Exploring Economical Influences on Germany's Crime Rates 1976-2020

Seyda Betul Aydin [*1] **Csenge Frater** [*2] **Melis Oktayoglu** [*3]

Abstract

Understanding crime is an important step in prevention and prediction of its occurrence. By properly doing so we can alleviate a huge burden from society, considering not just financial concern but also the state of general well-being. In this project paper we are attempting to better understand the phenomena by looking into German crime data Federal Office of Statistics (Statistisches Bundesamt), covering years 1976 to 2020 with annual aggregation. First we turned to clustering methods to better understand the relationship between the different kinds of criminal activities, then after observing a peak around 2007 on our data, lead by the notoriously known Great Recession, we looked at economical factors from the World Bank in the hope of better explainability.

1. Contribution Plan

Literature research, limitations, and manuscript revision was done by Csenge Frater. Melis Oktayoglu contributed to data description, discussion and correlation analysis. Most of the analysis and the discussion was made by Seyda Betul Aydin. All members equally contribute to initial exploration of the data and the final evaluation of the methods and results.

Code for our analysis can be found on github.

2. Introduction

2.1. Influences on crime

Crime is a complex phenomenon driven by multiple factors. Common sense and clear evidence point towards socio-economic indicators as the main candidate. More precisely, for example low legitimate earning prospects, joblessness, and greater inequality seem to be related to higher rates [1].

But even those can be inconsistent in some cases. One such counter example is the Great Recession in the United States and most developed countries, which with its financial collapse unexpectedly curbed crime rates. According to Rosenfeld [2] this could be explained away by inflation as an intermediate variable. If inflation is present in addition to recession, the value of goods rise with an incentive on theft, which introduces a domino-like effect of pushing other kinds of crimes up. If the value of money doesn't drop significantly, increase of criminal activity cannot be observed either. This highlights the fact that these factors are not invariant, can have a complex relationships hard to disentangle.

2.2. Looking at the individual level - Rational choice theory

Ehrlich, Block, Heineke, and Wolpin [4] introduced a theoretical framework for predicting crime rates, which looks at the expected utility of a person allocating time on illegal and legal acts. Each individual is assumed to maximise expected utility, i.e. it depends on their initial wealth, legal income, as well as legal income after previous conviction, weighted by the probability of detection and punishment. This also highlights the fact that the time spent on illegal activity highly depends on the person's monetary state, according to economic models.

2.3. Problems with handling crime data

One paper [5] specifically addresses the problems with the German Police Crime Statistics. The main problem with this dataset -, and usually with crime data in general - is that it contains only the officially reported cases. This problem could be solved partly by the so-called "dark-figure" studies with subjective narratives and large sample sizes, where the reported crime proportions could be matched to the officially recorded data. However, this method is fairly biased due to subjective experiences of victimization and relies on only the individual's memory. Also these studies have a minimum age requirement with not institutionalized and convicted people, which rules out the possibility of a representative measure. Unfortunately with our limited resources we couldn't sooth the bias of under-reporting by

[*]Equal contribution [1]6323572, seyda-betul.aydin@student.uni-tuebingen.de, MSc. Quantitative Data Science [2]6639491, csenge.frater@student.uni-tuebingen.de, MSc. Machine Learning [3]6620677, melis.oktayoglu@student.uni-tuebingen.de, MSc. Machine Learning.

Project report for the "Data Literacy" course at the University of Tübingen, Winter 2023/24 (Module ML4201). Style template based on the ICML style files 2023.

victims and under-recording by the police [6].

According to [5] adding up the total number of cases with equal weights can also be misleading as an indicator of true crime risk. For the latter problem the paper examines three types of weight calculations: monetization weighting, opinion-based weighting and data-based weighting. Looking at their results all methods show similar results. With balancing, the overall crime risk is affected more by less frequent forms of offenses.

3. Data and Methodology

3.1. Data Description

We retrieved our crime data from the Federal Office of Statistics (Statistisches Bundesamt). The dataset provides a comprehensive overview of various criminal offenses reported over a span of 45 years, from 1976 to 2020, with the total number of crime events in Germany. These crime types vary from financial crimes such as forgery, theft, aggregated theft, and fraud, to sexual crimes such as rape and sexual abuse of children, to various traffic offenses, to violent crimes such as murder, manslaughter, and offenses causing bodily harm. Overall we had 26 types with relatively few missing data. Before any clustering analysis, we grouped the crime types by subjective choice into categories. During initial exploration of the data we have seen a discernible peak in the year 2007. (Figure 1.)

Based on our initial knowledge of the Great Recession happening around 2007, we wanted to test whether these fluctuations in the crime rates are driven by the economic and governmental political changes in Germany . Therefore we further collected percentage change data from the World Bank: Germany's inflation and % GDP growth for our year range.

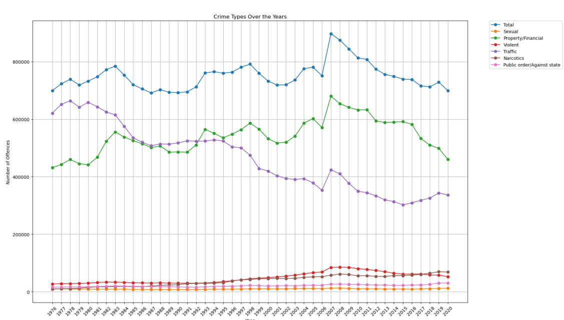

Figure 1. Yearly changes of the crimes, aggregated to categories by hand

3.2. Methodology

For the few missing values, we imputed the mean of the data by each crime. We chose percentage change normalization for further analysis. For inflation, we did not make year to year change because CPI index by itself is providing yearly change.

We tried a couple of clustering algorithms: k-means, PCA preprocessing with k-means, Gaussian Mixture Models, and Random Trees Embedding with hierarchical clustering. Only slight common patterns could be observed, so we decided to use the latter because of it's non-parametric nature, for the reason that there is no initial parameter that would introduce a bias in our analysis. We set a random seed to a specific number for reproducibility.

In our analysis, we employ RandomTreesEmbedding function from scikit-learn library that creates an embedding of the data using a forest of 100 random trees. This is a way of transforming the data into a high-dimensional, sparse binary representation. Each estimator (tree) in the forest is a decision tree that is fit on a bootstrap sample of the data and partitions the feature space into non-overlapping regions. What is more during the process, we use linkage function which performs hierarchical agglomerative clustering. The 'ward' method is an approach that minimizes the total within-cluster variance. At each step, the two clusters that lead to the minimum increase in total within-cluster variance after merging are combined.

We combine those crimes which have similar trends and make a correlation analysis of crimes with GDP and inflation. In order to combine crimes, we take average of the level values of crimes then we take average of them. Then we carried out Pearson correlation analysis with each variable against every other.

We have also conducted regression analysis (Ordinary Least Squares) to see R squared results which is a statistical measure in regression analysis that represents the proportion of the variance in the dependent variable that is predictable from the independent variables. Equations are:

$$Financialcrimes = \beta_0 + \beta_1 inflation + \varepsilon$$

$$Financialcrimes(\%) = \beta_0 + \beta_1 inflation + \varepsilon$$

where the dependent variable is financial crime - grouped by the clustering analysis - and the independent variable is inflation.

4. Results

In the clustering analysis we are able to see which financial crimes have similar trends. Figure 2 visualizes these merges. The height of the branches (distance) represents how different the clusters are. The greater the height, the more dissimilar the clusters. If two crime types are very similar, they will be merged at a lower height, indicating less distance between them. What we observe is, in line with our expectations, theft, theft and misappropriation, fraud and other property offences fell into same cluster. We progressed with these four types further as ”financial crimes”.

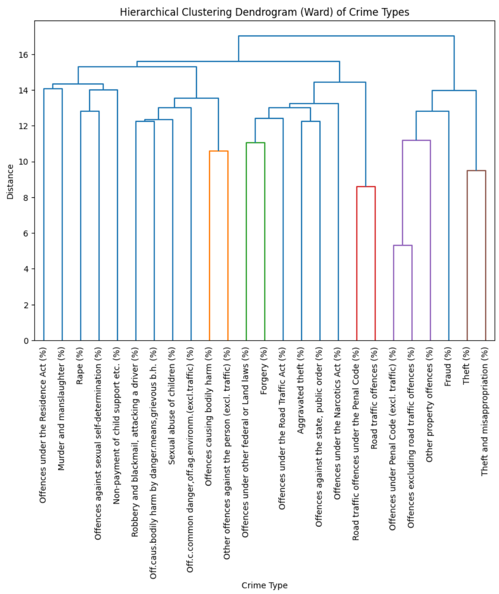

Figure 2. Clustering analysis with RandomTreesEmbedding

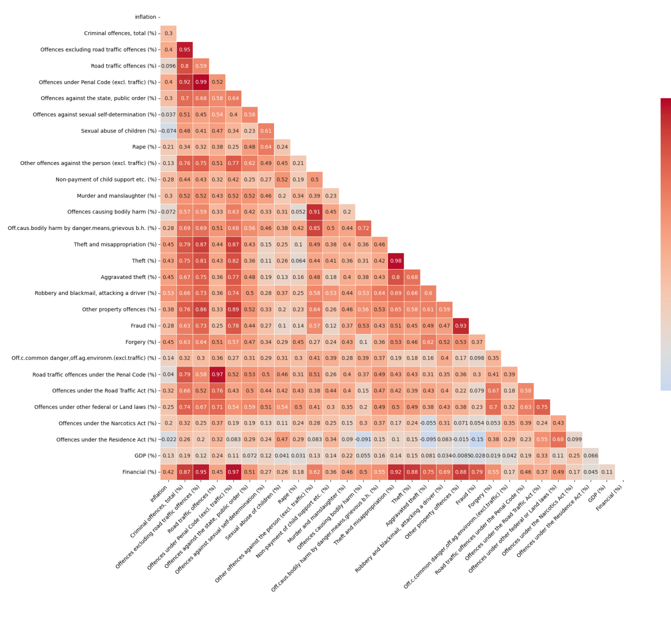

Figure 3. Correlation heatmap of individual crime types, grouped financial crimes by clustering, and inflation

As seen from Figure 3, we find is we find a moderate 0.42 correlation rate between the crimes we grouped with inflation. It shows that people are tend to commit to financial crimes more as there is a concern coming from price increases. On the other hand, we find a weak correlation between GDP (%) with financial crimes which is 0.11. However the latter is considered very weak, both of them are positive rates which aligns to our expectation.

To further investigate the relationship of inflation with our subjective grouping, we also checked their yearly pattern (ie. theft and misappropriation, theft, aggravated theft, robbery and blackmail, attacking a driver, other property offences, fraud and forgery). The data presented on Figure 4 shows there is a relationship between our by-hand grouping of financial related crimes and inflation. Here we also did not use level variables of the crimes but year-to-year change.

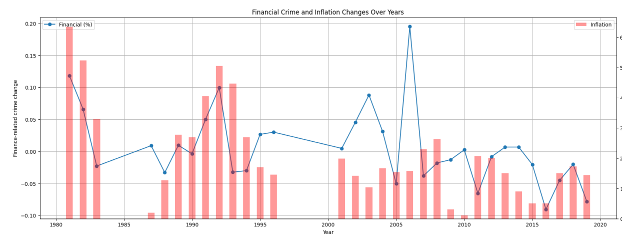

Figure 4. Financial Crime by subjective grouping and Inflation Changes Over Years

Lastly, for the regression analysis we get statistically significant results with the first equation R-squared = 0.223 ($p = 0.001$), and with the second one 0.173 ($p = 0.005$). However these values are not considered high, which could mean - due to our significant results - that the independent variable is somewhat associated with the dependent one but does not explain much of the variability on its own. In theory the model with percentage change can capture the volatility and relative changes in financial crime better. This can be crucial when the absolute number of crimes can be misleading due to changes in population, reporting practices, or other scaling factors. Additionally, percentage change is less affected by the scale of the data, which is helpful when the dataset includes different magnitudes. However, in our regression analysis the one with percentage change explains the variance in inflation with a smaller extent. R squared is not always good indicator, for example it can be misleading as it tends to increase with more variables, regardless of whether those variables are significant predictors or not. For further results of the regression analysis see the Appendix.

5. Discussion

5.1. Influence of Inflation Change to Crime

When we particularly focus on the financial crimes we listed previously, from Fig 3. we can see that there is a 0.41 positive correlation with year-to-year % change inflation and 0.11 with GDP % growth. This indicates that there is a moderate influence of particularly the yearly changes in the inflation to financially motivated crimes, as expected after our literature research [2].

Legal acts try to minimize crime with respect to these assumed influences. Long-term solutions includes extending labour market opportunities, especially for the less-skilled [1]. One seemingly successful policy in Germany was the investment in education and social system, which provides social support for high-school dropouts and thus prevents the possible downward cycle of the less prospective youth. [3] Considering that after the pandemic inflation has become a global problem with varying severity in many countries [7], this indicates to policy leaders that the efficient control of inflation will also be an important task when tackling crimes of these types we got after random trees clustering.

Particularly the highest correlation with the year-to-year % change inflation was to ”Robbery and Blackmail, attacking a driver” crime type with 0.53 correlation. Thinking about the planning that goes into such crime when compared to its counterparts in the financial crimes group, this might affect the higher correlation.

Apart from financial crimes, year-to-year % change inflation has $<$0.3 positive correlation with other crime types. Therefore, we can say that inflation is not enough to explain the changes in the crimes of other types, which have most likely more complex motivations than simply finances.

5.2. Influence of % GDP Growth Change to Crime

Yearly change of % GDP Growth has failed to prove any influence on the German crime types with almost all of them having $<$0.4 correlation, with most of them correlating $\sim$0.1. The highest two correlations it gave to us were with ”Offenses under the Road Traffic act” (0.33) and ”Offenses under Narcotic Act” (0.25).

It has been a discussion that GDP, although in cases a useful measure, seems to fail at describing the well-being of a country [8]. We should also note that it does not describe the income gap between the rich and poor.

5.3. Limitations and possible directions

According to our literature research [5], plotting the aggregated total number of crimes each year (Figure 1.) and the subjective groupings could be weighted by one of the methods mentioned above for better indication of crime risk. Here because of our limited resources we omitted this methodology, partly because we were interested in the number of events that happened, not the actual riskiness of each year. Further analysis with the ”financial crimes” - subjective grouping as well as the resulted group from clustering - could potentially be improved by one of these weighting methods in the future, to see whether this risk factor plays a role in the relationship of these variables.

The clustering analysis could be revised by a method that leverages more on the temporal characteristics of our data, one such method is Dynamic time warping. Rani and Rajasree [9] examines this methodology with different distance measures and finds Euclidean distance to be the best for their crime data. Also the latter method is deterministic and does not depend on random intialization, which would further help with reproducibility.

We have progressed with four types lead by our clustering analysis, while the subjective grouping included eight kinds of types. This gap, involving ”Robbery and blackmail, attacking a driver”, ”Forgery”, ”Aggravated theft”, and ”Non-payment of child support etc.”, could be further investigated.

We have different magnitudes in our data for different crimes, so we could carry out 0-1 normalization before clustering instead of percentage change. This is important mainly because with clustering we are exploring the relationship between different types and percentage change could be misleading with high magnitude differences. However, in respect to our correlation analysis inflation data was provided in yearly change so choosing the same preprocessing method on the crime types seems reasonable.

For further certainty of our results, the initial assumptions of Pearson correlation could be checked with also the significance levels. Both correlation and OLS has the assumption of linear relationships which is not neccessarily the case.

References

[1] Cox, R. (2010). Crime, Incarceration, and Employment in Light of the Great Recession. The Review of Black Political Economy, 37(3-4), 283-294. https://doi.org/10.1007/s12114-010-9079-6

[2] Rosenfeld, R. (2014). Crime and Inflation in Cross-National Perspective. Crime and Justice, Vol 43. Why Crime Rates Fall, and Why They Don’t, 43, 341–366. 10.1086/677665

[3] Hanslmaier, M., Kemme, S., Stoll, K. et al. (2015). Forecasting Crime in Germany in Times of Demographic Change. Eur J Crim Policy Res 21, 591–610. https://doi.org/10.1007/s10610-015-9270-1

[4] Entorf, Horst and Spengler, Hannes, (2008). Is Being ’Soft on Crime’ the Solution to Rising Crime Rates?: Evidence from Germany. DIW Berlin Discussion Paper No. 837, Available at SSRN: https://ssrn.com/abstract=1428782 or http://dx.doi.org/10.2139/ssrn.1428782

[5] Bug, M., Meier, K. (2014). How to obtain a more

accurate picture of crime through crime statistics: Proposals and methods, DIW Economic Bulletin, ISSN 2192-7219, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin, Vol. 5, Iss. 3, pp. 13-21

[6] MacDonald, Z., (2002). Official Crime Statistics: Their use and Interpretation, The Economic Journal, Volume 112, Issue 477, Pages F85–F106, https://doi.org/10.1111/1468-0297.00685

[7] Harding, M., Lindé, J., Trabandt, M. (2023). Understanding post-COVID inflation dynamics. Journal of Monetary Economics, 140, S101–S118. https://doi.org/10.1016/j.jmoneco.2023.05.012

[8] Stiglitz, J. E. (2021). GDP is the wrong tool for measuring what matters. Scientific American. https://www.scientificamerican.com/article/gdp-is-the-wrong-tool-for-measuring-what-matters/

[9] Rani, A., Rajasree, S. (2014). Crime Trend Analysis and Prediction Using Mahanolobis Distance and Dynamic Time Warping Technique.

6. Appendix

$$Financialcrimes = \beta_0 + \beta_1 inflation + \varepsilon$$

Dep. Variable:	Financial crimes (Avg.)	**R-squared:**	0.223
Model:	OLS	**Adj. R-squared:**	0.204
Method:	Least Squares	**F-statistic:**	11.76
Date:	Wed, 24 Jan 2024	**Prob (F-statistic):**	0.00139
Time:	19:32:17	**Log-Likelihood:**	-438.82
No. Observations:	43	**AIC:**	881.6
Df Residuals:	41	**BIC:**	885.2
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	7.241e+04	1729.086	41.876	0.000	6.89e+04	7.59e+04
inflation	-2216.0478	646.267	-3.429	0.001	-3521.210	-910.886

Omnibus:	2.985	**Durbin-Watson:**	0.405
Prob(Omnibus):	0.225	**Jarque-Bera (JB):**	2.542
Skew:	0.593	**Prob(JB):**	0.280
Kurtosis:	2.891	**Cond. No.**	4.96

$$Financialcrimes(\%) = \beta_0 + \beta_1 inflation + \varepsilon$$

Dep. Variable:	Financial (%)	**R-squared:**	0.173
Model:	OLS	**Adj. R-squared:**	0.154
Method:	Least Squares	**F-statistic:**	8.807
Date:	Wed, 24 Jan 2024	**Prob (F-statistic):**	0.00494
Time:	19:38:55	**Log-Likelihood:**	70.709
No. Observations:	44	**AIC:**	-137.4
Df Residuals:	42	**BIC:**	-133.8
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	-0.0254	0.013	-2.027	0.049	-0.051	-0.000
inflation	0.0141	0.005	2.968	0.005	0.004	0.024

Omnibus:	24.579	**Durbin-Watson:**	2.005
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	51.529
Skew:	1.474	**Prob(JB):**	6.47e-12
Kurtosis:	7.407	**Cond. No.**	4.86