MADINA KURMANGALIYEVA

CRIMINAL JUSTICE AND WEALTH INEQUALITY: HOW MUCH FREEDOM CAN MONEY BUY IN RUSSIA?[1]

Abstract. In many countries, the public would like to know whether their criminal judicial system is more lenient towards wealthier citizens and if so, by how much. The calculation of the relevant statistics requires knowledge of defendants' wealth, which is not observed in most circumstances. To address this issue, this paper proposes to base the analysis on criminal traffic accidents and use the information on the car of the defendant as the proxy to wealth and other proxies, if available. Utilizing the multiple proxy approach based on the data from Russia, the analysis finds that the Russian judicial system is more lenient to the defendants with higher wealth. The inequality partially comes from the design of the legal system which provides for certain legal channels that naturally create those disparities. However, the data suggests that the inequality persists even after accounting for those legal channels.
Keywords. Sentencing disparities, Wealth inequality, Criminal justice

1. INTRODUCTION

In many countries, especially developing ones, the public is concerned whether their judicial system treats all citizens equally, irrespective of their wealth. Perhaps, the area that interests people the most is criminal justice, given the high stakes of someone's freedom and future. Unfortunately, there are no readily available

[1] I would like to thank Centro Einauidi for the award, Andrea Ichino and Andrea Mattozzi for their supervision, as well as my peers at the European University Institute. I am especially grateful to Vadim Volkov, Dmitriy Skougarevskiy, Irina Chetverikova and other members of the Institute for the Rule of Law at the European University at Saint-Petersburg for data access, fruitful discussions and valuable advice. 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 Gözde Çörekçioğlu, Ada Gonzalez-Torres, Anastasia Antsygina, Sylta Cornils, Matteo Sostero.

measures of the court rulings disparities for different wealth groups that would allow the public to monitor such discrepancies officially and transparently. Some countries provide access to court rulings, but the wealth of the defendant is not recorded. This paper proposes to base the analysis on criminal traffic offenses, which provide a handy proxy to wealth – the value of the car of the defendant. Using a unique database on the population of criminal cases for Russia, this paper answers the question of how strong is the effect of wealth on court rulings in this country, and even sheds light on some of the channels of the effect.

The mistrust in the impartiality of the judiciary is high in Russia: more than a half of Russians do not trust the judicial system and perceive it as corrupt.[2] According to the World Justice Project, Russia ranks 74th among 102 countries for the effectiveness of the criminal justice system, which summarizes public and expert opinion on criminal justice's impartiality, freedom from corruption and improper influence.[3] Mass media focuses on cases that involve the rich and the powerful with the public discussion in social media calling for justice.

The Russian constitution guarantees the equality of all citizens before the law. At the same time, the judicial system provides for certain legal channels that naturally create disparities among different wealth groups. For non-severe crimes, which include traffic offenses, the judicial system allows the defendant and the victim to settle, where the settlement terminates the criminal prosecution of the defendant. If no settlement can be reached, the fact that the defendant has voluntarily compensated the victim's civil claims is regarded as a mitigating circumstance. Moreover, the quality of legal representation may differ between the specialized private lawyer and the defense attorney appointed by the court to those unable to pay. Wealthier defendants have higher ability to compensate victims and to afford better lawyers. This paper aims to provide a method on how to capture the extent of disparity that is contributed by legal channels and, most importantly, the disparity that goes beyond them. It does not attempt to provide normative judgments on the

[2] According to the survey by Levada center in 2014 and 2013: http://www.levada.ru/sbornik-obshhestvennoe-mnenie/.
[3] http://worldjusticeproject.org/rule-of-law-index.

optimality of these legal channels and the inequality they produce. If policy-makers can justify the disparity created by the legal channels, the residual disparity is more controversial.

This study relies on multiple sources of information: the administrative court data on the population of criminal traffic offenses; the texts of court rulings available for a subsample of cases; and the additional sources of information on lawyers and car prices.[4]

I use the multiple-proxy approach by employing GMM/IV regressions to measure the effect of wealth on settlement and incarceration rates.[5] [6] The wealth is proxied by three variables: whether the defendant has a tertiary education degree, whether he is a company executive or a company owner, and the estimated price of his or her car. The car price estimates are available only for a subset of cases, which represents one tenth of the population. So, first I use the restricted sample that has information on car prices and show that the GMM/IV results are robust. However, the results may not be valid for the whole population due to possible sample selection. So I validate the results by using the full sample and just two proxies. The results obtained using the subsample of cases and the results that use the full sample are similar, alleviating the concerns about sample selection.

The results show that defendants in Russia indeed get unequal outcomes in criminal courts. The disparity in the outcomes exist due to the legal channels, but it seems to go beyond them. For instance, the owner of a car at the bottom 5 percent of the distribution of car prices observed in the data (c. 1200 US Dollars) is associated with the settlement rates of 7 percent, while the owner of a car at the top 5 percent of the distribution (c. 24.500 US Dollars) is associated with the settlement rates of 43 percent. Among the non-settled cases – when the judge chooses between probation and incarceration – the former group of defendants are imprisoned in 42 percent of cases, while the latter group of defendants are imprisoned only in 25 percent of

[4] Access to the administrative dataset was provided by the Institute for the Rule of Law at the European University at St. Petersburg.

[5] It is inspired by the work of Black and Smith (2006) and Lubotsky and Wittenberg (2006) that generalize the classical measurement error models.

[6] I would like to thank Dmitriy Skougarevskiy for suggesting to use the multiple proxy approach.

cases. All the comparisons are for the group of traffic accidents with one fatality, caused by a sober driver, first-time offender, without dependents. The gap in the incarceration rates becomes slightly smaller if accounted for the voluntary compensations, but remains substantial. Controlling for the lawyer's experience as a way to capture her quality/specialization, the estimates of the gap barely changes.

In fact, the actual gap may be even larger, considering that the pre-court selection process and in-court settlements may lead to different quality of cases that reach the sentencing stage for different wealth groups. Similar to other empirical papers analyzing court data (see for example Ichino *et al.*, 2003), I develop a theoretical model, specific to the setting, and use it to predict the characteristics of cases that reach the sentencing stage. Wealthier defendants are able to settle for higher severity of cases, so the cases that reach the sentencing stage are more severe on average for the wealthy. The same mechanism is likely to be at play for pre-court selection. In fact, the data seems to confirm the hypothesis based on the observable characteristics of case severity. Hence, the selection bias is likely causing a bias in the estimates, making the gap look smaller than it is.

The residual disparity raises questions about the judicial impartiality, corruption or some other forms of discrimination. Although the data does not give any further indication of what other channels are at play, this information might nevertheless be a good starting point for further investigation. Also, one can trace how this gap changes over time or reacts to the judicial reforms. Moreover, such statistics can be calculated for other countries that give open access to court rulings, which may be used for a cross-country comparison.

This study complements the study by Volkov (2016), who analyzes the judicial disparities in Russia with respect to different demographic characteristics and the social status of offenders. His paper focuses on violent crimes, theft, drugs, and fraud. Volkov finds that judges tend to incarcerate the college-educated less and the unemployed more, which he links to the judges' expectations of the probability of recidivism. At the same time, he finds that judges incarcerate entrepreneurs and top managers more often and for longer duration, which the author links to the judges' bias against people in "the position of trust and authority". This paper differs in that

it does not classify the defendants into different groups based on employment, education, gender, citizenship, etc., but rather focuses on the common underlying factor – wealth – that confounds with all of them. For the questions concerning the effect of wealth, it can be argued that the defendants in traffic accident cases provide a more representative sample of the overall population compared to the defendants who committed felonies with intent. For example, executives who are involved in the latter case are likely to be a very limited and specific sample of the whole population of executives. Also, the current paper provides some intuition behind the process of case selection which may also explain why Volkov (2016) observes harsher outcomes for entrepreneurs and top managers.

This study also complements those by Paneyakh (2016) and Paneyakh (2014), which discuss how the system of performance evaluation create perverse incentives for prosecutors to indict disproportionately people who lack resources to fight against the conviction. According to the author, the same evaluation system makes judges avoid acquittals and use settlements and probation as quasi-acquittals instead. This paper provides some evidence on unequal case selection to court based on the employment status of the defendant. However, I do not find that the defendants with lower socio-economic status are more likely to get the quasi-acquittals, but quite the opposite. Hence, it seems that inequality in indictment is exacerbated further in court.

This paper also contributes to the existing empirical literature on court outcome disparities in criminal justice for different wealth categories (Hartley *et al.* 2010; Rattner *et al.* 2008; Champion 1989) and to the empirical papers on judicial inequality, of which there are many (see for example Alesina and La Ferrara 2014; Shayo and Zussman 2011).

The paper is organized as follows: Section 2 provides information on the judicial system in Russia for criminal traffic offenses, it also gives the description of data sources and samples used in the analysis; Section 3 describes the model of settlements and its implications for the empirical approach; Section 4 provides the setup for the analysis based on the multiple proxies and discusses the results; and Section 5 concludes.

2. JUDICIAL PROCESS AND DATA

2.1 Judicial process for criminal traffic offenses

Russia has a civil law legal system. The core principles of its criminal law are summarized by the Criminal Code that categorizes types of crimes and prescribes possible punishments. According to the Code, the criminal traffic offenses include traffic accidents that have led to severe injuries or death.

The Code categorizes criminal traffic offenses based on the number of fatalities – severe injuries only; one death; multiple fatalities. It also differentiates between sober and drunk offenders. The Code prescribes the upper bound of a prison sentence depending on the offense category: it starts at two years for *severe injuries* + *sober* and rises up to nine years for *multiple fatalities* + *drunk*.[7] Besides the prison sentence, the driver license might be temporarily revoked.[8] Moreover, the judge that rules on the criminal case usually decides simultaneously on the outcome of the civil case: the amount of compensation liable by the defendant to the victim.

Suppose there is a traffic collision where at least one person is severely injured or dies. Before court, this case is processed by an investigator in the police department who collects and analyze evidence, after which she transfers all the materials to the prosecutor. Based on the evidence, the prosecutor decides whether, and whom, to indict. Then the case is transferred to the court.[9] Given that the acquittals are very rare in Russian courts, the indictment in most cases is equivalent to being found guilty, whereas court's task reduces to the decision on the type and duration of the punishment (Volkov 2016; Shklyaruk 2014; Trochev 2014).

Importantly, the cases that reach court can be already subject to non-random selection. According to Paneyakh (2014), the investigators and prosecutors aim at minimizing the probability of case acquittals in court, since every acquittal results in

[7] In legal practice, the offense categories are referenced by numerals, from 1 to 6. In this paper, I reference them by the combination they represent, e.g. offense type 1 is *severe injuries* + *sober*.
[8] In case of real imprisonment, the license revocation starts after the date of the release from prison.
[9] If a deceased person was indicted, this cases are likely to be closed before reaching the court.

a written reprimand to the prosecutor. So they select "easy" cases where the culpability of the potential defendant is obvious or the defendant is not expected to be able to fight against the conviction. According to the author, this creates unequal selection of cases where those who have resources to fight are less likely to get indicted.

In the case of traffic offenses, one may expect that the degree of selection issues should be less pronounced: after all, they involve serious harm. Nevertheless, the data shows some suspicious patterns of case selection into court. Table 1 shows that for sober drivers there are usually around 23 cases with severe injuries for every multiple-fatality case. However, for executives the ratio drops to 18 and for the law enforcement officials to 10.[10] Either they are predisposed to cause much larger harm when they get into accidents, or this is the result of asymmetric selection into court, where these groups are more likely to avoid indictment for less severe crimes. Notice, that law enforcement officials might not be wealthier than the average citizen, but they possess other kind of resources: higher ability to exert informal influence. The ratio of single deaths to multiple fatalities shows a similar, but less pronounced pattern. Overall, it seems that the asymmetric attrition is more likely to happen when the severity of the offense is lower

TABLE 1 • NUMBER OF CASES BY TYPE OF HARM AND JOB CATEGORY OF THE DEFENDANT, AS A MULTIPLE OF TO THE NUMBER OF CASES WITH MULTIPLE FATALITIES WITHIN EACH JOB CATEGORY

Job category	Severe injuries	One death	Multiple fatalities
Not working	24.65	8.82	1
Worker	23.28	9.29	1
Office worker	22.73	8.32	1
Executive	18.25	7.06	1
Law enforcer	10.41	5.81	1

Note: Only the cases where the defendant was sober.

[10] The definition of the law enforcement official is the same as in Volkov (2016).

For certain crimes, including traffic offenses, the court has the right to stop the criminal prosecution if the defendant and the victim reached reconciliation. Since the reconciliation assumes monetary compensation of the victim, this is equivalent to a settlement in a civil litigation. The victim-defendant settlement in the criminal case involves the following actions: (1) the defendant completely compensates the victim's moral damages, (2) the victim forgives the offender and officially asks in written form the court to stop criminal prosecution, (3) the judge allows the criminal charges to be dropped and waives any punishment for the defendant.[11] In case of the settlement, the victim cannot make any compensation claims for moral damages against the defendant afterwards.[12] The compensation of medical expenses and material damage, however, is a separate civil suit, which usually involves the insurance company.

TABLE 2 • TYPES OF OUTCOMES FOR THE DEFENDANT IN A CRIMINAL TRAFFIC OFFENSE

	Found guilty	Criminal record	Incarcerated	Compensation	License
Acquittal	No	No	No	No	No
Settlement	No	No	No	Yes	No
Prison sentence:					
Probation	Yes	Yes	No	Yes[B]	Yes/No[C]
Real imprisonment	Yes	Yes	Yes	Yes[B]	Yes/No[C]

[A] *The voluntary compensation of the victim's moral damages is a pre-requisite for settlement;*
[B] *The judge rules on the amount of compensation if defendant disagrees with the victim's demands;*
[C] *On the judge's discretion.*

[11] If the initial victim dies, close relatives are recognized as victims.
[12] Although the investigator with the permission of prosecutor also has the right to drop the case due to the settlement, the official statistics suggests it happens not so often.

TABLE 3 • THE CRIMINAL CODE'S CLASSIFICATION OF TRAFFIC OFFENSES AND THE SUMMARY STATISTICS OF OUTCOMES

The Criminal Code classification				Summary statistics			
#	Harm	Under influence?	Max prison (years)	Settled (%)	Probation (%)	Incarcerated (%)	Incarcerated (mean, years)
1	Severe injuries	Sober	2	43	30	3	1.4
2	Severe injuries	Drunk	3	22	50	26	1.9
3	Single fatality	Sober	3	23	44	31	2.3
4	Single fatality	Drunk	7	5	23	70	3.1
5	Multiple fatalities	Sober	7	7	29	62	3.5
6	Multiple fatalities	Drunk	9	2	6	91	5.0

Note: The summary statistics is based on the official database of the defendants' statistical cards; averaged over 2009-2013 years.

In court, if no settlement is reached, the judge acquits less than 0.5 percent of cases. The rest get a prison term, but the judge may decide to suspend the sentence, except for repeat offenders. If the defendant compensates voluntarily the claims asked by the victim, this is considered to be a mitigating circumstance. The potential outcomes for the defendant are summarized in Table 2. Table 3 classifies the Criminal traffic accidents and provides descriptive statistics on the frequency of real imprisonment, suspended sentences and in-court settlements.

2.2 Data

The paper uses multiple sources of information. The two main sources are the database of the statistical cards on defendants and the collection of the texts of court rulings. The defendants' statistical cards contain information about each criminal case: the name of the judge, court, region, the judge's ruling, the defendant's demographic characteristics and prior criminal records, and other data. Most importantly, it provides information about the educational and employment status of defendants. The texts of court rulings additionally provide information on the lawyer's name, the car model driven by the defendant, and the presence of a voluntarily compensation.

The database of statistical cards contains the universe of the criminal traffic offenses for the period from January 2009 to December 2013. The database is unified by the Judicial Department at the Supreme Court of the Russian Federation, which requires courts to collect statistical information about each case in a standardized manner.[13,14] This is more than forty-six thousand cases, excluding the *severe injuries* + *sober* offense category.[15] Unlike other criminal traffic offenses which are punished by incarceration, the *severe injuries* + *sober* offense category assumes milder punishment types such as fines and the deprivation of freedom with no isolation from society. These defendants are rarely incarcerated, so this offense category is excluded from the analysis.

The database of the rulings texts consists of around twenty-five thousand cases for the period from June 2010 to the end of 2013. The texts of court rulings were found online.[16] The Law "On ensuring of access to courts' activity in the Russian Federation" obliged all courts to post the texts of all public rulings online. The law became effective in June 2010. In reality, compliance with the law was not perfect

[13] See their website: http://www.cdep.ru/index.php?id=70.
[14] The access to the anonymized database was kindly provided by the Institute for the Rule of Law at the European University at Saint-Petersburg: http://enforce.spb.ru/en/.
[15] Except for some omissions and errors, for detailed explanations see Volkov (2016).
[16] Accessed through the websites that aggregate the texts of rulings that have been posted by courts: primarily rospravosudie.com, but also gcourts.ru, sudebniyeresheniya.rf.

and varied from court to court (Pozdnyakov 2013a; Pozdnyakov 2013b). The ratio of number of texts to statistical cards is around 70 percent for June 2010 to December 2013.

The legislators provided contradicting and vague directives to courts on what should be considered sensitive information, so the courts sanitized the texts to varying degree before posting them. They left the lawyers' names in most of the cases. They often stripped monetary values, including the amount of compensations, but one can always tell whether voluntary compensation took place or not. They often cut the information on the defendants' cars. Only one quarter of the cases with the information on cars (excluding trucks, buses, motorbikes, etc.) have complete information on the brands and models. This raises concerns of non-random sample selection. I assess this issue in Section 2.4, by looking at the summary statistics, and in Section 4.4, by checking the robustness of results in the full sample, which I find to be robust.

The statistical cards and the texts of court rulings were merged based on the combination of common variables: the registry number of the case, the court name, the judge's name, case outcomes, etc. After the merge, one half of the statistical cards for the period from June 2010 to the end of 2013 have been successfully linked to their texts.

I collected car prices for different car models in October 2014 from the secondary market car sales website *auto.ru*. Based on the three hundred most recent advertisements at the time of collection, I calculate the average of the offer prices for each car model, which I call the estimated car price. For the cars belonging to companies or for the expensive cars driven by chauffeurs, I consider the price to be missing. The resulting sample that has the information on car prices represents one tenth of the population of cases.

I used other texts of court rulings to find additional information for each lawyer that appear in the sample. For each lawyer, I count the average annual caseload of criminal cases and criminal traffic offenses. This additional information is available for 80 percent of cases from the subsample with the car prices. Table 4 provides the descriptive statistics for each sample for the period from 2010 to 2013.

TABLE 4 • THE DESCRIPTIVE STATISTICS FOR 2010-2013 BY SAMPLES

Statistics	Population	Sample with car prices	Sample with car prices and lawyers
N. obs.	38910	3976	3239
N. courts	2359	1157	1041
N. courts by traffic offense case load:[A]			
(1,32]	1295	464	397
(32,61]	523	314	288
(61,108]	333	240	225
(108,302]	181	139	131
Type of offense: (sample share, %)			
severe injuries + drunk	28.7	26.1	27.1
one fatality + sober	46	47.3	45.5
one fatality + drunk	17.6	19.3	20.2
mult. fatalities + sober	5.5	4.9	4.8
mult. fatalities + drunk	2.2	2.4	2.4
Year: (sample share, %)			
2010	22.8	11.4	11.2
2011	24.2	26.3	26.5
2012	27.1	31.5	32.4
2013	25.8	30.8	29.9
Trial outcome: (sample share, %)			
settled	18.9	16.5	15.6
incarcerated	38	37.3	37.7
suspended	41.1	44.6	45.1
First-time offenders (sample share, %)	83.9	86.2	85.5
Education: (sample share, %)			
college	19.2	21.2	20.3
vocational	36.1	36.5	36.8
Occupational status: (sample share, %)			
executive	3.3	3.2	2.8
office or civil worker	7.5	7.4	7.4
manual worker	38.8	36.6	37.9
no employment	38.4	40	39.6
Males (sample share, %)	94.7	93.9	93.9
Marital status: (sample share, %)			
single	44.2	45	45.3
married	45.6	44.7	44.1
Age:			
age, mean	33.5	33.1	33
age, median	30	30	30
age, st. dev.	11.6	11.6	11.5

A The case load is calculated as the sum of all criminal traffic offences in a given court during 2009 to 2013, including the category of 'severe injuries + sober'.

2.3 Proxies of wealth

The executive status, college degree, and car prices are used as proxies to wealth. The executive status dummy equals one if the defendant is the owner of a firm or the chief executive of a commercial organization. The share of executives in the samples is around 3 percent. The college degree dummy equals one if the defendant has complete or incomplete college degree. The share of college graduates is around 19 percent in the population and around 21 percent in the restricted sample. The average of the estimated car prices is 291 thousand Rubles, which was equal to 7.3 thousand US Dollars at the time of data collection.[17] The median of the prices is twice lower than the average. Table 5 reports the percentiles of the distribution of the estimated car prices.

Table 5 • The distribution of the estimated car prices in the court rulings

Quantiles	5%	25%	50%	75%	90%	95%	99%
Price (in RUB mln)	0.048	0.075	0.147	0.382	0.584	0.959	1.907

Based on 7052 observations (the texts of court rulings)

The three proxies positively correlate with each other (see the matrix of correlations in Table 6). Figure 1 compares the distribution of the estimated car prices conditional on the executive status or college degree. The distribution shifts to the right both for the college graduates and for the executives, however there is no perfect sorting. The degree of correlations is relatively weak, since each variable is only a noisy signal of wealth. The estimated car price is likely to be the noisiest measure since it is itself a noisy measure of the true car price.

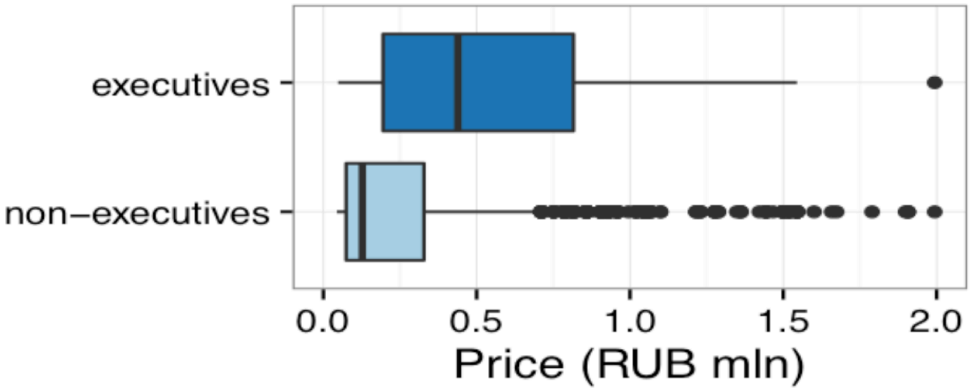
[17] The average exchange rate for the period from September 16, 2014, to October 15, 2014, was 39.22 Rubles per USD. Source: Russian Central Bank; www.cbr.ru/currency_base.

TABLE 6 • THE MATRIX OF CORRELATIONS

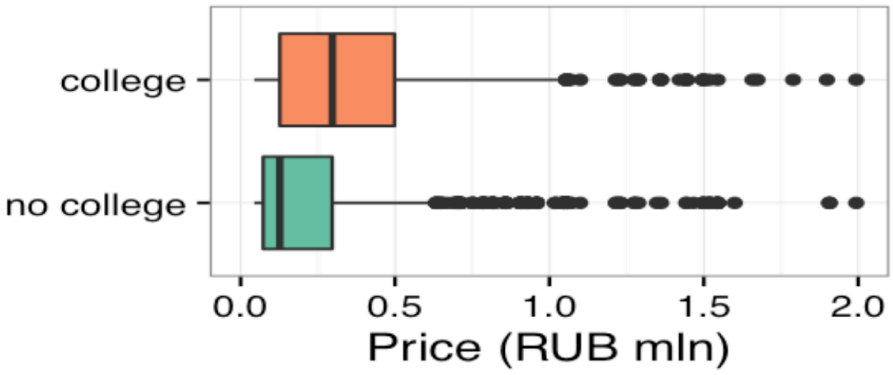
	car price	executive	college
car price	1		
executive	0.1729	1	
college	0.1916	0.1798	1

Based on 3987 observations (only those cases that have information on all three variables).

FIGURE 1 • BOX PLOTS FOR CAR PRICE DISTRIBUTIONS BY:



A) THE EXECUTIVE STATUS



b) The college degree

2.4 The restricted sample

The analysis starts with the restricted sample that contains both the estimated car prices and the information from the statistical cards. The restricted sample captures only one tenth of all observations, so the courts that have higher caseload are more likely to remain in the sample. Table 4 shows that at the lowest quartile of the caseload the restricted sample loses two thirds of the courts. At the highest quartile the restricted sample loses only one fifth of the courts. According to the table, all other characteristics do not differ dramatically across the samples.

The estimation based on the restricted sample might be not valid for the whole population if the low caseload courts are substantially different from the high caseload courts in the parameter of interest – the incarceration rate disparity due to wealth. Later, Section 4.4 checks whether the estimates for the restricted sample are the same for the population and shows that they are quite similar.

3. SELECTION INTO THE SENTENCING STAGE

Since incarceration rates are calculated using the non-settled cases only, it is important to understand the selection of cases into the settlement. I argue that if I could account for the possible bias this would make the estimates for the gap in incarceration rates even larger, reinforcing the findings. Ever since Priest and Klein, 1984, the mechanisms of the selection of cases into court and the understanding of the direction of bias have been the important part of the judicial research (e.g., Waldfogel 1998; Eisenberg and Faber 1996; Shavell 1996; Waldfogel 1993). Similar to Ichino *et al.* 2003, I develop a theoretical model, specific to the setting, and use it to predict the characteristics of cases that reach the sentencing stage. I assume that the probability of settlement declines with the severity of offense.

This prediction is consistent with the data. The settlement rate drops with the number of victims and with culpability of the defendant: for example, drunk drivers

are less likely to get settlement (see the settlement rates summary statistics in Table 2). Since wealthier defendants are able to settle more often, all other things being equal the severity of cases that reach the sentencing stage is expected to be higher on average for the wealthy. Some aspects of the severity of the case can be observed in the data, but some cannot. Hence, the unobserved component will bias the estimates. If judges cared about the severity of offense only, they would incarcerate the wealthy more frequently. If the data shows that the wealthy are incarcerated less often, it must be due to indeed more lenient outcomes prescribed by the judges. Moreover, if one accounts also for a similar pattern of case selection into court, the estimate of the gap in incarceration rates is likely to increase even further.

3.1 The basic model

This section presents the model that helps to understand the direction of the estimator's bias that may arise due to settlements. The model solves optimal decisions of the defendant and the victim right after the indictment but before the court decision.

The settlement model is a perfect information sequential game, described by the game tree in Figure 2. There are two risk-neutral rational players: the defendant and the victim. The defendant is initially endowed with wealth w, the victim – with zero. The defendant moves first and chooses the amount of compensation C to offer, subject to the initial endowment constraint: $C \in [0, w]$. The victim moves second, observes C and decides whether to *Accept* or *Decline* the offer. If the victim accepts, the game ends with the payoff of $w - C$ for the defendant and C for the victim. If the victim declines, the game reaches lottery called Court.

The lottery portrays the ex-ante uncertainty faced by the players. Outcomes of the lottery are *Prison* with the probability p and *Suspend* with the probability $1 - p$.[18] In the case of *Suspend*, the payoffs are $(w, 0)$, i.e., every player remains with the initial

[18] Since in-court acquittals are almost non-existent, the defendant's prospects in court are limited to real imprisonment or suspended sentence.

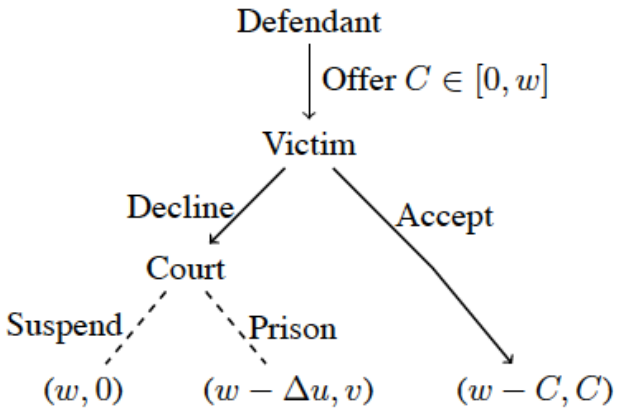
endowment. In the case of *Prison*, the defendant suffers the disutility from going to prison □, while the victim gets vengeance satisfaction $v > 0$.[19]

The defendant first solves the victim's optimal strategy. The victim maximizes her utility given the compensation offer: $\max\{C, pv\}$. The victim accepts the offer if $C \geq pv$. The defendant maximizes $\max\{w - C, w - pd\}$ subject to the following constraints: $C \leq w$ and $C \geq pv$. If $pv > w$, Defendant offers zero. If $pv \leq w$, Defendant finds it optimal to settle only if $d \geq v$.

Thus, the settlement takes place ($S = 1$) if the settlement is feasible ($pv \leq w$) and the settlement is a better alternative to the lottery ($pv \leq pd$).

$$\begin{cases} 1 & \text{if } v \leq \min\left\{\frac{w}{p}, d\right\} \\ 0 & \text{Otherwise.} \end{cases} \tag{1}$$

FIGURE 2 • GAME TREE



3.2 The selection bias

Assume that the probability of incarceration, p, and the vengeance utility of the victim, v, increase with the gravity of offense, as measured by the indices b and g.

[19] Notice that since the defendant is not acquitted, he will be liable to compensate the victim. In the game, the victim's endowment of zero already includes the expected compensation, and C offered by the defendant is a surplus over the expected compensation he is ready to pay to settle.

The index b is correlated with w, while g is distributed independently of b and w. For example b includes defendant's reckless behavior at the time of the accident, while g includes behavior and personal characteristics of the victim, which are random ex-ante, but legally relevant ex-post.[20] Moreover, d and p depend on w, assuming that the wealthier prisoners experience a greater decline in utility and that wealth may improve chances of the defendant in court. The assumptions on the functional forms of p, v, and d and their partial derivatives are presented in Table 7 below.

TABLE 7 • ASSUMPTIONS ON FUNCTIONAL FORMS

Function $f(.)$	$\partial f(.)/\partial b$	$\partial f(.)/\partial g$	$\partial f(.)/\partial w$
$\overline{p}(b,g,w)$	≥ 0	≥ 0	≤ 0
$\overline{v}(b,g)$	≥ 0	≥ 0	
$\overline{d}(w)$			≥ 0

Restating the settlement condition 1, the settlement takes place if:

$$v(b,g) \leq min\{\frac{w}{\overline{p}(b,g,w)}, \overline{d}(w)\} \quad (2)$$

The left-hand side of the inequality is increasing with g, while the right-hand side is decreasing with g. So there exists a threshold $\tau(b,w)$, such that the two sides are equal. Then the settlement condition can be rewritten as in the system of equations 3.

[20] For instance, the death of a child is likely to be perceived by the judge as a graver incident than a similar offense with an adult victim. Glaeser and Sacerdote (2013) show in their study of the criminal traffic offenses in Alabama, U.S., that the characteristics of the victim such as race and criminal history tend to affect judicial sentencing, despite the random nature of the victim-offender match.

$$S = \begin{cases} 1 & \text{if } g \leq \tau(b,w); \\ 0 & \text{Otherwise} \end{cases} \tag{3}$$

$$\tau(b,\mathrm{w}): \quad v(b, g = \tau) = \min\{\frac{w}{\overline{p}(b, g = \tau, w)}, \overline{d}(w)\} \tag{4}$$

Suppose that there is a continuum of cases that differ in b, g and w. The cumulative density of g is denoted as $F^g()$ and by definition $Pr(g \leq a \mid b,w) = Pr(g \leq a) = F^g(a)\ \forall a,b,w$. Settlements truncate the density of g, and the new density of the non-settled cases is represented as:

$$\begin{aligned} Pr(g < a \mid b,w,S=0) &= Pr(g < a \mid g \geq \tau(b,w)) \\ &= \max\left\{\frac{\int_{\tau(b,w)}^{a} d\,F^g(x)}{\int_{\tau(b,w)}^{\infty} d\,F^g(x)}, 0\right\} \\ &\equiv \Theta^{NS}(a \mid b,w) \end{aligned} \tag{5}$$

Notice that for any two levels of wealth, $w' > w$, the function $\Theta^{NS}(b,w')$ first order stochastically dominates $\Theta^{NS}(b,w)$. Hence, the conditional expected difference in the gravity of offense is positive (see Equation 6). This is the selection bias.

$$\begin{aligned} E(g \mid b,w',S=0) &- E(g \mid b,w,S=0) \\ &= \int_{-\infty}^{+\infty} x\, d\Theta^{NS}(x \mid b,w') - \int_{-\infty}^{+\infty} x\, d\Theta^{NS}(x \mid b,w) \geq 0 \end{aligned} \tag{6}$$

I assume that the true functional form of the probability of prison $p(g,w,b)$ can be linearly approximated with parameters α and ξ as in equation 7. If b is observed, while g is not, I can run the regression 8 on the sample of the non-settled cases ($S = 0$).

$$p_i(w_i,b_i,g_i) = \alpha_0 + \alpha w_i + \xi_b b_i + \xi_g g_i + \varepsilon_i \tag{7}$$

$$p_i(w_i,b_i,S_i=0) = \beta_0 + \beta w_i + \gamma_b b_i + \epsilon_i \tag{8}$$

Then, the parameter β will be identified as the difference in the conditional expectations of p for the discrete change in the level of wealth from $w = 0$ to $w = 1$ averaged across all b (Equation 9). The parameter β will capture α, the true effect of wealth, plus some extra term, which is the positive selection bias ($\xi_g > 0$ is assumed).

$$\begin{aligned}\beta &= \int_{-\infty}^{+\infty} (\mathbb{E}(p \mid b, w = 1; S = 0) - \mathbb{E}(p \mid b, w = 0; S = 0))db \\ &= \alpha + \xi_g \int_{-\infty}^{+\infty} (\mathbb{E}(g|g \geq \tau(b, w = 1) - \mathbb{E}(g|g \geq \tau(b, w = 0))db\end{aligned} \tag{9}$$

PROPOSITION 3.1: β is an upper bound of α, i.e., $\beta \geq \alpha$. If β is below zero, then it must be that α is also below zero, i.e., judges incarcerate wealthier defendants with a lower frequency.

The next section will be devoted to estimating the disparities in settlements rates and in the incarceration rates for the non-settled cases. For the latter, the estimated β will be a conservative estimate for the true effect α.

4. RESULTS

4.1 The econometric model with multiple proxies

Suppose that the dependence of incarceration rates or settlements rates on wealth, as well as the relation between wealth and their proxies, can be approximated with the following non-causal linear model:

$$y = \beta x^* + u \tag{10a}$$

$$x_1 = x^* + \epsilon_1 \tag{10b}$$

$$x_2 = \rho_2 x^* + \epsilon_2 \tag{10c}$$

$$x_3 = \rho_3 x^* + \epsilon_3 \tag{10d}$$

where y can be either the dummy for settlements or the dummy for incarceration; $x^\star$ is the true measure of wealth, which is unobserved to the econometrician; x_1, x_2,

and x_3 are the noisy measures of wealth. In this paper, x_1 is the car price in millions of Rubles, x_2 is the dummy for executive status, and x_3 is the dummy for college degree.

As in Lubotsky and Wittenberg (2006), I normalize the scale of wealth to the scale of the first proxy, i.e., $\rho_1 = 1$. This normalization is harmless, since the true scale of wealth is non-identifiable in the given framework. The normalization to the car price should be understood in the following way: we do not know how much wealth levels are different between someone whose car costs $Value_1$ and someone who owns $Value_2$ car, but we can estimate the effect associated with that increase in wealth.

In the presence of other covariates, Z, (e.g. the constant, the type of offense, etc.), the variables y, x_1, x_2, and x_3 in equations 10a-10d should be substituted by their Yulized residuals – the residuals obtained after regressing each variable on Z. Black and Smith (2006) argue that the econometrician must be careful of what is included in Z. The inclusion of other covariates that also correlate with wealth will strip the Yulized proxies from their signaling power. This is why I am reluctant to add the rich set of demographic controls into Z, leaving only those that are legally relevant.

The measurement errors make the OLS estimators biased downward in the magnitude. The attenuation bias can be remedied by the instrumental variable approach, but additional assumptions are required (see more at Browning and Crossley 2009; Black and Smith 2006; Griliches 1986). The regression of y on x_1 using x_j as an instrument, where $j \in \{2,3\}$, would provide the estimator, which in probability converges to $\beta_{1,j}^{IV}$ (eq. 11). If the two measurement errors are uncorrelated, $cov(\epsilon_j, \epsilon_1) = 0$, and the measurement error of the j-th proxy does not affect y, $cov(\epsilon_j, u) = 0$, then $\beta_{1,j}^{IV}$ will be equal to the parameter of interest, β. If the above mentioned assumptions are true both for $j = 2$ and $j = 3$, then β can be efficiently estimated by the General Method of Moments estimator, whereas the validity of the instruments can be tested since the model is overidentified (see Wooldridge 2010, on the GMM estimator).

$$\beta_{1,j}^{IV} = \frac{cov(y, x_j)}{cov(x_1, x_j)} = \frac{\rho_j \beta \sigma_x^2 + cov(\epsilon_j, u)}{\rho_j \sigma_x^2 + cov(\epsilon_j, \epsilon_1)} \quad (11)$$

4.2 The GMM results

Table 8 shows the estimates of $\hat{\beta}^{GMM}$. The additional regressors, Z, control for the offense type (see Table 2), criminal priors, other simultaneous crimes, and dependents. The optimal weighting matrix takes into account the possibility of clustering at the regional level. All estimates are statistically significant at 95 percent confidence level. According to the first and second columns, the one million Rubles increase in the car price is associated with the forty-nine percentage points rise in the settlement rate, and for the non-settled cases it is associated with the thirty-two percentage points drop in the incarceration rate. The intercepts are at 5 percent and 50 percent, respectively. These figures do not control for the presence of voluntary compensations. If the regression controls for it, the size of the associated effect decreases in the magnitude, but only insignificantly (see column 3). Remember that the estimates of $\hat{\beta}$ for the probability of incarceration are based on the sample of the non-settled cases and are likely to be affected by the selection bias. According to Proposition 3.1 the estimated parameter is the upper bound for the true parameter, so the real gap in the incarceration rates can be even larger.

Moreover, I recalculate the incarceration probabilities by adding the variables that associate with the quality of the lawyer – their average caseload per year, its square, and the yearly caseload of the criminal traffic offenses. The sample shrinks since not all observations have this additional information on lawyers. In order to make a meaningful comparison, I reestimate the previous regression – with no controls for the lawyer quality – on the smaller sample; the results are reported in column 4. The movement from the previous sample, column 3, to the more constrained one, column 4, does not affect the intercept, but the point estimate of $\hat{\beta}^{GMM}$ increases in magnitude (not significantly). After adding the controls for the lawyer's caseload (column 5), the estimate of $\hat{\beta}^{GMM}$ does not change much, but the intercept grows: now the intercept captures the defendant with a completely inexperienced lawyer (as

opposed to the average caseload lawyer). Having an inexperienced lawyer is associated with higher incarceration rates.

Why do not the additional controls for the lawyers experience reduce the magnitude of the estimate of $\hat{\beta}$? One would expect to see at a decrease in the magnitude of the wealth effect if the quality of the lawyer is one of the channels of the disparity. Nevertheless, the magnitude seems to go into the opposite direction. It might be so because the wealthier defendants choose experienced lawyers when their cases are relatively easier compared to the less wealthy who choose the experienced lawyers only when their prospects are especially grim. Then, when the controls for the lawyer's experience are added, the change in $\hat{\beta}^{GMM}$ captures two effects: the decrease in the wealth effect and the change in the quality of the cases – selection bias. Alternatively, it could be that the lawyer's caseload does not capture the true quality of her work. The regression misses the information on whether the attorney is hired or appointed by court. The same experienced lawyer may put much more effort when he is hired as opposed to being appointed.[21] Also, the lawyer's quality might not be an important channel.

[21] More on the role of defense attorney in Russia see Moiseeva (2016).

TABLE 8 • THE GMM ESTIMATES OF THE EFFECT OF WEALTH ON THE SETTLEMENT RATE AND THE INCARCERATION RATE

	Pr(settlement)	Pr(incarceration)			
	(1)	(2)	(3)	(4)	(5)
$\hat{\beta}^{GMM}$	0.489 (0.086)	-0.325 (0.115)	-0.266 (0.113)	-0.324 (0.121)	-0.367 (0.129)
Intercept	0.048 (0.030)	0.499 (0.041)	0.541 (0.040)	0.549 (0.045)	0.611 (0.050)
F statistic for weak identification	55.5	45.1	45.7	43.9	40.6
Test underidentification, p-value	0.0000	0.0000	0.0000	0.0000	0.0000
Test overidint. restr., p-value	0.469	0.678	0.863	0.799	0.965
N. obs.	3986	3329	3329	2736	2736
Sample ID	A	B	B	C	C
Main controls	+	+	+	+	+
Volunt. compensation	-	-	+	+	+
Lawyer's experience	-	-	-	-	+

$\hat{\beta}^{GMM}$ is the feasible GMM estimate for β when x_1 is instrumented and x_2 and x_3 are excluded instruments (See eqns 10a-10d). Samples A, B, and C include the matched dataset of the statistical cards and the court rulings (see Section 2.2). Samples B and C exclude the settled cases. Sample C drops cases with missing information about the lawyers' caseloads. F statistics for weak identification is based on Kleibergen-Paap rk Wald F statistic, robust for clustered correlation. For reference, the Stock and Yogo (2005) critical value for Cragg-Donald F statistic and i.i.d. errors, 10 percent maximal IV size bias, is 19.93 (valid for iid errors). Underidentification test is based on Kleibergen-Paap rk LM statistic, robust for clustered correlation. Test for the overidentifying restrictions is based on Hansen J statistic test; the joint null hypothesis is that the instruments are valid. Main controls: drunk, severe injuries only, multiple fatalities, drunk*multiple fatalities, first-time offender, multiple crimes, the presence of dependents. Lawyer's experience: the average number of cases per year, the square of the average number of cases per year, the average number of traffic offense cases per year. All regressions include regional fixed effects. Standard errors are in parenthesis, clustered at the regional level. Number of clusters = 80.

4.3 Robustness

The test for the overidentifying restrictions fails to reject the null hypothesis that the two instruments are jointly valid (Table 8). In fact, the just-identified IV estimation with either of the two instruments gives similar results (compare $\hat{\beta}^{IV}_{1,2}$ and $\hat{\beta}^{IV}_{1,3}$ in Table 9). The standard errors of the IV estimates are substantially larger, which

makes $\hat{\beta}_{1,3}^{IV}$ for the incarceration probability insignificant at 95 percent confidence level.

The results of Table 8 are replicated using the continuously updated GMM estimator (CUE), proposed by Hansen et al. (1996) and the limited information maximum likelihood estimator (LIML). The CUE estimator is the GMM estimator with the weighting matrix being continuously updated with each estimation of $\hat{\beta}^{GMM}$ until convergence. Hansen et al. (1996) show that this estimator is robust to the small sample biases and more reliable for testing the overidentifying assumptions. CUE estimator takes into account the clustered structure of errors, while LIML estimator assumes homoscedasticity. Both are robust to the small sample biases of IV estimates (see discussion in Angrist and Pischke 2008). Both estimators give results almost identical to the one estimated by GMM with optimal weighting matrix. All estimates are reported in Table 9.

TABLE 9 • THE GMM ESTIMATES OF THE EFFECT OF WEALTH ON THE SETTLEMENT RATE AND THE INCARCERATION RATE

	Pr(settlement)	Pr(incarceration)		
	(1)	(2)	(3)	(4)
$\hat{\beta}^{GMM}$	0.489 (0.086)	-0.325 (0.115)	-0.266 (0.113)	-0.367 (0.129)
F statistic for weak identification	55.5	45.1	45.7	40.6
$\hat{\beta}^{CUE}$	0.491 (0.086)	-0.325 (0.115)	-0.266 (0.113)	-0.367 (0.129)
F statistic for weak identification	55.5	45.1	45.7	40.6
$\hat{\beta}^{LIML}$	0.493 (0.087)	-0.313 (0.118)	-0.261 (0.117)	-0.365 (0.141)
F statistic for weak identification	55.5	45.1	45.7	40.6
$\hat{\beta}^{IV}_{1,2}$	0.430 (0.115)	-0.356 (0.140)	-0.278 (0.133)	-0.371 (0.148)
F statistic for weak identification	41.7	24.9	25.2	17
$\hat{\beta}^{IV}_{1,3}$	0.557 (0.132)	-0.268 (0.177)	-0.243 (0.172)	-0.360 (0.209)
F statistic for weak identification	98.5	77.1	25.2	60
N. obs.	3986	3329	3329	2736
Sample ID	A	B	B	C
Main controls	+	+	+	+
Volunt. compensation	-	-	+	+
Lawyer's experience	-	-	-	+

$\hat{\beta}^{GMM}$ is the feasible GMM estimate for β when $x1$ is instrumented and $x2$ and $x3$ are excluded instruments (See eqns 10a-10d). $\hat{\beta}^{CUE}$ is the continuously updated GMM estimator of Hansen et al., 1996. $\hat{\beta}^{LIML}$ is the limited information ML estimator. $\hat{\beta}^{IV}_{1,j}$ is the IV estimate with x_1 is an endogenous variable and x_j is the excluded instrument, where $j \in \{2,3\}$. The Stock and Yogo (2005) critical value for Cragg-Donald F statistic and i.i.d. errors, 10 percent maximal IV size bias, is 19.93 for $\hat{\beta}^{GMM}$ and $\hat{\beta}^{CUE}$ and 16.38 for $\hat{\beta}^{IV}_{1,j}$. All regressions include regional fixed effects. Standard errors are in parenthesis, clustered at the regional level. Number of clusters = 80. For more information on the definitions of controls, samples, and F statistic, please see the footnotes for Table 8

The GMM/IV approach substantially improves the identification with respect to the one-proxy OLS results. Table 10 reports estimates by using each proxy separately in an OLS equation. To make comparison between the three proxies meaningful, I scale the estimates for the second and the third proxies by using their respective $\hat{\rho}$, calculated as:

$$\widetilde{\rho_j} = \frac{cov(x_j, x_k)}{cov(x_1, x_k)} = \frac{\rho_j \rho_k \sigma_x^2 + cov(\epsilon_j, \epsilon_k)}{\rho_k \sigma_x^2 + cov(\epsilon_k, \epsilon_1)}$$

where $j = 2$ and $k = 3$, or $j = 3$ and $k = 2$. Notice, that $\widetilde{\rho}_j$ is a consistent estimator of ρ_j only if $cov(\epsilon_2, \epsilon_3) = 0$, i.e. the measurement errors of the second and the third proxies are uncorrelated.

TABLE 10 • THE OLS ESTIMATES OF THE EFFECT OF WEALTH ON SETTLEMENT AND INCARCERATION RATES.
x_1= CAR PRICE, x_2=EXECUTIVE POSITION, x_3=COLLEGE DEGREE

	Pr(settlement)	Pr(incarceration)		
	(1)	(2)	(3)	(4)
$\hat{\beta}_1^{OLS}$	0.026 (0.018)	0.012 (0.022)	0.019 (0.022)	0.040 (0.026)
N	3987	3330	3330	2738
$\hat{\beta}_2^{OLS} \star \widetilde{\rho}_2$	0.041 (0.006)	-0.019 (0.006)	-0.027 (0.010)	-0.028 (0.010)
N	46087	37434	15430	12931
$\hat{\beta}_3^{OLS} \star \widetilde{\rho}_3$	0.082 (0.007)	-0.036 (0.008)	-0.046 (0.010)	-0.051 (0.012)
N	46087	37434	15430	12931
$\widetilde{\rho}_2$	0.49	0.426	0.426	0.389
$\widetilde{\rho}_3$	1.09	1.048	1.055	1.089
Main controls	+	+	+	+
Volunt. compensation	-	-	+	+
Lawyer's experience	-	-	-	+

$\hat{\beta}_j^{OLS}$ is the OLS estimate for β when y is regressed on x_j separately from other proxies (See eqns 10a-10d). The regressions use maximum observations available for each proxy. $\tilde{\rho}_j$ is estimated as in equation 12. All regressions include regional fixed effects. Standard errors are in parenthesis, clustered at the regional level. Number of clusters = 80. The sample for Pr(incarceration) excludes the settled cases. For more information on the definitions of controls and F statistic, please see the footnotes for Table 8

In general, the OLS estimates have the correct sign and even statistical significance (except for the first proxy and the probability of incarceration), but the magnitude of the effect is much smaller compared to the GMM. This is due to the attenuation bias resulting from the measurement errors. According to the OLS results, the first proxy – car prices – is the noisiest measure out of the three, as expected.

4.4 The IV estimation using the full sample

Assuming that the estimate of ρ_2 is valid for the overall population, I estimate $\hat{\beta}_{2,3}^{IV}$ by instrumenting the executive status by the college degree, i.e., x_2 by x_3. This approach will be less efficient than the GMM, because it uses just one instrument, but if valid, it can gain efficiency and external validity by being able to use all forty-six thousand observations, the whole population.

First of all, I check how the estimator performs on the restricted sample with car prices in order to compare the results with the GMM estimates. Table 11 shows $\hat{\beta}_{2,3}^{IV}$, translated into the same scale as car prices. The standard errors have increased compared to the standard errors of the GMM estimator, especially for the samples with non-settled cases only, but the point estimates are quite close to the GMM estimates. This boosts confidence in applying this estimation approach on the larger sample.

TABLE 11 • THE COMPARISON OF THE GMM ESTIMATES TO THE IV ESTIMATES THAT INSTRUMENT THE EXECUTIVE STATUS BY THE COLLEGE DEGREE.

x_1= CAR PRICE, x_2=EXECUTIVE POSITION, x_3=COLLEGE DEGREE

	Pr(settlement)	Pr(incarceration)		
	(1)	(2)	(3)	(4)
$\hat{\beta}^{GMM}$	0.489 (0.086)	-0.325 (0.115)	-0.266 (0.113)	-0.367 (0.129)
F statistic for weak identification	55.5	45.1	45.7	40.6
$\hat{\beta}^{IV}_{2,3} \star \tilde{\rho}_2$	0.557 (0.106)	-0.268 (0.183)	-0.243 (0.180)	-0.360 (0.215)
F statistic for weak identification	35.9	16.4	16.7	14.1
$\tilde{\rho}_2$	0.49	0.426	0.426	0.389
N. obs.	3986	3329	3329	2736
Sample ID	A	B	B	C
Main controls	+	+	+	+
Volunt. compensation	-	-	+	+
Lawyer's experience	-	-	-	+

$\hat{\beta}^{GMM}$ is the feasible GMM estimate for β when $x1$ is instrumented and $x2$ and $x3$ are excluded instruments (See eqns 10a-10d). $\hat{\beta}^{IV}_{2,3}$ is the IV estimate where x_2 is instrumented by x_3. $\tilde{\rho}_2$ is estimated as in 12. All regressions include regional fixed effects. Standard errors are in parenthesis, clustered at the regional level. Number of clusters = 80. For more information on the definitions of controls, samples, and the critical values for F statistic, please see the footnotes for Tables 8 and 9

Table 12 reports the IV estimates for the maximum available samples. In columns 1 and 2, $\hat{\beta}^{IV}_{2,3}$ is estimated using all the population of cases. The estimates are now more efficient than the GMM thanks to the larger sample size. The results show that the IV estimates for the population are not significantly different from the GMM estimates on the restricted sample in Table 8, however the point estimates based on the full sample have shrunk in magnitude.

Columns 3, 4, and 5 represent results for the probability of incarceration with additional controls, information on which is available only in the texts of court

rulings. Hence, I am unable to calculate the results using the full population of cases, but have to use the dataset of the matched cases. However, this still represents a considerable fivefold increase in the sample size, compared to the sample which was available for the GMM estimates. The results are similar to the GMM estimates.

Thanks to the sample size, now the regressions can be reestimated with the court fixed effects instead of the regional fixed effect. In general, the results are very similar with the estimated gap in incarceration rates slightly increasing in magnitude. This serves as an additional robustness check to the GMM estimates on the restricted sample.

Overall, this exercise suggests that the estimates based on the restricted sample are in general similar to the ones obtained for the overall population, and that the omission of the court fixed effects does not alter results. It also shows that the use of only two proxies on the larger sample helps in gaining efficiency compared to the GMM estimates, but there is a tradeoff: I have to assume that the estimate of ρ_2 is a valid estimate for the population data. Nevertheless, the GMM approach with the three proxies was an important step for testing the validity of the identifying restriction.

5. CONCLUSION

Overall, the empirical results suggest that there is a substantial difference in the court outcomes among different wealth groups in Russia. Unsurprisingly, the wealthier defendants tend to settle much more often. If the defendant at the bottom 5 percent of wealth distribution settles only in 7 percent of the cases, the defendant at the top 5 percent settles in 43 percent of the cases. Wealthier defendants are able to afford higher compensation, which makes victims more willing to accept their offers. The disparities created by the different abilities to afford settlements cannot be given any straightforward normative judgment. From the social welfare point of

view, perhaps, settlements serve important role as a way to improve the wellbeing of victims, decriminalize certain acts, and save resources for the judicial system, but it may well decrease the deterrent effect of criminal prosecution. Whether settlements are efficient or not in this setup should be a subject of another study which goes beyond the aims of this paper.

More striking, however, is the finding that the disparity persists even accounting for the actual settlements, voluntary compensations, and lawyer experience. Among the non-settled cases, when the judge decides whether to suspend the prison term or not, the defendants at the bottom 5 percent of wealth distribution are imprisoned in 42 percent of cases, while the defendants at the top 5 percent are imprisoned only in 25 percent of cases. Keeping in mind that the wealthier group is also expected to be more culpable on average, given the selection bias, the fact that we find this gap at all is telling.

I am cautious to interpret the gap in incarceration rates as the evidence of judicial bias. There may still be some defense-attorney effect, which has not been perfectly captured by the controls. Importantly, the regression misses the information on whether the attorney is hired or appointed by court. The same attorney may exert much less effort when she is appointed by court. So even the lawyer's fixed effects would not help to overcome this difficulty. Nevertheless, it does not undermine the results, if one believes that there should not be any disparity neither due to the judicial bias, nor to the quality of legal counsel (after all, court appoints the attorney as an attempt to smooth the inequality).

The estimates are robust to the choice of instruments. The paper shows that having two proxies for wealth may be already enough to test the disparities and having three proxies is useful for testing the validity of the identifying restrictions. The approach proposed in this paper can be applied to analyze the judicial systems of other countries that provide access to their judicial data.

6. REFERENCES

Alesina A. and La Ferrara E. (2014), *A test of racial bias in capital sentencing*, «The American Economic Review», 104, 11, pp. 3397-3433

Angrist J.D. and Pischke J.-S. (2008), *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton, Princeton University Press

Black D.A. and Smith J.A. (2006), *Estimating the returns to college quality with multiple proxies for quality*, «Journal of labor Economics», 24, 3, pp. 701

Browning M. and Crossley T. (2009), *Are two cheap, noisy measures better than one expensive, accurate one?*, «The American Economic Review», 99, 2, pp. 99-103

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

Eisenberg T. and Farber H.S. (1996), *The Litigious Plaintiff Hypothesis: Case Selection and Resolution*, Tech. rep. National Bureau of Economic Research

Glaeser E.L. and Sacerdote B. (2003), *Sentencing in homicide cases and the role of vengeance*, «The Journal of Legal Studies», 32, 2, pp. 363-382

Griliches Z. (1986), *Economic data issues*, «Handbook of econometrics», 3, pp. 1465-1514

Hansen L.P., Heaton J. and Yaron A. (1996), *Finite-sample properties of some alternative GMM estimators*, «Journal of Business & Economic Statistics», 14, 3, pp. 262-280

Hartley R.D., Miller H.V. and Spohn C. (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., Polo M. and Rettore E. (2003), *Are judges biased by labor market conditions?*, «European Economic Review», 47, 5, pp. 913-944

Lubotsky D. and Wittenberg M. (2006), Interpretation of regressions with multiple proxies, «The Review of Economics and Statistics», 88, 3, pp. 549-562

Moiseeva E. (2016), *Plea bargaining in Russia: the role of defence attorneys and the problem of asymmetry*, «International Journal of Comparative and Applied Criminal Justice», pp. 1-22

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

– (2016), *The practical logic of judicial decision making: Discretion under pressure and compromises at the expense of the defendant*, «Russian Politics & Law», 54, 2-3, pp. 138-163

Pozdnyakov M. (2013a), *O neobhodimosti reformirovanija Federal'nogo zakona ot 22.12.2008 N262-FZ «Ob obespechenii dostupa k informacii o dejatel'nosti sudov v Rossijskoj Federacii» [The necessity of reforming of the Federal Law "On ensuring of access to the court activity in the Russian Federation]*, Tech. rep. preprint IRL-01/2013, The European University at Saint-Petersburg, the Institute for the Rule of Law

– (2013b), *Prakticheskaja realizacija principa otkrytosti pravosudija v Rossijskoj Federacii [The Practical Implementation of the Principle of Open Justice in the Russian Federation]*, SPB: Institute for the Rule of Law

Priest G.L. and Klein B. (1984), *The selection of disputes for litigation*, «The Journal of Legal Studies», 13, 1, pp. 1-55

Rattner A., Turjeman H. and Fishman G. (2008), *Public versus private defense: Can money buy justice?*, «Journal of Criminal Justice», 36, 1, pp. 43-49

Shavell S. (1996), *Any frequency of plaintiff victory at trial is possible*, «J. Legal Stud.», 25, pp. 493-501

Shayo M. and Zussman A. (2011), *Judicial ingroup bias in the shadow of terrorism*, «Quarterly Journal of Economics», 126, pp. 1447-1484

Shklyaruk M. (2014), *Traektoria ugolovnogo dela v oficial'noi statistike [The criminal case trajectory in official statistics]*, SPB: Institute for the Rule of Law

Stock J.H. and Yogo M. (2005), *Testing for weak instruments in linear IV regression*, Identification and inference for econometric models: Essays in honor of Thomas Rothenberg

Trochev A. (2014), *How judges arrest and acquit*, «Historical Legacies of Communism in Russia and Eastern Europe», p. 152

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

Waldfogel J. (1993), *Criminal sentences as endogenous taxes: Are they just or efficient*, «JL &Econ», 36, pp. 139-151

– (1998), *Reconciling asymmetric information and divergent expectations theories of litigation*, Tech. rep. National Bureau of Economic Research

Wooldridge J.M. (2010), *Econometric Analysis of Cross Section and Panel Data*, Cambridge (Ma), The MIT Press