

STOCKHOLM SCHOOL OF ECONOMICS

Department of Economics

5350 Master's thesis in economics

Academic Year 2017-2018

Analysis of municipal violations committed in Slovakia: Evidence from administrative panel data

Veronika Horváthová (40904)

Abstract

Using Slovak administrative data on municipal violations committed between years 2011-2016, several violation rates are computed and analyzed. One task is to corroborate the results collected from the already existing literature. Therefore, violation rates are computed controlling for age, education, gender, unemployment or occupation type and the results are consequently compared with the stylized facts about crime rates from the literature. It can be concluded that only some fraction of the stylized facts presented in the previous literature on crime rates can be replicated using the Slovak municipal violations data. Moreover, these violation rates are investigated controlling for income in order to see whether the behavior of the rich differs from the observed behavior of the poor. The data shows that individuals from the lowest or the highest income groups are significantly more likely to commit a violation than individuals with only an average income rate. Finally, no significant effect with respect to the timing of the pay day is found, neither by income, nor by occupation type.

Keywords: crime, violation rate, demographic factors, socio-economic factors, panel data

JEL: I21, I32, J19, K42

Supervisor: Abhijeet Singh

Date submitted: December 11, 2017

Date examined: January 9, 2018

Discussants: Malin Ed and David Ceasar-Heymann

Examiner: Maria Perrotta Berlin

Acknowledgements

I am grateful for everybody who supported me in my research. A special thanks to my supervisor, Assistant Professor Abhijeet Singh, for sharing his expertise, and sincere and valuable guidance in the process.

I wish to express my sincere thanks to my colleges for unlimited support, attention and their comments on the matter.

I am also grateful to my parents, who have provided me with moral and emotional support in my life. I wish to express special thanks to my other family members and friends who have supported me along the way.

Contents

List of Figures	4
List of Tables	4
1 Introduction	1
2 Literature review	3
2.1 Crime and age	4
2.2 Crime and socio-economic factors	4
2.2.1 Crime and education	4
2.2.2 Crime and unemployment	5
2.2.3 Crime and income	6
2.3 Crime and gender	7
3 Research questions and hypotheses	8
4 Data description	9
4.1 Available information	9
4.2 Final dataset	11
4.3 Possible drawbacks	12
5 Results	15
5.1 Descriptive statistics	15
5.2 Violation rates	17
5.2.1 Age	17
5.2.2 Education	19
5.2.3 Unemployment	21
5.2.4 Income	24
5.2.5 Occupation	27
5.2.6 Gender	29
5.3 Seasonal effects	30
5.3.1 Effect of the pay day	31
6 Discussion	32
7 Conclusion	36
Bibliography	38
A Appendix	41

List of Figures

1	Distribution of violations by year	16
2	Distribution of fines by face value	16
3	Frequency of violations (% of all obs.)	16
4	Average fine (in EUR)	17
5	Violation rates by age cohorts (per 100,000 individuals)	18
6	Violation rate by age (2012-2015)	19
7	Average fine by age	19
8	Violations by education level	20
9	Average fine by education level	20
10	Violation rates by education cohorts by year (per 100,000 individuals)	21
11	Violations by unemployment	22
12	Average fine by unemployment	22
13	Violation rates by unemployment (2012-2015 data, per 100,000 individuals)	23
14	Distribution of wages by year	25
15	Average fine by income cohorts	25
16	Violation rates by income cohorts by year (per 100,000 individuals)	26
17	Violation rates by income per month (2012-2015, per 100,000 individuals)	27
18	Violation rates by NACE (12-15)	29
19	Average fine by NACE	29
20	Share on violations by gender	30
21	Average fine by gender	30
22	Violation rates by gender (2012-2015, per 100,000 individuals)	30
A.1	Number of duplicates grouped by personal identification number	41
A.2	NACE classification of occupations	42
A.3	Number of violations by day in a month (NACE subset)	42
A.4	Violations by day in a month by income pctlles (NACE subset)	43
A.5	Violations by day in a month by income pctlles for <i>Public administration</i> (NACE O)	43
A.6	Violations by day in a month by income pctlles for <i>Education</i> (NACE P)	44
A.7	Violations by day in a month by income pctlles for <i>Human health</i> (NACE Q)	44

List of Tables

1	Legislative fines for the most common types of municipal violations	10
2	Average income by year in EUR	24

1 Introduction

Obedience to law is not something that can be taken for granted, as criminal behavior and repeating violations of law can be observed everywhere. As significant public resources are currently being spent to prevent and punish unlawful behavior, it is worth knowing why such behavior occurs in the first place. Once this is thoroughly investigated, that might provide the necessary background for the discussion about what can be done to minimize the costs and maximize the efficiency of the existing law system. Possibly extending the knowledge required for this discussion, the subsequent study tries to corroborate whether already observed facts about crime rates are true for data about Slovak violation rates, as violations comprise only a subset of all criminal deeds. And moreover, it investigates whether a pay day is correlated with the total number of observed violations.

First to propose a formal model explaining the motivation behind committing different crimes was Becker (1968). Back in the days, his paper was revolutionary as not many scholars were thinking about crime in economic terms. He was the first to introduce the concept of rationality into explaining why criminal behavior occurs. This might seem to be a very obvious link now, but at that time this thought was considered to be very unconventional and truly out of the box. Most of the literature only focused on providing descriptive statistics on crime rates with respect to different factors. In other words, these studies were mostly focused on correlation instead of causality. His paper was the first to implement a basic cost-benefit analysis and it proposed that criminal behavior can be observed when the benefits of committing a crime simply outweigh the costs. By costs are meant the probability of being caught and consequently, the punishment itself. Nowadays, due to technical progress, sophisticated models can be proposed and these try to causally link crime to different individual demographic and socio-economic factors.

This study tries to analyze violation rates, instead of crime rates, which are typically discussed in the literature. By the term crime rate is expressed the number of criminal deeds (overall or of a certain type) that are expected to be observed and committed per 100,000 individuals. As this study focuses on the calculation of several violation rates, the important thing to know is how a violation is defined. Municipal violations comprise only a subset of all crimes. By this term is understood a criminal deed which is committed and is against the law, but it is not as severe to be considered a crime. For example, municipal violations comprise traffic violations like speeding over the limit or drunk driving (under the assumption that no people are hurt or die). Another example might be being too loud after a specific hour or consumption of alcohol at public places. Such cases are typically not considered crimes and they are not taken to court. The result is such that perpetrators receive fines, that have to be paid off before a certain date.¹ Consequently, any violation rate accounts for the expected number of municipal violations being committed per 100,000 individuals.

¹If this condition is not satisfied, then such case might eventually end up in court. However, empirical data show, that most of the fines are paid off immediately when the deed was committed. Based on the dataset at hand, more than 88 % of cases were resolved immediately.

In line with the previous argument, one of the main goals is to investigate whether it is possible to replicate already defined stylized facts observed when more severe crimes were analyzed also when one includes only a subset of all crimes. More precisely, when one investigates only the unskilled crimes or in other words, municipal violations. To provide an example, the existing literature can present a lot of evidence that in general, the crime rate decreases with age. The corresponding research question this thesis tries to answer is whether it is possible to observe the same pattern with respect to age as one restrict the sample and includes only municipal violations.

This analysis tries to answer several questions. First of all, it intends to provide an investigation of violation rates with respect to different factors such as age, unemployment, education, gender or type of occupation. Based on the already existing literature, stylized facts are collected and summarized in the third part of this thesis and the subsequent analysis tries to corroborate these results in the Slovak environment.

Secondly, this thesis investigates whether it is possible to associate different income levels with different behavioral patterns. This part is inspired by a very influential paper written by Mani et al. (2013) in which the authors discuss how poverty influences cognitive functions. In general, several papers found that bad decisions could be traced to the fact that the decision makers were poor and hence, unable to optimize their behavior efficiently. It might be interesting to analyze municipal violation patterns from this perspective in order to see which income group tends to commit these municipal violations. Whether it is the poor being involved in a very self-damaging type of behavior or whether it is the rich, who do not face such severe consequences because paying a fine is of a small burden for them.

Thirdly, the purpose of this study is also to investigate whether the exact timing of the pay day plays any role whatsoever and whether it affects the time distribution of committed violations. This part is inspired by a paper written by Anderoni et al. (2017) in which the authors performed a natural field experiment. The experimenters intentionally misdelivered envelopes to rich and poor household in order to investigate the return rates. In other words, they wanted to analyze the extent to which were these misdelivered letters in the end resent to the original receiver. They found that these rates depended strongly on the exact days when the individuals included in the experiment received their wages. Therefore, it might be interesting to analyze whether this result can be extended and corroborated in other social situations.

As for the motivation why such thesis is of a practical importance, it might be helpful to shed some light into this topic to understand what drives different violation patterns. From the theoretical micro perspective, this is an interesting question itself, just to untangle something more about the human nature. Moreover, there might be a relationship between violation and crime

rates. It might be the case that individuals more likely to commit a less severe type of a crime (speeding, consumption of alcohol at public places) might consequently be more likely to commit more severe crimes. Then, the investigation of violation rates might serve as some sort of a predictor of the crime rate, as well. Also, this paper can be viewed as a case study for one particular country for which this type of analysis has not been performed yet. Therefore, it might serve as an important starting point for further research in this specific area. In practice, this might be of a particular importance when designing smarter public policies. For example, more effective control mechanisms might be designed based on the conclusions stemming from analysis.

Practically, the analysis is performed using administrative individual Slovak municipal violation data spanning from 2011 to 2016. These data are matched with different demographic variables such as age, gender and other indicators of personal socio-economic status (education, employment status, average yearly income and occupation type by the statistical classification of economic activity). Consequently, the violation rates are calculated and the number of violations is analyzed with respect to the known pay days.

The structure of the paper is as follows. The second chapter presents a literature review of different topics related to violations and crime in general. The third chapter provides a review of hypotheses this thesis tries to corroborate and a summary of research questions. The fourth chapter provides description of the dataset used in the analysis while the fifth chapter presents descriptive statistics of the data and obtained analytical results. The sixth chapter discusses the validity of the study and provides ideas for further research. The final chapter concludes.

2 Literature review

A vast literature analyzing different crime patterns can be found as crime has always been an interesting element to study. First, the studies started presenting mostly descriptive results providing calculations of violation or crime rates with respect to different age, education or income cohorts. As for the formal definition, violation/crime rates are typically calculated as the total number of criminal deeds one would expect to observe per 100,000 inhabitants. This definition is also used in the following parts of this paper. As for the practical difference between a violation rate and a crime rate, while the violation rate accounts for criminal deeds such as speeding or being too loud after a particular hour (unskilled crimes), the crime rate typically accounts for more serious deeds such as stealing or homicides.

As it was already mentioned, the literature originally provided descriptive statistics on observed violation and crime rates. Only papers from recent years tried to search for causal links between individual factors and committed criminal deeds. Yet, much research is still left to be done, especially with respect to gender, as the literature never really focused on this particular factor.

2.1 Crime and age

One of thoroughly studied and very well documented relationships in criminology is the age-crime profile. In general, there is a strong agreement on the fact that the crime rate peaks in the late teen years or in the early twenties and that it decreases as individuals age (Hirschi and Gottfredson, 1983). While Hirschi and Gottfredson (1983) claim that there is a substantial part to the age-crime profile which can only be explained by age, newer papers go further in order to understand what it is about the age that causes this profile to be fairly invariant, even when controlling for different factors such as gender, race or type of crime.

One obvious problem with this variable is that it is strongly correlated to other important factors that evolve as individuals age. Just to mention a few, existing literature naturally links age to education, family, maturity, stronger social network or income (Blonigen, 2010; Glaeser et al., 1996; Sweeten et al., 2013; Williams and Sickles, 1998). Controlling for these effects can explain much of the variation previously assigned to be caused solely by the variable age (Hansen, 2003; Williams and Sickles, 1998). In practice, it can be argued that all of these factors are expected to have a negative effect on the crime rate as all of these factors increase the opportunity costs of committing a crime (Hirschi and Gottfredson, 1995; Lochner, 2004).² In general, it is important to disentangle different effects that were previously assigned to be caused by age, but are in fact caused by individual socio-economic status factors such as education, unemployment or income, in order to get unbiased analytical results. And also, correctly estimated causal links allow policy makers to propose fitting prevention and correction mechanisms. For example, it might be the case that policies focusing solely on age could induce some unexpected behavioral responses that might worsen the overall situation.

2.2 Crime and socio-economic factors

2.2.1 Crime and education

Multiple studies focused on the descriptive statistics of crime rates with respect to education. As a result, many studies found correlation between lower achieved education level and higher crime rates (Farrington et al., 1986; Freeman, 1996; Hansen, 2003; Lochner, 1999). As for the younger age cohorts, it is argued that schooling is important because it simply keeps the students out of streets. Nevertheless, schooling allows individuals to acquire knowledge necessary for further success (Hansen, 2003).

²More precisely, higher education, stronger family connections (starting own family), more experience gained through trial and error, stronger social network in terms of having made own friends and business connections or higher income due to more years of experience - all of these factors practically increase the opportunity costs of committing a crime because older individuals can lose more in terms of all of the aforementioned factors. Hence, then it is more difficult for the benefits to outweigh the costs and as a result, lower crime rate is recorded.

In order to further understand the aforementioned relationship between age, crime and education, Lochner (2004) proposes that these patterns should be analyzed within a human capital framework, which he developed in his paper. He argues that older, more intelligent and more educated people accumulate higher human capital which can consequently be translated into higher wages and therefore higher opportunity costs of committing an unskilled crime. As a result, the age-crime profile decreases with increasing age, at least for the unskilled crimes.³ Moreover, the overall age-crime profile is expected to be fairly flat in environments with little skill formations. As for the dynamics, the drop in crime rate is steeper within younger age cohorts as the marginal value of work increases the fastest within these particular age groups early in their work careers.

2.2.2 Crime and unemployment

The evidence regarding crime and unemployment shows that in many countries some form of correlation between these two variables can be found. More precisely, several papers found that unemployment does have an effect and worsens the situation especially in relation with property crimes (Lin, 2008; Raphael and Winter-Ebmer, 2002). However, a general link between unemployment and crime has not been established (Freeman, 1999). Related to this, criminal patterns were also found to react to changes in income incentives (Entorf and Spengler, 2000; Gould et al., 2002). Additionally, a group especially affected by unemployment is the youth (Farrington et al., 1986; Fougère et al., 2009; Gould et al., 2002). In general it can be observed that as the labor market prospects improve, the crime rate decreases and vice versa. The effect of unemployment is often explained by a Becker/Ehrlich model which directly links labor income and crime.

Overall, the models on crime and unemployment typically focus on individuals who are on a margin of crime, deciding whether to participate in illegal activities or not. The standard tool applied in analysis is a Becker/Ehrlich model which has been used for analytical purposes for decades now. It introduces a concept of rationality into individual's decision making process and proposes that the decision between working and cheating might be a rational one, influenced by present incentives. Hence, this model takes into account a basic trade-off between labor income and monetary income stemming from cheating under some probability of being caught and punished. Then, the final result obviously depends on which scenario yields greater expected utility and that is chosen by the agent (Becker, 1968; Draca and Machin, 2015). Naturally, direct implication of this theory is that the unemployed are expected to be significantly more prone to crime and violent behavior since there is much for them to gain, when accounting for the fact that their labor income is practically zero. In some cases, income earned through illegal activity might be the only available source of

³By an unskilled violation or crime can be understood larceny or different traffic violations. The distinction between skilled and unskilled crimes is important to note because better education or higher intelligence levels may actually increase expected profits of some while collar crimes such as fraud or embezzlement. In return, that may imply that these crime rates actually peak later than in early twenties as they provide substantial skill premium. Lochner's estimates actually display increasing arrest rates for while collar crimes in relation to increasing education level.

income and hence, participation in a crime might be a very rational option. It should be noted that even though the basic Becker/Ehrlich model is oversimplifying the reality, it can be extended by introducing dynamic properties of time allocation (Lochner, 2004) or crime specialization in terms of different risks and different outcomes (Draca et al., 2015).

2.2.3 Crime and income

Overall, the literature on unethical behavior and crime associated with income presents two prominent arguments. One of them explains why the poor are the social class which is more likely to participate in this type of behavior while the other one argues that in fact the rich are the group more likely to commit unethical and criminal deeds.

First of all, there is some literature suggesting that the poor might be the social group more likely to behave unethically or to break the law. The argument is that their motivation to increase their revenue or resources is relatively high (due to the high marginal utility of income) which might drive them to commit criminal deeds. Empirical evidence based on the data on the police arrests from the US and the UK indeed shows that crime decreases with increasing income (Freeman, 1996; Grogger, 1998; Machin and Meghir, 2000). This argument is supported by the evidence collected in a natural field experiment in the Netherlands as presented in the paper by Anderoni et al. (2017). The experimenters misdelivered letters on purpose in order to investigate whether the incorrect recipients would resend the letters to the original receivers. Since it was possible to see that the letters contained some monetary amount, the subjects could realize that they could earn money by not returning the envelopes. This experiment showed that the rich households returned the misdelivered envelopes to a relatively larger extent indicating that the rich in fact behaved more pro-socially. The authors argued that this is caused by the fact that the rich simply have more money and that the rich households did not care much about keeping additional misdelivered 20 EUR while even such a small amount of money was highly valued by the poor households.

Additionally, as for the income, another important factor is the exact pay off day. Multiple studies showed that the behavior of subjects depended highly on the fact, whether the subjects were supposed to make an important decision before or after the pay off day. As for the empirical evidence, Anderoni et al. (2017) found out that the rate of pro-social behavior (rate of return of the misdelivered letters) was the highest shortly after the pay day and that the rate was decreasing as it was getting closer to the date of the following pay day. Similar results were observed by Carvalho et al. (2016). They found that participants were more present-biased when making intertemporal choices a couple of days before the pay day as compared to the group making these decisions after the pay day. The channel causing such behavior was famously described in a paper by Mani et al. (2013). The central argument is that poverty can be traced to induced higher levels of stress which in turn might affect the individual decision making process. This naturally yields a hypothesis that even criminal deeds might be subject to this type of behavior. Certainly, there are other papers

showing that the poor struggle with making strategic decisions (Haushofer and Fehr, 2014; Karlan et al., 2014; Mani et al., 2013). Therefore, it might be the case that committed crimes might be to some extent attributed to poor and more present-biased decision making process while unfortunately, the consequences might be long-lasting. Not accounting for the fact itself that basically any monetary punishment is relatively more harsh for the poor due to the obvious lack of resources.

The second line of argument proceeds as follows. In fact, the upper income class might be the one more prone to unethical behavior or crime. In general, this could be due to the fact that these people do not truly comprehend the concept of scarcity of resources because often they do not face the same severe trade-off as the poor. Overall, the rich face fewer risks and enjoy greater freedom. For example, paying a fine is typically not seen as a very harsh punishment since these additional costs are easy to cover due to diminishing marginal utility of income of the rich. Indication that a fine can be merely viewed as a price is presented in a paper by Gneezy and Rustichini (2000). Here, the experimenters thought that an introduction of fine for a late arrival to a day-care center was supposed to reduce the rate of late arrivals but instead the behavioral response was exactly the opposite as the parents participating in this experiment viewed the fine as a fee allowing them to be late. In line with this evidence, the members of the upper-class might be more likely to break the law since paying a fine can be seen as a fair bargain covering up for the unlawful behavior. This argument is supported by a paper written by Piff et al. (2012). They run several field and laboratory experiments and the evidence showed that the rich behaved relatively more unethically than the poor. More precisely, as for the field experiment, the rich were more likely to break the law while driving (cutting off other cars, cutting off pedestrians). Additionally, the laboratory experiment showed that the upper class members endorsed unethical behavior at work or cheated in order to increase their chances of winning relatively more often than lower-class individuals. These results were also replicated by Dubois et al. (2015). Another example is that the rich face higher marginal tax rates and hence, they have higher motivation to avoid paying taxes as they might be left with greater income. The empirical evidence does indeed support this line of reasoning (Andreoni et al., 1998; Cox, 1984; Wang and Murnighan, 2014). In other words, the literature suggests that individuals with higher income status tend to be more self-focused and greedy when discussing unethical and criminal behavior (Kraus et al., 2009, 2011).

2.3 Crime and gender

Several studies investigated the relationship between crime and gender. One stylized fact is that women are significantly underrepresented in the collected crime datasets when compared to men (Entorf and Spengler, 2000; Graham and Bowling, 1995; Hirschi and Gottfredson, 1983). Nevertheless, the general age-crime pattern is still observed, although the line is flatter for females, but it still peaks around late teen years or in the early twenties (Entorf and Spengler, 2000; Hirschi and Gottfredson, 1983). Interestingly enough, there is a significant lack of evidence on why it is the case that women are less likely to break the law and commit a crime or a violation. Most of

the papers only analyze behavior of men or propose formal theories with no further solid empirical evidence. Hence, the whole concept of the crime gender gap is still being eagerly disputed (Heidensohn, 2012).

3 Research questions and hypotheses

Based on the literature review, this thesis tries to corroborate the following descriptive results:

Hypothesis 1 (H1): *Violation rate peaks with the youngest cohort and decreases with age.*

Hypothesis 2 (H2): *Violation rate decreases with higher education.*

Hypothesis 3 (H3): *Violation rate is higher for the unemployed than for the employed.*

Hypothesis 4 (H4): *Violation rate decreases with age for the employed and the unemployed.*

Hypothesis 5 (H5): *Males are overrepresented in the sample of perpetrators.*

Hypothesis 6 (H6): *Violation rate decreases with age for males and also females.*

As for the following Hypothesis 7, the literature argues that the crime rate can increase but also decrease with income. Based on the nature of the dataset, the subsequent H7 states that the rich are more likely to commit violations. This is due to the fact that violations comprise criminal deeds such as speeding and consuming alcohol at public places. It can be argued that these deeds cannot really help the poor to increase their resources, there is not much for them to gain when they commit these violations. Unlike the rich, who can simply afford to pay for it, as was illustrated by an experiment by Gneezy and Rustichini (2000). As a result, the fine was merely seen as a price and the following H7 tries to cover that particular case exactly.

Hypothesis 7 (H7): *Violation rate increases with income.*

Moreover, as it was already mentioned, this paper tries to investigate possibly existing pay day effects. Hence, two additional hypotheses, not exactly overtaken from the existing literature, but based on the arguments proposed above, are:

Hypothesis 8 (H8): *Number of violations committed by the poor decreases after the pay day.*

Hypothesis 9 (H9): *Number of violations committed by the rich is not affected by the pay day.*

The first of the two hypotheses is based on the fact that as it was shown, the poor tend to engage in worse decisions before the pay day when compared to the situation after the pay day, as they are capable of thinking more pro-socially and they are less present-biased. Therefore, the Hypothesis 8 assumes that the poor can account for the fact that deeds as speeding are not truly worth it and that they would commit these types of violations to a smaller extent as these can be seen as a

waste of money yielding no particular gain (as no crimes yielding monetary gains are included in the database comprising municipal violations). Moreover, it is assumed that the pay day should not have any effect on the rich as they face decreasing marginal utility of income and their position does not change drastically with respect to their own pay day. In most of the cases, it can probably be assumed that they are as likely to have enough resources to pay for a fine a couple of days before the pay day as they are a couple of days after.

4 Data description

The following chapter provides a short description of all the available data and how the final dataset is created and adjusted for. That includes explanations why some observations are dropped altogether and possible risks related to these changes. This chapter also discusses the internal validity of the study and presents possible drawbacks in terms of quality of the data that could affect the final results. Some projections on how the internal validity could be increased in case more detailed data was available are also presented towards the end of this chapter.

4.1 Available information

The final dataset, as described in the following subsection, is created by merging several smaller datasets together. All of the datasets consist of administrative data collected and provided by different governmental organizations. The main dataset is the one containing information about municipal violations. These data are available on the individual level and one observation contains information about one particular violation. The dataset includes the following information about every unique observation: the fine in nominal terms, the date when the deed was committed and the date when the case was closed. Unfortunately, this dataset does not include information about the type of violations committed. In other words, as for now, it is impossible to tell why these individuals received their fines. But for the illustration, Table 1 presents some examples of the most common violations committed in Slovakia and the corresponding legislative fines. It can be seen that the face value somewhat represents the severity of the deeds.⁴

Furthermore, every observation left in the dataset includes a coded identification/personal number which allows for merging with other datasets. In case that one person committed several violations, all of them are reported and included as separate observations.

Secondly, the data on municipal violations is merged with demographic data. Therefore, it is possible to identify the gender of the perpetrator, the nationality, the number of kids at the end of each year and the year of birth. The date when the violation occurred and the year of birth together allow to impute the exact age at the time when the individual committed the violation. Other (less

⁴Summary of fines for traffic violations can be found here, fine for drinking at public places here and fine for being too loud after a certain hour here.

Speeding (max. +20 km/h)	up to 50 EUR
Speeding (+21-50km/h)	30 - 400 EUR
Speeding (+51 km/h and more)	250 - 800 EUR
Drunk driving	150 - 800 EUR
Driving over the STOP sign	100 EUR
Driving over the NO ENTRY sign	50 EUR
Drinking at public places	up to 33 EUR
Being too loud after a certain hour	up to 33 EUR

Author rendering of existing legislative acts.

Table 1: Legislative fines for the most common types of municipal violations

detailed⁵) information, that can possibly be used in the analysis, is entries about the county of the permanent residence (however, only available for years 2015 and 2016) and whether the individual is married. Unfortunately, not all indications about marriages include additional information about the month and the year of the wedding. That implies that it is not always possible to indicate whether the individual was married at the exact time when one committed the violation in question.

Finally, the information about socio-economic status (SES) is incorporated into the final dataset. At first, the data on the highest achieved educational level is added. This data consists only of binary variables indicating whether the individual of interest finished the given educational level. Unfortunately, the dataset lacks information on the exact dates when the given level was achieved. Another part includes information about unemployment. This data includes information about the exact date when the individual was registered at the Office of Labor, hence, included in the database of unemployed and also the date when such individual was dropped from this database. If an individual was registered in this database more than once, all of these information is collected and available for use, hence, this data is very detailed and precise. Due to this it is fairly simple to impute a binary variable indicating whether an individual was unemployed at the time when the violation was committed.

Lastly, there is data on the statistical classification of economic activity (OKEC), which is information similar to NACE and indicates the type of profession of the given individual. If an individual changes jobs in the course of one year, then OKEC indicates the profession where the individual spent most of the working time of that given year. For example, if an individual spent three months working as a bus driver and another nine as a shop assistant, OKEC would identify this individual under the category shop assistant and the corresponding OKEC code would be assigned to this individual. Moreover, the data also includes the total yearly income on the individual level earned during the year when the violation was committed. If an individual changes

⁵By term *less detailed* is meant that these information are not available for some fraction of observations.

jobs, then this yearly income is equal to the total monetary amount earned while working under different professions. However, the last two variables are not provided for a substantial part of the dataset (for around one half of the entries in the dataset).

4.2 Final dataset

The final dataset comprises only municipal violations that occurred between 2011 - 2016. This span is chosen mostly due to technical reasons, the data from the time before 2011 is simply not available in the database just yet. All observations with the face value of the fine below zero or over 10,000 EUR are dropped from the database because clearly these are only typos. Also, observations with no indication of the face value of the fine are dropped because such observations do not add much value to the analysis. Cases with no identification number are dropped because without this number it is impossible to match any other demographic or socio-economic data. Cases lacking entries about the exact date when the violation was committed are also dropped. Practically, many dropped observations suffer from more of the aforementioned issues at once (for example lacking information about the personal number and also the fine), therefore dropping them does not affect the results much as such observations do not have much of an added value for the analysis. Theoretically, this could be an issue if there existed a structural or a technical problem in the reporting. Such problem could cause that one particular city or region would be incorrectly recorded in the database and hence, such set of adjustments and drops would create a structural bias. Unfortunately, as for the nature of the available data, it cannot be resolved whether these mistakes in reporting are structural. Fortunately, the number of observations is large enough and that can compensate for dropped cases to at least some extent. Another minor thing is that observations where the day when the case was closed is before the day when the deed actually occurred are also dropped for being obvious mistakes. Other technical adjustments comprise dropped cases due to the fact that the recorded month is an integer greater than 12 or when the recorded day when the deed occurred is an integer greater than 31, but these issues affect only a couple of cases. Fortunately, the sample is still large enough, hence, even after all of these adjustments there are around 4 million observations left that can be used. More precisely, the final dataset consists of 4,014,089 unique observations.

As for some restrictions based on the data on demographics, violations committed by foreigners are dropped from the analysis. As foreigners are defined people that are not eligible for having a permanent residence in Slovakia. Foreigners eligible for a permanent residence must pass language and other tests, hence, there is no reason to drop them from the database as they are tested to fit the general population. This ensures that the results can provide some insight into how the majority population behaves in terms of municipal violations. In practice, a couple of thousands observations are dropped due to this adjustment, which is a minor adjustment with respect to the total number of observations. Moreover, the dataset includes only the adult population, hence, individuals over 18. This is due to the fact that many other characteristics such as SES cannot really

be matched to people below 18 since they are technically not unemployed nor working and there is not much information known about them other than their age. Hence, the analysis focuses only on the adults. As for some minor adjustment, some observations are dropped because of obvious reasons. For example, observations where the individual died before the violation was committed.

As for the data on education, some information is changed manually in the dataset. Education status of people below 22 at the time when the violation was committed with reported achieved education level higher than high school is changed to education status *high school*. That is the highest level these individuals could have achieved by the time the violation was committed. Similarly, the status of people below 24 with more than a bachelor degree is changed to the status *bachelor degree*. Third change is that the status of people below 28 with recorded PhD diploma is changed to status *master degree*. These changes are based on the assumption that there is low prevalence of taking gap years while studying in Slovakia and these age limitations are just natural limitations arising from the standard time requirements of different programs.

4.3 Possible drawbacks

Ideally, one would include many other information, which would increase the precision of the analysis. As for the data on municipal violations, the most significant drawback is that at the moment it is impossible to say why all of the individuals received a fine. Although this information is available for some observations in the dataset, it is not well coded. The problem is that at the time of collection of the data there was no systematic approach on how the data should be reported. If there was such approach, that would allow different police districts to collect the data in a standardized format and that would consequently allow for a further analysis of deeds instead of an analysis of fines and the dates only. Hence, if one wanted to include this type of information in the model, one would have to go through all of the observations manually in order to create at least a categorical variable which would indicate the most common violations committed in Slovakia. Due to limited manpower and obvious time restrictions, this was very well beyond the scope of this analysis.

Another interesting thing would be to know where the violations were committed (ideally on a county level). Again, this information is available to some extent in the database, there is information about the police district being responsible for the investigation. It can be assumed that due to the nature of the data, the police district recorded as the one running the investigation represents the county where the violation occurred. Unfortunately, the data is far from being usable as there is no standardized way of how different districts are named.⁶ One would have to go again observation by observation in order to recode the information about a district which was

⁶To illustrate the problem, if there is a police office in a particular city (in Bratislava, for instance) then sometimes there would be entry as *Bratislava* only, then it could be *Police district Bratislava*, *PD BA* and many other combinations. Moreover, there are many typos which further complicate the situation.

in charge of the investigation into a usable categorical variable. The time did not allow me to perform this transformation of data and therefore these information had to be dropped from the analysis altogether.

Unfortunately, not even the Register of Individuals can shed some light on the likely locations of the committed municipal violations. The aforementioned register provides information about the place of permanent residence, not the actual place of residence of individuals. In practice, there might be large discrepancies between the number of residents with the permanent residence and the actual number of people living and working in a given city. For example, according to the Statistical Office of the Slovak Republic, slightly more than 425 thousands of people live in the capital city.⁷ According to a study from this year, which is based on data from mobile sim-cards spending the night in the capital city, the actual number of residents might be close to 650 thousands.⁸ This clearly indicates that many people do not report their actual place of residence. Therefore, it is impossible to impute the location of the committed violation and the data on the permanent residence should be viewed merely as a proxy for the data on actual place of residence. As a result, one could expect that merging these data with data on fines would result in under-representation of violations committed in larger cities as many cases would be assigned to people only commuting to larger cities (as illustrated with the example with the capital city with the official and actual number of citizens) while these data would probably overstate the total number of violations committed in smaller cities and villages (as many people tend to keep their original address even after moving to a different city). One other minor issue which is mentioned in the previous section is that the data on the permanent residence of individuals is only available for years 2015 and 2016. This provides additional problems when controlling for the actual residence during the period of interest, but it is difficult to form any expectations on how this might bias the results.

As for the data about highest achieved education level, this dataset consists only of dummies. There is no information about the exact graduation dates, therefore this causes some problems when matched to the municipal violation data. This might cause some discrepancies in cases when the reported highest achieved level in the dataset is greater than it was at the time when the violation was committed. It is difficult to state the total number of problematic cases due to this issue. As is stated earlier, some transformation has been done in order to increase the precision of the data, but there might still be some discrepancies, especially within the category of people up to approximately 28 years. Likely, this is only a minor problem given the size of the dataset, but it still makes controlling for education a little imprecise. This might bias the results, especially within the youngest cohort, the size and the direction of the possible bias are disputable and very difficult to predict. Moreover, it is not possible to match all observations with education, only around 7/8 of the total sample. There is no known objective reason why this is the case and also,

⁷Data available from the Statistical Office of the Slovak Republic here.

⁸Case study available from Market Locator.

how this could possibly bias the result. But this could create additional issues if the unavailability of data is a structural problem with one group being systematically missed out or being incorrectly reported in the data.

There is another problem partially linked to education. As it was mentioned before, the data can be matched to a very precise information about unemployment. It was also mentioned that a dummy is created, indicating whether the individual in question was unemployed at the time when the violation was committed. However, if there is no information about unemployment, one cannot directly imply that the individual was employed at the time when the violation was committed. Such individuals could have been studying (hence, being out of the labor force), taking parental leave or being retired.⁹ Therefore, the implied violation rate of the employed is probably overstating the real violation rate, as there are individuals included in the labor force that do not belong there. Unfortunately, there is not much data available that would allow to control over this particular issue. It can only be controlled for this effect to some extent with education and it is done so in a way that if the education level has to be recoded as explained in the previous subsection, then such individual is dropped from the analysis with respect to the unemployment as it can be assumed that such individuals were studying at the time when the violation was committed and therefore, these individuals should not be accounted for in the analysis.

As for the data on wages and the statistical classification of economic activity (OKEC), the biggest technical issue is that only about one half of all of the observations from the dataset on violations could be matched with the data on wages and OKEC. There is no clear explanation why this is the case, for some reason this data is simply not available for matching for the other half of the sample. Unfortunately, it cannot be identified whether there is some structural problem in reporting of the data linked to some specific characteristic of a subset of a population. But if it was linked to a specific part of population with a specific feature, that could bias the results in an unpredictable way and it could invalidate the whole analysis. Unfortunately, there is not much information available to predict the size or the direction of the bias.

There are two more factors which could be interesting to control for if the data was being collected and made available. The first factor is religiosity and the second one is political preferences. The former is interesting because statistically, around 75 percent of citizens admit to have religious beliefs.¹⁰ Since religion in general does not particularly support any criminal behavior and moreover promotes upholding to social norms, it could be interesting to see whether there could be an effect when controlling for this factor or not. Another factor interesting to control for could be political preferences just to see whether members or supporters of any political beliefs are more prone to violating these social norms.

⁹For example, women having more children were allowed for early retirement, where the exact year of eligibility for early retirement depended on the total number of own children.

¹⁰Data available from the web page of the Ministry of Culture of the Slovak Republic here.

5 Results

The main goal of this section is to present stylized analytical results and to critically evaluate the accuracy of hypotheses stated in the Section 3 by means of providing evidence either in favor or against them. First of all, the descriptive statistics of the final dataset is presented which is followed by the analysis of the violation rates with respect to different individual and socio-economic factors, as were discussed in the literature review. Final subsection comprises exploratory part of the study where the total number of violations is studied with respect to the identified pay day.

5.1 Descriptive statistics

In total, 4,014,089 unique observations are included in the final sample. However, this final number of observations does not consist only of violations committed by unique perpetrators. Many perpetrators included in the sample committed more than one municipal violation, as can be seen from the Figure A.1 included in the Appendix section of this paper. Here, the individuals are grouped by the total number of violations assigned to a particular identification number. In other words, if the number in the column copies equals 5, that means that such an individual included in the corresponding row committed 5 different municipal violations. The total number of unique perpetrators committing 5 violations can be calculated as the difference between columns observations and surplus. Total number of unique perpetrators included in the sample is equal to 1,404,392.

The first panel provides a basic overview of the data on municipal violations. Figure 1 provides an overview of all committed violations presented as shares on the total number of observations included in the sample. It can be seen that years 2011 and 2016 are significantly underrepresented in the sample. As it was discussed, year 2011 is problematic because it is the year when the digitalization started and hence, not all observations are included. Apparently, the missing observations are not added retrospectively and what is collected seems to be around one half of the normal amount when compared to the subsequent years. As for 2016, since the dataset was obtained in early 2017, this low level is probably caused by an administrative lag. From the data at hand it seems like the standard level of violations can be associated with levels recorded in 2014 and 2015 (around 880,000 observations per year). Figure 2 then presents the cumulative distribution of fines.¹¹ It can be seen that more than 90% of all fines are below 100 EUR with mean being equal to approximately 36 EUR. Around one third of all fines are equal to 20 EUR with another more than 20% observations with fine being equal to 10 EUR, these are the most common values found in the data set. Otherwise it can be seen that the fines are grouped around some numbers such as 30 EUR, 40 EUR etc. which is not surprising given the legislation and how the fines for criminal deeds are defined in different acts. Table 1 from the previous section then illustrates some examples of the most common committed municipal violations and the corresponding fines.

¹¹Fines with a face value greater than 300 EUR are not shown in this figure (around top 1% of all fines is not presented in the graph) in order to present a more detailed view on the distribution. Using all the extreme values such as fine being equal to 6,000 EUR would make the graph completely illegible.

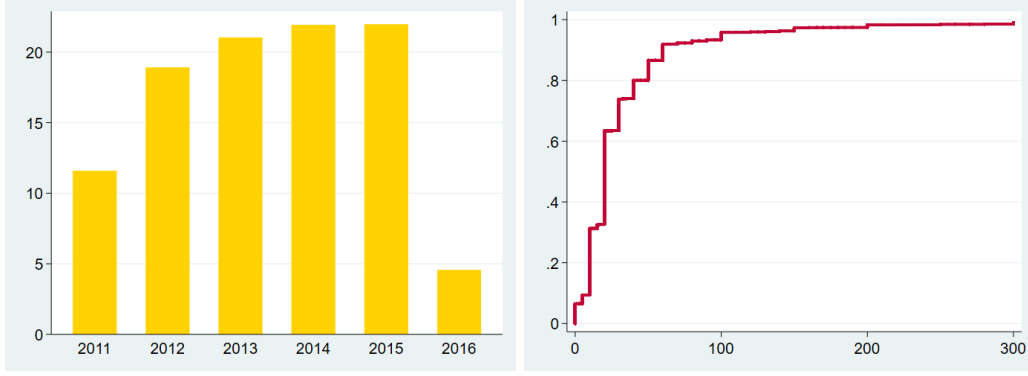


Figure 1: Distribution of violations by year Figure 2: Distribution of fines by face value

Secondly, the following panel displays distribution of violations by weekday, by day in a month and by month. The left panel provides an overview of committed violations based on the weekday. It can be seen that while the number of observations is fairly constant for all working days, the total level of recorded violations is significantly lower during weekends. The middle panel presents the total number of violations by day in a month. It can be seen that there are no large differences among different days, there seems to be no significant pattern. Although, it can be noted that the beginning and the end of the month seem to be slightly less prone to violations. The only significant outlier is the 31st but since this date is underrepresented in the calendar, this drop goes in line with the rest of the distribution. Similarly, recorded violations are fairly evenly distributed across different months. Figure Fig. 3c shows that there are no large differences in the number of municipal violations committed throughout the year besides December. This month seems to be significantly less prone to violations than other months since the proportion of violations committed during this month is by 2 percentage points lower than the overall monthly average. Otherwise, two months with shares above the average are March and October.

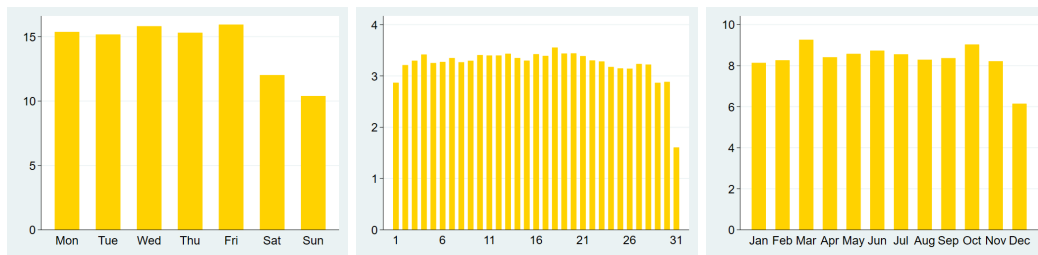


Fig. 3a: Frequency of violations (% of all obs.): By weekday Fig. 3b: Frequency of violations (% of all obs.): By day/month Fig. 3c: Frequency of violations (% of all obs.): By month

Figure 3: Frequency of violations (% of all obs.)

Below, the average fine is plotted by weekday, by day in a month and by month. First of all, it can be seen that the average for fines on Saturday and Sunday is by approximately 4 EUR higher than on any other work day. This is presented in Figure Fig. 4a. This is itself interesting as Saturday

and Sunday are days that can be associated with lower shares on the total number of violations (see Figure Fig. 3a). Moreover, Figure Fig. 4c then displays low average values in the beginning of the year followed by an increase in the consequent months with the average fine peaking in August and then decreasing again towards the end of the year. Notice that while March can be associated with highest share on the total number of violations and one of the lowest average fine, October on the other hand can be associated with second highest share on the total number of violations and also one of the highest average fine. January and February can be associated with low share on the total violations and also mean fine below the overall average. As for the Figure Fig. 4b, no clear pattern can be observed.

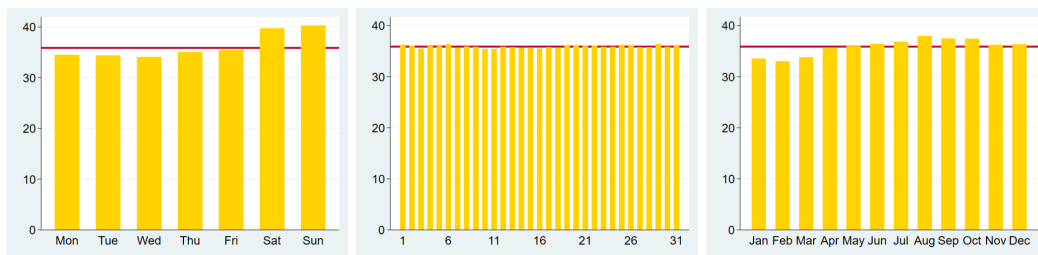


Fig. 4a: Average fine (in EUR): Fig. 4b: Average fine (in EUR): Fig. 4c: Average fine (in EUR):
By weekday By day/month By month

Figure 4: Average fine (in EUR)

5.2 Violation rates

5.2.1 Age

In order to corroborate Hypothesis 1, the municipal violations data has to be analyzed by age cohorts. Figure 5 provides an overview of violation rates by different cohorts and by year. These violation rates express the number of violations per 100,000 individuals in different years committed by individuals from different age cohorts.¹² It should be noted that years 2011 and 2016 do differ from the rest of the dataset. This is probably a data collection issue (see Figure 1 for the overall picture). As the digitalization was only introduced in 2011, not all cases were managed to be reported in the sample. There is no other known objective reason why 2011 would be expected to diverge from the general trend to such a significant extent, when compared to the following years. As for the year 2016, only a small fraction of committed violations was recorded into the general database. It seems like there is a significant administrative lag causing these small recorded violation rates in 2016. This effect can be observed for all cohorts and again, there is no known objective reason for why there would be expected such a notable decrease.

Returning to H1, Figure 5 together with Figure 6 do not fully corroborate the general age-crime pattern discussed in the previous section. While the violation rate does indeed decrease with age,

¹²Demographic data provided by the Statistical Office of the Slovak Republic here.

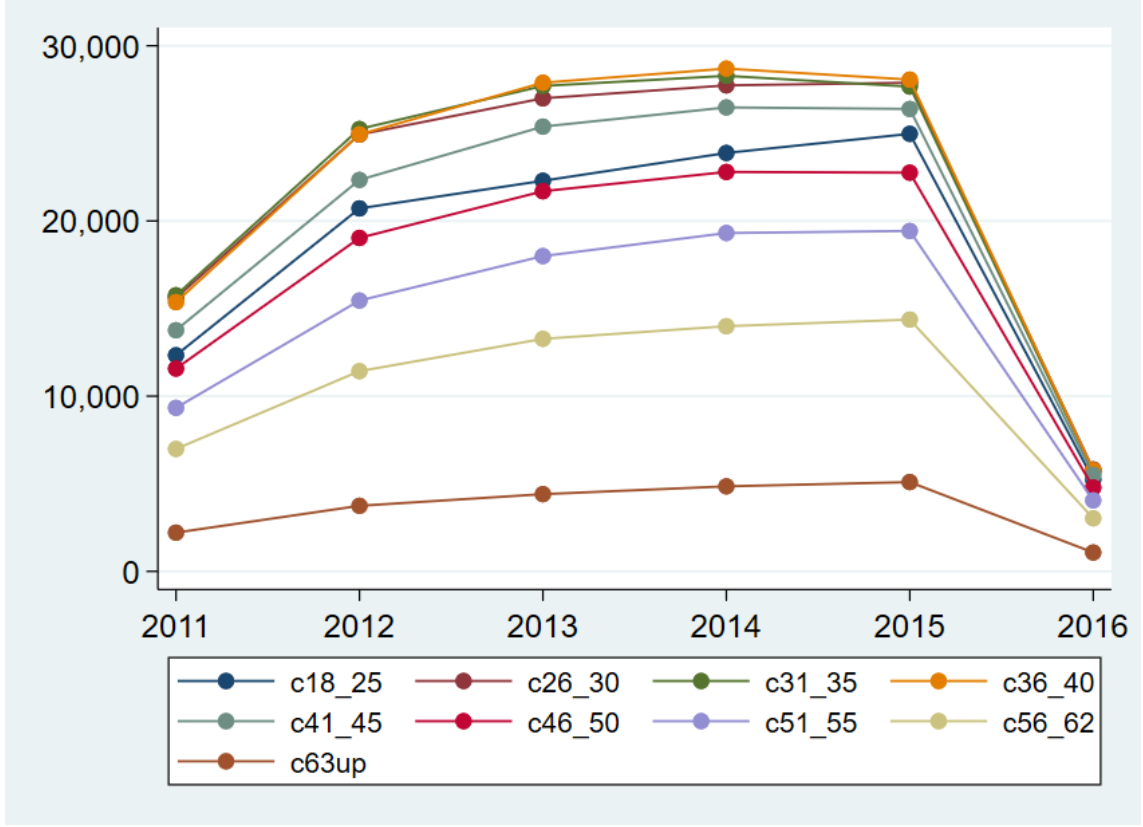


Figure 5: Violation rates by age cohorts (per 100,000 individuals)

this observed rate peaks significantly later than in late teens or early twenties, as that is the common result in the literature. From the data at hand, the overall violation rate rate peaks in late thirties and that is true for all years included in the sample. Moreover, the rate for the youngest cohort is even smaller than the observed rate for the age cohort 41-45. Hence, this observation suggests that there might be differences to the general age-crime pattern once the analysis focuses on violations instead of more serious property or violent crimes. To conclude, the Slovak violation rate by age can be better described with a reversed U-shape than by a gradually decreasing line.

Additionally, Figure 7 reveals that there is only one significant difference in the average fine received by individuals when the mean fine is plotted by different age cohorts. The data shows that on average, the individuals within the oldest cohort receive a fine lower by almost 10 EUR when compared to the average fine over the whole sample. The drop in mean fine between the age cohorts 56-62 and 63+ is statistically significant with p-value being equal to 0.¹³ Moreover, due to large number of observations, almost all pairwise comparisons of the mean difference in average fine for the adjoining age cohorts are statistically different from 0.¹⁴ However, only the drop associated with the oldest age cohort seems to be of practical importance.

¹³Mean fine for the cohort 56-62 is equal to 33.67 EUR, mean fine for the cohort 63+ is equal to 27.45 EUR, using t-test the p-value on the difference is 0.

¹⁴Differences between the cohorts 18-25 vs. 26-30 and 41-45 vs. 46-50 are not statistically significant on 5% level.

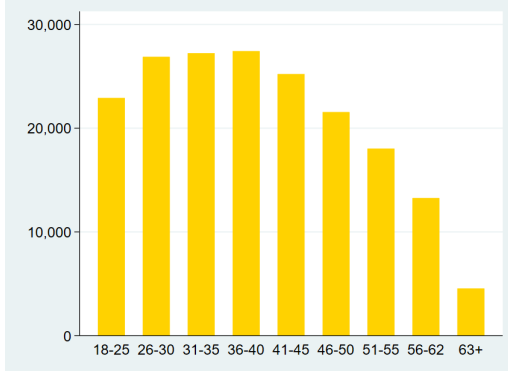


Figure 6: Violation rate by age (2012-2015)

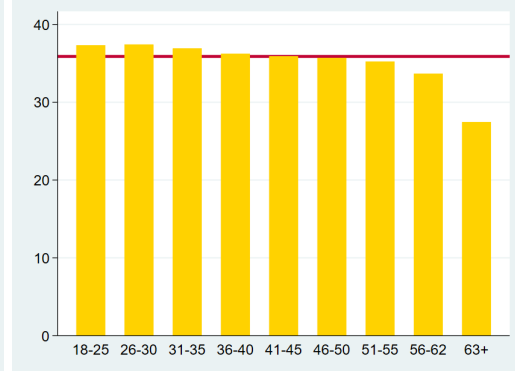


Figure 7: Average fine by age

5.2.2 Education

Moving on to different factors defining the total socio-economic status, it is of interest to analyze fines with respect to education. It is worth noticing that the dataset is restricted, because not all observations could have been matched to the highest achieved educational level. For some observations the level is simply not available. Unfortunately, it is again impossible to examine why some observations had to be dropped (probably it was some sort of an administrative problem with the data collection, but it is difficult to uncover what exactly it was). Moreover, another restriction is that there is no statistical data available that would describe the educational level over the whole population. The best available is the statistics of the highest achieved educational level of all of the individuals in the labor force (either employed or unemployed).¹⁵ For that reason, all observations that could have been traced to individuals still studying at the time when the violation was committed, were dropped from the database for the purpose of this subsection. If kept in the data, such observations would overstate the true violation rates for some education cohorts as such individuals are not officially in the labor force.¹⁶ Therefore, the total number of observations used in the following subsection is restricted to 3,439,667.

The following panel provides an overview of the distribution of violations and the average fine when accounting for the highest achieved educational level. Categories included in the graph are people with no education (*No educ*), the ones who finished the basic 9 year elementary school (*Elem*), people with finished high school with or without graduation diploma (*HS_nograd* and *HS_grad*)¹⁷ and then people with different university diplomas (*Bachelor*, *Master* or *PhD*). It can

¹⁵The data is available from the Statistical Office of the Slovak Republic here.

¹⁶Observations dropped from the database were the ones that had higher achieved educational level than could have been possible, given their age. More precisely, if there was an 25 years old individual with a PhD degree, such individual was dropped because it is extremely improbable that such case would occur in a real life as normally people finish their PhD being 28 and more years old. Therefore, individuals younger than 28 years with PhD, younger than 24 with MSc. and younger than 22 with BSc. degree are dropped and it is implicitly assumed that they were studying at the time when the violation was committed.

¹⁷Practically, individuals included in both of these two categories finished high school. The difference is the type of education they receive. Individuals with finished high school but with no graduation diploma typically obtain

be seen that most of the violations are committed by people with finished high school, but without graduation. Almost 50% of violations are committed by such individuals. As for the average fine, Figure 7 presents mean fines by highest finished education level. No clear pattern can be observed here, it seems like the educational level is not highly correlated with fine. Although, it can be concluded that the mean fine for the people with lower educational background is slightly higher than the total average while the opposite is true for the individuals with higher achieved educational level.

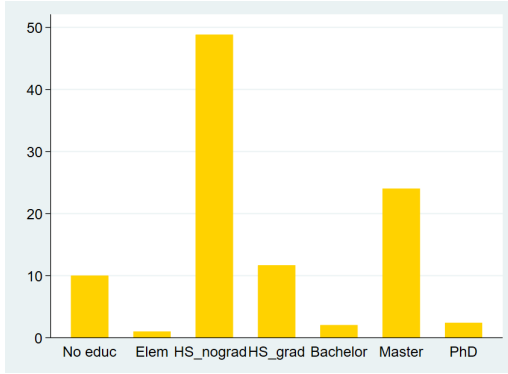


Figure 8: Violations by education level

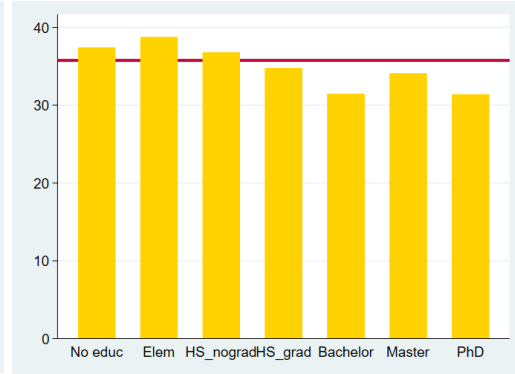


Figure 9: Average fine by education level

Figure 10 presents the violation rates by year by different highest achieved educational levels. It should be noted that all rates are drawn with rates presented on the left y-axis while only the implied PhD violation rate is shown with values linked to the right y-axis. Therefore, the first striking thing is that the violation rate of people with PhD is significantly higher than all the other rates. There is no obvious reason for why this should be the case, not even according to the literature. If anything, the existing literature argues for the exact opposite than what can be observed given the data. Moreover, the existing literature argues that the crime rate can be higher for some very sophisticated crimes where the outcome depends on special skills than can be gained through better education and tenure. But that does not seem to be relevant in this case as this paper analyzes only unskilled crimes.

Otherwise, it can be seen that the highest rate can be associated with people with finished high school but with no graduation diploma. In Slovakia, if one wants to study further in a university, one needs a graduation diploma. Practically, people with no such diploma start working directly after the high school. Typically, they work in services or provide other jobs that are less demanding more practical education. Such students are trained for very specific types of occupations and they do not receive such an intensive training in general knowledge. These individuals are trained to be chefs, car-mechanics, waiters, seamstresses, rangers etc. Therefore, people with high school but no diploma typically start working directly after school when they are 18 years old. On the other hand, individuals with high school and a graduation diploma had to successfully pass the school leaving examination, which is typically required for further studies at universities. Such people obtain more general training in many theoretical subjects and many people with graduation tend to continue with their studies further on. Typically, these individuals are not trained in one particular field or subject.

in terms of skills. Nevertheless, this result is in line with the existing literature that has repeatedly shown that higher crime rates are typically associated with lower educational background. Therefore, it seems like Slovak data can corroborate Hypothesis 2 to only a limited extent. No decreasing pattern with increasing education can be observed. Also, violation rates of people with only elementary education or with passed school leaving examinations are significantly lower than corresponding rates of people with BSc. or MSc. degrees. Therefore it does not seem like violation rates can be associated with some sort of pattern as was true with variable age only.

Interestingly enough, there is a significant difference in the violation rates of people with and without graduation. Another interesting observation is a fact that the violation rate of people with finished master degree is fairly high and can be compared with the rate of people with high school and no graduation. From the first sight, it truly seems like no clear pattern can be observed, when the violation rates are analyzed with respect to education only.

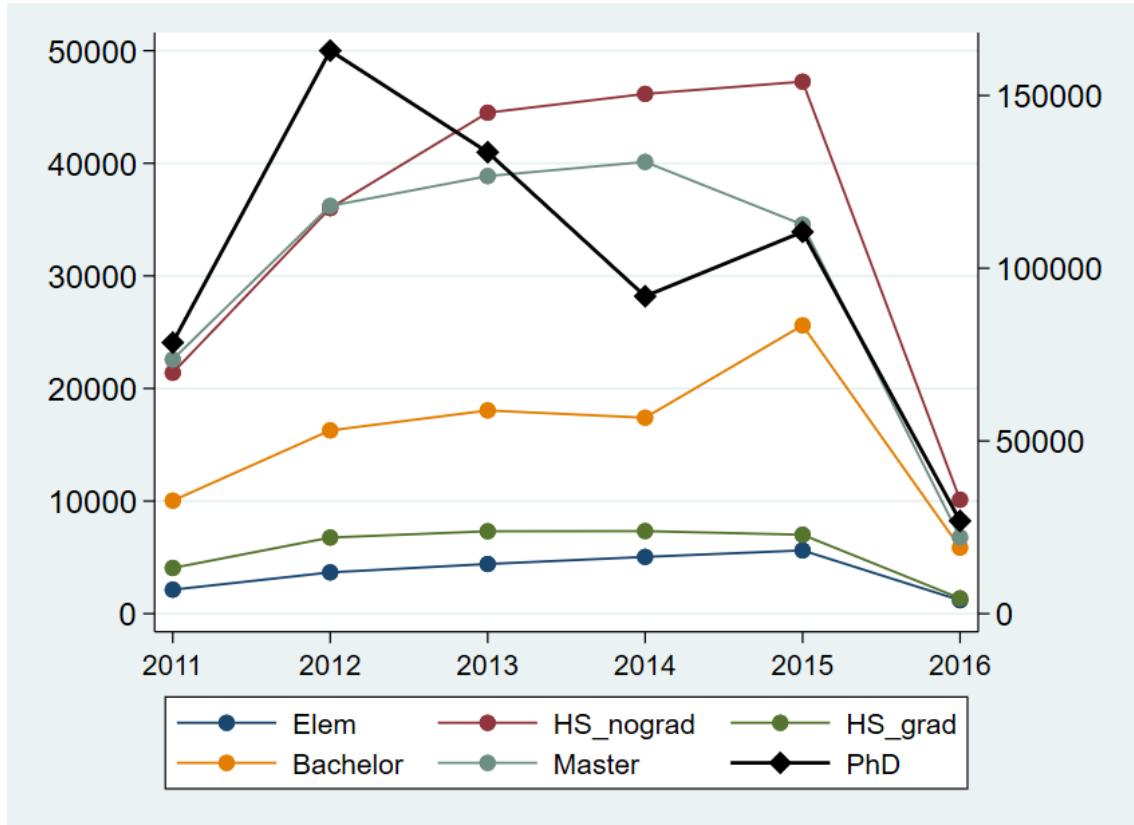


Figure 10: Violation rates by education cohorts by year (per 100,000 individuals)

5.2.3 Unemployment

The following section presents a stylized overview of the relationship between committed violations and unemployment and it provides evidence that allows to check the validity of Hypothesis 3 and Hypothesis 4. From the Figure 11 it is clear that only around 10% of municipal violations are

committed by unemployed individuals.¹⁸ It can be noted that unemployed are actually under-represented in the sample when compared to the general unemployment rate over that period of time. The unemployment rate was decreasing during that period from around 14% in 2011 to 10% in 2016.¹⁹ Moreover, there is no difference between the average fine given to an employed or an unemployed individual, at least based on the data in the sample.

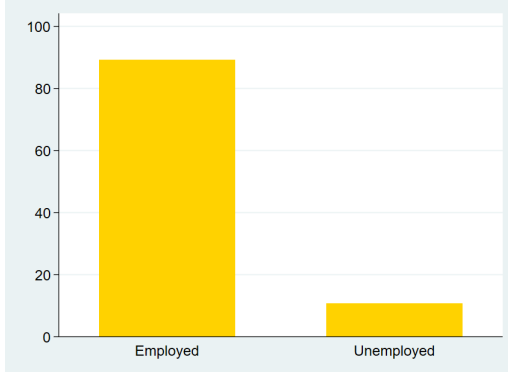


Figure 11: Violations by unemployment

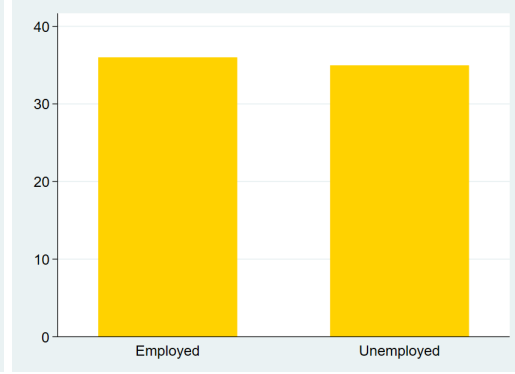


Figure 12: Average fine by unemployment

The following Figure 13 challenges the validity of H3 and H4. First of all, it should be noted that for technical reasons it is not possible to use the whole age distribution of individuals and that the youngest and the oldest cohorts are dropped from the Figure 13. The youngest cohort is problematic because it consists of many people that are still studying. As defined by the law, these students are technically not unemployed nor employed, they are simply not part of the labor force. Therefore, the violation rate of especially employed individuals cannot really be imputed for this age cohort since only a fraction of individuals is already working and there is no analytical way to disentangle which violations were caused by then students and which violations were committed by employed individuals. Hence, the net employed violation rate for the given age cohort cannot be calculated. Since the number of employed people within this cohort is relatively small, that overstates the real violation rate to a great extent. It is slightly easier with the unemployed, since these individuals must be registered at the Office of Labor, so the count here should reflect the reality (and the violation rate fits the general pattern, even though it is not plotted). But since there is no comparison group, this group is dropped for this part of the analysis as well. Similarly, since the retired are not employed nor unemployed, but dropped entirely from the labor force calculations, the violation rate cannot be calculated in this case either.

Again, it can be seen that the overall pattern is in line with H4. In other words, the violation rate decreases with age. However, at least the unemployed violation rate peaks again in late thirties, so significantly later than stated in the literature. The employed violation rate is fairly in line with the already existing literature. One interesting fact to notice is that there is a relatively high violation rate within the oldest (plotted) employed age cohort. It should be noted

¹⁸These individuals were unemployed at the time when the violation was committed.

¹⁹Data available from the Statistical Office of the Slovak Republic and Eurostat.

that this number is probably overestimated as the total number of violations committed by the employed is probably overstated. Again, there is the same problem that while it is fairly easy to identify violations committed by the unemployed (since these people must be in the database actively searching for a job and regularly reporting to the Office of Labor), it is not so clear cut with the employed because some observations currently assigned to not unemployed people might be committed by already retired individuals. For example, by law, mothers are entitled to early retirement and the exact eligibility age depended on the number of children such mother had. So the age of retirement was decreasing as the number of own children was increasing. Individuals are also eligible for early retirement when they turn 60, if they can afford it. Mathematically speaking, this might overstates the total number of violations committed by the employed (since the ones committed by the retired individuals cannot be identified and dropped). Therefore is this result probably overestimating the overall rate within that age cohort and it might be the case that if the data was adjusted for all of the aforementioned factors, that the rate would actually fit the overall observed profile.

As for the H3, the data at hand does not really support this hypothesis. The observed violation rate is higher for the employed individuals in the sample and that holds for all age cohorts. This is an interesting observation, because based on the Becker/Ehrlich model, the unemployed have more to win (or better said, less to lose) when committing crimes and therefore the crime rate is expected to be higher. However, it is important to note that this paper provides an analysis of violation rate. The dataset consists mostly of violations like speeding or being too loud after a specified time. Since it is not so straight forward to say what one might gain in terms of breaking the law in form of speeding or being too loud and bothering own neighbours, it might actually make sense that the employed are the group more prone to committing these types of unskilled crimes as they can simply afford to pay fines for such deeds. From this perspective, it is fairly reasonable to assume that as there is not much to gain under a relatively high probability of being caught and paying the fine, the costs might actually outweigh the benefits for the unemployed and hence, that results in lower observed violation rate.

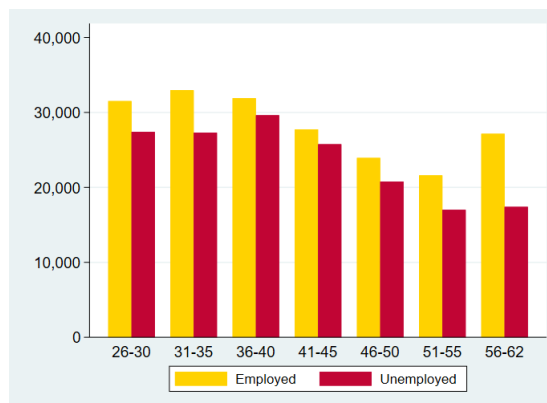


Figure 13: Violation rates by unemployment (2012-2015 data, per 100,000 individuals)

5.2.4 Income

In order to provide some more evidence in favor of or possibly against the Hypothesis 7, this subsection analyzes the available data on municipal violations by yearly income. First of all, it should be noted that the sample is restricted in this case because around one half of all of the observations has to be dropped. That is due to the fact that the dropped observations cannot not be matched to data on incomes. Unfortunately, the reasons explaining this unavailability of data are not at the disposal. This might cause some problems if there is a structural issue causing the data unavailability and hence, this part of analysis should be viewed as an exploratory part for the time until the reason explaining the existing lack of data is known. To conclude, because of this the number of observations used in this part is equal to 1,903,293.

As for some descriptive statistics by income, Figure 14 provides a cumulative distribution of income by different years included in the analysis.²⁰ Income average for the whole sample is equal to 9,943 EUR. That is equivalent to an average monthly income of approximately 829 EUR. A closer detail is presented in the following Table 2, the data introduce sample and economy income averages by years, which are included in the database.²¹ It can be seen that the sample average by year provides a reasonable estimate of the average income in the economy by year.

Year	2011	2012	2013	2014	2015	2016
Monthly average (sample)	760	786	851	844	858	893
Monthly average (economy)	786	805	824	858	883	912
Yearly average (sample)	9 120	9 432	10 212	10 128	10 296	10 716
Yearly average (economy)	9 432	9 660	9 888	10 296	10 596	10 944

Table 2: Average income by year in EUR

Figure 15 then presents average fine by different income cohorts. These cohorts represents nominal monthly income levels. The red line presents the sample average which is equal to 36 EUR. It can be seen that the differences in the average fine are fairly small among different income cohorts and as for now it seems like there is only small correlation between these two variables in the sample. Although, it is fairly interesting to note that the highest average fine can be associated with the second (200-400 EUR) and the tenth (1,800+ EUR) income group while the minimum average is linked to middle parts of the income distribution (nominal monthly income of around 1,000 EUR).

The following Figure 16 presents a stylized overview of the violation rates by income groups and by year.²² Again, it should be noted that it is clear that years 2011 and 2016 are underreported,

²⁰The graph does not plot the whole distribution, but only approximately 90% of it. This is to provide a detailed overview on yearly income of the lower income groups, which dominate the dataset. Adding the highest income classes makes the graph completely illegible.

²¹The data on average income in the whole economy can be found here.

²²Income distribution data available from the Statistical Office of the Slovak Republic here and information about

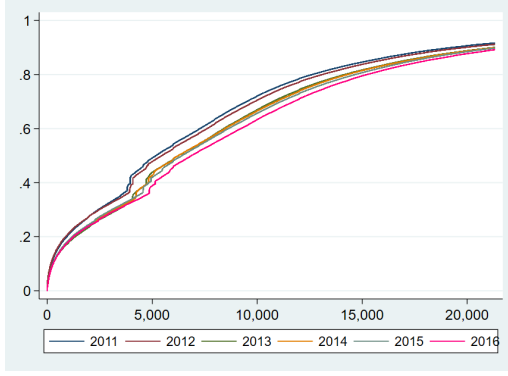
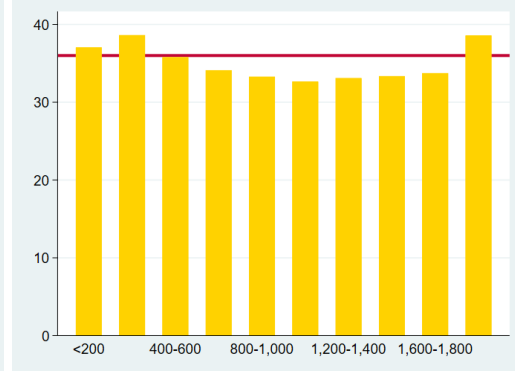


Figure 14: Distribution of wages by year Figure 15: Average fine by income cohorts

when compared to the rest of the sample. However, the data for 2011 shows that this time this year fits the overall pattern by income group better than before. Unfortunately, year 2016 suffers from significant reporting issues. Otherwise, there is no objective reason why it would be reasonable to observe such a significant drop in the violation rate for all income cohorts, as no significant behavioral nor legislative change can be traced back to 2016.

Moreover, there seems to be another reporting issue. It can immediately be seen that the overall pattern is not fitted by the lowest income group (with a monthly reported income below 200 EUR). Notably, the black line (with y-values represented on the right axis instead of the left one, which is used for all the other income cohorts) demonstrating the violation rate by year for the lowest income group attains significantly higher values than all the other income groups. Most likely it is a data issue as there is no other reason why it would be reasonable to believe that the poorest individuals would commit at least 20-times more violations than any other income group. Mathematically, it must be the case that this income group is significantly underrepresented in the official statistical data on incomes or overrepresented in the database on violations, which might cause the violation rate to reach such large and unrealistic value (probably it is a combination of both problems).

As for the rest of the sample, the highest observed violation rates can be associated with income groups 200-400 EUR and 1,800+ EUR (which are the same groups that can be associated with the highest average fine levels, more in Figure 15). While the former dominates the later years (2014-2016), the latter dominates in the beginning of the sample (2011-2013). Based on the data at hand, it seems like in the beginning of the period of interest, the rich were more likely to break the law while the violation rate of the poor was slowly increasing over years until it outrun the rich. So it can be concluded that the data for the poor and the rich tell a completely different story from the rest of the sample. That being said, the rest of the data is fairly nicely ordered as it can be seen that the violation rate then decreases with decreasing income. This pattern is violated when accounting for the income group 400-600 EUR as this rate is greater than the rate the total number of employed people can be found here.

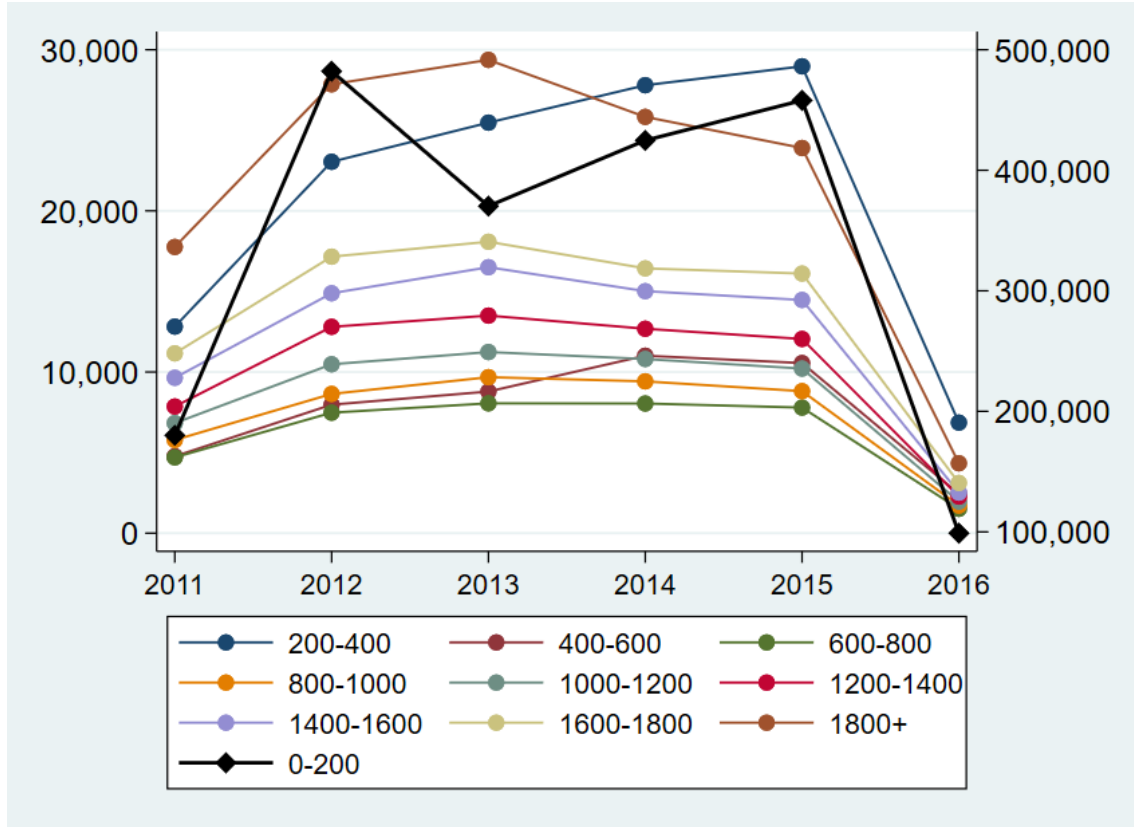


Figure 16: Violation rates by income cohorts by year (per 100,000 individuals)

of the income group 600-800 EUR. Moreover, there is a sudden increase in the violation rate of the group 400-600 EUR from 2013 to 2014. Overall, the lowest violation rate is recorded for the income group 600-800 EUR and that holds throughout the whole period of interest.

Figure 17 below provides a standardized overview of the violation rate by different income cohorts using data from 2012-2015 (2011 and 2016 are dropped due to underreporting as before). Additionally, it should be noted that the lowest income cohort is also dropped from this Figure 17. This is due to the fact that this rate is extremely high (as it is presented in the Figure 16 above in black) and that would make the graph illegible. Mathematically and based on the data at hand, implied violation rate for the income cohort with an average monthly income below 200 EUR is equal to 431,190 violations per 100,000 individuals and that seems to be very unrealistic.

Overall, it can be seen that the distribution is somehow in line with both types of arguments, that were presented above in the literature review. Notably, one can observe a significant U-shape of this rate with peaks at the both ends of the income distribution. First line of the argument claimed that the poor are more prone to crime as they have more to gain. It should be noted, however, that municipal violations consist of a very specific subset of deeds breaking the law and that it is overall difficult to argue what one gains when speeds or is too loud after a specific hour. But still it seems like the lowest income cohorts can be associated with significantly higher violation

rates. Secondly, the data is also in line with the opposite argument that the rich are more prone to crime as they can simply afford it. Indeed, the violation rate increases with income. Interestingly enough, the violation rate for the lowest and the highest plotted income cohort is almost the same. Hence, it can be concluded that H7 seems to be valid but the poor are also very likely to commit a violation so this Hypothesis 7 should be adjusted for.

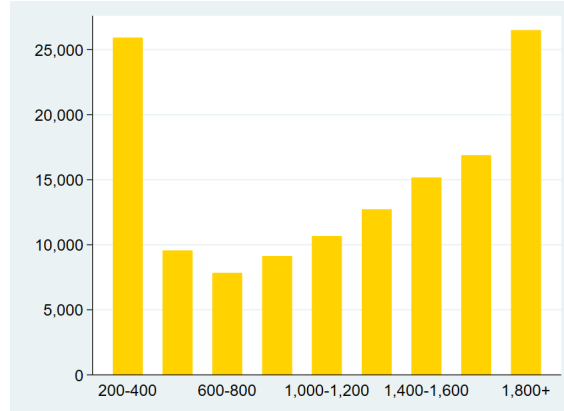


Figure 17: Violation rates by income per month (2012-2015, per 100,000 individuals)

5.2.5 Occupation

Available data allows for an analysis of violation rates by occupation. For this purpose, the data on OKEC is available. OKEC is an older type of Slovak statistical classification of economic activity and practically, it sorts different occupations into broader categories.²³ OKEC can be consequently transformed into NACE classification, which is a classification created by Eurostat (broad structure can be found in the Appendix, Figure A.2).²⁴ NACE classification is used for the following parts of the analysis as it provides a logical grouping of occupations by type. Moreover, it is convenient to use this classification also due to the fact that the Statistical Office of the Slovak Republic currently reports the numbers of employed within different occupation types in line with the NACE classification and these are used for calculations of violation rates within different occupations.

In order to perform this part of the analysis, the final dataset is restricted to a significantly smaller number of observations (1,903,239). There are two reasons why this must be the case. First of all, one reason is that the *OKEC* code is simply not available for all observations. Again, it is not known why this is the case and whether it is a structural problem linked to how the data is collected and consequently reported. Also, since this data is collected together with data on income, it is impossible to disentangle whether this undercoverage stems from issues linked to income reporting or OKEC reporting. That is itself unfortunate, on the other hand, at least the number of observations is fairly large so that should compensate for possible structural problems

²³The list of all *OKEC* codes can be found here.

²⁴Broad structure and a manual can be found here.

to at least some extent. Moreover, as for the structure of people included in the dataset by OKEC code, shares of different employment groups reported in the data are fairly in line with the overall distribution of people by NACE code of the whole country (as it has been mentioned, it is possible to transform OKEC into NACE, which is a standardized format used in the European Union nowadays).²⁵ Hence, at least some factors indicate that the problem is probably not structurally linked to only one occupation type. However, there might be some regional effects present that cannot be accounted for due to the nature of the data. Occupation types by NACE are not really divided into subgroups depending on counties, such data is not available, therefore it is impossible to check whether there is a region systematically misreporting information, which could bias the results.

The following graph provides a basic descriptive statistics, as it was the case with other variables before. Figure 18 presents violation rates by occupation type, as defined by NACE (in gold). Note that two groups from the NACE classification are not reported (so the total number of observations shrinks to 1,849,281). This concerns groups T (*Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use*) and U (*Activities of extraterritorial organizations and bodies*). Group T is ignored because no violation is even recorded for this group. Hence, implied violation rate is 0, but this seems to be a mistake in reporting. It might be the case that many people practically working within this sector are not really recorded in the data as they might choose not to report to the government. Even the group reported in the official data consists of couple hundred of people per year. As for the other group, implied violation rate of group U is more than 1 million cases per 100,000 so that indicates data issue.

Moreover, this graph also presents the average yearly income (in red) per particular occupation type (in thousands of EUR). Only data from 2012-2015 are used here, as the number of observations available for years 2011 and 2016 is underreported and that might bias the violation rates downwards (so Figure 18 is based on 1,542,388 observations). It can be seen that three sectors can be associated with significantly higher levels of violation rates. More precisely, it is the groups L (*Real estate activities*), N (*Administrative and support service activities*) and S (*Other service activities*). On the other hand, the lowest levels can be linked to groups E (*Water supply; sewerage, waste management and remediation activities*), J (*Information and communication*), O (*Public administration and defence; compulsory social security*), P (*Education*), Q (*Human health and social work activities*) and R (*Arts, entertainment and recreation*). Based on these aggregates, there seems to be no particular correlation between average income and the violation rate by occupation type.

As for the Figure 19, it can be seen that a significantly greater variability can be observed, when compared to previous graphs presenting average fine by some specification of cohorts. Occupation

²⁵The data can be downloaded from the Statistical Office of the Slovak Republic here.

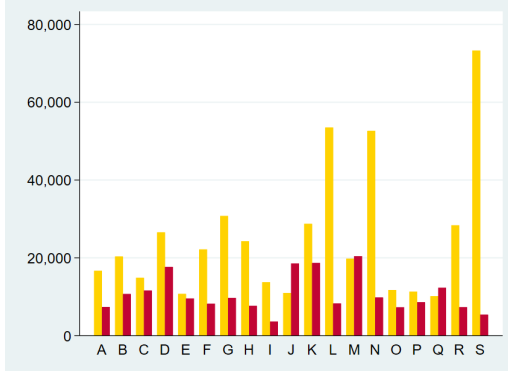


Figure 18: Violation rates by NACE (12-15)



Figure 19: Average fine by NACE

types with the highest average fine are groups H (*Transportation and storage*) and E (*Water supply; sewerage, waste management and remediation activities*) with average being greater than 40 EUR. On the other hand, group P (*Education*) can be associated with the lowest average fine per group. Interestingly enough, group P scored low in both categories. People working in sector *Education* are subject to one of the lowest violation rates and also to the lowest average fine per cohort. Similar is true for people working within the sector *Human health and social work activities*, which is basically the sector of health care. On the contrary, there is no group that can be associated with high violation rate and also high average fine. However, group E can be linked to one of the highest average fines per occupation type while this group has one of the lowest violation rate in the sample.

5.2.6 Gender

The analysis by gender is supposed to check the validity of Hypothesis 5 and Hypothesis 6. Hence, based on the collected information (full sample is used), males are indeed significantly more likely to commit a municipal violation than females. The data shows that around 80% of all recorded violations are committed by men. This means that men are significantly overrepresented in this dataset, assuming that the ratio of men to women is approximately 1:1 in the general population.²⁶ Hence, the evidence is in line with H5.

Moreover, the Figure 21 shows that men receive higher fine on average than females. This difference in mean fine is equal to approximately 12 EUR. Testing for a difference in mean fine between genders is statistically significant with p-value being practically 0.

As for the H6, the following Figure 22 presents the age-crime profiles by gender. The overall trend is similar to the one observed in Figure 5, although here it is decomposed to demonstrate the differences between males and females. Obviously, the female violation rate is significantly lower than the male counter part. To be more precise, males violation rate is approximately four times higher than the female violation rate. Overall, it can be concluded that the violation rates do

²⁶Source of the demographics data by gender: Statistical Office of the Slovak Republic

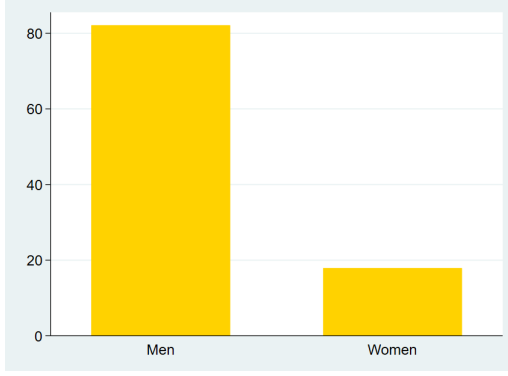


Figure 20: Share on violations by gender

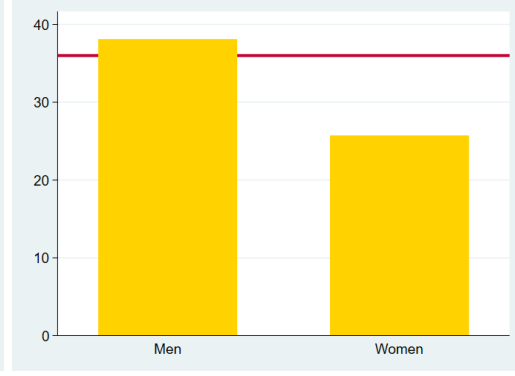


Figure 21: Average fine by gender

decrease with age, which supports H6. However, both rates peak significantly later than is stated in the aforementioned literature. Moreover, there are also differences between genders since the male rate peaks in the late twenties and the female rate peaks in late thirties. Therefore, the data supports the overall trend as claimed by H6, but peaks are observed for older than expected age cohorts.

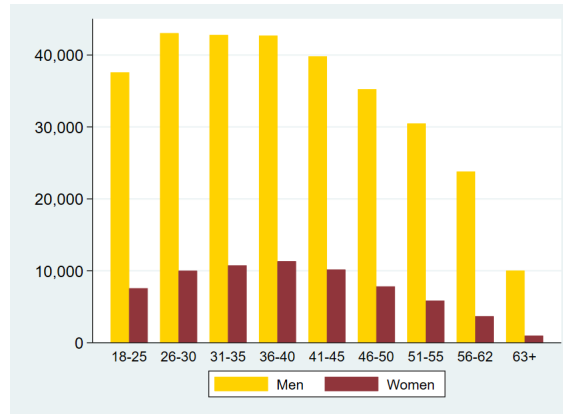


Figure 22: Violation rates by gender (2012-2015, per 100,000 individuals)

5.3 Seasonal effects

As for now, the analysis provides mostly descriptive results which are trying to corroborate the already observed stylized facts collected in a literature review. This subsection tries to analyze the data on municipal violations from the perspective that the timing of income might affect the likelihood of committing these violations. Overall, the idea is to look at two different effects. First of all, sectors for which the exact pay day can be identified are kept in the dataset and then the total number of violations is plotted by day in a month. That is to see whether the total number of violations depends on the pay day or not. Under this setting, the data set is at first divided into two subgroups - the poor and the rich - with the cut off line being the average wage. Naturally, as the poor is identified a group of individuals with incomes below the average wage. The rich

are defined vice versa. After that, all individuals are divided into smaller subgroups by income distribution percentiles in order to see whether possibly there is an effect linked to a particular income group only. This is to account for the fact that basic division into the poor and the rich can oversimplify the reality. Using the data at hand, people with yearly incomes of a couple hundreds of EUR are in the same group as someone earning around nine thousands per year. Both of these cases are below the average wage cut off line. But obviously, these groups are not poor to the same extent, the latter group faces significantly other problems than people with only a couple of hundreds EUR. Therefore, the income distribution is analyzed to a greater detail.

The total number of observations used in this analysis is 1,903,239. That is the total number of observations with identified OKEC code. Unfortunately, even some observations with known OKEC/NACE code must be dropped later on, because in many cases it is simply not possible to impute the pay day (but this concerns mostly the following subsection, not the one investigating seasonality).

5.3.1 Effect of the pay day

As it was already discussed previously, there might be an effect of the exact pay day on the behavior of people (Anderoni et al., 2017; Carvalho et al., 2016; Mani et al., 2013). Hence, this part tries to corroborate the hypothesis whether it is possible to observe a similar effect with respect to the total number of committed violations when one accounts for the exact pay day. Therefore, the final sample must be restricted to include only several NACE codes for which the pay day is known.

Sectors, where it is possible to identify the exact pay day are mostly the ones which are accounted for as a public sector. Therefore it is the NACE sector O (*Public administration and defence; compulsory social security*), P (*Education*) and Q (*Human health and social work activities*). Unfortunately, it must be noted that not all employees registered in these sectors do necessarily work for the government as civil servants. Some fraction of employees works for private run schools or hospitals. However, it is impossible to determine individuals working within the private sector. Fortunately, this fraction should be relatively small so the results should not be biased to a significant extent.

Based on a common knowledge, payments from the government typically arrive at the bank accounts of the subordinate organizations around 10th day every month. Depending on the number of organizations between the central government (for example the Ministry of Finance) and the final subordinate organization paying off wages (a local school), the wages are being paid off between the tenth and the fifteenth day of every month. Naturally, the more organizations participate in the transfers of money, the longer it takes for the workers to receive their wages. Therefore, wages are typically paid off before 15th day every month.²⁷

²⁷More about the timing of payments from the central government can be found here (wages of teachers) or here.

As for the whole dataset, no clear pattern is visible with respect to days in a month. There is also no significant pattern visible when the number of violations is plotted by day in a month for individuals with own income below and above the average wage separately (for the overview, see Figure A.3). However, a closer inspection shows that the 14th and 15th day in a month can be associated with a small decrease in the total number of violations (when compared to the rates during the previous couple of days). There is also a consequent increase in the total number of violations on the following days. Moreover, this effect can be observed within both groups. But this fact should be interpreted with caution as there are also other larger drops and increases visible that one cannot really account for. Moreover, Figure A.4 presents the shares of violations committed on different days by income percentiles. Income percentile 1 consists of people with the lowest incomes while 9th income percentile consists of people with the highest incomes. As it can be seen, the distributions of shares of violations do not vary much across different income cohorts. Hence, it does not seem that there would be an effect on the aggregate level. However, it must be noted that the pay off day varies for different firms and occupations therefore it is not surprising that no strong effect is visible on the aggregate data.

Moving on to the analysis within different occupation types, Figure A.5, Figure A.6 and Figure A.7 present shares on the total number of violations by day in a month and by sector. As it was explained before, these three sectors are to large extent publicly run and hence, they can be analyzed under the simplifying assumption that the pay day is somewhere between 10th and 15th every month. Moreover, these graphs project the proportion of violations committed on a given day by people within a certain income percentile. A brief look indicates no clear pattern with respect to different income cohorts nor with respect to different occupation types by NACE. Although, when compared to the overall representation showed in Figure A.4, these graphs present greater variability in shares of violations with respect to different days, the overall case can be associated with fairly flat trend for all income cohorts during the whole month. So this detail reveals that the evidence is very mixed. While for some income cohorts it seems like the same as for the case presented in Figure A.2 can be identified, the effect cannot even be consistently observed for the same income cohort across different occupation types. There is no clear evidence that a pattern, which could be specifically linked to the pay day, is present in the data.

6 Discussion

As the obvious data limitations have been discussed already in one of the previous chapters, this section presents hypotheses which might explain some striking observations found based on the data at hand. Most of the topics discussed below should be viewed as open questions for further research and explanations provided here are subject to discussion and critical revision.

First of all, it is interesting to notice Figure Fig. 3a together with Figure Fig. 4a. It can be seen that significantly smaller volumes of violations are committed during weekends. On the other hand, these violations can be associated with significantly higher average fine per case. That might indicate that the structure of violations committed during weekends might be different from the structure observed during working days. Another part of the explanation might be that the police force operates differently during weekends and that they might focus on different crimes during the working days as compared to the weekend. These two factors might explain the observed difference in average fine of almost 4 EUR. That represents a statistically significant increase as the weekend average fine is by more than 10 percent higher when compared to the average observed fine during working days. This result can be considered as a fairly robust one. It might be relevant to uncover the true reason causing this pattern so that the police and other resources can be used more efficiently. In other words, it might be more efficient for them to focus on some specific types of violations during the working days and to prevent other violations from happening during the weekends, if it is true that different violations are being committed. Similar argument can be made for the observations as plotted by month (see Figure Fig. 3c and Figure Fig. 4c). These two graphs also signalize possible structural breaks with respect to different months and it might be worth to investigate this further.

As for the age-crime pattern and Hypothesis 1, the overall pattern is in line with the existing literature. However, it is interesting to observe that this rate peaks later than is typically observed for crime rates in the literature. It might be due to the nature of the sample, which is analyzed in this paper. As the dataset consists mostly of speeding tickets, it might be the case that the youngest cohort simply does not drive that much, as not so many cars are owned within this age cohort. As many students are included there, they typically do not tend to use cars that often as for example parents who drive their kids around to school and other extracurricular activities. Therefore, here it might be in line with an old Slovak saying which can be paraphrased such that one simply needs an opportunity to commit a violation. Maybe, if the young had more opportunities, the violation rate would be similar as is for the people in their thirties. And as the youngsters have significantly less opportunities to commit these violations, the observed rate is therefore lower.

Overall, the average fine decreases as the age increases. Moreover, it is interesting that Figure 7 presents that the oldest age cohort can be associated with significantly lower level of average fine than all the other age cohorts. That again indicates that the structure of committed violations is probably different when compared to the overall distribution of violations. Unfortunately, due to the nature of data it is impossible to define how the structure differs. But this might be an interesting question for further research.

Figure 10 presents violation rates by the highest achieved education level. It can be seen that the PhD rate does not fit the overall pattern very well. More precisely, the recorded violation rate

is several times higher than observed rates for all the other education levels. This is probably a statistical discrepancy as there is no obvious reason for why the PhD rate should be significantly out of scale with respect to all the other rates. It might be the case that people with PhD are underrepresented in the sample retrieved from the Statistical Office of the Slovak Republic. That would in turn cause the violation rate to sky-rocket. As individuals must report to the database their highest achieved education degree, it might be the case that many people decided not to report and as a result, the rate is overestimated.

Another interesting observation is the fact that the violation rate of people with finished master degree is fairly high and can be compared with the rate of people with high school and no graduation. These might be caused by a statistical discrepancy, if the data provided by the Statistical Office understates the true levels. That might be the case if it is assumed that many adults are not accounted for in the Slovak labor force if they work in foreign countries, but might cause violations here. Another problem might be that the number of people within different groups might be overstated and maybe there are some groups that the should have been dropped additionally. Also, as people can report information about their highest achieved education level, but it is not mandatory to do so, the overall recorded levels of people with a master degree might understate the true levels. Finally, if these rates are truly higher and the Figure 10 is a correct representation of the reality, than that could be in line with the hypothesis that higher income stems from better education and these individuals might actually just afford to commit these deeds and pay the fine. That is also a relevant option, but it is very difficult to disentangle these different effects based on the data at hand.

Interestingly enough, there is a significant difference in the violation rates of people with and without graduation (see also Figure 10). There is again no obvious reason for why this should be the case, as both group have been studying for the same period and people in both groups started working directly after school (otherwise they would be recorded in other group). Therefore, as for the background, these groups are fairly comparable. Yet, the observed violation rates are significantly different.

One of the most interesting results is presented in Figure 17. As it was mentioned in the literature review, overall there are two theories about the crime rates. One argues that the poor are more likely to crime as they have more to gain. The second line argues that the rich might also be prone to crime as they can afford it. In case of municipal violations, the latter makes sense as the dataset consists of speeding tickets and similar deeds against the law. Therefore, the second line of argument is easily applicable to this particular situation, it might be the case that the fine is seen as a fair price to pay for this unlawful behavior. It is however way more difficult to explain the former. Logically, the poor do not fit the pattern as much. Yet, the violation rate is the same as for the rich. The dataset does not consists of crimes that would benefit them to any significant

extent. If such an open case is taken to court, then they possibly might face significantly worse monetary consequences than the rich who just pay it immediately. It is difficult to explain how the poor might benefit from speeding as they cannot afford it. It is not the case that the Becker model would be directly applicable to this particular situation, this model does not really explain the behavior of the poor. One reason might be that the poor simply do not care. As they do not have the resources to pay the fine anyway, they do not face any immediate trade-off. The potential risk of being convicted and punished can be viewed as a very distant future. Then, in line with the argument that the poor might be more present-biased, this risk of punishment is discounted to a significant extent that it is not considered that important when making a decision whether to break the law or not. Hence, maybe even the pure joy of speeding might outweigh the potential risks, which might take their time due to many administrative lags. Another explanation might be strategic inattention. The poor might suffer from significant scarcity of time and committing a violation might be seen as a fair price for coping with other difficult situations. Also, they might simply live in a more risky environment where many people commit such traffic or other violations. Then it might be considered socially acceptable to commit unlawful deeds. It would be interesting to decompose this result into factors causing it to be present.

Linked to the same Figure 17, it is interesting to notice that the group least prone to crime consists of individuals with an average monthly income around 600 EUR. This amount is below the average income in the economy during the whole period of interest. It is difficult to explain, why this might be the case. One reason might be that that these individuals are less present-biased than the poorest group. But this can probably explain only one part of the final effect, because even with income around 600 EUR, these people cannot really be considered a middle class just yet and this income level can still be associated with significant stress levels. However, it might be the case that due to limited resources, these individuals can correctly evaluate their financial position and might decide not to commit violations as they yield little gain under a great costs.

Interestingly enough, when put into context, it can be seen that people in the middle of the whole income distribution are not only significantly less likely to commit a violation. They also receive on average lower fines than the poor and the rich (see Figure 15). That might again signalize that there is a structural change in the type of deeds which are committed by the poor, by the rich and by the people in the middle of the distribution. This might be an interesting topic for further research, when the data describing the exact type of deeds are available.

One income group is not included in Figure 17. Figure 16 also presents the implied violation rate of individuals with yearly incomes of less than 200 EUR. It can be seen that this rate suffers from similar issue as the implied violation rate of people with PhD diplomas. This violation rate of the income cohort with the lowest yearly wages is significantly higher than all the other rates. As the rest of the violation rates for all the other cohorts are reasonable, then there must be some

sort of statistical discrepancy linked to this particular income cohort. Again, they are probably underrepresented in the data downloaded from the Statistical Office. It might be the case that many people with low recorded income in the database collected on fines are not at all included in the database provided by the Statistical Office. For example, if such people receive unemployment benefits, there is a probability that these individuals are not recorded in that particular database as such transfers are not accounted for as labor incomes. Therefore, many people might simply not be reported in the used sample provided by the Statistical Office, which in turn causes the rate to sky-rocket.

It is interesting to return to Figure 18 and to Figure 19. As it was mentioned already, two groups can be associated with low average fine and also fairly low violation rates and that is the people working in sector P (*Education*) and Q (*Healthcare*). This is itself interesting as these two sectors cover occupations, which are significantly pro-socially oriented. Also, these two sectors can be associated with least anti-social behavior. It might be the case that due to the pure nature of these occupation types, spillover effects can be observed in other sectors, too. In this case that would imply lower violation rates and also lower fines, which could be understood as that these individuals commit less severe violations and that they tend to behave less anti-socially.

Finally, due to a lack of data about the pay day on the individual level, the available data could have been analyzed only on the aggregate level. Naturally, more thorough analysis would be required to analyze this effect, but that would only be possible if there was data on the individual level. At first, it would be enough to at least control for the public sector only but OKEC did not allow for that. Hence, the analysis could have been run only under a very simplifying assumption that the pay day is on the 15th every month and for the entire sector (even the private institutions). This can definitely be improved in further research.

7 Conclusion

The goal of this paper was to analyze the data on municipal violations committed in Slovakia during the period 2011-2016. These administrative data were matched with different demographic and socio-economic data which allowed for analysis of different violation rates. First of all, the task was to investigate whether different stylized facts typically found in the literature on crime rates could be replicated based on the Slovak data. This served as an instructive empirical exercise as the data at hand comprised only information on municipal violations. Therefore it was interesting to investigate whether these already standardized results about crime rates could be replicated using only a subset of all crime data. Secondly, the analysis tried to uncover whether the rich and the poor tend to behave differently. Finally, the last part of this paper tried to search for effects that could be attributed to the timing of the individual pay day.

First of all, the analysis focused on the variable age and the typical age-crime pattern could have been found. However, it should be noted that the observed violation rate peaks significantly later than the typically observed crime rate. Secondly, as for the education, the highest violation rate could be attributed to individuals with finished high school without a graduation. Nevertheless, no clear pattern between the violation rate and the education could be observed as the second highest rate could be attributed to people with finished master degree and on the exact opposite, the lowest rate could be attributed to people with elementary education. Thirdly, the unemployed could be associated with lower violation rates than the employed, but the employed violation rate was overstated due to the nature of the available data. Moreover, the overall age-crime profile could be observed but the peak was still observed later than typically in the literature. As for the gender, here the observed results replicated the standard results to quite some extent. It was true that men were significantly more likely to commit a violation but again, the age-crime profiles peaked later than is normally the case with crime rates. Moreover, it can be concluded that both hypotheses providing explanations whether the rich or the poor are more likely to commit violations seems to be true as both extreme income cases could be linked to significantly higher violation rates than the rest of the sample. And as for the effect of the pay day, no evidence supporting the existence of this effect has been found. But unfortunately, the data did not allow for a very thorough analysis of this effect as the data on pay day was not available on the individual level.

As for the future research, the aforementioned descriptive results raised many interesting questions, that could not have been answered within the scope of this thesis, but could serve as new starting points for further investigation on the matter. For example, it would be interesting to explain why the poor could be associated with significantly higher violation rates than individuals with average incomes. Moreover, if there was more detailed data available about individual pay days, that would allow for more thorough analysis of this effect as the overall analysis was run under the simplifying assumption that they pay day was in the middle of every month. More detailed data would allow for a more precise analysis. All of these factors could help to explain what drives different violation patterns and this can eventually be used to optimize the use of resources within the sector of public safety.

Bibliography

- James Anderoni, Nikos Nikiforakis, and Jan Stoop. Are the rich more selfish than the poor, or do they just have more money? A natural field experiment. *National Bureau of Economic Research*, w23229, 2017.
- James Andreoni, Brian Erard, and Jonathan Feinstein. Tax compliance. *Journal of Economic Literature*, 36(2):818–860, 1998.
- Gary S. Becker. Crime and punishment: An economic approach. *Palgrave Macmillan UK, The economic dimensions of crime*, pages 13–68, 1968.
- Daniel M. Blonigen. Explaining the relationship between age and crime: Contributions from the developmental literature on personality. *Clinical Psychology Review*, 30(1):89–100, 2010.
- Leandro S. Carvalho, Stephan Meier, and Stephanie W. Wang. Poverty and economic decision-making: Evidence from changes in financial resources at payday. *American Economic Review*, 106(2):260–284, 2016.
- Dennis Cox. Raising revenue in the underground economy. *National Tax Journal*, 37(3):283–288, 1984.
- Mirko Draca and Stephen Machin. Crime and economic incentives. *Economics*, 7(1):389–408, 2015.
- Mirko Draca, Theodore Koutmeridis, and Stephen J. Machin. The changing returns to crime: Do criminals respond to prices? *London School of Economics*, 2015.
- David Dubois, Derek D. Rucker, and Adam D. Galinsky. Social class, power, and selfishness: When and why upper and lower class individuals behave unethically. *Journal of Personality and Social Psychology*, 108(3):436, 2015.
- Horst Entorf and Hannes Spengler. Socioeconomic and demographic factors of crime in Germany: Evidence from panel data of the German states. *International Review of Law and Economics*, 20(1):75–106, 2000.
- David P. Farrington, Bernard Gallagher, Lynda Morley, Raymond J. St. Ledger, and Donald J. West. Unemployment, school leaving, and crime. *The British Journal of Criminology*, 26(4): 335–356, 1986.
- Denis Fougère, Francis Kramarz, and Julien Pouget. Youth unemployment and crime in France. *Journal of the European Economic Association*, 7(5):909–938, 2009.
- Richard B. Freeman. Why do so many young American men commit crimes and what might we do about it? *National Bureau of Economic Research*, No. w5451, 1996.
- Richard B. Freeman. The economics of crime. *Handbook of Labor Economics*, 3:3529–3571, 1999.

- Edward L. Glaeser, Bruce Sacerdote, and Jose A. Scheinkman. Crime and social interactions. *The Quarterly Journal of Economics*, 111(2):507–548, 1996.
- Uri Gneezy and Aldo Rustichini. A fine is a price. *Journal of Legal Studies*, 29(1):1–17, 2000.
- Eric D. Gould, Bruce A. Weinberg, and David B. Mustard. Crime rates and local labor market opportunities in the United States: 1979-1997. *The Review of Economics and Statistics*, 84(1): 4–61, 2002.
- John Graham and Benjamin Bowling. Young people and crime. *Home Office Research Study*, 145, 1995.
- Jeff Grogger. Market wages and youth crime. *Journal of Labor Economics*, 16(4):756–791, 1998.
- Kirstine Hansen. Education and the crime-age profile. *British Journal of Criminology*, 43(1): 141–168, 2003.
- Johannes Haushofer and Ernst Fehr. On the psychology of poverty. *Science*, 344(6186):862–867, 2014.
- Frances Heidensohn. The future of feminist criminology. *Crime, Media, Culture*, 8(2):123–134, 2012.
- Travis Hirschi and Michael Gottfredson. Age and the explanation of crime. *The American Journal of Sociology*, 89(3):552–584, 1983.
- Travis Hirschi and Michael R. Gottfredson. Control theory and the life-course perspective. *Studies on Crime and Crime Prevention*, 1995.
- Dean Karlan, Lakshmi A. Ratan, and Jonathan Zinman. Savings by and for the poor: A research review and agenda. *Review of Income and Wealth*, 60(1):36–78, 2014.
- Michael W. Kraus, Paul K. Piff, and Dacher Keltner. Social class, sense of control, and social explanation. *Journal of Personality and Social Psychology*, 341(6149):976–980, 2009.
- Michael W. Kraus, Paul K. Piff, and Dacher Keltner. Social class as culture: The convergence of resources and rank in the social realm. *Current Directions in Psychological Science*, 20(4): 246–250, 2011.
- Ming-Jen Lin. Does unemployment increase crime? Evidence from US data 1974-2000. *Journal of Human Resources*, 43(2):413–436, 2008.
- Lance Lochner. Education, work, and crime: Theory and evidence. *Rochester Center for Economic Research, Working Paper No. 465*, 1999.
- Lance Lochner. Education, work, and crime: A human capital approach. *International Economic Review*, 45(3):811–843, 2004.

- Stephen Machin and Costas Meghir. Crime and economic incentives. *Journal of Human Resources*, 39(4):958–979, 2000.
- Anandi Mani, Sendhil Mullainathan, Eldar Shafir, and Jiaying Zhao. Poverty impedes cognitive function. *Science*, 341(6149):976–980, 2013.
- Paul K. Piff, Danie M. Stancato, Stéphane Côté, Rodolfo Mendoza-Denton, and Dacher Keltner. Higher social class predicts increased unethical behavior. *Proceedings of the National Academy of Sciences*, 109(11):4086–4091, 2012.
- Steven Raphael and Rudolf Winter-Ebmer. Identifying the effect of unemployment on crime. *The Journal of Law and Economics*, 44(1):259–283, 2002.
- Gary Sweeten, Alex R. Piquero, and Laurence Steinberg. Age and the explanation of crime, revisited. *American Economic Review*, 106(2):260–284, 2013.
- Long Wang and Keith J. Murnighan. Money, emotions, and ethics across individuals and countries. *Journal of Business Ethics*, 125(1):163–176, 2014.
- Jenny Williams and Robin C. Sickles. On the role of social capital in youth crime: A dynamic structural approach. *Adelaide, SA: University of Adelaide*, 1998.

A Appendix

Duplicates in terms of rc

copies	observations	surplus	copies	observations	surplus
1	612207	0	62	682	671
2	590730	295365	63	252	248
3	495186	330124	64	192	189
4	399404	299553	65	325	320
5	325090	260072	66	528	520
6	262182	218485	67	268	264
7	213325	182850	68	68	67
8	174248	152467	69	276	272
9	144036	128032	70	140	138
10	118140	106326	71	213	210
11	98648	89680	72	144	142
12	82992	76076	73	292	288
13	69095	63780	74	222	219
14	56714	52663	75	300	296
15	49155	45878	76	228	225
16	41376	38790	77	154	152
17	35139	33072	78	156	154
18	28530	26945	79	158	156
19	24928	23616	80	240	237
20	22480	21356	81	81	80
21	19131	18220	82	164	162
22	16214	15477	83	83	82
23	13478	12892	84	168	166
24	12768	12236	86	344	340
25	9975	9576	89	267	264
26	9230	8875	90	180	178
27	8370	8060	93	93	92
28	6972	6723	95	95	94
29	5394	5208	96	96	95
30	5940	5742	98	196	194
31	4650	4500	99	297	294
32	3968	3844	105	210	208
33	3927	3808	106	318	315
34	3672	3564	107	107	106
35	3535	3434	108	108	107
36	2196	2135	109	109	108
37	2479	2412	110	110	109
38	2394	2331	112	112	111
39	2496	2432	114	228	226
40	2040	1989	117	117	116
41	1804	1760	121	121	120
42	1638	1599	125	125	124
43	1763	1722	128	128	127
44	1848	1806	129	129	128
45	1620	1584	130	130	129
46	1426	1395	134	268	266
47	940	920	137	137	136
48	960	940	139	139	138
49	980	960	145	145	144
50	900	882	147	147	146
51	663	650	153	153	152
52	468	459	159	159	158
53	742	728	167	167	166
54	432	424	172	172	171
55	770	756	185	185	184
56	840	825	186	186	185
57	456	448	208	208	207
58	464	456	211	211	210
59	472	464	260	260	259
60	300	295	312	312	311
61	366	360			

Figure A.1: Number of duplicates grouped by personal identification number

Broad Structure of NACE Rev. 2

Section	Title	Divisions
A	Agriculture, forestry and fishing	01 – 03
B	Mining and quarrying	05 – 09
C	Manufacturing	10 – 33
D	Electricity, gas, steam and air conditioning supply	35
E	Water supply; sewerage, waste management and remediation activities	36 – 39
F	Construction	41 – 43
G	Wholesale and retail trade; repair of motor vehicles and motorcycles	45 – 47
H	Transportation and storage	49 – 53
I	Accommodation and food service activities	55 – 56
J	Information and communication	58 – 63
K	Financial and insurance activities	64 – 66
L	Real estate activities	68
M	Professional, scientific and technical activities	69 – 75
N	Administrative and support service activities	77 – 82
O	Public administration and defence; compulsory social security	84
P	Education	85
Q	Human health and social work activities	86 – 88
R	Arts, entertainment and recreation	90 – 93
S	Other service activities	94 – 96
T	Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use	97 – 98
U	Activities of extraterritorial organisations and bodies	99

Figure A.2: NACE classification of occupations

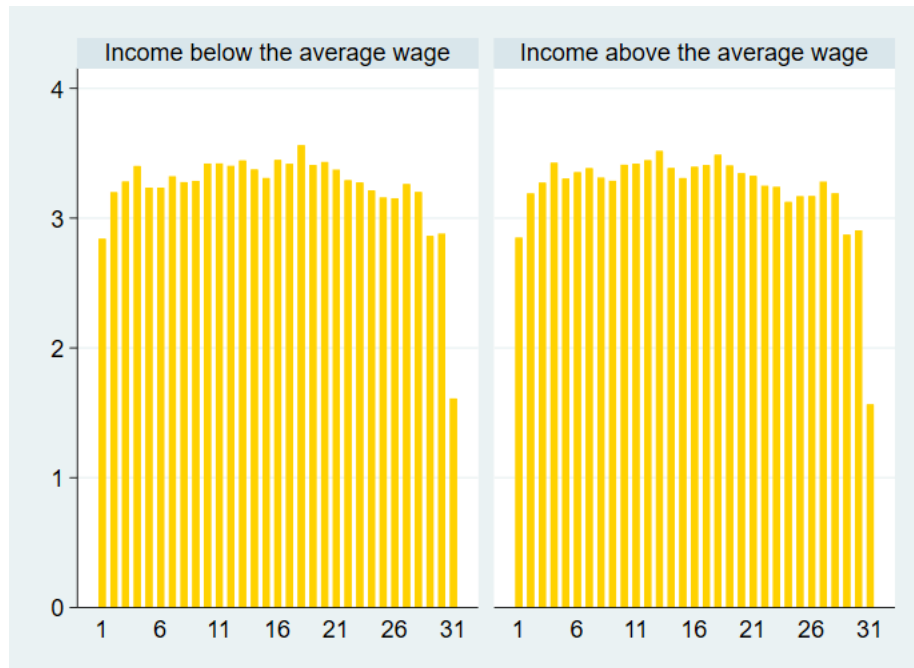


Figure A.3: Number of violations by day in a month (NACE subset)

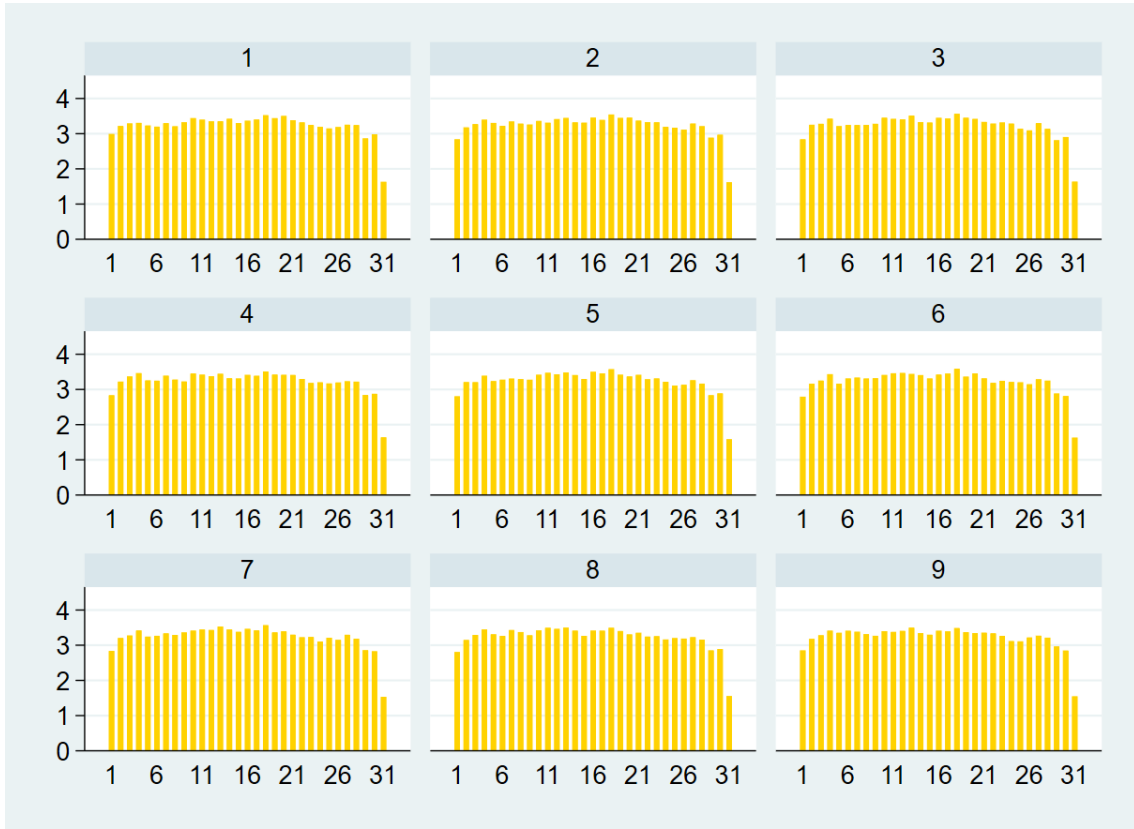


Figure A.4: Violations by day in a month by income pctlles (NACE subset)

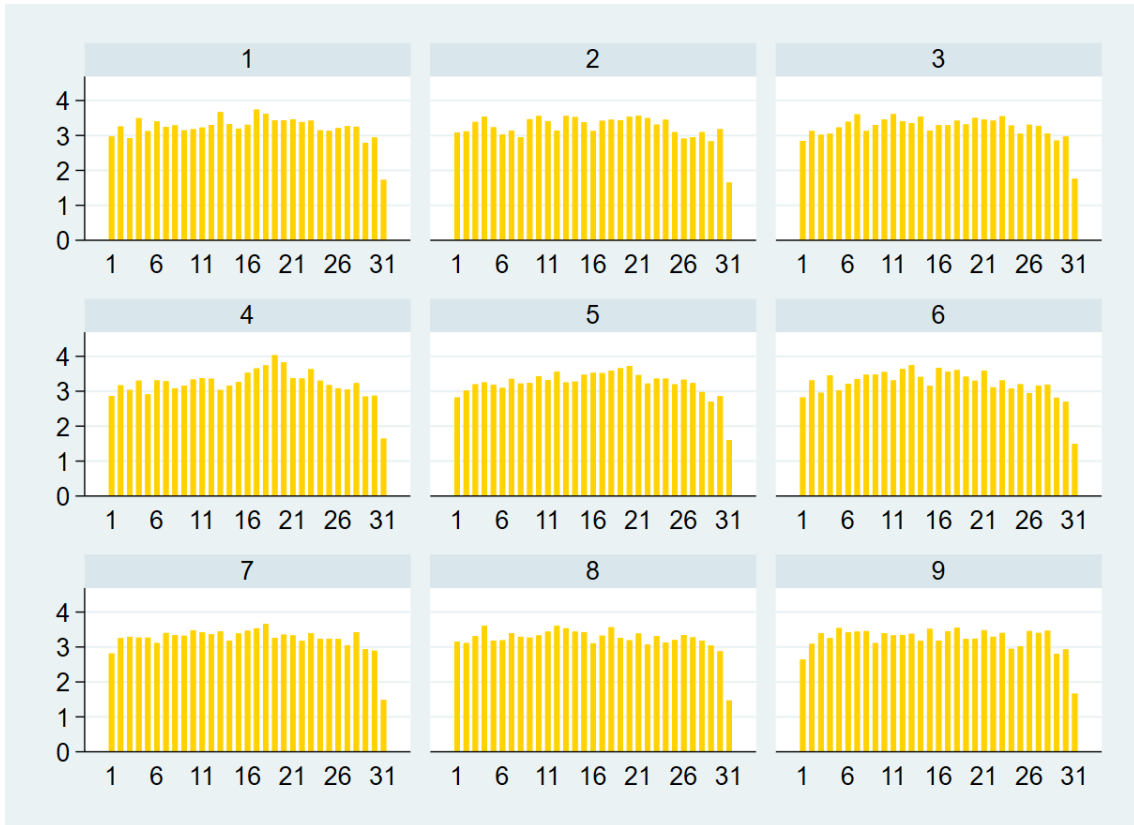


Figure A.5: Violations by day in a month by income pctlles for *Public administration* (NACE O)

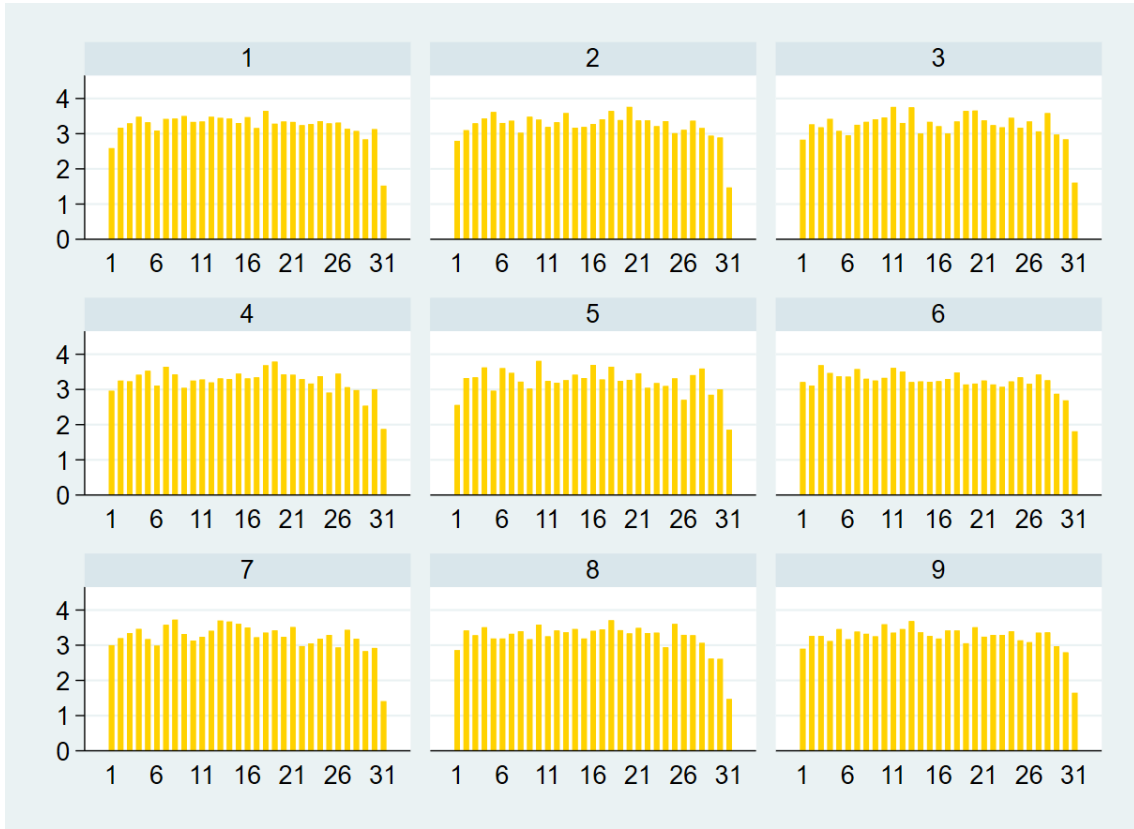


Figure A.6: Violations by day in a month by income pctlles for *Education* (NACE P)

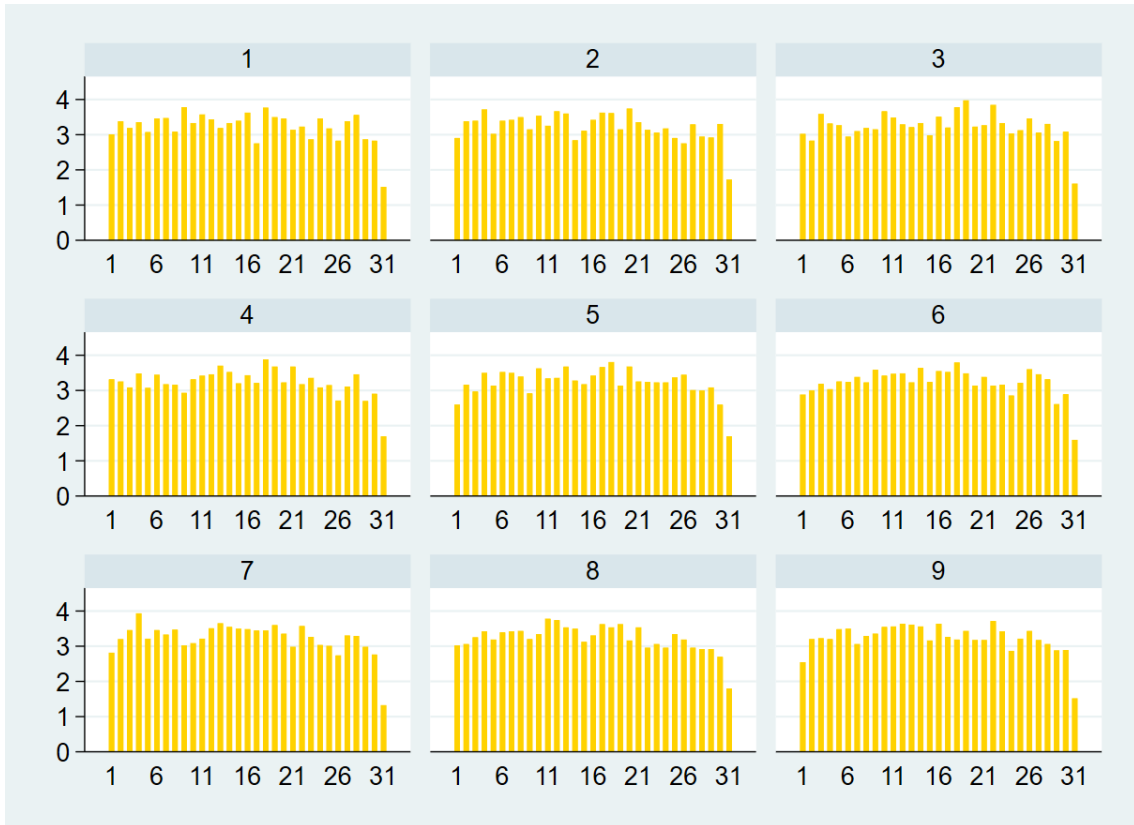


Figure A.7: Violations by day in a month by income pctlles for *Human health* (NACE Q)