THE GRADUATE INSTITUTE GENEVA

INSTITUT DE HAUTES ÉTUDES INTERNATIONALES ET DU DÉVELOPPEMENT

GRADUATE INSTITUTE OF INTERNATIONAL AND DEVELOPMENT STUDIES

ESSAYS IN LAW AND ECONOMICS OF ENFORCEMENT

THESIS
submitted at the Graduate Institute
in fulfillment of the requirements of the
PhD degree in International Economics

by

Dmitriy SKUGAREVSKIY

Thesis № 1226

Geneva

2017

ESSAYS IN LAW AND ECONOMICS OF ENFORCEMENT

INSTITUT DE HAUTES ETUDES INTERNATIONALES ET DU DEVELOPPEMENT
GRADUATE INSTITUTE OF INTERNATIONAL AND DEVELOPMENT STUDIES

ESSAYS IN LAW AND ECONOMICS OF ENFORCEMENT

THESIS
submitted at the Graduate Institute
in fulfillment of the requirements of the
PhD degree in International Economics

by

Dmitriy SKUGAREVSKIY

Thesis № 1226

Geneva

2017

Dmitriy SKUGAREVSKIY

Sur le préavis de M. Jean-Louis ARCAND, professeur à l'Institut et directeur de thèse, de M. Nicolas BERMAN, professeur adjoint à l'Institut et membre interne du jury, et de M. Roberto GALBIATI, professeur, département d'économie, Sciences Po, Paris et expert extérieur, le directeur de l'Institut de hautes études internationales et du développement autorise l'impression de la présente thèse sans exprimer par là d'opinion sur son contenu.

Le dépôt officiel du manuscrit, en 6 exemplaires, doit avoir lieu au plus tard le 24 novembre 2017.

Genève, le 2 novembre 2017

Philippe Burrin
Directeur

Thèse N° 1226

RESUME / ABSTRACT

Titre de la thèse / Title of thesis: Essays in Law and Economics of Enforcement

Résumé en français: Cette thèse étudie comment le droit pénal affecte la prise de décision et le comportement dans la deuxième plus grande juridiction du monde, la Russie. L'information sur la prise de décision judiciaire provient d'une nouvelle data set sur les accusés criminels universels traités dans le pays en 2009-2013. Le premier chapitre demande si le fait de plaider coupable à un crime entraîne une réduction de la durée de la peine. Les résultats révèlent une grande hétérogénéité des avantages individuels à plaider coupable et ce plaidoyer est le plus gratifiant pour ceux qui choisissent de ne pas plaider coupable. Le deuxième chapitre examine si les récidivistes commettent des crimes plus graves aux derniers stades de leur carrière criminelle. J'observe une escalade prononcée de la gravité des crimes dans le pays, qui est observée dans d'autres pays. Troisième chapitre découvre une discontinuité dans la durée de la peine pour les crimes de drogue en Russie. J'utilise un régression sur discontinuité pour trouver que la durée de l'incarcération augmente de 0,84 ans lorsque le quantité de drogue augmente de significatif à important. La discontinuité observée est étrangère à la loi mais se manifeste fortement dans la pratique, mettant en évidence la fonction expressive de loi.

English summary: This dissertation studies how criminal law affects decision-making and behaviour in the world's second largest jurisdiction, Russia. The information on judicial decision-making comes from a novel data set on the universe criminal defendants processed in the country in 2009–2013. First chapter asks whether pleading guilty to a crime leads to a reduction in sentence length. Results reveal high heterogeneity of individual benefits to pleading guilty and that pleading is most rewarding for those who choose not to plead guilty. Second chapter examines whether repeat offenders commit more serious crimes at the later stages of their criminal career. I find pronounced escalation in offence severity in the country which is robust to modelling assumptions and is observed in other jurisdictions. Third chapter uncovers a discontinuity in punishment for drug crimes in Russia. I employ a regression discontinuity design to find that length of unconditional real incarceration increases by 0.84 years when the drug weight crosses the significant-large weight threshold. The observed discontinuity is extraneous to the law but is strongly manifested in practice, highlighting the expressive function of the said law.

Contents

To P.

Acknowledgments

Writing is a long endeavour, and the support of people who have kept me company in the journey of creating this dissertation is precious. I am thankful to Professor Neven for igniting my interest in Law & Economics, Professor Arcand for his guidance, and Professor Volkov for demonstrating that the world can be bettered through academic research.

Cordial gratitude goes to my colleagues at the Institute for the Rule of Law at the European University at St. Petersburg. Conversations with Kirill Titaev, Maria Shklyaruk, Irina Chetverikova, Aryna Dzmitryieva, Alexey Knorre, Ekaterina Khodzhaeva, Vladimir Kudryavtsev, and Ella Paneyakh vastly enriched my knowledge of the subject. I thank Laura Cyron, Rahul Mehrotra, Anastasia Smirnova, and my classmates and colleagues at the Graduate Institute. I should also thank Wolfgang Alschner for demonstrating me the value of intellectual curiosity that transcends disciplinary divides.

I am indebted to my parents and family for lending me a moral and financial impetus to pursue a doctoral degree. I also appreciate help and hospitality of Serguei Zelepoukine and Elena Lubimova.

Financial support from the Graduate Institute is gratefully acknowledged.

Geneva — Saint Petersburg, June 2017

Introduction

This dissertation studies how criminal law affects decision-making and behaviour. In doing so, it views the law as a policy tool that shapes agents' actions to match the expected outcomes. In contrast to the black-letter law of legal documents and instruments, in this study I examine the law in action as manifested in judicial decision-making.

First chapter asks whether pleading guilty to a crime leads to a reduction in sentence length. To answer this question I examine case outcomes and characteristics of defendants from 7 jurisdictions around the world, including civil and common law countries. The wealth of information comes from a novel data set on the universe of 2.2+ million eligible criminal defendants processed in the 2011–2013's Russia, the world's second largest jurisdiction. With rich data at hand, I investigate a defendant's decision to plead guilty and its ramifications in the framework of essential heterogeneity (Heckman and Vytlacil, 1999, 2005, 2007). I identify and estimate the Marginal Treatment Effect of pleading guilty on length of unconditional real incarceration along the distribution of unobserved willingness to go to trial. This is done with a new instrumental variable that capitalises on court docket information, is relevant, and is universally available in the studied jurisdictions. Results reveal (i) high heterogeneity of individual benefits to pleading guilty, (ii) that pleading is most rewarding for those who choose not to plead guilty. These results are observed in every studied jurisdiction and are not sensitive to modelling assumptions, thereby demonstrating high internal and external validity. Uncovered heterogeneity in the benefits of a plea bargain sheds new light on the design and functioning of this legal institution.

Second chapter examines whether repeat offenders commit more serious crimes at the later stages of their criminal career. I trace criminal behaviour of the universe of offenders who committed their first crime and subsequently recidivated in Russia in 2009–2013. I create a novel measure of offence seriousness based on convictions and punishments assigned in the country and then view the computed offence seriousness as a trend-stationary autoregressive process where the upward trend captures the escalation effect while the negative autoregressive component reflects the stabilising incapacitation effect. In identifying this model I exploit the exogenous relationship between weather and crime. My findings show pronounced escalation in the country, are robust, and are replicable in other jurisdictions. The results allow me to identify groups of offenders who are more likely to escalate. Such risk-based measures point to more efficient ways to organise post-release supervision in the country.

Third chapter uncovers a discontinuity in punishment for drug crimes in Russia. To this end, I construct a data set on 35,125 seizures of cannabis or heroin from police records in 2013–2014 and link them to sentencing outcomes for the charged defendants. I build on the fact that severity of sanctions for drug offences is graduated with the weight seized. Legal thresholds stipulate significant, large, and extra-large weights of drugs seized by type. I employ a regression discontinuity design to find that length of unconditional real incarceration increases by 0.84 years when the drug weight crosses the significant-large threshold of 100 grams for cannabis or 2.5 grams for heroin. Since the Criminal Code prescribes no discontinuity in the punishment schedule at the threshold, this chapter uncovers the effect that is extraneous to the law but is strongly manifested in practice. The revealed discontinuity is robust to manipulations in drug weights by the police or defendants, size of weight bandwidth around the threshold, imperfect compliance with the law, differences in observable crime and defendant characteristics in the vicinity of the threshold, or case facts appearing in verdict texts.

Chapter 1

Is Pleading a Bargain under Essential Heterogeneity?*

Introduction

Plea bargaining is at the forefront of modern debate on criminal justice system design. Having emerged in the 19th-century England and Wales (Vickers, 2012), it has witnessed an increase in popularity in many jurisdictions around the world ever since. In the 1990s–2000s a number of countries undertook procedural reforms to enshrine plea bargaining in their criminal procedure (Langer, 2004). This constituted an export of the United States' legal institutions throughout the world.

Prevalent academic view of American-style plea bargaining emphasises its private na-

*I thank Shawn Bushway for help with Miller et al. (1980) data, Kathryn Hendley and seminar participants at UW–Madison, IHEID, and European University at St. Petersburg for their input. I am also indebted to Jean-Louis Arcand for helping me navigate the subtleties of MTE estimation.

ture (Scott and Stuntz, 1992). By engaging in it, as argument goes, a defendant trades probability of punishment (a verdict of guilt) for its severity, making a rational calculation.[1] In sharp contrast, continental European legal tradition restricts prosecutorial discretion (Merryman and Pérez-Perdomo, 2007). In civil law countries plea bargaining is limited by judicial oversight and is no longer a private contract between the prosecution and the defence.[2] In the latter system the defendant "simply 'throws himself on the mercy of the court' by pleading guilty to the original charge under the expectation of receiving a more lenient sentence thereby" (Padgett, 1985, p. 756). Plea bargaining in civil law tradition becomes a trilateral agreement between the judge, the prosecution, and the defence.

Such marked difference in views towards plea bargaining in civil and common law traditions may hinder any quantitative study of the key parameter of plea regime: plea discount, also known as "trial penalty" (for a comprehensive review of the literature estimating plea discount see Tata and Gormley (2016)). Plea discount is a differential in sentence length a defendant receives at trial and when pleading guilty, *ceteris paribus*. In American-style private regime of plea bargaining the decision to plead guilty hinges on expected plea discount granted by the prosecution (Rhodes, 1979) whereas the civil law tradition of judicial oversight of the procedure grants the judge the discretion to assign

[1] The argument was introduced by Landes (1971), Grossman and Katz (1983) demonstrated the welfare-improving effects of plea bargaining, Harris and Springer (1984) emphasized the trade-off between probability of punishment and its severity in a toy model.

[2] Adelstein and Miceli (2001) go as far as to argue that plea bargaining is inconsistent with the fundamental values of inquisitorial system.

the sentence in case of a guilty plea. Participation of sentencing judge, an officer of the court, therefore adds public elements and safeguards to a private contract between the prosecution and the defence on punishment severity. This difference in institutional design invariably amounts to differences in plea discount to be recovered from sentencing data in civil and common law jurisdictions. The comparability problem emerges even when assuming away legal differences by studying jurisdictions with similar institutional design. Givati (2014) finds that the society's value system influences the prevalence of plea bargaining, requiring researchers to take into account factors extraneous to criminal procedure when performing comparisons.

External validity concerns aside, internal validity of estimates of plea discount within jurisdictions has also been questioned. Smith (1986) points that any study of plea discount should be mindful of the measure of sentence length. In estimating the size of plea discount the literature has been comparing the length of custodial sentence (sentence resulting in real incarceration) for those who pleaded guilty and those who went to trial (Rhodes, 1979, Brereton and Casper, 1982, Spohn and Cederblom, 1991, Albonetti, 1997, Mustard, 2001, Ulmer and Bradley, 2006, Ulmer et al., 2010). This comparison assumes away (i) the determination of guilt, (ii) the choice of punishment by restricting the sample to the individuals who were found guilty by the court and were sentenced to real incarceration. Juxtaposition of conditional-on-real-incarceration sentences for those who plead guilty and those who go to trial may not offer a complete characterisation of plea discount even if we hold legal and extralegal characteristics of the jurisdiction constant.

Abrams (2011, 2013) advances this argument by focusing on sentence length unconditional of trial. To construct this measure, he replaces with nil sentences for the defendants who were dismissed or acquitted or were not sentenced to real incarceration. After disaggregating unconditional sentences for the full sample, he showed that expected sentences are not longer at trial than for plea bargain. This finding of zero to negative plea discount has prompted discussion on credible estimation of plea discount among criminal justice scholars and professionals (Kim, 2015, footnote 9).

In his pioneering study of unconditional length of real incarceration, Abrams (2011, p. 218) acknowledges that the produced estimate is the Local Average Treatment Effect (LATE) of pleading guilty (Imbens and Angrist, 1994). By construction, the LATE is defined by the instrumental variables (which drive the treatment take-up) and is not necessarily a parameter of policy interest (Heckman, 1997, Deaton, 2009). In case of the Abrams study, the LATE captures the plea discount for people that were induced to plead guilty by the seniority of the judges adjudicating their cases. This parameter is relevant only for a small share of population of the defendants. Furthermore, legal scholars and criminal justice professionals need to understand the relationship between the Average Treatment Effect (ATE) of pleading guilty, the Average Treatment Effect on the Treated (ATT), and the Average Treatment Effect on the Untreated (ATU).

This chapter contributes to both strands of literature on plea discount by performing credible plea discount estimation in multiple jurisdictions. It first gathers the data on unconditional sentence lengths and other observables from 7 jurisdictions around the world, including both samples and the universe of adjudicated criminal cases in civil and common law countries under different time periods. The wealth of information comes from a novel data set on the universe of 2.2+ million eligible criminal defendants processed in the 2011–2013's Russia, the world's second largest jurisdiction. Additional evidence comes from 6 jurisdictions in the 1970's United States. With the aid of this data I then propose a new instrumental variable — number of days elapsed from court receiving a case to it issuing a verdict — that relies on court docket information, is relevant, and is universally available in the said jurisdictions. This instrument enables me to estimate a continuum of LATEs for small changes in the number of days a case spends in court that are associated with people pleading guilty. This continuum is also known as the Marginal Treatment Effect (MTE, Heckman and Vytlacil (1999, 2005, 2007), see Cornelissen et al. (2016) for a review).

From the estimated MTE schedules I conclude that the marginal benefit of pleading guilty is non-linear in the unobservable case and defendant characteristics that lead the defendants to plead guilty. In other words, the accused with unobservables that make them less likely to plea enjoy the largest plea discount. This argument is reinforced when I aggregate the estimated MTEs into conventional treatment effect parameters and find in all jurisdictions that ATU < ATE < ATT of pleading guilty on sentence length. In other words, plea discount is largest for those who do not plead guilty. A series of robustness checks demonstrate internal validity of this finding whereas estimation across jurisdictions ensures its external validity. Uncovered heterogeneity in the benefits of a plea bargain sheds new light on the design of this legal institution and warrants future, possibly qualitative, examination of the decision to plea and its outcomes along the entire profile of the treatment status.

The chapter proceeds as follows. Section 1.1 introduces the model, estimation technique, and the instrument, Section 1.2 describes the gathered data, institutional contexts, and treatment variables. Section 1.3 presents the results and offers a discussion.

1.1. Model of plea discount

1.1.1. Set-up

Potential outcomes model I closely follow Arcand and Bassole (2011) in notation. Let Y_i be the unconditional (on guilt or punishment type) length of real incarceration for defendant i. This amounts to setting $Y_i = 0$ for cases resulting in anything but real incarceration.[3] Now consider an additive separable Roy (1951) model where outcome

[3]Equating cases resulting in non-carceral outcomes to nil sentence length allows one to include such outcomes into estimation, but this parametrisation comes at a price of imposing equal severity for acquittals, case dismissals, fines, mandatory or correctional labour, or other punishments not resulting in real incarceration. In other words, I assume a uniform ordinal preference ranking of punishment types by defendants. In reality, however, a low-income offender might view real incarceration as preferable over a large fine in terms of its discounted value (Lott, 1992).

equations for individual i sentenced by judge j if s/he pleads guilty (Y_1) or not (Y_0) are:

$$\begin{cases} Y_{1,i,j} & = \alpha_1 + \beta_1 X_{i,j} + U_{1,i,j} \text{ if } D = 1 \\ Y_{0,i,j} & = \alpha_0 + \beta_0 X_{i,j} + U_{0,i,j} \text{ if } D = 0 \end{cases}, \tag{1.1}$$

where X_{ij} are the individual-, case-, and judge-level observable characteristics that contribute to sentence length decided by a judge. They can include legal characteristics (e.g. mitigating circumstances or case facts) as well as extralegal ones that cannot influence the sentence severity under equality before the law principle but do affect judicial decisions in practice (e.g. defendant's gender, age, socio-economic status).

Pleading guilty is determined by a latent variable

$$D^*_{i,j} = [X_{i,j}, Z_{i,j}]\,\gamma - V_{i,j}, \tag{1.2}$$

where $Z_{i,j}$ is a set of observables that determine only the decision to plead guilty and not the outcome sentence length (excluded instrument). $V_{i,j}$ is an error term that contains unobserved characteristics that make the defendants less likely to plead guilty (as it enters (1.2) with negative sign). In the literature $V_{i,j}$ is referred to as the "unobserved resistance to treatment" and can be interpreted as the defendant's unobserved willingness to go to trial.

Outcome equation error terms contain judge characteristics, case facts, as well as the unobservables that affect both the individual's decision to plead guilty (or go to trial) and the judge's decision to assign more severe punishment:

$$\begin{aligned} U_{1,i,j} &= \lambda_j + \xi_1 V_{i,j} + \varepsilon_{1,i,j} \\ U_{0,i,j} &= \lambda_j + \xi_2 V_{i,j} + \varepsilon_{0,i,j} \end{aligned} \tag{1.3}$$

Therefore, $cov\left(U_{0,i,j}, V_{i,j}\right) \neq cov\left(U_{1,i,j}, V_{i,j}\right) \neq 0$ in the general case of $\xi_1 \neq \xi_2$.

I impose the *conditional independence condition* $\left(U_{0,i,j}, U_{1,i,j}, V_{i,j}\right) \perp Z_{i,j} | X_{i,j}$. This is a relaxed version of the traditional excluded instrument assumption of full independence because here $X_{i,j}$ can be correlated with the unobservables. Such relaxation comes at a price of an additional assumption of linear additive separability of the unobservables in (1.1) (Brinch et al., 2015). I also assume that the conditional (on $X_{i,j}$) distribution of $Z_{i,j}\gamma$ is non-degenerate ($Z_{i,j}$ is not-constant).

Now rewrite (1.1) in the switching regression framework:

$$\begin{aligned} Y_{i,j} &= D_{i,j} Y_{1,i,j} + \left(1 - D_{i,j}\right) Y_{0,i,j} \\ &= D_{i,j}\left(\alpha_1 + \beta_1 X_{i,j} + U_{1,i,j}\right) + \left(1 - D_{i,j}\right)\left(\alpha_0 + \beta_0 X_{i,j} + U_{0,i,j}\right) \\ &= \alpha_0 + \beta_0 X_{i,j} + D_{i,j} \underbrace{\left(\left(\alpha_1 - \alpha_0\right) + \left(\beta_1 - \beta_0\right) X_{i,j} + \left(U_{1,i,j} - U_{0,i,j}\right)\right)}_{\Delta} + U_{0,i,j}, \end{aligned} \tag{1.4}$$

where Δ is plea discount, the parameter of interest. I note that it is determined by an additive constant, observable characteristics of defendants $X_{i,j}$, and, crucially, unobserved differences in sentence lengths.

1.1.2. Conventional estimators of Δ

OLS The first impulse is to uncover $\hat{\Delta}$ with ordinary least squares. As per Heckman and Vytlacil (2007), the covariate-specific OLS estimate of plea discount for a random individual with observables x can be decomposed into:[4]

$$\begin{aligned}
\hat{\Delta}^{OLS}(x) &= E\left[Y_{i,j}|X_{i,j}=x, D_{i,j}=1\right] - E\left[Y_{i,j}|X_{i,j}=x, D_{i,j}=0\right] \\
&= E\left[\alpha_1+\beta_1 X_{i,j}+U_{1,i,j}|X_{i,j}=x, D_{i,j}=1\right] \\
&- E\left[\alpha_0+\beta_0 X_{i,j}+U_{0,i,j}|X_{i,j}=x, D_{i,j}=0\right] \\
&= (\alpha_1-\alpha_0)+(\beta_1-\beta_0)\,x+E\left[U_{1,i,j}|D_{i,j}=1\right]-E\left[U_{0,i,j}|D_{i,j}=0\right] \\
&= E\left[\Delta_{i,j}|X_{i,j}=x\right]+E\left[U_{1,i,j}|D_{i,j}=1\right]-E\left[U_{0,i,j}|D_{i,j}=0\right] \\
&= \text{ATE}(x)+\underbrace{E\left[U_{1,i,j}-U_{0,i,j}|D_{i,j}=1\right]}_{\text{Sorting on Gains}^{U}_{1,i,j}}+\underbrace{E\left[U_{0,i,j}|D_{i,j}=1\right]-E\left[U_{0,i,j}|D_{i,j}=0\right]}_{\text{Selection Bias}_{1\to 0,i,j}}
\end{aligned}$$

OLS estimation uncovers the average treatment effect of plea discount under three assumptions:

1. $E\left[U_{1,i,j}-U_{0,i,j}|D_{i,j}=1\right] \Rightarrow cov\left(\Delta, D\right)=0$ (no sorting on the gains effect).
2. $E\left[U_{0,i,j}|D_{i,j}=1\right]-E\left[U_{0,i,j}|D_{i,j}=0\right] \Rightarrow cov\left(D_{i,j}, U_{0,i,j}\right)=0$ (no selection bias effect).
3. $cov\left(\Delta, U_{0,i,j}\right)\neq 0$ (orthogonality of unobservables).

These are incredible assumptions in practice. When it comes to sorting on the gains effect, it is more realistic to assume that those defendants who decide to plead guilty have

[4]Notation and derivation borrows from Kyui (2016).

unobservables (e.g. case facts, private information on culpability) that ensure lower sentence if they plead guilty (screening effect of plea due to Grossman and Katz (1983)). Existence of such self-selection amounts to the negative sign before the Sorting on $\text{Gains}^{U}_{1,i,j}$ term. The setting when people take-up treatment based on their unobservable characteristics is also known as "essential heterogeneity" (Heckman et al., 2006).

Selection bias effect emerges when unobservable case facts or defendant characteristics affect both the individual's decision to plead guilty and the court's decision on sentence length. This is equally incredible to assume a zero selection bias. Eisenstein and Jacob (1977) offer a seminal account of judicial decision-making through the lens of working groups. Informal groups of discretionary actors that emerge in the courtroom were found to influence judicial behaviour. The configuration of relationship between judges, prosecutors and defence attorneys affects court outcomes, as many qualitative studies have found. Obviously, courtroom working group configuration is one of many unobservables that result in non-zero selection bias in real settings.

When it comes to the unobservables, a source of OLS bias might arise when $cov\,(\Delta, U_{0,i,j}) \neq 0$. From the error structure equation (1.3) it follows that even when the idiosyncratic component of the unobservables in no-plea case is $\varepsilon_{0,i,j} = \text{const}$ we can still have non-zero covariance $cov\,(D_{i,j}, U_{0,i,j})$ between the decision to plea and unobservables in case of trial because $U_{1,i,j}$ may enter the scene through the common error term $V_{i,j}$.

IV Mindful of the shortcomings of the OLS estimator, one could apply instrumental variables (IV) estimator instead. For expositional clarity, assume that the excluded instrument $Z_{i,j}$ is a binary vector. This allows to write the covariate-specific Wald estimator of plea discount for a random individual with observables x and excluded instrument $Z_{i,j}$ in terms of covariances:

$$
\begin{aligned}
\hat{\Delta}^{IV}(x, Z_{i,j}) &= \frac{cov(Z_{i,j}, Y_{i,j})}{cov(Z_{i,j}, D_{i,j})} \\
&= \frac{cov(Z_{i,j}, \alpha_0 + \beta_0 x + D_{i,j}((\alpha_1 - \alpha_0) + (\beta_1 - \beta_0)x + (U_{1,i,j} - U_{0,i,j})) + U_{0,i,j})}{cov(Z_{i,j}, D_{i,j})} \\
&= \frac{\begin{array}{c} cov(Z_{i,j}, \alpha_0 + \beta_0 x + D_{i,j}((\alpha_1 - \alpha_0) + (\beta_1 - \beta_0)x)) \\ + cov(Z_{i,j}, D_{i,j}(U_{1,i,j} - U_{0,i,j}) + U_{0,i,j}) \end{array}}{cov(Z_{i,j}, D_{i,j})} \\
&= \mathrm{ATE}(x) \times \frac{cov(Z_{i,j}, D_{i,j})}{cov(Z_{i,j}, D_{i,j})} + \frac{cov(Z_{i,j}, (U_{1,i,j} - U_{0,i,j}) D_{i,j})}{cov(Z_{i,j}, D_{i,j})} \\
&= \mathrm{ATE}(x) + \frac{cov(Z_{i,j}, (U_{1,i,j} - U_{0,i,j}) D_{i,j})}{cov(Z_{i,j}, D_{i,j})}
\end{aligned}
\tag{1.5}
$$

Instrumental variables estimator requires either of the two assumptions to uncover the ATE of plea discount:

1. $(U_{1,i,j} - U_{0,i,j}) = 0$ (no unobserved heterogeneity in sentencing)
2. $(U_{1,i,j} - U_{0,i,j}) \perp D_{i,j}$ (no sorting on the gains)

In practice, both assumptions imply absence of essential heterogeneity in plea discount. If this is not the case and the defendants' plea discount varies with unobservables, IV estimator would be biased. Given the practices of administrative data collection in the judiciary, measurement errors are not uncommon in data on sentencing outcomes. Furthermore, collection of many observables requires financial and labour input in developing the court record form and properly populating it with accurate information in the courts. Budgetary constraints in collecting the information on sentencing keep many observables that influence Δ in $U_{D,i,j}$, exacerbating omitted variable bias and stimulating essential heterogeneity.

1.1.3. Proposed excluded instrument $Z_{i,j}$

IV estimation in studying judicial decision-making has capitalised on random assignment of cases between judges that is present in some jurisdictions. The identification strategy rests on the observation that cases are assigned randomly to judges that are heterogenous in their severity (Nagin and Snodgrass, 2013, Aizer and Doyle, 2015, Dobbie and Song, 2015). Then instrumentation of the parameter of interest with judge fixed effects might offer (aside from issues that arise with many instrument asymptotics (Kolesár, 2013)) a credible estimate for such parameter. Another approach is due to Abrams (2011), Abrams and Fackler (2016) that instrument their parameter of interest (also Δ) by seniority (tenure length) of sentencing judge. Their strategy comes from a Priest and Klein (1984) model of settlement with which they argue that defendants can better infer unobserved judge severity for more senior judges because their probability density function of sentences is observed through prior decisions.

While randomisation of case assignments across judges is a desirable setting for any causal study, few jurisdictions can offer true random distribution of cases. First, judges are clustered within courts where randomisation occurs, whereas little randomisation between courts can be present. Second, judges do not have uniform workload or employment throughout the period they serve: in the US federal system, for instance, 10% of judicial seats are vacant (Yang, 2016).

I propose a different case-specific instrumental variable $Z_{i,j}$ to identify plea discount in this chapter: the number of days the case has spent in court since it was received by its clerks from the prosecutor's office. Advocates of plea bargaining are continuously pointing to it as a means of reducing the backlog in disposition of cases, eradicating bottlenecks in the procedure, and reallocating the resources to more complex cases. In many jurisdictions defendants waive the right to appeal when they plead guilty, and the hearings proceed without examination of evidence. Such arguments suggest that the speed of adjudication upon receipt of case might be a relevant and strong instrument. This is indeed the case in the data, as I will demonstrate in Section 1.3. As an aside, such instrument is readily available in many jurisdictions due to docket management concerns and requirements that judges face. Such requirements ensure that many stages of case handling by the court officers are duly documented.

What remains untestable, though, is the conditional independence condition $(U_{0,i,j}, U_{1,i,j}, V_{i,j}) \perp Z_{i,j} | X_{i,j}$. One could argue, for instance, that defence tactic of stalling might not only influence the decision to plea but also irritate the judge to the point of $Z_{i,j}$ entering into $U_{1|0,i,j}$. At this stage it is important to invoke the conditionality of the exogeneity assumption and state that the proposed instrument assumes that $X_{i,j}$ includes the number of days between the date of crime and indictment (case being sent to court) as an observable covariate. The latter variable proxies for case complexity and, as I will show below in Table 1.2, captures information that is different from what is communicated by $Z_{i,j}$ as adjudication speed.

Another commentator might point out that the $Z_{i,j}$ is unobserved at the time the defendant decides to plead guilty. Indeed, the literature argues to restrict information in $Z_{i,j}$ to what is available at the time of the decision to take up the treatment (Eisenhauer et al., 2015) and not include the (known in the future only) length of adjudication. However, the defendant is cognisant (through interaction with police or investigators and the fact that s/he waives the right to appeal) that $E[Z|D=1] < E[Z|D=0]$. This observation on differentials in expected speed of disposition hints at the underlying mechanism. One source of heterogeneity in unobserved resistance to pleading guilty $V_{i,j}$ (which will be formally analysed below) comes from the differences in discount factors. Those who have a preference for a prompt disposition of his/her case would favour lower $Z_{i,j}$.

Final benefit of the proposed instrument is that it is continuous and exhibits sufficient variation by treatment status and observables to identify the marginal treatment effect of pleading guilty that I will shortly introduce.

1.1.4. Local average and Marginal treatment effects

LATE To further show that the IV estimator (1.5) uncovers the local average treatment effect, recall from (1.2) that pleading $D_{i,j} = 1$ occurs when $[X_{i,j}, Z_{i,j}]\,\gamma > V_{i,j}$ (or, equivalently, $D_{i,j} = \mathbb{I}_{D^*_{i,j} \geq 0}$, where $\mathbb{I}_{\bullet}$ is an indicator function). I can apply the cumulative distribution function F of V to both sides of this inequality, which yields $F\left[[X_{i,j}, Z_{i,j}]\,\gamma\right] > F\left[V_{i,j}\right]$. Both sides of this equation are now bounded in $[0, 1]$ interval. The left-hand-side shows the propensity of pleading guilty based on the observable characteristics which I will refer to as $P\left(X_{i,j}, Z_{i,j} | X_{i,j} = x, Z_{i,j} = z\right)$. The right hand-side shows the quintiles of the distribution of unobserved resistance to pleading guilty (Cornelissen et al., 2016), and I will refer to it as $F\left[V_{i,j}\right] \equiv U_{D,i,j}$. To reiterate, for an offender with observables x, z, and unobserved resistance to pleading guilty u_D:

$$
\begin{aligned}
[X_{i,j}, Z_{i,j}]\,\gamma &> V_{i,j} \\
F\left[[X_{i,j}, Z_{i,j}]\,\gamma\right] &> F\left[V_{i,j}\right] \\
P\left([X_{i,j}|Z_{i,j}]\,\gamma | X_{i,j} = x, Z_{i,j} = z\right) &> U_{D,i,j} = u_D
\end{aligned}
\tag{1.6}
$$

Individual decides to plead guilty when the encouragement for a guilty plea based on her observable characteristics is larger than her unobserved resistance to pleading guilty u_D bound in $[0, 1]$ interval. I further impose the *common support condition* that states that for each defendant with observables X who decides to plea there should exist at least one defendant with same observables X who decides to go to trial (Heckman and Vytlacil, 2007). This condition ensures imperfect separability of the decision to plead guilty in terms of the observable characteristics of the defendants. When it is satisfied the instrumental variables estimator (1.5) can be rewritten as

$$\hat{\Delta}^{IV}(x,z) = \text{ATE}(x) + \frac{cov\left(z,\left(U_{1,i,j} - U_{0,i,j}\right)|D_{i,j} = 1\right) P\left(X_{i,j}, Z_{i,j}|X_{i,j} = x, Z_{i,j} = z\right)}{cov\left(z, D_{i,j}\right)}$$

The local nature of the IV-estimated effect becomes apparent when I compare two distinct values of the instrument z and $z^{'}$ such that $P(x,z) < U_D \Rightarrow D = 0$ (not pleading guilty) and $P\left(x,z^{'}\right) > U_D \Rightarrow D = 1$ (guilty plea). For brevity, consider a Wald estimator with excluded instrument and sole endogenous treatment variable (i.e. no covariates: $X = \emptyset$):

$$\begin{aligned} \hat{\Delta}^{IV}_{LATE}\left(z,z^{'}\right) &= \frac{cov\left(Z_{i,j}, Y_{i,j}\right)}{cov\left(Z_{i,j}, D_{i,j}\right)} = \frac{cov\left(Z_{i,j}, \Delta D_{i,j}\right)}{cov\left(Z_{i,j}, D_{i,j}\right)} \\ &\overset{def}{=} \frac{E\left[\Delta D_{i,j} Z_{i,j}\right] - E\left[\Delta D_{i,j}\right] E\left[Z_{i,j}\right]}{E\left[D_{i,j} Z_{i,j}\right] - E\left[D_{i,j}\right] E\left[Z_{i,j}\right]} \quad , \\ &= E\left[\Delta|z < V_{i,j} \leq z^{'}\right] \end{aligned} \tag{1.7}$$

where I use the definition of covariance in the second line. The IV estimator manages to uncover the plea discount averaged over compliers — individuals who decide to plead guilty based on the extra encouragement coming from the value of the excluded instrument $Z_{i,j}$ shifting from z to $z^{'}$ (Imbens and Angrist, 1994). However, the IV does not communicate any information about plea discount for the accused who would always plead guilty or always go to trial regardless of the incentive coming from the shift in the value of the instrument $Z_{i,j}$. This is an important limitation in criminal justice setting where one can observe high separability of propensity $P(x,z)$ to plead guilty with respect to such observables X as socio-economic or employment status, gender, or income. In particular, LATE of Abrams (2011) captures the plea discount for the defendants that were induced to plead guilty by the seniority of the judge adjudicating their cases and is silent on the plea discount for the defendants whose decision to plead guilty is orthogonal to judge's tenure.

Definition of MTE When essential heterogeneity (selection into pleading guilty based on unobservable characteristics $U_{D,i,j}$) is present one cannot arrive at the conventional treatment parameters with OLS or IV estimation. Instead, one can estimate a schedule of LATEs for small changes in $Z_{i,j}$ that induce the defendants to plead guilty (Heckman and Vytlacil, 1999, 2005, 2007). First, rearrange the outcome equation (1.4) as

$$Y_{i,j} = \alpha_0 + X_{i,j}\beta_0 + D_{i,j}\left((\alpha_1 - \alpha_0) + X_{i,j}(\beta_1 - \beta_0)\right) + D\left((U_{1,i,j} - U_{0,i,j})\right) + U_{0,i,j}$$

Then replace the treatment dummy with its propensity from (1.6) and take the conditional (on observables) expectation in terms of the unobservables:

$$E\left[Y_{i,j} | X_{i,j} = x, Z_{i,j} = z, P(X_{i,j}, Z_{i,j}) = p\right] = \alpha_0 + x\beta_0 + px_{i,j}(\beta_1 - \beta_0) + K(p), \quad (1.8)$$

where all non-linear terms are aggregated in $K(p) \equiv p(\alpha_1 - \alpha_0) + E[U_{0,i,j}|P(X_{i,j}, Z_{i,j}) = p] + pE[(U_{1,i,j} - U_{0,i,j})]$. The Marginal Treatment Effect (MTE) is defined as the derivative of the outcome equation conditional on observables x, z w.r.t. the propensity to plead guilty:

$$\begin{aligned} \hat{\Delta}^{MTE}(x, z, u_D) &\equiv \left.\frac{\partial E[Y_{i,j}|X_{i,j} = x, Z_{i,j} = z, P(X_{i,j}, Z_{i,j}) = p]}{\partial p}\right|_{p=u_D} \\ &= x(\beta_1 - \beta_0) + \left.\frac{\partial K(p)}{\partial p}\right|_{p=u_{D,i,j}} \end{aligned} \tag{1.9}$$

Intuition behind MTE $\hat{\Delta}^{MTE}$ shows the plea discount at certain levels of the unobserved resistance to pleading guilty. Adopting the example of Cornelissen et al. (2016), consider a case of p, propensity to plead guilty based on observable characteristics, taking a certain value $p = p_0$. Then all individuals with unobserved resistance to treatment $u_D < p_0$ decide to plea, ones with $u_D = p_0$ are indifferent. Now increase p_0 by a small amount ∂p. This increase will shift the indifferent individuals into pleading guilty. The change in the outcome sentence length for them is $\partial Y = \partial p \times \text{MTE}(u_D = p_0)$. I could gradually shift the excluded instrument $Z_{i,j}$ and first estimate plea discounts for those defendants who are likely to plead guilty based on their unobservables (low unobserved resistance to plead u_D). Then I could find plea discounts for the individuals with unobservables such that they are indifferent between pleading guilty and going to trial. Finally, I could estimate plea discounts for the defendants who are not likely to plea. This exercise would give me the schedule of treatment effects at different values of u_D. When the *common support condition* is fully satisfied, this u_D will encompass a near-unit interval of all quintiles of unobserved resistance to plead guilty (or, equivalently, willingness to go to trial).

Estimation of MTE Since (1.8) is non-linear in p, taking its derivative (1.9) requires non-, semi- or fully-parametric estimation. Heckman et al. (2006) details existing approaches. I build on their Semi-parametric Method 2 that models the non-linear term $K(p)$ in the outcome equation semi-parametrically. However, I depart from the said approach in several aspects which are enumerated in Supplementary appendix on page 50.

Inference on MTE Heckman et al. (1997) notes that "the bootstrap provides a better approximation to the true standard errors than asymptotic standard errors for the estimation of β_1, β_0 and $K(P)$" (as cited by (Carneiro et al., 2011, footnote 21)). In light of this observation, I construct confidence interval around $\widehat{MTE}$ with percentile bootstrap. Unlike Heckman et al. (2006), though, I bootstrap not the outcome equation (1.8) in isolation, but jointly with the propensity equation (1.6). This allows me to incorporate the sampling uncertainty arising both at the decision stage and at the outcome stage. This has important implications for the definition of common support (where the MTE can be estimated rather than interpolated): this is no longer $\begin{bmatrix} \max\left\{\min\widehat{p}\,|D=0,\ \min\widehat{p}\,|D=1\right\}, \\ \min\left\{\max\widehat{p}\,|D=0,\ \max\widehat{p}\,|D=1\right\} \end{bmatrix}$ anymore since $\widehat{p}$ is also bootstrapped. I plug in the global minimum/maximum $\widehat{p}$ from the bootstrap replications in the above equation to define the common support. In practice, this support is narrower than the one with fixed $\widehat{p}$. Finally, to make the estimation feasible i.t.o. the universe of defendants in the Russian data described below, I bootstrap on 20% random sample.

Evaluating other treatment effects with MTE Once $\widehat{MTE}$s given by (1.9) are obtained, one can integrate it at specific regions where $\widehat{P}(Z) < u_D$,$\widehat{P}(Z) > u_D$, or $\widehat{P}(Z) \lesseqqgtr u_D$ to get ATU, ATT, ATE, respectively. This can be done by computing weighted averages of $\widehat{MTE}$ with weights specified in Heckman and Vytlacil (2005, Table 1). However, the procedure for estimating the weights proposed in Heckman et al. (2006) involves probit estimation at every MTE grid point, which is not feasible in large data sets. Carneiro et al. (2017) propose to evaluate the treatment parameters in a simple and scalable procedure:

1. For each defendant obtain its $\widehat{P}(X, Z)$ and store it in a new column.
2. Repeat each observation n times, and create a column with n gridded values of u_D in common support after MTE bootstrap (I use 0.01 grid if the number of observations is less than 100,000 and a grid of 50 points otherwise).
3. Evaluate $\widehat{MTE}(X = x, U_D = u_D)$ at each row's observables X and u_D. This will give a schedule of $\widehat{MTE}$ for every quintile of unobserved resistance to plead guilty.
4. Obtain treatment effects:
   (a) ATE is the average of all the $\widehat{MTE}$s,
   (b) ATT is the average of the $\widehat{MTE}$s for observations where $\widehat{P}(Z) > u_D$,
   (c) ATU is the average of the $\widehat{MTE}$s for observations where $\widehat{P}(Z) < u_D$,
   (d) AMTE is the average of the $\widehat{MTE}$s for observations where the $\widehat{P}(Z) < u_D$ changes to $\widehat{P}(Z) > u_D$.

Inference on the obtained treatment parameters immediately follows with percentile bootstrap. However, as in the case of inference on MTE, it seems equally feasible to bootstrap only Step 2–4 above or include plea propensity estimation $\widehat{P}(X, Z)$ Step 1 in bootstrap as well. To maintain reasonable computation time in large problems, I leave $\widehat{P}(X, Z)$ outside the bootstrap for the Russian data described below.

1.2. Data

This chapter seeks to estimate plea discount under various institutional designs of the criminal procedure. To this end, I gather information on sentencing outcomes and related covariates from 7 jurisdictions across the world. Apart from providing the description of the collected data, this section also offers its the legal and institutional context.

1.2.1. Russia

Court records Russian court system is unlike any counterpart of comparable size: in this country any court is a federal entity with uniform structure. The lion's share of judges adjudicating criminal cases are formally appointed by the president and enjoy a federal status, as well as the courts.[5] The Judicial Department at the Supreme Court of Russia is responsible for administrative aspects of the judiciary. It has a separate line in the federal budget and pays salaries to court officials, maintains the infrastructure, gathers and publishes statistical reports, and provides informational support to courts.

[5] Strictly speaking, judges of peace that deal with misdemeanours do not have federal status, but are still reporting information to the federal authority in centralised and uniform fashion.

In an effort to increase information technology penetration in the Russian judiciary, it launched a country-wide court record collection system in 2009. Ever since then every single court record is expected to be digitised by a local court clerk, stored into a local data base which is then transferred to one of 83 regional offices of the Judicial Department. The regional offices gather the incoming local records into a regional database and upload it to Moscow, where the Supreme Court is located. In Moscow the central Judicial Department merges countrywide individual records into one data base which is then used to produce aggregated statistics, e.g. number of cases when the accused was sentenced to real incarceration for a given charge by region or number of minor offenders by charge. Such centralised arrangement is unprecedented in the world. In many federations court administration is delegated to its subdivisions and no uniform bottom-top data gathering procedure has ever been established.

The Institute for the Rule of Law at the European University at St. Petersburg was granted access to the data set on over 5 million depersonified court records on adult offenders processed by criminal courts in 2009–2013 that comprise the universe of cases and defendants. I identified this source of disaggregated data for academic use and led the Institute's effort to prepare the data on which this chapter now builds.

Data cleaning The accessed data are of high granularity and turn out to contain errors. I have developed a data cleaning procedure[6] that removed approximately 5% of records.

[6] This routine includes removal of records where primary punishment (punishment for the gravest charge) is not equal to overall punishment for an individual, or where primary punishment type is not

Further removal was due to duplicate detection: the data collection system had no version tracking system, so most appeals in higher courts prompted new court records with same observable characteristics but for non-empty appeal outcome fields. I also manually cleaned the sentencing judge name variable for over 25,000 judges, encountering and fixing the problems very similar to those Hauser (2012, p. 32) documented in his Florida state data: little consistency in judge name format, misspelled names and abbreviations, omission of everything but last name. I additionally rolled back surname changes when judges married in the said period and decided to take the names of their spouses. This cleaning enabled me to create a unique judge identifier based on his/her regularised surname and region. This identifier will be used to control for case-invariant unobserved heterogeneity in sentencing or plea propensity between judges.

Institutional context In 2001 Russia adopted a new Criminal Procedure Code that enabled plea bargaining.[7] The Russian reform introduced adversarial principles in the Soviet inquisitorial system, but some of them have remained dormant ever since (Burnham and Kahn, 2008). Plea bargaining was not among the unsuccessful innovations. In 2011–13 61.5% of eligible cases were disposed in the fast-track mode of trial following guilty plea (Table 1.1). Criminal Procedure Code stipulates that by pleading guilty the

equal to overall punishment type; removal of records where overall sentence size is more than 2 times as large as the primary sentence size while being more than 1.5 times as large as its upper bound; removal of records if sentence size is less than 0.7 of its lower bound; removal of caseload and judge variables based on the fact that judge caseload exceeds an reasonably set upper bound of 25 criminal cases per month.

[7] Detailed information is relegated to Supplementary appendix on page 52.

Variable	Mean	Median	SD	Min	Max	Observations
age	32.870	31	11.034	18	89	2,264,209
male	0.837	1	0.369	0	1	2,264,209
citizen	0.969	1	0.174	0	1	2,264,209
resident	0.921	1	0.270	0	1	2,264,209
education:						
(incomplete) higher education	0.089	0	0.285	0	1	2,264,209
vocational school	0.316	0	0.465	0	1	2,264,209
high school	0.374	0	0.484	0	1	2,264,209
incomplete high school	0.202	0	0.401	0	1	2,264,209
elementary school or no	0.019	0	0.136	0	1	2,264,209
socio-economic status:						
unemployed	0.640	1	0.480	0	1	2,264,209
worker	0.243	0	0.429	0	1	2,264,209
prisoner	0.006	0	0.078	0	1	2,264,209
student	0.024	0	0.153	0	1	2,264,209
office worker	0.030	0	0.170	0	1	2,264,209
official	0.009	0	0.092	0	1	2,264,209
top manager	0.010	0	0.099	0	1	2,264,209
entrepreneur	0.016	0	0.124	0	1	2,264,209
law enforcer	0.000	0	0.010	0	1	2,264,209
other	0.023	0	0.149	0	1	2,264,209
married	0.264	0	0.441	0	1	2,264,209
has dependants	0.338	0	0.473	0	1	2,264,209
crime under alcohol	0.251	0	0.433	0	1	2,264,209
crime under drugs	0.007	0	0.082	0	1	2,264,209
# charges per crime	1.199	1	0.583	1	5	2,264,209
crime stage:						
finished crime	0.933	1	0.249	0	1	2,264,209
preparation	0.001	0	0.036	0	1	2,264,209
attempt	0.065	0	0.247	0	1	2,264,209
crime in group:						
no group	0.875	1	0.331	0	1	2,264,209
group without intent	0.010	0	0.099	0	1	2,264,209
group with intent	0.113	0	0.317	0	1	2,264,209
organised group	0.002	0	0.041	0	1	2,264,209
role in crime group:						
actual doer	0.121	0	0.326	0	1	2,264,209
organiser	0.001	0	0.036	0	1	2,264,209
instigator	0.000	0	0.013	0	1	2,264,209
accomplice	0.003	0	0.052	0	1	2,264,209
first-time offender	0.599	1	0.490	0	1	2,264,209
pretrial detention	0.089	0	0.285	0	1	174,633
days elapsed from crime to court	144.461	72	220.203	0	1,483	2,264,209
days elapsed from court to verdict	38.650	23	47.354	0	328	2,264,209
unconditional length of real incarceration	0.979	0	1.419	0	28	2,264,209
conditional length of real incarceration	2.164	2	1.373	0	28	1,024,517
plea	0.615	1	0.487	0	1	2,264,209

TABLE 1.1: Summary statistics for the universe of the accused adult individuals with criminal charges eligible for fast-track mode of trial (Chapter 40 of Criminal Procedure Code) and adjudicated by Russian district, territory courts, and judges of peace in 2011–2013. Data exclude list-wise-deleted missing observations and 55,965 singleton observations after running a 2SLS regression of unconditional length of real incarceration on the above regressors with judge and primary charge fixed effects. (Correia, 2015). "days elapsed from crime to court" is the number of days between the crime date and the date of case being sent by prosecution to court. "days elapsed from court to verdict" is the number of days between the court receiving the case and issuing the verdict. "unconditional length of real incarceration" is the yearly size of real incarceration when non-custodial sentences or dismissals are replaced with zeros. Conversely, "conditional length of real incarceration" is the yearly size of real incarceration when non-custodial sentences or dismissals are removed from consideration. The latter four variables are right-winsorised at 99%. "plea" is a dummy equal to unity when individual pleaded guilty and entered fast-track mode of trial (Chapter 40 of Criminal Procedure Code).

defendant waives the right to appeal.[8] What does the accused person receive in return? The Code provides that the sentence for those pleading guilty shall not exceed the 2/3 of the sentencing range.[9] By pleading guilty, the defendant makes the judge exclude the upper third of the sentence length from consideration.

Such plea bargaining arrangement is an import of the Italian procedure by an American professor, an excellent account of Solomon (2012) suggests. In 2000 Russian Criminal Procedure Code drafting group invited Professor Stephen Thaman (St. Louis University) to provide a comparative perspective on plea bargaining in Germany, Spain, Italy, and the US. Later he drafted the said Section 40. He proposed a plea discount of 1/3: "the judge was to follow normal sentencing procedure and then subtract 1/3" (Solomon, 2012, p. 288). This is the sentencing discount that is found in Italy's *giudizio abbreviato* (abbreviated trial) special procedure (Fabri, 2008, p. 14) that was introduced during the country's criminal procedure reform of 1989. The difference between the draft's 1/3 and the Code's final "not more than 2/3" might seem to be slight at first glance, but in reality the provision in the final text of the Code gives the sentencing judge an immense discretion in determining plea discount: it is only weakly bounded from below.

Data restrictions & extensions I restrict the data to 2011-2013 because the key variable $Z_{i,j}$ — days elapsed between court receiving the case and issuing the final verdict — was introduced only in that period. I then limit the data to offenders eligible for plead-

[8] Criminal Procedure Code of Russia. Article 317.

[9] Criminal Procedure Code of Russia. Article 316, part 7

ing guilty[10] and right-winsorise conditional and unconditional sentence lengths at 99%. Finally, I perform list-wise deletion of missing observations and remove singleton (Correia, 2015) observations in terms of judges or primary charges. This brings the size of the data to 2,264,209 cases. Its summary statistics is given in Table 1.1.

However rich the data may be, they lack two important variables: pretrial detention of the defendant and the private/public type of defence counsel. To remedy this shortcoming, I perform a one-to-one match of the studied universe of court records with a sample of court texts gathered by `RosPravosudie.com` project and placed in the public domain. This match (detailed in Supplementary appendix on page 58), allows me to extract information on presence or absence of pretrial detention for 174,633 cases. I also extract word counts of introductory and factual part of verdict texts by counting the number of words before the phrases "HAS RULED/DECIDED THAT" in the matched verdict texts.

COVARIATES My outcome variable Y is the length of unconditional real incarceration, $X_{i,j}$ include the variables stated in Table 1.1 as well as judge, primary charge, and half-year time fixed effects. Note that I follow Volkov (2016) in creating socio-economic status variables from the present formal occupational and positional characteristics of the

[10] Eligibility criteria of upper bound of the length of real incarceration for the charge not exceeding 10 years renders the following charges (as of 2013) as not eligible for the fast-track mode of trial: 10501, 10502, 11103, 11104, 12602, 12603, 12713, 12723, 13103, 13203, 16103, 16203, 16204, 16303, 16402, 16604, 17414, 18602, 18603, 18804, 20404, 20501, 20502, 20503, 20512, 20602, 20603, 20901, 20902, 20903, 21001, 21003, 21102, 21103, 22603, 22604, 22702, 22703, 22812, 22813, 22903, 23003, 27500, 27600, 27700, 27800, 27900, 28101, 28102, 28103, 29004, 29500, 31700, 32103, 35301, 35302, 35601, 35602, 35700, 35800, 35902.

accused. In alternative specifications I include a dummy equal to unity when the defendant was under pretrial detention that is uncovered from matched verdict texts. $Z_{i,j}$ is the number of days elapsed between the date of crime and the date the case file was received by court.

1.2.2. Common law jurisdictions

In an effort to ensure the external validity of my findings on Δ, I extend my data by considering common law jurisdictions with publicly available information.

US sample

First source of data comes from Miller et al. (1980) study of plea bargaining in 6 American jurisdictions in 1978. I follow Bushway and Redlich (2012) in excluding El Paso from consideration, which leaves me with 5 jurisdictions. I also consider the data on burglaries and observables only as they form the lion's share of cases in the data. Unlike Bushway and Redlich (2012), though, I do not restrict my sample to male offenders who pled guilty. List-wise deletion of missing observations produces 2,018 cases to consider. The summary statistics in offered in Table 1.A.1.

Covariates My outcome variable Y is the length of unconditional real incarceration, $X_{i,j}$ include the variables stated in Table 1.A.1. It should be noted that this data set includes information on strength of evidence and type of defence counsel available to the defendant. $Z_{i,j}$ is days elapsed from indictment to disposition. Crucially, the information on days elapsed from crime to indictment is not available and is replaced with the number of days elapsed from arrest to indictment. Also, the data do not include the identity of the sentencing judge.

Alaska sample

Second source of data is due to Clarke et al. (1982). This is a study of disposition of felony cases throughout Alaska in 1974–76. What makes this period particularly interesting is an explicit ban of plea bargaining by the state attorney in July, 1975 (Rubinstein and White, 1978). Even though is beyond the scope of this chapter to engage in a discussion on the motivation or outcomes of this ban, such natural experiment brings important and much needed temporal variation to credibly estimate the Δ. List-wise deletion of missing observations leaves 2,318 data points summarised in Table 1.A.2.

Covariates Similarly, the outcome variable Y is the length of unconditional real incarceration, $X_{i,j}$ include the variables stated in Table 1.A.2. $Z_{i,j}$ is days elapsed from indictment to disposition. As in Miller et al. (1980) data, the information on days elapsed from crime to indictment is replaced with the number of days elapsed from arrest to indictment; identity of the sentencing judge is also unknown.

1.3. Results & Discussion

1.3.1. Descriptive statistics

As a point of departure, consider adoption rates for plea bargaining across jurisdictions (Tables 1.1, 1.A.1, 1.A.2). Whereas such rate is 61.5% for the universe of criminal cases in Russia, it is expectedly larger (85.8%) in Miller et al. (1980) sample since the latter focuses on burglaries and robberies only. Markedly lower adoption rates in the case of Alaska data (38.9%) can be attributed to the institution of moratorium on plea bargaining that occurred in the middle of the studied period.

What unifies the three data sources is the socio-economic status of the accused. In case of Russia 64.0% of the accused eligible for fast-track mode of trial were unemployed, in Alaska — 50.8%, while in Miller et al. (1980) data this figure reaches 78.0%. This observation implies that the vast majority of the defendants who are eligible for plea bargaining have low socio-economic status and might well face monetary, temporal, and informational constraints when weighing the benefits of pleading guilty versus going to trial.

More to that, 25.1% of the defendants were alleged to have committed their crimes under the influence of alcohol in Russia. Such crimes were found to have been committed in sole fashion (87.5%), by predominantly first-time offenders (59.9%). Such configuration of a median crime — committed by a sole unemployed first-time offender — explains the median time of 72 days from crime to indictment (case being sent from prosecution to court). The variation in this indicator is quite large (mean is 144.4 days, standard deviation — 220.2), suggesting pronounced right tail in the distribution of crime complexity and police effort that this variable is proxying for.

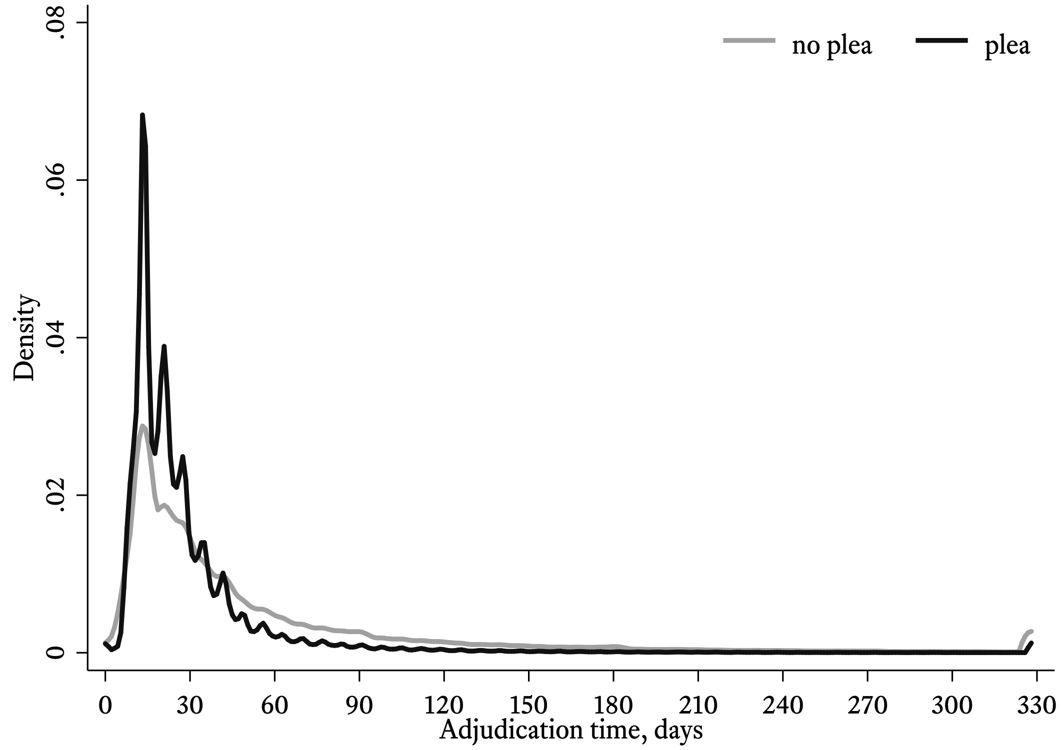
Upon receipt of a median case, Russian court spends 23 days adjudicating it. The variation is expectedly large, equally indicative of fat tails in the distribution of case complexity and defence tactics (or lack thereof). What is more surprising is the magnitude of this metric in common law jurisdictions: in Miller sample it amounts to 60 days whereas in Alaska it is found to be 101 days. Clearly, some part of the difference may be explained by the focus of these studies on burglaries & robberies or felonies, respectively. However, at least in Alaska case, Rubinstein and White (1978) put this figure in the context of a backlog of cases at courts in this jurisdiction — the primary reason for growing use of plea bargaining.

How does the speed of adjudication vary with pleading guilty? Figure 1.1 reports the density of speed of adjudication by the incidence of pleading guilty for every jurisdiction. In case of Russia and the US sample (Figures 1.1a, 1.1b) I observe that adjudication times for no-plea cases are higher and display fatter right tails — an expected result given my interpretation of this variable as a proxy of heterogenous discount factors of individuals. Interestingly, Russian data exhibit peaks at certain lengths of adjudication times (larger densities at 15 and 30 days are most visible). This may be related to docket management concerns and procedural restrictions that courts face. Alaska data are a noticeable outlier, where adjudication times are *lower* for non-plea cases. This regularity holds in the subsample of cases adjudicated before the ban on plea bargaining was instituted. However, as in the other two jurisdictions, the tails of the distribution of adjudication speed are fatter for the cases where the defendants went to trial in Alaska, still supporting my interpretation of this variable.

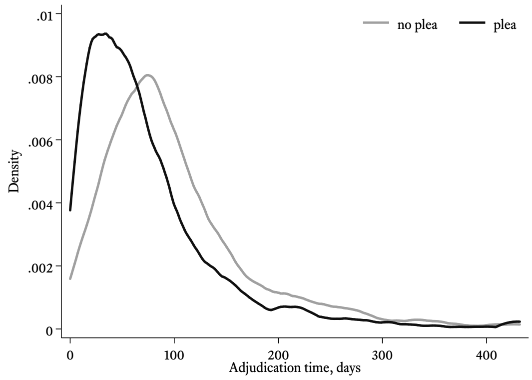
When I turn to the outcome variable of interest — years of real incarceration — any jurisdiction displays a large gap in its conditional and unconditional measure. In Russia, for instance, the average length of real incarceration for those 1,024,517 individuals who were sentenced to it is 2.16 years. When I replace with zeros non-custodial punishments and consider the universe of 2,264,209 defendants, the mean unconditional length of real incarceration drops to 0.98 years (Miller data: 6.64→2.77, Alaska: 3.11→0.62). This is indicative of the fact that the majority of cases result in non-custodial sentences or dismissals. Conditioning the outcome variable on guilt and real incarceration assumes aways such modes of disposition of cases and produces an inflated measure of expected punishment severity.

1.3.2. OLS/LATE evidence

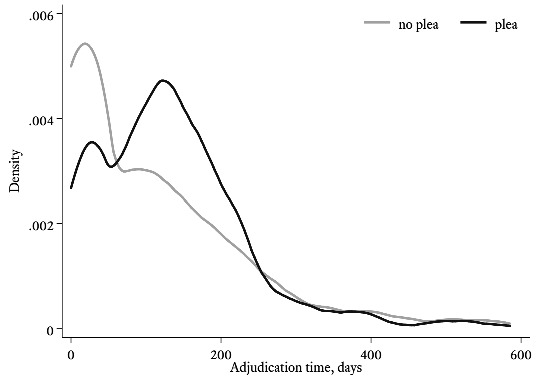
To highlight the importance of studying unconditional lengths of real incarceration, I conduct a simple experiment in Table 1.3 by linearly regressing (un)conditional length of real incarceration on plea dummy and other covariates in the universe of Russian defendants. Importantly, this exercise controls for unobservable case-invariant heterogeneity between judges and charges with the aid of respective fixed effects. Column (1) of Table 1.3 offers a finding in line with Abrams (2011): OLS of unconditional length of real incarceration produces a *positive* $\Delta = 0.017$ years which is non-significantly different



(a) Russia, universe of eligible accused individuals, 2011–2013



(b) US sample from Miller et al. (1980)



(c) Alaska sample from Clarke et al. (1982)

Figure 1.1: Kernel density estimate of days elapsed from indictment to disposition by guilty plea status, winsorised at 99%. Note: figures have different *y*-scale.

from zero. This suggests negative-to-zero plea discount, or plea penalty. The finding is reversed when I condition my dependent variable on guilt and real incarceration in column (2). This way, the Δ for the subset of defendants sentenced to real incarceration becomes a highly significant 0.261 years, or 3.13 months. Conditioning on real incarceration reverses the inference on Δ as it introduces severe selection bias.

As I have demonstrated in Subsection 1.1.2, OLS estimator rests on incredible assumptions. For this reason, I use my excluded instrument $Z_{i,j}$ — days case spends in court — to arrive at the two-stage least squares estimates of Δ. Those estimates are reported in columns (3)–(4) of the table and now appear to be (i) much higher in magnitude in relation to the OLS estimates, (ii) similar for conditional and unconditional definition of the dependent variable. First finding is expected because OLS would produce a $\hat{\Delta}$ which is downward biased in presence of negative selection on the gains. Second finding hints at validity of the chosen instrument. Well-defined instrumental variable $Z_{i,j}$ would eliminate the selection bias that arises in the OLS of conditional sentence length and drive the 2SLS estimates closer. This finding is sustained when I introduce one key omitted variable — pretrial detention — into the model with the aid of verdict texts. First-stage diagnostics reported in Table 1.3 signals that my $Z_{i,j}$ is highly relevant: first-stage R^2 is over 30%, F-statistic on $Z_{i,j}$ exceeds its critical value for the null of no significance. The first-stage behaviour and effect size is expected: an additional one hundred days of case staying in court is associated with 28.3% reduction in propensity to plead guilty.

Dependent variable	(1)	(2)	(3)
	Log word count of factual part in verdict text		
log(days elapsed in court)	0.143***		0.141***
	(0.009)		(0.009)
log(days elapsed from crime to court)		0.044***	0.040***
		(0.002)	(0.002)
Judge fixed effects	yes	yes	yes
Primary charge fixed effects	yes	yes	yes
Observations	171,464	171,917	171,367
R^2	0.500	0.480	0.504

Table 1.2: This table shows OLS estimates of regressing natural logarithm of word count of introductory and factual part of matched verdict texts on the number of days between the court receiving the case and issuing the verdict ("days elapsed in court") and the number of days between the crime date and the date of case being sent by prosecution to court ("days elapsed from crime to court"). I count the number of words before the phrases "HAS RULED/DECIDED THAT" in matched verdict texts to arrive at the dependent variable. Control variables identical to those in Table 1.3 are included in the model but not reported. Huber-Eicker-White standard errors clustered at region level in parentheses.

Additional evidence in favour of the proposed $Z_{i,j}$ comes from Table 1.2 where I regress the word counts of introductory and factual parts of verdict texts (that list case facts and details of the crime) on two of my measures of time: (i) number of days between the crime date and and the date of case being sent by prosecution to court, (ii) number of days between the court receiving the case and issuing the verdict. A 10% increase in the number of days elapsed in court is estimated to be associated with a 1.4% increase in the length of the factual part of verdict text. This association operates separately from the association between word counts and days elapsed between crime and indictment (an estimated 0.4% increase in verdict word count after 10% increase in number of days elapsed since crime).

The observed separability highlights the difference in the information that the indicators of time carry. I posit that "days elapsed from crime to court" is indicative of crime complexity and police effort whereas "days elapsed in court", that is my $Z_{i,j}$, captures defence strategy and willingness to go to trial that is different in the universe of offenders due to discount factor heterogeneity.

When it comes to samples from other jurisdictions (Tables 1.A.3, 1.A.4), I find a similar downward bias of OLS in estimating Δ in comparison with 2SLS results. Smaller number of observations and focus on felonies (Alaska sample) or burglaries and robberies (Miller study) precludes direct comparisons with results from 2011-13's Russia. However, in all settings the chosen instrument $Z_{i,j}$ is relevant (albeit exercising unexpected positive association with the propensity to plead guilty in Alaska). What is illuminating, though, is that neither in Miller nor Alaska data $\widehat{\Delta}^{IV}$ is positive: in the former estimation it amounts to insignificantly different from zero 3.32 years of plea discount; in the latter it is a significant 3.65 years of plea *penalty*. Lack of external validity of 2SLS estimates highlights its local nature. As the produced estimates are LATE, they are representative of different samples of defendants who are encouraged to plead guilty with a shift in $Z_{i,j}$.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable	Years of (un)conditional real incarceration							
Estimator	OLS	OLS	2SLS	2SLS	OLS	OLS	2SLS	2SLS
	Universe of all eligible accused individuals				*Accused with pretrial detention info from verdict texts*			
			SECOND STAGE				SECOND STAGE	
plea	0.017	-0.261***	-0.793***	-0.730***	0.044***	-0.164***	-0.637***	-0.656***
	(0.013)	(0.009)	(0.024)	(0.020)	(0.015)	(0.014)	(0.070)	(0.068)
			FIRST STAGE				FIRST STAGE	
100×days elapsed in court			-0.208***	-0.283***			-0.173***	-0.224***
			(0.007)	(0.010)			(0.014)	(0.026)
KP rk LM statistic *p*-value			0	0			0	0
KP rk Wald F statistic			905.47	757.98			142.23	75.19
Conditional on real incarceration	no	yes	no	yes	no	yes	no	yes
Judge fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Primary charge fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Observations	2,264,209	1,023,988	2,264,209	1,023,988	174,633	59,458	174,633	59,458
$R^2_{1st\ stage}$			0.305	0.304			0.341	0.328
$R^2_{2nd\ stage}$	0.686	0.597	0.630	0.578	0.709	0.637	0.653	0.616

Table 1.3: Russian universe. This table reports coefficients from a regression of conditional (on guilt and real incarceration, columns (2), (4), (6), (8)) or unconditional (columns (1), (3), (5), (7)) length of real incarceration with OLS of pleading guilty and other covariates (columns (1), (2), (5), (6)) or two-stage least squares (columns (3), (4), (7), (8)) where pleading guilty is instrumented with number of days between court receiving the case and issuing the final verdict. See Table 1.1 for the list of covariates and note that I do not interact plea dummy with them in this regression. Standard errors are Huber-Eicker-White, clustered at region level, and are reported in parentheses. KP rk LM statistic *p*-value and KP rk Wald F statistic are due to Kleibergen and Paap (2006). Columns (5)–(8) report the results for the sub-sample of cases with known information on pretrial detention extracted from verdict texts (see Supplementary appendix on page 58).

1.3.3. Evidence from MTE

Since Δ^{LATE} may not be comparable across jurisdictions, given the chosen $Z_{i,j}$, I evaluate the schedule of Δ^{MTE} for mean offenders within each jurisdiction and focus on comparing its profile. Before doing that, one should ensure that two conditions on plea propensity $P(X, Z)$ are satisfied.

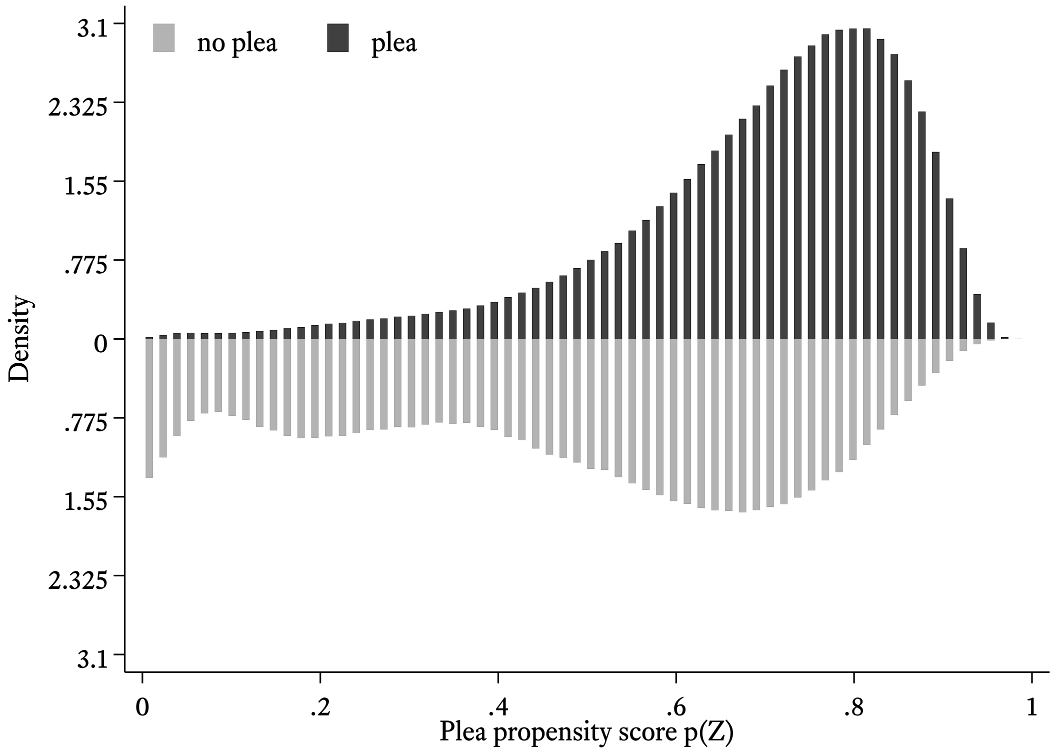
Non-separability of $P(X_{i,j}, Z_{i,j})$ $\widehat{MTE}$ can be evaluated only in the region of common support of the estimated propensity score where I observe both the decision to plead guilty and go to trial. Outside this region there exists no information on the alternative decisions (given the observables). In the ideal case of unit common support one is able to observe decision-making across the entire schedule of plea propensities.

This condition is testable by estimating $\hat{P}(X_{i,j}, Z_{i,j})$ and plotting its density by observed decision to plead guilty. This is performed for every jurisdiction in Figure 1.2. Expectedly, the universe of Russian data produces a near-unit common support $[0.01, 0.97]$. Same cannot be concluded about the two samples from common law jurisdictions. For this reason in displaying MTE for those samples I will specify the common support region over which it is estimated.

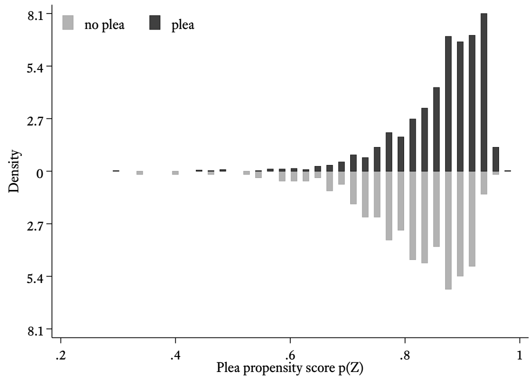
Sufficiency in identifying variation of $Z_{i,j}$ The excluded instrument $Z_{i,j}$ should be continuous and should exhibit enough variation to allow me to identify a schedule of $\widehat{MTE}$s. This condition is also testable. In Figure 1.2 I compare the predicted $\widehat{P}(X_{i,j}, Z_{i,j})$ (black line) and $\widehat{P}(\overline{X}, Z_{i,j})$, where $\overline{X}$ is mean value of observables. In other words, grey lines report the density of plea propensity for a mean offender when only $Z_{i,j}$ is varying. This exercise allows to assess how different offenders are in plea propensity when only $Z_{i,j}$ is shifting. As before, the benefit of considering the universe of offenders in Russian case becomes apparent with this test since $\widehat{P}(\overline{X}, Z_{i,j})$ covers almost 75% of the unit interval of plea propensity. The variation is less rich in Miller or Alaska data where the identifying variation of $Z_{i,j}$ is responsible for approximately 30% coverage of plea propensity.

Estimated $\widehat{MTE}$ profiles Having passed all the necessary checks, I estimate MTE for the common support, by jurisdiction. This result is presented in Figure 1.4 that plots $\hat{\Delta}^{MTE}$ by unobserved resistance u_D to pleading guilty. To the left, one could observe plea discounts for individuals with low resistance to treatment u_D who are, consequently, more likely to plead guilty. To the right one could see $\hat{\Delta}$ for defendants with large u_D who are less likely to plead guilty and have unobservables that make going to trial their preferred choice.

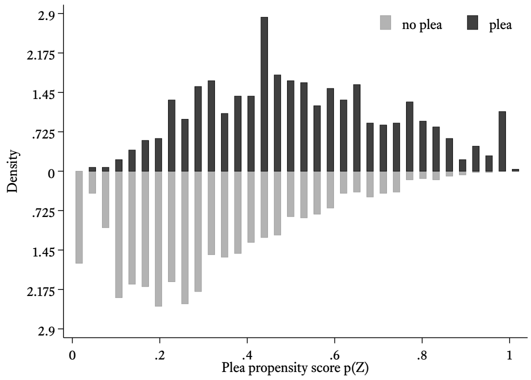
First lesson from the estimated profile in Figure 1.4 is that it is not flat in any jurisdiction. To see it more formally, I conduct an F-test by comparing the full model (1.8) where



(A) Russia, universe of eligible accused individuals, 2011–2013

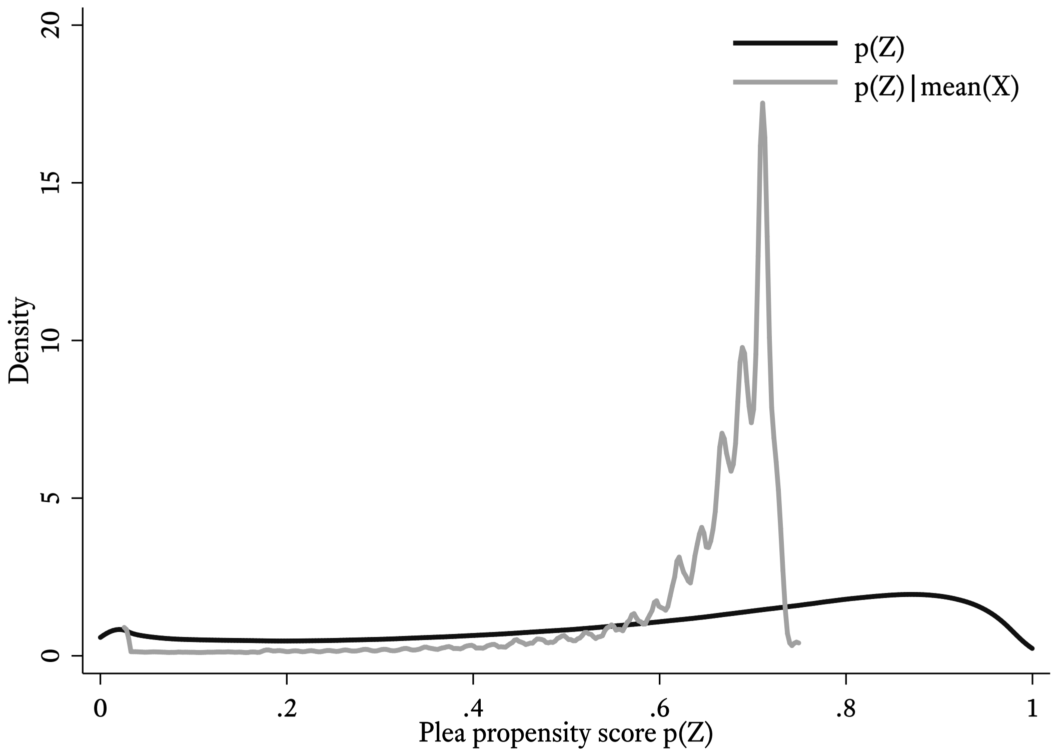


(B) US sample from Miller et al. (1980)

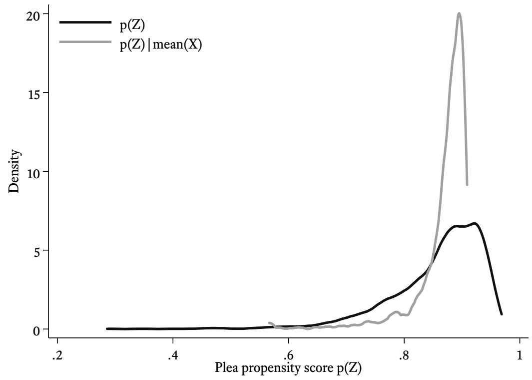


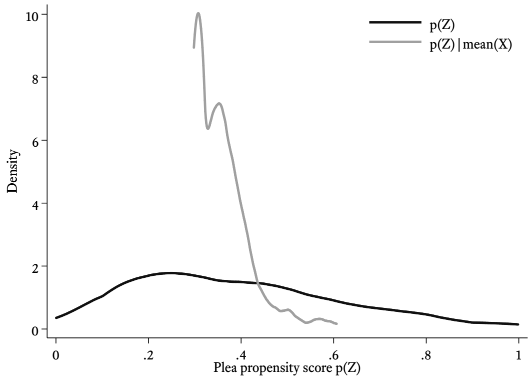
(C) Alaska sample from Clarke et al. (1982)

Figure 1.2: Distribution of estimated plea propensity score $\hat{P}\left(X_{i,j}, Z_{i,j}\right)$. Results are after coordinate descent logit that includes all covariates (and primary charge and judge dummies in Russian case). This binary classifier yields 76.6% correctly predicted, precision 77.1%, recall 88.0% for Russian data. Note: figures have different y-scale.



(a) Russia, universe of eligible accused individuals, 2011–2013





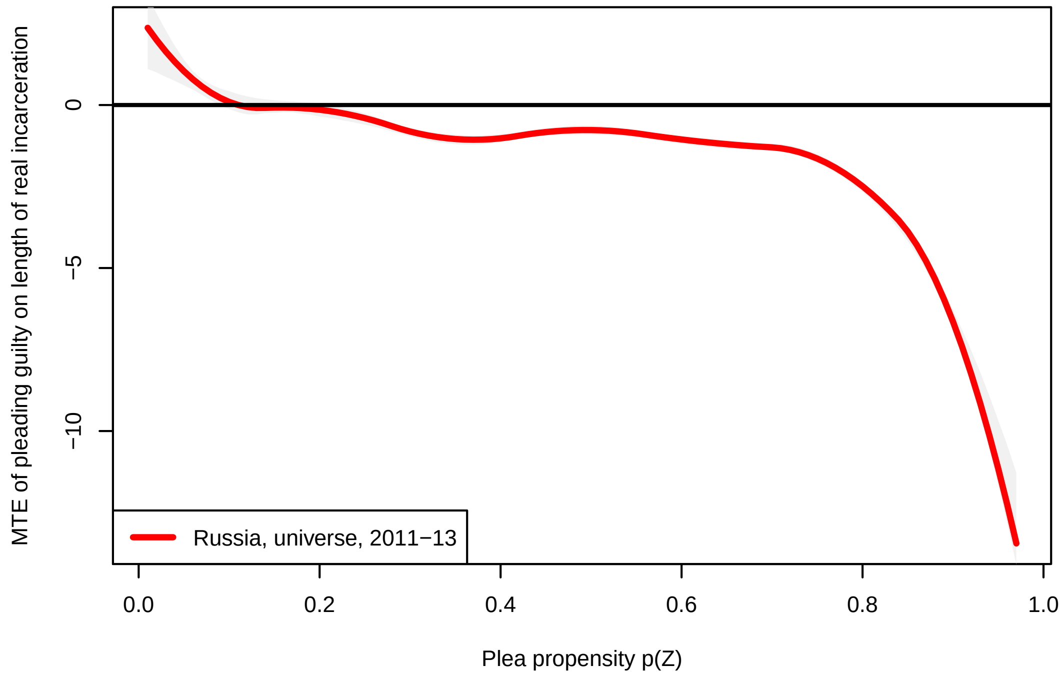
(b) US sample from Miller et al. (1980) **(c) Alaska sample from Clarke et al. (1982)**

Figure 1.3: Identifying variation in the data. This figure shows kernel density estimates of predicted plea propensity scores by actual incidence of pleading guilty, evaluated at observable characteristics of the accused (black line) or mean characteristics of the accused and observable adjudication speed (grey line). Note: figures have different y-scale.

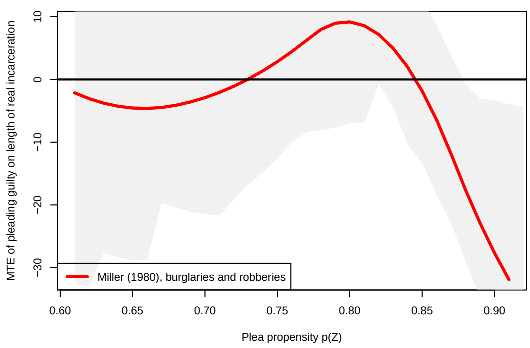
$K(p)$ is modelled with P-splines with a restricted model where $K(p) = p$ is assumed to be linear. This test strongly rejects the hypothesis of linearity of $K(p)$ (F = 152.71, p-value < 0.001) in Russia. Non-linearity of $K(p)$ signals the presence of essential heterogeneity in the model and implies that plea discount is varying in unobserved heterogeneity to plead guilty. This corroborates with the finding of Abrams and Fackler (2016, p. 31) that "the benefits acquired via a plea bargain may vary substantially depending on the nature of the crime the defendant is facing."

Second lesson from $\widehat{MTE}$ concerns its slope. $\widehat{MTE}$ is found to be *increasing* in unobserved resistance to treatment u_D. This signfies that people who are less likely to plea (right of MTE profile) are enjoying larger plea discount Δ. Those who are most likely to plea receive plea *penalty* instead. Therefore, I observe negative sorting on the gains. Such negative slope of MTE is present in all jurisdictions, even though the common support in Miller or Alaska data does not span the near-unit interval. To the best of my knowledge, this fact has not been previously documented in the literature.

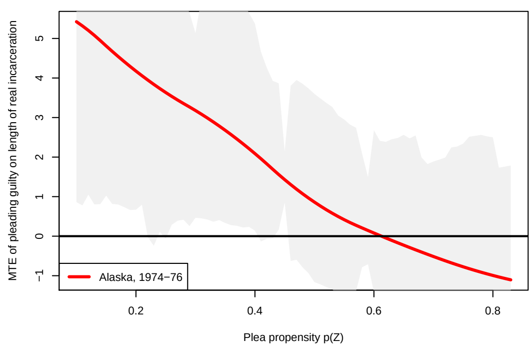
Treatment effects of Δ Another way to express negative sorting on the gains is to evaluate the $\widehat{MTE}$s at appropriate values of u_D to obtain conventional treatment parameters. This is performed in Table 1.4 which holds the main result of this chapter.



(a) Russia, universe of eligible accused individuals, 2011–2013



(b) US sample from Miller et al. (1980)



(c) Alaska sample from Clarke et al. (1982)

Figure 1.4: This figure shows $\widehat{MTE}$s for an offender with observable characteristics at means in selected jurisdictions. Estimation procedure in listed in Supplementary appendix on page 50 and involves approximation of $K(p)$ with P-splines, inclusion of covariates listed in summary statistics tables in $X_{i,j}$ and primary charge fixed effects where appropriate. Grey areas are 90% confidence intervals from percentile bootstrap with 300 replications (conducted on 20% sub-sample of data in Russian case). $\widehat{MTE}$s for Russia do not include judge fixed effects but rather region fixed effects. Note: figures have different y-scale and common support for u_D (reflected in differing x-scale).

Data	Specification	ATU	ATE	ATT	AMTE
US, sample $u_D = [0.61, 0.92]$	P-splines, no judge FE	-15.70 (-27.20, -0.23)	-5.04 (-12.13, 3.29)	-1.25 (-10.04, 11.14)	-12.12 (-20.26, -1.63)
Alaska, sample $u_D = [0.10, 0.83]$	P-splines, no judge FE	0.52 (-0.28, 2.52)	1.59 (0.37, 3.87)	3.20 (0.51, 6.75)	2.19 (0.45, 4.50)
Russia, universe $u_D = [-0.71, 0.52]$	P-splines, judge & charge FE	-0.79 (-0.85, -0.71)	-0.56 (-0.60, -0.51)	-0.39 (-0.44, -0.34)	-0.59 (-0.60, -0.56)
ALTERNATIVE SPECIFICATIONS					
Russia, universe $u_D = [0.01, 0.97]$	P-splines, region & charge FE	-4.16 (-4.31, -3.85)	-1.71 (-1.76,-1.61)	-0.27 (-0.35, -0.18)	-1.90 (-1.96, -1.81)
Russia, universe $u_D = [-0.78, 0.55]$	P-splines, judge & charge FE, pretrial detention	-0.54 (-0.74, -0.34)	-0.27 (-0.51,-0.18)	-0.08 (-0.47, 0.12)	-0.12 (-0.26, -0.17)
Russia, universe $u_D = [0.01, 0.97]$	cubic splines, judge & charge FE	-0.79 (-0.89, -0.70)	-0.57 (-0.61, -0.52)	-0.40 (-0.46, -0.36)	-0.58 (-0.61, -0.54)
US, sample $u_D = [0.58, 0.92]$	cubic splines, no judge FE	-15.04 (-25.40, 0.03)	-3.57 (-8.69, 2.51)	0.14 (-6.82, 7.85)	-12.02 (-19.36, -0.94)
Alaska, sample $u_D = [0.10, 0.86]$	cubic splines, no judge FE	-0.20 (-0.87, 2.23)	1.10 (-0.11, 3.47)	3.12 (0.17, 6.21)	1.88 (0.20, 4.36)

TABLE 1.4: This table reports treatment effects of pleading guilty evaluated from $\widehat{MTE}$ s for selected jurisdictions under various model specifications. Estimation procedure in listed in Supplementary appendix on page 50. Common support of propensity score at which the effects are evaluated is reported in the first column as $u_D[\ldots, \ldots]$. Note that in case of estimation with judge fixed effects I do not include judge dummies in the model but rather proceed with estimation on judge-demeaned data. This changes the interpretation of u_D to individual's deviation in unobserved resistance to treatment in relation to its mean value for the sentencing judge and by virtue of this no longer bounds $P(X, Z)$ in the unit interval. 90% confidence intervals in parentheses come from percentile bootstrap with 100 replications (conducted on 20% sub-sample of data in Russian case, 25 replications on 10% sub-sample for cubic splines in Russian case).

In every studied jurisdiction ATU < ATE < ATT of pleading guilty on length of unconditional real incarceration. I further estimate the treatment effects under alternative specifications, parametrising $K(p)$ with cubic splines, or running the estimation on subsample of data with available information on pretrial detention. I also find that my departure from Semi-parametric Method 2 of Heckman et al. (2006) in estimation does not alter the results qualitatively (Figure 1.A.1). The results also holds when I remove all unobserved invididual-invariant heterogeneity with defendant fixed effects instead of judge fixed effects.

The finding of negative sorting on unobserved gains to pleading guilty contributes to the debate on normativity and size of Δ. Instead of asking why LATE of pleading guilty is estimated at a particular value for the defendants who are encouraged by the shift in the value of instrument, I reverse the question and show how plea discount varies when the representative sample of defendants (in terms of their unobserved distaste for pleading guilty) is changed. This reveals high heterogeneity of Δ for different populations and, in turn, suggests that future studies of plea discount need to examine plea decisions and their outcomes along the entire profile of the unobserved heterogeneity. Additional lesson of this chapter is that plea discount cannot be summarised in one Δ due to inherent essential heterogeneity of defendants' decisions and outcomes. Finally, the chapter highlights the importance of taking into consideration the full repertoire of sentencing outcomes and dismissals that defendants face. Merely conditioning the outcome variable on custodial sentence assigned yields biased estimates of plea discount. Future work is required to investigate the structural (and, possibly, sequential) decision-making of defendants in presence of essential heterogeneity.

Supplementary appendix

Variable	Mean	Median	SD	Min	Max	Observ.
age	23.888	22	6.475	16	62	2,018
male	0.968	1	0.177	0	1	2,018
white	0.562	1	0.496	0	1	2,018
married	0.139	0	0.346	0	1	2,018
full-time employed prior to arrest	0.220	0	0.414	0	1	2,018
juvenile records	0.418	0	0.493	0	1	2,018
number of prior felony convictions	1.181	0	1.783	0	8	2,018
private defence counsel	0.209	0	0.406	0	1	2,018
pretrial detention	0.443	0	0.497	0	1	2,018
misdemeanor conviction	0.057	0	0.233	0	1	2,018
crime:						
burglary	0.748	1	0.434	0	1	2,018
robbery	0.252	0	0.434	0	1	2,018
jurisdiction:						
Norfolk, VA	0.131	0	0.338	0	1	2,018
Seattle, WA	0.297	0	0.457	0	1	2,018
Tucson, AZ	0.155	0	0.362	0	1	2,018
New Orleans, LA	0.146	0	0.353	0	1	2,018
Delaware county, PA	0.271	0	0.445	0	1	2,018
eyewitness	0.715	1	0.451	0	1	1,788
number of witnesses	5.682	5	3.231	0	18	1,994
days elapsed from arrest to indictment	35.841	21	40.105	0	229	2,018
days elapsed from indictment to disposition	79.093	60	71.146	0	432	2,018
unconditional length of real incarceration	2.772	0	5.457	0	75	2,018
conditional length of real incarceration	6.643	4	6.758	0	75	842
plea	0.858	1	0.349	0	1	2,018

Table 1.A.1: Summary statistics for the sample of the accused individuals with criminal charges eligible for plea and adjudicated in selected US jurisdictions in 1978, from Miller et al. (1980). Data are pre-processed following Bushway and Redlich (2012) and includes only males or females charged with robbery or burglary felony offences. Missing observations after list-wise deletion were excluded from consideration. "eyewitness" is a dummy equal to unity when there was any positive eyewitness identification of the accused. "number of witnesses" is an integer specifying the number of witnesses in the case. "days elapsed from arrest to indictment" is the number of days from person being arrested to indictment. "days elapsed from indictment to disposition" is the number of days from person being indicted to final case disposition (in court or elsewhere). "unconditional length of real incarceration" is the yearly size of real incarceration when non-custodial sentences or dismissals are replaced with zeros. Conversely, "conditional length of real incarceration" is the yearly size of real incarceration when non-custodial sentences or dismissals are removed from consideration. The latter five variables are right-winsorised at 99%. "plea" is a dummy equal to unity when individual pleaded guilty.

Variable	Mean	Median	SD	Min	Max	Observ.
age	26.532	23	9.333	17	74	2,398
male	0.868	1	0.339	0	1	2,398
race:						
black	0.139	0	0.346	0	1	2,398
native american/indian/eskimo	0.175	0	0.380	0	1	2,398
white/caucasian/other	0.686	1	0.464	0	1	2,398
married	0.254	0	0.435	0	1	2,398
occupation:						
unemployed	0.508	1	0.500	0	1	2,398
student	0.023	0	0.148	0	1	2,398
military	0.043	0	0.202	0	1	2,398
length of residency in Alaska:						
$\leq$ 6 months	0.100	0	0.300	0	1	2,398
6 months – 2 years	0.233	0	0.423	0	1	2,398
3 years – 7 years	0.180	0	0.384	0	1	2,398
$\geq$ 8 years	0.487	0	0.500	0	1	2,398
# prior felony convictions	0.574	0	1.861	0	21	2,398
pretrial detention	0.723	1	0.448	0	1	2,398
location:						
Anchorage	0.616	1	0.487	0	1	2,398
Fairbanks	0.308	0	0.462	0	1	2,398
Juneau	0.076	0	0.266	0	1	2,398
period:						
15.08.1974 – 14.02.1975	0.198	0	0.399	0	1	2,398
15.02.1975 – 14.08.1975	0.291	0	0.454	0	1	2,398
16.08.1975 – 15.02.1976	0.272	0	0.445	0	1	2,398
16.02.1976 – 15.08.1976	0.239	0	0.426	0	1	2,398
police witness	0.305	0	0.460	0	1	2,328
eyewitness	0.871	1	0.335	0	1	2,318
days elapsed from arrest to indictment	26.636	1	64.555	0	389	2,398
days elapsed from indictment to disposition	119.313	101	112.951	0	585	2,398
unconditional length of real incarceration	0.624	0	2.804	0	40	2,398
conditional length of real incarceration	3.111	1	5.613	0	40	481
plea	0.389	0	0.488	0	1	2,398

Table 1.A.2: Summary statistics for the sample of the accused adult individuals with criminal charges eligible for plea and adjudicated in Anchorage, Juneau, and Fairbanks, Alaska in August, 1974 – August 1976, from Clarke et al. (1982). Missing observations after list-wise deletion were excluded from consideration. "eyewitness" is a dummy equal to unity when there was any positive eyewitness identification of the accused. "police witness" is a dummy equal to unity if police officer was witness to the crime. "days elapsed from arrest to indictment" is the number of days from person being arrested to indictment. "days elapsed from indictment to disposition" is the number of days from person being indicted to final case disposition (in court or elsewhere). The latter two variables are right-winsorised at 99%. "unconditional length of real incarceration" is the yearly size of real incarceration when non-custodial sentences or dismissals are replaced with zeros. Conversely, "conditional length of real incarceration" is the yearly size of real incarceration when non-custodial sentences or dismissals are removed from consideration. "plea" is a dummy equal to unity when individual pleaded guilty.

Dependent variable	(1)	(2)	(3)	(4)
	Years of (un)conditional real incarceration			
Estimator	OLS	OLS	2SLS	2SLS
			Second stage	
plea	-0.962**	-1.522**	-3.332	-10.623**
	(0.413)	(0.682)	(2.248)	(4.381)
			First stage	
100×days elapsed in court			-0.068***	-0.050***
			(0.013)	(0.018)
KP rk LM statistic p-value			0	0
KP rk Wald F statistic			26.86	7.94
Conditional on real incarceration	no	yes	no	yes
Judge fixed effects	no	no	no	no
Primary charge fixed effects	yes	yes	yes	yes
Observations	2,018	842	2,018	842
$R^2_{1st\ stage}$			0.051	0.042
$R^2_{2nd\ stage}$	0.231	0.511	0.209	0.287

Table 1.A.3: Miller sample. This table reports coefficients from a regression of conditional (on guilt and real incarceration, columns (2), (4)) or unconditional (columns (1), (3)) length of real incarceration with OLS of pleading guilty and other covariates (columns (1), (2)) or two-stage least squares (columns (3), (4)) where pleading guilty is instrumented with number of days between court receiving the case and issuing the final verdict. See Table 1.A.1 for the list of covariates and note that I do not interact plea dummy with them in this regression. Standard errors are Huber-Eicker-White, clustered at region level, and are reported in parentheses. KP rk LM statistic p-value and KP rk Wald F statistic are due to Kleibergen and Paap (2006).

Dependent variable	(1)	(2)	(3)	(4)
	Years of (un)conditional real incarceration			
Estimator	OLS	OLS	2SLS	2SLS
			Second stage	
plea	0.540***	-1.662***	3.655***	4.217
	(0.101)	(0.532)	(1.220)	(5.800)
			First stage	
100×days elapsed in court			0.042***	-0.032
			(0.010)	(0.021)
KP rk LM statistic p-value			0	.09
KP rk Wald F statistic			19.04	2.28
Conditional on real incarceration	no	yes	no	yes
Judge fixed effects	no	no	no	no
Primary charge fixed effects	yes	yes	yes	yes
Observations	2,398	481	2,398	481
$R^2_{1st\ stage}$			0.214	0.449
$R^2_{2nd\ stage}$	0.413	0.808	0.180	0.700

Table 1.A.4: Alaska sample. This table reports coefficients from a regression of conditional (on guilt and real incarceration, columns (2), (4)) or unconditional (columns (1), (3)) length of real incarceration with OLS of pleading guilty and other covariates (columns (1), (2)) or two-stage least squares (columns (3), (4)) where pleading guilty is instrumented with number of days between court receiving the case and issuing the final verdict. See Table 1.A.2 for the list of covariates and note that I do not interact plea dummy with them in this regression. Standard errors are Huber-Eicker-White, clustered at region level, and are reported in parentheses. KP rk LM statistic p-value and KP rk Wald F statistic are due to Kleibergen and Paap (2006).

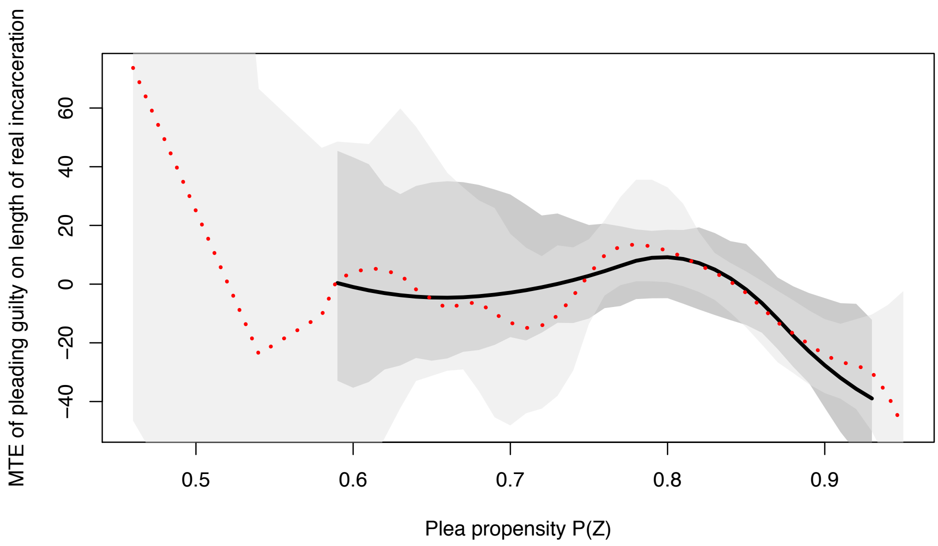


Figure 1.A.1: Estimated MTE for Miller data under different approaches to estimation. $\widehat{MTE}^{Heckman}$ (dotted red line) is estimated with Heckman et al. (2006)'s Semiparametric Method 2 implemented by Brave and Walstrum (2014), where $\widehat{P}(X, Z)$ is estimated with logit, 3-degree local polynomial is used to approximate $K(p)$. $\widehat{MTE}^{current\ approach}$ (solid black line) is estimated with coordinate descent logit, P-splines are used approximate $K(p)$, as described above. Shaded areas are 90% confidence interval from percentile bootstrap (100 replications). My approach also bootstraps at the decision stage, restricting the common support. Note that the grid of u_D at which $\widehat{MTE}$s are evaluated has different granularity across methods and is higher for $\widehat{MTE}^{current\ approach}$. This can partially explain larger wiggliness of $\widehat{MTE}^{Heckman}$.

Procedure for marginal treatment effects estimation

Estimation of $P(x, z)$ To obtain propensity scores (1.6) Heckman et al. (2006) use probit. Its convergence might be problematic in near-separability case, or when inclusion of fixed effects (e.g. when controlling for unobserved case-invariant heterogeneity between judges by including judge fixed effects) gives rise to the incidental parameters problem (Greene, 2004), or when the problem is simply too large. I model propensity to plea with coordinate descent logit of Friedman et al. (2009) along a LASSO regularisation path but impose no regularisation when drawing predictions from the estimated model.

Parametrisation of $K(p)$ I model $K(p)$ with P-splines in lieu of local polynomials. While the local polynomial estimation is an industry standard when it comes to derivative estimation, but its performance may be dubious is some instances. In simulated data drawn from Cox family, P-splines exhibit best behaviour i.t.o. RMSE but over-reject the null of linearity (Govindarajulu et al., 2009, p. 15). I use fast Restricted Maximum Likelihood to choose their tuning parameters with a logic implemented in `mgcv::bam` (Wood, 2004). As a robustness check, I also use penalised (to ensure that the ends match) cubic splines. To maintain compatibility with previous literature, I also estimate a fully parametric model where $(U_{0,i,j}, U_{1,i,j}, V_{i,j}) \sim$ *Multivariate Normal* in another robustness check.

Calculation of $\partial\widehat{K}(p)/\partial p$ Having obtained the $\hat{\alpha}_0, \hat{\boldsymbol{\beta}}_0, (\widehat{\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0})$ in P-spline estimated (1.8), Heckman et al. (2006) partial them out: $\tilde{Y} = K(p) + \tilde{\varepsilon}$, where $\tilde{Y} \equiv Y - \hat{\alpha}_0 - X'\hat{\boldsymbol{\beta}}_0 - X'(\widehat{\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0})P(X,Z)$ so that $E\left[\tilde{Y}|P(X,Z) = p\right] = K(p)$. Then they estimate this partialled-out regression with local polynomial and obtain $\partial\widehat{K}(p)/\partial p$ analytically. Such two-step approach requires to find the tuning parameters for the semi-parametric smoother twice and independently, which doubles computation time and does not take advantage of the same nature of the problem. I estimate only (1.8) [with fREML-optimal splines] and find $\partial\widehat{K}(p)/\partial p$ numerically with finite-difference method on a 0.01 grid of plea propensities.

Russian criminal procedure and sentencing: an overview[11]

The Russian legal system belongs to the continental European tradition of civil law. It relies on codified statute laws and procedural codes that regulate the application of laws. Despite the new Criminal Code (adopted in 1996) and the new Criminal Procedure Code (adopted in 2002), the procedure preserves a strong continuity with the Soviet criminal justice. The key features of the latter are the highly formalised investigation procedure and the domination of the investigator-prosecutor tandem and, consequently, a highly accusative bias with diminishing acquittal rate (Solomon, 1987). The criminal procedure system in Russia is often called neo-inquisitorial or investigatory, referring to the fact that the state in the face of its public officials objectively and on behalf of everyone concerned carries out the investigation of a crime to determine what happened (Burnham and Kahn, 2008).

The Criminal Code of the Russian Federation (2012) divides all criminal offences into four categories of seriousness, or gravity: low, medium, high, and top gravity. This classification determines the type of criminal procedure and sentencing rules. Low gravity crimes are handled by judges of peace and several of these, such as minor injuries or insults are processed in the mode of private prosecution. The plaintiff brings the case directly to the court and the law does not require formal investigation and support by the public prosecution. In contrast to that, medium, high and top gravity crimes trigger

[11] This appendix comes from Skougarevskiy and Volkov (2014) and was reproduced in abridged form in Volkov (2016).

the complex formalised procedure maintained by several organisationally distinct actors: police operatives (until 2009 known as "militia"), investigators, prosecutors (procurators), and judges of federal district courts. Police operatives are responsible for reacting to criminal acts or information about them, conducting detective work, finding, detaining, and interrogating suspects. All information about the crime is then passed over to the investigator, who is the key actor in the process. The investigator makes the decision concerning the initiation of the formal investigation procedure and brings charges against the suspect. The initiation of a criminal case (*ugolovnoe delo*) is the decisive move that often seals the fate of the suspect, because the investigator makes this move only if he or she is highly confident of having enough proof to convince the prosecutor and the judge about the blameworthiness of the suspect.

Centred on the case file, the heavily formalised pretrial investigation procedure is the centre-piece of the Russian criminal justice. The investigator has to record details of the crime, produce protocols of interrogation, testimonies, and proof according to strict procedural norms. The content of case file and the conclusion of guilt written by the investigator are the key sources of judgment for both public prosecutor and the judge. Once the investigation is completed, the case file is submitted to the prosecutor's office for approval. On the basis of the conclusion of guilt the prosecutor makes decision to support the charges and requests the type of punishment and the size of sanction for the accused.

The numbers and social composition of defendants in Russia's criminal courts is a combined result of several organised activities preceding the trial. These include the anti-crime activity of police employees and their policies of selective registration of offences; the discretion of the investigation agencies concerning the initiation of criminal procedure against suspects and the qualification of offences; the prosecutorial discretion in bringing cases to courts and requesting the type and severity of punishments.

In contrast to the adversarial Anglo-American tradition where prosecution and defence present their evidence in trial before the judge, in the Russian system the judge is presented first of all with a written file that accumulates the previous work of investigators and the prosecution. The judge can consider only that which is included in the case file, the content of which is determined by the investigation side. The defence side can collect its own evidence and proof, but these rarely make their way into the case file before the trial. The evidence of the defence side is presented at the trial, leaving it to the discretion of the judge to formally include it into the case file and thus be taken into account. The judge, however, can request additional expertise and information during the trial at the request of one of the sides.

After the hearings the judge has to make two interrelated decisions. First, to assess the proof and decide whether the crime took place and whether the defendant is guilty of committing it. Second, to select the type of punishment and the sanction if the first decision is positive. Ranked by the cost to the defendant in the ascending order the main accusative sentencing decisions are the following: no punishment; non-carceral punishment (a fine, mandatory or correctional works, occupational restrictions); restriction of freedom; arrest; suspended incarceration; real incarceration (from 2 months to life sentence). Still, the principal choice is that between incarceration and alternative punishments (the in/out decision).

The Criminal Code gives the judge a rather wide discretion in determining the sanction. Each degree of gravity of offence is defined with reference to the maximum possible length of incarceration measured in years. For low gravity this is 3 years; 5 for medium, 10 for high and over 10 years — for especially high (top) gravity. The qualification of the offence, including the degree of gravity, is the duty of the investigation, and the judge can only either accept it or reduce it. Besides four degrees of gravity, each article of the Criminal Code describes a particular offence and prescribes an upper bound or both a lower and an upper bound of sanction for a fine or incarceration. For example, according to Part 1 of the Article 161 "Robbery", this crime is "punishable by community service for a term of up to four hundred and eighty hours, or by correctional works for a term of up to two years, or by restriction of freedom for a term of two to four years, or by an arrest for a term of up to six months, or by deprivation of freedom for a term of up to four years", Criminal Code (2012). Within the same article of the Code there may be several parts (subsections) designating different degrees of gravity of the same crime. For example, Part 3 of Article 161 designates robbery committed by an organised group and sets the sanction from six to twelve years with or without a fine of up to one million rubles.

So the judge has a wide sentencing discretion in assigning the length of incarceration as well as various non-carceral alternatives for the same crime. What are the main considerations guiding the sentencing decision? According to the Criminal Code, the judge shall consider the nature and degree of social danger of crime (which in part are reflected in the degree of gravity), the personality of the convicted, including any mitigating or aggravating circumstances, and also the influence of the imposed sanction on the rehabilitation of the convicted and on the conditions of life of his family. There are 15 different aggravating circumstances, including repeated offence, a leading role in committing the crime, participation in a group or organisation, and so on. Repeated offence classified as recidivism is also specified in a separate article that sets the sanction no less than the lower third of the sanction interval, but allows a more lenient punishment in case the judge identifies mitigating circumstances. The list of mitigating circumstances includes such things as committing a crime for the first time as a result of a combination of circumstances; minor age; responsibility for infant children; self-defence, physical or mental coercion; giving oneself up; cooperation with investigation; medical help to the victim of the crime. Legal scholars note, the list of mitigating circumstances is open-ended (Smirnov and Kalinovskiy, 2012, p. 598). The Criminal Code also compels the judge to account for the stage of committing a crime (preparation, attempt, or completed criminal act) and the role of the convict as accomplice. Despite the specification of mitigating and aggravating circumstances, their identification and documentation to a large extent depends upon judicial discretion. The law requires the judge to take into account the personality of the defendant, but does not specify how this should be done and which particular indicators should affect the sentencing decision. This gives the judge the legal opportunity to take into account extralegal characteristics of the defendant, but we do not know how judges use this discretion. Interview sources indicate that they look at occupation, employment, family status, and use reference letters from one's place of work or from the local community to justify an increase or reduction of sanction.

Verdict texts matching procedure

This appendix documents the merge of the universe of criminal court records data obtained from the Judicial Department at the Supreme Court of the Russian Federation for the years 2009–2013 with the verdict texts gathered by `RosPravosudie.com` project and placed in the public domain.[12]

I start with 9,130,283 verdict texts issued by the courts of general jurisdiction (territory courts and district courts) and 9,398,643 verdict texts for the judges of peace. Out of these texts, I select only criminal cases in courts of first instance (based on the verdict text metadata created by RosPravosudie staff). Thereby, the starting number of verdicts to consider reduces to 916,387 for the courts of general jurisdiction and 508,511 for the judges of peace.

Verdict texts (meta)data cleaning

The data have undergone a comprehensive cleaning exercise:

1. Only judge surname was extracted from the relevant data fields in both sources.
2. Region names and verdict dates were transformed to conformable formats in both data sets.
3. Charge name in verdict text metadata was transformed to court records data format such that "art. 159.2 p. 1 para 5" became «15921»
4. A new variable was created in both data sets, extracting the bare-bone number from criminal case numbers. This way, both the record "1-254/09" and "1-254 (2009)" would be transformed to "254", the case number net of year or other special symbols.

[12] `https://rospravosudie.com/society/33m`

Merge strategies

I developed the following strategies for effecting the one-to-one merge between court records and publicly available verdict texts:

1. merge on region name, court type, judge surname, and vanilla criminal case number. This yields 42,962 records merged.
2. merge on region name, court type, judge surname, verdict date, and the bare-bone criminal case number. This yields 83,604 additional records merged.
3. merge on region name, court type, judge surname, verdict date, and first two charges (with parts). This yields 68,998 additional records merged.
4. merge on region name, court type, judge surname, verdict date, and first two charges (without parts). The reasoning behind this merge is that for some verdict texts metadata does not specify charge parts. This yields 42,509 additional records merged.
5. merge on region name, court type, judge surname, verdict date, and first two charges (with parts), changing the order of charges, conditional on case number being empty in the verdict texts metadata. This yields 6,720 additional records merged.
6. merge on region name, court type, judge surname, and verdict date. This yields 1,252 additional records merged.

It is important to note that the merge was deemed successful if it produces a one-to-one relationship. In other words, if any merge yielded two or more candidate records, both of them were discarded. For instance, it is natural to expect that the merge strategy (6) above produces one-to-many relation: a judge can adjudicate multiple cases per day. I have discarded such ambiguous cases and was left only with the definite merges when one judge ruled on one case per day.

The merge effort rendered 246,045 one-to-one merges in total in 2009–2013. However, further examination revealed an error in RosPravosudie data: 3500+ verdicts had (pairwise or more) identical texts but differing metadata that was supposedly extracted from those texts. Such erroneous verdicts were removed form the analysis, and I was left with 242,527 verdicts to consider.

Chapter 2

Escalation in Crime Seriousness, Incapacitation, and Post-Release Supervision: Evidence from Russia*

Introduction

Oleg Belov of the city of *Nizhny Novgorod* killed his wife, mother, and 6 children in July–August, 2015.[1] This crime was a dire result of escalating pattern of domestic violence in that family. The late wife had raised her grievances with local police inspectors 6 times in October–July, filing written complaints for domestic violence and abuse. However, the police inspectors did not respond to her complaints, citing work

[1]Criminal case 2-1/2016 (2-26/2015) adjudicated by territory court of Nizhny Novgorod.

*I thank Brian Francis and Jiayi Liu for helping me make sense of UK data as well as Jeffrey Dickinson and seminar participants at IHEID and European University at St. Petersburg for their feedback on earlier drafts of this chapter.

overload and hearsay nature of the statements that had been refuted by the husband. In the aftermath of the tragedy, 9 police inspectors left the job, thwarted by negligence charges pressed against their colleagues by the prosecution (RIA Novosti, 2015).

Substantial chunk of local police work in Russia and other jurisdictions is post-release monitoring of convicted individuals or at-risk groups (Volkov et al., 2015). This monitoring is normatively set to cover the entire population of such individuals: when there resides a number of released offenders in the area of responsibility of a local police officer (*uchastkovii*), the inspector is expected to monitor the entire group regularly. However, qualitative studies have shown that on-site documentary work takes up to 60% of daily time budget of local police officers (Volkov et al., 2015, p. 27). In such setting the normative requirement of full monitoring coverage of offenders is unrealistic.

Social scientists have long been studying patterns of criminal behaviour over time through the lens of criminal careers and developmental criminology (Blumstein et al., 1986, Farrington, 2003). In that strand of literature two definitions of escalation emerge: the first concerns the growing frequency of individual's criminal behaviour over the course of life (Sherman et al., 1991) whereas the second definition views escalation as the growing seriousness/severity of the committed crimes with time. This chapter studies the escalation in its second definition. Escalation defined that way has become the focus of research only recently (Le Blanc, 2002, Armstrong and Britt, 2004, Berg and DeLisi, 2005, Piquero and Chung, 2001, Francis et al., 2005, Piquero et al., 2006, Kazemian et al., 2009, Ramchand et al., 2009, Liu et al., 2011, Francis and Liu, 2015, Cihan et al., 2017). So far quantitative examination of escalation has been limited in either the size of individuals observed or the length of time periods under study.

This chapter contributes to the literature on escalation of crime seriousness in three ways. First, having gauged offence seriousness in the spirit of Francis et al. (2005), the chapter traces the criminal behaviour in a novel jurisdiction, Russia, with a large-N, small-T universe of 449,967 offenders who committed their first crime and subsequently recidivated in 2009–2013. Second, it identifies the conditional-on-reoffending effect of individual characteristics on crime escalation with a Mundlak (1978) device. Third, the chapter suggests that offence seriousness should be considered as a trend-stationary autoregressive process where the upward trend captures the escalation effect while the negative autoregressive component reflects the stabilising incapacitation effect. In identifying this dynamic panel model I depart from the traditional assumption of instrumental orthogonality of the first difference and exploit the exogenous relationship between weather and crime severity.

The results point to more efficient ways to organise post-release supervision than random monitoring. Predictions from the created model could be used to identify at-risk groups of offenders and target monitoring efforts. Since exposure to violent behaviour is shown to reduce the human capital of the victims (Brown and Velásquez, 2017), risk-based post-release supervision might appear to be welfare-improving: it reallocates the time budget of police officers in a fashion that minimises escalation in the group under monitoring and reduces societal costs of potential crimes.

The chapter proceeds as follows. Section 2.1 introduces the novel data on criminal behaviour of individuals, Section 2.2 offers a review of approaches to measuring offence severity with observable information, Section 2.3 joins the data and the measured severity in a model of escalation and presents its results, Section 2.4 shifts the focus to de-escalation by suggesting a novel decomposition that separates the effect of escalation from incapacitation.

2.1. Data

Institutional context Russia's judiciary is unique in relation to other federations of comparable size. Firstly, all courts adjudicating criminal cases are under direct (courts of general jurisdiction) or indirect (judges of peace) federal management by the Judicial Department of the Supreme Court. The Judicial Department is responsible for maintenance and administration of operations of the judiciary. It is also tasked with financial and informational management of courts. Due to its mandate, the Department strives to achieve centralisation of the judiciary when it comes to information flows and reporting. In 2009 it launched an all-Russia system that gathers court records that are filled by the clerks for every accused individual in every court. Figure 2.A.1 showcases those court records in paper and digital form. Information is then transferred from courts to the regional offices of the Judicial Department and, finally, to the federal office in Moscow in a process that is depicted in Figure 2.1.

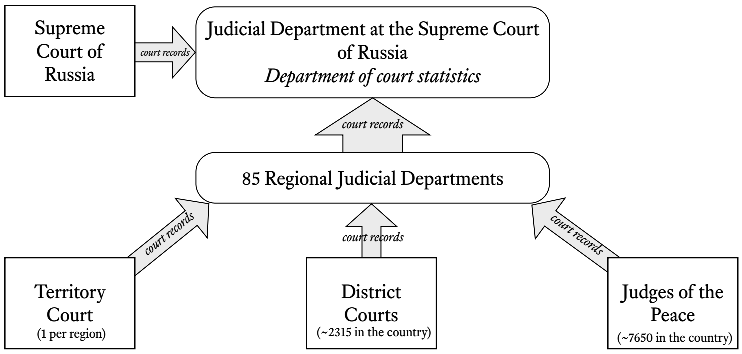


Figure 2.1: This figure shows information flow of court records in the Russian judiciary. Data on the size of the judiciary is from Bocharov et al. (2016, p. 15).

Secondly, criminal procedure of Russia maintains continuity with its Soviet counterpart. While its detailed description is relegated to Supplementary appendix on page 52, it is important to note that every criminal investigation in the country is formalised. This means that any accident report that is deemed to constitute a crime by the police results in formal initiation of a criminal case file. This case file is populated with evidence, witness reports, and procedural documents in the process of pretrial investigation conducted by the police. In contrast to common law jurisdictions where detective work is deformalised, the contents of a Russian case file bear procedural significance. When the investigation is completed, the case file is sent to the prosecutor who examines the gathered evidence and either approves the file and sends it to court for adjudication or, alternatively, returns the case file for additional pretrial investigation. The investigators can also drop the charges and not send the case to prosecution. At court the judge conducts the trial stage of investigation by examining the evidence gathered in the case file by police investigators at the pretrial stage.

To put the described procedure in perspective, note that in 2013 70,657 case files were dismissed at the pretrial stage while the remaining 905,616 files were sent to the prosecutor's office in the country (Shklyaruk, 2014). Out of those forwarded to prosecution, 33,609 cases were returned for further investigation while 872,007 files were sent to court for adjudication. Jointly with cases of private prosecution brought directly to courts by victims, in 2013 Russian courts received 946,474 criminal cases concerning 985,805 accused individuals. Low observed rate of charge dismissals at pretrial stage (7.2% in 2013) ensures that the majority of crimes reach courts and are therefore stored in the unified system of court records. However, one limitation is still in place. Least serious offences (petty disorderly conduct, traffic offences, loitering) are not criminalised and exist in the Administrative Code. Such offences, even though still adjudicated by courts, are not registered in the unified system and fall outside the scope of the court records on criminal prosecutions.

Obtained data The Institute for the Rule of Law at the European University at St. Petersburg was granted access to the universe of over 5 million depersonified court records from 2009–2013 by the Judicial Department of the Supreme Court of Russia for adult offenders. I relied on this information to construct an individual$\times$crime panel of defendants who were first brought before court *and* then recidivated in the said period.

Crucially, I consider not only those who were sentenced for the charged crimes but also those with dismissed charges (for rehabilitating or non-rehabilitating reasons).[2] The data are unusually rich in terms of observable characteristics of individuals and committed crimes which are listed in Table 2.1. I consider 1,058,870 crimes by 449,967 adult offenders in an unbalanced individual$\times$crime panel. This panel comes after trimming at 7 crimes per individual, a threshold that comprises over 99% of crimes by repeat offenders in the examined period. I also remove singleton observations (Correia, 2015) in terms of individual-region clusters.

It is important to note that since the data are organised in longitudinal individual$\times$crime observations, inference can be drawn both on individuals and their crimes. For instance, the mean value of variable "is 2nd crime" in panel A of Table 2.1 of 0.425 suggests that among those individuals who recidivated, 42.5% have committed up to two crimes in the period under study. This indicator is observed to be decreasing in crime number: only 3.1% of individuals who committed their first crime in 2009 have committed up to 4 crimes in the following five years. When it comes to panel B of Table 2.1 it is important to note that the aggregation is done at crime level. Such data organisation necessitates less conventional interpretation: a mean of 0.851 for the male dummy means that 85.1% of crimes in my individual$\times$crime panel are committed by male offenders, not that 85.1% of individuals in the data are male. The latter interpretation may be incorrect because

[2] Limiting the focus to convicted individuals only strengthens the findings of this chapter, as a robustness check in Figure 2.5b shows.

Variable	Mean	Median	SD	Min	Max
PANEL A: CRIME CHARACTERISTICS					
is 2nd crime	0.425	0	0.494	0	1
is 3rd crime	0.107	0	0.309	0	1
is 4th crime	0.031	0	0.172	0	1
is 5th crime	0.009	0	0.095	0	1
is 6th crime	0.003	0	0.053	0	1
is 7th crime	0.001	0	0.027	0	1
age at crime	30.831	29	9.896	18	88
charges per crime	1.199	1	0.562	1	5
under alcohol	0.278	0	0.448	0	1
under drugs	0.007	0	0.083	0	1
group crime	0.124	0	0.330	0	1
finished crime	0.933	1	0.250	0	1
crime day of year	179.988	179	104.158	1	366
PANEL B: OFFENDER CHARACTERISTICS					
male	0.851	1	0.356	0	1
higher education	0.059	0	0.236	0	1
high school education	0.920	1	0.272	0	1
married	0.208	0	0.406	0	1
resident	0.940	1	0.237	0	1
unemployed	0.700	1	0.458	0	1
worker	0.209	0	0.407	0	1
prisoner	0.005	0	0.072	0	1
student	0.021	0	0.142	0	1
office worker	0.021	0	0.144	0	1
official	0.004	0	0.065	0	1
top manager	0.006	0	0.078	0	1
entrepreneur	0.012	0	0.109	0	1
law enforcer	0.000	0	0.017	0	1
offence severity	24.219	23	14.334	1	100

TABLE 2.1: Summary statistics for 1,058,870 crimes by 449,967 offenders in an unbalanced panel of individuals who committed their first crime and recidivated in 2009–2013 in Russia. Panel A reports characteristics of crimes that the said individuals were charged with in the corresponding period. "is *i*-th crime" variables are dummy variables equal to unity when the crime in question was the *i*-th one. I consider only the primary article (most severe in terms of sanctions) of charges for each crime. "charges per crime" is the number of criminal actions the individuals were charged with. "under alcohol/drugs" is an indicator of a crime committed under alcohol/narcotic intoxication. "crime day of year" is an integer indicating the day of year when the crime was committed. Panel B reports the characteristics of individuals who were charged with the crimes, at crime-observation level. Socio-economic status variables are constructed as in Volkov (2016). The individual characteristics are recorded at crime dates. The data are trimmed upwards at 7 crimes per individual: this encompasses over 99% of repeat offenders.

of different drop-out rates for male and female repeat offenders. This may occur due to inherent gender differences in propensity to recidivate.

Data extensions I further enrich the court records data with socio-economic status of the accused based on recorded occupational characteristics using a procedure of Volkov (2016). Additionally, I attach geographic location to each of over 2,400 district/territory courts in the data.[3]

2.2. Measuring offence seriousness

With longitudinal data on criminal behaviour at hand, I proceed to construct the dependent variable of interest that reflects seriousness (severity) of the committed crimes.

2.2.1. An overview of approaches

I start by adopting instrumental view of the law as a price system for good and bad behaviours (Bénabou and Tirole, 2011, Gneezy and Rustichini, 2000). But where can this "price" be found?

Offence gravity in the Code Russia is a civil law country that codifies its crimes in federal law. Furthermore, Criminal Code (2012) attaches a category of gravity (seriousness) to every offence. This classification has four levels — low, medium, high, and top gravity — and has procedural implications when it comes to the rules for charge

[3] Results of this effort are reported at `http://atlasjustice.com`

dismissals or fast-track mode of trial and the repertoire of punishments that could be assigned. Low gravity offences are adjudicated by judges of peace and often proceed in private mode of prosecution wherein the victim bring the case directly to the judge.

Gravity classification in the Code — made on procedural rather than criminological grounds — might appear problematic when interpreted as a measure of offence seriousness. First, with mere four classes, it is too broad and does not capture potential differences in crime seriousness. For instance, grave sexual violence against a minor would fall into the same top gravity category with sexual violence against an adult person. Second, this grouping may be contaminated by the so-called "dormant" charges — charges that criminalise certain behaviour but are rarely enforced due to organisational constraints or low incidence of such behaviour in society. Russian Criminal Code has been amended over 1,500 times since its entry into force in 1996 (Esakov et al., 2017, Nikonov, 2017). Such sway of amendments will result in a substantial measurement error when interpreting gravity grouping as an indicator of seriousness. Despite those difficulties, some studies have relied on offence legislation to infer crime seriousness (Kyvsgaard, 2002, Burton et al., 2004).

Average custodial sentence lengths A number of studies (Gibbs, 1968, Wolpin, 1978, Reilly and Witt, 1996, Carrington et al., 2005, Carrington, 2013) have adopted an inductive approach to gauging severity. In those studies the authors measure average lengths of custodial (real incarceration) sentences for charges to arrive at a continuous indicator of severity. This metric, while producing a schedule of comparable seriousness scores for crimes, has two drawbacks. First, it is a conditional-on-incarceration measure. With a view to avoid an apples-and-oranges problem of comparing monetary fines, hours of correctional labour, and years of real incarceration, the average is computed over the latter sentences only. Second, given its conditional nature, it is either zero-inflated or produces missing results for charges that never result in real incarceration. When considering such charges a researcher can either exclude them from the analysis, exacerbating the conditionality problem and potentially biasing the results, or replace sentence lengths for non-carceral verdicts with zeros. Since real incarceration is not the major mode of disposition of criminal cases in any jurisdiction, including Russia (Skougarevskiy et al., 2014), this would cause a distribution of offence seriousness to be skewed to zero.

This chapter: Factorisation of convictions matrix Francis et al. (2005) propose a metric of offence seriousness that is rooted in correspondence analysis. They organise the decisions of the judiciary in a charge$\times$punishment type contingency matrix $\mathbf{N}$. Each cell of this matrix represents the number of individuals sentenced to a column punishment type for a row offence. The benefit of the contingency table approach comes from the fact that the columns can include a mix of punishment types (fines, mandatory labour, real or suspended incarceration) and discretised sentence lengths (e.g. zero to one year of real incarceration, one to three years, etc.). In other words, this contingency table carries information on the entirety of sentencing options for convicted individuals available to judges. Additional benefit of the approach is that it relies on aggregate data on convictions which are readily available in some jurisdictions. It is beyond the scope of this chapter to conduct a comparative study of offence seriousness across jurisdictions, but this is a research avenue that may hold much promise.

The constructed contingency table is then factorised with the aid of correspondence analysis (Greenacre, 2017). This procedure performs a singular value decomposition of an appropriately transformed contingency table to yield row (offence) scores that capture variation in punishment schedule and are scaled to belong to $[1, 100]$ interval. In this chapter I extend and depart from the procedure of Francis et al. (2005) in ways described in Supplementary appendix on page 101 to compute such row scores for Russia in 2013. Since I observe the universe of criminal cases adjudicated by the judiciary, my contingency table is rich, encompassing 508 offences (articles and article parts of Criminal Code) in rows and 11 punishment types in columns. I differ from Francis et al. (2005) in introducing of inference on the computed severity metric with parametric (Poisson) bootstrap (Ringrose, 2012) and square-root stabilisation of the underlying contingency table.

2.2.2. Obtained seriousness measurements

I now proceed to an examination of offence seriousness scores derived from correspondence analysis of a contingency table of offence$\times$punishment type conviction frequencies in 2013's Russia.

Table 2.2 lists top- and bottom-10 offences in terms of the estimated seriousness scores. The results in Panel A suggest that the chosen approach produces meaningful ranking of severity scores: top-10 offences are indeed crimes of top gravity related to homicide, death due to violent actions of sexual nature, organised crime, or distribution of drugs in extralarge quantities. The top seriousness of 100 is assigned to Article 132, part 5 of the Criminal Code that concerns violent sexual behaviour toward a minor victim of an offender who has been previously convicted of sexual violence. Interestingly, Article 132, part 4 that criminalises violent sexual behaviour toward a minor victim for a first-time sexual offender is estimated to have a severity of 77.3. While one should exercise caution when it comes to cardinal interpretation of estimated seriousness scores (Wagner and Pease, 1978, Liu et al., 2011) and not conclude that crime under Art. 132 p. 4 is 22.7% less severe than under Art. 132 p. 5, such scores offer a useful measure of relative position of crimes on the scale of offence severity.

When it comes to the bottom-ten offences, the ranking is also in line with intuition. The offence with minimum severity of 1 is poaching. Other bottom-10 offences either rest in the light gravity categorisation of the Criminal Code or relate to "dormant" charges that are rarely enforced. This may be indirectly inferred from the minimal standard deviation of such charges: when judges rarely assign punishments for a charge, column variability for its row in a charge$\times$punishment type contingency matrix would be tiny.

Offence	Article and charge name	Seriousness	St.dev
	TOP-10 OFFENCES		
13205	Violent Sexual Actions: by repeated sexual offender, towards minor	100	2.6
20901	Banditry	96.1	0.0
31700	Encroachment on the Life of an Officer of a Law-enforcement Agency	93.8	4.9
22914	Stealing or Exportation of Narcotic Drugs or Psychotropic Substances	90.5	5.1
10502	Aggravated murder	88.6	3.2
13104	Rape: leading to victim death / towards minor	87.1	5.2
32103	Disorganisation of the Activity of Institutions Providing Isolation from Society	81.0	5.9
20902	Banditry	78.7	9.4
13204	Violent Sexual Actions: towards minor / leading to victim death	77.3	5.2
13203	Violent Sexual Actions: towards minor / leading to grave injury	74.8	6.3
	BOTTOM-10 OFFENCES		
25601	Illegal Procurement (Catching) of Aquatic Biological Resources	1	0.0
13901	Violation of the Inviolability of the Home	1.1	0.0
32502	Theft or Damage of Documents, Stamps, and Seals or the Stealing of Excise Tax Marks, Special Marks or Marks of Conformance	1.1	0.1
24401	Outrages upon Bodies of the Deceased and Their Burial Places	1.2	0.3
32400	Acquisition or Sale of Official Documents and Government Awards	1.2	0.2
33001	Arbitrariness	1.2	0.1
30301	Falsification of Evidence	1.2	0.3
23601	Violation of Sanitary and Epidemiological Rules	1.3	0.4
29702	Contempt of Court	1.4	0.3
11501	Intentional Infliction of Light Injury	1.4	0.0

Table 2.2: Offences with largest or lowest estimated seriousness scores. This table shows top- and bottom-10 Russian Criminal Code offences by estimated seriousness in the [1,100] range (see page 101 for details on computation). "St.dev" is the standard deviation of the estimated seriousness. Numbers code offences: e.g. "10501" means Art. 105 p. 1 of the Russian Criminal Code in its 2013 edition, "Murder without aggravating circumstances".

This explanation is confirmed when I turn to Table 2.A.2 expressing top-10 offences in terms of their bootstrapped severity scores. Most of the charges there relate to bribery. Skougarevskiy (2014) documents large heterogeneity in bribe sizes by occupation (median size of a bribe received by a teacher or a doctor is reported at RUB2,000–3,000 [2013 USD60–90] whereas for law enforcement officers it is 1.5 times as large). Given such heterogeneity in bribe sizes, punishment schedules adjust accordingly which results in large standard deviation of the estimated severity scores.

It is also illuminating to study the distribution of the estimated offence severity by type of crime, offered in Figure 2.2. The taxonomy of types, defined in Skougarevskiy et al. (2014), seeks to group offences in a manner that maximises between-type variation while maintaining legally meaningful within-type variation. One can observe that ecological crimes have lowest mean and severity and limited within-group variability. In contrast, violent crimes span almost the entire interval of severity. The leading crime type in terms of mean seriousness score comprises sexual crimes.

Finally, when I consider the full profile of offence severity, it bears a resemblance to the family of Extreme value distributions. Figure 2.A.2 plots empirical versus theoretical seriousness quantiles from a maximum-likelihood-fitted Weibull distribution and shows that they are in close conformity. This allows me to argue that a theoretical study of crime seriousness should entertain a possibility of modelling it with the aid of the said family.

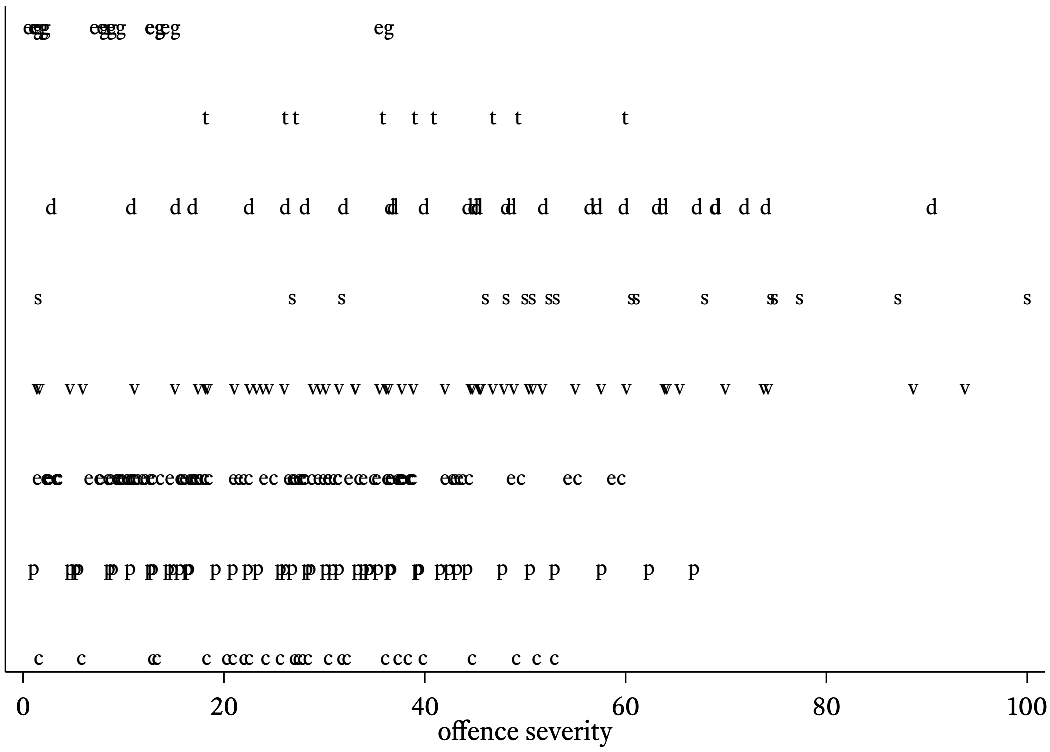


Figure 2.2: Distribution of estimated offence seriousness scores within and across crime types. Each letter position shows estimated offence seriousness while letter label signifies its crime group: **p** — property, **v** — violent, **ec** — economic, **s** — sexual, **eg** — ecological, **t** — traffic, **d** — drug, **c** — corruption. See Skougarevskiy et al. (2014) for details on crime grouping used.

2.3. Identifying escalation in seriousness

Having operationalised and estimated my seriousness variable, I now turn to modelling its evolution over time.

2.3.1. Model

A number of researchers of escalation have viewed its evolution over time as a Markov process (Rojek and Erickson, 1982, Britt, 1996, Osgood and Schreck, 2007). Operating in that framework, they compute offence switching probabilities for offenders with different observable characteristics and then contrast the obtained differences. Given the longitudinal nature of my data, I adopt a different approach due to Liu et al. (2011), Francis and Liu (2015). In this framework I parametrise offence seriousness of offender i committing his/her t-th crime as

$$\begin{aligned} \mathit{severity}_{i,t} = \alpha + \phi \sum_{j \in [1,7]} \mathbb{I}_{j=t} + \beta \mathbf{X}_i^{constant} + \gamma \mathbf{X}_i^{varying} + \delta \mathbf{X}_t^{crime} \\ + \kappa \sum_{j \in [1,7]} \mathbb{I}_{j=t} \times \mathbf{X}_{i,t} + \varepsilon_i + \nu_t + \xi_{i,t}, \end{aligned} \tag{2.1}$$

where $\mathbb{I}_{\bullet}$ is an indicator function equal to unity when its condition is satisfied and nil otherwise, $\mathbf{X}_i^{constant}$ is a matrix of observable individual-level characteristics that are highly unlikely to change over his/her criminal trajectory. Such characteristics typically include gender and other phenotype. $\mathbf{X}_i^{varying}$ is, in contrast, a set of observed individual-varying characteristics such as age at crime or employment. Finally, $\mathbf{X}_t^{crime}$ is a matrix of observable features of the committed t-th crime. They may include crime stage or preparedness. ε_i, ν_t, $\xi_{i,t}$ are unobservable individual-, crime-, or individual-crime-level characteristics, respectively.

Trend in severity is captured in (2.1) non-parametrically with the aid of dummy variables $\sum_{j\in[1,7]} \mathbb{I}_{j=t}$ which are equal to unity for respective crime orders. Parametrisation with dummies rather than polynomials of crime order allows for any empirical trend configuration. Since my panel is trimmed upward at 7 crimes, inclusion of dummies does not substantially reduce the degrees of freedom of the model. What does reduce the degrees of freedom, though, is the $\sum_{j\in[1,7]} \mathbb{I}_{j=t} \times X_{i,t}$ term. With this term, I allow severity trend to vary with every observable characteristic of individuals or crimes. This inflicts substantial damage on the degrees of freedom of the model which might make estimation without regularisation infeasible. However, since I operate in a large-N, small-T setting, examining the universe of repeat offenders for 5 years, conventional estimation still appears practical thanks to the richness of the data.

The key parameters of interest in the model are ϕ and κ. I now turn to their identification.

2.3.2. Estimation and identification

Random effects estimator To identify and estimate model (2.1), it is necessary to formulate assumptions on the unobserved characteristics. Liu et al. (2011) assume that

$\varepsilon_i, \nu_t, \xi_{i,t} \sim$ *Normal* and estimate the model with Feasible Generalized Least Squares (also known as "random effects"). Such random intercepts model is identified only when its exogeneity assumption is satisfied: $\varepsilon_i, \nu_t, \xi_{i,t} \perp X_{i,t}$.

This assumption might appear problematic in reality. Whenever unobserved individual characteristics (e.g. his proficiency in criminal career) are varying with the order of crime, ε_i is correlated with crime order dummies $\sum_{j\in[1,7]} \mathbb{I}_{j=t}$, rendering random effects model invalid.

Within estimator A "nuclear" option is to eliminate ε_i by using the Within-offender estimator (also known as "fixed effects"). While appealing in theory, this approach comes at a price of no point estimates or inference on $\boldsymbol{\beta}$ and $\boldsymbol{\kappa}$ for individual-constant characteristics. This is not desirable since my initial task is to conduct inference on escalation trajectories of different individuals given their crime-constant observable characteristics.

Mundlak estimator Inference on $\boldsymbol{\beta}$ and $\boldsymbol{\kappa}$ is possible without the troubling assumption of $\varepsilon_i, \nu_t, \xi_{i,t} \perp X_{i,t}$, Mundlak (1978) suggests (also known as "correlated random effects" estimator). It is beyond the scope of this chapter to provide a detailed discussion of this well-known estimator (its properties are given, *inter alios*, by Krishnakumar (2006)). In practice, the Mundlak estimator augments model (2.1) with individual-specific means of crime order-varying variables to account for the between-offender heterogeneity. Feasible GLS estimate of such augmented model is shown to be equivalent to the Within estimate, with the benefit of simultaneously identifying crime order-invariant offender-level coefficients β and κ.

Identification I adopt a kitchen-sink approach to identification of ϕ and κ. Severity escalation model (2.1) is anything but parsimonious: non-parametric escalation trend dummies are interacted with observables. To control for ε_i I use the Within-individual/Mundlak estimator while ν_t is subsumed by $\phi \sum_{j \in [1,7]} \mathbb{I}_{j=t}$ crime order dummies. However, it is crucial to note that **all inference is conditional on reoffending**. I do not investigate the decision to commit a new crime in this chapter, therefore selection-into-reoffending effects are still in $\xi_{i,t}$.

My $\mathbf{X}_i^{constant}$ includes gender, higher education or high school dummy, married indicator, resident indicator, and socio-economic status due to Volkov (2016). $\mathbf{X}_i^{varying}$ is a quadratic polynomial in offender age at crime. $\mathbf{X}_t^{crime}$ contains crime committed under the influence of alcohol or drugs indicator, group characteristic of crime, crime stage (finished or in preparation), role in crime (actual doer, organiser, accomplice), quadratic polynomials in number of charges for crime and number of crime day in year, consecutive crime half-year dummies. The latter two terms seek to account for seasonality of crime. Since I interact every observable characteristic with crime order dummies $\sum_{j \in [1,7]} \mathbb{I}_{j=t} \times X_{i,t}$ I have to be more parsimonious in the seasonality specification and cannot include crime day dummies. Lower moments of $\mathbf{X}_{i,t}$ are listed in Table 2.1 with summary statistics.

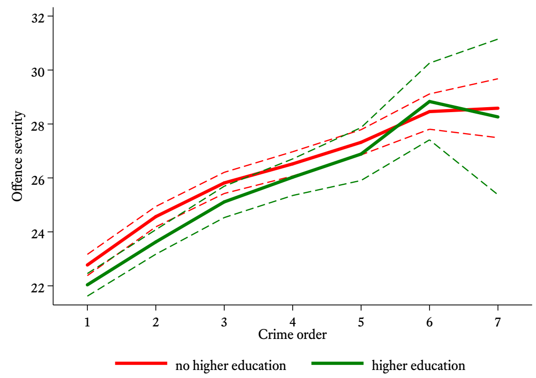
Inference Unless indicated otherwise, standard errors come from region-block bootstrap (there were 83 regions in 2009–2013 in Russia).

2.3.3. Results

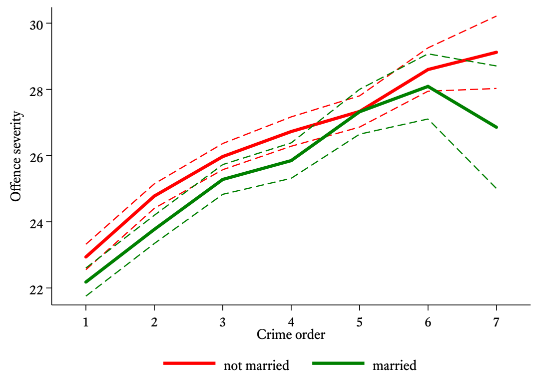
Since model (2.1) has hundreds of parameters, it makes little sense to present the results on $\hat{\phi}$ and $\hat{\kappa}$ directly.[4] Instead, I compute adjusted predictions of severity at representative values of observable regressors over the schedule of crime orders. Results of this exercise from the Mundlak estimator for mean offenders (i.e. when $X_{i,t}$ take their average values) that are different in one observable characteristic are presented in Figure 2.5.

Subfigures of Figure 2.5 contrast such observable characteristics of offenders or crimes as (A) education, (B) marriage, (C) crime committed under alcohol intoxication, (D) socio-economic status. Upward trend of offence severity with order of crime is observed in any mean characterisation, suggesting pronounced escalation, since the latter is the derivative of the severity trend line w.r.t. crime order. I find limited to no pacifying effect of higher education or marriage on escalation. In contrast, the state of alcohol intoxication at crime contributes to a pronounced pattern of escalation. Finally, individual's socio-economic status influences his/her escalation trajectory. Unemployed individuals are predicted to commit crimes with highest severity and that are becoming more and more severe. Entrepreneurs, in contrast, are found to start their criminal career by

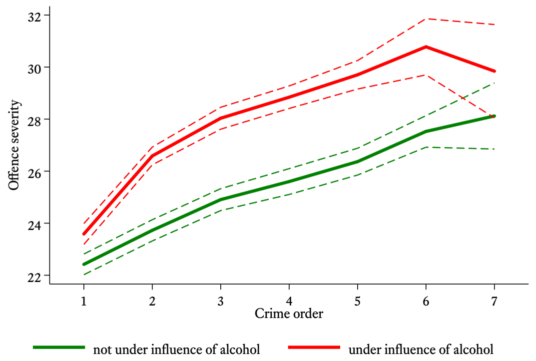
[4] For completeness, I do a variant of this in Table 2.A.2 with regression results.



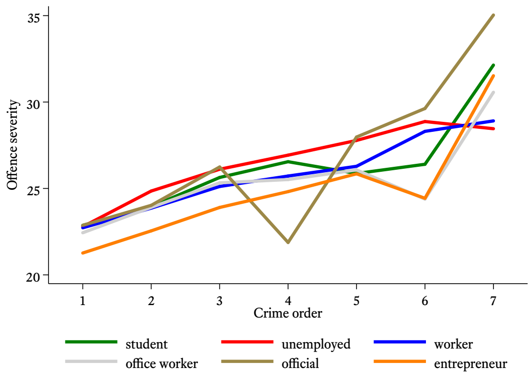
(A) Higher education



(B) Marriage



(C) Alcohol



(D) Socio-economic status

Figure 2.3: Adjusted predictions from Mundlak (1978)-estimated (2.1) for mean offenders that are different in one observable characteristic: presence of higher education (panel A), marital status (panel B), commission of crime under the influence of alcohol (panel C), and socio-economic status of the offender (panel D, confidence intervals not reported). 90% region-block bootstrapped CI are reported as dashed lines. Note: figures have different y-axis scale.

committing crimes of lowest severity. Wiggly behaviour of predicted severity for certain types of offenders (officials or entrepreneurs) can be attributed to their limited presence in the data: officials committed only 0.4% of crimes in the studied panel.

Age profile Age at crime is known to be one of the most potent drivers of criminal behaviour (Farrington, 1986). This chapter finds that same can be stated when it comes to offence severity. Figure 2.4 plots escalation trajectories of mean offenders with different age at crime onset. 18-year-old individuals are predicted to have the largest severity of their first crime whereas 30-year-old offenders are projected to commit the first crime with lowest severity. This age-severity relationship manifests itself in a spectral gradient of respective severity curves in Figure 2.4. This finding is consistent with existing literature. In particular, Liu et al. (2011, p. 180) state that change in severity with age "can be thought of as a developmental or maturational process." However, this statement does not apply to escalation since the slopes of all age curves are estimated to be similar. This is a novel finding that sheds light on the reoffending behaviour with age over the course of criminal careers.

2.3.4. Robustness checks and external validity

Uncertainty in measurements of offence severity My novel measure of offence severity is shown to exhibit behaviour in line with criminological intuition and offence gravity categorisation in the Criminal Code in Subsection 2.2.2. However, factorisation

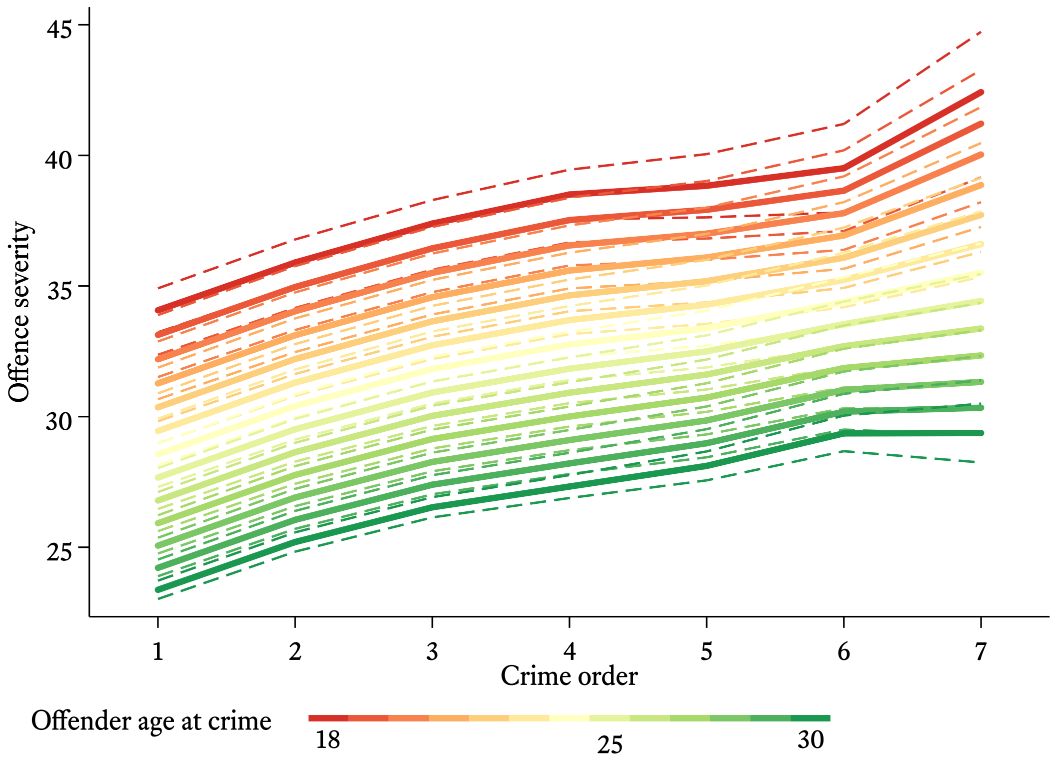
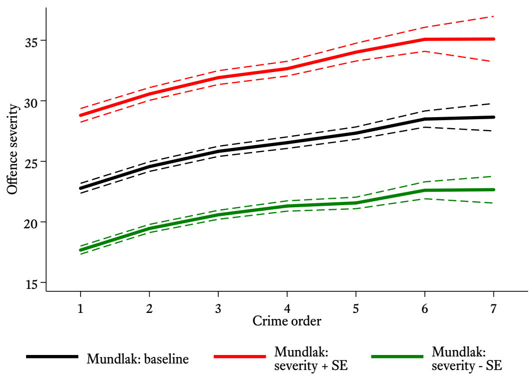


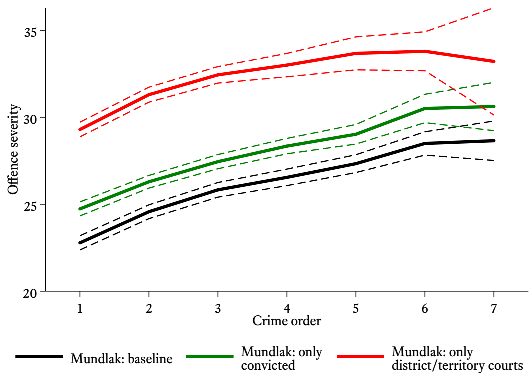
FIGURE 2.4: Adjusted predictions from Mundlak (1978)-estimated (2.1) for mean offenders that are different in age at crime. Spectral gradient reports predicted escalation trajectories for individuals with age at crime in [18,30] interval. 90% region-block bootstrapped CI are reported as dashed lines.

of offence$\times$punishment type contingency table of convictions might appear problematic for rows with a limited number of convictions, or the so-called "dormant" charges. In such settings even a unit change in a number of convicted individuals under a "dormant" offence would vastly increase the standard deviation of the row score from the correspondence analysis, as parametric bootstrap (Ringrose, 2012) shows. In Figure 2.5a I compute adjusted predictions for a mean offender when my offence $severity_{i,t}$ is shifted by one bootstrapped standard deviation of its row score upward (red line), downward (green line), or remains at the baseline estimated level. This exercise demonstrates that the effect of uncertainty on $severity_{i,t}$ is additive rather than multiplicative. Predicted severity curves are shifted upward or downward, in line with the nature of this exercise, but do no appear to change their slopes, signalling that escalating behaviour is robust to uncertainty in measurement of severity.

Alternative definition of individuals under study The data that form the backbone of this study contains the information on the universe of the defendants on criminal cases. My baseline estimate capitalises on the richness of this information by including all charges adjudicated by courts in the model. This means that I treat case dispositions not resulting in convictions as crimes. For instance, when a case for an individual was dismissed for non-rehabilitating reasons (e.g. due to reconciliation with the victim, Smirnov and Kalinovskiy (2012)) it is still included in the panel as an event of crime. The reason for treating such dispositions as crimes is two-fold. First, they amount to a large



(A) Uncertainty in severity scores



(B) Only convictions or crimes processed by territory or district courts

Figure 2.5: Adjusted predictions from Mundlak (1978)-estimated (2.1) for a mean offender. Panel A shows the baseline predictions (black line) in relation to the predictions when $severity_{\bullet,\bullet}$ is replaced with $severity_{\bullet,\bullet} \pm \mathrm{SE}\,(severity_{\bullet,\bullet})$ where SE $(\bullet)$ is the standard error of the estimated seriousness scores (see page 101 for details on computation and Table 2.A.2). Panel B shows the baseline predictions (black line) in relation to predictions where I consider as observations only instances of sentences and do not treat dismissals (for non-rehabilitating reasons) as crimes (green line). Red line in panel B reports the predicted offence severity trajectory when I consider only cases adjudicated by district or territory courts and exclude judges of peace from consideration. 90% Huber-Eicker-White region-clustered CIs are reported as dashed lines. Note: figures have different y-axis scale.

number of outcomes (in 2009 19.3% of cases resulted in reconciliation, Skougarevskiy et al. (2014)). Second, their omission is hypothesised to bias the predicted severity upward since 47.0% of reconciliation outcomes concerned cases of private prosecution that deal with misdemeanours.

In Figure 2.5b I display predicted severity for mean offender when non-convictions are included in the model (my baseline estimate, black line) or when the model is estimated on a panel of convictions only (green line). As expected, the latter model produces larger estimates of the level of offence seriousness but does not alter the slope of the severity curve. Finally, I limit the sample to the point of excluding all charges adjudicated by judges of peace (that deal with misdemeanours) from my panel. The resulting severity is the red line in Figure 2.5b. Unsurprisingly, limitation of the data to felonies drives the severity curve upward but still results in a positive slope of severity trend. This slope, however, is less steep in relation to the baseline specification with all dispositions and charges adjudicated by judges of peace. I will investigate this behaviour in the following section.

External validity of findings Estimator of escalation presented in this chapter achieves high degree of internal validity due to its focus on the universe of offenders and the modelling abstraction from unobserved crime order- or offender-invariant heterogeneity in severity. However, two problems might hinder generalisability of the findings. First, my panel is short as I follow the individuals for 5 years only. Second, it is estimated in a jurisdiction previously unknown to quantitative scholars of crime. This requires establishing external validity of the findings with extraneous information. To this end, I replicate the analysis of Liu et al. (2011) on England and Wales Offenders Index 1953 birth cohort sample that was followed from 1963 till 1999. I bring the formulations of model (2.1) as close as possible given the differences in the data and report the Within estimates of predicted severity for a mean offender in Russian and UK data in Figure 2.6. Even though $severity_{i,t}$ is measured differently in the UK data (log transformation after correspondence analysis) and my Russian data (square-root transformation before correspondence analysis), the two metrics can be compared in terms of the predicted slope, i.e. escalation. The estimators lie remarkably close in that regard, with Russia's within-offender severity trajectory estimated more efficiently thanks to the large-N nature of the data. This allows me to conclude that my findings exhibit external validity.

2.4. De-escalation and incapacitation

2.4.1. Within- vs. Between-offender seriousness and incapacitation

So far my estimates of severity have reflected its within-individual trajectory. In other words, they show the expected change in an offender's crime seriousness if the number of offences s/he commits increases by 1. Another policy-relevant metric is the expected difference between crime seriousness of offender A and offender B if they differ

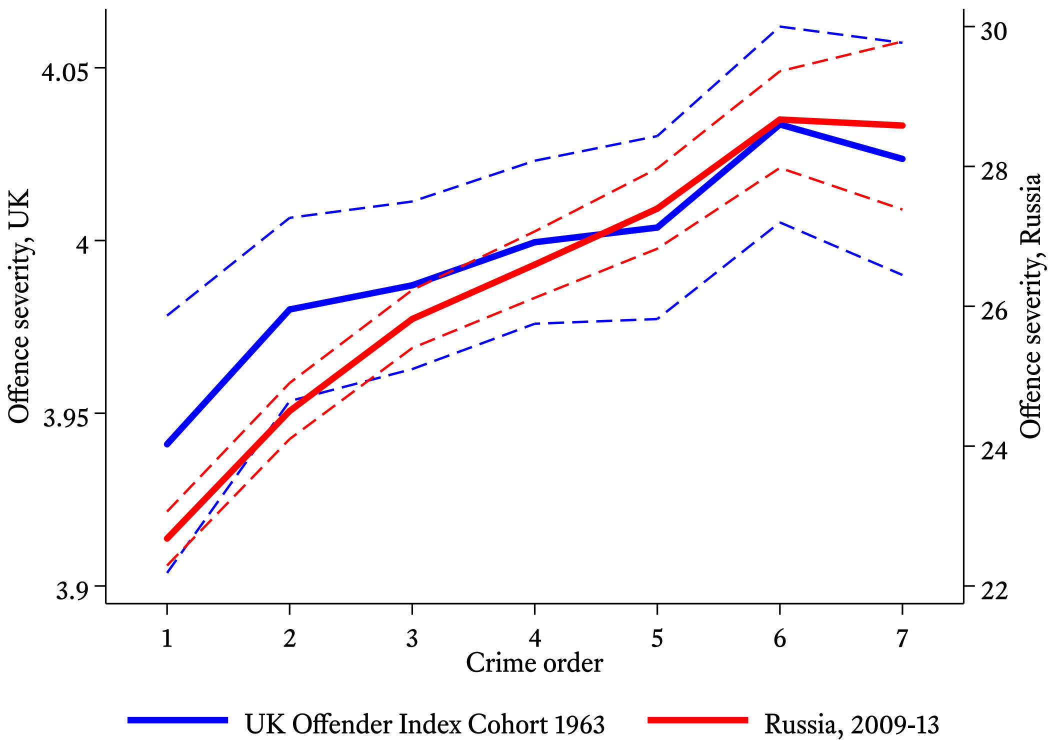


Figure 2.6: Adjusted predictions from the Within-estimated (2.1) for UK (blue line) / Russia (red line) for a mean offender. Dashed lines show Huber-Eicker-White 90% robust (UK) or region-block bootstrapped (Russia) CIs. I use the data on 4,396 offenders from England and Wales Offenders Index 1953 birth cohort sample followed till 1999 (Francis et al., 2005, Liu et al., 2011, Francis and Liu, 2015). Control variables for UK data include quadratic polynomials in age and offence number, and number of charges per crime. Offence seriousness is estimated by Francis et al. (2005) with the same procedure as on page 101 on 2001 contingency table of convictions. However, instead of pre-transforming the **N** with square root, Francis et al. (2005) log-transform the resulting **F** row scores to stabilise the estimates. Note: scales of y-axes are different.

in number of offences by 1. This metric is given by the Between estimator which entails running a regression of (2.1) on individual-level means of the covariates. Policy interest aside, random effects estimator is a weighted average of the Within and Between estimators (Baltagi, 2008), therefore an exploration of the predicted severity trajectory for all the estimators might help to better understand the nature of unobserved heterogeneity within and between offenders.

Figure 2.7 reports predicted mean severity over crime order for a battery of estimators. In line with theoretical predictions, Mundlak estimator (green line) produces results nearly identical to the Within-offender estimator (red line). Slight discrepancy between the two estimators is attributed to the fact that $\mathbf{X}_i^{constant}$ includes the observables that can be individual-varying. Random effects estimator (orange line) offers a severity trajectory that is larger than the Within estimate for the first two crimes and noticeably smaller for the 3rd–7th crimes. This is due to the fact that the Between-offender estimator (blue line) and the Within-estimator yield different seriousness trend dynamics.

Two reasons can explain the mirror trajectory of the Within and Between estimators. First, it might be a mere artefact of a short-T panel used in this study. One could argue that since offenders were followed for 5 years only, the time frame is not enough to identify the between-individual effect of a unit change in the crime order. However, external validity check in Subsection 2.3.4 allows to refute this explanation since the results are in line with estimates from a medium-T panel of individuals followed for 37 years.

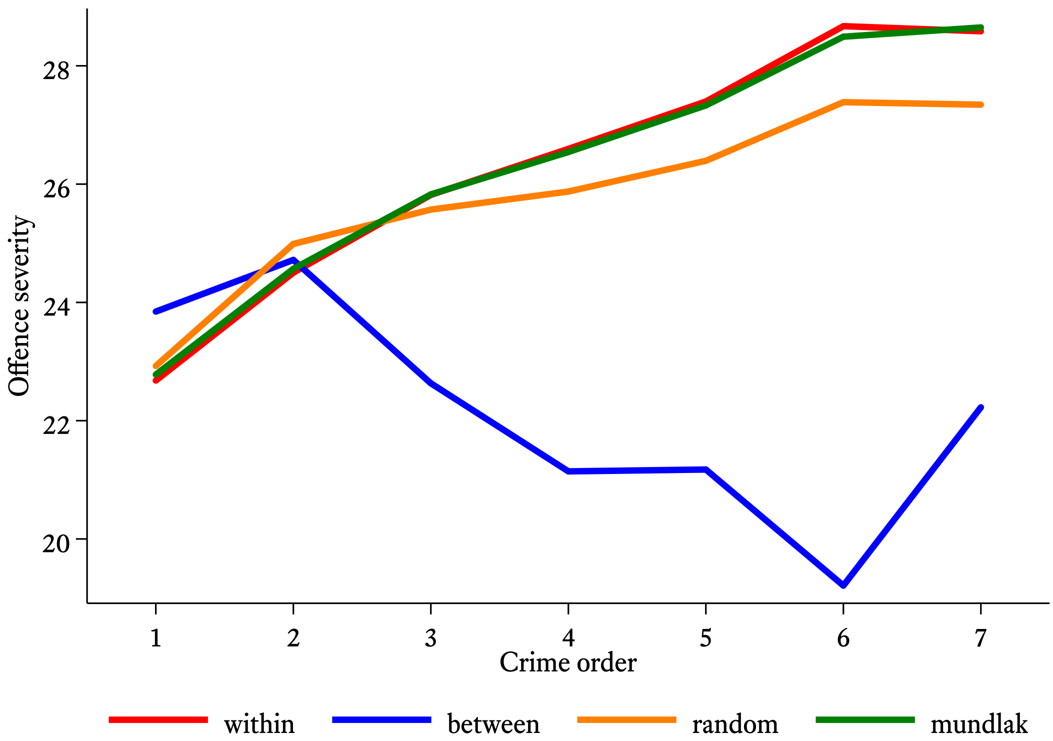


FIGURE 2.7: Adjusted predictions from the within-, between, or random effects-estimated (2.1) for a mean offender, CIs not reported.

Second reason relates to incapacitation. Offenders with more offences commit less serious crimes because they are selected into longer custodial sentences and, therefore, cannot commit serious crimes by virtue of incapacitation, the argument goes. Indeed, decreases in offence seriousness are often explained by termination in offending behaviour (Le Blanc, 2014). This observation complicates my analysis and warrants a change in the modelling strategy.

2.4.2. Is de-escalation prompted by incapacitation?

I propose a dynamic model

$$\begin{aligned} \textit{severity}_{i,t} = \alpha + \rho \textit{severity}_{i,t-1} + \phi \sum_{j \in [1,7]} \mathbb{I}_{j=t} + \beta \mathbf{X}_i^{\textit{constant}} \\ + \gamma \mathbf{X}_i^{\textit{varying}} + \delta \mathbf{X}_t^{\textit{crime}} + \varepsilon_i + \nu_t + \xi_{i,t}, \end{aligned} \tag{2.2}$$

where $\textit{severity}_{i,t-1}$ contains the severity of the previous crime committed by offender i and other terms are defined as in model (2.1). The idea behind inclusion of crime-lagged severity term is that it captures the incapacitating effect of the previous crime seriousness. I expect ρ to be negative since individuals who commit more severe crimes are assigned higher sanctions and penalties for their behaviour which are more likely to amount to real incarceration. Whilst incarcerated, such individuals are less likely to commit more severe crimes in the future.

The lagged term in (2.2) gives rise to the familiar Nickell (1981) bias in ρ when it comes to the Within-individual estimator. Elimination of ε_i through demeaning automatically creates a correlation between $severity_{i,t-1}$ and the value of ε_i, violating the exogeneity assumption. The true negative ρ will be biased in the Within estimation that does not account for the created endogeneity problem. In columns (1) and (2) of Table 2.3 I report that the Within and Pooled estimates bound $\hat{\rho}$ in $[-.386, .217]$ interval. Since the Pooled estimate suffers from the upward Hurwicz (1950) bias present in any dynamic regression problem, additional evidence is required to infer the true ρ.

Approaches relying on internal instruments

Anderson and Hsiao (1982) Popular treatment of the problem of Nickell bias involves estimating (2.1) on first differenced data to eliminate ε_i and instrumenting the lagged differenced regressor $\Delta severity_{i,t-1}$ with internal instruments based on levels of the past lag $severity_{i,t-2}$. This is the approach of Anderson and Hsiao (AH) levels estimator. This approach relies on the allegedly exogenous behaviour of the second lag of the severity to identify the influence of its first lag on severity. Due to the fact that the instrument is based on the data it is often referred to as an "internal" instrument.

Arellano and Bond (1991), Blundell and Bond (1998) In medium-to-large-T panels consistent estimation can be achieved by inclusion of additional differenced lags of $severity_{i,t-h}$, where $h \geq 2$, as instruments and the Generalised Method of Moments (GMM) estimation. This is the proposition of Arellano and Bond (AB) differences estimator. Blundell and Bond (1998) (BB) suggest that further efficiency gains can be

	(1) Within	(2) Pooled	(3) AH	(4) AB	(5) BB	(6) Everaert	(7) Reanalysis weather	(8) Actual weather
$\hat{\rho}$	-.386	.217	.050	.052	.056	-.342	-.338	-.333
	(.006)	(.008)	(.012)	(.010)	(.016)	(.015)	(.408)	(1.038)
							FIRST STAGE	
$\hat{\omega}^{tempe-}_{rature}$							-.009	-.004
							(.008)	(.019)
$\hat{\omega}^{precipi-}_{tation}$							-.544	-.551
							(.294)	(.654)
F-stat	—	—	3717.4	—	—	8551.4	2.5	0.7
Observations	225,037	225,037	43,872	43,872	225,037	225,037	225,037	46,056

Table 2.3: ρ in (2.2) under different estimators, region-clustered standard errors reported in parentheses. Only instances of individuals sentenced at district or territory courts (not judges of peace) are included. This is due to no locational information on judges of peace that renders $\rho^{Reanalysis|Actual\ weather}$ estimation unfeasible. ρ^{Within} is the Within OLS-estimated value that controls for unobserved crime order-invariant heterogeneity across offenders. ρ^{pooled} is a value from Pooled OLS estimator. ρ^{AH} is estimated value from Anderson and Hsiao (1982) estimator with lagged levels of $severity_{i,t-2}$ as instruments. ρ^{AH} comes from Arellano and Bond (1991) difference GMM estimator. ρ^{BB} is given by applying Blundell and Bond (1998) System GMM estimator. $\rho^{Everaert}$ is estimated value from Everaert (2013) orthogonal to backward mean transformation-based estimator. $\rho^{Reanalysis\ weather}$ comes from two-stage least squares estimated (2.3) where $severity_{i,t-1}$ is instrumented with weather conditions at the crime location at its date from ECMWF-ERA Interim global atmospheric reanalysis data (Dee et al., 2011). Used weather conditions are daily mean 2-metre surface temperature and total daily precipitation at 0.25° grid. $\rho^{Actual\ weather}$ is from 2SLS-estimated (2.3) where $severity_{i,t-1}$ is instrumented with actual weather conditions in courts within 20km radius of weather stations as provided by Veselov and Pribylskaya (2015). "F-stat" is Kleibergen and Paap (2006) rk Wald F statistic of the first stage. $\omega^{precipitation|temperature}$ is the first-stage coefficient on excluded instrument from (2.3).

achieved by inclusion of levels of $severity_{i,t-h}$ as instrumental variables.

Difference/System GMM approaches are problematic in unbalanced panels where they magnify the gaps; weak internal instruments lead to finite sample bias. What is more, all of the above estimators rely on the assumption of the instrumental orthogonality of the first difference. This assumption might be incredible in practice. If three crimes are committed in a short period of time, the severity of the earliest crime might affect the severity of the latest crime directly, and not exclusively through the severity of the second crime.

Transformation-based approaches An alternative strand of the literature seeks to identify a transformation of (2.1) that eliminates ε_i and does not introduce the Nickell bias. Examples include forward orthogonal differences of Arellano and Bover (1995), X-differencing of Han et al. (2014), or orthogonal to backward mean transformation of Everaert (2013).

Columns (3)–(5) of Table 2.3 report $\hat{\rho}$ from AH, AB, and BB estimators, respectively. I observe that they produce almost identical estimates of *positive* ρ. Such results defy the incapacitation explanation suggesting that larger severity of previous crime contributes to larger severity of the current crime. However, when I entertain a transformation-based approach of Everaert (2013) in column (6), the resulting $\hat{\rho}$ is negative, in line with the incapacitation theory. Estimators of (2.1) are in stark disagreement regarding $\hat{\rho}$.

Instrumentation with weather conditions

Facing conflicting evidence on ρ, I next rely on extraneous variation in offence severity to aid identification. This variation comes from weather conditions at the location of crime on its date. Rational choice theory or routine activities theory both suggest that criminal behaviour is influenced by extraneous factors (Cohn, 1990). Empirical research has found the weather to be a significant factor for certain crime types (Baron and Bell, 1976, Perry and Simpson, 1987, DeFronzo, 1984, Cohn, 1990, Anderson et al., 1995, Jacob et al., 2007, Butke and Sheridan, 2010, Ranson, 2014). I rely on the daily mean 2-metre surface temperature and total daily precipitation at 0.25 grid ($\sim 28 \times$ 28km) at the adjudicating court location at crime date from the European Center for Medium-range Weather Forecasting ERA-Interim global atmospheric reanalysis model (Dee et al., 2011). The severity model then becomes

$$\begin{aligned} severity_{i,t} &= \alpha + \rho \widehat{severity}_{i,t-1} + \phi \sum_{j \in [1,7]} \mathbb{I}_{j=t} + \beta \mathbf{X}_i^{constant} \\ &\quad + \gamma \mathbf{X}_i^{varying} + \delta \mathbf{X}_t^{crime} + \varepsilon_i + \nu_t + \xi_{i,t} \\ severity_{i,t-1} &= \omega \left[temperature_{i,t-1}, precipitation_{i,t-1} \right] + \Xi \mathbf{X}_{i,t} + \varsigma_{i,t} \end{aligned} \tag{2.3}$$

This equation reportes second and first stage of the two-stage least squares estimation (after the Within-transformation).

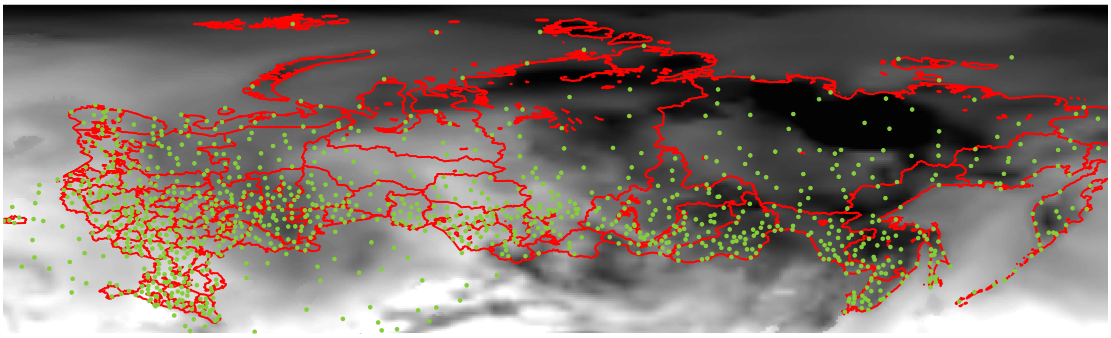
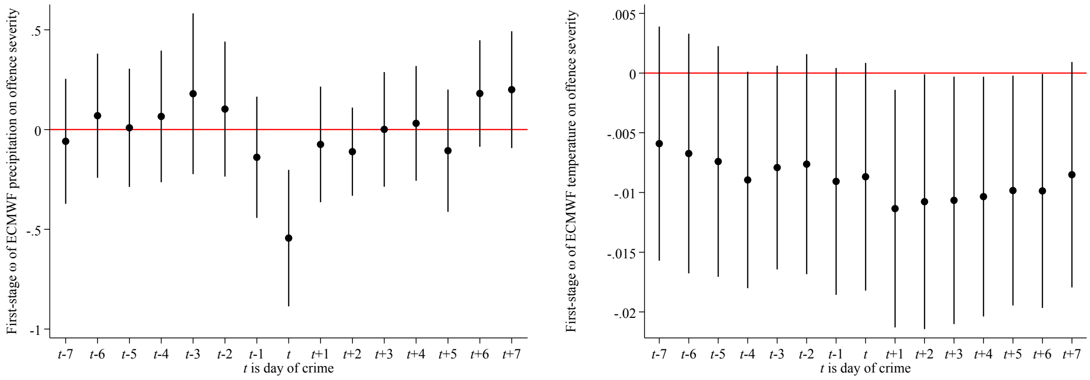


Figure 2.8: This figure shows predicted 2 metre surface temperature gradient from the ERA-Interim global atmospheric reanalysis (Dee et al., 2011) at 0.25° grid over Russia and neighbouring countries, on October 3, 2014, 12:00 UTC. Darker shades indicate grids with lower temperature. Red lines show Russian regional and water boundaries. Green dots are weather stations from Veselov and Pribylskaya (2015).

Weather data sources Most studies exploiting variation in weather rely on observed conditions at weather stations. This is problematic in case of Russia. Figure 2.8 shows locations of weather stations in ex-USSR countries as green dots (Veselov and Pribylskaya, 2015). It can be immediately seen that the network of weather stations is limited in its territory coverage, leaving many habitable locations outside its scope. Global Reanalysis models, in contrast, use observed weather conditions as inputs to a climate model that interpolates the conditions on a fine grid (Dee et al., 2011, Auffhammer et al., 2013). The benefit of the latter approach comes from temporal and spatial scale of the model that relies on a worldwide network of stations to produce present-time weather conditions from observed information.

Instrument validity and relevance To establish validity of the proposed instrumental variable, in Figure 2.9 I perform a placebo test on the first-stage $severity_{i,t-1}$ model. In that test I shift the weather conditions in past or future, thereby regressing lagged offence seriousness on weather conditions on days preceding or following the actual crime date. Subfigure 2.9a reports the schedule of estimated $\hat{\omega}$ with shifted precipitation information. It turns out that the only date when precipitation affects crime seriousness (at 99% level of statistical significance) is the date of crime, giving credence to the proposed identification strategy. Interestingly, the results of the same placebo test for temperature (Subfigure 2.9b) do not display a significant relationship between surface temperature and offence severity. This is the opposite of what Ranson (2014) established in the US. Nevertheless I use both precipitation and temperature as excluded instruments because they are likely to be correlated, and omission of the latter might induce omitted variable bias (Auffhammer et al., 2013).

Results on ρ Column (7) of 2.3 reports the results of 2SLS-Within estimated model (2.3) using the weather conditions from reanalysis data. $\hat{\rho}$ is now much closer to the Within/Everaert-estimated value and is negative, confirming the incapacitation theory. Inference on it, however, appears to suggest that ρ is not significantly different from zero. I posit that this finding should be interpreted with caution: I am using the Within-offender model with rich seasonality controls (quadratic polynomials in day of year and half-year dummies) in $X_{i,t}$. All this limits the observable variation in weather conditions and renders inference more challenging. Furthermore, since I trace the universe of offenders, hypotheses tests relate to the external validity of the findings whereas their



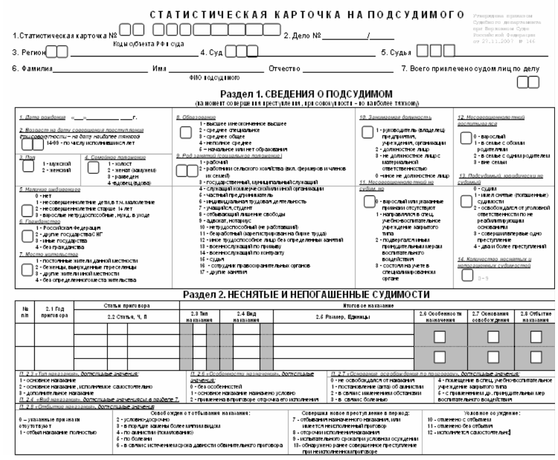
(A) Precipitation **(B) Surface temperature**

Figure 2.9: Placebo test: using weather conditions from days preceding or following the date of crime. Estimated $\hat{\omega}$ from the first stage of (2.3) where $severity_{i,t-1}$ is instrumented with weather conditions at the crime location from ECMWF-ERA Interim global atmospheric reanalysis data (Dee et al., 2011). Panel A reports $\hat{\omega}$ for daily mean 2-metre surface temperature, Panel B reports $\hat{\omega}$ for total daily precipitation at 0.25° grid. Dots are point estimates whereas lines show 99% Huber-Eicker-White region-clustered CIs. Note: figures have different y-axis scale.

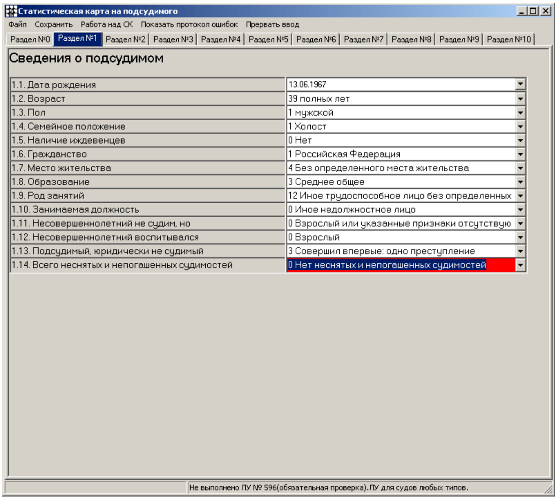
internal validity is satisfied by virtue of studying the population.

Finally, in column (8) I perform the analysis using observed conditions at the weather stations from Veselov and Pribylskaya (2015). Due to sparsity of its network, the number of observations displays almost a five-fold drop. Nevertheless, the point estimate of $\hat{\rho}$ is virtually unchanged. Its standard error more than doubles, which can also be attributed to limited observable variation in weather conditions after controlling for within-offender heterogeneity and seasonality in crime seriousness.

Supplementary appendix



(a) An excerpt from court statistical card on the accused in paper form



(b) Early programme interface with an excerpt from court statistical card

Figure 2.A.1: This figure exhibits excerpts from statistical cards on the accused under criminal cases filled by clerks of Russian courts. Panel a shows the paper version whereas panel b displays an interface of a programme (circa 2009) that was developed to digitise court record data entry operations. Source: Judicial Department at the Supreme Court of Russian Federation.

Procedure for calculating offence severity scores

1. Let $\mathbf{N}$ be a 508 offences $\times$ 11 punishment types count matrix for convictions in 2013's Russia:

	Fine	Mandatory works	Limitation of freedom	Suspended incarceration	Correctional works	Real incarceration, years					
						<1	1–3	3–5	5–10	10–20	>20
Murder (10501)	0	0	0	65	0	3	18	179	4651	1379	4
Aggravated murder (10502)	0	0	0	6	0	0	1	11	161	920	189
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Theft (15801)	17023	10818	1803	11675	8006	4871	5026	719	107	10	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Note: life sentences are top-coded at 30 years, only suspended incarceration punishments are considered.

2. Following Greenacre (2009), power-transform $\mathbf{N}$ by taking element-wise square roots to stabilise the variance of the contingency table (Bartlett, 1936).

3. Obtain the correspondence matrix $\mathbf{P} = \frac{1}{n}\mathbf{N}$, where n is the grand total. Let $\mathbf{r}$ and $\mathbf{c}$ be row/column marginal totals of $\mathbf{P}$.

4. Construct $\mathbf{D}_r$, $\mathbf{D}_c$ — diagonal matrices of $\mathbf{r}$ and $\mathbf{c}$.

5. Calculate the matrix of standardised residuals: $\mathbf{S} = \mathbf{D}_r^{-\frac{1}{2}} (\mathbf{P} - \mathbf{rc}') \mathbf{D}_c^{-\frac{1}{2}}$.

6. Calculate its singular value decomposition: $\mathbf{S} = \mathbf{U}\mathbf{D}_\sigma\mathbf{V}'$, where $\mathbf{U}'\mathbf{U} = \mathbf{V}'\mathbf{V} = \mathbf{I}$.

7. Obtain principal coordinates of rows as $\mathbf{F} = \mathbf{D}_r^{-\frac{1}{2}}\mathbf{U}\mathbf{D}_\sigma$, scale them in 1–100 range and declare them as measures of offence seriousness.

8. Extending Francis et al. (2005), perform 1000 replications of parametric (Poisson) bootstrap of steps 1–7 to compute standard deviations of the principal coordinates $\mathbf{F}$ (Ringrose, 2012).

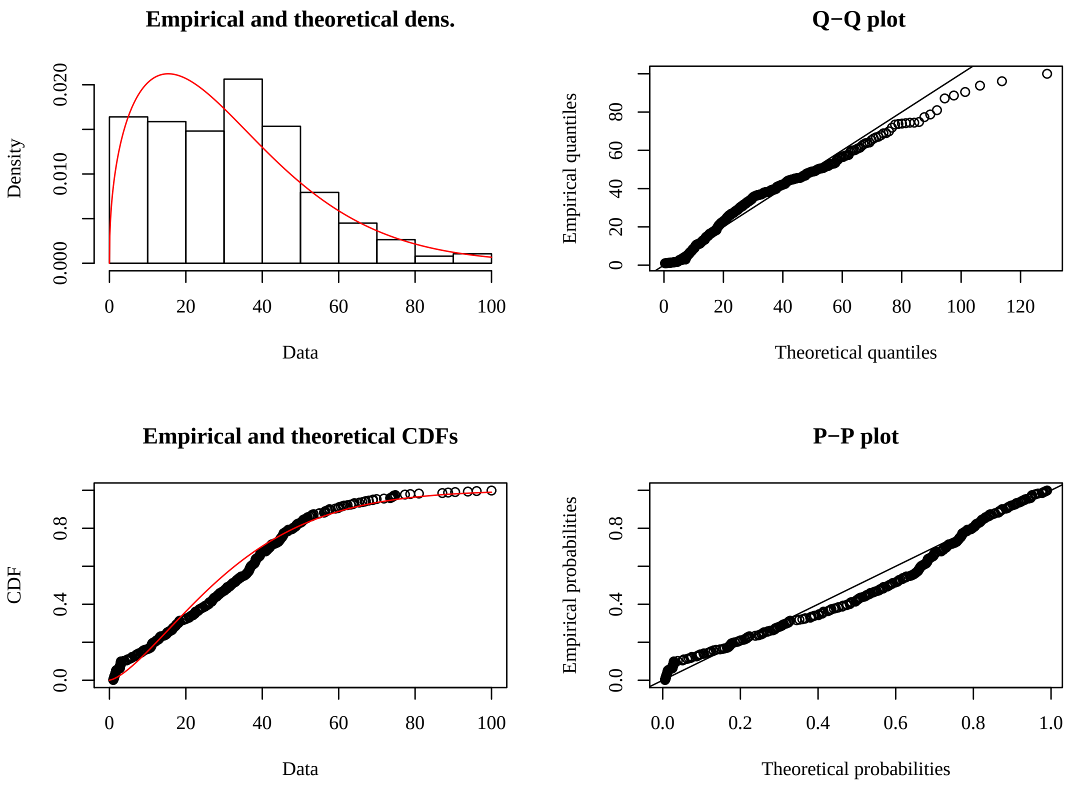


Figure 2.A.2: Empirical (black bars and dots) and fitted Weibull (red lines) distribution of estimated offence seriousness of Russian Criminal Code charges. Full distribution is given in Figure 2.A.3.

Offence	Article and charge name	Seriousness	St.dev
20402	Bribery in a Profit-Making Organisation	30.4	9459.8
28502	Abuse of Official Powers	30.4	4578.4
15923	Swindle related to Social Payments	30.1	347.4
19600	Deliberate Bankruptcy	30.9	300.3
19902	Evading Payment of Taxes (large-scale)	29.9	263.0
31801	Use of Violence Against a Representative of the Power	30.1	253.4
23802	Violation of Safety Standards Regulation in production	30.8	165.2
29113	Mediation in Bribery	32.2	135.8
16003	Embezzlement (large-scale)	30.8	126.3
29005	Bribe-Taking (large-scale)	31.7	125.9

Table 2.A.1: Offences with largest bootstrapped seriousness score standard deviations. This table shows top-10 Russian Criminal Code offences by the standard deviation of the estimated seriousness (see page 101 for details on computation). Numbers code offences: e.g. "10501" means Art. 105 p. 1 of the Russian Criminal Code in its 2013 edition, "Murder without aggravating circumstances".

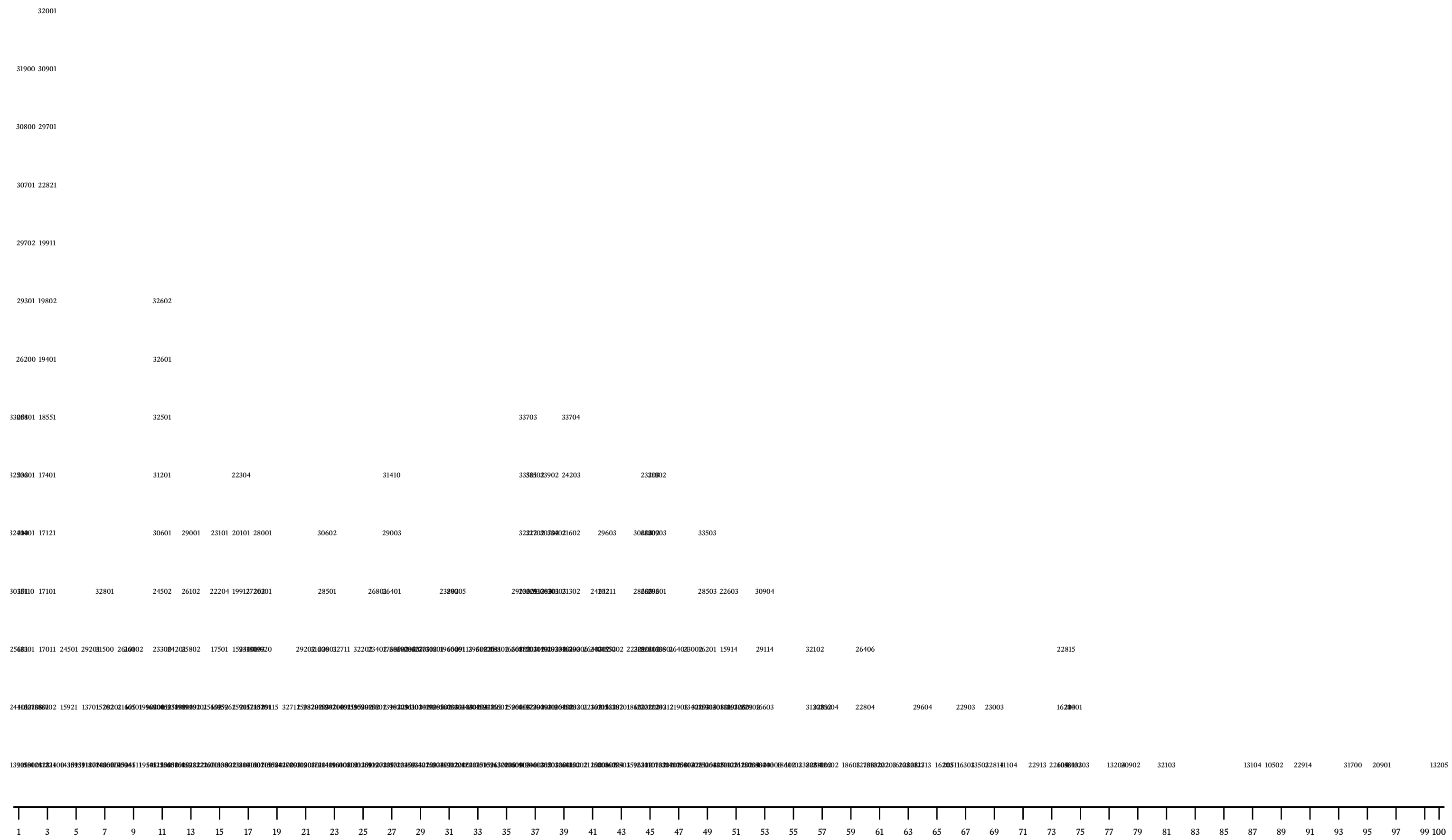


FIGURE 2.A.3: Stripplot of estimated offence seriousness of Russian Criminal Code charges. x-axis reports the estimated offence seriousness in the [1,100] range that was rounded to the nearest .5. Offences with similar rounded seriousness scores are stacked. Numbers code offences: e.g. "10501" means Art. 105 p. 1 of the Russian Criminal Code, "Murder without aggravating circumstances".

Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Estimated offence severity on [1,100] scale								
Crime order									
is 2nd crime	-0.609** (0.306)		-0.903*** (0.262)	-0.489 (0.382)	-0.489 (0.375)	-0.937 (0.576)	0.014 (0.417)	-1.475*** (0.495)	-0.184 (0.564)
is 3rd crime	-0.772 (0.707)	-4.162*** (1.469)	-1.262* (0.748)	-1.494 (0.963)	-1.494* (0.875)	-2.949** (1.245)	-0.665 (0.818)	-1.256 (0.847)	-0.335 (0.883)
is 4th crime	1.124 (1.010)	-3.074 (3.757)	0.499 (0.850)	-1.346 (1.095)	-1.346 (1.175)	-2.158 (1.643)	-1.256 (1.186)	-2.985** (1.411)	-0.490 (1.755)
is 5th crime	-0.358 (1.979)	-15.788* (8.703)	-1.677 (1.813)	-2.790 (2.448)	-2.790 (2.058)	-7.914** (3.155)	0.373 (1.963)	-2.074 (2.874)	-2.535 (3.328)
is 6th crime	-6.105** (2.697)	-18.899 (18.160)	-7.667*** (2.673)	-11.803*** (3.393)	-11.803*** (3.483)	-18.396*** (4.627)	-7.317* (3.935)	-9.700** (3.880)	1.275 (6.171)
is 7th crime	9.409 (5.962)	-42.119 (39.531)	5.963 (5.453)	7.878 (8.157)	7.878 (6.693)	-4.796 (10.429)	13.669** (6.922)	9.307 (10.462)	-9.194 (11.055)
invididual-constant characteristics									
male				1.637*** (0.117)	1.637*** (0.143)	2.157*** (0.195)	0.679*** (0.131)	1.762*** (0.189)	1.143*** (0.216)
higher education				-3.149*** (0.241)	-3.149*** (0.280)	-1.840*** (0.398)	-3.594*** (0.277)	-1.354*** (0.311)	-1.234*** (0.271)
high school education				-2.413*** (0.194)	-2.413*** (0.189)	-2.787*** (0.264)	-2.155*** (0.196)	-1.735*** (0.239)	-1.175*** (0.186)
married				-0.758*** (0.098)	-0.758*** (0.096)	-0.433*** (0.148)	-0.925*** (0.073)	-0.140 (0.127)	-0.346*** (0.103)
local resident				-1.334** (0.639)	-1.334** (0.619)	-2.218*** (0.814)	-0.825* (0.485)	-0.688 (0.702)	-0.781*** (0.242)
invididual-varying characteristics									
unemployed				-0.025 (0.287)	-0.025 (0.325)	-0.345 (0.388)	0.129 (0.299)	-1.100*** (0.343)	0.272 (0.247)
worker				-0.089 (0.283)	-0.089 (0.326)	-0.098 (0.393)	-0.095 (0.298)	-0.471 (0.353)	0.436* (0.258)
prisoner				4.658*** (0.419)	4.658*** (0.418)	5.627*** (0.562)	4.305*** (0.388)	3.446*** (0.390)	2.754*** (0.364)
student				0.055 (0.290)	0.055 (0.343)	-0.402 (0.426)	0.188 (0.322)	-0.142 (0.377)	0.593** (0.274)
office worker				-0.376 (0.315)	-0.376 (0.336)	0.786* (0.428)	-0.786** (0.327)	-0.389 (0.387)	0.535 (0.327)
official				0.052 (0.428)	0.052 (0.447)	6.430*** (0.789)	-3.317*** (0.424)	0.210 (0.532)	-1.698*** (0.494)
top manager				0.466 (0.342)	0.466 (0.423)	5.698*** (0.643)	-2.225*** (0.362)	0.838* (0.446)	-0.428 (0.394)
entrepreneur				-1.549*** (0.338)	-1.549*** (0.368)	0.012 (0.562)	-2.368*** (0.325)	-1.030*** (0.396)	-0.077 (0.343)
law enforcer				2.181* (1.235)	2.181** (1.105)	7.919*** (1.808)	-1.131 (1.140)	3.996*** (1.348)	1.167 (1.125)
age	-1.113*** (0.083)	-0.066 (0.084)	-0.432*** (0.033)	-1.056*** (0.085)	-1.056*** (0.075)	-1.388*** (0.117)	-0.940*** (0.076)	-0.908*** (0.087)	-1.633*** (0.088)
age^2	0.005*** (0.001)	-0.007*** (0.001)	0.003*** (0.000)	0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)	0.005*** (0.001)	0.005*** (0.001)	0.011*** (0.001)
crime characteristics									
charges per crime	3.660*** (0.334)	5.413*** (0.245)	4.424*** (0.399)	3.698*** (0.370)	3.698*** (0.341)	5.981*** (0.516)	2.453*** (0.296)	5.650*** (0.306)	3.708*** (0.269)
charges per $crime^2$	-0.265*** (0.056)	-0.341*** (0.055)	-0.289*** (0.066)	-0.271*** (0.065)	-0.271*** (0.060)	-0.634*** (0.093)	-0.087 (0.055)	-0.626*** (0.057)	-0.382*** (0.049)
under alcohol	1.123*** (0.132)	1.038*** (0.081)	0.976*** (0.209)	1.168*** (0.123)	1.168*** (0.136)	2.203*** (0.177)	0.262** (0.126)	1.852*** (0.163)	1.636*** (0.153)
under drugs	5.353*** (0.439)	17.196*** (0.430)	9.887*** (0.541)	5.393*** (0.448)	5.393*** (0.459)	3.964*** (0.531)	6.240*** (0.443)	4.437*** (0.418)	2.985*** (0.431)
group crime	6.627*** (0.176)	8.044*** (0.104)	7.357*** (0.235)	6.633*** (0.193)	6.633*** (0.193)	5.976*** (0.274)	7.194*** (0.161)	6.278*** (0.215)	3.001*** (0.223)
finished crime	-3.573*** (0.455)	-7.506*** (0.146)	-5.078*** (0.582)	-3.605*** (0.547)	-3.605*** (0.495)	-2.885*** (0.602)	-3.630*** (0.429)	-2.953*** (0.503)	-3.297*** (0.372)
crime day of year	0.002* (0.001)	0.001 (0.001)	0.002* (0.001)	0.002* (0.001)	0.002* (0.001)	0.006*** (0.002)	-0.001 (0.001)	0.000 (0.001)	-0.003* (0.002)
crime day of $year^2$	-0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000** (0.000)	0.000** (0.000)	0.000 (0.000)	0.000** (0.000)
Constant, crime half-year dummies, interactions of crime order with all above regressors are included in the model but not reported here									
Individuals×crimes	1,058,870	1,058,870	1,058,870	1,058,870	1,058,870	1,058,870	1,058,870	742,830	493,464
Individuals	449,967	449,967	449,967	449,967	449,967	449,967	449,967	318,138	218,980
estimator	Within	Between	Random	Mundlak	Mundlak	Mundlak	Mundlak	Mundlak	Mundlak
var-cov	bootstrap	bootstrap	bootstrap	bootstrap	cluster	cluster	cluster	cluster	cluster
model	baseline	baseline	baseline	baseline	baseline	+ severity SE	- severity SE	only sentenced	only district or territory courts

Table 2.A.2: Estimates of selected coefficients of (2.1) under different estimators, region-clustered or region-block-bootstrapped standard errors reported in parentheses. Dependent variable is the estimated seriousness (see page 101 for details on computation). “model” indicates the characteristics of the underlying data: “baseline” means the model with all individuals who committed their first crime in 2009 in Russia and recidivated in the said period, including those whose charges were dismissed for non-rehabilitating reasons; in models “± severity SE” I add/subtract estimated standard errors of severity scores to the dependent variable (see Table for top-10 offences in terms of standard deviations of their severity scores); in model “only sentenced” I consider as observations only instances of sentences and do not treat dismissals (for non-rehabilitating reasons) as crimes; in model “only district or territory courts” I include only instances of individuals sentenced at district or territory courts (not judges of peace). “Within” is the Within OLS-estimated value that controls for unobserved crime order-invariant heterogeneity across offenders. Offences with largest bootstrapped seriousness score standard deviations. “Between” is OLS-estimated value that controls for unobserved crime order-invariant heterogeneity within offenders. “Random” is the FGLS-estimated model. “Mundlak” is the FGLS-estimated model of Mundlak (1978) with individual-specific means of crime-varying variables included as additional regressors (not reported in the table).

Chapter 3

What do graduated sanctions tell us about the functions of the law: a case of drug crimes*

Introduction

Judicial decision-making is influenced by many factors (Ulmer, 2012). Empirical studies of sentencing have uncovered the combination of legal factors, such as prior record and offence seriousness, and extralegal factors, such as race, age, gender, or employment that shape the decisions of judges (Steffensmeier et al., 1993, 1998). The research has pointed to what judges look at when they make sentencing decisions and how they gather information, given the time and attention constraints as well as tacit

*I thank Maria Shklyaruk, Kirill Titaev, Alexey Knorre, and seminar participants at the European University at St. Petersburg for their feedback on earlier drafts of this chapter. I am also grateful to David Bjerk for his help with US federal sentencing data. I thank Minister Mikhail Abyzov and Open Government of Russia for granting me access to the police data analysed herein.

pressure from the court workgroup of prosecutors and defence counsel (Eisenstein and Jacob, 1977).

On the opposite side rests the decision to engage in criminal behaviour given the potential legal sanction. Early scholarship in Law & Economics has emphasised the imperative function of the law as a set of constraints that change the incentives of agents (Posner, 1983, 1997). This approach has come under criticism from legal scholars who point to the expressive function of the law (Sunstein, 1996, Cooter, 1998) as a signalling device for a social norm and what constitutes deviant behaviour (Nance, 1997). Experimental evidence has supported the expressive view of the law (Cooter and Bohnet, 2001, Galbiati and Vertova, 2008, Sacconi and Faillo, 2010).

This chapter sets out to understand the function of the law in the context of sentencing of drug offenders. Such crimes present an appealing testing ground for two reasons. First, many crimes related to drug possession without intent to sell or distribute can be viewed as victimless (Famega and Gaines, 2013) or associated with minimal social harm. Second, in many jurisdictions punishments for such crimes are graduated with the gravity of the offence expressed in the weight thresholds that specify different offence categories by the amount of seized drugs. The relationship between the perceived offence severity (manifested in mandatory minimum sanctions assigned to the thresholds) established by the legislator and the executive and the defendant's blameworthiness (expressed in the weight and type of drug seized) as viewed by the court can be examined empirically to understand whether the sanctions applied by judges are continuous in the perceived severity or carry an additional expressive function.

In this chapter I study the use of graduated sanctions for possession, storage, and transportation of drugs without intent to sell in Russia. Russian Criminal Code (2012) Article 228 stipulates in its parts 1, 2, and 3 sentencing ranges for significant, large, and extra-large weights of drugs seized, respectively. Those ranges are continuous in years of real incarceration assigned to the left or right of the weight thresholds provided in government orders. I gather a novel data set of 35,125 cannabis or heroin seizures reported in criminal case files opened by the police in 2013–14.[1] The weights of seized drugs are then linked to case outcomes following the investigation, prosecution, and sentencing of the defendants. This exercise is conducted for many regions and law enforcement agencies in the country, ensuring internal validity of the constructed data.

I then employ a regression discontinuity design (RDD) to study the judicial decision-making in the neighbourhood of the significant→large weight cutoff (100 grams for cannabis or 2.5 grams for heroin). I detect a highly significant discontinuity of additional 0.84 years of unconditional real incarceration assigned by judges when the weight of drugs seized crosses this threshold. Since the Criminal Code prescribes no discontinuity in the punishment schedule at the threshold, this chapter uncovers the effect that is extraneous to the law but is strongly manifested in practice. The revealed discontinuity

[1] Complete data on seizures can be examined at a companion website `http://atlasjustice.com/drugmap/`.

is robust to manipulations in drug weights by the police or defendants, size of weight bandwidth around the threshold, imperfect compliance with the law, differences in observable crime and defendant characteristics in the vicinity of the threshold, or case facts appearing in verdict texts.

The studies closest in design and motivation to this chapter are Bjerk (2014, 2017) that examine the role of mandatory minimums in US federal drug convictions. Unlike Bjerk, though, my focus is not limited to convictions for drug offences. By virtue of examining the police case files, I am able to include information on defendants whose charges were dismissed by investigators or judges. I follow Smith (1986) and Abrams (2011) by treating case outcomes resulting in charge dismissals, acquittals, or non-carceral punishments as zeros and including them in the study rather than conditioning the data on convictions and real incarceration sentences only, as in previous research. Secondly, I benefit from the fact that all drug offences are federal crimes in Russia. By focusing on charges with no intent to sell, I examine the outcomes of the war on drugs in the world's second largest jurisdiction at the lowest level of perceived social harm of offences.

The chapter proceeds as follows. Section 3.1 introduces the data on drug crimes and its context, Section 3.2 builds an RDD model and shows its baseline results, Section 3.3 subjects the results to a number of robustness checks, Section 3.4 closes with a discussion of potential channels of the uncovered threshold effect.

3.1. Data and context

Background 92,497 people were accused of drug crimes in Russia in 2009, comprising 9.9% of all defendants (Skougarevskiy et al., 2014, Table 8). 66.2% of those defendants were charged under Article 228 of Russian Criminal Code (2012) that penalises possession, transportation, making, or distribution of drugs with no intent to sell. In 2013–2014 natural cannabinoids (marijuana, hashish, and hash oil) were the primary drug seized in 33.8% of all drug crimes with known drug type, heroin — 24.9% (Knorre and Skougarevskiy, 2015, Table 2). Median weight of seized drugs was 20–39 grams of natural cannabinoids or 1–2 grams of heroin per crime (depending on the law enforcement agency).

These figures portray a criminal justice system that is extremely punitive toward drug offenders. Even though in 2004 Russia decriminalised possession of small amounts of narcotics, the cutoff values following the 2006 reversal were extremely low (Levinson, 2008). Table 3.1 juxtaposes the thresholds and sentencing ranges for drug crimes in Russia and for federal crimes in the United States Code. Even though this exercise is contaminated with an apples-and-oranges problem of comparing thresholds for different types of crimes and offenders across the jurisdictions, it still carries value when one considers the sheer difference in the magnitudes of the thresholds.[2]

[2] I have further traced government orders amending the Russian thresholds over the period under study and found no evidence of upward revisions of the cutoff points, only inclusion of new drug types (primarily synthetic cannabinoids with different chemical formulas).

	Russian Criminal Code			United States Code		
	Significant	Large	Extra-large	21 USC § 841 (b)(1)(D)	21 USC § 841 (b)(1)(B)	21 USC § 841 (b)(1)(A)
Cannabis	6	100	100,000	<50,000	100,000	1,000,000
Hash	2	25	10,000	<10,000	—	—
Heroin	0.5	2.5	1000	—	100	1000
Cocaine	0.5	5	1500	—	500	5000
Incarceration range, years	0.167–3	3–10	10–15	...–5	5–40	10–life

Table 3.1: Left panel shows gram weight thresholds for selected drugs in Russian Criminal Code, effective from 1 January, 2013, as amended by federal law 18-FZ of March 1, 2012 and stipulated in Government Resolution N1001 of October 1, 2012. See excerpts from Article 228 in Table 3.A.1. Thresholds for the United States are for first-time offenders under federal jurisdiction for unlawful distribution, possession with intent to distribute, manufacture, importation and exportation, etc. and come from Yeh (2015, p. 1). "Cocaine" means cocaine powder.

Small effective size of significant drug thresholds means that almost the entirety of drug possession in the country is criminalised (drug use, however, is an administrative offence). Commentators observe a lack of health and harm reduction services (e.g. methadone maintenance treatment) and over 1% of HIV-positive population, driven by injective drug use, in the country (Eastwood et al., 2016, Macfarquhar, 2016).

All police forces in Russia are under federal authority (Volkov et al., 2013). Russian criminal procedure is also heavily formalised.[3] In practice, this means that every accident report that is deemed to constitute a crime leads to initiation of a case file which is populated with evidence in the course of investigation. When it comes to drug crimes, the procedure provides that any seized substance that is suspected to be prohibited is sent for forensic expertise to determine its type and weight. If the inferred weight ex-

[3]See Supplementary appendix on page 52 for a description of criminal procedure in the country.

ceeds the "significant" threshold value (see Table 3.1), criminal case file is opened with the expert conclusion contained therein. Through detective work information on suspect might emerge, also recorded in a case file. When the investigation is completed, the case file is sent to prosecution for approval. The prosecutor can either agree with the results of investigation and send the case to court or return the case file for further investigation. Finally, the judge has the discretion to dismiss the charge or assign a monetary fine, hours of correctional labour, or years of suspended or real incarceration.

Obtained data Since criminal procedure in Russia is formalised, investigators have to fill out statistical cards whenever they initiate a case file or press charges against suspects. An example of a statistical card on criminal case file is reported in Figure 3.A.1. As the police is federal, the format of the cards and rules of reporting are uniform. Seeking to streamline the data collection process, in 2013 the Ministry of Internal Affairs of Russia launched a prototype unified data base "MOST-R" containing source information from such statistical cards on criminal case files. The Institute for the Rule of Law at the European University at St. Petersburg was granted access to the universe of over 5 million collected depersonified cards over 2013–2014 from this data base.

Knorre and Skougarevskiy (2015, `http://atlasjustice.com/drugmap/`) extracted and studied the information on drug crimes, primary drug types, and drug weights seized from statistical cards on crime characteristics (Form #1) using the aforementioned data base. I extend their logic in this chapter to match information on drug crimes with information on charged individuals (Form #2) from accompanying statistical cards as well amounts of seized drugs (Form #4) and case outcomes at trial (Form #6). This match allows me to build a comprehensive defendant-level data set on seized drugs and convictions. I then consider only individuals that had just one drug-related charge and no other charges, charged with possession of cannabis or heroin in 2013–2014 (these two drug types comprise 60.6% of defendants). I next limit my focus to Article 228 part 1 and part 2 only, excluding part 3 that deals with extra-large weights of drugs seized. Drug weights are right-winsorised at 99% within each drug type separately. Such narrow focus on two drug types and individuals who were charged with a sole instance of a drug crime allows me to minimise potential measurement errors in drug weights arising from aggregation of multiple charges or drug types with low average weights (e.g. MDMA). Finally, I conduct balance checks with the universe of 2013 court convictions on drug offences and exclude regions where the discrepancy between the total number of convicted individuals in the court records and police data is too large. This removes only a limited number of regions.

Summary statistics Table 3.2 reports lower moments of the collected data by mode of disposition of defendants' cases and part of Article 228. In particular, panel A shows the summary statistics for the entire sample of the accused while panel B only reports the information on the convicted individuals who were sentenced to real incarceration.

Six remarks are due. First, the majority of the defendants for Art. 228 p. 1 have low

	Art. 228, p. 1 (signif. weight)				Art. 228, p. 2 (large weight)			
	Mean	Min	Max	Obs.	Mean	Min	Max	Obs.
PANEL A: ALL ACCUSED INDIVIDUALS IN THE DATA								
male	0.922	0	1	21,096	0.916	0	1	14,029
citizen	0.979	0	1	21,096	0.970	0	1	14,029
russian	0.845	0	1	18,473	0.882	0	1	12,317
age group:								
14–17	0.014	0	1	21,096	0.018	0	1	14,029
18–24	0.153	0	1	21,096	0.175	0	1	14,029
25–29	0.194	0	1	21,096	0.212	0	1	14,029
30–39	0.417	0	1	21,096	0.419	0	1	14,029
40–49	0.160	0	1	21,096	0.134	0	1	14,029
50+	0.063	0	1	21,096	0.041	0	1	14,029
occupation:								
unemployed	0.770	0	1	21,094	0.730	0	1	14,028
blue-collar worker	0.181	0	1	21,094	0.213	0	1	14,028
white-collar worker	0.008	0	1	21,094	0.011	0	1	14,028
repeat offender	0.605	0	1	21,096	0.632	0	1	14,029
had administrative offences	0.097	0	1	21,096	0.075	0	1	14,029
under influence of alcohol	0.064	0	1	21,096	0.037	0	1	14,029
under influence of drugs	0.393	0	1	21,096	0.471	0	1	14,029
crime preparation	0.001	0	1	21,096	0.017	0	1	14,029
pretrial detention	0.030	0	1	21,096	0.193	0	1	14,029
agency initiating the case:								
police	0.896	0	1	21,096	0.770	0	1	14,029
federal narcotics service	0.104	0	1	21,096	0.229	0	1	14,029
other agencies	0.000	0	1	21,096	0.001	0	1	14,029
seized drug type:								
cannabis	0.713	0	1	21,096	0.579	0	1	14,029
heroin	0.287	0	1	21,096	0.421	0	1	14,029
case reached trial	0.978	0	1	21,096	0.982	0	1	14,029
punishment type:								
mandatory works	0.112	0	1	20,293	0.001	0	1	13,641
correctional works	0.074	0	1	20,293	0.002	0	1	13,641
limitation of freedom	0.043	0	1	20,293	0.001	0	1	13,641
suspended incarceration	0.319	0	1	20,293	0.555	0	1	13,641
real incarceration	0.191	0	1	20,293	0.438	0	1	13,641
fine	0.257	0	1	20,293	0.002	0	1	13,641
other punishment	0.004	0	1	20,293	0.002	0	1	13,641
sentence lower than mandatory minimum	0.000	0	1	20,205	0.010	0	1	13,622
conditional real incarceration, years	1.384	0	10	3,874	3.336	0	10	5,968
unconditional real incarceration, years	0.254	0	10	21,096	1.419	0	10	14,029
	10 25 50 75 90				10 25 50 75 90			
cannabis weight percentiles, grams:	[7 10 21 47 84]				[118 166 324 1053 2219]			
	10 25 50 75 90				10 25 50 75 90			
heroin weight percentiles, grams:	[0.56 0.67 0.88 1.30 1.90]				[2.68 3.00 4.10 7.30 24.36]			
PANEL B: ONLY SENTENCED TO REAL INCARCERATION								
	Art. 228, p. 1 (signif. weight)				Art. 228, p. 2 (large weight)			
male	0.926	0	1	3,874	0.902	0	1	5,968
citizen	0.941	0	1	3,874	0.939	0	1	5,968
russian	0.850	0	1	3,314	0.861	0	1	4,954
age group:								
14–17	0.001	0	1	3,874	0.003	0	1	5,968
18–24	0.098	0	1	3,874	0.116	0	1	5,968
25–29	0.178	0	1	3,874	0.210	0	1	5,968
30–39	0.470	0	1	3,874	0.492	0	1	5,968
40–49	0.180	0	1	3,874	0.144	0	1	5,968
50+	0.074	0	1	3,874	0.036	0	1	5,968
occupation:								
unemployed	0.862	0	1	3,874	0.801	0	1	5,968
blue-collar worker	0.114	0	1	3,874	0.167	0	1	5,968
white-collar worker	0.007	0	1	3,874	0.011	0	1	5,968
repeat offender	0.867	0	1	3,874	0.781	0	1	5,968
had administrative offences	0.119	0	1	3,874	0.065	0	1	5,968
under influence of alcohol	0.080	0	1	3,874	0.031	0	1	5,968
under influence of drugs	0.457	0	1	3,874	0.529	0	1	5,968
crime preparation	0.002	0	1	3,874	0.029	0	1	5,968
pretrial detention	0.131	0	1	3,874	0.412	0	1	5,968
agency initiating the case:								
police	0.903	0	1	3,874	0.765	0	1	5,968
federal narcotics service	0.097	0	1	3,874	0.235	0	1	5,968
other agencies	0.000	0	1	3,874	0.000	0	1	5,968
seized drug type:								
cannabis	0.584	0	1	3,874	0.326	0	1	5,968
heroin	0.416	0	1	3,874	0.674	0	1	5,968
sentence lower than mandatory minimum	0.001	0	1	3,874	0.018	0	1	5,968
conditional real incarceration, years	1.384	0	10	3,874	3.336	0	10	5,968
	10 25 50 75 90				10 25 50 75 90			
cannabis weight quantiles, grams:	[7 10 18 43 400]				[116 159 325 1151 3005]			
	10 25 50 75 90				10 25 50 75 90			
heroin weight quantiles, grams:	[0.55 0.69 0.91 1.35 2.02]				[2.69 3.05 4.32 8.71 35.37]			

TABLE 3.2: Summary statistics of the collected data on individuals charged with a single cannabis- or heroin-related drug crime under Art. 228 p. 1 or 2 in 2013–2014.

socio-economic status (77.0% were not employed). Second, the share of repeat offenders is noticeably high for drug crimes (60.5%). Third, the lion's share of cases reached trial (97.8%). This is due to organisational constraints that investigators face when weighing a decision to dismiss a case (Shklyaruk and Skougarevskiy, 2015, Titaev and Shklyaruk, 2016). Fourth, the leading mode of case disposition is suspended incarceration (31.9%), followed by a monetary fine (25.7%). Whereas drug offences are heavily criminalised in the country, the majority of defendants are assigned non-carceral punishments. This suggests a more lenient drug policy than it might appear from cross-country comparison of significant weight thresholds. Fifth, downward departures from mandatory minimum sentences are virtually non-existent in Russia, in contrast to US federal drug crime sentencing practices (Albonetti, 1997, Hartley et al., 2007). Sixth, the distribution of weight of drugs seized is skewed to the left region of the range given by the thresholds. 50% of cannabis-related defendants under Article 228 p. 1 had less than 20 grams of this drug seized while the significant→large weight range is 6–100 grams. What is more, the quantiles of the weight of drugs seized are similar for all defendants and those who were sentenced to real incarceration. This is an important observation pointing to the judicial logic on the choice of punishment type. Evidence from Table 3.2 suggests that drug weight may not be the leading factor in the judge's decision to assign real incarceration to a particular individual.

The finding on predominantly non-custodial disposition of drug crimes requires me to make unconditional (on conviction and real incarceration) length of real incarceration the focus of this study. In doing so, I follow Smith (1986) and Abrams (2011) in replacing non-carceral outcomes as zero years of real incarceration. The respective means are reported in rows "unconditional real incarceration, years" and "conditional real incarceration, years" of Table 3.2. Clearly, such transformation drives the average down (1.384→0.254 years for Art. 228 p. 1), thereby enriching the sentence length variable with information on non-custodial verdicts and case dismissals.

3.2. Model and results

I now set up my regression discontinuity model (Imbens and Lemieux, 2008, Lee and Lemieux, 2010). Consider a defendant i arrested with W_i grams of drugs. She receives $Y_{2,i}$ years of unconditional real incarceration if she is charged under Art. 228 p. 2 (large drug weight) and $Y_{1,i}$ if she is charged under Art. 228 p. 1 (significant weight). Investigation decides under which part of Article 228 to charge the defendant. Let T_i be the indicator of the article part, equal to 1 for part 2 and to 0 for part 1.

A government order prescribes a weight cutoff $\bar{w}$ that determines the choice of the part charged by the investigator so that $T_i = \mathbb{I}_{W_i \geq \bar{w}}$, where $\mathbb{I}_\bullet$ is an indicator function equal to unity when its condition is satisfied. In words, whenever the seized drug weight exceeds the prescribed cutoff value $\bar{x}$, the defendant is charged with part 2 and with part 1 otherwise. I assume perfect compliance of investigation with the government cutoff (sharp RDD design) so that $\Pr\left[T_i = 0 | W_i < \bar{w}\right] = \Pr\left[T_i = 1 | W_i \geq \bar{w}\right] = 1$. I will relax this assumption in a robustness check.

The parameter of interest is

$$\tau \equiv \left. \frac{\partial E\left[Y_i | W_i = w\right]}{\partial w} \right|_{w=\bar{w}}$$

Under two relatively weak assumptions[4] the parameter τ will express the difference in sentence lengths at the weight cutoff that defines the part of article charged:

$$\lim_{\epsilon \to 0+} \left\{ E\left[Y_i | W_i = \bar{w} + \epsilon\right] - E\left[Y_i | W_i = \bar{w} - \epsilon\right] \right\} = E\left[\tau | \bar{w}\right].$$

For cannabis, $\bar{w}$ is 100 grams. If I compare the sentences for cannabis defendants charged under part 1 of Article 228 and those charged under part 2 where weight seized was slightly below or slightly above 100 grams, I can in the limit estimate the difference τ in sentences induced solely by the change in the part of Article 228 charged. τ can also be seen as the Local Average Treatment Effect (LATE) of being charged with part 2 of Article 228 vs. part 1 on length of unconditional real incarceration (Imbens and Angrist, 1994).

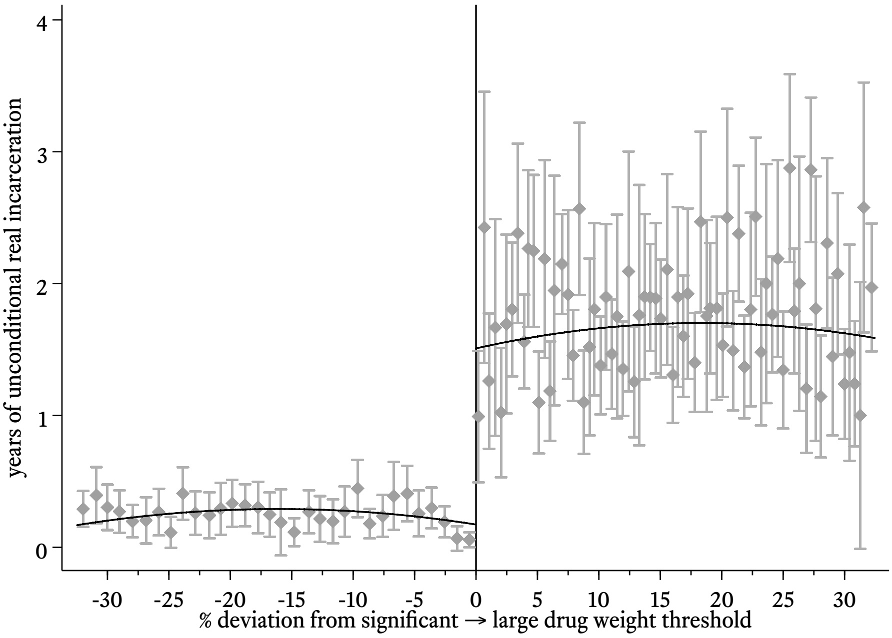
What is the normative content of τ? Table 3.A.1 compares the texts of parts 1 and 2 of Article 228 to find that the two parts punish the "same deeds" that are different only in the amount of drugs seized. The schedule of available sanctions is continuous, from

[4](i) continuity of $E\left[Y_i | W_i = w\right]$ around $\bar{w}$ (or, equivalently, positive density of W in the neighbourhood of $\bar{w}$) and (ii) well-defined $\lim_{\epsilon \to 0+} \left\{ E\left[\tau | W_i = \bar{w} + \epsilon\right] \right\}$.

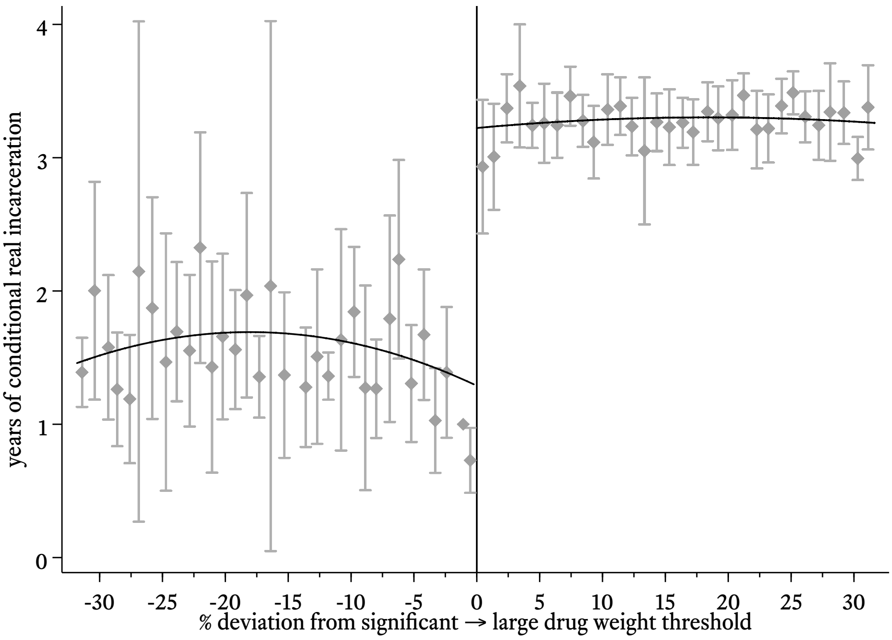
0.167 to 3 years for part 1 and from 3 years to 10 years for part 2. This gives a strong legal reason to expect τ to be zero if the repertoire of punishment types available to a judge is constant.

Figure 3.1 offers the visual rendition of τ in observed data. Panel A plots as squares the averages of unconditional sentences by drug weight bins in the neighbourhood below (left of) or above (right of) the significant$\rightarrow$large weight threshold for cannabis and heroin. The x-axis expresses percentage deviation of the weight of drugs seized from the cutoff value. Ticks show 95% confidence intervals of the binned averages. Bin widths and the size of the neighbourhood are optimally determined with a mean-squared-error-minimising approach of Calonico et al. (2017). Solid line is a second-order local polynomial. Informally, τ can be inferred as the difference between the estimated polynomials at 0% deviation from the cutoff.

Even a cursory examination of Figure 3.1 suggests that estimated τ is large, in contrast to the normative expectation of it being zero. Panel B of Figure 3.1 reports the τ when I redefine the outcome variable of interest as the conditional length of real incarceration and plot only average sentences for defendants sentenced to real incarceration. This is done to refute the obvious explanation of the non-zero τ due to different punishment types available to judges for Article 228, p. 1 and 2. Even when I condition on one punishment type, the difference in average sentences induced by sheer change of part of Article 228 remains equally large.



(A) Unconditional real incarceration: cannabis and heroin



(B) Conditional real incarceration: cannabis and heroin

Figure 3.1: Solid lines in the sub-figures show the second-order local polynomial regression of the length of unconditional (panel A) or conditional (panel B) real incarceration on weight of seized cannabis or heroin to the left or right of the significant → large seized drug weight threshold. Local regression is estimated conditioning on observable covariates with MSE-optimal symmetric bandwidth (Calonico et al., 2017) and triangular kernel function. Mean lengths of incarceration by seized drug weight bins (in percentage points relative to the threshold) are reported as diamonds, vertical lines with ticks show their 95% confidence intervals. Results diaggregated by drug type are reported in Table 3.A.3

3.3. Robustness checks

With τ effect appearing surprisingly potent in a graphical examination, I will now show that this finding is unaffected by conventional explanations. Table 3.3 organises the results, with column (2) showing the τ of 1.31 years being the formal equivalent of τ reported in panel A of Figure 3.1.

Conditional expectation of observables at cutoff A criminal justice practitioner could argue that the defendants under part 1 and 2 of Article 228 are inherently different in their observable characteristics. Formally speaking, conditional expectation of covariates might be different below and above the cutoff, rendering RDD estimate inconsistent (Calonico et al., 2016). In Figure 3.A.4 I test this argument by searching for discontinuities in the observable characteristics of defendants around the significant→large cutoff. I detect discontinuity in the indicator of a minor or repeated defendant, pretrial detention, case being initiated by the federal narcotics service (FSKN), or heroin being seized at the cutoff. The signs on the latter three covariate RDD estimates are positive, indicating higher probability of pretrial detention, investigation by the FSKN, or incidence of heroin for the charges related to part 2 of Article 228. For this reason in column (1) of Table 3.3 I report the covariate-adjusted RDD estimate. It is markedly lower (0.84) than the unadjusted results (1.31) in column (2) but is still large in magnitude in relation the the expected τ of zero. In what follows, covariate-adjusted RDD estimate will serve as

Model	(1) baseline	(2) baseline	(3) two-sided bandwidth	(4) HC0 var-cov	(5) with region dummies	(6) 1st-order polynomial	(7) 3rd-order polynomial	(8) Fuzzy RDD
LATE, years of incarceration	0.84	1.31	1.10	0.84	0.83	0.81	0.86	0.93
Clustered 95% CI	[.60; 1.13]	[.88; 1.81]	[.94; 1.30]	[.64; 1.09]	[.61; 1.08]	[.64; .947]	[.62; 1.14]	[.64; 1.25]
Bandwidth, p.p.	[±32.50]	[±38.36]	[-22.54; +3128.55]	[±31.33]	[±29.62]	[±35.80]	[±39.32]	[±38.02]
Observations: below	19380	19382	19380	19380	19380	19380	19380	19380
Observations: above	14444	14445	14444	14444	14444	14444	14444	14444
Effective observations: below	2277	2859	1481	2179	2017	2562	2959	2855
Effective observations: above	2952	3273	13348	2875	2751	3108	3322	3271
Cond. on real incarceration	no	no	no	no	no	no	no	no
Cond. on covariates	yes	no	yes	yes	yes	yes	yes	yes

Table 3.3: This table shows the local average treatment effect of crossing the significant → large seized drug weight threshold on length of unconditional real incarceration under different settings. Column (1) is the baseline regression discontinuity estimate from second-order local polynomial regression conditioned on observable covariates with MSE-optimal symmetric bandwidth (Calonico et al., 2017) and triangular kernel function. Column (2) is the baseline estimate without conditioning the local regression on observable covariates. Column (3) reports the LATE when MSE-optimal bandwidths are selected independently to the right and to the left of the cutoff. Column (4) is identical to column (1) in point estimates whereas variance-covariance matrix is Huber-Eicker-White without any additional weights imposed. Column (5) includes region indicators as additional covariates in the local polynomial regression. Columns (6)–(7) report the LATEs from first- or third-order local polynomial regression, respectively. Column (8) shows estimates from fuzzy regression discontinuity design where I instrument the significant → large seized drug weight treatment dummy with a dummy equal to unity if the charged crime was under Art. 228 p. 2 (large weight) and nil if it was under Art. 228 p. 1 (significant weight). Confidence intervals come from region-cluster-robust nearest neighbour variance-covariance matrix (Calonico et al., 2014).

my baseline.

Custom bandwidth One could argue that the MSE-optimally set size of the neighbourhood of 38.3 percentage points (p.p. henceforth) around the significant→large weight cutoff is incredibly wide. Figure 3.2 reports τ when the neighbourhood size is set in the 1 (1) 15p.p. range. If anything, narrowing the size of the neighbourhood around the cutoff drives the estimated τ upwards.

Drug weight manipulation So far I have assumed that seized drug weight W_i has no measurement errors. However, the defendants could potentially manipulate the weight shifting it to the left of the cutoff to enjoy more lenient sentencing under part 1 of Article 228, as the estimated τ suggests. The police, in contrast, could shift the weight to the right of the cutoff, thereby moving the charge to part 2 of Article 228. This could happen with entrapment, falsification of weights at the forensic lab, or dilution or concentration of the active ingredient of seized drugs. Figure 3.3 reports the density of convictions by bins of percentage points of drug weight deviations from the cutoff. It further demonstrates with the aid of Cattaneo et al. (2017) test that the densities of defendants appearing to the right or to the left of the cutoff are not equal. Such strong evidence for weight manipulation around the cutoff could potentially invalidate my $T_i = \mathbb{I}_{W_i \geq \bar{w}}$ identification assumption because individuals appear to self-select into the part of Article 228.

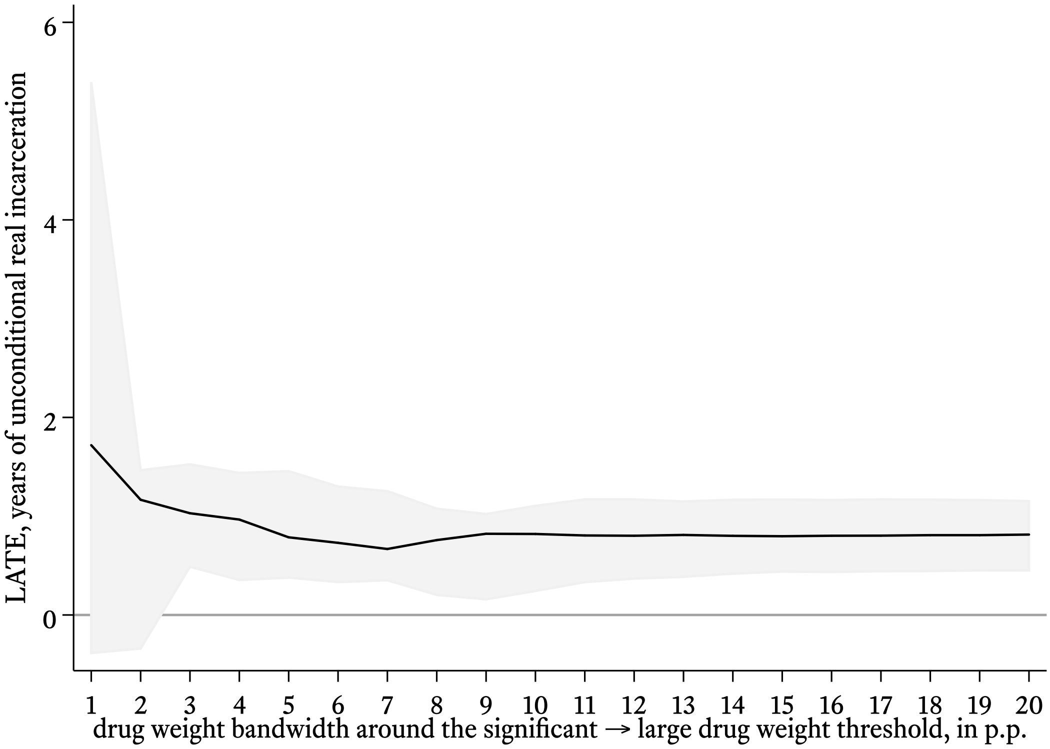


Figure 3.2: Custom bandwidth. This figure shows the local average treatment effect of crossing the significant → large seized drug weight threshold on length of unconditional real incarceration under different bandwidth sizes. *x*-axis reports the percentage-point size of the bandwidth around the threshold. Grey area is 95% confidence interval from region-cluster-robust nearest neighbour variance-covariance matrix (Calonico et al., 2014).

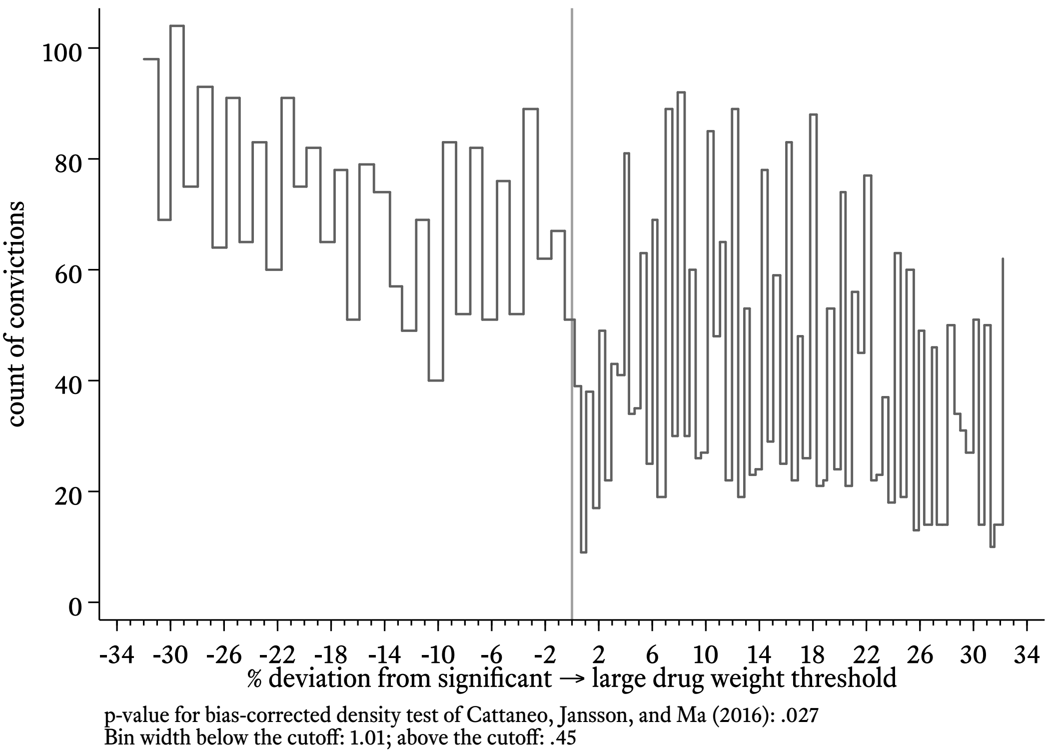


FIGURE 3.3: This figure reports density of heroin- and cannabis-related convictions for Art. 228 p. 1 (significant weight seized, left to the zero-line) or Art. 228 p. 2 (large weight seized, right to the zero-line) by percentage point deviation from the significant → large drug seized weight cutoff. I also report *p*-value for bias-corrected density test of Cattaneo et al. (2017) with the null hypothesis of no manipulation in drug weight around the threshold. Note: figure has asymmetric bin width to the left and right of the cutoff. Supplementary appendix Figure 3.A.2 reports densities by drug type.

Doughnut RDD Unequal density around the cutoff is also known as a specific case of "heaping" (Barreca et al., 2011, 2016). The authors propose a "donut-RD" approach as a robustness check to understand the magnitude of bias induced by potential manipulations. This amounts to removing observations in the immediate vicinity of the cutoff and estimating τ free of such potentially manipulated data points. Figure 3.4 constructs a schedule of $0\ (0.05)\ 2.5$p.p.-sized doughnuts and reports the estimated τ when observations in the specified neighbourhood of the cutoff are omitted. I do not observe any marked change in the estimated τ as a result of this exercise.

Placebo test Instead of removing observations in the immediate vicinity of the cutoff, I could also shift the cutoff itself in a series of falsification tests. The estimated τs when the significant$\rightarrow$large drug weight cutoff is shifted to the left or right of its true value are reported in Figure 3.5. The cutoffs shifted by -1, 1, and 2 percentage points indeed reproduce the discontinuity estimate. However, the effect size is much lower. This is additional evidence that supports my identifying assumptions.

Asymmetric bandwidth In column (3) of Table 3.3 I report τ when the size of the neighbourhood is allowed to be determined independently to the right and to the left of the cutoff. This increases the estimated τ in relation to the baseline.

Within-region balancing In column (5) of Table 3.3 I impose additional balance on covariates by including region dummies, τ is unchanged.

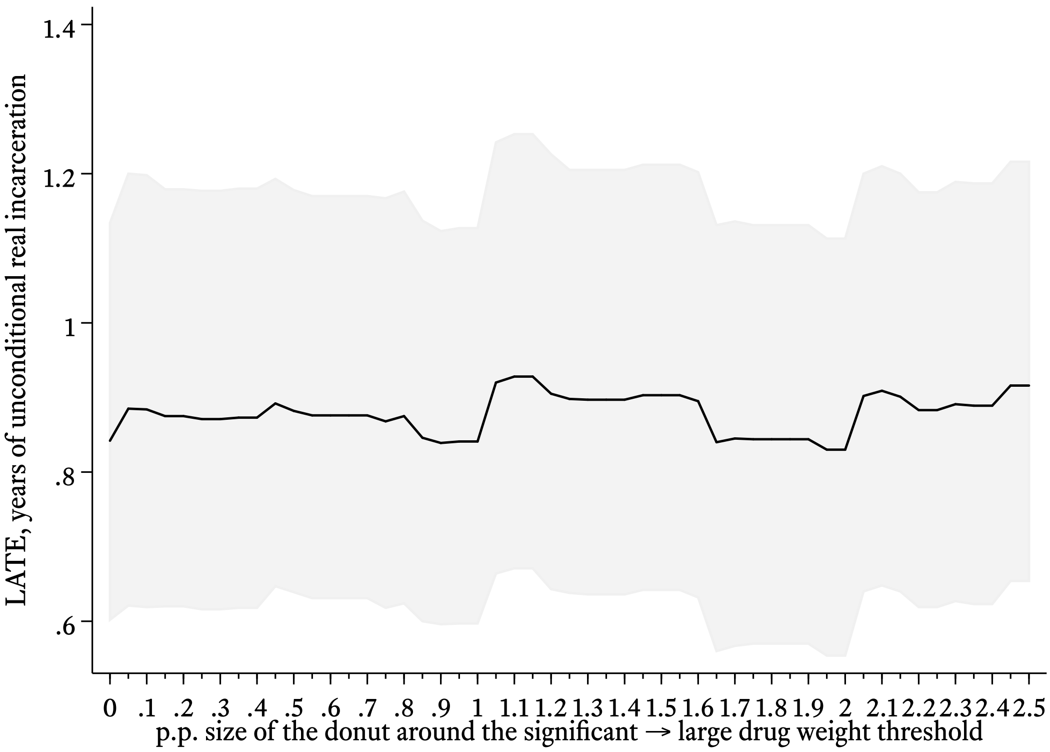


FIGURE 3.4: Doughnut regression discontinuity. This figure shows the local average treatment effect of crossing the significant → large seized drug weight threshold on length of unconditional real incarceration when I exclude the observations in the vicinity of the stipulated weight threshold (Barreca et al., 2011, 2016). *x*-axis reports the percentage-point size of the window where observations are removed. Grey area is 95% confidence interval from region-cluster-robust nearest neighbour variance-covariance matrix (Calonico et al., 2014). LATE estimates at indicative levels of doughnut size are reported in Table 3.A.2.

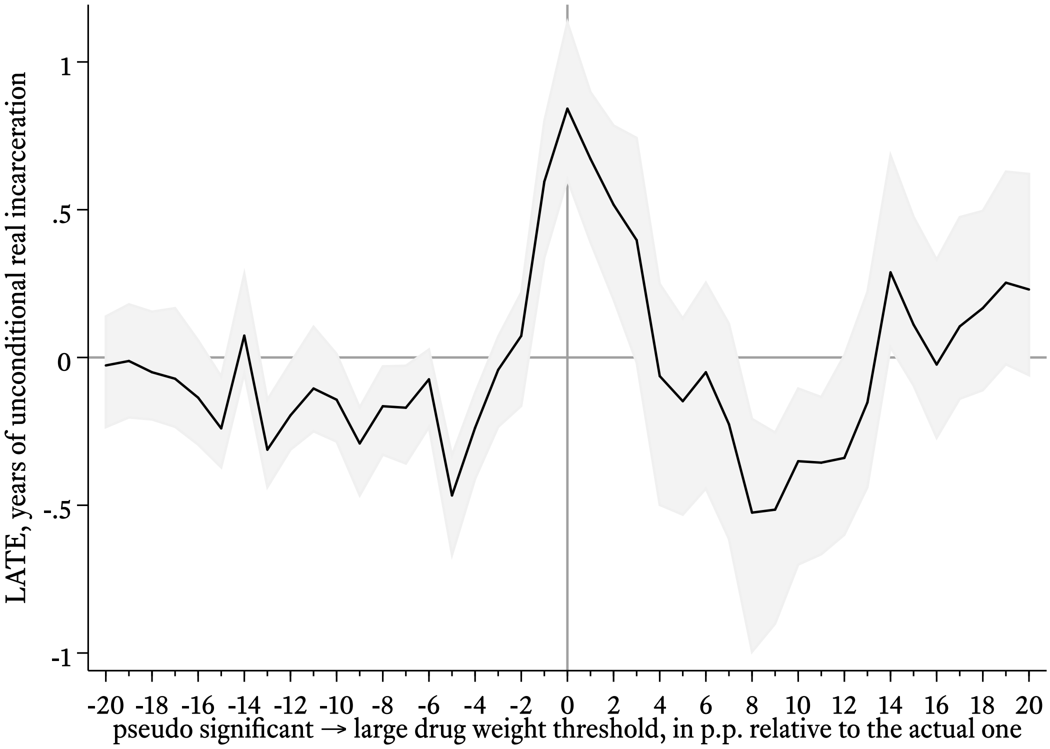


Figure 3.5: Placebo test. This figure shows the local average treatment effect of crossing the significant → large seized drug weight threshold on length of unconditional real incarceration when I impose a pseudo threshold that is shifted in relation to the actual value. *x*-axis reports the percentage-point size of the shift, where zero indicates no shift (actual value). Grey area is 95% confidence interval from region-cluster-robust nearest neighbour variance-covariance matrix (Calonico et al., 2014).

Order of local polynomial regression My baseline specification uses 2nd-order local polynomial to estimate τ in the optimally set neighbourhood. Cognisant of variance-bias trade-off when choosing the order of polynomial, in columns (6) and (7) of Table 3.3 I report τ estimated with 1-st or 3-rd order polynomial, respectively. The results are unchanged.

Imperfect compliance I finally relax my assumption of perfect compliance with the government order $\Pr[T_i = 0|W_i < \bar{w}] \neq \Pr[T_i = 1|W_i \geq \bar{w}]$ in a Fuzzy RDD design. Column (8) shows the result which is slightly more potent that the baseline estimate. Assumption of imperfect compliance allows me to partially alleviate the troublesome findings of the density manipulation test.

3.4. Potential channels

Since the findings on τ are established to be robust to a number of potential explanations, I now ask why the mere fact of being charged with part 2 of Article 228 could lead to additional 0.84 years of unconditional real incarceration in relation to part 1 of this article.

Regional variation Bjerk and Mason (2017) document variation in judicial discretion in adjudicating federal drug cases. While my τ is not the measure of judicial discretion but rather an indicator of a discontinuity in their decision-making, I have no normative prior on its variation. In Figure 3.6 I decompose the τ by region running location-specific

RDDs for cannabis. I find marked regional variation in τ to the extent of it being not significantly different from zero in one group of regions (top of the figure) and it being extremely pronounced in another group of regions (bottom of the figure). What is more, I observe spatial clustering of the discontinuity: the regions with most pronounced positive τ are located in Southwest Russia and include republics of North Caucasus.

While the result cannot be replicated on the heroin sample due to limited number of its seizures around the cutoff per region, uncovered regional variation in cannabis alone communicates important information about the nature of the discontinuity. It is hard to reconcile non-zero τ with the imperative function of the law since the discontinuity shows regional variation. In contrast, expressive reading of the law as a means to announce socially desirable behaviour (Bénabou and Tirole, 2011) is in conformity with the uncovered discontinuity and its regional variation. τ is found to be largest in predominantly Muslim regions of Russia where per capita consumption of alcohol or drugs is lowest in the country (Republic of Dagestan, Kabardino-Balkaria Republic, Republic of Ossetia (Alania)). This might be weakly indicative of the fact that τ is a stigma imposed on drug users by criminal justice system.

Temporal variation in adoption of legislation The period under study (2013–2014) was an aftermath of an amendment to Article 228. Before 1 January, 2013 Article 228 consisted of two parts rather than three. Prior to the amendment the significant drug weight was called "large", while the large weight was referred to as "extra-large".

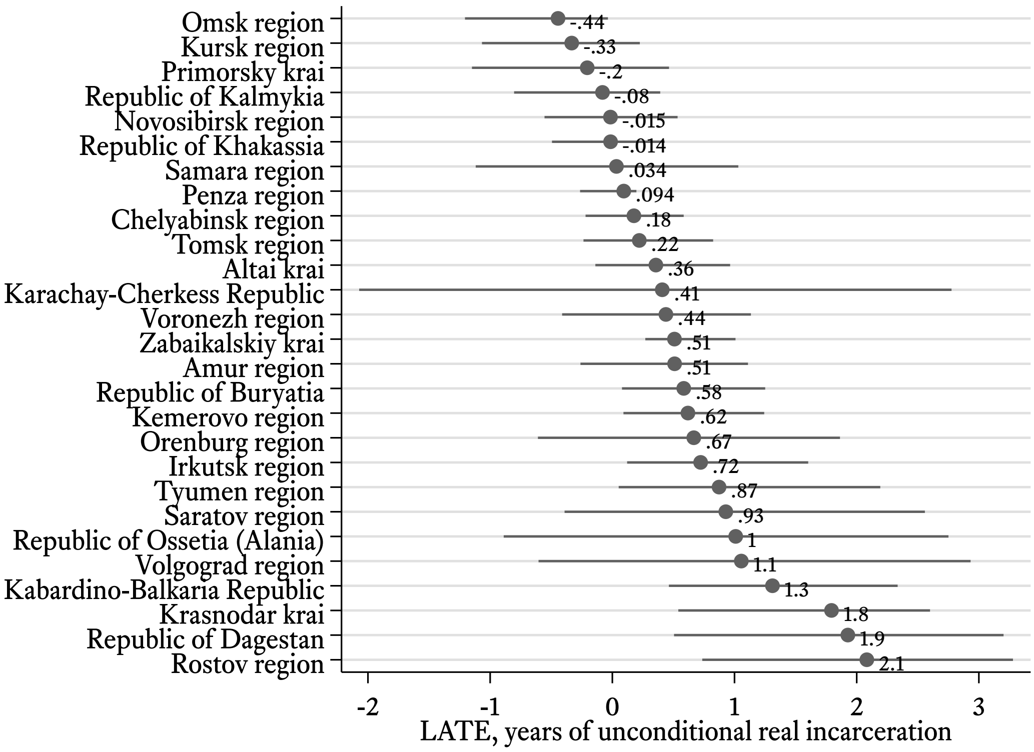


FIGURE 3.6: This figure shows the local average treatment effect of crossing the significant → large seized cannabis weight threshold on length of unconditional real incarceration by region. I consider only regions with more than 100 crimes and consider linear local polynomial regression. Lines are 95% confidence intervals from Huber-Eicker-White variance-covariance matrix.

The threshold values were the same, so the only change induced by the amendment was renaming "large" weight to "significant", "extra-large" to "large", and including a new "extra-large" category in part 3 of Article 228. While the jurisprudence has continuously highlighted the neutrality of the amendment, in Figure 3.7 I test it empirically by computing τ only for cases adjudicated shortly after the changes came into effect on 1 January, 2013. I do not observe any difference in τ when the post-reform window is sufficiently large to conduct credible inference.

Unobservable case facts Every defendant faces trial with a unique set of facts unobservable to an econometrician. As seen in Table 3.2, for example, drug weight is not the predominant factor driving the decision of a judge to assign real incarceration. Therefore, non-zero τ might be driven by unequal density of unobservables around the cutoff. I propose to test this channel with the aid of information available from verdict texts.

To this end, I start with a completely different source of information on drug weights, observable characteristics, and punishments. From the universe of court cases adjudicated in 2013 and available as court records at the Judicial Department at the Supreme Court of Russia[5], I extract the information on verdicts for parts 1, 2 of Article 228 in the country. I then match this information with verdict texts available in the public domain.[6] As a next step, I develop a set of regular expressions to extract drug type and weight from verdict texts matched with court records. Finally, I restrict my attention to charges on

[5]See Section 1.2 for a complete description of this data source.
[6]See Supplementary appendix on page 58 for a description of matching procedure.

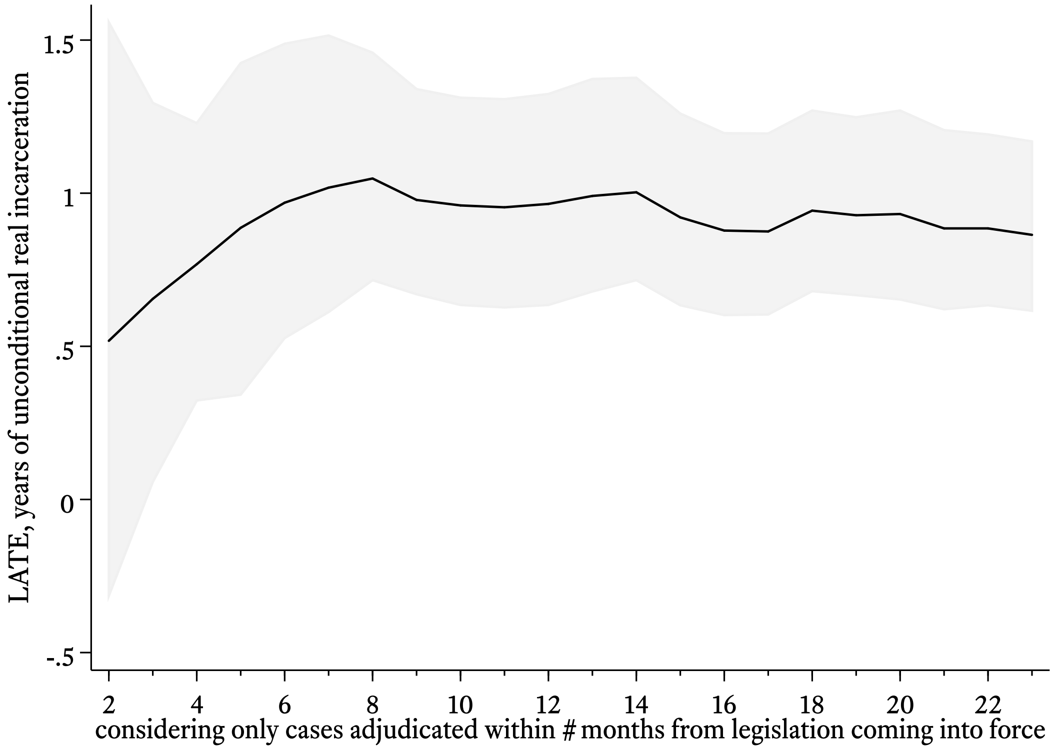


Figure 3.7: This figure shows the local average treatment effect of crossing the significant → large seized drug weight threshold on length of unconditional real incarceration by months elapsed from the date of new edition of Art. 228 p. 1–2 coming into force. *x*-axis shows the month cutoff for observations to be included in the regression. Grey area is 95% confidence interval from region-cluster-robust nearest neighbour variance-covariance matrix (Calonico et al., 2014).

heroin and cannabis in 2013 to maintain comparability with my police data.

This exercise produces just 807 observations (further reduced to 794 after list-wise deletion). However, they serve as an important source of alternative information on drug weights that does not rely on police records. In columns (1)-(2) of Table 3.4 I re-run my baseline specification without or with covariate balancing, respectively. The estimated τ effects appear much more potent in relation to police records-based τ, providing additional evidence in favour of the uncovered discontinuity.

However, the main benefit of my matching exercise comes into play when I consider extraneous information from the attached verdict texts. In particular, I view verdict texts from the bag-of-words perspective (Gentzkow et al., 2017), and construct their unigram or bigram counts. Having pruned the grams that occur in less than 1% of verdicts and removing stop-words in my unigram computation, I produce a 807 verdicts$\times$4721 unigram or 8136 bigram count matrix. These are the counts of terms and phrases describing the facts of the cases in verdict texts. I next employ the Multinomial inverse regression of Taddy (2013, 2015) and regress count of each unigram on the Article 228 part 2 indicator. This gives me the relationship between uni/bigram counts in verdict texts and the charge being part 1 or part 2 of Article 228. Then I compute the Sufficient Reduction (Cook and Ni, 2005) of my Article 228 part 2 dummy in terms of verdict text unigrams or bigrams.

Intuitively, with inverse multinomial logit I provide a mapping between terms and

	(1)	(2)	(3)	(4)	(5)	(6)
conditional on text sufficient reductions	no	no	unigrams	bigrams	no	unigrams
LATE, years of incarceration	2.19	2.13	1.51	1.70	2.23	1.46
Clustered 95% CI	[1.21; 3.13]	[1.37; 2.81]	[.71; 2.28]	[.94; 2.46]	[1.45; 2.90]	[.57; 2.25]
Bandwidth, p.p. or grams	[±37.76]	[±32.66]	[±34.44]	[±35.63]	[±33.91]	[±35.12]
Observations: below	508	505	505	505	470	470
Observations: above	286	283	283	283	256	256
Effective observations: below	55	47	48	50	44	46
Effective observations: above	49	46	47	48	41	43
Cond. on real incarceration	no	no	no	no	no	no
Cond. on covariates	no	yes	yes	yes	yes	yes
Drug type	cannabis & heroin	cannabis & heroin	cannabis & heroin	cannabis & heroin	cannabis	cannabis

Table 3.4: This table shows the local average treatment effect of crossing the significant → large seized drug weight threshold on length of unconditional real incarceration. Data are obtained from the alternative source: I match verdict texts that have been made public on court websites with the universe of court records of the Judicial Department at the Supreme Court in 2013. From the matched texts I extract information on weight and type of drug (cannabis or heroin) seized. Column (1) is the baseline regression discontinuity estimate from second-order local polynomial regression with MSE-optimal symmetric bandwidth (Calonico et al., 2017) and triangular kernel function, not conditioning on observable covariates. Column (2) is the baseline estimate when conditioning the local regression on observable covariates. Column (3) reports the LATE when I include as a covariate a Sufficient Reduction of multinomial inverse regression (Taddy, 2013) of verdict text unigrams on a dummy equal to unity if the charged crime was under Art. 228 p. 2 (large weight) and nil if it was under Art. 228 p. 1 (significant weight). The inverse multinomial regression is estimated for 4721 unigrams that occur in more than 1% of documents (after removing stop words) factorised into independent Poisson regressions (Taddy, 2015). Then the Sufficient Reduction (Cook and Ni, 2005) of verdict texts in terms of the choice of Art. 228 p. 2 over p. 1 is obtained from the estimates and included in the RDD local regression as a covariate. Column (4) is identical to column (3) in estimation procedure but uses 8136 verdict text bigrams as dependent variables in the multinomial inverse regression step. Columns (5)–(6) are identical to columns (3)–(4), respectively, but consider only verdicts on seizures of cannabis. Confidence intervals come from region-cluster-robust nearest neighbour variance-covariance matrix (Calonico et al., 2014).

their relation to verdict text being about Article 228 part 2. Then I reduce this mapping to a unidimensional measure that communicates to what extent a given verdict text is on part 2 or part 1 of Article 228. Importantly, this information is based on verdict text uni/bigrams only and does not rely on any case observables but the text. Finally, I include the computed Sufficient Reduction of verdict texts as an additional covariate in my RDD estimation. Columns (3)–(4) report the τ after this exercise for unigram or bigram count matrix, respectively. After forcing balancing on sufficient reduction of verdict texts my discontinuity estimate goes down but still remains positive and large. Clearly, verdict texts do not capture the entirety of unobservables driving verdict decisions. However, accounting for textual content of verdicts allows me to give less credence to any explanation of τ that involves unobservables since the estimates turn out to be unaffected when I control for a chunk of unobserved variation with the aid of verdict texts.

Supplementary appendix

Article 228 part 1	Article 228 part 2
Illegal acquisition, storage, transportation, making, or processing of narcotic drugs, psychotropic substances, or analogues thereof on a large scale, as well as illegal acquisition, storage, and transportation without the purpose of selling plants containing narcotics, or psychotropic substances, or parts thereof containing narcotics, or psychotropic substances **on a significant scale**, without the purpose of sale shall be punishable with a fine in an amount of up to forty thousand roubles, or in the amount of the wage or salary, or any other income of the convicted person for a period of up to three months, or by compulsory works for a term of up to four hundred and eighty hours, or by corrective labour for a term of up to two years, or *by deprivation of liberty for a term of up to three years, or by incarceration for the same term.*	The same deeds committed on a **large scale** shall be punishable by *incarceration for a term of three to ten years* with or without a fine in an amount of up to five hundred thousand thousand roubles or in the amount of the wage or salary, or any other income of the convicted person for a period of up to three years.

Table 3.A.1: Texts of Art. 228 p. 1–2 of Russian Criminal Code, as amended by federal law 18-FZ of March 1, 2012. My emphasis is added. Weight thresholds are reported in Table 3.1.

Donut size	baseline (no donut)	0.5pct donut	1pct donut	2.5pct donut
LATE, years of incarceration	0.84	0.88	0.92	0.92
Clustered 95% CI	[.60; 1.13]	[.64; 1.18]	[.66; 1.24]	[.65; 1.22]
Bandwidth, p.p.	[±32.50]	[±33.75]	[±34.65]	[±38.17]
Observations: below	19380	19375	19329	19249
Observations: above	14444	14404	14376	14274
Effective observations: below	2277	2366	2417	2724
Effective observations: above	2952	2965	2990	3102
Cond. on real incarceration	no	no	no	no
Cond. on covariates	yes	yes	yes	yes

Table 3.A.2: Doughnut regression discontinuity. This table shows the local average treatment effect of crossing the significant $\rightarrow$ large seized drug weight threshold on length of unconditional real incarceration when I exclude the observations in the vicinity of the stipulated weight threshold (Barreca et al., 2011, 2016). The percentage-point size of the window where observations are removed is reported in the header row. Confidence intervals come from region-cluster-robust nearest neighbour variance-covariance matrix (Calonico et al., 2014). LATE estimates at intermediate values of doughnut size are reported in Figure 3.4.

СТАТИСТИЧЕСКАЯ КАРТОЧКА НА ВЫЯВЛЕННОЕ ПРЕСТУПЛЕНИЕ

1. ______________________
орган: внутренних дел (01), прокуратуры (02), таможенный (04), суд (05), ФСИН (06), ФСБ (07), ПС ФСБ (11), ФССП (08), ГПС МЧС (10), ФСКН (14), иной (09)

2. Учесть: основная (1), дополнительная (2), изменить (корректирующая) (3) 02

РАЗДЕЛ 1. ЗАПОЛНЯЕТСЯ СОТРУДНИКАМИ ПРАВООХРАНИТЕЛЬНОГО (ПРАВОПРИМЕНИТЕЛЬНОГО) ОРГАНА ПО ВЕДЕНИЮ РЕГИСТРАЦИОННО-УЧЕТНОЙ И СТАТИСТИЧЕСКОЙ РАБОТЫ

вид год номер

3. Номер уголовного дела (1), материала (2) год _____ №__________ 03

4. Порядковый номер преступления в уголовном деле ___________ номер 04

5. Номер регистрации сообщения о преступлении в регистрационном документе
№ _____________ Дата "____" __________ 200__ г. 05

6. Дата направления карточки в ИЦ "____" ____________________ 200__ г.
Сотрудник органа ________________________________ 06
(фамилия, подпись) год мес. чис.

7. Дата поступления карточки в ИЦ "____" ________________ 200__ г.
Данные карточки учтены в государственной форме отчетности "____" ________________ 200__ г. 07
Сотрудник информационного центра ________________________________
(регистрационно-учетного подразделения) (фамилия, подпись)

РАЗДЕЛ 2. ЗАПОЛНЯЕТСЯ ЛИЦОМ, ВЕДУЩИМ РАССЛЕДОВАНИЕ УГОЛОВНОГО ДЕЛА ИЛИ РАЗРЕШИВШИМ МАТЕРИАЛ

8. ПРЕСТУПЛЕНИЕ ПРЕДОТВРАЩЕНО на стадии приготовления, покушения: сотрудниками ОВД (01), по их инициативе (02); сотрудниками таможенного органа (07), ФСИН (09), ФССП (10), ФСБ (08), ПС ФСБ (12); ГПС МЧС(11); ФСКН (85); прокуратуры (13); силами общественности (03), частным детективом (04), частным охранником (05) 08

9. ПРЕСТУПЛЕНИЕ ВЫЯВЛЕНО (наличие состава преступления установлено): **следователем:** прокуратуры (001), органа внутренних дел (002), в т. ч. по расследованию организованной преступной деятельности (003), налоговых преступлений (087), ГУ МВД России по ФО (088), ФСБ (083), ФСКН (085); **сотрудниками:** *криминальной милиции:* подразделений: уголовного розыска (007), ОБППГ (008), ОРЧ на транспорте (057), подразделений по борьбе с организованной преступностью (005), в т.ч. ОРБ ГУ МВД России по ФО (055), подразделений по экономическим (006), налоговым (049) преступлениям, ОРБ ЭП (058), НП (059) ГУ МВД России по ФО, ДЭБ МВД России (069), УСТМ (056), собственной безопасности (050), НЦБ Интерпола (033); *милиции общественной безопасности:* ГИБДД (011), участковым уполномоченным (012), дознания (013), ППС (014), ПДН (015), БППРИАЗ (020), ОМОН (016), вневедомственной охраны (017), ЦВСНП (052), спецприемника (053); ИВС (051); ПВС (032); дежурной части (018), других служб органов внутренних дел (039); **сотрудниками:** ФСИН (070): ИУ (061), ЛИУ (062), СИЗО (064); таможенного органа (042); органа ФССП (082); ГПС МЧС (031); ФСБ (043), ПС ФСБ (084);ФСКН (086), в т.ч.: оперативных (098), по контролю за легальным оборотом наркотиков (099) подразделений; суда (071); частным детективом (044), частным охранником (045); *следственно-оперативной группой (080);* по оперативным данным (100); с помощью оперативно-технических мероприятий (200) 09

10. УГОЛОВНОЕ ДЕЛО ВОЗБУЖДЕНО: **прокурором** (001), в т. ч. при отмене им постановления об отказе в возбуждении уголовного дела (002), в том числе по инициативе ОВД (029), **следователем:** прокуратуры (003), органа внутренних дел (004), в т. ч. по расследованию организованной преступной деятельности (005), налоговых преступлений (028), ГУ МВД России по ФО (088); ФСБ (010); ФСКН (085); **судьей** (015); **дознавателем** (006), в т. ч.: органов внутренних дел (007); сотрудником органов внутренних дел, на которого возложены полномочия по проведению дознания: милиции общественной безопасности (021), криминальной милиции (022), в т.ч. подразделений по БОП (012); таможенных органов (009); органов ФССП (025); ГПС МЧС (026); ФСКН (086); ФСБ (014), ПС ФСБ (027); ФСИН (018) 10

11. Дата возбуждения уголовного дела (1), вынесения постановления об отказе в возбуждении уголовного дела (3)
"___" ____________ 200__ г. 11 год мес. чис.

12. ОПИСАНИЕ (КРАТКАЯ ФАБУЛА) ПРЕСТУПЛЕНИЯ, МЕСТО, ДАТА И ВРЕМЯ ЕГО СОВЕРШЕНИЯ ________ год мес. чис. часы 12

ст. зн. ч. п.

13. КВАЛИФИКАЦИЯ ПРЕСТУПЛЕНИЯ ст. _______ зн. ______ ч. _______ п. _______ УК 13

13.1 КВАЛИФИКАЦИЯ преступления, предшествовавшего легализации (отмыванию) денежных средств и иного имущества
кол-во _______ по ст. _______ зн. ______ ч. _______ п. _______ УК 13.1.1
кол-во _______ по ст. _______ зн. ______ ч. _______ п. _______ УК 13.1.2
кол-во _______ по ст. _______ зн. ______ ч. _______ п. _______ УК 13.1.3
кол-во

15. КАТЕГОРИЯ ПРЕСТУПЛЕНИЯ (ст. 15 УК): небольшой тяжести (2), средней тяжести (3), тяжкое (1), особо тяжкое (4) 15

16. По ст. 30 УК: приготовление (1), покушение (2) 16

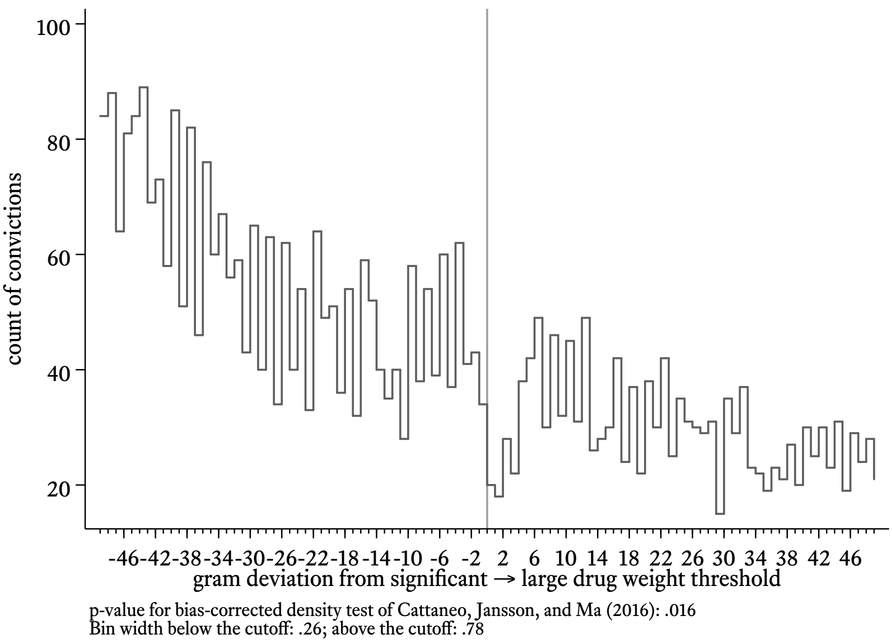
17. ПРЕСТУПЛЕНИЕ СОВЕРШЕНО в крупном (1), особо крупном (2) размере; причинен значительный (3), крупный (4), особо крупный (5) ущерб 17

18. Преступление общеуголовной (1), экономической (2) направленности, в т.ч. налоговое преступление (3) 18

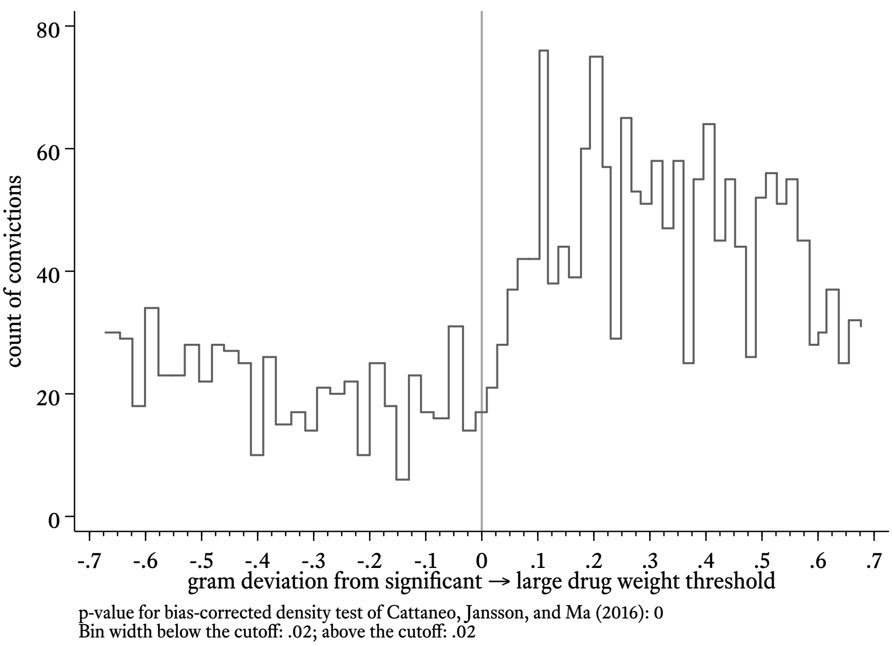
19. МЕСТО СОВЕРШЕНИЯ ПРЕСТУПЛЕНИЯ (по справочнику № 2)______________ 19

20. ОРГАНИЗАЦИОННО-ПРАВОВАЯ ФОРМА ХОЗЯЙСТВУЮЩЕГО СУБЪЕКТА (по справочнику № 11) __________ 20

FIGURE 3.A.1: This figure shows the first page of a statistical card on criminal cases filled by police investigators. Displayed page contains a form with basic crime characteristics. Source: Prikaz Genprokuratury Rossii N 39, MVD Rossii N 1070, MChS Rossii N 1021, Minjusta Rossii N 253, FSB Rossii N 780, Minjekonomrazvitija Rossii N 353, FSKN Rossii N 399 ot 29.12.2005 (red. ot 20.02.2014) "O edinom uchete prestuplenij".

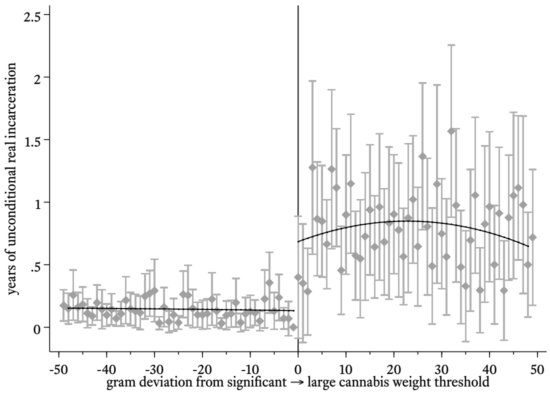


(A) CANNABIS

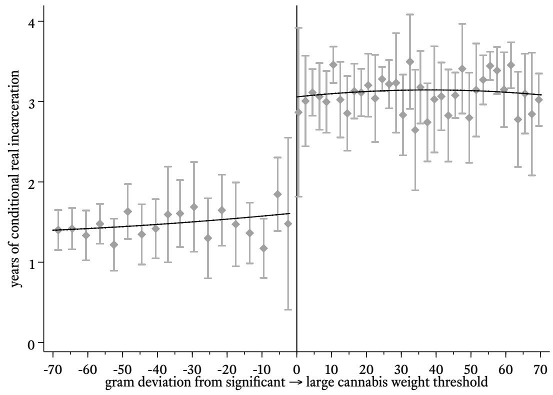


(B) HEROIN

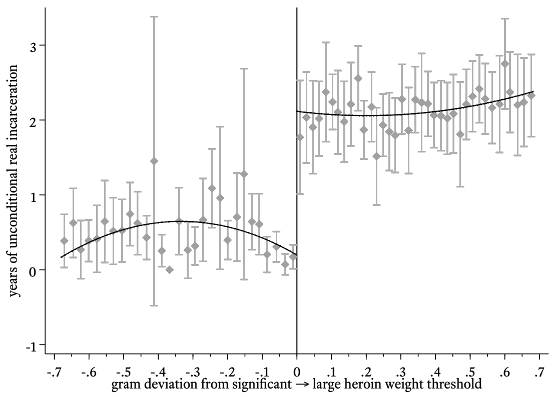
FIGURE 3.A.2: This figure reports density of cannabis- (panel A) or heroin- (panel B)-related convictions for Art. 228 p. 1 (significant weight seized, left to the zero-line) or Art. 228 p. 2 (large weight seized, right to the zero-line) by gram deviation from the significant → large drug seized weight cutoff. I also report p-values for bias-corrected density test of Cattaneo et al. (2017) with the null hypothesis of no manipulation in drug weight around the threshold. Note: figures have asymmetric bin width to the left and right of the cutoff and different x, y-scale.



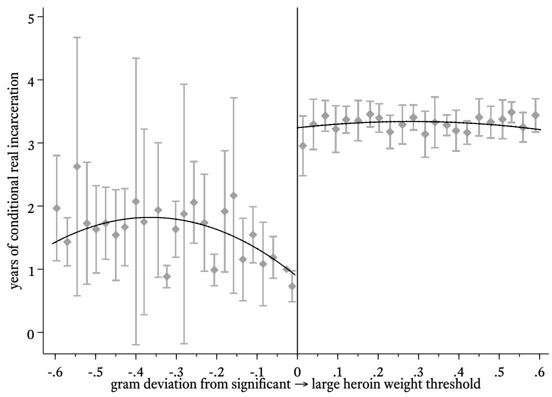
(a) Unconditional real incarceration, cannabis



(b) Conditional real incarceration, cannabis



(c) Unconditional real incarceration, heroin



(d) Conditional real incarceration, heroin

Figure 3.A.3: Solid lines in these figures show the second-order local polynomial regression of the length of unconditional (panels a, c) or conditional (panel b, d) real incarceration on weight of seized cannabis (panels a, b) or heroin (panels c, d) to the left or right of the significant → large seized drug weight threshold. Local regression is estimated conditioning on observable covariates with MSE-optimal symmetric bandwidth (Calonico et al., 2017) and triangular kernel function. Mean lengths of incarceration by seized drug weight bins (in percentage points relative to the threshold) are reported as diamonds, vertical lines with ticks show their 95% confidence intervals.

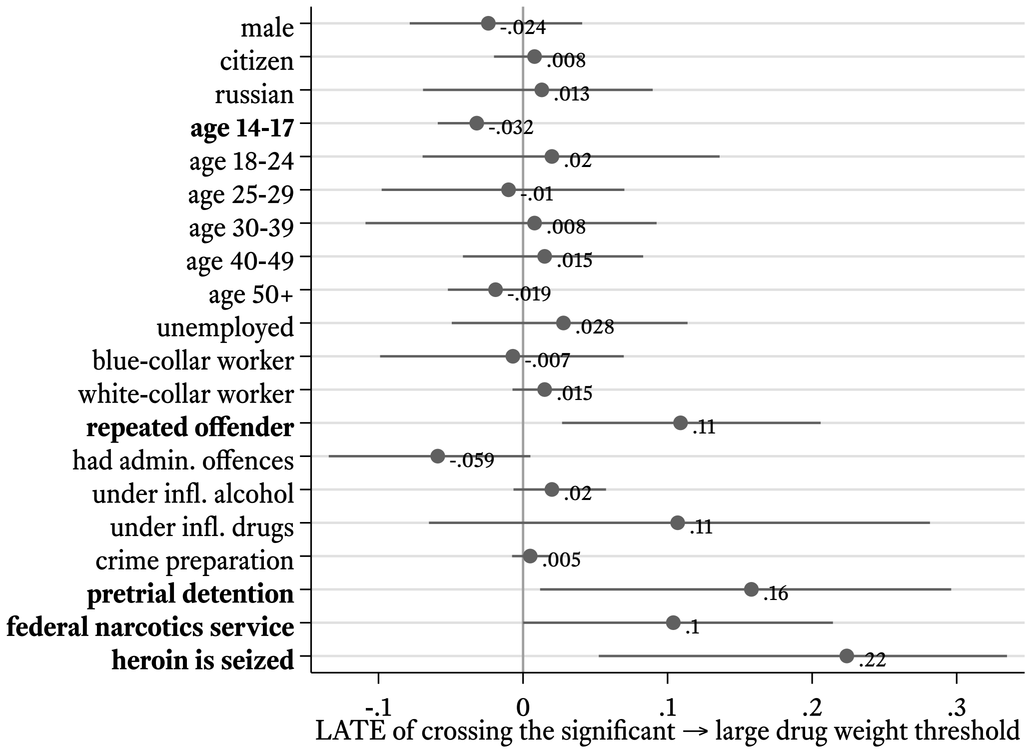


Figure 3.A.4: This figure shows the local average treatment effect of crossing the significant → large seized drug weight threshold on observable characteristics of crime or individual. Lines are 95% confidence intervals from region-cluster-robust nearest neighbour variance-covariance matrix (Calonico et al., 2014). Observable characteristics that are significantly (at 95% level) different around the cutoff are bolded.

References

Abrams, D. (2011). Is pleading really a bargain? *Journal of Empirical Legal Studies 8*(1), 200–221.

Abrams, D. (2013). Putting the trial penalty on trial. *Duquesne Law Review 51*, 777.

Abrams, D. and R. Fackler (2016). Is pleading really a bargain?: Evidence from North Carolina. mimeo.

Adelstein, R. and T. Miceli (2001). Toward a comparative economics of plea bargaining. *European Journal of Law and Economics 11*(1), 47–67.

Aizer, A. and J. Doyle (2015). Juvenile incarceration, human capital, and future crime: Evidence from randomly assigned judges. *Quarterly Journal of Economics 130*(2), 759–803.

Albonetti, C. (1997). Sentencing under the federal sentencing guidelines: Effects of defendant characteristics, guilty pleas, and departures on sentence outcomes for drug offenses, 1991-1992. *Law and Society Review 31*(4), 789–822.

Anderson, C., W. E. Deuser, and K. DeNeve (1995). Hot temperatures, hostile affect, hostile cognition, and arousal: Tests of a general model of affective aggression. *Personality and Social Psychology Bulletin 21*(5), 434–448.

Anderson, T. and C. Hsiao (1982). Formulation and estimation of dynamic models using panel data. *Journal of Econometrics 18*(1), 47–82.

Arcand, J.-L. and L. Bassole (2011). Essential heterogeneity in the impact of community driven development. Technical report, The Graduate Institute of International and Development Studies Working Paper.

Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies 58*(2), 277–297.

Arellano, M. and O. Bover (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics 68*(1), 29–51.

Armstrong, T. and C. Britt (2004). The effect of offender characteristics on offense specialization and escalation. *Justice Quarterly 21*(4), 843–876.

Auffhammer, M., S. M. Hsiang, W. Schlenker, and A. Sobel (2013). Using weather data and climate model output in economic analyses of climate change. *Review of Environmental Economics and Policy 7*(2), 181–198.

Baltagi, B. (2008). *Econometric analysis of panel data*. John Wiley & Sons.

Baron, R. and P. Bell (1976). Aggression and heat: The influence of ambient temperature, negative affect, and a cooling drink on physical aggression. *Journal of Personality and Social Psychology 33*(3), 245.

Barreca, A., M. Guldi, J. Lindo, and G. Waddell (2011). Saving babies? revisiting the effect of very low birth weight classification. *Quarterly Journal of Economics 126*(4), 2117–2123.

Barreca, A., J. Lindo, and G. Waddell (2016). Heaping-induced bias in Regression-Discontinuity designs. *Economic Inquiry 54*(1), 268–293.

Bartlett, M. (1936). The square root transformation in analysis of variance. *Supplement to the Journal of the Royal Statistical Society 3*(1), 68–78.

Bénabou, R. and J. Tirole (2011). Laws and norms. Technical report, National Bureau of Economic Research.

Berg, M. and M. DeLisi (2005). Do career criminals exist in rural America? *Journal of Criminal Justice 33*(4), 317–325.

Bjerk, D. (2014). How mandatory are mandatory minimum sentencing laws and to what extent do they impact sentence length? Evidence from United States federal drug convictions. `http://www1.cmc.edu/pages/faculty/dbjerk/MandMins4.pdf`.

Bjerk, D. (2017). Mandatory minimum policy reform and the sentencing of crack cocaine defendants: An analysis of the Fair Sentencing Act. *Journal of Empirical Legal Studies 14*(2), 370–396.

Bjerk, D. and C. Mason (2017). How uniform is justice across judicial districts? Evidence from United States federal drug convictions. `http://www1.cmc.edu/pages/faculty/dbjerk/InterDistrictSentencingVariation2.pdf`.

Blumstein, A., J. Cohen, R. J, and C. Visher (Eds.) (1986). *Criminal careers and "career criminals"*, Volume 2. National Academy Press.

Blundell, R. and S. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics 87*(1), 115–143.

Bocharov, T., V. Volkov, A. Dzmitryieva, K. Titaev, I. Chetverikova, and M. Shklyaruk (2016). Diagnostika raboty sudebnoj sistemy v sfere ugolovnogo sudoproizvodstva i predlozhenija po ee reformirovaniju. Technical report, The Institute for the Rule of Law at the European University at St. Petersburg.

Brave, S. and T. Walstrum (2014). Estimating marginal treatment effects using parametric and semiparametric methods. *Stata Journal 14*(1), 191–217.

Brereton, D. and J. D. Casper (1982). Does it pay to plead guilty? Differential sentencing and the functioning of criminal courts. *Law and Society Review 16*(1), 45–70.

Brinch, C., M. Mogstad, and M. Wiswall (2015). Beyond LATE with a discrete instrument. *Journal of Political Economy*, forthcoming.

Britt, C. (1996). The measurement of specialization and escalation in the criminal career: An alternative modeling strategy. *Journal of Quantitative Criminology 12*(2), 193–222.

Brown, R. and A. Velásquez (2017). The effect of violent crime on the human capital accumulation of young adults. *Journal of Development Economics 127*, 1–12.

Burnham, W. and J. Kahn (2008). Russia's Criminal Procedure Code five years out. *Review of Central and East European Law 33*(1), 1–94.

Burton, S., M. Finn, D. Livingston, K. Scully, W. Bales, and K. Padgett (2004). Applying a crime seriousness scale to measure changes in the severity of offenses by individuals arrested in Florida. *Justice Research and Policy 6*(1), 1–18.

Bushway, S. and A. Redlich (2012). Is plea bargaining in the "shadow of the trial" a mirage? *Journal of Quantitative Criminology 28*(3), 437–454.

Butke, P. and S. Sheridan (2010). An analysis of the relationship between weather and aggressive crime in Cleveland, Ohio. *Weather, Climate, and Society 2*(2), 127–139.

Calonico, S., M. Cattaneo, M. Farrell, and R. Titiunik (2017). `rdrobust`: Software for Regression Discontinuity Designs. *Stata Journal*.

Calonico, S., M. Cattaneo, and R. Titiunik (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica 82*(6), 2295–2326.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik (2016). Regression discontinuity designs using covariates.

Carneiro, P., J. Heckman, and E. Vytlacil (2011). Estimating marginal returns to education. *American Economic Review 101*(6), 2754–2781.

Carneiro, P., M. Lokshin, and N. Umapathi (2017). Average and marginal returns to upper secondary schooling in Indonesia. *Journal of Applied Econometrics 32*(1), 16–36.

Carrington, P. (2013). Trends in the seriousness of youth crime in Canada, 1984–2011. *Canadian Journal of Criminology and Criminal Justice 55*(2), 293–314.

Carrington, P., A. Matarazzo, and P. DeSouza (2005). *Court careers of a Canadian birth cohort*. Statistics Canada, Canadian Centre for Justice Statistics.

Cattaneo, M., M. Jansson, and X. Ma (2017). `rddensity`: Manipulation testing based on density discontinuity. Technical report, University of Michigan Working Paper.

Cihan, A., J. Sorensen, and K. Chism (2017). Analyzing the offending activity of inmates: Trajectories of offense seriousness, escalation, and de-escalation. *Journal of Criminal Justice 50*, 12–18.

Clarke, S., U. of North Carolina at Chapel Hill School of Government, and U. S. of America (1982). *Alaska Plea Bargaining Study, 1974-1976*. Inter-university Consortium for Political and Social Research #07714.

Cohn, E. (1990). Weather and crime. *British journal of Criminology 30*(1), 51–64.

Cook, R. and L. Ni (2005). Sufficient dimension reduction via inverse regression: A minimum discrepancy approach. *Journal of the American Statistical Association 100*(470), 410–428.

Cooter, R. (1998). Expressive law and economics. *Journal of Legal Studies 27*(S2), 585–607.

Cooter, R. and I. Bohnet (2001). Expressive law: Framing or equilibrium selection? Berkeley Program in Law and Economics, Working Paper #31.

Cornelissen, T., C. Dustmann, A. Raute, and U. Schönberg (2016). From late to mte: Alternative methods for the evaluation of policy interventions. *Labour Economics 41*, 47–60.

Correia, S. (2015). Singletons, cluster-robust standard errors and fixed effects: A bad mix. `http://scorreia.com/research/singletons.pdf`.

Criminal Code, . (2012). Criminal Code of the Russian Federation No. 63-FZ of June 13, 1996 (as last amended on March 1, 2012). http://legislationline.org/download/action/download/id/4247/file/RF_CC_1996_am03.2012_en.pdf.

Deaton, A. (2009). Instruments of development: Randomization in the tropics, and the search for the elusive keys to economic development. Technical report, National Bureau of Economic Research.

Dee, D., S. Uppala, A. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. Balmaseda, G. Balsamo, P. Bauer, et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society 137*(656), 553–597.

DeFronzo, J. (1984). Climate and crime: Tests of an FBI assumption. *Environment and Behavior 16*(2), 185–210.

Dobbie, W. and J. Song (2015). Debt relief and debtor outcomes: Measuring the effects of consumer bankruptcy protection. *American Economic Review 105*(3), 1272–1311.

Eastwood, N., E. Fox, and A. Rosmarin (2016). A quiet revolution: Drug decriminalisation across the globe. *London: Release*.

Eisenhauer, P., J. Heckman, and E. Vytlacil (2015). The Generalized Roy Model and the cost-benefit analysis of social programs. *Journal of Political Economy 123*(2), 413–443.

Eisenstein, J. and H. Jacob (1977). *Felony justice: An organizational analysis of criminal courts*. Little, Brown Boston.

Esakov, G., M. Dolotov, M. Filatova, M. Redchic, P. Stepanov, and K. Tsay (2017). Ugolovnaja politika: Dorozhnaja karta (2017–2025 gg.). Technical report, Centr strategicheskih razrabotok.

Everaert, G. (2013). Orthogonal to backward mean transformation for dynamic panel data models. *Econometrics Journal 16*(2), 179–221.

Fabri, M. (2008). Criminal procedure and public prosecution reform in Italy: A flash back. *International Journal for Court Administration 1*(1), 3–15.

Famega, C. and C. Gaines (2013). *Drugs, Crime, and Justice: Contemporary Perspectives* (3 ed.)., Chapter Explaining Drug Crime with Criminological Theory, pp. 74–104. Waveland Press.

Farrington, D. (1986). Age and crime. *Crime and Justice 7*, 189–250.

Farrington, D. (2003). Developmental and life-course criminology: Key theoretical and empirical issues — the 2002 Sutherland Award address. *Criminology 41*(2), 221–225.

Francis, B. and J. Liu (2015). Modelling escalation in crime seriousness: A latent variable approach. *Metron 73*(2), 277–297.

Francis, B., K. Soothill, L. Humphreys, and A. Cutajar Bezzina (2005). Developing measures of severity and frequency of reconviction. Technical report, Lancaster University.

Friedman, J., T. Hastie, and R. Tibshirani (2009). `glmnet`: Lasso and elastic-net regularized generalized linear models. *R package 1*(4).

Galbiati, R. and P. Vertova (2008). Obligations and cooperative behaviour in public good games. *Games and Economic Behavior 64*(1), 146–170.

Gentzkow, M., B. Kelly, and M. Taddy (2017). Text as data. Technical report, National Bureau of Economic Research.

Gibbs, J. (1968). Crime, punishment, and deterrence. *Southwestern Social Science Quarterly 48*(4), 515–530.

Givati, Y. (2014). Legal institutions and social values: Theory and evidence from plea bargaining regimes. *Journal of Empirical Legal Studies 11*(4), 867–893.

Gneezy, U. and A. Rustichini (2000). A fine is a price. *Journal of Legal Studies 29*, 1.

Govindarajulu, U., E. Malloy, B. Ganguli, D. Spiegelman, and E. Eisen (2009). The comparison of alternative smoothing methods for fitting non-linear exposure-response relationships with Cox models in a simulation study. *International Journal of Biostatistics 5*(1).

Greenacre, M. (2009). Power transformations in correspondence analysis. *Computational Statistics & Data Analysis 53*(8), 3107–3116.

Greenacre, M. (2017). *Correspondence analysis in practice*. CRC press.

Greene, W. (2004). The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects. *Econometrics Journal 7*(1), 98–119.

Grossman, G. and M. Katz (1983). Plea bargaining and social welfare. *American Economic Review 73*(3), 749–757.

Han, C., P. Phillips, and D. Sul (2014). X-differencing and dynamic panel model estimation. *Econometric Theory 30*(1), 201–251.

Harris, R. and F. Springer (1984). Plea bargaining as a game: An empirical analysis of negotiated sentencing decisions. *Review of Policy Research 4*(2), 245–258.

Hartley, R., S. Maddan, and C. Spohn (2007). Prosecutorial discretion: An examination of substantial assistance departures in federal crack-cocaine and powder-cocaine cases. *Justice Quarterly 24*(3), 382–407.

Hauser, W. (2012). *Do Judges' Experiences And Indelible Traits Influence Sentencing Decisions? New Evidence From Florida.* Ph. D. thesis, Florida State University.

Heckman, J. (1997). Instrumental variables: A study of implicit behavioral assumptions in one widely used estimator. *Journal of Human Resources 32*(3), 441–462.

Heckman, J., H. Ichimura, and P. Todd (1997). How details makes a difference: Semiparametric estimation of the partially linear regression model. Unpublished manuscript, University of Chicago, Department of Economics.

Heckman, J., S. Urzua, and E. Vytlacil (2006). Estimation of treatment effects under essential heterogeneity. `http://jenni.uchicago.edu/underiv/documentation_2006_03_20.pdf`.

Heckman, J. and E. Vytlacil (1999). Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the National Academy of Sciences of the United States of America 96*(8), 4730–4734.

Heckman, J. and E. Vytlacil (2005). Structural equations, treatment effects, and econometric policy evaluation. *Econometrica 73*(3), 669–738.

Heckman, J. and E. Vytlacil (2007). Econometric evaluation of social programs, part II: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments. *Handbook of econometrics 6*, 4875–5143.

Hurwicz, L. (1950). Least squares bias in time series. *Statistical inference in dynamic economic models 10*, 365–383.

Imbens, G. and J. Angrist (1994). Identification and estimation of local average treatment effects. *Econometrica 62*(2), 467–475.

Imbens, G. and T. Lemieux (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics 142*(2), 615–635.

Jacob, B., L. Lefgren, and E. Moretti (2007). The dynamics of criminal behavior evidence from weather shocks. *Journal of Human resources 42*(3), 489–527.

Kazemian, L., D. Farrington, and M. Le Blanc (2009). Can we make accurate long-term predictions about patterns of de-escalation in offending behavior? *Journal of Youth and Adolescence 38*(3), 384–400.

Kim, A. (2015). Underestimating the trial penalty: An empirical analysis of the federal trial penalty and critique of the Abrams study. *Mississippi Law Journal 84*(5), 1195–1255.

Kleibergen, F. and R. Paap (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics 133*(1), 97–126.

Knorre, A. and D. Skougarevskiy (2015). Kak MVD i FSKN borjutsja s narkotikami: Sravnitel'nyj analiz rezul'tativnosti dvuh vedomstv. Technical report, The Institute for the Rule of Law at the European University at St. Petersburg, (ed. Maria Shklyaruk), http://enforce.spb.ru/images/analit_zapiski/FSKN_MVD_memo_2015_web.pdf.

Kolesár, M. (2013). Estimation in an instrumental variables model with treatment effect heterogeneity. *mimeo*.

Krishnakumar, J. (2006). Time invariant variables and panel data models: A Generalised Frisch–Waugh theorem and its implications. *Contributions to Economic Analysis 274*, 119–132.

Kyui, N. (2016). Expansion of higher education, employment and wages: Evidence from the russian transition. *Labour Economics 39*, 68–87.

Kyvsgaard, B. (2002). *The criminal career: The Danish longitudinal study*. Cambridge University Press.

Landes, W. (1971). An economic analysis of the courts. *Journal of Law and Economics 14*(1), 61–107.

Langer, M. (2004). From legal transplants to legal translations: The globalization of plea bargaining and the Americanization thesis in criminal procedure. *Harvard International Law Journal 45*(1), 1–64.

Le Blanc, M. (2002). The offending cycle, escalation and de-escalation in delinquent behavior: A challenge for criminology. *International Journal of Comparative and Applied Criminal Justice 26*(1), 53–83.

Le Blanc, M. (2014). Developmental criminology: Thoughts on the past and insights for the future. In J. Morizot and L. Kazemian (Eds.), *The Development of Criminal and Antisocial Behavior*, pp. 507–537. Springer.

Lee, D. and T. Lemieux (2010). Regression discontinuity designs in economics. *Journal of Economic Literature 48*, 281–355.

Levinson, L. (2008). Half a gram — a thousand lives. *Harm Reduction Journal 5*(1), 22.

Liu, J., B. Francis, and K. Soothill (2011). A longitudinal study of escalation in crime seriousness. *Journal of Quantitative Criminology 27*(2), 175–196.

Lott, J. (1992). Do we punish high income criminals too heavily? *Economic Inquiry 30*(4), 583.

Macfarquhar, N. (2016). HIV cases surpass a million in Russia, but little is done. *New York Times*.

Merryman, J. and R. Pérez-Perdomo (2007). *The civil law tradition: An introduction to the legal systems of Europe and Latin America*. Stanford University Press.

Miller, H., W. McDonald, and J. Cramer (1980). *Plea Bargaining in the United States, 1978*. Inter-university Consortium for Political and Social Research #07775.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica 46*(1), 69–85.

Mustard, D. (2001). Racial, ethnic, and gender disparities in sentencing: Evidence from the us federal courts. *Journal of Law and Economics 44*(1), 285–314.

Nagin, D. and M. Snodgrass (2013). The effect of incarceration on re-offending: Evidence from a natural experiment in Pennsylvania. *Journal of Quantitative Criminology 29*(4), 601–642.

Nance, D. (1997). Guidance rules and enforcement rules: A better view of the cathedral. *Virginia Law Review 83*(5), 837–937.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica 49*(6), 1417–1426.

Nikonov, M. (2017). Razvitie sistemy ugolovnogo pravosudija: Vektory, mery reformirovanija, osnovnye igroki. Technical report, Centr strategicheskih razrabotok.

Osgood, W. and C. Schreck (2007). A new method for studying the extent, stability, and predictors of individual specialization in violence. *Criminology 45*(2), 273–312.

Padgett, J. (1985). The emergent organization of plea bargaining. *American Journal of Sociology 90*(4), 753–800.

Perry, J. D. and M. Simpson (1987). Violent crimes in a city: Environmental determinants. *Environment and Behavior 19*(1), 77–90.

Piquero, A., R. Brame, J. Fagan, and T. Moffitt (2006). Assessing the offending activity of criminal domestic violence suspects: Offense specialization, escalation, and de-escalation evidence from the spouse assault replication program. *Public Health Reports 121*(4), 409–418.

Piquero, A. and H. L. Chung (2001). On the relationships between gender, early onset, and the seriousness of offending. *Journal of Criminal Justice 29*(3), 189–206.

Posner, R. (1983). *The economics of justice*. Harvard University Press.

Posner, R. (1997). Social norms and the law: An economic approach. *American Economic Review 87*(2), 365–369.

Priest, G. and B. Klein (1984). The selection of disputes for litigation. *Journal of Legal Studies 13*(1), 1–55.

Ramchand, R., J. MacDonald, A. Haviland, and A. Morral (2009). A developmental approach for measuring the severity of crimes. *Journal of Quantitative Criminology 25*(2), 129–153.

Ranson, M. (2014). Crime, weather, and climate change. *Journal of Environmental Economics and Management 67*(3), 274–302.

Reilly, B. and R. Witt (1996). Crime, deterrence and unemployment in England and Wales: An empirical analysis. *Bulletin of Economic Research 48*(2), 137–159.

Rhodes, W. (1979). Plea bargaining: Its effect on sentencing and convictions in the District of Columbia. *Journal of Criminal Law and Criminology 70*, 360.

RIA Novosti (2015). Devjat' policejskih reshili uvolit'sja posle tragedii v Nizhnem Novgorode. `https://ria.ru/incidents/20150810/1175814859.html`.

Ringrose, T. (2012). Bootstrap confidence regions for correspondence analysis. *Journal of Statistical Computation and Simulation 82*(10), 1397–1413.

Rojek, D. and M. Erickson (1982). Delinquent careers a test of the career escalation model. *Criminology 20*(1), 5–28.

Roy, A. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers 3*(2), 135–146.

Rubinstein, M. and T. White (1978). Alaska's ban on plea bargaining. *Law & Society Review 13*, 367.

Sacconi, L. and M. Faillo (2010). Conformity, reciprocity and the sense of justice. how social contract-based preferences and beliefs explain norm compliance: The experimental evidence. *Constitutional Political Economy 21*(2), 171–201.

Scott, R. and W. Stuntz (1992). Plea bargaining as contract. *Yale Law Journal 101*(8), 1909–1968.

Sherman, L., J. Schmidt, D. Rogan, and P. Gartin (1991). From initial deterrence to long-term escalation: Short-custody arrest for poverty ghetto domestic violence. *Criminology 29*(4), 821–850.

Shklyaruk, M. (2014). Traektorija ugolovnogo dela v oficial'noj statistike na primere obobshhennyh dannyh pravoohranitel'nyh organov. Technical report, The Institute for the Rule of Law at the European University at St. Petersburg, `http://enforce.spb.ru/products/papers/6265-traektoriya-ugolovnogo-dela-v-ofitsialnoj-statistike`.

Shklyaruk, M. and D. Skougarevskiy (2015). Kriminal'naja statistika: Mehanizmy formirovanija, prichiny iskazhenija, puti reformirovanija. Technical report, The Institute for the Rule of Law at the European University at St. Petersburg, `http://enforce.spb.ru/images/Products/Crimestat_report_2015_IRL_KGI_web.pdf`.

Skougarevskiy, D. (2014). Extra jus: Nenuzhnaja gumanizacija vzjatok. Technical report, Vedomosti #3730 of 04.12.2014, `https://www.vedomosti.ru/opinion/articles/2014/12/04/nenuzhnaya-gumanizaciya-vzyatok`.

Skougarevskiy, D. and V. Volkov (2014). Criminal justice in Russia: Towards a unified sentencing model. The Institute for the Rule of Law at the European University at St. Petersburg Working Paper #4.

Skougarevskiy, D., V. Volkov, A. Dzmitryieva, K. Titaev, I. Chetverikova, and Y. Shesternina (2014). *Ugolovnaya uisticiya Rossii v 2009 gody: Kompleksnii analiz sudebnoi statistiki*. The Institute for the Rule of Law at the European University at St. Petersburg.

Smirnov, A. and K. Kalinovskiy (2012). *Ugolovnii process: uchebnik dlya vuzov*. Moscow: Norma.

Smith, D. (1986). The plea bargaining controversy. *Journal of Criminal Law and Criminology 77*(3), 949–968.

Solomon, P. (1987). The case of the vanishing acquittal: Informal norms and the practice of Soviet criminal justice. *Europe-Asia Studies 39*(4), 531–555.

Solomon, P. (2012). Plea bargaining Russian style. *Demokratizatsiya 20*(3), 282.

Spohn, C. and J. Cederblom (1991). Race and disparities in sentencing: A test of the liberation hypothesis. *Justice Quarterly 8*(3), 305–327.

Steffensmeier, D., J. Kramer, and C. Streifel (1993). Gender and imprisonment decisions. *Criminology 31*(3), 411–446.

Steffensmeier, D., J. Ulmer, and J. Kramer (1998). The interaction of race, gender, and age in criminal sentencing: The punishment cost of being young, black, and male. *Criminology 36*(4), 763–798.

Sunstein, C. (1996). Social norms and social roles. *Columbia Law Review 96*(4), 903–968.

Taddy, M. (2013). Multinomial inverse regression for text analysis. *Journal of the American Statistical Association 108*(503), 755–770.

Taddy, M. (2015). Distributed multinomial regression. *The Annals of Applied Statistics 9*(3), 1394–1414.

Tata, C. and J. Gormley (2016). Sentencing and plea bargaining: Guilty pleas versus trial verdicts. *Oxford Handbooks Online*.

Titaev, K. and M. Shklyaruk (2016). *Rossijskij sledovatel: Prizvanie, professija, povsednevnost*. Moscow: Norma.

Ulmer, J. (2012). Recent developments and new directions in sentencing research. *Justice Quarterly 29*(1), 1–40.

Ulmer, J. and M. Bradley (2006). Variation in trial penalties among serious violent offences. *Criminology 44*(3), 631–670.

Ulmer, J., J. Eisenstein, and B. Johnson (2010). Trial penalties in federal sentencing: extra-guidelines factors and district variation. *Justice Quarterly 27*(4), 560–592.

Veselov, V. and I. Pribylskaya (2015). AISORI: Automated information system for processing of regime data. `http://meteo.ru/it/178-aisori`, `http://aisori.meteo.ru/ClimateR`.

Vickers, C. (2012). Plea bargaining and sentencing discrimination: Evidence from England and Wales, 1870-1910. unpublished, Northwestern University.

Volkov, V. (2016). Legal and extralegal origins of sentencing disparities: Evidence from Russia's criminal courts. *Journal of Empirical Legal Studies 13*(4), 637–665.

Volkov, V., A. Dzmitryieva, K. Titaev, E. Khodzhaeva, I. Chetverikova, and M. Shklyaruk (2015). Diagnostika raboty pravoohranitel'nyh organov po ohrane obshhestvennogo porjadka i perspektivy sozdanija municipal'noj milicii v Rossii. Technical report, The Institute for the Rule of Law at the European University at St. Petersburg, `http://enforce.spb.ru/images/Products/report_municipal_IRL_KGI_2015_web.pdf`.

Volkov, V., I. Grigoriev, A. Dzmitryieva, E. Moiseeva, E. Paneyakh, M. Pozdnyakov, K. Titaev, I. Chetverikova, and M. Shklyaruk (2013). Koncepcija kompleksnoj organizacionno-upravlencheskoj reformy pravoohranitel'nyh organov Rossijskoj Federacii. Technical report, The Institute for the Rule of Law at the European University at St. Petersburg.

Wagner, H. and K. Pease (1978). On adding up scores of offence seriousness. *British Journal of Criminology 18*(2), 175–178.

Wolpin, K. (1978). An economic analysis of crime and punishment in England and Wales, 1894-1967. *Journal of Political Economy 86*(5), 815–840.

Wood, S. (2004). Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association 99*(467), 673–686.

Yang, C. (2016). Resource constraints and the criminal justice system: Evidence from judicial vacancies. *American Economic Journal: Economic Policy 8*(4), 289–332.

Yeh, B. (2015). Drug offenses: Maximum fines and terms of imprisonment for violation of the Federal Controlled Substances Act and related laws. Technical report, Congressional Research Service.