 |
| Possession of weapons offences | 155.93 | 13 | 1,056 | 79 |
| GDP/capita (£) | 10,249.20 | 14,383 | 82,164 | 26,030.5 |

Block 3. Table with the results of the OLS regressions and its interpretation

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Regression Results** | | | | |
|  | | | | |
|  | *Dependent variable:* | | | |
|  |  | | | |
|  | Conservative vote share | | | |
|  | | | | |
| Theft offences | -0.002\*\*\* | -0.001\*\*\* | -0.001\*\* | -0.002\*\*\* |
|  | (0.0002) | (0.0003) | (0.001) | (0.001) |
|  |  |  |  |  |
| Drug offences |  | -0.006\*\*\* | -0.006\*\*\* |  |
|  |  | (0.002) | (0.002) |  |
|  |  |  |  |  |
| Possession of weapons offences |  |  | 0.010 |  |
|  |  |  | (0.014) |  |
|  |  |  |  |  |
| GDP/capita |  |  |  | -0.0001 |
|  |  |  |  | (0.0001) |
|  |  |  |  |  |
| Interaction effect |  |  |  | 0.00000 |
|  |  |  |  | (0.00000) |
|  |  |  |  |  |
| Constant | 48.687\*\*\* | 48.655\*\*\* | 48.733\*\*\* | 53.091\*\*\* |
|  | (1.054) | (1.030) | (1.036) | (3.241) |
|  |  |  |  |  |
|  | | | | |
| Observations | 180 | 180 | 180 | 180 |
| R2 | 0.248 | 0.285 | 0.288 | 0.256 |
| Adjusted R2 | 0.243 | 0.277 | 0.275 | 0.244 |
| Residual Std. Error | 9.986 (df = 178) | 9.760 (df = 177) | 9.772 (df = 176) | 9.983 (df = 176) |
| F Statistic | 58.557\*\*\* (df = 1; 178) | 35.317\*\*\* (df = 2; 177) | 23.676\*\*\* (df = 3; 176) | 20.233\*\*\* (df = 3; 176) |
|  | | | | |
| *Note:* | \*p\*\*p\*\*\*p<0.01 | | | |

Four models are presented here: three with different amounts of variables and one interaction model. According to Adjusted R2 the model of best fit is the second model with 27,7% of explanation and Drug offences and Theft offences independent variables. Other models have worse results. Also the AIC and BIC has the lowest result for 2nd model which means that it has the biggest predictive capabilities as well.

1. Are there any statistically significant independent variables in the model of the best fit?

Yes, both Drug offences and Theft offences are statistically significant with the level of confidence of bigger than 99%.

2. What is the direction of relationship between each of the independent variables and dependent

variable and what does the estimate for each of the IVs essentially mean?

The relation is negative for both variables.

A 1% increase in the Theft offences is associated with an average decrease of 0.1% in the Conservative party result, all things being equal (p-value = 0.00566).

A 1% increase in the Drug offences is associated with an average decrease of 0.6% in the Conservative party result, all things being equal (p-value = 0.00260).

3. What is the predictive capability of the model of the best fit?

The explanatory power of second model is 27,7% according to the adjusted R2.

4. What does F-statistic say about the model’s characteristics?

The F-statistic has the p-value of 1.24e-13 that signalizes that the overall model is statistically significant.

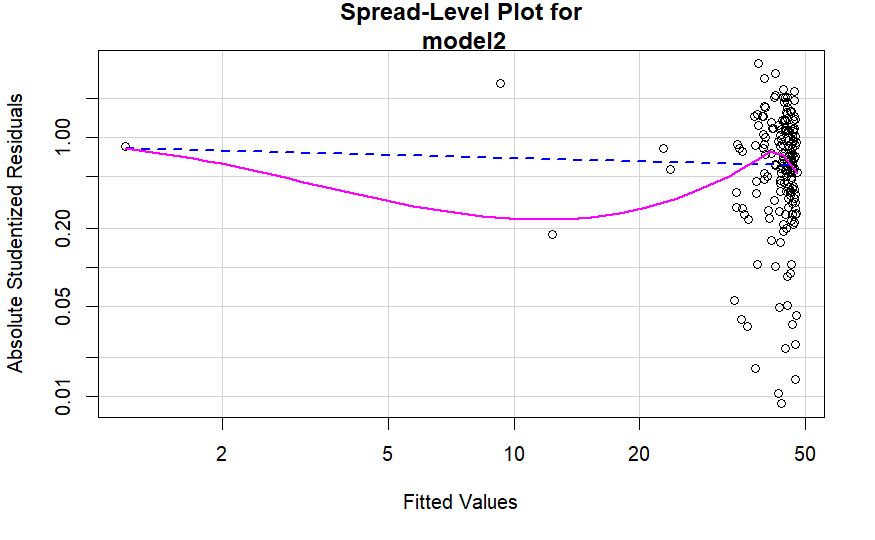
Block 4. OLS regression diagnostics

1. Multicollinearity

|  |  |
| --- | --- |
| **Multicollinearity analysis** | |
|  | |
| `Theft offences` | `Drug offences` |
|  | |
| 2.344 | 2.344 |
|  | |

As the values for both variables are smaller than 10, we can say that we do not face a multicollinearity problem in this model.

2. Heteroscedasticity



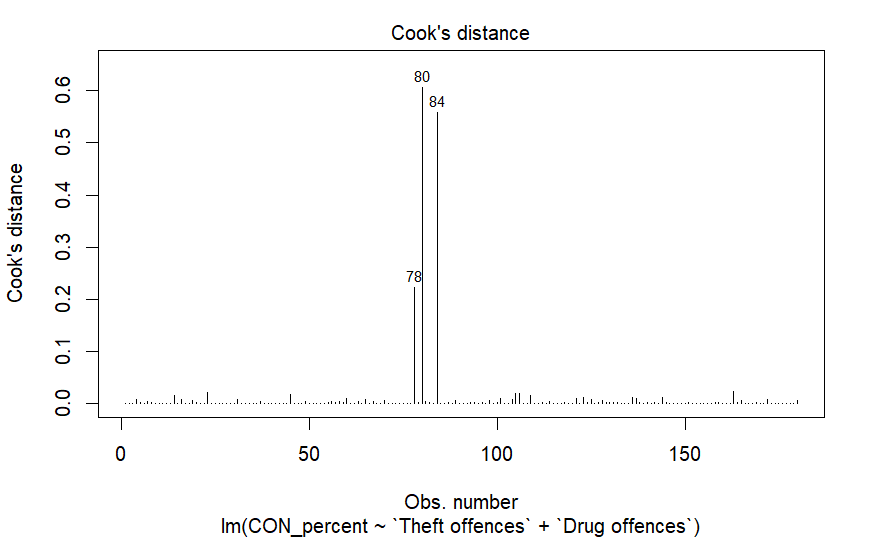
The p-value is bigger than 0,05 (0.21613), therefore there is no heteroscedasticity in the model.

However, the plot shows that the observations are concentrated at one place and therefore it is probably not the best situation. The pink line goes a lot from the blue one and the blue is almost horizontal. Therefore, I would assume that the problem is present.

3. Outliers

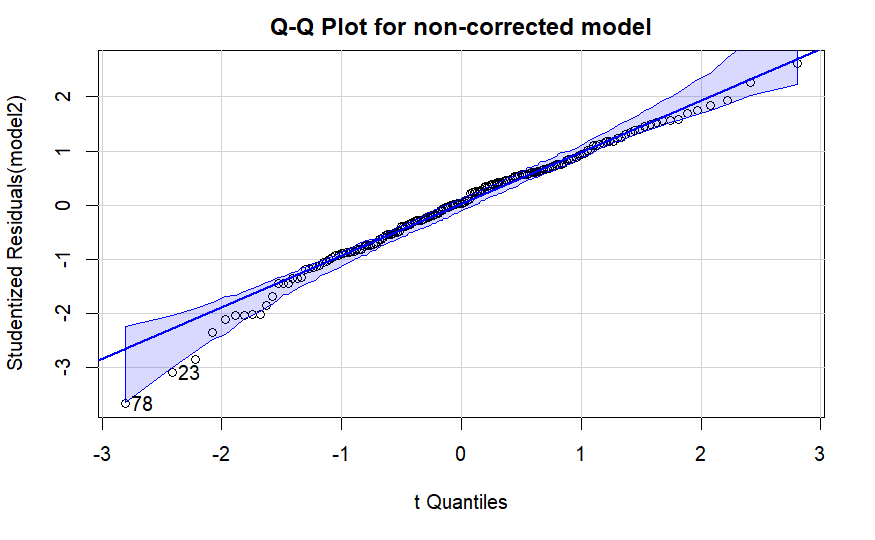
The test has shown the only outlier – observation 78 that stands for Knowsley. It is a Liverpool surbub that is a Labourist stronghold. However, as we will see later it is not an influential observation therefore it does not really harm the model.

4. Influential observations



We can see that there are three observations that stand out. However, they do not cross the Cook’s distance line and cannot be called influential. They stay for Leeds and Liverpool – urban Labour strongholds.

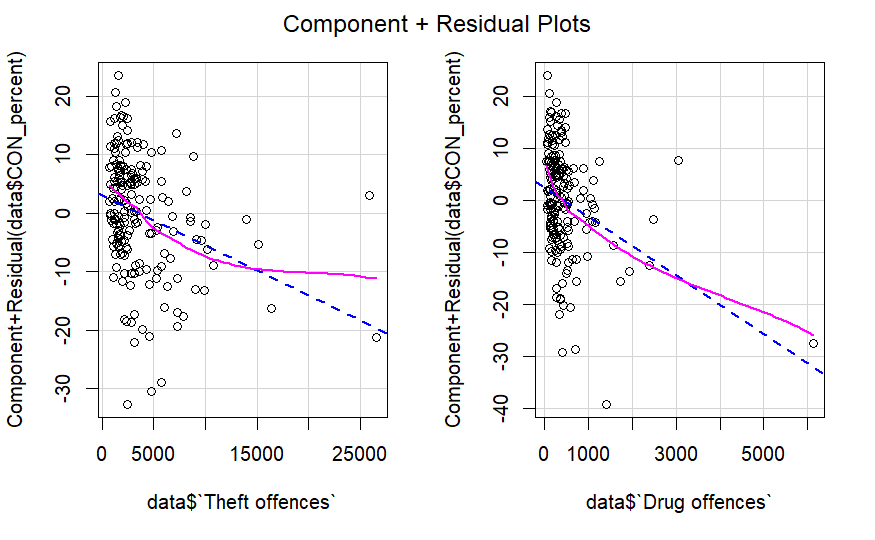
5. Non-normality of regression residuals distribution



The problem exists as the p-value for lambda(1) is smaller than 0,05 (0.0010587) and therefore we can reject the H0 (that residuals are normally distributed) so they are not normal and the problem exists.

According to the graph while some observations like 78 and 23 are out of range the majority stand within the range and we can assume that the problem is not present. However, according to the formal test the problem does exist.

6. Non-linearity analysis



P-value is >0.05 for `MLE of lambda` of the both IV (0.1397 and 0.1313), therefore the problem of non-linearity statistically does not exist.

The visualization shows us that the pink line does not correspond the blue line in either plot so I would assume the presence of the problem, however it is absent.