Missing Rich Offenders: Traffic Accidents and the Impartiality of Justice

Madina Kurmangaliyeva * [a,b]

[a]European University Institute
[b]Institute for the Rule of Law at the European University at Saint-Petersburg

December 6, 2017

(download the latest version at https://bit.ly/madina_jmp)

Abstract

This paper estimates the effect that wealth and power have on criminal justice outcomes by exploiting the random matching of drivers to pedestrians in traffic accidents. If justice is impartial, we should observe the same share of rich offenders both for poor and rich victims, conditional on location and time. Rich victims act as a control group to estimate the proportion of *missing* rich offenders whose victims are less powerful. I use this estimation approach on data from Russia, and find that its justice system is not impartial. The same approach can be applied not only to other countries but also to other characteristics that should be irrelevant to judicial outcomes in an impartial legal system, such as race and gender.

*I would like to thank the anonymous reviewers for their helpful comments. I am grateful to Andrea Ichino and Andrea Mattozzi for their supervision, and my peers at the European University Institute for the constructive feedback. This project would not have been possible without Vadim Volkov, Dmitriy Skougarevskiy, Irina Chetverikova and other members of the Institute for the Rule of Law at the European University at Saint-Petersburg who provided invaluable advice and data access. Data used in this study was created with support of the Russian Science foundation grant 17-18-01618. This paper has benefited from the comments by reviewers and conference participants at the Annual Conference on Empirical Legal Studies at Duke Law School, November 2016; seminar participants at Nazarbayev University, August 2015. Special thanks to Jérôme Adda, Gözde Çörekçioğlu, Ada González-Torres, Anastasia Antsygina, Sylta Cornils, and Matteo Sostero.

1 INTRODUCTION

An impartial legal system is an indicator of good institutions. In particular, the legal system should be impartial to the influence of personal power, whatever its provenance: wealth, political authority, or social connections.[1] Measuring impartiality empirically, however, is a challenging task. There are two main reasons for this. First, powerful offenders may evade justice early in the legal process, and thus may not appear in any records. Second, their crimes may be incomparable in nature to the crimes committed by the less powerful. In this paper, I propose a simple methodology to address these challenges. I test this methodology by applying it to Russian data on criminal traffic offenses.

My identification strategy relies on the random matching of drivers with pedestrians in traffic accidents in a given location and time. Since drivers typically do not choose whom to hit, the share of rich – and hence powerful – offenders should be the same across all types of victims, rich or poor. It should also remain the same among the cases chosen for prosecution, if the justice system is impartial. Under impartial justice, different offenders are more or less likely to be prosecuted – for example because of different degrees of culpability – but the identity of the victim must be irrelevant for the decision to prosecute, resulting in identical profiles of prosecuted offenders across all victims.[2]

Specifically, if rich offenders are *de facto* more likely to evade justice, rich victims should also be more successful in resisting such evasion (Hagan, 1982). In this case, I expect the share of rich offenders among the prosecuted cases to be lower for poor victims than for rich ones. Rich victims can thus act as a control group to quantify the extent of *missing* rich offenders for poorer victims. In other words, I estimate how many cases involving rich offenders and poor victims need to be added to the sample so that rich offenders are balanced across all types of victims.[3]

[1] See Glaeser et al. (2003) for a discussion on how a legal system subverted by the powerful can lead to weaker property rights protection and the propagation of economic inequality. For a discussion on the link between the quality of institutions and economic development, see for example Acemoglu and Robinson (2013), Nunn and Puga (2012), Dell (2010), Coatsworth (2008), and Glaeser et al. (2004).

[2] The identification strategy is similar in spirit to Levitt and Porter (2001) who exploit the random matching of drivers in two-car crashes to estimate the risks posed by drunk drivers. If sober and drunk drivers are equally likely to result in a fatal car crash, then they should be "equally mixed" in the sample of fatal crashes. The distributional asymmetries allow them to recover how deadly drunk drivers are.

[3] The identification strategy is related to the literature on "missing" girls in China (Shi and Kennedy, 2016; Johansson and Nygren, 1991, e.g.), which uses the observed distortions in sex ratios among the newborn children to reveal the number of girls that have been either aborted, killed in infancy, or underreported as a result of the one-child policy and the social preference for boys.

I apply the method on police data from Russia, which ranks low in the effectiveness and the impartiality of its criminal justice system, according to The World Justice Project (2014).[4] In Russia, a traffic offense is criminal if it has resulted in a grave (usually life-threatening) injury. The data represents all prosecuted cases for criminal traffic offenses for 2013–2014. I restrict the sample to offenders of working age who have hit one working-age pedestrian, in total $6,600$ cases from $1,200$ police departments with the median of 2 observations per police department.

I proxy the level of resources that give power by employment status. I classify individuals in three groups. Low-resource individuals are those with no permanent employment; high-resource individuals are entrepreneurs, white-collar executives, and government officials. All remaining individuals are in the middle group. As an alternative proxy for resources, I also use the offenders' education level and impute the prices for their cars based on their resale value. I observe the hour and date of the accident, which allows to account for intra-day variation in the population of offenders and victims. I proxy the location of an accident by the police department identifier, since police departments only investigate offenses that happened on their territory.

I find two key results that raise doubts about the impartiality of justice in Russia. First, rich offenders are prosecuted less often. Comparing the profiles of offenders when victims are poor relative to when victims are rich, there are about 50% fewer top-group offenders and 40% fewer middle-group ones than there should be, after accounting for police department and hour fixed effects. I infer that these missing rich offenders have avoided prosecution at some earlier stage. Because of the inherent gravity of the accidents in my sample, I can rule out that offenders compensate the victims on the spot to avoid police investigation. Actually, the imbalance is even more evident for the cases involving the victim's death. The results are robust to using alternative proxies for resources, or to different regional breakdowns.

Second, I find that rich offenders are less likely to be punished in court. Among the prosecuted cases, I see that rich offenders are incarcerated less often when their victims are less powerful. This imbalance might be ascribed to settlements between parties, as allowed in the Russian justice system for unintentional crimes (albeit with the approval of a judge). Indeed, when I estimate the combined share of offenders who have either settled or been imprisoned, I find that this share is the same across all victims. Taken at face value, this would misleadingly suggest that the court system is impartial. However, since rich offenders are already less likely to be prosecuted in the first place, at the time of the accident the probability of punishment actually

[4] See Paneyakh (2014) and Sklyaruk and Skougarevsky (2015) for the discussion on how Russian prosecutors, facing wrong incentives to maximize their conviction statistics, may be more willing to prosecute low-resource offenders exploiting their low level of resistance.

does vary by type of victim: the middle-group offenders settle or are imprisoned in 25% of cases when the victim is low-group versus 40-50% for middle- or top-group victims. Top-group offenders settle or are imprisoned in 8% of cases when the victim is less resourceful versus roughly around 50% when the victim is also in the top group.

I acknowledge that the geographic area supervised by police departments may not be fine-grained enough to eliminate all the spatial correlation. To check the quality of the controls, I run placebo tests using other spatially correlated characteristics of offenders and victims (e.g., foreign citizenship) which show that police-level fixed effects successfully eliminate spatial correlation for them. Also, results at large are robust to elimination of the biggest – by the number of observations – police departments from the sample. I also argue that the stronger evidence of *missing* rich offenders for more severe cases cannot be explained by an omitted spatial segregation and is suggestive of other channels. While all of the above does not rule out the residual presence of spatial correlation, it increases my confidence in attributing the imbalanced distribution of offenders to the lack of impartiality in the justice system.

Methodologically this paper contributes to available empirical tests of unjustified biases in police and court decisions (e.g., Alesina and La Ferrara, 2014; Knowles et al., 2001; Shayo and Zussman, 2011).[5] Glaeser and Sacerdote (2003) use traffic offenses for their exogenous pairing of offenders with victims to test for gender and racial biases in criminal sentencing, but they do not address the non-random selection of cases into court. My approach allows to estimate the disparities at earlier (unobserved) stages of criminal justice, which is an important challenge in the literature on sentencing disparities (Zatz and Hagan, 1985; Ulmer, 2012).[6] Since the distribution of offenders should be independent from the distribution of victims also in terms of attributes other than wealth, one can apply the same identification method to study racial, gender, and ethnic biases at different stages of criminal justice.[7]

This study is closely related to the literature on justice disparities for different socio-economic status of offenders and in particular to the study by Volkov (2016).[8]

[5]Alesina and La Ferrara (2014) and Shayo and Zussman (2011) also exploit the identity of the victim in finding judicial bias. Alesina and La Ferrara (2014) test racial bias by comparing the rates with which appellate courts reverse capital sentences for white victims versus minority victims in the United States. Shayo and Zussman, 2011 exploit a random assignment of judges to study in-group bias through the variation in judges' and plaintiffs' ethnicities.

[6]In general, empirical papers analyzing court data usually develop a theoretic model, specific to the setting, and use it to predict the characteristics of cases that reach the sentencing stage and, thus, the direction of the bias (see for example, Ichino et al., 2003).

[7]For example, for Russia I find that females tend to be prosecuted slightly more often when their victims are females as well, revealing a pro-female bias in prosecution. However, the difference in prosecution rates does not translate into ex-ante disparities in incarceration or settlement rates.

[8]See D'Alessio and Stolzenberg, 1993, for a general overview of the literature. Other studies

Basing his study on court data for violent crimes, theft, drugs, and fraud, Volkov (2016) finds that judges in Russia tend to incarcerate the college-educated less often and the unemployed more often. At the same time, the author observes that judges incarcerate entrepreneurs and top managers more often, which he attributes to the judges' bias against people in "the position of trust and authority". I find the opposite and I believe that the difference between our results is driven by *missing* offenders and a selection bias. In particular, crimes committed by entrepreneurs and top managers may be qualitatively different from the same-category crimes committed by lower-status offenders, which is a general concern in the literature. Rather than comparing judicial outcomes for rich to poor offenders directly, traffic offenses allow us to measure the disparity throughout the justice system via the random variation of victims' resources.

This paper is also related to the normative studies of whether wealthy offenders should be allowed to "buy" justice, which usually omit the role of the victim, assuming that the interaction is between the offender and the prosecutor (e.g., Garoupa and Gravelle, 2003).[9] My results show that the interaction can be between the offender and the victim, which may have policy-relevant implications.

The rest of the paper is organized as follows. Section 2 presents the identification strategy. Section 3 provides institutional framework and summarizes the data. Section 4 presents results for prosecution and for incarceration and settlement rates, and discusses possible mechanisms behind missing rich offenders. The last section concludes.

2 Identification strategy

This section provides definitions and describes the identification strategy used in the paper. I first present the simplest case, in which resources are binary, and then I explain how it adapts to more resource levels. The general case with continuous resources is presented in Appendix A.

2.1 Binary resources

In a given location and time g, an offending driver (O) hits a random pedestrian, called the victim (V). Offenders and victims are endowed with resources that give power, denoted as r_o and r_v respectively. The level of resources can be low (L), or high (H): $r_o, r_v \in \{L, H\}$. In location g, the probability that the offender is

that investigate the effect of resources on justice include the literature on corporate advantage (e.g., Yoav, 1999; Bacher et al., 2005) and on the effect of private versus public defense counseling (e.g., Champion, 1989; Rattner et al., 2008; Hartley et al., 2010).

[9] See also Lott Jr (1987), Arlen (1992), Kobayashi and Lott (1996), and Clark (1997)

rich is denoted as $\Pr(r_o = H)$, while $\Pr(r_v = H)$ denotes the probability that the victim is rich. There are four possible combinations of offender-victim resources, $r_o \times r_v \in \{LL, LH, HL, HH\}$. Since drivers are randomly matched to victims, the distributions of their resources are independent. Hence, the joint probability function is the multiplication of the marginal distributions:

$$\Pr(r_o, r_v) = \Pr(r_o)\Pr(r_v) \quad \text{for } r_o, r_v \in \{L, H\} \tag{1}$$

Because of independence, the share of H offenders is the same for H or L victims in the population of accidents and equal to the unconditional share of the H-type among offenders:[10]

$$\Pr(r_o = H \mid r_v = L) = \Pr(r_o = H \mid r_v = H) = \Pr(r_o = H) \tag{2}$$

After the accident, the legal system decides whether to prosecute (or, more generally, punish) the offender, $P = 1$, or not, $P = 0$.[11] Only prosecuted cases are observed. The probability of prosecution, $\Pr(P = 1 \mid r_o, r_v)$, is denoted as $\pi(r_o, r_v)$, which is a function of the victim's and offender's resources. I denote the relative prosecution rates for H versus L offenders for a given r_v as:

$$\rho_L^H(r_v) \equiv \frac{\pi(r_o = H, r_v)}{\pi(r_o = L, r_v)} \tag{3}$$

For example, $\rho_L^H(r_v = L) = 2/3$ means that for every three poor offenders the system prosecutes two rich offenders, given the victim is poor. The value of $\rho_L^H(r_v = L)$ does not have a normative interpretation, because H offenders may differ from L offenders in their culpability or the propensity to run away. However, there is no reason why the relative prosecution rate should differ across *victims* in an impartial legal system.

Definition 1. *A justice system is* ***impartial*** *if the relative prosecution rate does not depend on the resources of the victim:*

$$\rho_L^H(r_v = L) = \rho_L^H(r_v = H)$$

Let $\delta^H(r_v)$ denote the expected share of H offenders among the prosecuted cases conditional on the victim's resource level. I can express it as a function of $\rho_L^H(r_v)$ and $\Pr(r_o = H)$:[12]

$$\delta^H(r_v) \equiv Pr(r_o = H | r_v, P = 1) = \frac{\rho_L^H(r_v)}{\rho_L^H(r_v) + \frac{1 - \Pr(r_o = H)}{\Pr(r_o = H)}} \tag{4}$$

[10]Proof: $\Pr(r_o = H \mid r_v) = \frac{\Pr(r_o = H, r_v)}{\Pr(r_o = H, r_v) + \Pr(r_o = L, r_v)} = \frac{\Pr(r_o = H)\Pr(r_v)}{\Pr(r_o = H)\Pr(r_v) + \Pr(r_o = L)\Pr(r_v)} = \Pr(r_o = H)$

[11]Other types of punishment may include the decision to report, indict, incarcerate, etc.

[12]Proof: Use Bayes' rule and Equations (1), (2), and (3)

Then, the odds ratio of observing an H offender given the victims are L-type as opposed to H-type captures the ratio of the relative prosecution rates:

$$\frac{\delta^H(r_v = L)/\big(1-\delta^H(r_v = L)\big)}{\delta^H(r_v = H)/(1-\delta^H(r_v = H))} = \frac{\rho_L^H(r_v = L)}{\rho_L^H(r_v = H)} \tag{5}$$

Proposition 1. *The justice system is impartial if and only if the odds ratio of observing a rich offender between poor and rich victims is equal to one:*

$$\rho_L^H(r_v = L) = \rho_L^H(r_v = H) \iff \frac{\delta^H(r_v = L)/\big(1-\delta^H(r_v = L)\big)}{\delta^H(r_v = H)/(1-\delta^H(r_v = H))} = 1 \tag{6}$$

which is equivalent to observing the same share of rich offenders for poor and rich victims:[13]

$$\rho_L^H(r_v = L) = \rho_L^H(r_v = H) \iff \delta^H(r_v = L) = \delta^H(r_v = H) \tag{7}$$

Otherwise, if justice is not impartial, I expect H offenders to be prosecuted less frequently when their victims are L, rather than H, i.e. $\rho_L^H(r_v = L) < \rho_L^H(r_v = H)$. Hence, I expect lower odds of observing a rich offender for the less powerful victims:

$$\begin{aligned}\rho_L^H(r_v = L) < \rho_L^H(r_v = H) &\iff \delta^H(r_v = L) < \delta^H(r_v = H)\\ &\iff \frac{\delta^H(r_v = L)/\big(1-\delta^H(r_v = L)\big)}{\delta^H(r_v = H)/(1-\delta^H(r_v = H))} < 1\end{aligned}$$

2.2 Empirical approach when resources are binary

Using the prosecuted traffic offenses in location g, I can fit the following population regression function:

$$\mathbb{1}\{r_o = H\}_c = \alpha + \beta \mathbb{1}\{r_v = H\}_c + u_c \tag{8}$$

where $\mathbb{1}\{r_o = H\}_c$ and $\mathbb{1}\{r_v = H\}_c$ are the indicator functions for the rich offender and the rich victim, respectively; c is the case identifier, and u is the error term. The regression parameters α and β help capturing the following parameters of interest:

$$\alpha = \delta^H(r_v = L) \tag{9}$$

[13] *Proof.* Condition (6) follows from Equation (5). Condition (7) is the result of

$$\frac{\delta^H(r_v = L)/\big(1-\delta^H(r_v = L)\big)}{\delta^H(r_v = H)/(1-\delta^H(r_v = H))} = 1 \iff \delta^H(r_v = L) = \delta^H(r_v = H)$$

$$\beta = \delta^H(r_v = H) - \delta^H(r_v = L) \tag{10}$$

$$\frac{\alpha/(1-\alpha)}{(\alpha+\beta)/(1-\alpha-\beta)} = \frac{\rho_L^H(r_v = L)}{\rho_L^H(r_v = H)} \tag{11}$$

Using Proposition 1, I can test the impartiality by testing the following hypothesis:

$$H0 : \beta = 0 \tag{12}$$

If $\beta = 0$, then the odds ratio $\frac{\alpha/(1-\alpha)}{(\alpha+\beta)/(1-\alpha-\beta)} = 1$.

Because of Equation (11), the odds ratio has a direct interpretation. The lower its value, the greater is the share of missing H offenders. For example, $\frac{\rho_L^H(r_v=L)}{\rho_L^H(r_v=H)} = 0.75$ means that only three out of four cases involving rich offenders and poor victims have been prosecuted relative to the number observed for rich victims – i.e., one quarter of cases involving rich offenders and poor victims is missing.

The identification strategy is robust to allowing risky behavior of pedestrians and drivers to correlate with the level of their resources. Risk-taking behavior changes the marginal distribution of types for offenders and victims, but the two distributions remain independent, as long as risky behaviour of drivers does not depend on which pedestrians are around, and vice versa. Potential problems arise if the system prosecutes only those cases where the offender was relatively more culpable than the victim, leading to a violation of the independence assumption. I test this hypothetical prosecution rule for Russia in Appendix E.1 and do not find reasons for such concerns.

2.3 Three or more levels of resources

When there are more than two levels of resources, the procedure can always be reduced to the binary case, by picking just two levels of r_o and two levels of r_v at a time and finding the corresponding odds ratio. Alternatively, several adjacent levels can be combined into one.

I denote the observed share of i-type offenders conditional on the victim's resource level in the sample restricted for i-type and j-type offenders only as:

$$\delta_j^i(r_v) \equiv Pr(r_o = i | r_o \in \{i, j\}, r_v; P = 1)$$

Then, equation (5) can be restated as:

$$\frac{\delta_j^i(r_v = k)/\left(1 - \delta_j^i(r_v = k)\right)}{\delta_j^i(r_v = l)/\left(1 - \delta_j^i(r_v = l)\right)} = \frac{\rho_j^i(r_v = k)}{\rho_j^i(r_v = l)} \tag{13}$$

For example, when there are three levels of resources – L for low, M for middle, and H for high, $L < M < H$ – it gives nine possible combinations of offender-victim resources and nine odds ratios to estimate: $\frac{\rho_j^i(r_v=k)}{\rho_j^i(r_v=l)}$, such that $i,j,k,l \in \{L,M,H\}$, $i > j$, $k < l$. When $i = l$ and $j = k$, I call such odds ratios symmetric, otherwise, the odds ratios are asymmetric. For example, an asymmetric odds ratio that captures $\frac{\rho_L^H(r_v=L)}{\rho_L^H(r_v=M)}$ tells us the share of non-missing H offenders for L victims using M victims (and L offenders) as the control group.

3 Institutional setup and data

This section introduces the institutional setup of Russia and summarizes the information about the data.

3.1 Criminal traffic offenses and punishment

Russia has a civil law legal system. The Criminal Code of Russia classifies bodily injuries into "light", "average", and "severe", where severe injuries are usually life-threatening or leading to disability. A traffic offense is considered to be *criminal* if the driver has caused someone else a "severe" injury.

The Code distinguishes criminal traffic offenses based on the number of fatalities it has caused: for the cases involving only one pedestrian, there are either no death or one death. The Code further differentiates between sober and intoxicated offenders, giving a total of four different offense groups, each of which is potentially punishable by incarceration. The maximum prison sentence varies with the offense type: it starts at two years for *no death + sober* and rises up to seven years for *one death + intoxicated*. The *no death + sober* offense type also allows for milder forms of punishments including the "limitation of freedom", which imposes some restrictions on movement at night, on leaving the municipality and requires regular check-ups with the authorities, among other things.

The criminal court usually revokes the driver's license and decides on the amount of compensation of damages for pain and suffering the defendant has to pay to the victim. The compensation of the medical expenses and property damage almost always involves the insurance company of the offender in a separate civil case, which depends on the results of the criminal.

3.2 Legal process

The timeline of events in the legal process for a criminal offense can be split into three parts: (1) the police investigation and the decision of the prosecutor; (2)

decision to settle, and (3) the court trial.

When a traffic accident happens, the local road police arrives at the scene. If someone is severely injured, an investigator from the local police department joins the investigation. She registers the case as a criminal offense after collecting the medical reports about severe injuries. At this stage, the information about the circumstances of the criminal case and the information on the victim enters the police records. Based on the evidence collected by the investigator, the prosecutor decides whether to prosecute the suspected offender, if there is one. If the suspect is prosecuted, the information about him enters the police records.

After the decision to prosecute, defendants may settle with the victims. A settlement between the offender and the victim involves the following steps: (1) the defendant compensates the the victim or the close family members for the damages for pain and suffering; (2) the victim forgives the offender and officially asks the criminal charges to be dropped; (3) subject to the approval of the judge (or earlier with the permission of the prosecutor) the offender (a) gets no criminal conviction, since his guilt is not ruled by court, (b) keeps his driver's license, and (c) may settle again in future even for the same offense. The information about the settlement, nevertheless, enters the police database.

If no settlement is reached, the case is forwarded to court and the judge decides whether the defendant is guilty or not. A prominent feature of the Russian criminal system is that acquittals in court are very rare, less than 1% out of total traffic offenses in court. Thus, the court is *de facto* the sentencing stage of the criminal justice (Volkov, 2016; Shklyaruk, 2014; Trochev, 2014). In practice, incarcerations are rare for *no death + sober* offenses, but quite common for the rest. Table 1 summarizes the statistics of outcomes for all criminal traffic offenses tried in Russian courts in 2009–2013 by the offense groups. The table shows that the probability of settlements (imprisonment) drop (increase) with the severity of the offense.

3.3 Data

The data comes from the Russian centralized police-database for 2013–2014.[14] Each investigation is recorded by different personnel in police and prosecutor offices through different statistical forms.

The first set of forms represents the entirety of the criminal cases that have been registered by the police. The data includes a short description of each case provided by the investigators for their own easy reference. Hence, the style and the amount of detail contained in the description vary across police departments. I identify the

[14] The access is provided by the Institute for the Rule of Law at the European University at Saint Petersburg. The Institute has cleaned and transformed the raw administrative data into a Stata database with the support of Russian Science Foundation grant 17-18-01618.

Table 1: The classification of traffic offenses

The Criminal Code classification				Summary statistics*			
#	Fatalities	Offender's state	Max prison	Settled	Probation	Incarceration	
			(yrs)	(%)	(%)	(%)	(avg. yrs)
1	No death	Sober	2	43	30	3	1.4
2	No death	Intoxicated	3	22	50	26	1.9
3	One death	Sober	3	23	44	31	2.3
4	One death	Intoxicated	7	5	23	70	3.1

*Source: Official database of court data; averaged over 2009–2013;

cases that involve pedestrians by using an automated regular expressions parser I developed for this project. The parser captures the patterns of texts associated with pedestrians in the description: either a direct use of the word 'pedestrian' with its derivatives or patterns like 'hit [the name of the victim] who was crossing the street'. The search has identified $21,300$ such cases out of more than $70,000$ criminal traffic offenses, 96% of which involve a single victim.

The information on the hour of the accident is available in 80% of cases and the presence of this information is not correlated with any particular combination of offender-victim resources (see Table 15 in Appendix E.2). I interact the hour of the accident with the day of the week, differentiating between the weekdays, Saturdays, and Sundays. Since accidents are rarer during weekends and nights, I recode some hours into bigger groups. For weekdays, I group the accidents after midnight and before 5 am. For weekends, I divide the hours into four groups: 1 am to 7 am, 8 am to noon, 1 pm to 5 pm, and 6 pm to midnight. For example, all accidents that happened at 2 pm on Monday are in the same group as other 2 pm accidents from other weekdays, but not together with 2 pm accidents from Saturday or Sunday.

Among the pedestrian victims, 38% died in the accident, 50% are females, 9% are minors, 22% are retired, 40% have no permanent employment, 20% are blue-collar workers, 2% are white-collar workers, less than 1% each are government officials and businessmen. The police has failed to find the offender in 25% of cases, which includes the official reports about the failure and cases with the missing information on the suspect. In 6% of cases the initial charges have been dropped. The remaining 69% have been prosecuted.

The second type of forms provide information on defendants. I have merged these forms with the previous using the case identifiers, the police department number and the year of the registration of the case. For cases involving one pedestrian victim, it amounts to $14,100$ prosecuted offenders, who are 90% males, 5% retirees, 34% with no permanent employment, 45% blue-collar workers, 5% white-collar workers (out of which 1% are executives), 3% businessmen, and 1% government officials. About 16% of offenders have a criminal record. Among offenders, 16% settled, 12% were punished by the limitation of freedom, 10% received a suspended prison term, and 12% were incarcerated. In another 21% of cases, the offender was found guilty, but not punished thanks to the amnesty for *no deaths + sober* offenses in one of the years. In 25% of cases, the records state that the case is before the court but does not specify the outcome.

As the primary proxy for resources, I use the employment status (or simply the *status*), denoted as s_o for offenders and s_v for victims, where $s_o, s_v \in \{L, M, H\}$. I exclude cases that involve children, students, and retirees. The L type, which includes individuals that do not have permanent employment, represents 56% of victims and 37% of offenders. The H type, which includes white-collar top-managers, entrepreneurs, and government officials, including law enforcement officials, accounts for 2% of victims and 7% of offenders. The officials are considered to have greater knowledge about the judicial system and be better connected to the decision-makers on their case, even if they are not wealthier than the white-collar workers. The M type, mostly blue-collar workers, constitute 42% of victims and 57% of offenders.

In total, there are $8,100$ cases, but only $6,600$ cases has the information on the hour of the accident. Table 2 shows the distribution of cases across offender-victim status groups. The H type is the least populous, with just 16 observations for H-type offenders and H-type victims. The mass of observations is for L and M types.

Table 2: Number of observations in each group

		Victim		
		L	M	H
Offender	L	1,748	695	49
	M	1,829	1,821	83
	H	192	176	16

Additional proxies for offenders' resources are based on their education and the imputed car prices. I classify offenders into three groups based on educational achievement, $e_o \in \{sch, voc, col\}$. The lowest level includes school graduates with no further education (10% of offenders). The middle level includes vocational training or other types of degrees that do not constitute a college degree (70% of offenders). College graduates, the highest level, represent 20% of offenders.

Information on cars driven by offenders comes from the case description. Usually, the text mentions the car driven by the offender before any other participant. For those first-mentioned cars I impute the estimated value using the online advertisements of secondary car sales on *auto.ru* retrieved in October 2014. Car prices are available for $4,400$ offenders. I impute the prices based on the mean observed for a certain car brand and model in the advertisements. Additionally, there are 700 truck, bus, and motorbike drivers, for whom the prices are not available. I allocate offenders into three groups based on the cars they drive, denoted $c_o \in \{inf, nor, lux\}$. The lowest, *inferior*, group includes trucks, buses, motorbikes, and cars with the imputed price below $250,000$ RUB (about $4,500$ USD at the time of writing), which represents 59% of offenders. The middle group includes *normal* cars: those with the imputed price above $250,000$ but below $500,000$ RUB (25% of offenders). The top group includes all other cars, which I refer to as *luxury* cars (16% of offenders).

Table 3 summarizes the statistics for the $8,100$ cases by the employment status of victims and offenders. It shows averages of the variables for each status and the two normalized differences for M versus L, and H versus M. The normalized difference is calculated as the difference in averages divided by the square root of the average of the standard deviations.[15] This measure is more relevant in assessing imbalances across groups than the t-statistic, and the normalized difference greater than 0.25 indicates an imbalance (Imbens and Wooldridge, 2009).

I observe more deaths among L victims than for higher-status victims. The higher-status offenders are more likely to be college graduates and tend to drive more expensive cars, although the mean difference between L and M offenders is not as pronounced as that between M and H. L offenders are also on average younger than the rest and tend to come from areas with a lower load of traffic offenses per police department, possibly linked to rural areas.

The statuses of victims and offenders are correlated. H victims are more likely to be hit by expensive cars and college-educated offenders. Nevertheless, offenders and victims across all status levels do not differ in the general characteristics of the accidents: time, day, month, and location within or outside the city (as captured by regular expressions from the case descriptions).

Figure 1 shows the density of police departments by number of observations. Among $6,600$ cases that have full information, the average number of observations per police department is 5.45 and the median is 2. There are six departments with more than 100 observations each, with a single one having 269 cases. It could be that some departments oversee a territory too large to be used as the location control. In Section E.5, I show that the main results are mostly robust to the gradual removal of the largest police departments from the sample. Figure 2 in Appendix

[15] $nd = (\bar{X}_j - \bar{X}_k) / \sqrt{\frac{S_j^2 + S_k^2}{2}}$ where $\bar{X}_i$ is the mean and S_i is the standard deviation of a variable X for group i.

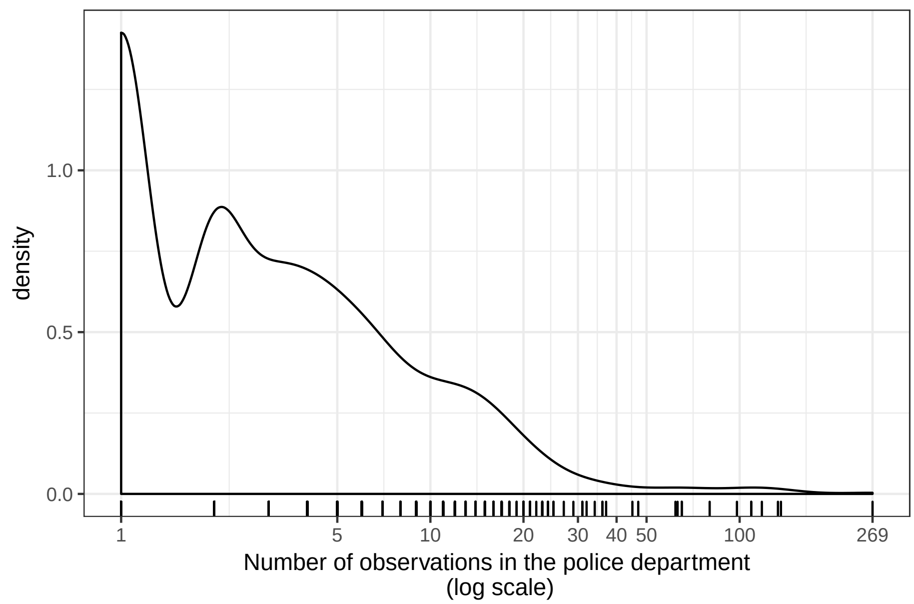
Table 3: Descriptive statistics for 8,100 prosecuted cases by the employment status of victims or offenders

	By Victim type:					By Offender type:				
	L	*M*	*H*	*M*-*L*	*H*-*M*	*L*	*M*	*H*	*M*-*L*	*H*-*M*
	$\hat{\mu}$	$\hat{\mu}$	$\hat{\mu}$	nd	nd	$\hat{\mu}$	$\hat{\mu}$	$\hat{\mu}$	nd	nd
Victim info:										
$s_v \in \{M,H\}$						0.30	0.52	0.51	0.44	-0.02
$s_v = H$						0.02	0.02	0.06	0.02	0.18
died	0.37	0.27	0.24	-0.22	-0.06	0.30	0.33	0.37	0.05	0.09
female	0.38	0.43	0.30	0.10	-0.27	0.38	0.41	0.41	0.07	-0.01
intoxicated	0.06	0.04	0.01	-0.08	-0.18	0.05	0.04	0.06	-0.04	0.06
Offender info:										
$s_o \in \{M,H\}$	0.53	0.74	0.67	0.44	-0.14					
$s_o = H$	0.05	0.06	0.14	0.06	0.25					
college	0.21	0.20	0.29	-0.02	0.21	0.14	0.22	0.46	0.22	0.51
female	0.09	0.08	0.11	-0.02	0.08	0.07	0.10	0.11	0.10	0.04
age:										
<24	0.23	0.21	0.19	-0.05	-0.06	0.26	0.21	0.05	-0.11	-0.48
>40	0.31	0.34	0.35	0.07	0.02	0.26	0.36	0.37	0.22	0.03
intoxicated	0.13	0.12	0.17	-0.02	0.12	0.15	0.12	0.11	-0.12	-0.03
crime history	0.15	0.16	0.16	0.03	-0.00	0.20	0.14	0.13	-0.17	-0.02
car info present	0.60	0.55	0.59	-0.10	0.07	0.59	0.57	0.50	-0.03	-0.16
car price present	0.50	0.47	0.52	-0.06	0.09	0.51	0.49	0.42	-0.04	-0.13
imputed price	0.30	0.28	0.38	-0.06	0.38	0.27	0.30	0.41	0.12	0.43
Accident:										
hour info present	0.82	0.80	0.81	-0.04	0.02	0.80	0.82	0.85	0.04	0.08
hour (14 = 2 *pm*)	14.68	14.04	14.79	-0.10	0.12	14.76	14.21	14.41	-0.09	0.03
day (3 = Wed)	3.01	3.11	3.08	0.05	-0.01	2.95	3.10	3.23	0.07	0.07
month (7 = Jul)	7.37	7.52	7.41	0.04	-0.03	7.32	7.50	7.56	0.06	0.02
outside of city[1]	0.17	0.14	0.15	-0.09	0.02	0.17	0.15	0.17	-0.06	0.07
in city (streets)[1]	0.40	0.46	0.50	0.13	0.08	0.43	0.43	0.38	0.01	-0.10
# cases/police[2]	207	237	189	0.07	-0.11	155	262	220	0.26	-0.09
Outcomes:										
settlement	0.18	0.17	0.14	-0.03	-0.08	0.14	0.19	0.24	0.13	0.12
reached court	0.61	0.57	0.60	-0.08	0.06	0.60	0.59	0.58	-0.03	-0.01
out of which:										
lim. freedom	0.11	0.13	0.11	0.07	-0.08	0.11	0.12	0.10	0.02	-0.08
real incarcer.n	0.12	0.11	0.18	-0.04	0.20	0.14	0.10	0.11	-0.10	0.01

$\hat{\mu}$: mean; nd: normalized difference;
[1] if location is mentioned in the case description;
[2] number of all criminal traffic offenses registered per police department in 2013-2014.

B provides information on the distribution of the sample for each combination of offender-victim employment statuses across police departments' sample sizes.

Figure 1: The distribution of police departments by sample size



4 EMPIRICAL RESULTS

This section presents the empirical setup and results.

4.1 SETUP

I modify regression (8) to account for the unobserved variation in the marginal distributions of offenders and victims across locations and time by including indicators for police departments, p_c; hour&day, t_c; month, m_c; and year, y_c, as:

$$\begin{aligned} \mathbb{1}\{r_o = i \ vs. \ j\}_c = \alpha + \beta_M \mathbb{1}\{r_v = M\}_c + \beta_H \mathbb{1}\{r_v = H\}_c \\ + \gamma p_c + \tau_t t_c + \tau_m m_c + \tau_y y_c + u_c \\ i, j \in \{L, M, H\} \end{aligned} \tag{14}$$

where the dependent variable $\mathbb{1}\{r_o = i \ vs. \ j\}_c$ equals one if the offender in case c has status i, and zero, if it is j, so the sample is restricted just for two types of offenders, i and j. The regression estimates two slopes: β_M for M-type victims and

β_H for H-type victims. I use the linear regression specification because it provides consistent estimates for the population parameters even when the number of fixed effects is large, while logit or probit models are inconsistent (Greene, 2002; Angrist and Pischke, 2008). In total, I estimate the regression separately on three samples of offenders: M vs. L, H vs. L, and H vs. M.

Note that:

$$\begin{aligned}\alpha &= \delta_j^i(r_v = L)\\ \alpha + \beta_M &= \delta_j^i(r_v = M)\\ \alpha + \beta_H &= \delta_j^i(r_v = H)\end{aligned}$$

which allows to estimate the relative prosecution rates as in Equation (13) for $i, j, k, l \in \{L, M, H\}, i > j, k < l$.

Table 4: Regression results for prosecuted cases: with and without the fixed effects

	Dependent variable:								
	$\mathbb{1}\{s_o = M\ vs.\ L\}$			$\mathbb{1}\{s_o = H\ vs.\ L\}$			$\mathbb{1}\{s_o = H\ vs.\ M\}$		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$\hat{\beta}_M$	0.212	0.120	0.114	0.103	0.073	0.071	−0.007	0.012	0.014
	(0.014)	(0.017)	(0.017)	(0.017)	(0.021)	(0.020)	(0.010)	(0.012)	(0.013)
$\hat{\beta}_H$	0.117	0.035	0.046	0.147	0.106	0.092	0.067	0.081	0.089
	(0.050)	(0.057)	(0.057)	(0.057)	(0.085)	(0.083)	(0.038)	(0.047)	(0.046)
$\hat{\alpha}$	0.511	0.550	0.553	0.099	0.109	0.110	0.095	0.085	0.084
N	6,224	6,224	6,224	2,875	2,875	2,875	4,117	4,117	4,117
p-values for H0:									
$\hat{\beta}_M = 0$	0.000	0.000	0.000	0.000	0.000	0.001	0.482	0.315	0.272
$\hat{\beta}_H = 0$	0.019	0.545	0.421	0.010	0.212	0.268	0.081	0.082	0.050
$\hat{\beta}_M = \hat{\beta}_H = 0$	0.000	0.000	0.000	0.000	0.002	0.002	0.131	0.182	0.120
FIXED EFFECTS:									
Police department		✓	✓		✓	✓		✓	✓
Hour×Day, month, yr			✓			✓			✓

* These are the results for the OLS regression of the status of the offender on the status of the victim as in Equation (14). N is the number of observations.
** Standard errors are in parentheses, clustered at the police level.
*** α is reconstructed at the means of covariates.

4.2 PROBABILITY OF PROSECUTION

Table 4 presents the estimates for α, β_M and β_H for all three samples using the employment status as the proxy for resources, i.e., $r_o = s_o$ and $r_v = s_v$. Column 1, 4, and 7 provide the OLS results without the location and time fixed effects. Among the prosecuted cases, richer offenders are correlated with richer victims. Location fixed effects reduces the magnitude of the correlation, as predicted; however, not enough to eliminate all the correlation (See Columns 2, 5, and 8). The time fixed effects do not change the results (Columns 3,6, and 9).

According to the individual and joint significance of β-parameters (the lower part of Table 4), I can reject the null hypothesis that the prosecution is impartial. For the sample of M versus L offenders and H versus L offenders, β_M is significantly greater than zero. For the sample of H versus M offenders, it is β_H that is greater than zero at 5% significance level. Across all samples, β_H is estimated less precisely than β_M due to the smaller sample of H victims. As predicted, the signs of β_M and β_H are positive, implying that rich victims are better able to prosecute rich offenders.

Table 5: Prosecuted cases: symmetric odds ratios[†]

Definitions of H, L	Ratio	est.	se	p-val H0: Ratio= 1	N
Poor=L, Rich= M	$\frac{\rho_L^M(s_v = L)}{\rho_L^M(s_v = M)}$	0.618	(0.046)	0.000	6,224
Poor=L, Rich= H	$\frac{\rho_L^H(s_v = L)}{\rho_L^H(s_v = H)}$	0.488	(0.257)	0.047	2,875
Poor=M, Rich=H	$\frac{\rho_M^H(s_v = M)}{\rho_M^H(s_v = H)}$	0.518	(0.163)	0.003	4,117

[†] Symmetric odds ratios use the same definitions of *rich* and *poor* for both victims and offenders (see Section 2.3).

[*] Estimates of the odds ratios are based on equation (13) and the results in Table 4 with location and time fixed effects. For example, the first row is calculated using the results in Column 3 of Table 4, as $\frac{\hat{\alpha}/(1-\hat{\alpha})}{(\hat{\alpha}+\hat{\beta}_M)/(1-\hat{\alpha}-\hat{\beta}_M)}$. N is the number of observations.

[**] Standard errors are estimated using Delta method.

Table 5 reports the odds ratios for symmetric groups of offenders and victims.[16] It shows that the justice system prosecutes only six out of ten M-type offenders who hit an L-type victim in comparison to all ten who hit an M-type victim. In other words, four out of ten M-type offenders are missing for L-type victims. The test strongly rejects the null hypothesis that the justice system treats L and M individuals

[16] See the full table with the asymmetric groups in Appendix C.

impartially (i.e., that the odds ratio is equals one). Comparing L to H individuals, the odds ratio is lower – only every second H-type offender is not missing for L-type victims – but it is imprecise. Despite the imprecision, the test still rejects the impartiality under 5% significance level. Comparing M to H groups, every second H-type offender is missing for M-type victims, and the impartiality is rejected. Overall, the results show that a substantial number of offenders is missing when offenders are richer than their victims.

If I use other proxies for offenders' resources, I also find evidence that rich offenders are missing disproportionately for poor victims. Specifically, I find that around a half of luxury-car drivers is missing when victims are L-type. The estimates of the odds ratios for college-educated offenders are below the impartiality level of one when compared to offenders with no college degree, but they are not significant. At the same time, I find that offenders with the vocational training disappear disproportionately when compared to the offenders who have only a school degree. See Appendix D for more details.

4.3 Results by the severity of the offense

I perform the same analysis separately for the sample of cases with *no death + sober* offenses (less severe) and for the sample with all other cases that include deaths or intoxicated offenders (more severe). The results for both samples are in line with each other, see Column 1 and 2 of Table 6. However, the evidence of missing H-type offenders is much stronger for more severe cases. Every four out of five H offenders are missing when victim is L or M. The result is similar even if I restrict the sample only for cases in which victims died (Column 3 of Table 6). For less severe cases, for which incarcerations are very rare, the odds ratios are higher and imprecise, failing to reject the impartiality of justice there.

4.4 Probability of incarceration

Next, I look at the probability of incarceration and how it differs with the status of the victim. The probability of a settlement between the offender and the victim probably depends on the combination of each party's resources – i.e., richer offenders are more likely to settle with poorer victims. Instead, I do not distinguish between incarcerations and settlements, which I jointly name as *court punishment*. First, as is standard in the literature, I simply estimate the disparities in court punishment – disregarding the earlier stages of the justice process. I also show how the figures change if I correct the estimates for the missing rich offenders. Second, I use my identification strategy to directly estimate the share of rich offenders missing from court punishment.

Table 6: The odds ratios by case severity: Prosecuted cases

Groups:	Ratio	Sample 1 *No deaths+Sober*	Sample 2 *Deaths or Drunk*	Sample 3 *Deaths*
Poor=L, Rich=M	$\frac{\rho_L^M(s_v=L)}{\rho_L^M(s_v=M)}$	0.581	0.628	0.519
		(0.056)	(0.105)	(0.101)
		[0.000]	[0.000]	[0.000]
Poor=L, Rich=H	$\frac{\rho_L^H(s_v=L)}{\rho_L^M(s_v=H)}$	0.600	0.211	0.308
		(0.410)	(0.161)	(0.303)
		[0.329]	[0.000]	[0.022]
Poor=M, Rich=H	$\frac{\rho_M^H(s_v=M)}{\rho_L^M(s_v=H)}$	0.704	0.201	0.267
		(0.288)	(0.124)	(0.225)
		[0.303]	[0.000]	[0.001]

* The table reports the estimates of the odds ratios based on equation (13).
** Standard errors are in parentheses, estimated using Delta method.
*** p-values for testing H0: Odds ratio = 1 are in square brackets.

4.4.1 Conventional approach and the correction

First, I simply compare the differences in court punishment among the cases that have been selected for prosecution. It is a usual approach in the literature on socio-economic status disparities. I restrict the sample for the cases that have either intoxicated offenders or victim deaths, excluding second-time offenders. I estimate the following linear probability model:

$$\begin{aligned} y_c = &\sum_{j\in\{L,M,H\}}\sum_{k\in\{L,M,H\}} \beta_{j,k}\mathbb{1}\{s_v=j\}_c\mathbb{1}\{s_o=k\}_c+ \\ &\zeta_{0,1}O_c^{intox}+\zeta_{1,0}V_i^{dead}+\zeta_{1,1}O_c^{intox}\times V_c^{dead}+ \\ &\psi V_c^{intox}+\gamma police_c+u_c \end{aligned} \tag{15}$$

where $\beta_{j,k}$ captures the probability of incarceration for the victim of status j and the offender of status k; ζ-parameters capture the mean differences in the offense types (the intoxication state of the offender interacted with the victim's death); ψ estimate the difference in punishments when victims were intoxicated; and γ is a vector capturing police department fixed effects. The dependent variable y_c is an indicator function that equals one if the offender is punished.

Panel A of Table 7 presents the results for the probability of incarceration and Panel B for the probability of incarceration or settlement. The probability of incarceration differs across offenders and their victims. As expected, the incarceration

Table 7: Probabilities of incarceration or settlement for the prosecuted offenders

A. Probability of prison

		VICTIM		
		L	M	H
OFFENDER	L	0.256	0.309	0.488
		(0.018)	(0.038)	(0.164)
	M	0.219	0.247	0.493
		(0.019)	(0.022)	(0.142)
	H	0.197	0.277	0.345
		(0.057)	(0.104)	(0.240)

B. Pr. prison or settlement

		VICTIM		
		L	M	H
OFFENDER	L	0.440	0.405	0.508
		(0.021)	(0.044)	(0.147)
	M	0.402	0.403	0.507
		(0.019)	(0.024)	(0.111)
	H	0.391	0.416	0.524
		(0.056)	(0.110)	(0.282)

N obs 2,567. Sample includes cases where either victim has died or offender was intoxicated, or both. Predicted at mean values of covariates using regression (15).

rates are lower when victims are poorer, but the settlement rates are higher with such victims as well. In general, there might be reasons why some offenders are punished more severely than others, but we should expect that the punishment does not depend on the type of the victim. Indeed, if I look at Panel B, there is no significant difference across victims for any given offender status.

However, this analysis doesn't account for the fact that these are *conditional* probabilities, and they must be corrected for disparities at earlier stages. From the set of previous results (Column 2 of Table 6), I know that around 80% of H-type offenders are missing when victims have lower employment status, and 50% of M-type offenders are missing when victims are L. I multiply the conditional probabilities of incarceration for these three groups by the corresponding estimate of non-missing observations, i.e., the odds ratio:

$$\Pr(y|r_o = i, r_v = j) = \Pr(y|P = 1, r_o = i, r_v = j)\frac{\rho_j^i(r_v = j)}{\rho_j^i(r_v = i)} \tag{16}$$

I estimate the standard errors using Delta method, assuming that the covariance of the estimators is zero.

Table 8 presents the results for the ex-ante probabilities and shows that only 8% of H offenders are incarcerated or settled when the victim is L or M, which is lower than 50% estimate for H victim, but not statistically different. For M offenders, 25% are punished when the victim is L, which is substantially lower than 40% observed for M-victims. Notice that these differences wouldn't have been observed without the correction. It means that the bulk of the disparities happen at the very early stages of the judicial process.

Table 8: Ex-ante (corrected) probabilities of incarceration or settlement

		Pr. prison or settlement		
		VICTIM		
		L	M	H
OFFENDER	L	0.440	0.405	0.508
		(0.021)	(0.044)	(0.147)
	M	0.252^c	0.403	0.507
		(0.044)	(0.024)	(0.111)
	H	0.083^a	0.084^b	0.524
		(0.064)	(0.056)	(0.282)

[a,b,c] The estimates from Table 7 (Panel B) are multiplied by the following estimates from Table 6 (Column 2): (a) $\frac{\rho_L^H(r_v=L)}{\rho_L^H(r_v=H)}$, (b) $\frac{\rho_M^H(r_v=M)}{\rho_M^H(r_v=H)}$, (c) $\frac{\rho_L^M(r_v=L)}{\rho_L^M(r_v=H)}$

4.4.2 THE ODDS-RATIO APPROACH

Finally, I apply my identification strategy directly on the sample of cases in which the offender has been punished in court, and the results confirm the absence of impartiality – richer offenders are less likely to settle or be incarcerated. Table 9 presents the results for the symmetric odds ratios, which estimate the ratio of the relative court-punishment rates. For example, here $\rho_L^H(r_v = l)$ is the relative rate at which the system punishes H with respect to L offenders in court, when victims are L-type. It shows that one in two of M-type offenders avoid the punishment when their victims are L-type. It roughly corresponds to the estimates in Table 8, i.e. $0.252/0.403 \approx 0.6$. Moreover, H-type offenders are almost never punished when their victims are L-type and only 14% are punished when the victims are M-type. These estimates are broadly in line with the results in Table 8: $0.083/0.524 \approx 0.16$ and $0.084/0.524 \approx 0.16$. All odds ratios are significantly below one, rejecting the impartiality of the justice system.

4.5 HOW DO RICH OFFENDERS DISAPPEAR?

Although it is clear that many rich offenders are missing when victims are poor, it is hard to say why and how exactly they avoid prosecution. This section provides a few hypotheses and suggestive evidence on the matter.

For instance, more educated offenders may simply be better defenders of their own interests, providing a better witness account of events.[17] If so, this channel

[17] I would like to thank an anonymous referee for pointing out this possible channel.

Table 9: The symmetric odds ratios: *Incarceration+Settlement* cases

Groups	Ratio	est.	se	p-val H0: Ratio= 1	N
Poor=L, Rich=M	$\frac{\rho_L^M(s_v=L)}{\rho_L^M(s_v=M)}$	0.486	(0.136)	0.000	1,014
Poor=L, Rich=H	$\frac{\rho_L^H(s_v=L)}{\rho_L^H(s_v=H)}$	0.005	(0.026)	0.000	489
Poor=M, Rich=H	$\frac{\rho_M^H(s_v=M)}{\rho_M^H(s_v=H)}$	0.142	(0.142)	0.000	655

* The table reports the estimates of the odds ratios as in Table 5, but for the sample of cases with the intoxicated offenders or victim deaths which have resulted in a settlement or an incarceration of the offender.

** Standard errors are estimated using Delta method.

should have also benefited academics, doctors, or teachers. While they are more educated than the average workers, in Russia they usually do not earn more than average. According to the official statistics for 2013, the average monthly salaries varied from $21,000$ RUB (c. 680 USD) for the junior medical workers to $42,000$ RUB for medical doctors and science workers, which is comparable to the $30,000$ RUB average monthly salary across all professions, but substantially below $60,000$ RUB in financial and mining sectors. Table 10 shows that these workers are also much more likely to have a college degree than any comparable employment status group.

Table 10: The share of college graduates

	By sphere:	
s_o	Other	Medicine, education, science
M	0.205	0.620
H	0.430	0.778

Based on $61,000$ offenders.

I again estimate (14) on the sample of prosecuted cases but including a separate dummy for those highly-educated yet non-wealthy workers:

$$\begin{aligned}\mathbb{1}\{s_o = i\ vs.\ j\}_c = \alpha + \beta_M \mathbb{1}\{s_v = M\}_c + \beta_H \mathbb{1}\{s_v = H\}_c \\ + \hat{\beta}_{med,edu,sci} \mathbb{1}\{\text{V works in medicine, education, science}\} \\ + \gamma p_c + \tau_t t_c + \tau_m m_c + \tau_y y_c + u_c \\ i, j \in \{L, M, H\}\end{aligned} \tag{17}$$

and present the results in Table 10. I do not find supporting evidence for this hypothesis since the coefficient $\beta_{med,edu,sci}$ is not significantly different from zero, whereas the estimates for β_M and β_H have not changed with respect to the results in Table 4.

Table 11: Regression results: medical, education, and science workers.

	Dependent variable:		
	(1)	(2)	(3)
	$\mathbb{1}\{s_o = M\ vs.\ L\}$	$\mathbb{1}\{s_o = H\ vs.\ L\}$	$\mathbb{1}\{s_o = H\ vs.\ M\}$
$\hat{\beta}_M$	0.114	0.069	0.015
	(0.017)	(0.021)	(0.013)
$\hat{\beta}_H$	0.046	0.092	0.092
	(0.057)	(0.083)	(0.046)
$\hat{\beta}_{med,edu,sci}$	−0.004	0.022	−0.038
	(0.049)	(0.061)	(0.033)
p-value for H0: $\hat{\beta}_{med,edu,sci} = 0$	0.931	0.717	0.249
FIXED EFFECTS:			
Police department	✓	✓	✓
Hour×Day, month, yr	✓	✓	✓

* These are the results for the OLS regression of the status of the offender on the status of the victim including the dummy for medical, education, and science workers as in Equation (17).
** Standard errors are in parentheses, clustered at the police level.

The results also suggest that in more egregious cases rich offenders avoid prosecution at least as often as they are missing from less severe ones. I contend that this is more probably attributable to corruption, rather than other channels. If this result was driven solely by prosecutors choosing the easier-to-prosecute cases, there would be no reason to observe fewer rich offenders in the most egregious cases. The same reasoning applies if one believes that the quality of lawyers of both offenders and victims played a pivotal role. While it might be easier for rich offenders to be illegitimately reclassified from intoxicated offenders into sober ones by influencing the medical report, it is harder to imagine why rich offenders are less likely to be prosecuted when their victims die.

4.6 Checking for residual spatial correlation

So far I have attributed the low share of rich offenders among poor victims to the lack of impartiality of the legal system. However, we could also imagine that these results are driven by a substantial spatial segregation of rich and poor individuals within the boundaries of police departments. Such spatial segregation would imply that poorer victims are naturally less likely to encounter rich offenders in a given police department, i.e., $\Pr(r_o = H | r_v = L) < \Pr(r_o = H | r_v = H)$, which would violate the independence assumption in the identification strategy. While I cannot test for this residual spatial segregation directly – because accident location is reported only at police department resolution – I can provide some indirect evidence on the magnitude of the potential problem.

First, I test whether police-level location controls are enough to remove the spatial segregation with respect to other individual characteristics such as being a foreign citizen, a public employee, or a student. I assume that if the economic activity in Russian society is sharply segregated at a very localized level, my empirical test should also fail to remove the positive correlation between victims and offenders of that belong to the above-mentioned groups, which is not the case. Citizenship and the employment status as a public employee show substantial correlation, which disappears after controlling for police departments. See Appendix E.4 for more details.

Second, I test the robustness of my results by gradually removing large police department (those with the largest number of observations), assuming that the number of observations per department is positively dependent on the size of the territory the department oversees. The point estimates do not change substantially, although the exercise leads to a loss of precision as the result of a substantial decrease in the data variability. Nevertheless, the odds ratios that have been precisely estimated to begin with remain significantly below one – i.e., below the impartiality level – even after dropping the police departments with as little as twenty observations. See Appendix E.5.

5 Conclusion

I propose a new approach to measure unjustified disparities in judicial and police outcomes, by looking at criminal traffic offences and comparing the odds ratios of observing rich offenders when victims are poor and when they are rich. The quasi-experimental and unintentional nature of traffic accidents allows for causal interpretation of the observed asymmetries in the odds ratios. Unlike conventional measures of judicial disparities, this approach does not require to observe all legally relevant characteristics.

I apply the methodology on an original dataset compiled from Russian police records for 2013–2014 and classify offenders and victims of criminal traffic offenses into resource brackets based on their employment status. I find that rich offenders are prosecuted less frequently when their victims are poor than when their victims are rich. Moreover, the number of missing rich offenders does not decline with the severity of their offense.

The results can be interpreted as an indicator of the broader institutional quality of the Russian justice system. Indeed, whatever channels have contributed to the disparities in criminal traffic cases are also likely to be used by rich offenders in other types of criminal cases. Moreover, for intentional crimes, rich offenders could rationally choose to victimize less powerful individuals, increasing the extent to which rich offenders are missing.

The conventional method in the literature on disparities in socio-economic status would have failed to find any difference in court outcomes using the same data. Compared to this approach, my methodology offers the advantage of measuring directly the relative ex-ante probabilities of court punishment for different groups of offenders and victims. Moreover, offenders involved in unintentional crimes – like traffic offenses – are more representative of their underlying social group compared to those involved in intentional crimes.

The odds-ratios approach can be easily applied to other countries and settings. The analysis could be extended by exploiting the representative body of the texts of court rulings. In the absence of wealth proxies for victims, the methodology may potentially be extended for car-to-car collisions to use the imputed car prices both for offenders and for victims. Moreover, the odds-ratio approach can be used to test other types of police and judicial disparities – like gender, ethnic, or racial biases.

A IDENTIFICATION FOR CONTINUOUS RESOURCES

When resources are continuous, given the random matching and the impartiality of justice, the observed density of victim's resources, r_v, should be independent from the density of offender's resources, r_o. To test the independence, I can fit a following population regression function:

$$r_{o,c} = \alpha + \beta r_{v,c} + u_c \tag{18}$$
$$H0 : \beta = 0 \tag{19}$$

and test whether β is equal to zero. If it is not, it means that r_o is not independent from r_v, which rejects the impartiality of prosecution.

Denote the density of offenders in the population as $f(r_o)$. Then, the observed density of offenders among the prosecuted cases conditional on victim's resources is:

$$f(r_o|r_v, P = 1) = \frac{\pi(r_o, r_v) f(r_o)}{\int \pi(r_o, r_v) f(r_o) dr_o}$$

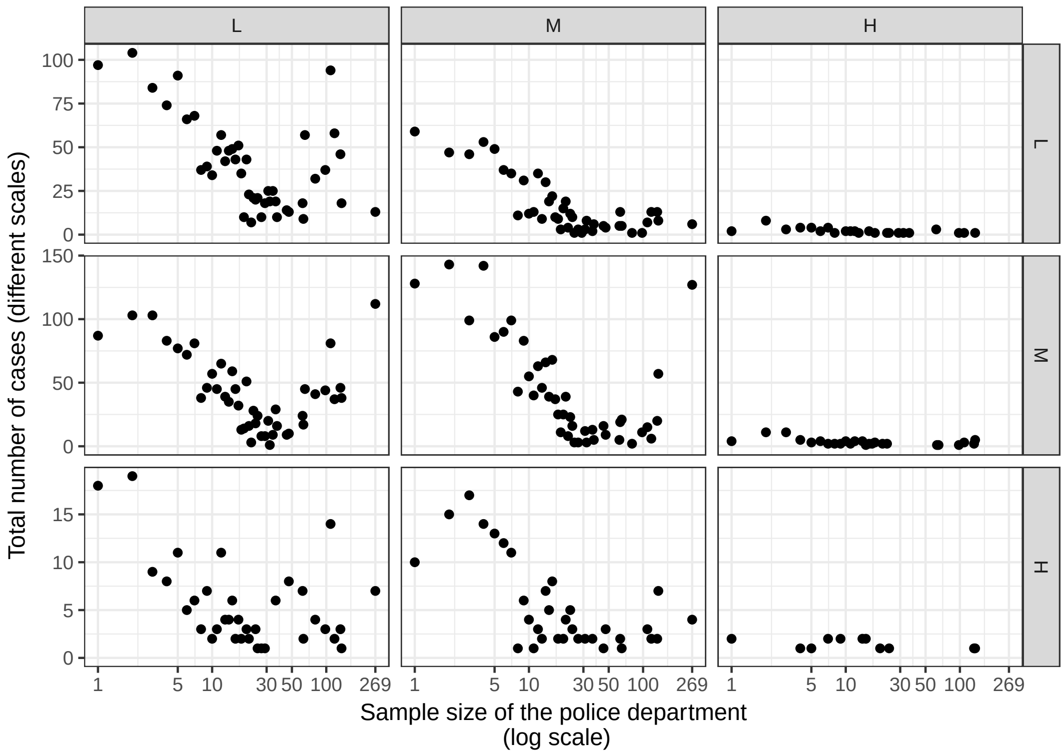
Thus, for any two levels of offenders resources $\underline{r}_o$ and $\bar{r}_o$ and any two levels of victim's resources $\underline{r}_v$ and $\bar{r}_v$, such that $\underline{r}_o < \bar{r}_o$ and $\underline{r}_v < \bar{r}_v$, the following ratio is a measure of relative prosecution ratios:

$$\frac{f(\bar{r}_o|\underline{r}_v, P = 1)/f(\underline{r}_o|\underline{r}_v, P = 1)}{f(\bar{r}_o|\bar{r}_v, P = 1)/f(\underline{r}_o|\bar{r}_v, P = 1)} = \frac{\pi(\bar{r}_o, \underline{r}_v)/\pi(\underline{r}_o, \underline{r}_v)}{\pi(\bar{r}_o, \bar{r}_v)/\pi(\underline{r}_o, \bar{r}_v))} = \frac{\rho_{\underline{r}_o}^{\bar{r}_o}(\underline{r}_v)}{\rho_{\underline{r}_o}^{\bar{r}_o}(\bar{r}_v)} \tag{20}$$

To estimate (20), I need to empirically approximate the conditional densities $f(r_o|\underline{r}_v, P = 1)$ and $f(r_o|\bar{r}_v, P = 1)$.

B Additional descriptive statistics

Figure 2: Number of observations by the police department's sample size
within each Offender-Victim resource combination (Columns correspond to the employment status of the victim: rows, to the offender)



C ALL ODDS RATIOS

Table 12: The odds ratios at different combinations of offenders' and victims' employment statuses

		ρ^i_j:		
		ρ^M_L	ρ^H_L	ρ^H_M
Victims	$\frac{\rho^i_j(s_v=L)}{\rho^i_j(s_v=M)}$	0.618	0.561	0.846
		(0.046)	(0.089)	(0.129)
		[0.000]	[0.000]	[0.233]
	$\frac{\rho^i_j(s_v=L)}{\rho^i_j(s_v=H)}$	0.828	0.488	0.439
		(0.197)	(0.257)	(0.147)
		[0.382]	[0.047]	[0.000]
	$\frac{\rho^i_j(s_v=M)}{\rho^i_j(s_v=H)}$	1.340	0.870	0.518
		(0.306)	(0.445)	(0.163)
		[0.267]	[0.770]	[0.003]

* The table reports odds ratios that estimate the relative prosecution rate, $\frac{\rho^i_j(r_v=k)}{\rho^i_j(r_v=l)}$ as in Equation (13), where the column indicates ρ^i_j, while the rows indicate the full ratio. For example, the ratio of 0.618 in the first row and the first column corresponds to $\frac{\rho^M_L(s_v=L)}{\rho^M_L(s_v=M)}$.

** Standard errors are in parentheses, estimated using Delta method.

*** p-values, in square brackets, for testing the null hypothesis that justice is impartial, i.e. 'the odds ratio = 1'

D RESULTS USING OTHER PROXIES

Table 13 presents selected odds ratios estimated using car and education as alternative proxies for offenders' resources, $r_o = c_o$ or $r_o = e_o$, while victims' resources are still proxied by the employment status. According to the results, every second luxury-car driver is missing for *L* victims using *H* victims and inferior cars as the controls. At the same time, there is no evidence of missing drivers of luxury cars for *M* victims using *H* victims and the drivers of normal cars as the controls. Also, the distribution of normal to inferior car drivers appears balanced across *L* and *M* victims.

According to Panel B of Table 13, every fifth offender with vocational training is missing for *L* victims, using *M* victims and offenders with a school degree as the controls. At the same time, there is no evidence that college graduates are disproportionately missing either for L or M victims.

Table 13: The odds ratios using alternative proxies: Prosecuted cases

Groups	Ratio	est.	se	p-val H0: Ratio= 1
A. Car as the proxy for r_o				
Offender: *Poor*= *inf*, *Rich*= *nor* Victim: $Poor = L$, $Rich = M$	$\frac{\rho_{inf}^{nor}(s_v=L)}{\rho_{inf}^{nor}(s_v=M)}$	1.112	(0.108)	0.303
Offender: *Poor*= *inf*, *Rich*= *lux* Victim: $Poor = L$, $Rich = H$	$\frac{\rho_{inf}^{lux}(s_v=L)}{\rho_{inf}^{lux}(s_v=H)}$	0.505	(0.182)	0.007
Offender: *Poor*= *nor*, *Poor*= *lux* Victim: $Poor = M$, $Rich = H$	$\frac{\rho_{nor}^{lux}(s_v=M)}{\rho_{nor}^{lux}(s_v=H)}$	1.201	(0.412)	0.626
B. Education as the proxy for r_o				
Offender: *Poor*= *sch*, *Rich*= *voc* Victim: $Poor = L$, $Rich = M$	$\frac{\rho_{sch}^{voc}(s_v=L)}{\rho_{sch}^{voc}(s_v=M)}$	0.792	(0.088)	0.017
Offender: *Poor*= *sch*, *Rich*= *col* Victim: $Poor = L$, $Rich = H$	$\frac{\rho_{sch}^{col}(s_v=L)}{\rho_{sch}^{col}(s_v=H)}$	0.786	(0.432)	0.620
Offender: *Poor*= *voc*, *Poor*= *col* Victim: $Poor = M$, $Rich = H$	$\frac{\rho_{voc}^{col}(s_v=M)}{\rho_{voc}^{col}(s_v=H)}$	0.825	(0.227)	0.441

* The table reports the estimates of the odds ratios as in equation 11.
** Standard errors are estimated using Delta method.

E Other robustness Checks

E.1 Victim's culpability in the accident

If we imagine that prosecutors select cases based on the culpability of offender's behavior relative to victim's behavior, it may lead to observed correlation of offenders and victims both likely to engage in reckless behavior. If reckless behavior correlates with resources, then it may naturally create correlation of rich victims with rich offenders. For example, the prosecutor does not prosecute the offender when his victim was drunk at the time of the accident, unless the offender himself was drunk. If poor individuals are more likely to be drunk, we will observe more cases with poor offenders and poor individuals. To check for this, I estimate model with an indicator variable for intoxicated victims on the indicator variable for intoxicated offenders, while constraining the sample only to offenders and victims of the same employment status. The estimates in Table 14 show no correlation between the intoxication of victims and offenders. I use it as an evidence that prosecutors mainly base their decisions on the culpability of offender in the accident.

Table 14: Intoxicated offenders and intoxicated victims

	Dependent variable is Victim was intoxicated	
	Est.	SE
$\hat{\beta}$: Offender was intoxicated	.007	(.016)

Sample: Cases where offenders and victims are of the same employment status: $s_d = s_v$.

E.2 Robustness to missing time information

I test whether the missing information on the hour of the traffic accident correlates with a particular combination of offender-victim statuses by running the following regression:

$$
\begin{aligned}
\mathbb{1}\{has\ hour\}_c = & \alpha + \beta_M \mathbb{1}\{s_o = M\}_c + \beta_H \mathbb{1}\{s_o = H\}_c \\
& + \gamma_M \mathbb{1}\{s_v = M\}_c + \gamma_H \mathbb{1}\{s_v = H\}_c \\
& + \sum_{j \in \{M,H\}} \sum_{k \in \{M,H\}} \zeta_{j,k} \mathbb{1}\{s_o = j\}_c \mathbb{1}\{s_v = k\}_c + u_c
\end{aligned} \tag{21}
$$

where the dependent variable $\mathbb{1}\{has\ hour\}_c$ is the dummy, which equals to one if the information on the hour is present, and the explanatory variables are dummies for statuses of offenders and victims. Importantly, all ζ-parameters should be equal to zero, lest it biases the results. Higher offender status is correlated the presence of information on the time of the accident. However, all the interaction terms – including the type of the victim – are jointly insignificant (Table 15). Only one of the interaction terms, for H-H group, seems to be significantly different from zero at 10% significance level. Since the sign is negative, the missing hour information makes H-type offenders disappear disproportionately for H-type victims from my sample. Such disappearance may bias my results but in the opposite direction: it will make the odds ratios seem higher and thus the results look closer to an impartial justice.

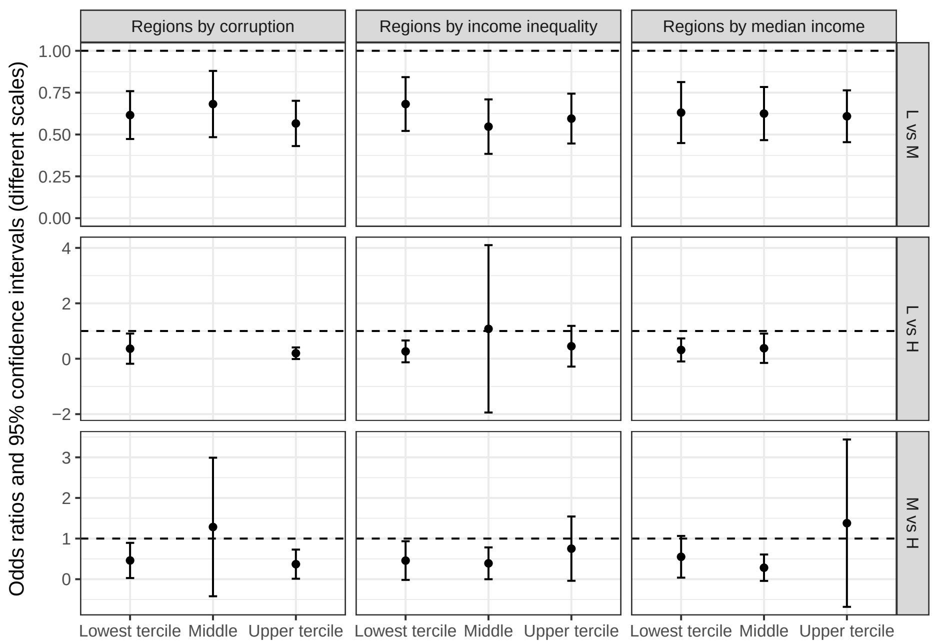
Table 15: Probability the time information is missing. N obs. $8,136$

	est.	se	p-val
s_o:			
β_M	.025	(.0110)	.0226
β_H	.057	(.0251)	.0226
s_v:			
γ_M	$-.021$	(.0171)	.2274
γ_H	.069	(.0567)	.2221
$s_o \times s_v$:			
$\zeta_{M,M}$	.011	(.0233)	.6400
$\zeta_{M,H}$	$-.078$	(.0777)	.3168
$\zeta_{H,M}$	.001	(.0422)	.9834
$\zeta_{H,H}$	$-.174$	(.0953)	.0683
α	.800	(.0077)	.0000
$H_0: \delta_{i,j} = 0, \forall i,j \in \{M,H\}$			
p-value $H_0 = 0.5161$			

E.3 Robustness to regional breakdowns

The degree of missing rich offenders does not vary across different regional breakdowns of Russia. First, I split regions into terciles based on an index of petty corruption.[18] Then, I divide the regions by income inequality, which I proxy with the ratio of the median to average wage incomes in the region.[19] Finally, I look at the regional breakdown based on the median wage income. Figure 3 reports the results of replicating the three odds ratios from Table 5 across regional breakdowns and shows that the estimates are not significantly different across them.

Figure 3: Relative probability of prosecution by different regional breakdown
(Columns correspond to the employment status of the victim: rows, to the offender)



The odds ratio "i vs j" estimates $\frac{\rho_i^j(s_v=i)}{\rho_i^j(s_v=j)}$ *as in Table 5; The odds ratio of* $\frac{\rho_L^H(s_v=L)}{\rho_L^H(s_v=H)}$ *is non-estimable for the middle tercile by corruption index and the upper tercile by median income due to the lack of variation.*

[18] The Ministry of Economic Development of the Russian Federation and the Public Opinion Foundation calculated the petty corruption index for 70 regions based on survey results in 2010. The report in Russian is available at `www.indem.ru/corrupt/doklad_cor_INDEM_FOM_2010.pdf`. I have imputed the index for the remaining 13 regions based on the index of the neighboring regions.

[19] The statistics on the median and average wage income is provided by the Federal State Statistics Service at `www.gks.ru/free_doc/new_site/population/bednost/tabl/3-1-5.doc`

E.4 PLACEBO TESTS

Table 16 presents the results of the tests for residual spatial segregation using other characteristics of offenders and victims. I expect these characteristics to be spatially correlated – e.g., student-drivers tend to drive closer to student-pedestrians, foreigners tend to live in certain neighborhoods where they cross roads and drive cars. At the same time I do not expect these characteristics to fail impartiality test.

Table 16: Placebo tests on other characteristics

	(1)	(2)	(3)
	A. Offender is a Russian citizen		
$\hat{\beta}$: Victim is a Russian citizen	0.082	0.023	0.022
	(0.027)	(0.030)	(0.030)
$\hat{\alpha}$	0.879	0.937	0.938
p-value for H0: $\hat{\beta} = 0$	0.003	0.433	0.455
	B. Offender is a public employee		
$\hat{\beta}$: Victim is a public employee	0.114	0.004	0.009
	(0.045)	(0.059)	(0.058)
$\hat{\alpha}$	0.026	0.029	0.029
p-value for H0: $\hat{\beta} = 0$	0.010	0.949	0.882
	C. Offender is a student		
$\hat{\beta}$: Victim is a student	0.013	-0.013	-0.015
	(0.016)	(0.017)	(0.017)
$\hat{\alpha}$	0.028	0.028	0.028
p-value for H0: $\hat{\beta} = 0$	0.418	0.448	0.383
FIXED EFFECTS:			
Police department		✓	✓
Hour×Day, month, yr			✓

* These are the results for OLS regression: $O_c = \alpha + \beta V_c + \gamma p_c + \tau_t t_c + \tau_m m_c + \tau_y y_c + u_c$, where p – police department, t – Hour×Day, m – month, y – year, O – dummy for the offender's characteristic, V – dummy for the victim's characteristic

** Standard errors are in parentheses, clustered at the police level.

*** The sample for the public employees consists of cases where offenders and victims are both M-types: $s_o = s_v = M$; The public employee category excludes law enforcement and other government officials; α is reconstructed at the means of covariates

First, I observe that Russian citizens as victims predict Russian citizens as offenders, see Panel A in Table 16. The highest concentration of foreign citizens is expected in big cities and the regions that border with other countries. Hence, the observed correlation is probably linked to the spatial allocation of foreigners in Russia. Of course, it may be also linked to general income differences between foreigners and citizens and the system's bias against foreign victims. Crucially, by adding controls for the police department and time of the accident, the correlation drops, and, while not dropping to zero, it becomes statistically insignificant.

I then repeat the exercise for public employees (excluding law enforcement and other government officials). I use only the cases where both offenders and victims are M-type to avoid contamination by the relative resource imbalances. As expected, public employees as victims show substantial correlation with public employees as offenders and the correlation disappears after the inclusion of the fixed effects (See Panel B). Similarly, I look at university students and find that they do not show statistically significant correlation before or after the inclusion of the fixed effects (Panel C).

Overall, it seems that the chosen controls for location and time work as intended, maintaining the question of why we see statistically significant disparities across the resource groups.

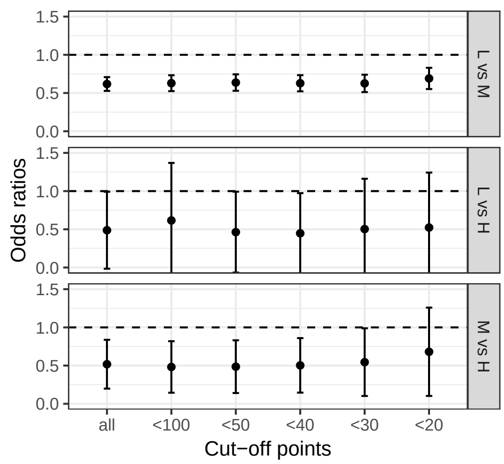
E.5 Robustness to excluding big police departments

Next, I check the robustness of my main results to a gradual removal of the observationally big police departments from the sample. First, I exclude the departments with a hundred or more observations. Then I lower the cut-off point to 50, 40, 30, and 20 observations.

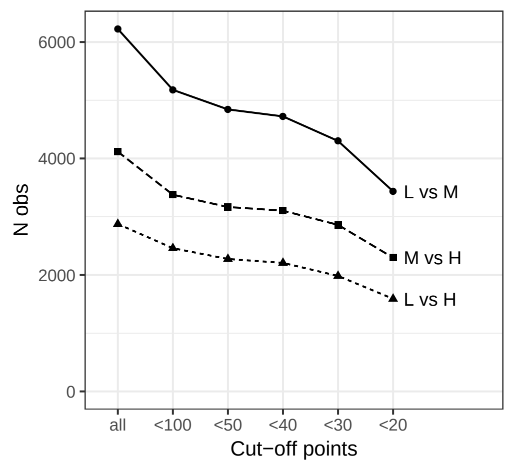
Figure 4 present the results of this exercise for the odds ratios for prosecution (Subfigure 4a) and for court punishment (Subfigure 4c). For convenience, the figure also provide the baseline estimates for the whole sample under label 'all'(the results of Table 5 and Table 9). The information on the remaining sample size at each cut-off point is in Subfigures 4b and 4d.

Subfigures 4a and 4c show that the point estimates for the odds ratios do not change much across different choices for cut-off points. The confidence intervals, however, expand, especially for the odds ratios that have been imprecisely estimated to begin with. Nevertheless, there are odds ratios that remain always robust: those that compare L to M individuals for the probability of prosecution and L to H offenders for the probability of court punishment.

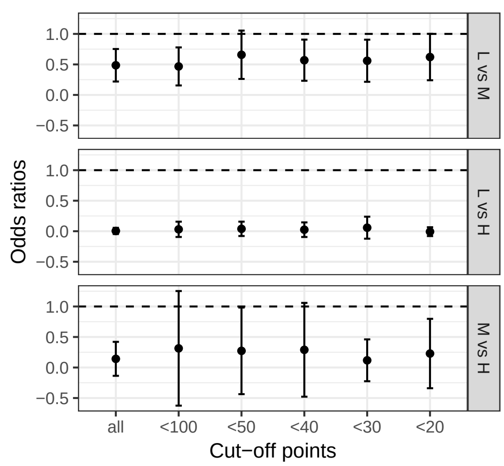
Figure 4: Odds ratios at different cut-offs for police departments



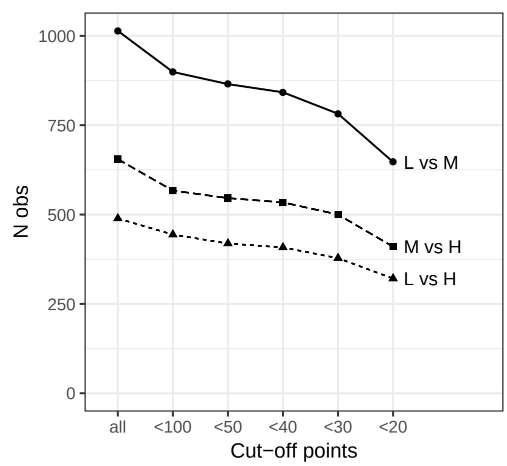
(a) Prosecution: est. with 95% CI



(b) Prosecution: sample size



(c) Court punishment: est. with 95% CI



(d) Court punishment: sample size

Cut offs: all – all police departments, $< x$ – only those police departments with less than x observations, $x \in \{100, 50, 40, 30, 20\}$; The odds ratio "$i$ vs j" estimates $\frac{\rho_i^j(s_v=i)}{\rho_i^j(s_v=j)}$ as in Table 5 or Table 9; CI = confidence interval

REFERENCES

Acemoglu, D. and J. A. Robinson (2013). *Why nations fail: The origins of power, prosperity, and poverty*. Crown Business.

Alesina, A. and E. La Ferrara (2014). "A test of racial bias in capital sentencing". *The American Economic Review* 104.11, pp. 3397–3433.

Angrist, J. D. and J.-S. Pischke (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Arlen, J. H. (1992). "Should Defendants' Wealth Matter?" *The Journal of Legal Studies* 21.2, pp. 413–429.

Bacher, J.-L. et al. (2005). "Another Look at the" Corporate Advantage" in Routine Criminal Proceedings". *Canadian Journal of Criminology and Criminal Justice* 47.4, pp. 685–708.

Champion, D. J. (1989). "Private counsels and public defenders: A look at weak cases, prior records, and leniency in plea bargaining". *Journal of Criminal Justice* 17.4, pp. 253–263.

Clark, L. D. (1997). "All Defendants, Rich and Poor, Should Get Appointed Counsel in Criminal Cases: The Route to True Equal Justice". *Marquette Law Review* 81.1, pp. 47–78.

Coatsworth, J. H. (2008). "Inequality, institutions and economic growth in Latin America". *Journal of Latin American Studies* 40.3, pp. 545–569.

D'Alessio, S. J. and L. Stolzenberg (1993). "Socioeconomic status and the sentencing of the traditional offender". *Journal of Criminal Justice* 21.1, pp. 61–77. ISSN: 00472352.

Dell, M. (2010). "The Persistent Effects of Peru's Mining Mita". *Econometrica* 78.6, pp. 1863–1903. ISSN: 1468-0262.

Garoupa, N. and H. Gravelle (2003). "Efficient deterrence does not require that the wealthy should be able to buy justice". ISSN: 09324569.

Glaeser, E., J. Scheinkman, and A. Shleifer (2003). "The injustice of inequality". *Journal of Monetary Economics* 50.1, pp. 199–222. ISSN: 03043932.

Glaeser, E. L. and B. Sacerdote (2003). "Sentencing in homicide cases and the role of vengeance". *The Journal of Legal Studies* 32.2, pp. 363–382.

Glaeser, E. L. et al. (2004). "Do Institutions Cause Growth?" *Journal of Economic Growth* 9.3, pp. 271–303. ISSN: 1573-7020.

Greene, W. H. (2002). "The behavior of the fixed effects estimator in nonlinear models".

Hagan, J. (1982). "The corporate advantage: A study of the involvement of corporate and individual victims in a criminal justice system". *Social Forces* 60.4, pp. 993–1022.

Hartley, R. D., H. V. Miller, and C. Spohn (2010). “Do you get what you pay for? Type of counsel and its effect on criminal court outcomes”. *Journal of Criminal Justice* 38.5, pp. 1063–1070.
Ichino, A., M. Polo, and E. Rettore (2003). “Are judges biased by labor market conditions?” *European Economic Review* 47.5, pp. 913–944.
Imbens, G. W. and J. M. Wooldridge (2009). “Recent developments in the econometrics of program evaluation”. *Journal of economic literature* 47.1, pp. 5–86.
Johansson, S. and O. Nygren (1991). “The Missing Girls of China: A New Demographic Account”. *Population and Development Review* 17.1, pp. 35–51. ISSN: 00987921, 17284457.
Knowles, J., N. Persico, and P. Todd (2001). “Racial Bias in Motor Vehicle Searches: Theory and Evidence”. *Journal of Political Economy* 109.1, pp. 203–229. eprint: `https://doi.org/10.1086/318603`.
Kobayashi, B. H. and J. R. Lott (1996). “In defense of criminal defense expenditures and plea bargaining”. *International Review of Law and Economics* 16.4, pp. 397–416. ISSN: 01448188.
Levitt, S. D. and J. Porter (2001). “How Dangerous Are Drinking Drivers?” *Journal of Political Economy* 109.6, pp. 1198–1237. eprint: `https://doi.org/10.1086/323281`.
Lott Jr, J. R. (1987). “Should the Wealthy Be Able to" Buy Justice"?” *Journal of Political Economy* 95.6, pp. 1307–1316.
Nunn, N. and D. Puga (2012). “Ruggedness: The blessing of bad geography in Africa”. *Review of Economics and Statistics* 94.1, pp. 20–36.
Paneyakh, E. (2014). “Faking performance together: systems of performance evaluation in Russian enforcement agencies and production of bias and privilege”. *Post-Soviet Affairs* 30.2-3, pp. 115–136.
Rattner, A., H. Turjeman, and G. Fishman (2008). “Public versus private defense: Can money buy justice?” *Journal of Criminal Justice* 36.1, pp. 43–49.
Shayo, M. and A. Zussman (2011). “Judicial ingroup bias in the shadow of terrorism”. *Quarterly Journal of Economics* (126), pp. 1447–1484.
Shi, Y. and J. J. Kennedy (2016). “Delayed Registration and Identifying the “Missing Girls” in China”. *The China Quarterly* 228, pp. 1018–1038.
Shklyaruk, M. (2014). *Traektoria ugolovnogo dela v oficial'noi statistike [The criminal case trajectory in official statistics]*. SPB: Institute for the Rule of Law.
Sklyaruk, M. and D. Skougarevsky (2015). *Kriminal'naya statistika: mekhanizmy formirovaniya, prichiny iskazheniya, puti reformirovaniya [Crime statistics: How it is constructed, tampered with, and could be reformed (analytical report)]*.
The World Justice Project (2014). *Rule of Law Index*. Tech. rep.

Trochev, A. (2014). “How Judges Arrest and Acquit.” *Historical Legacies of Communism in Russia and Eastern Europe*, p. 152.

Ulmer, J. T. (2012). “Recent developments and new directions in sentencing research”. *Justice Quarterly* 29.1, pp. 1–40.

Volkov, V. (2016). “Legal and Extralegal Origins of Sentencing Disparities: Evidence from Russia’s Criminal Courts”. *Journal of Empirical Legal Studies* 13 (4), pp. 637–665.

Yoav, D. (1999). “Do the haves still come out ahead-resource inequalities in ideological courts: the case of the Israeli High Court of Justice”. *Law & Soc’y Rev.* 33, p. 1059.

Zatz, M. S. and J. Hagan (1985). “Crime, time, and punishment: An exploration of selection bias in sentencing research”. *Journal of Quantitative Criminology* 1.I, pp. 103–126. ISSN: 07484518.