Is Pleading a Bargain under Essential Heterogeneity?*

Dmitriy Skougarevskiy†

September 5, 2017

Abstract

This paper 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.

*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. Financial support from the Russian Science Foundation grant 17–18–01618 is acknowledged.

†European University at Saint-Petersburg, 3 Gagarinskaya St., Saint Petersburg, 191187 Russia & Graduate Institute of International and Development Studies.

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 nature (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

[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.

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 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 paper 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 rele-

vant, 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 paper proceeds as follows. Section 1 introduces the model, estimation technique, and the instrument, Section 2 describes the gathered data, institutional contexts, and treatment variables. Section 3 presents the results and offers a discussion.

1. Model of plea discount

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

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

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).

[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).

Pleading guilty is determined by a latent variable

$$D^*_{i,j} = [X_{i,j}, Z_{i,j}]\,\gamma - V_{i,j}, \tag{2}$$

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 (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:

$$\begin{aligned} U_{1,i,j} &= \lambda_j + \xi_1 V_{i,j} + \varepsilon_{1,i,j} \\ U_{0,i,j} &= \lambda_j + \xi_2 V_{i,j} + \varepsilon_{0,i,j} \end{aligned} \tag{3}$$

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

I impose the *conditional independence condition* $(U_{0,i,j}, U_{1,i,j}, V_{i,j}) \perp Z_{i,j} | X_{i,j}$. 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) (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) in the switching regression framework:

$$\begin{aligned} Y_{i,j} &= D_{i,j} Y_{1,i,j} + (1 - D_{i,j})\, Y_{0,i,j} \\ &= D_{i,j}\left(\alpha_1 + \beta_1 X_{i,j} + U_{1,i,j}\right) + (1 - D_{i,j})\left(\alpha_0 + \beta_0 X_{i,j} + U_{0,i,j}\right) \\ &= \alpha_0 + \beta_0 X_{i,j} + D_{i,j}\underbrace{\left((\alpha_1 - \alpha_0) + (\beta_1 - \beta_0)\, X_{i,j} + (U_{1,i,j} - U_{0,i,j})\right)}_{\Delta} + U_{0,i,j}, \end{aligned} \tag{4}$$

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.

1.2. 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]

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

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

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

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

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 (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:

$$
\begin{aligned}
\hat{\Delta}^{IV}(x, Z_{i,j}) &= \frac{cov(Z_{i,j}, Y_{i,j})}{cov(Z_{i,j}, D_{i,j})} \\
&= \frac{cov(Z_{i,j}, \alpha_0 + \beta_0 x + D_{i,j}((\alpha_1 - \alpha_0) + (\beta_1 - \beta_0)x + (U_{1,i,j} - U_{0,i,j})) + U_{0,i,j})}{cov(Z_{i,j}, D_{i,j})} \\
&= \frac{\begin{array}{c} cov(Z_{i,j}, \alpha_0 + \beta_0 x + D_{i,j}((\alpha_1 - \alpha_0) + (\beta_1 - \beta_0)x)) \\ +cov(Z_{i,j}, D_{i,j}(U_{1,i,j} - U_{0,i,j}) + U_{0,i,j}) \end{array}}{cov(Z_{i,j}, D_{i,j})} \\
&= \text{ATE}(x) \times \frac{cov(Z_{i,j}, D_{i,j})}{cov(Z_{i,j}, D_{i,j})} + \frac{cov(Z_{i,j}, (U_{1,i,j} - U_{0,i,j})D_{i,j})}{cov(Z_{i,j}, D_{i,j})} \\
&= \text{ATE}(x) + \frac{cov(Z_{i,j}, (U_{1,i,j} - U_{0,i,j})D_{i,j})}{cov(Z_{i,j}, D_{i,j})}
\end{aligned}
\tag{5}
$$

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.

1.3. 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 paper: 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 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 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.

1.4. Local average and Marginal treatment effects

LATE To further show that the IV estimator (5) uncovers the local average treatment effect, recall from (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 $F\left[[X_{i,j}, Z_{i,j}]\,\gamma\right] > F[V_{i,j}]$. 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 $P(X_{i,j}, Z_{i,j}|X_{i,j} = x, Z_{i,j} = z)$. 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 $F[V_{i,j}] \equiv U_{D,i,j}$. To reiterate, for an offender with observables x, z, and unobserved resistance to pleading guilty u_D:

$$\begin{aligned} [X_{i,j}, Z_{i,j}]\,\gamma &> V_{i,j} \\ F\left[[X_{i,j}, Z_{i,j}]\,\gamma\right] &> F[V_{i,j}] \\ P\left([X_{i,j}|Z_{i,j}]\,\gamma|X_{i,j} = x, Z_{i,j} = z\right) &> U_{D,i,j} = u_D \end{aligned} \tag{6}$$

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 (5) can be rewritten as

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

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: $X = \emptyset$):

$$\begin{aligned}\hat{\Delta}_{LATE}^{IV}\left(z,z^{'}\right) &= \frac{cov\left(Z_{i,j},Y_{i,j}\right)}{cov\left(Z_{i,j},D_{i,j}\right)} = \frac{cov\left(Z_{i,j},\Delta D_{i,j}\right)}{cov\left(Z_{i,j},D_{i,j}\right)} \\ &\overset{def}{=} \frac{E\left[\Delta D_{i,j}Z_{i,j}\right] - E\left[\Delta D_{i,j}\right]E\left[Z_{i,j}\right]}{E\left[D_{i,j}Z_{i,j}\right] - E\left[D_{i,j}\right]E\left[Z_{i,j}\right]} \\ &= E\left[\Delta | z < V_{i,j} \leq z^{'}\right]\end{aligned} \quad , \tag{7}$$

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 (4) as

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

Then replace the treatment dummy with its propensity from (6) and take the conditional (on observables) expectation in terms of the unobservables:

$$E\left[Y_{i,j}|X_{i,j} = x, Z_{i,j} = z, P\left(X_{i,j}, Z_{i,j}\right) = p\right] = \alpha_0 + x\beta_0 + px_{i,j}\left(\beta_1 - \beta_0\right) + K\left(p\right), \quad (8)$$

where all non-linear terms are aggregated in $K\left(p\right) \equiv p\left(\alpha_1 - \alpha_0\right) + E\left[U_{0,i,j}|P\left(X_{i,j}, Z_{i,j}\right) = p\right] + pE\left[\left(U_{1,i,j} - U_{0,i,j}\right)\right]$. 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:

$$\begin{aligned}\hat{\Delta}^{MTE}(x, z, u_D) &\equiv \left.\frac{\partial E\left[Y_{i,j} | X_{i,j} = x, Z_{i,j} = z, P\left(X_{i,j}, Z_{i,j}\right) = p\right]}{\partial p}\right|_{p=u_D} \\ &= x\left(\beta_1 - \beta_0\right) + \left.\frac{\partial K(p)}{\partial p}\right|_{p=u_{D,i,j}}\end{aligned} \tag{9}$$

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 $\partial Y = \partial p \times \text{MTE}\left(u_D = p_0\right)$. 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 (8) is non-linear in p, taking its derivative (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 54.

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 (8) in isolation, but jointly with the propensity equation (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 (9) are obtained, one can integrate it at specific regions where $\widehat{P}(Z) < u_D$,$\widehat{P}(Z) > u_D$, or $\widehat{P}(Z) \lessgtr u_D$ 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.

2. Data

This paper 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.

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.

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

[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.

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 paper now builds.

Data cleaning The accessed data is of high granularity and turns out to contain errors. I have developed a data cleaning procedure[6] that removed approximately 5% of records. Further removal was due to duplicate detection: the data collection system had no ver-

[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 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.

sion 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). Criminal Procedure Code stipulates that by pleading guilty the 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

[7]Detailed information is relegated to Supplementary appendix on page 56.
[8]Criminal Procedure Code of Russia. Article 317.

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: 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 excludes 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).

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 pleading guilty[10] and right-winsorise conditional and unconditional sentence lengths at 99%.

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

[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,

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.

However rich the data may be, it lacks 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 62), 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 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 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.

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.

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.

2.2.1. 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 A.1.

Covariates My outcome variable Y is the length of unconditional real incarceration, $X_{i,j}$ include the variables stated in Table 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 does not include the identity of the sentencing judge.

2.2.2. 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 paper 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 A.2.

Covariates Similarly, the outcome variable Y is the length of unconditional real incarceration, $X_{i,j}$ include the variables stated in Table 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.

3. Results & Discussion

3.1. Descriptive statistics

As a point of departure, consider adoption rates for plea bargaining across jurisdictions (Tables 1, A.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.

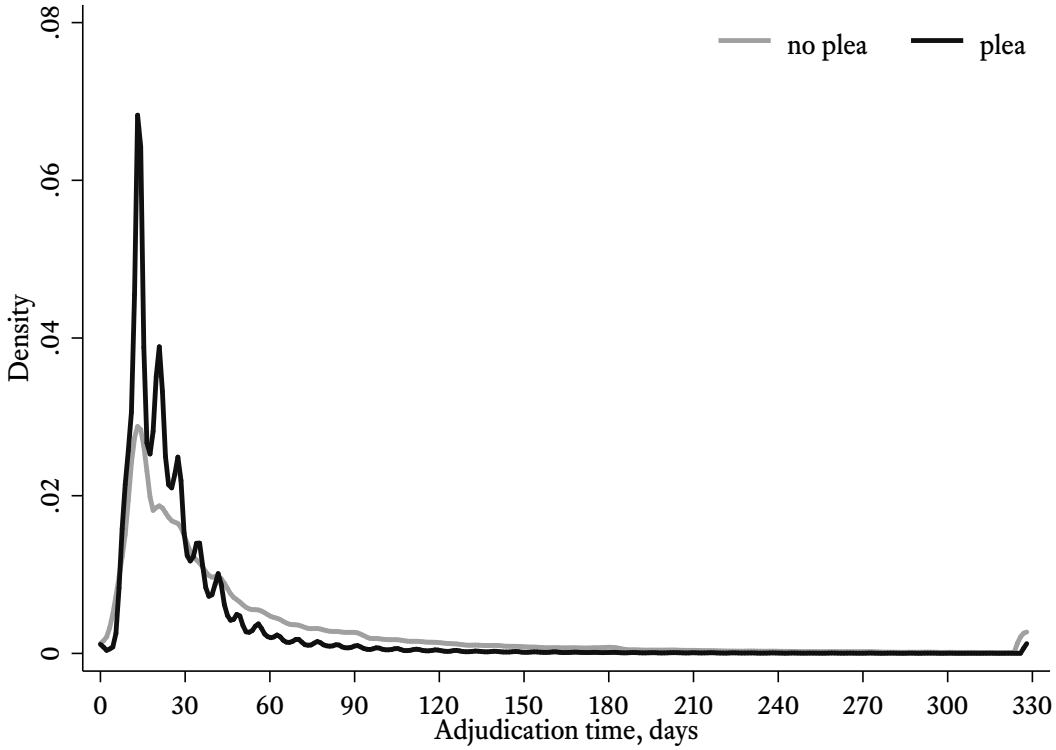
More to that, 25.1% of the defendants were alleged to have committed their crimes under the influence of alcohol in Russia. Such crimes were found to have been committed in sole fashion (87.5%), by predominantly first-time offenders (59.9%). Such configuration of a median crime — committed by a sole 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

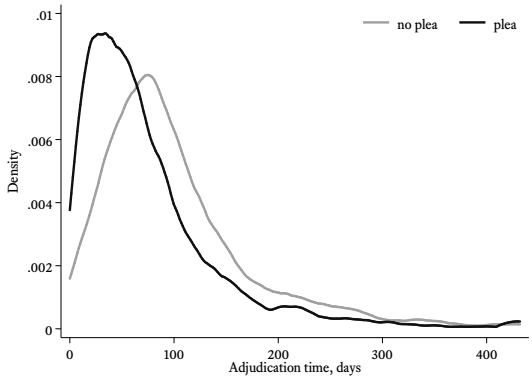
this metric in common law jurisdictions: in Miller sample it amounts to 60 days whereas in Alaska it is found to be 101 days. Clearly, some part of the difference may be explained by the focus of these studies on burglaries & robberies or felonies, respectively. However, at least in Alaska case, Rubinstein 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 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 1a, 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 exhibits 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 is 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.

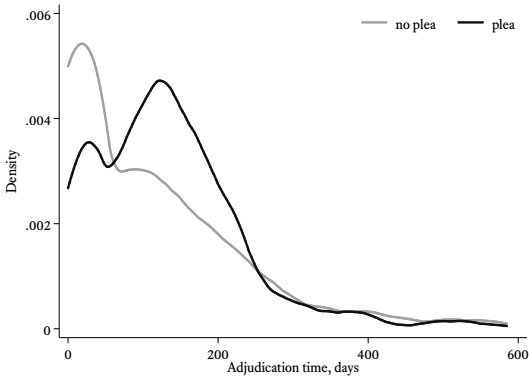
When I turn to the outcome variable of interest — years of real incarceration — any



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

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.

3.2. OLS/LATE EVIDENCE

To highlight the importance of studying unconditional lengths of real incarceration, I conduct a simple experiment in Table 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 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 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 1.2, 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 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.

Additional evidence in favour of the proposed $Z_{i,j}$ comes from Table 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

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 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 3 are included in the model but not reported. Huber-Eicker-White standard errors clustered at region level in parentheses.

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 A.3, 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 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 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 62).

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.

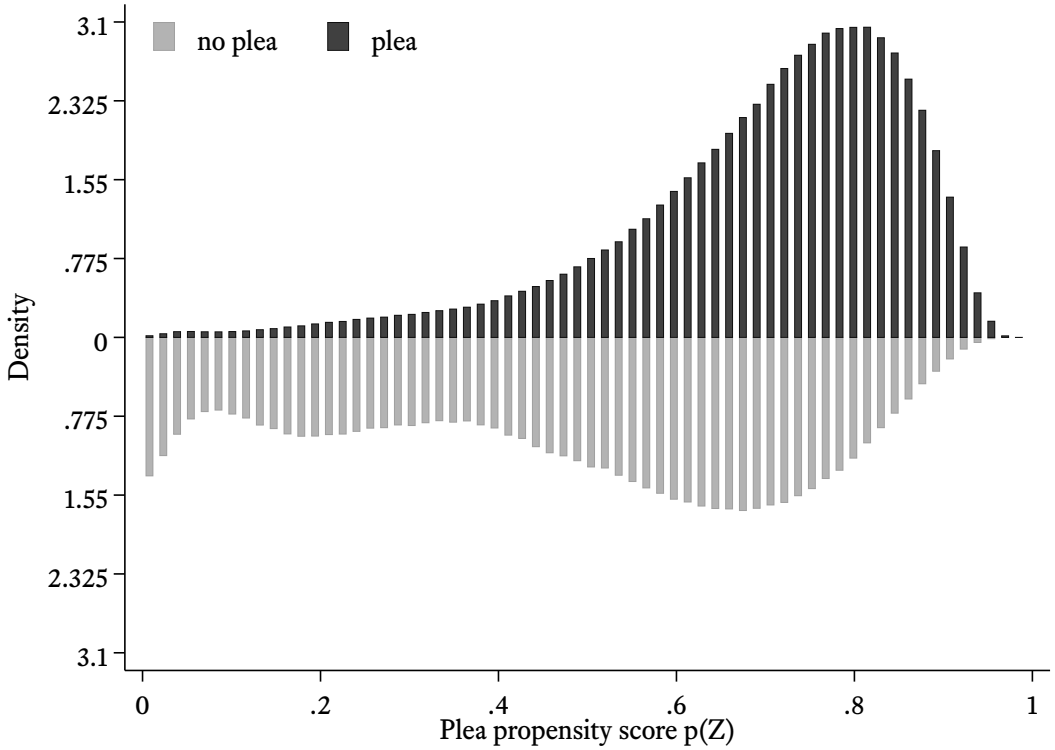
Non-separability of $P(X_{i,j}, Z_{i,j})$ $\widehat{MTE}$ can be evaluated only in the region of common support of the estimated propensity score where I observe both the decision to plead guilty and go to trial. Outside this region there exists no information on the alternative decisions (given the 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 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.

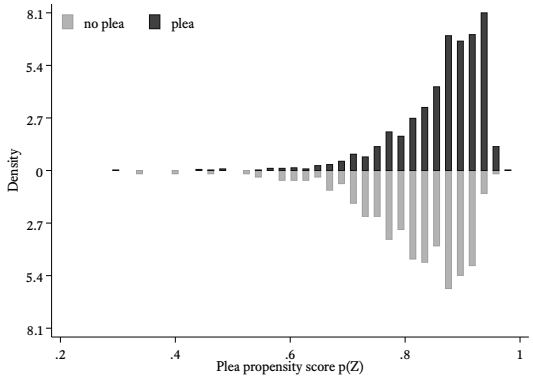
Sufficiency in identifying variation of $Z_{i,j}$ The excluded instrument $Z_{i,j}$ should be continuous and should exhibit enough variation to allow me to identify a schedule of $\widehat{MTE}$s. This condition is also testable. In Figure 2 I compare the predicted $\widehat{P}(X_{i,j}, Z_{i,j})$ (black 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 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.

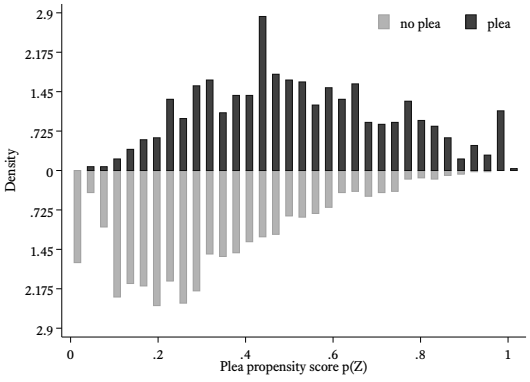
First lesson from the estimated profile in Figure 4 is that it is not flat in any jurisdiction. To see it more formally, I conduct an F-test by comparing the full model (8) where $K(p)$



(A) RUSSIA, UNIVERSE OF ELIGIBLE ACCUSED INDIVIDUALS, 2011–2013

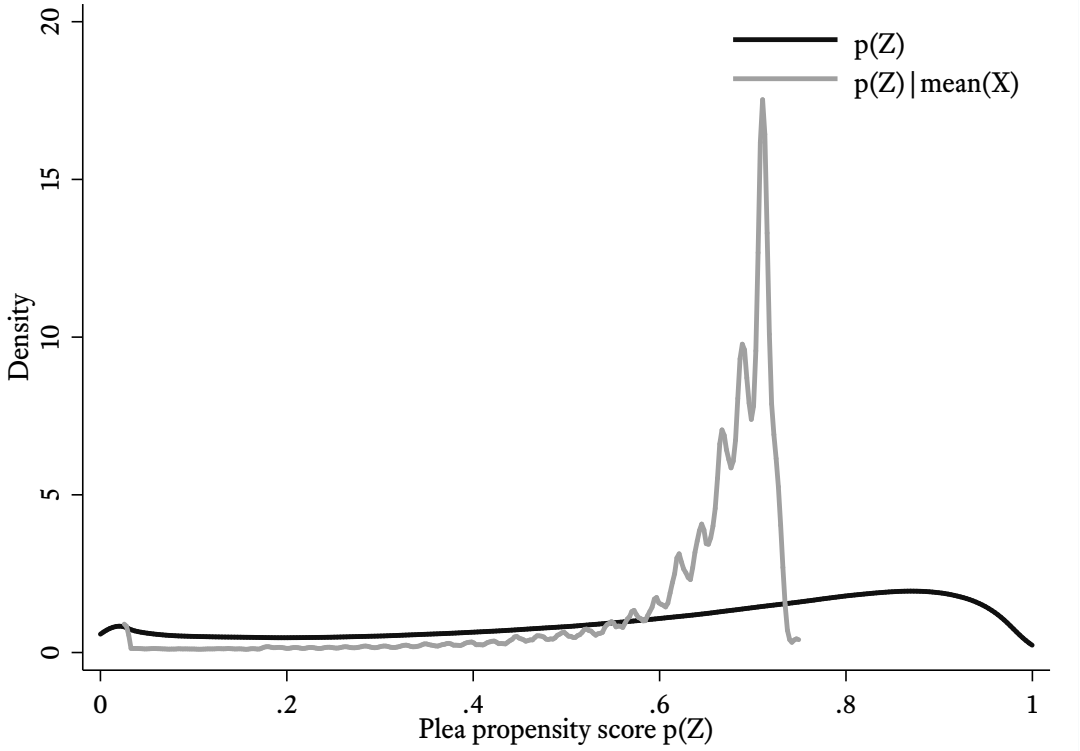


(B) US SAMPLE FROM MILLER ET AL. (1980)

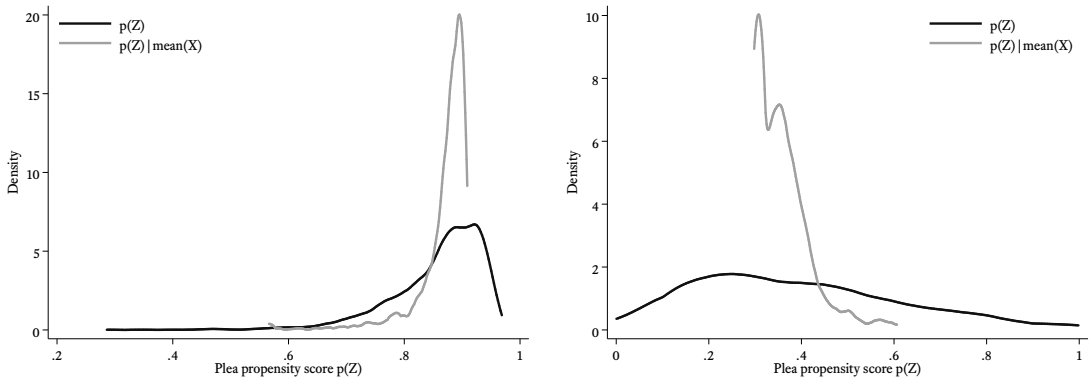


(C) ALASKA SAMPLE FROM CLARKE ET AL. (1982)

FIGURE 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



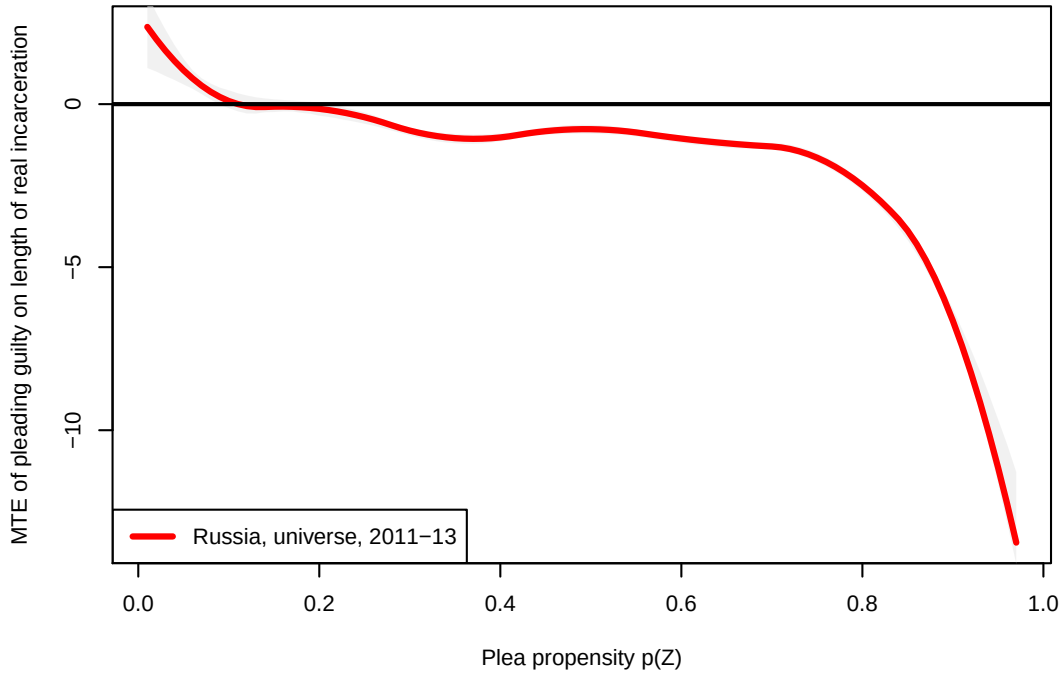
(B) US SAMPLE FROM MILLER ET AL. (1980) (C) ALASKA SAMPLE FROM CLARKE ET AL. (1982)

FIGURE 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.

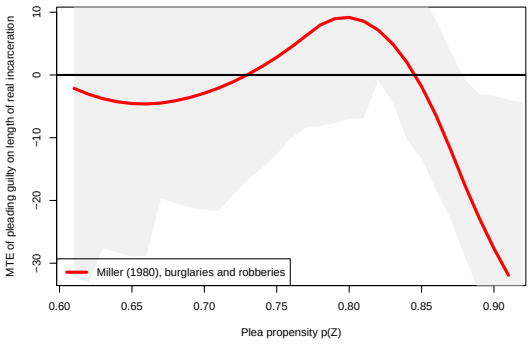
is modelled with P-splines with a restricted model where $K(p) = p$ is assumed to be linear. This test strongly rejects the hypothesis of linearity of $K(p)$ (F = 152.71, p-value < 0.001) in Russia. Non-linearity of $K(p)$ signals the presence of essential heterogeneity in the model and implies that 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.

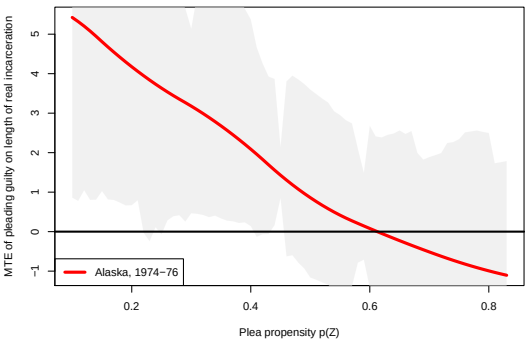
Treatment effects of Δ Another way to express negative sorting on the gains is to evaluate the $\widehat{MTE}$s at appropriate values of u_D to obtain conventional treatment parameters. This is performed in Table 4 which holds the main result of this paper.



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

TABLE 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 54. 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 de-

parture from Semi-parametric Method 2 of Heckman et al. (2006) in estimation does not alter the results qualitatively (Figure 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 paper is that plea discount cannot be summarised in one Δ due to inherent essential heterogeneity of defendants' decisions and outcomes. Finally, the paper 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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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

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

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

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).

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.

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.

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.

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

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.

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

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

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.

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

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.

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.

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

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

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

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.

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

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.

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.

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

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.

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.

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.

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 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 is preprocessed 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 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.

	(1)	(2)	(3)	(4)
Dependent variable	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 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 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).

	(1)	(2)	(3)	(4)
Dependent variable	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 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 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).

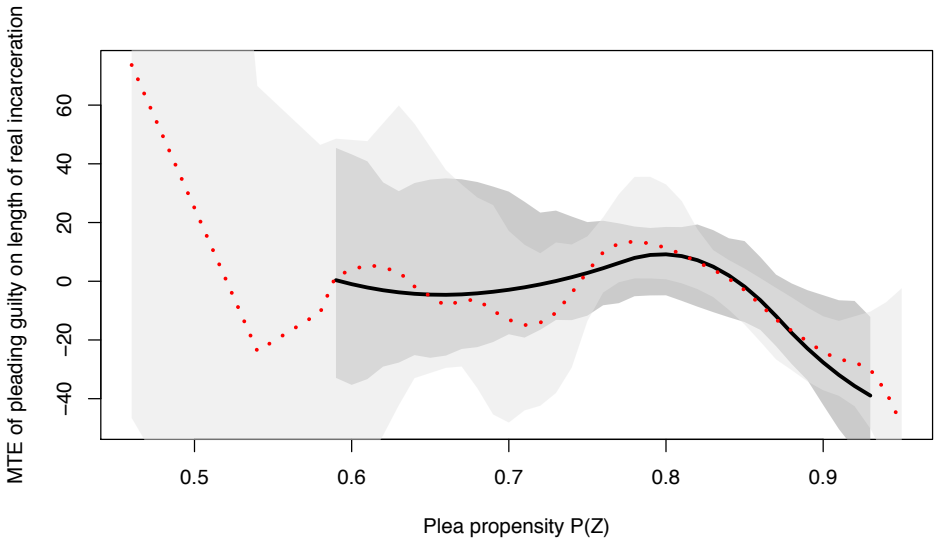


Figure A.1: Estimated MTE for Miller data under different approaches to estimation. $\widehat{MTE}^{Heckman}$ (dotted red line) is estimated with Heckman et al. (2006)'s Semiparametric Method 2 implemented by Brave and Walstrum (2014), where $\widehat{P}(X,Z)$ is estimated with logit, 3-degree local 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 (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 $\hat{\alpha}_0, \hat{\beta}_0, \widehat{(\beta_1 - \beta_0)}$ in P-spline estimated (8), Heckman et al. (2006) partial them out: $\tilde{Y} = K(p) + \tilde{\varepsilon}$, where $\tilde{Y} \equiv Y - \hat{\alpha}_0 - X'\hat{\beta}_0 - X'\widehat{(\beta_1 - \beta_0)}P(X,Z)$ 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 (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 has 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