The Cost of Bad Parents: Evidence from Incarceration on Children's Education

Carolina Arteaga*

July 4, 2018

Job Market Paper

[Most recent version]

Abstract

This paper provides evidence that parental incarceration increases children's educational attainment. I collect criminal records for 100,000 convicted poor parents in Colombia and combine it with administrative data on the educational attainment of their children. To overcome endogeneity concerns in incarceration I use exogenous variation that results from the random assignment of defendants to judges who differ in their stringency to convict and send defendants to prison. I write down an econometric model that features three treatment assignments, defined along two margins of selection and provide new identification results. I find that conditional on conviction, parental incarceration increases years of education by 0.6 years for the children whose parents were on the margin of incarceration. This positive effect is larger when the incarceration is for a violent crime, for boys and when the incarcerated parent is the mother.

JEL No. I24,J24,K42.

*Preliminary draft. I am very grateful for the support, guidance, and helpful comments of Adriana Lleras-Muney, Maurizio Mazzocco, Rodrigo Pinto, Sarah Reber, and Till von Wachter. I thank Stephen Billings, Leah Boustan, Moshe Buchinsky, Denis Chetverikov, Christian Dippel, Sam Norris, Rosa Matzkin, Matt Pecenco, Ricardo Perez-Truglia and Jeff Weaver for their feedback. I thank my colleagues Richard Domurat, Sepehr Ekbatani, Stefano Fiorin, Alex Fon, Keyoung Lee, Rustin Partow, and seminar participants at UCLA, CCPR, and the Central Bank of Colombia for insightful discussions. Natalia Cardenas, Mauricio Duran, and Juan D. Restrepo provided invaluable help in answering my questions about the institutional context. I gratefully acknowledge support from the Treiman Fellowship and CCPR. Comments are greatly appreciated.
Department of Economics, UCLA (caroartc@ucla.edu).

1 Introduction

Family environments during the early years, and especially parenting, are major determinants of human development (Heckman, 2013). It is estimated that over 1 million children in EU countries, and 2.7 million children in the U.S. have a parent in prison (Sykes and Pettit, 2014).[1] Despite the prevalence of this phenomenon, there is only a small literature on the causal effects of parental incarceration on children's outcomes. A large body of correlation-based evidence finds negative associations between parental incarceration and a host of important variables such as mental health, education, and crime (Wakefield, 2015). However, households where children experience parental incarceration are disadvantaged along many different dimensions[2] As a result, naive comparisons of outcomes would lead to negatively biased estimates.

In this paper I estimate the causal effects of parental incarceration on children's educational attainment. To address the endogeneity concerns, I exploit exogenous variation resulting from the random assignment of defendants to judges who differ in their leniency to convict and send defendants to prison. I do this in the context of the Colombian criminal justice system.

My data consist of criminal records of convicted parents in Colombia's census of potential beneficiaries of welfare: *SISBEN*. These data are collected by the government to characterize the poor population who may be eligible for different social programs.[3]SISBEN contains data on 39 million individuals (84% of Colombia's population), with national identification numbers (NIN), household members, and sociodemographic variables. Using the NINs I web-scrape criminal records for all 17 million parents who live in a judicial districts that has information available online.[4]This is the universe of cases in which the defendant was convicted. I find criminal records for approximately 100.000 parents in the years 2005 to 2016. I then track the educational outcomes of criminals' children using administrative data on

[1]Sykes and Pettit (2014) also estimate that for the U.S. 62% of black children born to high school dropouts will experience the imprisonment of a parent by age 17.

[2]Even prior to the incarceration event, these households are more likely to be poor and to experience domestic violence (Arditti, 2005; Arditti et al., 2012). In the US, Mumola (2000) estimates that 60% of parents in prison reported they used drugs in the month before their offense, and 25% reported a history of alcohol dependence, and about 14% of parents reported a mental illness. They also reported low levels of schooling; 70% did not have a high school diploma. Western (2018) also documents that around 60% of parents in prison had experienced childhood trauma, such as domestic violence or sexual abuse.

[3]Examples of these are such as free health insurance, conditional cash transfers, nutrition programs for their children, and subsidized housing, among others.

[4]17 out 33 judicial district have criminal records available online and this represents 67% of the population.

public school enrollment, and, also web-scrape the children's criminal records.

Identification of treatment effects is particularly challenging in my context. Unlike the previous literature using judge variation, I only observe those defendants who were convicted.[5] As a result, my sample of defendants is no longer balanced across judges, and the usual IV does not deliver interpretable treatment effects. This is because conviction is decided after a defendant is randomly assigned to a judge, and is a function of defendant's characteristics. To identify the causal effect of parental incarceration in this context, I need to compare children from parents who were convicted under the same leniency, and then use exogenous variation in incarceration. I do this by exploiting the differences in conviction and incarceration rates across judges. Specifically, I set the probability of conviction at a fixed threshold, and among those convicted defendants, I exploit the variation in judges' incarceration rates.

I estimate that conditional on conviction, parental incarceration increases years of education by 0.6 years for the children whose parents were on the margin of incarceration. This positive effect increases when unobserved defendant's quality decreases, which is identified from the leniency under which the parent was incarcerated. That is, defendants who are incarcerated under the most lenient judges have worse characteristics than those incarcerated under strict judges. I find that the positive effects of incarceration on schooling are driven by these defendants. In terms of observed heterogeneity, I estimate that the effect is larger when the child (i) is a boy, (ii) is not the first child, (iii) has parent who is convicted of a violent crime, and (vi) when the mother is incarcerated as opposed to the father. I also find a U shape pattern along the age of the child at the time of the event. Larger positive effects are estimated between ages 0 to 5, and 10 to 15, relative to 5 to 10. Even though the magnitudes of the differences in the coefficients are economically relevant, many of these differences are statistically indistinguishable from zero.

My findings suggest that on average, parents who are on the margin of incarceration are likely to reduce the amount of schooling their child attains if they instead remain in the household. This can be explained because the removal from the household of a violent parent or a negative role model can create a safer environment for a child (Johnson, 2008; Jaffee et al., 2003). Incarceration is also a mechanism that can limit the intergenerational transmission of violence, substance abuse, and crime to children.[6] This result also relates to findings in other fields that conclude that

[5]See Kling (2006), Aizer and Doyle, 2013; Di Tella and Schargrodsky (2013), Mueller-Smith, 2015; and Bhuller et al., 2016, among others.

[6]For example, using data from Sweden, Hjalmarsson and Lindquist (2007) report significant father-son correlations in criminal activity that begin to appear between ages 7 and 12, and are

the positive effects of being raised by one's parents depend on the quality of care they can provide (Jaffee et al., 2003). In addition, I find evidence that after the episode of parental incarceration, children move-in with their grandparents and also move to household outside SISBEN, which is a measure of upward mobility. These households might provide a better environment for the children and can help explain the positive effects.

I provide a new identification result for a setting in which treatment can take three values and is decided upon the crossing of two thresholds along distinct margins of selection. Specifically, given an instrument for each decision margin, treatment effects related to the crossing of the second threshold are identified upon; i) conditioning on the crossing of the first threshold given a fixed level of the first instrument, and ii) exploiting instrumental variation in the second margin. Unconditional treatment effects cannot be identified without further assumptions. This weaker result is, however, economically relevant. In my context it allows me to estimate the causal effect of incarceration conditional conviction under a given judge stringency. Examples of other contexts where this result applies are the decisions to participate in the labor force and work part-time or full time, attend college and enroll into stem or not stem majors, among many others.

Dobbie et al. (2018) and Bhuller et al. (2018) are the papers most closely related to mine. Using variation in judge leniency they study the intergenerational effects of incarceration in Sweden and Norway, respectively. Both of these papers find no effect on children's education outcomes. Many differences across our settings can explain the different results. First, the average prison sentence in Colombia is five years, compared with three and eight months in Sweden and Norway, respectively. A second key difference is the potential size of the treatment effects on schooling before college: In Colombia 47% of the population between 25 and 34 years old have less than a high school degree, whereas this number is 17% for both Norway and Sweden (OECD, 2016). Norway and Sweden have very generous welfare programs compared to those available in Colombia; these programs help insure disadvantaged populations and would also point towards smaller treatment effects. My results extend to other settings as a function of the similarity between contexts along these dimensions.

There are also other papers that find similar results to this study. For the US. Cho (2009) finds that children in Chicago's public schools whose mothers went to prison instead of jail for less than a week, are less likely to experience grade retention. Using an event study design, Billings (2018) finds that incarceration of a

fully established by the son's teen years.

parent improves end-of-grade exams and behavioral outcomes. He also finds larger benefits when the mother is incarcerated. Given the non-experimental variation of these two papers, their results should be taken as a lower bound of the effect of incarceration, but even in this case they are in line with my findings.

My paper also contributes to the literature on the effects of household structure on children's outcomes. Finaly and Neumark (2010) study whether marriage is good for children and find that unobserved factors drive the negative relationship between never-married motherhood and child education. For Hispanics, they find evidence that the children are better off living with a never-married mother. There is also a literature in sociology on the effects of parental marital conflict and divorce on children's well-being. Using 12 years of longitudinal data; Amato et al. (1995) find that in high-conflict families, children have higher levels of well-being as young adults if their parents divorced than if they stayed together. On the other hand, Doyle (2007, 2008) evaluates the welfare of children in foster care relative to staying with their parents and finds negative effects for children on the margin. Specifically, he finds that children placed in foster care have higher delinquency rates and teen birth rates, and lower earnings. My paper contributes to this body of literature with evidence that suggests that children may benefit from the absence of a convicted parent who is at the margin of incarceration.

Finally, in terms of policy implications my results call for greater support from the government in assisting children from fragile households. There is a strong body of experimental evidence on the powerful role of parenting and parenting supplements in shaping skills. Early childhood intervention have been remarkably successful in complementing parental care with positive economic, psychological, behavioral, and health benefits (Heckman et al, 2010). These programs can be a starting point to assist the parenting needs of this population.

The rest of the paper is structured as follows. Section 2 provides background on the judicial system in Colombia. Section 3 describes the data sources and provides summary statistics. Section 4 describes a model to identify causal effects in my setup, Section 5 presents my estimation and results. Section 6 discusses the results, the mechanism and external validity, and Section 7 concludes.

2 Background: The Colombian Court System

In this section, I describe the criminal justice system in Colombia. I go over how defendants are processed, how cases are assigned to judges, the types of crimes that are involved and the stages of the trial.

A criminal record is created when an arrest is made; once this happens, within 36 hours the police and a randomly assigned prosecutor must present the evidence that motivated the arrest in front a judge. This judge is randomly assigned from the lowest tier of a judge hierarchy, and determines if there are merits for the arrest and whether the defendant should await trial in prison or not.[7] Next, the case is randomly assigned to the judge who will preside over the trial —this is the judge who provides the exogenous variation in conviction and incarceration I use in this paper. In practice, once the first judge decides to continue with the prosecution of a defendant, the case is entered immediately into a software that assigns a judge at random between the judges in the judicial district and courtroom level the case was designated to; this constitutes what I will refer to as the randomization unit. Depending on the severity of the charges the prosecution pursues, a case will be randomized in one out of three possible court levels within the judicial district in which the crime was committed. Municipal courts receive simple cases, such as misdemeanors, property crimes involving small amounts, or simple assaults; these cases account for 38% of the data. More severe crimes, such as violent crimes, drug or gun-related crimes, and large property crimes are sent to circuit courts (56%). Lastly, crimes such as aggravated homicide or terrorism are assigned to a specialized judge (6%).[8] On average there are 20 judges per randomization unit, and in the largest district in Bogota the number of judges is 55. Figure 1 summarizes how defendants are processed in Colombia's criminal justice system.[9]

Once the judge is assigned, the prosecutor and defense present the case to the judge during multiple hearings. The purpose of the first hearing is to formally press charges. In the second hearing, the prosecutor and defense present all of the relevant evidence. In the third hearing, the judge decides on conviction; if the defendant is found guilty, the judge holds a final hearing to determine sentence length and incarceration. The penal code establishes minimum and maximum sentences for each crime but there is large scope for discretion on the judge side. The sentencing guidelines range is usually large, and the judge also determines the crime and the severity of the charge the defendant will be ultimately sentenced for.[10] The decision to send a defendant to prison is determined by the length of the sentence. To deal

[7] A defendant will go to prison before trial when at least one of the following conditions holds: i) the defendant is a danger to society, ii) the defendant can interfere with the judicial investigation, or iii) there is reason to believe the defendant will not appear in court for the trial. Art 308. Criminal Proceedings Code.

[8] Art 35-37, Criminal Proceedings Code.

[9] Acuerdo CSJ, 3329.

[10] For example prison time for possession of 100 grams of cocaine is 5 to 9 years. Art 376 Penal Code

with overcrowding in prisons, sentences are only served in prison when they are longer than a certain threshold. Currently, a sentence equal to 4 years or less is not served in prison; the only consequence of being convicted is that for the duration of the sentence, the judge must be notified of any changes of residence or when traveling outside the country.[11] In this setting, the population that faces a trial is divided into three groups: i) not convicted; ii) convicted and not incarcerated (those who are guilty but received a sentence shorter than 4 years), and iii) convicted and incarcerated (those guilty with a sentences longer than 4 years). The fact that a portion of the convicted population does not serve prison time is not a special feature of the Colombian system and is comparable to a sentence of probation in the US context for example.

While in prison, inmates can receive visits from adults once a week, and once a month from children. The government does not provide special social assistance to inmates' families. Unlike in the US, being convicted of a crime does not change the person's eligibility for welfare benefits, and in the labor market it is not common practice to ask about previous convictions even though, this information is available online.

3 Data and Summary Statistics

3.1 Data sources

I collect data from several sources. First, I use two waves of Colombia's census of potential beneficiaries of welfare (SISBEN). These data are collected by the government to characterize the poor population in the country and to target social programs. SISBEN has information on national identification numbers (NIN), household structure, age, gender, education, labor force participation and a large set of variables on the characteristics and the assets in the house (e.g fridge, stove, material of the floor, among others). With this information the government creates a score for each household which summarizes its level of wealth and uses this score to determine eligibility for most public programs. Example of these are free health insurance, conditional cash transfers, nutrition programs, subsidized housing and college loans, among many others. The first wave, conducted from 2003 to 2005, has data on 31.9 million individuals; the second wave conducted from 2008 to 2010, has data on 25.6 million people. There are 15.5 million individuals who appear in both waves. Attri-

[11] Art 63 Penal Code, and Ley 1709 de 2014.

tion from the first wave of the SISBEN is not random but related to an improvement in economic conditions in the household.

From this database, I obtain parent and children links when they live in the same household, and parents' NIN which I use to web-scrape criminal records. Anecdotal evidence for Colombia suggests that a large share of children with an incarcerated parent were not living with this parent at the time of incarceration. For the US, only half of the parents lived with their children at the time of incarceration (Parke and Clarke-Stewart, 2002).[12] My target population is, however, likely to be the most affected by incarceration.

In Colombia, criminal records from defendants who were convicted are public and available online for 17 out of 33 judicial districts. These 17 districts include the largest cities in the country, and are richer and more urban than districts without data online.[13] Criminal records have names and NIN of the defendant, court identifiers, crime and the date of crime and sentence information. Court identifiers are not linked to a specific judge. There is only one judge per court, but may change over time. For each judicial district I collect judges' directories with detailed judge tenure to construct leniency measures at the judge level. I complement these data with individual-level anonymized records from the Attorney General's Office. This database has information on the universe of criminal cases (including cases that did not result in conviction), with court identifiers, date and time of the trial, final verdict, and gender and age of the defendant, and allows me to have a measure of conviction stringency at the conviction stage.

Finally, I obtain education data from administrative records of public school enrollment for the years 2005-2016 with names and NIN's. Years of school is capped at 11 which is the last year of High School in Colombia.

3.2 Sample selection

To construct my sample I proceed as follows: from SISBEN I take the NIN for all parents who lived with their children in the 17 districts that have information online and web-scrape their criminal records. This adds up to 17 million adults. For computational reasons, I only search for records in the district the person was living in at the time of the survey. To asses the amount of records I miss due to

[12]Given how my parent to child links are constructed I focus on parents living with children rather than birth parents. This definition includes stepchildren when the parent identifies the child as his child instead of not-related to him or her.

[13]The universe of judicial sentences is public; however, they are only available in the national archives. The 17 districts represent 67% of the population, 69% of homicides, and 83% of property crimes. Criminal records are available here:
http://procesos.ramajudicial.gov.co/jepms/bogotajepms/conectar.asp

this restriction, I take a 5% random sample and look for their criminal records in all 17 districts. I estimate that I miss 8.6% of the sample due to crimes committed in districts different from the one in SISBEN. As a result, my data is not representative of the convicted population but it is selected to include only poor parents, who at the time of the SISBEN survey lived with their children, who lived in the largest districts of the country and who committed crimes in the district they were living in.

I find 328,579 criminal records for 256,108 individuals, of which 63,654 have missing fields in at least of one of the key variable such as court identifier, crime, year, and sentence. Half of these records with missing data correspond to Medellin, which is the second largest district and has missing court identifiers on all of their records. I keep only crimes committed after 2005, which leaves me with 193,520 records.[14] Next, I drop all records from court levels with only one judge (5,963 cases dropped), and in cases where the number of records in a year is less than 15 (44,806). I keep only courtrooms for which I have judge/year conviction rates from the Attorney General Office database. This leaves me with 128,792 criminal records from 105,133 adults. I keep only the first conviction in my sample and collect data on the crime, courtroom identifier, and decisions on sentence and incarceration. Once I merge back the criminal records to the SISBEN data, I keep only the first parental conviction in the household. My final database consists of criminals records for 91,032 parents. These parents are linked to 67,770 children, who were school-aged and experienced parental incarceration between ages 0 and 14.

I finish by collecting my two outcome variables for these children: educational attainment and criminal records. I find school records for 77% of them, which is close to the share of children between ages 12 and 17 who attend school in the census (76%). Table B.2 in the appendix shows evidence that having a missing education record is not related to parental incarceration, but to the child effectively not being at school or working, as reported in SISBEN. Missing values are also more prevalent for boys and for households with lower income and lower education of the head of the household, which are both predictors of being out of school. My final database consists of 52,419 children of convicted parents, born between 1990 and 2007. I also search for criminal records for all children of convicted parents who were 18 years of age by 2017. In the next section I characterize the population of convicted and incarcerated individuals, and their households and children.

[14] In 2005 there was a reform in the judicial system that renders the two periods incomparable. In the previous system the judge served as prosecutor and judge at the same time, and his was anonymous to the defendant. Also there were changes regarding sentencing guidelines.

3.3 Summary statistics

The population in my sample is negatively selected along two margins: income and crime. In Table 1 I show basic socioeconomic characteristics for adults in the overall population, for parents in the SISBEN with and without a conviction, and for parents with a conviction, by incarceration. By looking at column 1 vs. 2 and 3 we see that parents in the SISBEN have fewer years of education, are less likely to have a high school degree, live in larger households and are more likely to be single than adults in overall population. Within parents in the SISBEN, individuals with a conviction (column 3 relative to 2) are also negatively selected across a host of variables. Convicted adults have fewer years of schooling, are less likely to have a high school degree or more (23% vs. 31%), and have lower income scores. They also live in larger households and are more likely to be single (41% vs. 35%). Adults with criminal records are disproportionally males (84%), and given this disparity in gender, they are more likely to work and to be the head of the household than is the case for the population without criminal records.[15]

Next, within the sample of those with a conviction, I look at differences by incarceration (columns 4 and 5). Incarcerated parents have lower education and lower income, differences in the probability of incarceration are far smaller across gender. Incarceration is associated with lower probabilities of working and of being the head of the household. Table 2 splits the sample by gender. Convicted women have lower education relative to men, and come from poorer households. Compared to men, they are less likely to be the head of the household, but are still much more likely than in the overall population women (36% vs. 29%). Convicted women are also more likely to be single.

In terms of the distribution of crime for the convicted population in my data, I find that property crimes are the most common type of offense (25%) followed closely by drug-trafficking crimes (24%). Violent crimes account for 20% of the records, followed by gun-trafficking and misdemeanors with 18 and 12%, respectively. Incarceration rates vary substantially by crime. Figure 2 ranks crimes by incarceration rates, for a selected number of offenses. Serious crimes, such as kidnapping or rape, have the highest incarceration rates, whereas failure to pay child support, simple assault, and property damage have the lowest. In the middle of the distribution we find crimes such as drug trafficking, domestic violence, counterfeit currency trafficking, theft, and smuggling, among others.

[15] In the US context for example 29% of parents in state prisons have a HS degree or more, 48% are single, 92% are male and the median age is 32 (Mumola, 2000).

4 Identification

The literature on the effects of incarceration addresses the endogeneity of incarceration exploiting the random assignment of defendants to judges who differ in their leniency to incarcerate. In those papers, authors have data on the pool of cases that was randomly assigned across judges, and use this to construct their instrument. They compared incarcerated defendants with non incarcerated defendants which includes those who were not convicted, and also those who were but did not received a prison sentence.

In my case, I only observe cases of defendants who were convicted. With these data I can compute judge leniency on incarceration conditional on conviction. This leniency however, will not deliver a treatment effect we can interpret. To address this challenge I develop a framework to identify treatment effects in my data. I take a step back from the decision to incarcerate, and model treatment as the crossing of two thresholds; one that decides on conviction, and a second one that decides on incarceration. I exploit the fact that judges differ in their leniency regarding both conviction and incarceration decisions, and use this to derive causal effects of incarceration relative to conviction. In the next section I provide intuition for this identification result and explain why the simple IV fails, next I formalize the result.

4.1 A simplified framework

To fix ideas, let us first consider the following framework: Judges are randomly assigned to defendants to make conviction and incarceration decisions by evaluating two distinct attributes of the defendant. When deciding on conviction c, a judge assesses the strength of the evidence of the case at hand. Without loss of generality, the distribution of the strength of the evidence across defendants U^c is uniform [0,1], where zero is the smoking gun, and one is no evidence against a defendant. The judge can be one of two types: Harsh (h^c) or Lenient (l^c). Harsh judges do not require much evidence to convict a defendant. They have threshold of 0.8, and thus convict 80% of defendants; which corresponds to all defendants with a level of evidence below 0.8. Lenient judges require more evidence to convict a defendant, and choose a threshold such that they convict only 20% of them.

Next, if a defendant is convicted, the judge must make a decision about incarceration I. She/he makes this decision based on an assessment of how harmful the convicted defendant is to society, and how much punishment the defendant deserves. This trait which I call U^I is also distributed uniformly [0,1]. Very harmful defendants have low values of U^I, and respectively, non-harmful defendants have values close to 1. A harsh judge (h^I) would send 80% of the convicted defendants to prison, whereas a lenient one (l^I) would only incarcerate 20%. It is the same judge making both decisions. Figure 3 illustrates this situation, the x-axis traces the strength of the evidence on which the conviction decision is based. That is, we can order defendants along one relevant dimension, the strength of the evidence in the [0,1] interval. A judge splits the space in two when he sets his conviction rate: Defendants to the right are free, and defendants to the left are convicted. Similarly, the y-axis traces the defendant's punishment level, which is related to the assessment of predicted future criminal activity, unobserved (to the econometrician, not the judge) crime severity, and any mitigating/aggravating factors or family ties.[16] I will refer to this dimension as a measure of defendants' overall quality. For a fixed level of conviction, a judge's incarceration level splits the space of convicted individual in two: A defendant below the threshold will go to prison and a defendant above will not.

Due to randomization, all judges start with a statistically identical pool of defendants. However, after the conviction decision is made, the pool of convicted defendants is no longer comparable. Defendants convicted under the lenient judge will have, on average, a stronger case against them than those convicted under the harsh judge. Defendants convicted under a harsh judge can face two types of judges (h^c, l^I) and (h^c, h^I), where the first term refers to the type of the judge in the conviction stage, and the second one refers to the type in the incarceration stage. Similarly, those convicted under lenient judges can also have judges of types (l^c, h^I) and (l^c, l^I). Within these partitions, defendants are balanced across judges: First, because they were randomly assigned to their judge, and second, because they were selected into conviction under the same threshold. As a result, within partitions, there is exogenous variation in the probability of going to prison. For example, those convicted who were assigned to (h^c, l^I) judges face a 20% chance of incarceration, whereas those assigned to a (h^c, h^I) judge face a 80% probability. Figure 4 illustrates this argument. This means that for all those 60% of defendants whose harmfulness assessment is located above the worst 20% of the population, but still in the 80% bottom, incarceration is only a function of judge assignment. Thus, I will be able to estimate LATE-type parameters for defendants who fall in this range.

Specifically, for this example I will estimate the following two LATE parameters:

$$LATE_H = E[Y_I - Y_c | U^c < 0.8]$$

[16] In practice there are sentencing laws that guide the judge's incarceration decisions; however, there is a large scope for discretion, even within a specific crime. What this dimension tries to capture are the factors that cause a judge to make different incarceration decisions for criminals who have the same charges.

and,

$$LATE_L = E[Y_I - Y_c | U^c < 0.2]$$

Where $LATE_h$ is the causal effect of incarceration relative to conviction for those convicted under a harsh judge ($U^c < 0.8$), and $LATE_l$ is the one for conviction under a lenient judge. Y_I and Y_c represent counterfactual outcomes (years of education of the child) for incarceration (I) and conviction (c) and U^c traces the selection on the conviction stage.

$$LATE_h = \frac{E[Y|(h^c, h^I), U^c < 0.8] - E[Y|(h^c, l^I), U^c < 0.8]}{E[T = I|(h^c, h^L), U^c < 0.8] - E[T = I|(h^c, l^I), U^c < 0.8]}$$

Where $T = I$ in the denominator represents treatment assignment equal to incarceration. Similarly, we can have the analogous expression for $LATE_l$.

Lastly, to use the judge instrument as in the previous papers in the literature, I need to compute the share incarcerated for every judge. Recall that those paper define only two treatments: incarceration and everything else —which includes probation and not convicted. For a judge type (h^c, l^I), the probability of incarceration corresponds to: $0.8 \cdot 0.2 = 0.16$ which is the same as the one for (l^c, h^I). For judges type: (h^c, h^I) is $0.8 \cdot 0.8 = 0.64$, and for (l^c, l^I) equals $0.2 \cdot 0.2 = 0.04$. At first glance it looks like we have exogenous variation in incarceration that we can as an instrument. However, what this exercises ignores is that the pool of defendants is not being held constant across judges, and as a result differences will not only reflect the effect of incarceration but also the differences in samples. Figure 4 plots a situation in which I use the variation in incarceration rates from judges (h^c, h^I) and (l^c, l^I). From the graph it is clear that there are not well defined groups for a valid comparison. That is because we are not observing the same group of people across judges. Specifically, defendants with $U^c > 0.2$ are only observe for judge (h^c, h^I).

The next section formalizes the former identification result and extends the result to the case of continuous judge leniency.

4.2 Model

The model is described by the standard IV model that consists of five main random variables: $T, Z, Y, \mathbf{V}, \mathbf{X}$. Those variables lie in the probability space (Ω, F, P), where individuals are represented by elements $i \in \Omega$ of the sample space Ω. The variables are defined below:

- T_i denotes the assigned treatment of individual i, and takes values in

$supp(T) = \{t_f, t_c, t_I\}$. t_f stands for not convicted, t_c for convicted but not incarcerated, and t_I for convicted and incarcerated.

- Z_i is the instrumental variable in this analysis and takes values in $supp(Z)$, and represents judge assignment.
- Y_i denotes the outcome of interest for individual i, e.g. years of education of the child.
- $\mathbf{X}_i$ represents the exogenous characteristics of individual i.
- $\mathbf{V}_i$ stands for the random vector of unobserved characteristics of individual i, and takes values in $supp(\mathbf{V})$.

The random vector $\mathbf{V}$ is the source of selection bias in this model. It causes both the treatment T and outcome Y. The standard IV model is defined by two functions and an independence condition as follows:

$$\text{Outcome Equation: } Y = f_Y(T, X, V, \epsilon_Y) \tag{1}$$

$$\text{Treatment Equation: } T = f_T(Z, X, V) \tag{2}$$

$$\text{Independence: } Z \perp V, \epsilon_Y | X \tag{3}$$

where ϵ_Y is an unobserved zero-mean error term associated with the outcome equation.

In this notation, a counterfactual outcome is defined by fixing T to a value $t \in supp(T)$ in the outcome equation. That is, $Y(t) = f_Y(t, \mathbf{V}, \mathbf{X}, \epsilon_Y)$. The observed outcome for individual i is given by:

$$Y = Y(T) = \sum_{t \in \{t_f, t_c, t_I\}} Y(t) \cdot \mathbf{1}[T = t]. \tag{4}$$

The independence condition (3) implies the following exclusion restriction:

$$\text{Exclusion Restriction : } Z \perp Y(t) | \mathbf{X} \text{ for all } t \in supp(T). \tag{5}$$

For the sake of notational simplicity, I suppress exogenous variables $\mathbf{X}$ henceforth. All of the analysis can be understood as conditional on pre-treatment variables.

I assume that the treatment equation is governed by a combination of two threshold crossing inequalities. First, there is a conviction stage:

$$\begin{cases} \text{Free} & \text{if } \mathbf{1}[\phi_c(\mathbf{V}) > \xi_c(Z)] \\ \text{Convicted} & \text{if } \mathbf{1}[\phi_c(\mathbf{V}) \leq \xi_c(Z)] \cdot \end{cases}$$

where $\mathbf{1}[\cdot]$ denotes a binary indicator and $\phi_c(\cdot), \xi_c(\cdot)$ are real-valued functions. Function $\phi_c(\cdot)$ measures the degree of culpability assessed by the judicial system. This function looks at variables and information that are not observed by the econometrician but that are observed by the judge, such as the evidence, crime intensity, the effort of the defense and prosecutor lawyers, as well as unobserved characteristics of the defendant such as aggression, anti-social behavior, etc. The function $\xi_c(\cdot)$ assesses the judge leniency on conviction. This function can be understood as a threshold of reasonable doubt beyond which the defendant is convicted by the judge. Judges differ in their leniency and may set different threshold of evidence. The judge convicts defendant i whenever: $\phi_c(\mathbf{V}) \leq \xi_c(Z)$. If that is the case, a second stage is held and the judge makes a decision regarding incarceration:

$$\begin{cases} \text{Not incarcerated} & \text{if } \mathbf{1}[\phi_I(\mathbf{V}) > \xi_I(Z)] \\ \text{Incarcerated} & \text{if } \mathbf{1}[\phi_I(\mathbf{V}) \leq \xi_I(Z)] \end{cases}$$

Similarly, $\phi_I(\mathbf{V})$ is a function whose arguments are case and defendant's characteristics that are relevant for the assessment of the punishment level. Same as before, the judge compares $\phi_I(\mathbf{V})$ to her/his threshold to incarcerate $\xi_I(Z)$.

Treatment assignment can be summarized as follows:[17]

$$T = f_T(Z, \mathbf{V}) = \begin{cases} t_f & \text{if } \mathbf{1}[\phi_c(\mathbf{V}) > \xi_c(Z)] \\ t_c & \text{if } \mathbf{1}[\phi_c(\mathbf{V}) \leq \xi_c(Z)] \cdot \mathbf{1}[\phi_I(\mathbf{V}) > \xi_I(Z)] \\ t_I & \text{if } \mathbf{1}[\phi_c(\mathbf{V}) \leq \xi_c(Z)] \cdot \mathbf{1}[\phi_I(\mathbf{V}) \leq \xi_I(Z)] \end{cases}$$

This model relies on two separable threshold functions that play the role of the monotonicity condition. Consider two judges j and j', that see defendants i and i' who differ in their level of culpability. Say i' has more evidence against him than i, namely $\phi_c(i') < \phi_c(i)$. Supposed that judge j convicts defendant i' but not i. Then the threshold function implies that it can not be the case that judge j' convicts

[17] See example 4 in Lee and Salanie (2017).

defendant i, but not i'. More generally, let $D_i(j) = \mathbf{1}[T_i(j) = t_c]$ denote the binary indicator that judge j convicts defendant i. Thus if judge j convicts i' but not i, it implies:

$$D_i(j) > D_{i'}(j)$$

Then, it can not be the case that judge j' convicts defendant i, but not i'. Which means:

$$D_i(j) > D_{i'}(j) \to D_i(j') > D_i'(j')$$

which is equivalent to state that:

$$D_i(j) > D_i(j') \to D_{i'}(j) > D_{i'}(j')$$

We can generalize this to all individuals to arrive at the standard monotonicity assumption of Imbens and Angrist (1994).[18]

I assume the following standard regularity conditions: i) $E(|Y(t)|) < \infty$ for all $t \in supp(T)$, ii) $P(T = t|Z = z) > 0$ for all $t \in supp(T)$ and all $z \in supp(Z)$ and, iii) $(\phi_c(\mathbf{V}), \phi_I(\mathbf{V}))$ are absolutely continuous with respect to Lebesgue measure in $\mathbb{R}^2$. The first assumption guarantees the existence of the expectation, the second one assures that there is a share of the population assigned to each treatment group for every judge, and the third one allows me to apply the Lebesgue differentiation theorem.

Without loss of generality, it is useful to express treatment assignment using the following variable transformation:

$$\begin{aligned}
U^c &= F_{\phi^c(\mathbf{V})}(\phi^c(\mathbf{V})) \sim Unif[0,1], && (6)\\
U^I &= F_{\phi^I(\mathbf{V})}(\phi^I(\mathbf{V})) \sim Unif[0,1], && (7)\\
P_c &= F_{\phi^c(\mathbf{V})}(\xi^c(Z)); z \in supp(Z), && (8)\\
P_I &= F_{\phi^I(\mathbf{V})}(\xi^I(Z)); z \in supp(Z), && (9)
\end{aligned}$$

where $F_K(\cdot)$ denotes the cumulative distribution function of a random variable K. U^c, U^I, P_c, P_I are uniformly distributed random variables in $[0,1]$ due to assumption (iii). Let $P_c(z)$ denote the conditional random variable $P_c(Z = z)$ which is simply.

[18]This structure implicitly assumes that judges care about the same characteristics of the defendants and value them in the same way. I will address the validity of this assumption in the robustness section.

Moreover, independence condition (3) implies $P_c, P_I \perp (U^c, U^I)$. In this notation, the model can be expressed as:

$$T \equiv f_t(Z, V) = g_T(U^c, U^I, P_c, P_I) = \begin{cases} t_f & \text{if } \mathbf{1}[U^c > P_c(z)] \\ t_c & \text{if } \mathbf{1}[U^c \leq P_c(z)] \cdot \mathbf{1}[U^I > P_I(z)] \\ t_I & \text{if } \mathbf{1}[U^c \leq P_c(z)] \cdot \mathbf{1}[U^I \leq P_I(z)] \end{cases} \tag{10}$$

In the model, U^c and U^I have the same interpretation as in the previous section, and P_c is interpreted as the share convicted for judge z. Moreover, under the assumption that $U_c \perp U_I$, we can identify $P_I(z)$ from the data, that is:

$$P(U^I < P_I(z) | U^c \leq P_c(z)) = P(U^I < P_I(z)) = P_I(Z)$$

The left hand side is observed from the data, the first equality follows directly from the independence assumption and the last one the uniform distribution of U^I. P_I is interpreted as the share incarcerated. For ease of exposition, I will first explore identification under this assumption (see also Lee & Salanie, 2017) and then I will go over the results without it.

The goal is to identify and evaluate the treatment effect: $E(Y(t_I) - Y(t_c))$ which is a function of counterfactual variables $Y(t_I)$ and $Y(t_c)$. To achieve this goal, it is useful to express the observed expectations in terms of the variables that define the model:

$$E(Y \cdot \mathbf{1}[T = t_c] | P_c(Z) = p_c, P_I(Z) = p_I) = \tag{11}$$
$$= E(Y(t_c) \cdot \mathbf{1}[T = t_c] | P_c(Z) = p_c, P_I(Z) = p_I) \tag{12}$$
$$= E(Y(t_c) \cdot \mathbf{1}[U^c \leq p_c] \cdot \mathbf{1}[U^I > p_I] | P_c(Z) = p_c, P_I(Z) = p_I) \tag{13}$$
$$= E(Y(t_c) \cdot \mathbf{1}[U^c \leq p_c] \cdot \mathbf{1}[U^I > p_I]) \tag{14}$$
$$= \int_0^{p_c} \int_{p_I}^1 E(Y(t_c) | U^c = u^c, U^I = u^I) f_{u^c u^I}(u^c, u^I)\, du^c\, du^I \tag{15}$$
$$\tag{16}$$

$$= -\int_0^{p_c} \int_0^{p_I} E(Y(t_c) | U^c = u^c, U^I = u^I) f_{u^c, u^I}(u^c, u^I)\, du^c\, du^I + \int_0^{p_c} E(Y(t_c) | U^c = u^c) f_{u^c}(u^c)\, du^c$$

Equation (11) is an expectation observed in the data. Equality (12) comes from the definition of observed outcomes in Equation (4). Equality (13) expresses the indicator $\mathbf{1}[T = t_c]$ in terms of the inequalities of the choice model. Equality (14) uses the independence relation $Z \perp (U^c, U^I)$. Equality (15) expresses the expectation as the integral over the distribution of U^c, U^I where $f_{U^c,U^I}(u^c, u^I)$ stands for the probability density function of U^c, U^I at the point (u^c, u^I), and is equal to one. Equality (16) modifies the integration region. This change is useful to apply the Lebesgue differentiation theorem next;

$$\frac{\partial^2 E(Y \cdot \mathbf{1}[T = t_c] | P_c(Z) = p_c, P_I(Z) = p_I)}{\partial p_c \partial p_I} = -E(Y(t_c) | U^c = p_c, U^I = p_I) \tag{17}$$

Equality (17) arises as a direct application of the Lebesgue differentiation theorem. What this result gives me is a connection between the observed outcomes (Eq. 11) and the targeted counterfactual outcome (RHS Eq. 17). We can use the same steps applied to counterfactual $Y(t_c)$ to obtain counterfactual for $Y(t_I)$. Combining these two I obtain:

$$\frac{\partial^2 E(Y \cdot \mathbf{1}[T \in \{t_c, t_I\}] | P_c(Z) = p_c, P_I(Z) = p_I)}{\partial p_c \partial p_I} = E(Y(t_I) - Y(t_c) | U^c = p_c, U^I = p_I) \tag{18}$$

In the language of Heckman and Vytlacil (2005), Eq.18 defines the marginal treatment effect (MTE) of outcome Y with respect to treatment assignment t_c and t_I. It is interpreted as the causal effect of incarceration versus conviction only, for the share of defendants whose culpability and punishment assessments, U_c and U_I respectively, is set at quantiles p_c and p_I. The derivative in Equation (18) traces the MTE of incarceration relative to conviction throughout the unitary square of U^c, U^I. This result is an application of Lee and Salanie (2017) and extends the result of Heckman and Vytlacil (1999) to two dimensions. In Appendix A I explain graphically the intuition of this result. The main idea is that changes in P_c and P_I affect exogenously treatment assignment. Then, by looking at the derivative of the outcome variables with respect to P_c and P_I, we capture how the outcome variable changes when treatment changes at each point in the space of the unobservable confounding variables.

The average treatment effect (ATE) is the causal effect of t_c and t_I on Y in the population, and it corresponds to the integral of the MTE over the support of U_c and U_I.

$$E(Y(t_I) - Y(t_c)) = \int_0^1 \int_0^1 \frac{\partial^2 E(Y \cdot \mathbf{1}[T \in \{t_c, t_I\}]|P_c(Z) = p_c, P_I(Z) = p_I)}{\partial p_c \partial p_I} dp_c dp_I \tag{19}$$

Without the assumption of independence of U_c and U_I, variation in P_I is only identified once I fix the conviction threshold. Thus, I will be interested now in a different counterfactual object: $Y(t_I)$ and $Y(t_c)$ for those who were convicted under $P_c = p_c$. This means the objective if to identify causal effects of the form: $E(Y(t_I) - Y(t_c)|U^c < p_c)$, which is the the same exercise explained in Section 4.1. Let:

$$E(Y \cdot \mathbf{1}[T = t_c]|P_c(Z) = p_c, P_I(Z) = p_I, U^c < p_c) = \tag{20}$$
$$= E(Y(t_c) \cdot \mathbf{1}[T = t_c]|P_c(Z) = p_c, P_I(Z) = p_I, U^c < p_c) \tag{21}$$
$$= E(Y(t_c) \cdot \mathbf{1}[U^I > p_I]|P_c(Z) = p_c, P_I(Z) = p_I, U^c < p_c) \tag{22}$$
$$= E(Y(t_c) \cdot \mathbf{1}[U^I > p_I]|U^c < p_c) \tag{23}$$

Where I followed the same steps as before. Let:

$$P_I^* = Pr[U_I < P_I|U_c < P_c] = G(P_I) \tag{24}$$

P_I^* is the object I observe so I will define the observed expectations in terms of this variable:[19]

$$E(Y(t_c) \cdot \mathbf{1}[U^I > G^{-1}(p_I^*|U_c < p_c]|U^c < p_c) \tag{25}$$
$$\int_{P_I^*}^1 E(Y(t_c)|U^I = u^I, U_c < p_c) f_{u^{I*}|U^c<p_c}(p_I^*) du^I \tag{26}$$

And applying the Lebesgue differentiation theorem this results in:

$$\frac{\partial E(Y \cdot \mathbf{1}[T \in \{t_c\}]|p_c, p_I, U^c < p_c)}{\partial p_{I^*}} = -E(Y(t_c)|U^I = p_I, U^c < p_c) f_{u^I|U^c<p_c}(p_I^*) \tag{27}$$

And ultimately;

$$E(Y(t_I) - Y(t_c)|u^c < p_c) = \int_0^1 \frac{\partial E(Y \cdot \mathbf{1}[T \in \{t_c, t_I\}]|P_c(Z) = p_c, P_I^*(Z) = p_I^*, U^c < p_c)}{\partial p_I^*} dp_I^* \tag{28}$$

[19]Where $f_{u^{I*}|U^c<p_c}(p_I^*)$ in eq. (26) corresponds to:$f_{u^I|U^c<p_c}(p_I) \frac{\partial P_I((p_I^*)}{(p_I^*}$

What this result says is that we can trace the treatment effect of incarceration relative to conviction once we fix a threshold for conviction. We do this by evaluating the changes on the outcome variable when we change P_I^*. This delivers the MTE along the unobservable dimension $U^I|U^c < P_c$. The integral over the support of the instrument gives the LATE, or the ATE when the instrument has full support. In the next section I use this identification approach to estimate the effects of parental incarceration in my data.

5 Estimation

To apply the identification result of the previous section I start by estimating the sample analogs of $P_c(Z)$ and $P_I^*(Z)$ in the model. The interpretation of these variables is the probability of being convicted/incarcerated given the assignment to a specific judge. Following the literature, these are estimated as judge fixed effects from regressions after parsing out variation at the unit at which the randomization of judges occurred and crime-level conviction/incarceration rates. That is, the conviction/incarceration decision can be decomposed into a portion that is related to the individual, the judge, the crime, and the randomization unit/year. I do this as follows:

$$D_{iztcr} = \gamma_{rt} + \gamma_c + p_z + \epsilon_{iztcr}$$

Where D_{iztcr} corresponds to a conviction or incarceration dummy, i indexes individuals, z judges, t year, c crime, and r court/judicial district. γ_{rt} corresponds to randomization-level fixed effects, which is a courtroom tier-municipality and year-level fixed effect. γ_c is a crime level fixed-effect (161 different crimes); p_z is the judge fixed effect and is what we are after; and ϵ_{iztcr} is a zero mean term. Following the literature, I estimate the judge fixed effect $\widehat{p_{z-i}}$ for defendant i to be the following leave-out estimator:

$$\widehat{p_{z-i}} = \frac{1}{n_z - 1} \sum_{k \neq i} \widehat{res_{z,k}}$$

and

$$\widehat{res_{zk}} = D_{kztcr} - \widehat{\gamma_{rt}} + \widehat{\gamma_c}$$

where n_z is the number of cases of judge z, and res_{zk} is the residual from a regression of the conviction/incarceration dummy on γ_{rt} and γ_c.

Figure 6 shows the distribution of D_{iztcr} at the judge level, and $\widehat{p_z}$ for both conviction and incarceration. From the graph, we can see that although courtroom tier-year and crime-level fixed effects explain most of the variation, judge's fixed effects still represent a sizable share of the variance in conviction and incarceration.

5.1 Instrument Validity

Next, I examine how much do judge fixed effects predict individual-level decisions by estimating a first-stage regression for defendants as follows:

$$D_{iztcr} = \beta_0 + \widehat{p_{z-i}} + \beta_1 X_i + \epsilon_i$$

As before, D_{iztcr} corresponds to the conviction or incarceration dummy, and p_{z-i} is the leave-out mean of judge z assigned to person i. I run this regression with and without controls X_i. In the conviction regression, where I use anonymized data from the Attorney Generals Office, I can only control for age, gender and number of crimes. In the incarceration regression I control for schooling, income, occupation, gender, year of birth, and year in the survey. According to the results in Table 3, judge's fixed effects have a strong influence on conviction and incarceration decisions. The estimates are highly significant and suggest that being assigned to a judge with a 10 percentage point higher conviction/incarceration rate increases defendants probability of conviction/incarceration by roughly 8 and 12 percentage points, respectively. This relationship is robust to the inclusion of controls. Figure 7 depicts this first-stage relationship for conviction (left panel) and incarceration (right panel). These graphs show strong positive relationship between the instrument and the individual trial decisions. The F-stats on the first stage—corrected for the dimensionality of the instrument matrix, which is the number of judges—are above the critical value for the leave-out mean instrument for weak instruments (see Figure 4 in Stock et al, 2002). I discuss this later in the robustness section.

Figure 8 shows a scatter plot of both conviction and incarceration fixed effects. Key for the identification of treatment effects is the variation in incarceration stringency conditional on a level of conviction stringency. From the graph we can see there is substantial variation along the incarceration axis for each conviction rate.

For the instrument to be valid, the judge's fixed effects must be orthogonal to the defendant's characteristics. I test this in the anonymized data from the Attorney General's office, in which the universe of cases the judge hears is available. Table 4 checks the balance across defendants for my judge-stringency measures for conviction and incarceration. Across gender, age, and type of crime, which are the only variables available in these data, I find no individual or joint statistical significance. In addition, the identification result is supported on the observation that once P_c is fixed, the pool of convicted defendants is balanced across judges. I test whether covariates are associated with incarceration stringency for the convicted sample, once I split the sample by conviction group (low, medium or high), or control for the conviction level with a polynomial of P_c. In Table 5, I test individual and joint significance of variables associated with education, income and occupation status, and I find no evidence of a relationship with judge stringency.

To interpret the results of the IV as the causal effect of incarceration, judge stringency must only affect child's outcomes through incarceration. In this literature, it has been discussed that this may not be the case if the judge fixed effects on incarceration captures other dimensions of trial decisions such as fines, or guilt. In my setup, I believe it is less of a concern, first because in the case of Colombia fines are and rare and only associated with large property crimes, and second because I model the conviction decision directly.

Finally, an additional assumption is that conviction or incarceration decisions made by a lenient judge would also have been made by a stricter judge; this is called the monotonicity assumption. One testable implication is that first-stage estimates should be non negative for all sub samples. That is, if a judge is lenient, he or she is going to be lenient for both women and for men, and for both violent crimes and non-violent crimes. To test this assumption I construct judge fixed effects for one group in the population, for example men, and use this fixed effect in a first stage regression to predict individual conviction and incarceration for women. I do this for gender, type of crime and age groups. Table C1 in the appendix shows this first stage tests, I find positive first stage estimates across all slices of the data, which supports the monotonicity assumption.

5.2 Results

I now turn to the estimation of the effect of parental incarceration on children's educational attainment. I restrict attention to parental incarceration that occurs between ages 0 and 14, and only study cohorts born between 1990 and 2007, so that they are both young and old enough to appear in the educational attainment data, which I observe from 2005 to 2016. I consider only incarceration cases in households in which the person incarcerated was the parent of the child and not another household member.[20]

[20]The number of cases where this is the case is not large enough to study this population.

Following the identification result, I need to account for the different levels of conviction stringency at which defendants were found guilty. I do this in two ways: First, I sort my data by stringency in the conviction stage (P_c), and split the sample into three groups: low ($0.74 < P_c < 0.88$), medium ($0.88 < P_c < 0.9$) and high ($0.9 < P_c < 1$) conviction levels. And second, I pool the data and add a second degree polynomial on P_c. The first three columns of tables 6 and 7 have the regressions for the splitted sample, and the forth one has the pooled regression.

I begin by showing the OLS estimate of this design. Table 6 shows a regression of parental incarceration on years of education. Following Abadie et al. (2017), I cluster standard errors at the randomization unit level which is the level at which judge assignment is decided. Without controls, a child whose parent went to prison has 0.4 to 0.3 fewer years of schooling than a child whose parents did not. Once I add controls, this difference reduces drastically to less than 0.1 years. Still, we expect that incarcerated parents are negatively selected on unobservables that cannot be accounted for, so -0.1 years is a lower bound on the causal effect.

Next, Figure 9 shows a graphical representation of the reduced-form regression. This graph plots the distribution of judges' incarceration fixed effects against the predicted years of education from a local polynomial regression. From the graph, we can see that there is a strong positive relationship between judge stringency in incarceration and years of education. That is, as we move to the right and the probability of having a parent in prison exogenously increases, I estimate that years of education also increase. The top panel of Table 7 shows regression results for this reduced form: I estimate large increases in years of education for all specifications, and for all but the second column, the increase in years of education is statistically significant. Finally, the bottom panel of Table 7 shows results from the IV; I estimate that having an incarcerated parent increases years of schooling from 0.5 to 1 year. These estimates are statistically different from zero for the first and third tercile and for the pooled regression.

The effect on educational attainment I estimate combines two mechanisms. For children who finished their educational attainment, the effect captures the decision to drop out of school before high school completion, and for children who have not completed their schooling, the effect captures grade retention and continuous enrollment. I find evidence that suggests that both of these margins contribute to the positive effects; however, individually the data lacks power to estimate statistically significant effects. Tables B2 and B3 in the Appendix show the results. A second outcome variable I study is the probability of being convicted. For this exercise, I restrict the data to children who were 18 years old by 2017, so that they would have criminal records available. Figure C2 graphically depicts reduced-form estimates of judge stringency on conviction probability; we can see that the effect is close to zero. However, these estimates are very imprecise. This is natural, since conviction is a low incidence event; only 1.6% of children had a criminal record, and the difference in the OLS is only 0.1 pp.

5.3 Heterogeneity

In my context, marginal treatment effects (MTE) are particularly interesting because they trace the causal effect of incarceration along parents' unobserved characteristics (u^I), that matter for incarceration and that are correlated with parental quality, broadly defined. The intuition is the following: Parents who are incarcerated under the most lenient judges have worse characteristics than those incarcerated under strict judges. Basically, a strict judge incarcerates almost everyone, but a lenient judge incarcerates only the worst defendants, so those convicted under relatively lenient judges are more negatively selected.[21] I follow Heckman and Vitlacyl (2005) to estimate this MTE. I find that at the 10% level, there are heterogeneous treatment effects along parental quality (Figure 10). Specifically, I find that the positive effects of incarceration on schooling accrue when the worst parents go to prison.

The magnitude of the effect of parental incarceration on children's education is a function of the relationship between the parent and child prior to the incarceration episode, the type or quality of this parent, and the role of the child in the household. To document this heterogeneity, I estimate the IV regression for different subgroups in the data. Following previous literature in economics, but also in psychology and sociology, I estimate different regressions by gender of the child, gender of the parent, child's age at the time of the incarceration episode, birth order and the nature of the offense—violent *vs.* not violent. In Table 8 I show IV results for the pooled model for these different groups in the data.

According to the estimates, the benefits of parental incarceration are larger for boys than for girls, and this difference is statistically significant. Specifically, I find that boys' schooling increases by one year, whereas girls' schooling increases by 0.4 year, but the latter is not statistically significant. This result is consistent with previous research in psychology and economics, where it is documented that boys are more vulnerable than girls to negative experiences in the household (Bertrand & Pan (2013), Autor et al. 2016; Parke & Clarke-Stewart, 2002; Hetherington et al., 1998).

[21] I look at this empirically and find that among incarcerated defendants, those incarcted under more strict judges tend to have fewer and less severe charges. This follows almost directly from the definition of leniency, but helps to illustrate the way in which this defendants are "better".

Specifically, Autor et al. find that boys, relative to their sisters, have higher rates of disciplinary problems, lower achievement scores, and fewer high school completions when growing up in disadvantaged environments.

Next, I split the sample by gender of the parent and find that incarceration is more beneficial in the cases where the mother is the one going to prison. This result might be surprising at a first glance, but to understand it, it is important to bare in mind that children's wellbeing is more closely affected by their mother's behavior because of their main role as primary caregivers, and that criminal women are more negatively selected than criminal men (Table 2). Interestingly, Billings, 2018; and Turanovic et al., 2012; also estimate larger positive effects from maternal incarceration for the US; and it is also the case that in the U.S incarcerated women have worse socioeconomic backgrounds than incarcerated men (Harrison & Beck, 2006).

A source of heterogeneity that is associated with the quality of parent that goes to prison, is the type of crime they committed. Specifically, when looking at violent and non-violent crimes, I find that the benefits of incarceration are larger when the parent that leaves the households is convicted from a violent offense. This difference, however, is not statistically significant. It is nonetheless in line with the previous result on unobserved heterogeneity, in which the positive effects are a function of how good the defendant is as a parent.

Different roles of the child in the household might also lead to different effects. Specifically, I find larger positive effects for higher order children compared to the first born. The coefficient on the IV is still positive for first born children, but it reduces by more than 50%. This can be the result of two different mechanism that are hard to tell apart. On one side, it can the case that the positive effects from incarceration are attenuated by the need to cover for income and house production lost, and that this is more likely to be supplied by the older child. But on the other side, it could also be explained in the light of the literature on birth order and educational attainment, where it has been documented that first born children achieve higher schooling (Black, 2005). In this case, the younger siblings have grater scope to increase their education given their lower schooling in the baseline.

Lastly, I look at heterogeneous effects along the age of the child at the time of parental conviction. I split the sample in three groups: children who were 0 to 5 years, 6 to 10 years, and 11 to 15 years at the time of conviction. I find a U pattern in the effects on schooling. Specifically, I find very large positive effects for the first and third groups of around one year of education, and an increase of only 0.25 years for the second group, that is not statistically significant. Studies in developmental psychology conclude that children in the first age group are the most vulnerable, because they don't have yet the ability and skills to process trauma on their own —as opposed to older ages, Johnston (1995). This abilities develop over time helping children to cope with distress. On the other hand, the increase in the positive effect in the later years can be the result of how salient the decision to continue in school or drop out becomes.

5.4 Robustness

In this section I go over various exercises that evaluate the robustness of the results in the paper along different dimensions. First, the judge fixed effects approach (and in general all examiner fixed effects strategies) assumes that judges look at the same attributes of the defendant and weight these attributes in a similar fashion. Specifically, in my model the assumption is that I can write the conviction and incarceration decision as functions of $\xi^c(\cdot)$, and that $\xi^I(\cdot)$ is not judge specific. This is reasonable given that all judges go through the same training and have ultimately the same objective function, however it is still an assumption. To evaluate if the data support this assumption, for a handful of covariates, I estimate a random coefficient model, with different coefficients for each judge, and test whether this model provides a better fit, than a model with fixed judge coefficients. Table C.3 in the appendix shows that overall the data fails to reject the model with fixed coefficients per judge.

In Table 3 I report the first stage regression on incarceration and in the bottom of the table I report the F-test on the excluded instruments. This F-test corrects for the fact that the dimensionality of the instrument is the number of judges and not one (my measure of judge leniency). With this correction the F-stats are low, but above the critical values for weak instruments. The consequence of weak instruments is that the 2SLS-IV estimate will be biased towards the OLS (Stock et al., 2002). In my context, given that the OLS estimates are negative, the bias of the OLS is also negative, and the 2SLS IV estimates are positive, what this translates into is that we can expect even larger positive effects. To asses the size of this residual bias I perform two exercises; First, I estimate the IV using the LIML estimator which is less sensitive to weak instruments —the bias does not increase with the number of instruments (Rothenberg, 1993 and Stock et al., 2002)—, and second, I restrict my estimation to a sub-sample with a large number of cases and a small number of judges, case in which the F-stat is larger, and compare the estimates. Table C.6 in the Appendix shows the estimates for the LIML estimator, I find that as expected the estimated effect is larger but the 2SLS and the LIML are very close. For the second exercise I follow Belloni et al (2012), and use a LASSO approach to reduce the dimensionality of the instrument matrix. This results in dropping 50% of the judges and generates an 23% increase in the F-stat to 7. The bottom panel of Table C.6 shows the IV results for this subsample of judges. I find that the estimates are qualitatively the same as before.

In the results section, I show my preferred specifications for the estimates of the effect of parental incarceration on educational attainment. This choice of splitting the sample in three groups of P_c was arbitrary. To asses the robustness of this results, in Figure C.2, instead I order observation along P_c, and run multiple regression on a rolling window of 20.000 observations over P_c, moving the window 500 observations each time. Figure C.2 in the Appendix shows that for every sample I find a positive effect of incarceration on education.

Lastly, as a placebo check I evaluate whether there are differences in schooling from children of incarcerated and not incarcerated parents before the date of the sentence. Table C.7 in the appendix shows there is no evidence supporting that the positive effects I estimate are the results of pre-existing differences in educational attainment.

6 Mechanisms

6.1 What explains the positive effect?

The results presented here suggest that living with a convicted parent has negative consequences. There are many reasons to believe that this is the case. First, criminals are more likely to exert psychological and physical violence at home and this can be detrimental to the child's well being. Specifically, for the US, Western et al. (2004) find that incarcerated men engage in domestic violence at a rate about four times higher than the rest of the population. Literature in psychology documents that spending time with parents who engage in high levels of antisocial behavior is associated with more conduct problems for their children (Jaffee et al., 2003). This literature concludes that the salutary effects of being raised by two married, biological parents depend on the quality of care parents can provide.

Second, Chimeli and Soares (2017) document the causal effect of trading an illegal commodity on violence. In light of this, we can expect that households that take part in illegal businesses face constant violence or threats of violence related to guaranteeing property rights or resolving disputes within the business, all of which affect the quality of life in the household. There is also literature on the intergenerational transmission of violence, substance abuse and crime. Specifically, in the role-model hypothesis in psychology, in which children directly observe and model their parents' behavior, incarcerating parents could be beneficial by removing bad role models and by forcing children to update their beliefs about the consequences of criminal behavior (Hjalmarsson and Lindquist, 2012). Beyond intergenerational transmission, child exposure to these behaviors has been documented to have direct negative effects on outcomes in childhood and adulthood (Balsa (2008) and Chatterji and Markowitz (2000)).

6.2 How does the environment of the child changes?

To characterize the changes households and children experience after an episode of incarceration, for households for which I have two observations in the SISBEN (44% of them), I take those who had a parent convicted of a crime in between observations. Appearing in both waves of the SISBEN is not random, but to the contrary, leaving the sample is associated with an improvement in living standards. This is particularly relevant for children who might be moving to a household outside of SISBEN after the episode of parental incarceration. With this caveat, Table 9 shows that incarceration is associated with an increase in labor force participation (LFP) of the spouse, a worsening of the income score of the household, and a decrease in the probability of a male as the head of the household. I also find that the probability of living with grandparents increases and the probability of being in the second wave of SISBEN decreases, suggesting that incarceration induces children to moved-in with relatives who are better-off in terms of income.

Maternal and paternal incarceration are associated with very distinct changes in household living standards and composition. Table 9 regressions include interaction terms of the gender of the parent incarcerated. First, the increase in LFP of the spouse is completely driven by female spouses who enter the labor force after their male partner is incarcerated; this pattern is not observed when a women goes to prison. Second, the decrease in income after incarceration only comes from paternal incarceration; when the mother is incarcerated the members of the household do not see their income go down. When the father goes to prison, the probability of having a male as head of the household goes down by 9pp, whereas when a mother goes to prison this probability goes up by 3 pp. These changes have very different baseline levels. Whereas households in which the father is the defendant have a male as the head of the household 64% of the time, this number is only 35% when the mother is the defendant. The probability of a child living with his/her grandparents increases by 3pp when the father goes to prison, and decreases 6pp when the mother does. Again, the baseline levels are very different: 20% of households that experience paternal conviction have children living with grandparents, compared to 29% of those that experience maternal conviction.

6.3 The size of the parents at the margin

To derive policy implications, it is important to acknowledge the local feature of my estimates. This paper looks at the effects on children from convicted poor parents who are on the margin of incarceration. A large share of the population would be incarcerated regardless of judge assignment, and this paper does not provide any insights on the effects on educational attainment of the children of those individuals. Examples, of these are defendants guilty of murder or rape. On the other end of the distribution, defendants convicted of simple crimes will also avoid prison regardless of judge assignment. Defendants convicted of drug-and-gun trafficking, domestic violence and medium size property crimes compose the complier group in my estimation, and they are the group my estimates apply to. This marginal population, however, is particularly relevant because it is the population that is more likely to be affected by policy interventions to the criminal system.

To assess the size of the complier group, I follow Dahl, Kostøl, and Mogstad (2014). Parental compliers are defendants who would have received a different incarceration decision had their case been assigned to the most lenient judge instead of the strictest judge. We can define the size of this group (π_c) as follows:

$$\pi_c = Prob(Incarceration = 1|z_j = \bar{z}) - Prob(Incarceration = 1|z_j = \underline{z})$$

where $\bar{z}$ and $\underline{z}$ correspond to the incarceration rates of a judge at the 99th and 1st percentiles. Because of monotonicity, the share of parents who would go to prison regardless of the judge assigned to their case -always takers- is given by the incarceration rate for the most lenient judge and is equal to 22.5%. On the other hand, 47.7% of the sample are children of never takers who would not go to prison no matter which judge was assigned to their case. I estimate that children of compliers make up approximately 29.8% of the sample.

6.4 External validity and policy implications

To understand the external validity of my results, in Table C.2 I look at sentencing guidelines for a selected group of relevant crimes, and for reference I compare them to the guidelines for the state of New York.[22] The most salient feature of this

[22] These sentences can be decreased or increased whenever any mitigating or aggravating factors apply.

table is the amount of discretion judges have. In most cases, the guidelines for sentences span many years. Second, Table C.2 also shows an important difference across settings which is that sentences in Colombia tend to be longer than in the US.[23] This might imply that the relevant margin of comparison to the US is not the probation/short-term incarceration margin—because it is likely that most of these cases are never takers (of incarceration) in my set-up—but rather prisoners with medium-term sentences. In terms of the size of the treatment effects, my results apply to contexts in which dropping out of school in secondary education is relevant, the government offers little if any safety net to households, and the parental quality of the criminals at the margin of conviction.

Regarding policy implications these results call for greater support from the government in assisting children from fragile households. There is a strong body of experimental evidence on the powerful role of parenting and parenting supplements in shaping skills. Cunha et al. (2013) document the lack of parenting knowledge among disadvantaged parents. Early childhood intervention have been remarkably successful in complementing parental care with positive economic, psychological, behavioral, and health benefits (Heckman et al, 2010). The Perry program, which targeted very disadvantage kids from backgrounds where incarceration was a common feature, is an example of this. These programs have been successful in providing early supplements to parenting and can be a starting point to assist the parenting needs of this population.

7 Conclusions

The dramatic rise in incarceration in the US has led to an equivalent increase in the number of children growing up with a parent in prison. These trends have fueled a long-standing debate on the causal effects of parental incarceration on children. Children growing up with an incarcerated parent fare worse than those without one on a wide range of outcomes. Separating the causal effects of parental incarceration from pre-existing risk factors, has been a great challenge in the literature. In this paper I overcome this challenge and estimate the causal effects of parental incarceration on educational attainment. My results suggest that children benefit when their convicted parents are incarcerated. Specifically, I estimate that parental incarceration increases schooling by 0.6 years.

[23]This can be a consequence of the policy to deal with overcrowding in prisons. As a result, judges in Colombia have fewer instruments for punishment compared to the US. In the US, judges can decide among probation and incarceration length, whereas in Colombia, judges in practice only decide between no incarceration and medium-term incarceration.

I conclude with several important limitations of the paper. First, I only look at short-term effects of parental incarceration. This is important because these parents eventually leave prison and may return to live with their children. Even more, if incarceration decreases one's human capital and social and emotional skills, the type of parent that returns after incarceration, can be much worse than the one that left. In this case the long run effects can be very different than what I estimate. Another important shortcoming of this paper is that effectively I can only study one outcome variable. As shown by Dobbie et al, parental incarceration can have sizable effects on other variables such as earnings and teen pregnancy. These are important results that help characterize the complex shock that is having a parent incarcerated, but that due to data limitation I can not explore. Finally, given my sample selection my analysis is restricted to cases in which convicted parents are living with their children, which is not the majority of the cases, and to poor household. There are many reasons to believe my results do not extend to these other groups of children.

References

Abadie, A., Athey, S., Imbens, G.W. & Wooldridge, J., (2017). When Should You Adjust Standard Errors for Clustering? (No. w24003). National Bureau of Economic Research.

Abadie, A. (2003). Semiparametric Instrumental Variable Estimation of Treatment Response Models," Journal of Econometrics, 113(2), 231263.

Aizer, A. & J. J. Doyle (2015). Juvenile Incarceration, Human Capital and Future Crime: Evidence from Randomly-Assigned Judges. The Quarterly Journal of Economics 130 (2), 759–803.

Amato, P. R., Loomis, L. S., & Booth, A. (1995). Parental divorce, marital conflict, and offspring well-being during early adulthood. Social Forces, 73(3), 895-915.

Arditti, J.A., 2015. Family process perspective on the heterogeneous effects of maternal incarceration on child wellbeing. Criminology and Public Policy, 14(1), pp.169-182.

Arditti, Joyce (2012). Parental incarceration and the family: Psychological and social effects of imprisonment on children, parents, and caregivers. New York, NY: New York University Press.

Arditti, Joyce, Sara A. Smock, & Tiffaney S. Parkman (2005). It's been hard to be a father: A qualitative exploration of incarcerated fatherhood. Fathering 3:267–83.

Autor, D., Figlio, D., Karbownik, K., Roth, J., & Wasserman, M. (2016). Family disadvantage and the gender gap in behavioral and educational outcomes (No. w22267). National Bureau of Economic Research.

Balsa, A. I. (2008). Parental problem-drinking and adult children's labor market outcomes. Journal of Human Resources, 43(2), 454-486.

Bertrand, M. & Pan, J., 2013. The trouble with boys: Social influences and the gender gap in disruptive behavior. American Economic Journal: Applied Economics, 5(1), pp.32-64.

Bhuller, Manudeep, Gordon B. Dahl, Katrine V. Løken, & Magne Mogstad. 2017. "Incarceration, Recidivism and Employment." Quarterly Journal of Economics. 22648.

Billings, Stephen (2017) Parental Arrest and Incarceration: How Does it Impact the Children? (Preliminary draft)

Black, S.E., Devereux, P.J. & Salvanes, K.G., 2005. Why the apple doesn't fall far: Understanding intergenerational transmission of human capital. American Economic Review, 95(1), pp.437-449.

Black, S. E., Devereux, P. J., & Salvanes, K. G. (2005). The more the merrier? The effect of family size and birth order on children's education. The Quarterly Journal of Economics, 120(2), 669-700.

Chimeli, A. B., and Soares, R. R. (2017). The use of violence in illegal markets: Evidence from mahogany trade in the Brazilian Amazon. American Economic Journal: Applied Economics, 9(4), 30-57.

Cho, Rosa M. 2009a. "The Impact of Maternal Imprisonment on Children's Probability of Grade Retention: Results from Chicago Public Schools." Journal of Urban Economics, 65(1): 11-23.

Cho, Rosa M. 2009b. "The Impact of Maternal Incarceration on Children's Educational Achievement: Results from Chicago Public Schools." Journal of Human Resources, 44(3): 772-797.

Cunha, F., I. T. Elo, and J. Culhane (2013). Eliciting maternal expectations about the technology of cognitive skill formation. Working Paper 19144, NBER.

Currie, J. and Moretti, E., 2003. Mother's education and the intergenerational transmission of human capital: Evidence from college openings. The Quarterly Journal of Economics, 118(4), pp.1495-1532.

Dahl, G. B., A. R. Kostøl, and M. Mogstad (2014). Family Welfare Cultures. The Quarterly Journal of Economics 129 (4), 1711–1752.

Di Tella, R. and E. Schargrodsky (2013). Criminal Recidivism after Prison and Electronic Monitoring. Journal of Political Economy 121 (1), 28–73.

Dobbie, W., J. Goldin, and C. Yang (2016). The Effects of Pre-Trial Detention on Conviction,Future Crime, and Employment: Evidence from Randomly Assigned Judges. NBER Working Paper No. 22511.

Dobbie, W., H. Grönqvistz, S. Niknami, M. Palme and M. Priksk (2018). The Intergenerational Effects of Parental Incarceration. NBER Working Paper, January.

Doyle Jr, J. J. (2007). Child protection and child outcomes: Measuring the effects of foster care. American Economic Review, 97(5), 1583-1610.

Doyle Jr, J. J. (2008). "Child Protection and Adult Crime: Using Investigator Assignment to Estimate Causal Effects of Foster Care." Journal of Political Economy, 116(4): 746-770.

Ehrensaft, M.K., Cohen, P., Brown, J., Smailes, E., Chen, H. and Johnson, J.G., 2003. Intergenerational transmission of partner violence: a 20-year prospective study. Journal of consulting and clinical psychology, 71(4), p.741.

Furstenberg, F. F., Jr. (1995). Fathering in the inner city: Paternal participation and public policy. In W. Marsiglio (Ed.), Research on men and masculinities series, 7. Fatherhood: Contemporary theory, research, and social policy (pp. 119-147). Thousand Oaks, CA, US: Sage Publications, Inc.

Fomin, S. V. (1999). Elements of the theory of functions and functional analysis (Vol. 1). Courier Corporation.

Imbens, G.W., and D. B. Rubin (1997). Estimating Outcome Distributions for Compliers in Instrumental Variables Models," The Review of Economic Studies, 64(4).

Harrison, P. M. and A. J. Beck (2006). Prison and jail inmates at midyear 2005.

Hart, B. and T. R. Risley (1995). Meaningful differences in the everyday experience of young American children. Baltimore, MD: P.H. Brookes.

Heckman, James J., and Edward Vytlacil. 2005. "Structural Equations, Treatment Effects, and Econometric Policy Evaluation." Econometrica, 73(3): 669-738.

Heckman, J. J., S. H. Moon, R. Pinto, P. A. Savelyev, and A. Q. Yavitz (2010). Analyzing social experiments as implemented: A reexamination of the evidence from the HighScope Perry Preschool Program. Quantitative Economics 1 (1), 1–46.

Heckman, J. J., S. H. Moon, R. Pinto, P. A. Savelyev, and A. Q. Yavitz (2010b). The rate of return to the HighScope Perry Preschool Program. Journal of Public Economics 94 (1–2), 114–128.

Hetherington, E. M., Bridges, M., and Insabella, G. M. (1998). What matters? What does not? Five perspectives on the association between marital transitions and children's adjustment. American Psychologist, 53(2), 167.

Hjalmarsson, Randi, and Matthew J. Lindquist. 2011. "The Origins of Intergenerational Associations in Crime: Lessons from Swedish Adoption Data." Labour Economics, 20: 68-81.

Hjalmarsson, Randi, and Matthew J. Lindquist. 2012. "Like Godfather, Like Son: Exploring the Intergenerational Nature of Crime." Journal of Human Resources, 47(2): 550-582.

Hjalmarsson, Randi, Helena Holmlund, and Matthew J. Lindquist. 2015. "The Effect of Education on Criminal Convictions and Incarceration: Causal Evidence from Micro-data." Economic Journal, 125(587): 1290-1326.

Jaffee, S. R., Moffitt, T. E., Caspi, A., and Taylor, A. (2003). Life with (or without) father: The benefits of living with two biological parents depend on the father's antisocial behavior. Child development, 74(1), 109-126.

Kalil, A., 2015. Inequality begins at home: The role of parenting in the diverging destinies of rich and poor children. In Families in an era of increasing inequality (pp. 63-82). Springer, Cham.

Kim-Cohen, J., Moffitt, T.E., Taylor, A., Pawlby, S.J. and Caspi, A., 2005. Maternal depression and children's antisocial behavior: nature and nurture effects. JAMA Archives of general psychiatry, 62(2), pp.173-181.

Kling, J. R. (2006). Incarceration Length, Employment, and Earnings. The American Economic Review 96 (3), 863–876.

Lefgren, L., Sims, D. and Lindquist, M.J., 2012. Rich dad, smart dad: Decomposing the intergenerational transmission of income. Journal of Political Economy, 120(2), pp.268-303.

Lyle, D. S. (2006). Using military deployments and job assignments to estimate the effect of parental absences and household relocations on children's academic achievement. Journal of Labor Economics, 24(2), 319-350.

McLanahan, S., Tach, L., and Schneider, D. (2013). The causal effects of father absence. Annual review of sociology, 39, 399-427.

Mueller-Smith, M. (2017). The Criminal and Labor Market Impacts of Incarceration. University of Michigan Working Paper.

Murray, Joseph, and David P. Farrington. 2005. "Parental Imprisonment: Effects on Boys' Antisocial Behaviour and Delinquency Through the Life-Course." Journal of Child Psychology and Psychiatry, 46(12): 1269-1278.

Murray, Joseph, David P. Farrington, and Ivana Sekol. 2012. "Children's Antisocial Behavior, Mental Health, Drug Use, and Educational Performance After Parental Incarceration: A Systematic Review and Meta-analysis." Psychological Bulletin, 138(2): 175-210.

Murray, Joseph, Rolf Loeber, and Dustin Pardini. 2012. "Parental Involvement in the Criminal Justice System and the Development of Youth Theft Marijuana Use, Depression and Poor Academic Performance." Criminology, 50(1): 255-302.

Murray, Joseph, Carl-Gunnar Janson, and David P. Farrington. 2007. "Crime in Adult Offspring of Prisoners: A Cross-National Comparison of Two Longitudinal Samples." Criminal Justice and Behavior, 34(1): 133-149.

Parke, R. D. and Clarke-Stewart, K. A. (2003). The effects of parental incarceration on children, Prisoners once removed: The impact of incarceration and reentry on children, families, and communities pp. 189–232.

Pew Center. 2011. "State of Recidivism: The Revolving Door of America's Prisons." Washington, DC: Pew Charitable Trusts.

Stith, S.M., Rosen, K.H., Middleton, K.A., Busch, A.L., Lundeberg, K. and Carlton, R.P., 2000. The intergenerational transmission of spouse abuse: A meta-analysis. Journal of Marriage and Family, 62(3), pp.640-654.

Turanovic, J. J., Rodriguez, N., and Pratt, T. C. (2012). The collateral consequences of incarceration revisited: A qualitative analysis of the effects on caregivers of children of incarcerated parents. Criminology, 50(4), 913-959.

Western, Bruce. 2006. Punishment and Inequality in America. New York: Russell Sage Foundation.

Western, Bruce and Christopher Muller. 2013. "Mass Incarceration, Macrosociology, and the Poor." The Annals of the American Academy of Political and Social Science 647: 166–89.

Western, Bruce, Leonard M. Lopoo, and Sara S. McLanahan. 2004. "Incarceration and the Bonds between Parents in Fragile Families." Pp. 21–45 in Imprisoning America: The Social Effects of Mass Incarceration, edited by M. Patillo, D. Weiman, and B. Western. New York: Russell Sage Foundation

Western, B. and Pettit, B.: 2010, Collateral costs: Incarceration's effect on economic mobility, Washington, DC: The Pew Charitable Trusts .

Western, B., 2018. Homeward: Life in the Year After Prison.

Wildeman, Christopher. 2010. "Paternal Incarceration and Children's Physically Aggressive Behaviors: Evidence from the Fragile Families and Child Wellbeing Study." Social Forces, 89(1): 285-309.

Wildeman, Christopher, and Bruce Western. 2010. "Incarceration in Fragile Families." The Future of Children, 20(2): 157-177.

Wildeman, Christopher, Signe Hald Anderson, Hedwig Lee and Kristian Bernt Karlson. 2014. "Parental Incarceration and Child Mortality in Denmark." American Journal of Public Health, 104(3): 428-433.

Table 1: Population by conviction and incarceration

Sample:	Census: Adult population	SISBEN Criminal record No	Yes	SISBEN w/ conviction Incarcerated No	Yes
	(1)	(2)	(3)	(4)	(5)
Years of education	7.36	6.82	6.68	6.86	6.42
Finished High School D=1	44.0%	31.2%	22.8%	24.2%	20.8%
Income score		34.01	30.90	31.72	29.41
Gender (Male=1)	49.0%	47.6%	83.3%	84.5%	83.3%
# HH members	3.90	4.28	4.47	4.37	4.43
Occupation: Working D=1	48.0%	47.3%	65.4%	67.0%	63.9%
Head of the household D=1		41.2%	47.1%	46.9%	48.6%
Year of birth	1965	1966.9	1974.8	1975.0	1974.3
Marital status: Single D.	45.0%	34.7%	40.7%	45.0%	43.6%
Obs	26,757,687	16,195,178	89,257	55,790	33,467

Notes: Columns 1-5 are group means. Rows in column 6 and 7 correspond to separate regressions on a criminal record Dummy with controls. HHH: Head of the household, HS: High School. D: Dummy. Income Score: Score from 0 to 100, calculated using variables on income and education of the members of the household, size and characteristics of the house. Source: 2005 Census, SISBEN and criminal records.

Table 2: Convicted parents by incarceration and gender

Convicted sample: by gender and incarceration status	Women		Men	
	No	Yes	No	Yes
	(1)	(2)	(3)	(4)
Years of education	6.50	6.06	6.68	6.23
Dummy Has HS degree =1	20%	16%	22%	19%
Income Score	17.2	16.1	19.48	18.46
Occupation: Dummy Working=1	45%	40%	69%	68%
Dummy head of the household=1	36.2%	37.1%	47%	50%
Age at sentence	35.5	36.2	34.46	36.31
Marital status: Dummy Single=1	47.8%	45.1%	46%	44%
Obs	9,375	6,028	46,415	27,439

Notes: Columns 1-4 are group means. HHH: Head of the household, HS: High School. D: Dummy. Income Score: Score from 0 to 100, calculated using variables on income and education of the members of the household, size and characteristics of the house. Source: SISBEN and criminal records.

Table 3: First stage - Parents

Dep var: Decision Dummy	(1)	(2)	(3)	(4)
	Conviction	Conviction	Incarceration	Incarceration
Judge Stringency	1.033***	0.859***	1.280***	1.286***
	[0.0159]	[0.0235]	[0.0331]	[0.0336]
Controls		X		X
F stat	4,240	1,335	1,493	1,468
F stat c	5.059	1.614	5.68	5.59
Obs	226,085	74,885	89,199	88,274
Judges	835	818	262	262
R-sq	0.018	0.038	0.252	0.252
adj. R-sq	0.018	0.038	0.247	0.247

Controls column 2: gender, age number of crimes and crime category . Controls column 3: years of education, gender, income score, year of birth, occupation, year of survey. Standard errors clustered at the randomization unit -year level. Source: Attorney General's office, criminal records and poverty census. Fstat-c adjusts for the number of judges.

Table 4: Balance test-Trial sample

Dep. Var: Conviction / Incarceration stringency	Judge: Conviction stringency	Judge: Incarceration stringency
Age	0.0000024	0.00000971
	[0.0000208]	[0.0000355]
Gender	0.000324	-0.000212
	[0.000509]	[0.000768]
Number of charges	0.000867	0.00125
	[0.000835]	[0.00162]
Violent crime	-0.000293	0.0019
	[0.000805]	[0.00131]
Property crime	0.00203	0.00158
	[0.00224]	[0.00360]
Drugs related crime	-0.000927	-0.00242
	[0.00157]	[0.00272]
Guns related crime	-0.000666	-0.00127
	[0.00142]	[0.00215]
Misdeminour	-0.000867	0.0014
	[0.00112]	[0.00185]
Obs	187,231	162,960
Judges	1,272	683
F test	0.801	0.748

Standard errors clustered at the randomization unit/year level. Each rows corresponds to a different regression of judge leniency and defendant characterisitcs. When testing balance across crime categories I construct an alternative measure of conviction stringency that doesnt parse-out crime level conviction rates. The F-test corresponds to a regression where I inlcude all the variables at the same time. Source Attorney General's office and criminal records.

Table 5: Balance test II-Incarcerated sample

Dep var: Incarceration FE	(1) 0.74<Pc<0.88	(2) 0.88<Pc<0.9	(3) 0.9<Pc<1	(4) Pooled Pc
Years of education	-0.0000292	-0.0000215	0.000274	0.00011
	[0.000119]	[0.000136]	[0.000169]	[0.0000873]
Income score	-0.0000174	0.00000267	0.000013	0.0000106
	[0.0000283]	[0.0000292]	[0.0000364]	[0.0000175]
Age at sentence	0.0000218	-2.08E-08	0.0000107	0.0000197
	[0.0000338]	[0.0000320]	[0.0000435]	[0.0000266]
Gender	-0.00142	0.001	-0.00212**	-0.00104
	[0.00127]	[0.000793]	[0.00100]	[0.000633]
Years of education HH	-0.0000463	0.000106	-0.000153	-0.0000165
	[0.000157]	[0.000136]	[0.000162]	[0.0000996]
D: Working	-0.0000919	-0.000981	0.000137	-0.000126
	[0.000672]	[0.000763]	[0.00108]	[0.000493]
D: Studying	-0.0022	-0.000602	0.00103	0.00108
	[0.00316]	[0.00278]	[0.00364]	[0.00199]
D: Both census surveys	-0.000844	-0.000942	0.000587	-0.000305
	[0.000897]	[0.000634]	[0.000857]	[0.000488]
D: First survey	0.000355	0.000691	0.000648	0.000511
	[0.00124]	[0.00123]	[0.00162]	[0.000800]
Constant	0.178*	-3.04E-01	6.64E-02	0.360***
	[0.107]	[0.226]	[0.124]	[0.00594]
F Test	0.8494	0.5001	0.564	0.5763
Obs	16,684	17,416	15,845	49,945
R-sq	0.128	0.149	0.137	0.03

Additional controls: Pc, Municipality FE, sentence year FE. Standard errors clustered at the randomization unit year level.

Table 6: OLS Regression

OLS: no controls	(1)	(2)	(3)	(4)
Dep var: Years of education	0.74<Pc<0.88	0.88<Pc<0.9	0.9<Pc<1	Pooled Pc
Parental Incarceration Dummy	**-0.379*** **	**-0.269*** **	**-0.438*** **	**-0.368*** **
	[0.0935]	[0.0815]	[0.0912]	[0.0434]
Constant	6.639***	6.980***	6.799***	6.815***
	[0.0880]	[0.115]	[0.104]	[0.0435]
Obs	16,631	18,454	15,173	50,258
Clusters	405	380	386	661
R-sq	0.004	0.002	0.005	0.004
OLS: Adding controls				
Parental Incarceration Dummy	-0.0789*	-0.112***	-0.00749	-0.0766***
	[0.0450]	[0.0371]	[0.0402]	[0.0234]
Constant	7.888***	10.25***	7.486***	8.984
	[1.782]	[0.173]	[1.885]	[8.563]
Obs	16,631	18,454	15,173	50,258
Clusters	405	380	386	606
R-sq	0.391	0.375	0.374	0.372

Controls: Municipality FE, gender, YOB FE, year of sentence, birth order and year of survey. Column 4 controls for Pc nonparametrically. Sample: Children between 1990 and 2007 who had a convicted parent between ages 0 and 14. I cluster at the randomization unit and year level.

Table 7: Results: Reduced form and IV

Reduced form	(1)	(2)	(3)	(4)
Dep var: Years of education	0.74<Pc<0.88	0.88<Pc<0.9	0.9<Pc<1	Pooled Pc
Judge stringency	**0.996** **	**0.626**	**0.930** **	**0.794*** **
	[0.476]	[0.575]	[0.382]	[0.267]
Obs	16631	18454	15173	50258
Clusters	405	380	386	606
R-sq	0.391	0.375	0.374	0.372
IV Dep var: Years of education	(1)	(2)	(3)	(4)
	0.74<Pc<0.88	0.88<Pc<0.9	0.9<Pc<1	Pooled Pc
Parental Incarceration Dummy	**0.541** **	**0.604**	**1.195** **	**0.684*** **
	[0.276]	[0.566]	[0.543]	[0.249]
Obs	16,631	18,454	15,173	50,258
Clusters	405	380	386	606

Controls: Municipality FE, gender, YOB FE, year of sentence, birth order and year of survey. Column 4 controls add a second order polynomial on Pc. Sample: Children between 1990 and 2007 who had a convicted parent between ages 0 and 14. SE in brackets clustered at the randomization unit and year level.

Table 8: Heterogenous effects

IV	**Girls**	**Boys**	**Mother**	**Father**
Dep var: Years of education	(1)	(2)	(3)	(4)
Parental Inc.	0.4	1.009***	1.037**	0.646***
	[0.280]	[0.333]	[0.525]	[0.238]
Obs	24,803	25,455	11,249	39,009
	Type of crime		**Birth order**	
	Violent	Not violent	First	Rest
Parental Inc.	0.840**	0.594**	0.386	0.870***
	[0.356]	[0.281]	[0.318]	[0.293]
Obs	19,072	31,186	21,475	28,783
	Age of the child			
	0-5 years	5-10 years	10-15 years	
Parental Inc.	0.924***	0.246	1.058*	
	[0.297]	[0.337]	[0.634]	
Obs	17,728	22,345	8,671	
Pooled Pc	x	x	x	x

Pooled regression. Controls: Municipality FE, gende, YOB FE, year of sentence, birth order and year of survey. Not violent crimes: Property, drugs and misdemi-nours. Violente crimes: violent and gun related crimes. Stars in the label reffer to statistical significance of the differences in coefficients.

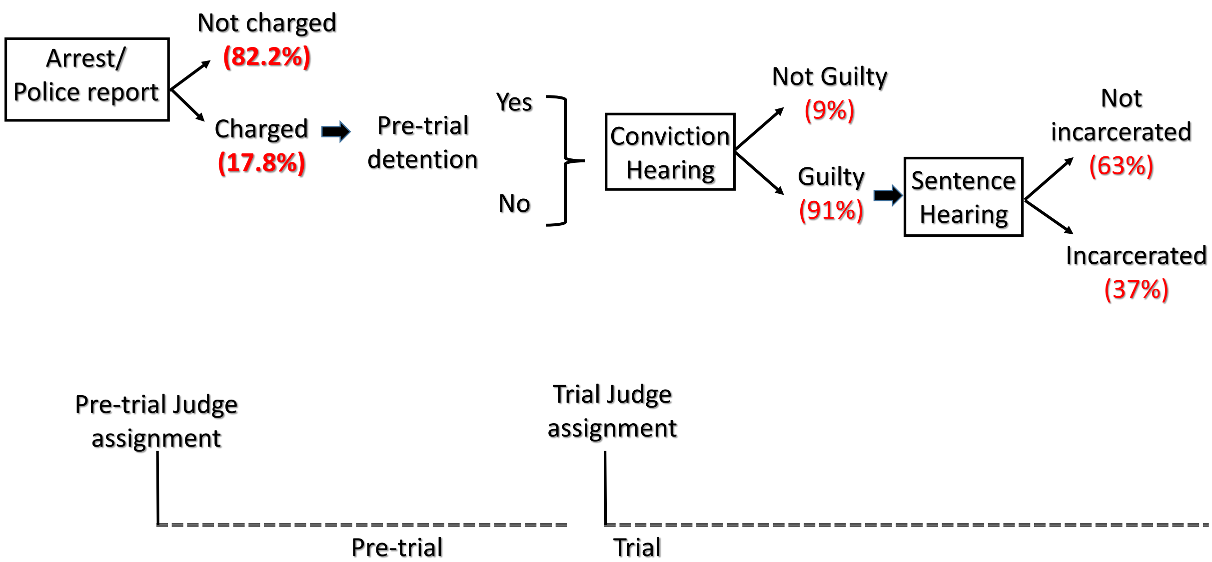
Table 9: Changes after incarceration

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep var:	LFP spouse	Income score	Years of educ. HHH	D: Male HHH	D: Lives w/ Grandpar-ents	# of peo-ple in HH	D: Both waves
Parental Inc.	0.0680***	-2.879***	0.0713**	-0.125***	0.0329***	-0.0905***	-0.0319***
	[0.0189]	[0.206]	[0.0315]	[0.00632]	[0.0118]	[0.0319]	[0.00546]
Parental Inc. # Mother	-0.069	3.573***	0.202***	0.346***	-0.100***	-0.0667	0.00593
	[0.115]	[0.549]	[0.0721]	[0.0144]	[0.0235]	[0.0733]	[0.0123]
Mean Dep. Var Father Inc.	39%	26.82	5.20	75%	4.61	19.9%	24.6%
Mean Dep. Var Mother Inc.	76%	23.81	4.49	40%	4.94	28.7%	22.6%
Obs	9,672	82,777	82,777	82,777	81,614	16,371	32,876
R-sq	0.22	0.75	0.20	0.20	0.33	0.11	0.08

Panel regressions. Controls: Poverty score, years of education of HHH, Municiaplity FE and year of survey FE. Dummy for living with grandparents also includes uncles and cousins. Households with data on both cross-sections of the poverty census and who had an conviction episode in between surveys. Source: SISBEN and criminal records.

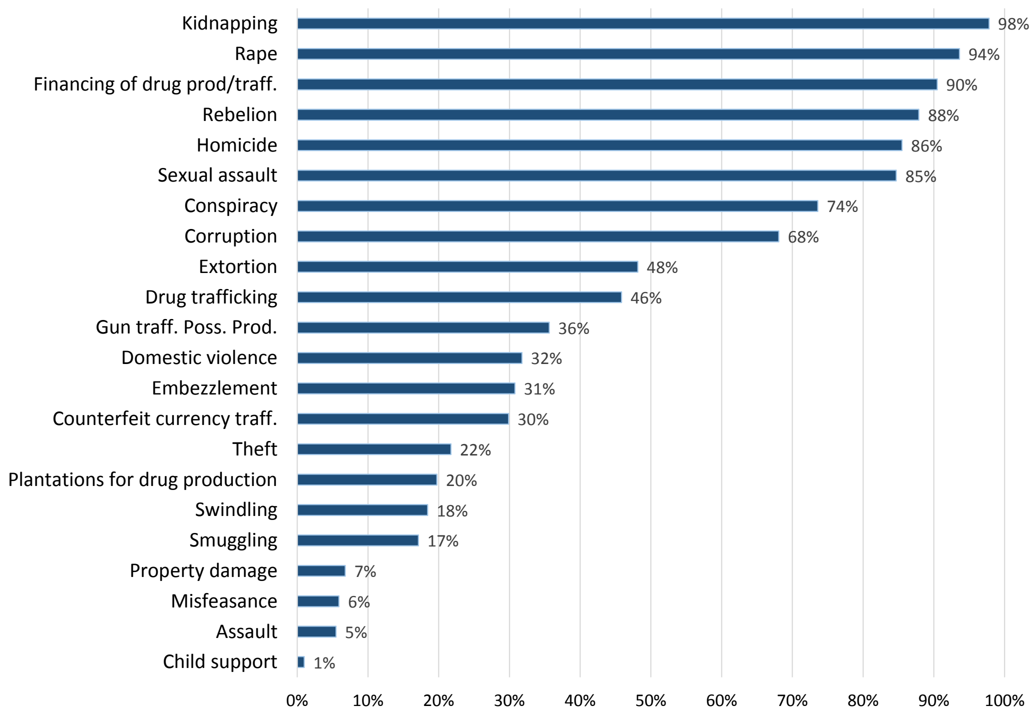
Figures

Figure 1: Prosecution and trial stages



Source: Colombian Penal proceedings code, Informe de la Comision Asesora de Politica Criminal (2012), SPOA and Criminal records.

Figure 2: Incarceration rates



Source: Criminal records. Selected crimes. I restrict to crimes with at least 100 cases.

Figure 3: Identification

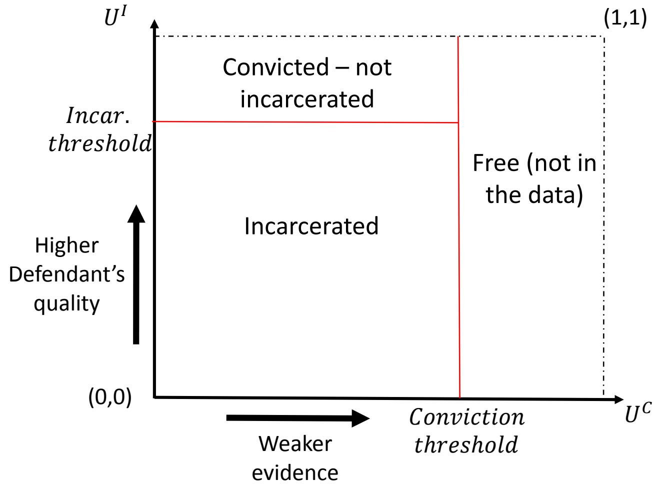


Figure 4: Identification

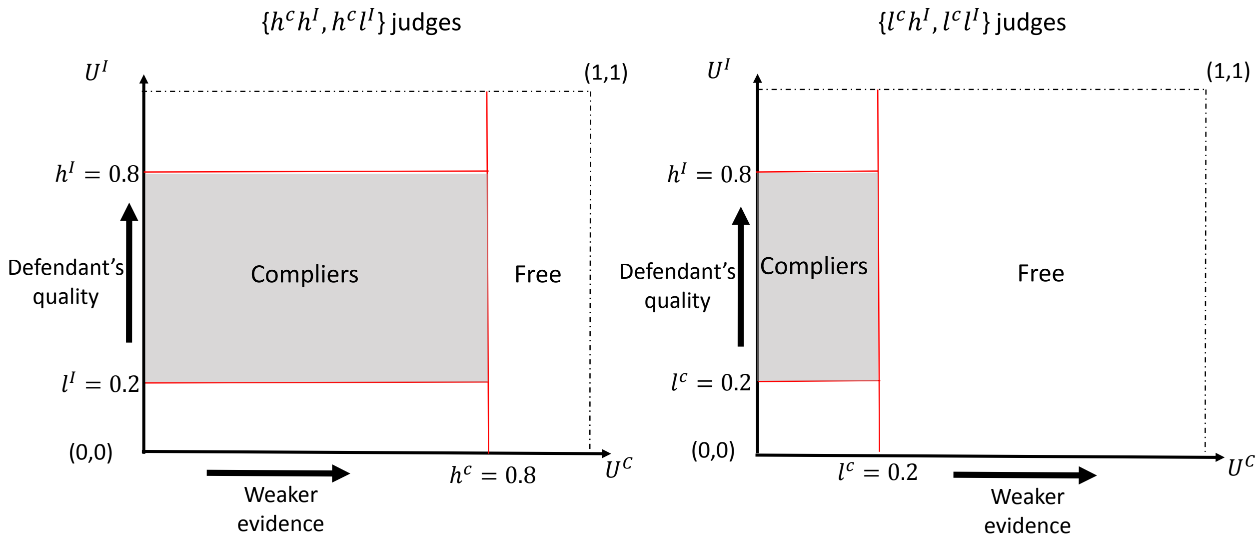


Figure 5: Identification

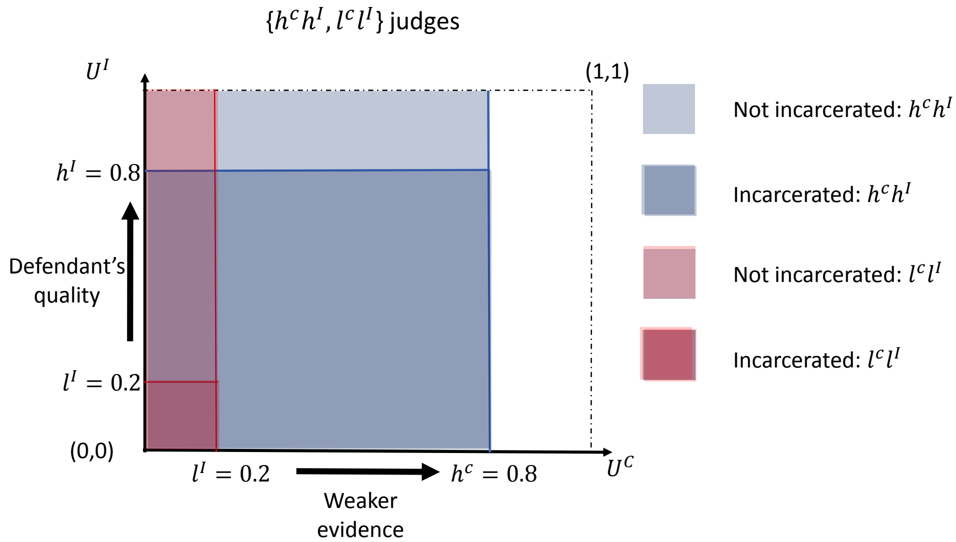
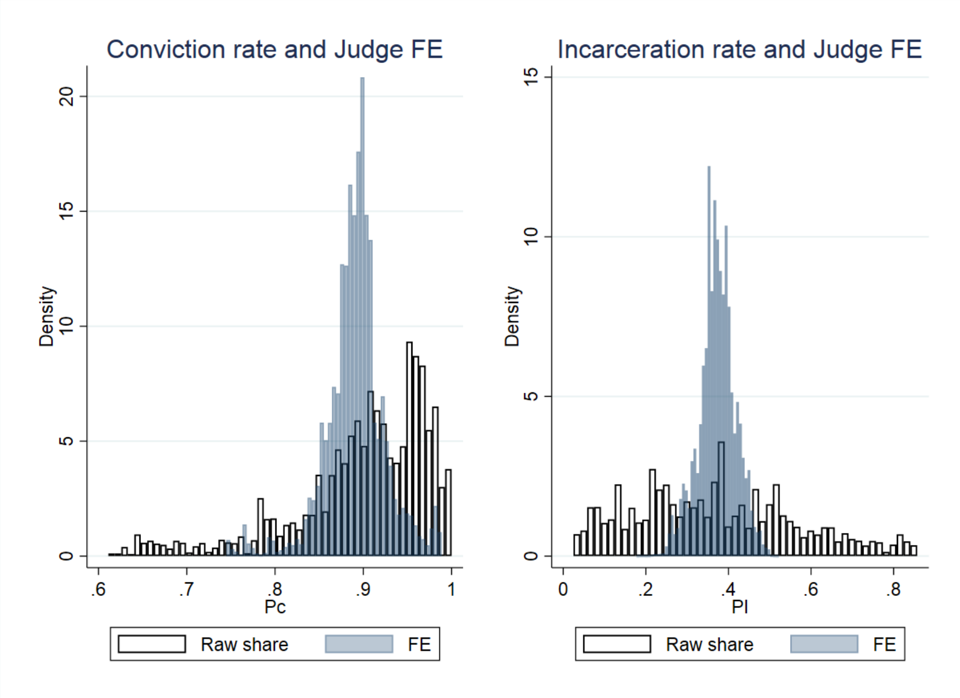
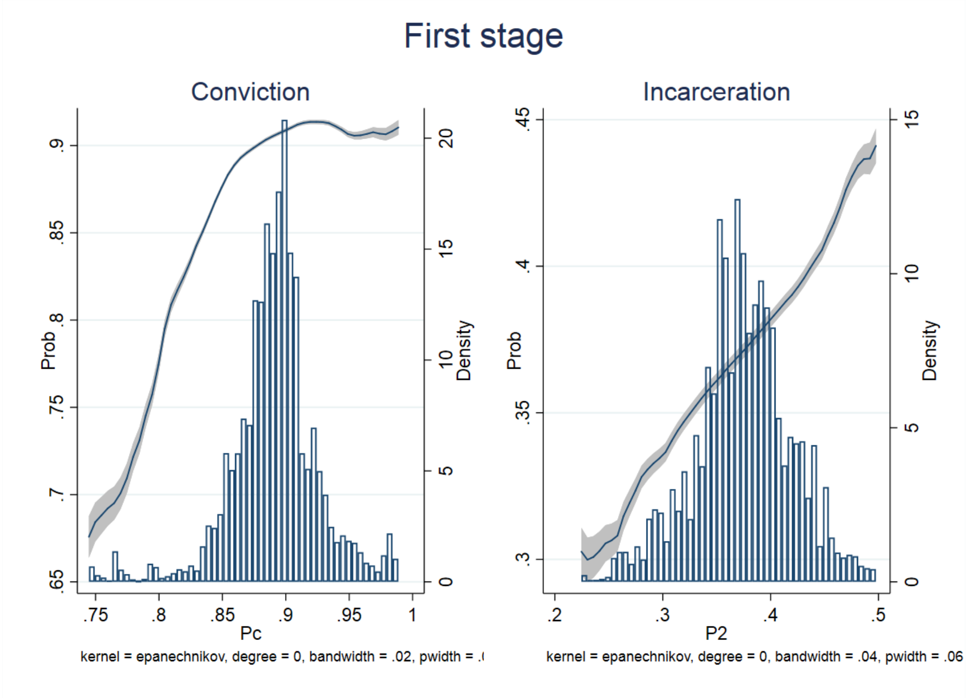


Figure 6: Judges' fixed effects



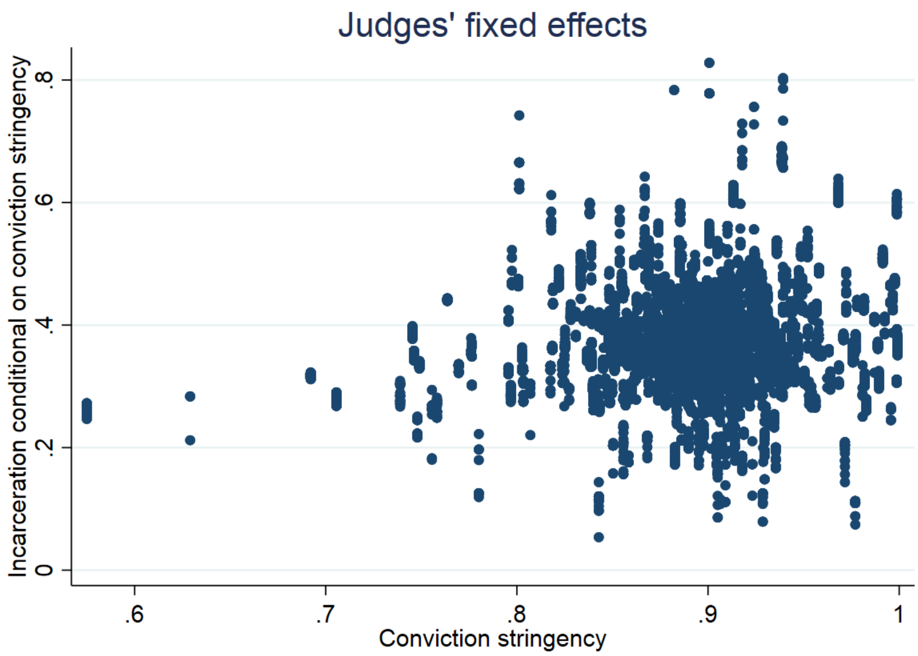
Source: Attorney General's office and criminal records. Raw rates are conviction/incarceration averages by judge. To construct the judge's fixed effect I take the residual at the judge level after regressing conviction/incarceration on (demeaned) randomization unit/year dummys, (demeaned) crime-level conviction/incarceration rates, without a constant.

Figure 7: First stage



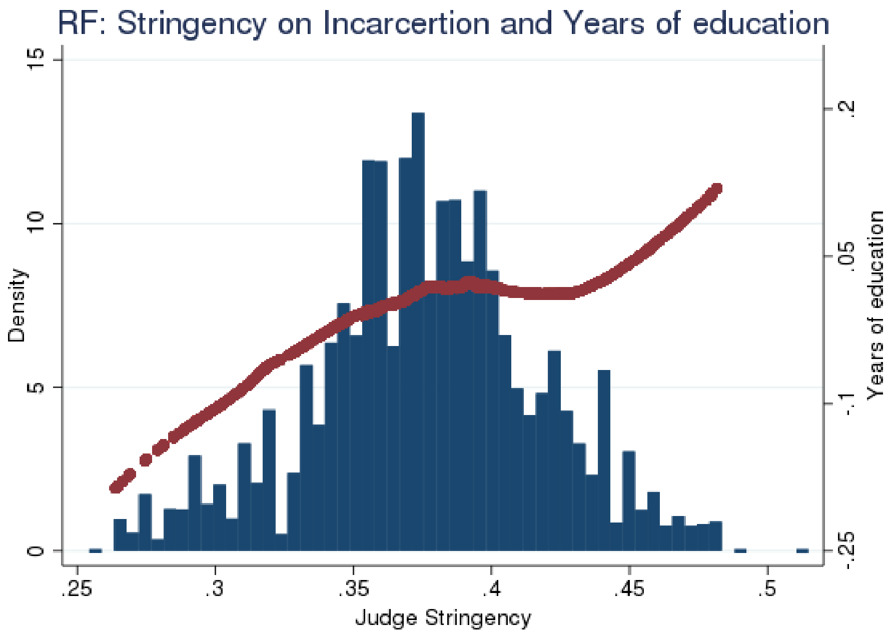
Source: Attorney General's office and criminal records. Raw rates are conviction/incarceration averages by judge. To construct the judge's fixed effect I take the residual at the judge level after regressing conviction/incarceration on (demeaned) randomization unit/year dummys, (demeaned) crime-level conviction/incarceration rates, without a constant.

Figure 8: Scatter plot: Judges' fixed effects



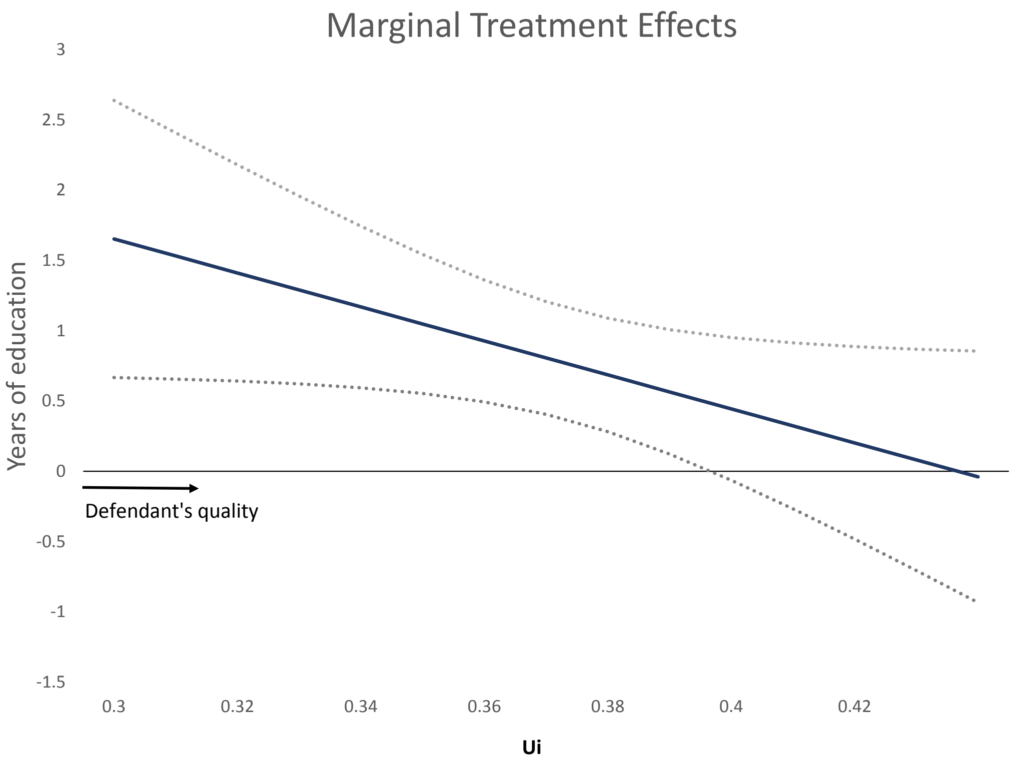
Source: Attorney General's office and criminal records. To construct the judge's fixed effect I take the residual at the judge level after regressing conviction/incarceration on (demeaned) randomization unit/year dummys, (demeaned) crime-level conviction/incarceration rates, without a constant.

Figure 9: Reduced form



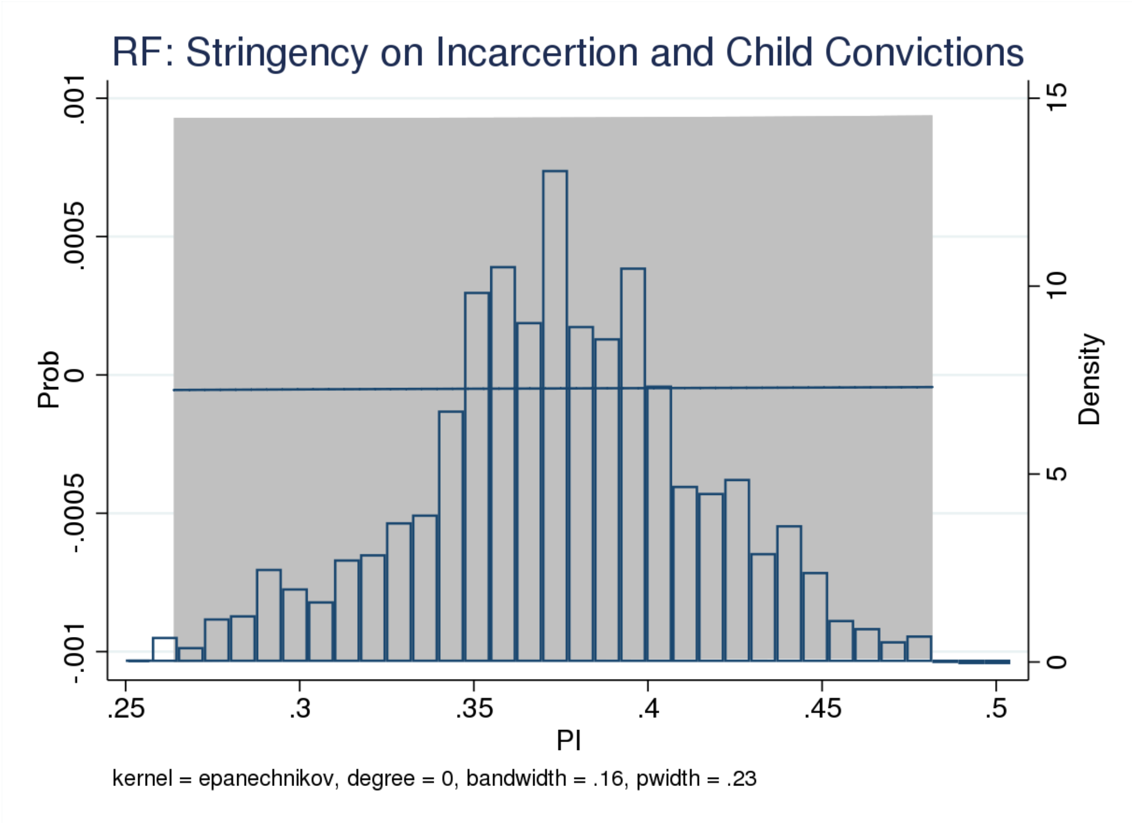
Notes: Histograms of parental incarceration judge stringency and the fitted value of local polynomial regressions of children's educational attainment on judge stringency. Pooled regression I control for p_c.

Figure 10: MTE



Notes: Following the LIV approach in Heckman and Vytlacil (2005) I regress $Yeduc = \alpha + \beta_1 P_i + \beta_2 P_i^2 + \beta_3 X$. This graphs plots: $\beta_1 P_i + 2\beta_2 P_i$ for the pooled regression.

Figure 11: Reduced form



Notes: Histograms of parental incarceration judge stringency and the fitted value of local polynomial regressions of children's criminal records on judge stringency.

A Appendix: Intuition for the 2 dimension LATE

In this section I go over the intuition of the results in eq. 18 and eq.19. This result extends the intuition behind LATE to a two-dimensional space. To make this point clear, let us think in discrete terms and use an example with 4 judges with threshold levels $\{P_c^1,P_I^1\}$, $\{P_c^1,P_I^2\}$, $\{P_c^2,P_I^1\}$, and $\{P_c^2,P_I^2\}$.[24]

For notation purposes, let:

$$f(p_c, p_I) = E(Y\mathbf{1}[T \in \{t_c\}]|P_c(Z) = p_c, P_I(Z) = p_I) \tag{29}$$

and

[24] Equivalent to {HL}, {HH}, {LH}, and {LL} in Section 4.

$$g(p_c, p_I) = E(Y\mathbf{1}[T \in \{t_I\}]|P_c(Z) = p_c, P_I(Z) = p_I) \tag{30}$$

Next, I can rewrite, in discrete terms, the identification result in equation 5 as:

$$\begin{aligned} &\frac{\Delta f(p_c, p_I)}{\Delta p_c \Delta p_I} + \frac{\Delta g(p_c, p_I)}{\Delta p_c \Delta p_I} = \\ &\qquad [f(p_c^2, p_I^2) - f(p_c^1, p_I^2)] - [f(p_c^2, p_I^1) - f(p_c^1, p_I^1)] + \\ &[g(p_c^2, p_I^2) - g(p_c^1, p_I^2)] - [g(p_c^2, p_I^1) - g(p_c^1, p_I^1)] = E(Y(t_I) - Y(t_c)|u^c = p_c, u^I = p_I) \end{aligned} \tag{31}$$

Now, let us go over each term in (31). First, $f(p_c^2, p_I^2)$ represents the outcomes of convicted but not incarcerated individuals who had a judge with thresholds $\{P_c^2, P_I^2\}$. Panel a in Figure C.3 shades the area in the u^c, u^I square that identifies these individuals. The next panels in Figure C.4 highlight the following terms in equation 8 and their differences. Ultimately, what equation (31) is doing is identifying the complier range in a two-dimensional space, which instead of an interval is a rectangle.

I estimate (18) by fitting a polynomial on p_I and p_c and evaluating the cross-derivative on the support of the instruments. Figure C5 shows the MTE in the relevant segment of the (u^c, u^I) square. There are some interesting features of these results; first, as before, as we increase u^I (defendants' quality), the effect on years of schooling decreases, confirming that this positive effect is accrued when incarceration removes a bad parent from the household. What is new in Figure C.5 is that now we can also move along the u^c margin, or the "strength of the evidence" margin. The data also show that as evidence becomes weaker, the positive effects also decrease. Ultimately, what this exercise shows is that the effect on children is very sensitive to the type of case a judge is deciding on. In the case of Colombia, marginal incarcerations are of defendants still very negatively selected and with sufficient evidence against them, so that their children are better off without that parent. How this result extends to other settings is a function of the location of the marginal cases in the u^c, u^I square.

B Appendix : Data construction

In this appendix, I explain in detail the construction of the sample and variables I use throughout the paper. The starting point for my data construction are the two SISBEN surveys. These data are collected by the government to target social programs for the poor. The survey is conducted at the household level, and consists of two modules. In the first, it asks about the characteristics of the house (flooring material, number of bedrooms, etc), access to utilities, and assets in the households (TV, refrigerator, car, etc.). In the second part, all members of the household are listed with names and national identification numbers, and their relationship to the head of the household is specified. The questionnaire then asks about gender, age, education level, marital status, disability status, and occupation. This survey is applied to everyone living in a municipality with a population of 30,000 or less, and in larger municipalities local authorities target households who could be potential beneficiaries of welfare programs. If a household is not targeted by local authorities and wishes to be surveyed, it can easily request to be included. The government uses this information to create a formula that measures the household's ability to provide resources for its members, and computes a score for each household that determines eligibility for different social programs. These data provide me with i) identification numbers with municipality location to web-scrape criminal records and, ii) parent-to-child links.

I select the population of adults who lived in the 17 out of 33 municipalities that have criminal records online. These districts represent 67% of the population, and 69% of homicide and 83% of property crimes.[25] I then web-scrape criminal records (from http://procesos.ramajudicial.gov.co/consultaprocesos/) by selecting the district and then searching individually for records with the ID numbers. From a 5% sample in which I look for criminal records in all 17 districts I estimate that I will miss 8.6% of the sample due to crimes committed in districts different from the one in the SISBEN.

I find 328,937 criminal records that belong to 256,366 individuals. I start by dropping observations that have missing values in year of sentence, crime or courtroom identifier (81,049 observations deleted). Next, I drop all records before 2005 (59,872 observations deleted), and all cases in which there is only one judge per district (4,635 observations deleted). I keep only the courtrooms for which there is data on convictions (14,786 observations deleted). Finally, I drop all observations where there are less than 15 cases in a year/judge cell (56,268 observations deleted). After this, I end up with 112,696 criminal records which correspond to 93,676 individuals. Table B.1 shows differences between the characteristics of individuals in the final data-set and those who were dropped. For the set of observations that

[25] Judicial districts with online data: Armenia, Barranquilla, Bogota, Bucaramanga, Buga, Cali, Ibague, Florencia, Manizales, Medellin, Neiva, Palmira, Pasto, Pereira, Popayan, Tunja, and Villavicencio.

have sentence data, I find that there is no evidence of differential incarceration rates across samples.

To assess how representative my sample is of the prison population, I compare counts of individuals sentenced by year from my data with counts of new inmates from official records of the Prison Authority (INPEC). I only have information available for 2015; according to INPEC, there were 27,287 new immates that year, from my data, I find that 5,932 defendants were sent to prison, which would suggest that I have data on 22% of the prison population. This number, however, should be taken with caution, because INPEC data include flows of inmates across prisons, and I don't have data on the size of these flows.

I then link these convicts to the 436,309 individuals living in their households, of whom 179,699 are in the relevant cohort years (1991-2007), and 106,465 are the child of a convict. Of this, 67,770 experienced the sentencing episode between ages 0 and 14. Finally, I have education data for 52,419 (77%) of these children. This rate is close to the share of children between ages 12 and 17 who attend school, according to the census (76%). Table B.2 in the appendix shows evidence that a missing education record is not related to parental incarceration, but to the child's not being at school or being working. Missing values are also more prevalent for boy, and for household with lower income and lower education of the head of the household.

Table B1: Sample selection-Defendants

Dep var: Out of sample D.	(1)	(2)
Incarceration		0.00141
		[0.00204]
Years edu.	0.0018	0.00118
	[0.00150]	[0.00157]
Income score	0.00118***	0.000837***
	[0.0000822]	[0.0000879]
Male D.	-0.0400***	-0.0209***
	[0.00279]	[0.00290]
Head HH D.	0.00877**	0.00771**
	[0.00370]	[0.00389]
Single	-0.0298***	-0.0213***
	[0.00222]	[0.00239]
Years edu. HHH	0.0004	0.000919
	[0.00150]	[0.00157]
D: Studying	0.0264***	-0.00653
	[0.00490]	[0.00486]
D: Working	0.0177***	0.0154***
	[0.00209]	[0.00226]
Yob	-0.00708***	-0.00312***
	[0.0000877]	[0.0000956]
Constant	14.55***	6.55E+00
	[0.173]	[3279.3]
Obs	260,968	196,314
R-sq	0.14	0.306

Additional controls: Municipality FE and survey year FE. The first column includes all criminal records and the second restricts to the ones that have data on sentence length.

Table B2: Sample selection

Dep var: Missing Educ.	(1)	(2)	(3)
Parental incarceration	-0.00245	-0.00314	-0.00335
	[0.00309]	[0.00309]	[0.00308]
Gender	0.00851***	0.00853***	0.00758***
	[0.00282]	[0.00282]	[0.00280]
Yob	0.0205***	0.0200***	0.0125***
	[0.000478]	[0.000479]	[0.000582]
Gender of the parent	-0.00366	0.00251	0.00186
	[0.00344]	[0.00347]	[0.00346]
Income score		-0.000886***	-0.000582***
		[0.000136]	[0.000136]
Years edu. HHH		-0.00493***	-0.00469***
		[0.000620]	[0.000617]
D: Studying			-0.0872***
			[0.00384]
D: Working			-0.0633
			[0.0583]
Constant	-40.43***	-39.54***	-24.40***
	[0.958]	[0.959]	[1.167]
Obs	65,125	65,125	65,125
R-sq	0.279	0.281	0.286

Additional controls: Municipality FE, survey year FE and birth order.

Table C1: Monotonicity

Monotonicity test: Out of sample First stage

	Males	Females	Violent	Not violent	Young	Old
Conviction-Judge FE	0.789***	0.194***	0.164***	0.376***	0.334***	0.310***
Out of sample	[0.0520]	[0.0102]	[0.00870]	[0.0208]	[0.0278]	[0.0198]
Obs	20,665	147,066	143,567	75,345	50,267	70,042
Incarceration-Judge FE	0.587***	0.163***	0.0517***	0.189***	0.360***	0.451***
Out of sample	[0.0565]	[0.0148]	[0.0163]	[0.0275]	[0.0237]	[0.0336]
Obs	23,345	104,672	78,652	48,582	75,710	50,387

I compute out of sample judge stringency measures and estimate first stage regressions.

Table C2: Effects on grade retention

Grade retention

Dep var: Expected - Actual grade OLS	(1) Sentence year	(2) + 1 year	(3) + 2 years	(4) + 3 years	(5) + 4 years
Parental incarceration	0.0089	0.0239**	0.0360***	0.0467***	0.0628***
	[0.00576]	[0.00934]	[0.0117]	[0.0135]	[0.0168]
Reduced form					
Parental incarceration	-0.00273	-0.0538	-0.13	-0.0652	-0.128
	[0.0682]	[0.100]	[0.137]	[0.133]	[0.159]
IV					
Parental incarceration	-0.00229	-0.0461	-0.115	-0.0573	-0.115
	[0.0565]	[0.0848]	[0.119]	[0.115]	[0.144]
Obs	29,833	28,399	26,451	24,076	21,078

Positive number of the dependant variable means the child is below the grade he should be given the grade he started with before convicion. Controls: Municipality FE, gender, YOB FE, year of sentence, birth order and year of survey, difference in expected and actual grade one year before sentence, pc and squared pc. Sample: Children between 1990 and 2007 who had a convicted parent between ages 0 and 14. I cluster at the randomization unit and year level.

Table C3: High-School completion

HS completion	OLS	RF	IV
Parental Incarceration	-0.0237***	0.0211	0.0206
	[0.00581]	[0.0591]	[0.0572]
Obs	35,643	35,643	35,643
R squared	0.121	0.12	0.119

Controls: Municipality FE, gender, YOB FE, year of sentence, birth order and year of survey, pc and squared pc. Sample: Children between 1990 and 2000 who had a convicted parent between ages 0 and 17. I cluster at the randomization unit and year level.

Table C4: Sentencing guidelines

Sentencing guidelines	Prison time	
Crime	Colombia	US NY
Possesion of cocaine: 14 grams -100 grams	5 to 9 years	1 to 9 years
Assault		
Simple/third degree	1 to 3 years	Up to 1 year
2nd degree	2 to 7 years	3 to 7 years
Theft		
Simple	2 to 9 years	Up to 1 year
Aggraveted theft	6 to 14 years	2-7 years
Domestic violence	4 to 8 years	Less than a year to 25 years

Source: Colombia articles 376, 112 239, 240 of the penal code, respectively. For New York: 220.16, 120.00, 120.00, 155.25 or 165.40, 155.30 and 120.00 to 120.12 sections of New York penal law code, respectivelty.

Table C5: Random coefficients test

Random coefficients for:	LR test	p-value
Years of education	2.81	0.2452
Income score	1.82	0.4031
Head of the HH Dummy	0.00	0.9999
Single Dummy	5.49	0.0641
Working Dummy	-4.94	0.9999
Male is head of HH Dummy	5.81	0.0548
Sex Dummy	34.22	0.0000

Due to computational constraints I run this mixed effects logistic regression only for judges in Bogota which is the largest district.

Table C6: Placebo check

Placebo test

Dep var: Years of education	OLS	RF	IV
Parental Inc.	-0.0190***		0.0414
	[0.00705]		[0.116]
Judge Leniency		0.0582	
		[0.143]	
Constant	4.138	3.979	4.299
	[4.107]	[4.105]	[4.039]
Obs	46,257	46,257	46,257

Controls: Pc, Municipality FE, gender, YOB FE, year of sentence, birth order and year of survey. Sample: Children who had a convicted parent before their first schooling record. SE in brakets, clustered at the judge level.

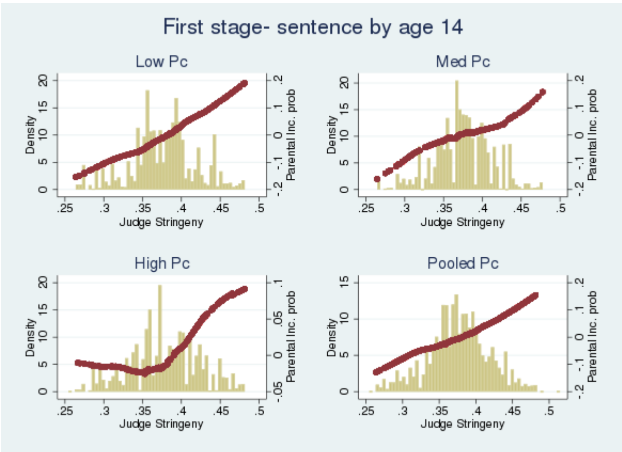
Table C7: LIML and LASSO estimates

IV LIML Dep var: Years of education	(1) 0.74<Pc<0.88	(2) 0.88<Pc<0.9	(3) 0.9<Pc<1	(4) Pooled Pc
Parental Incarceration	**0.996****	**0.626**	**0.930****	**0.786*****
	[0.476]	[0.575]	[0.382]	[0.266]
Obs	16,631	18,454	15,173	50,258
LASSO judges	**0.845*****	**0.105**	**2.053***	**0.692*****
Parental Incarceration	[0.265]	[0.336]	[1.155]	[0.208]
Obs	8,228	8,908	7,159	24,295

Controls: Municipality FE, gender, YOB FE, year of sentence, birth order and year of survey. Column 4 controls add a second order polynomial on Pc. Sample: Children between 1990 and 2007 who had a convicted parent between ages 0 and 14. SE in brackets clustered at the randomization unit and year level.

C Appendix: Extra tables and figures

Figure C1: First stage-children



Notes: Histograms of parental incarceration judge stringency and the fitted value of local polynomial regressions of parental incarceration on judge stringency. I divide the sample by terciles of judge stringency in the conviction stage, and in the pooled regression I control for p_c.

Figure C2: Reduced form

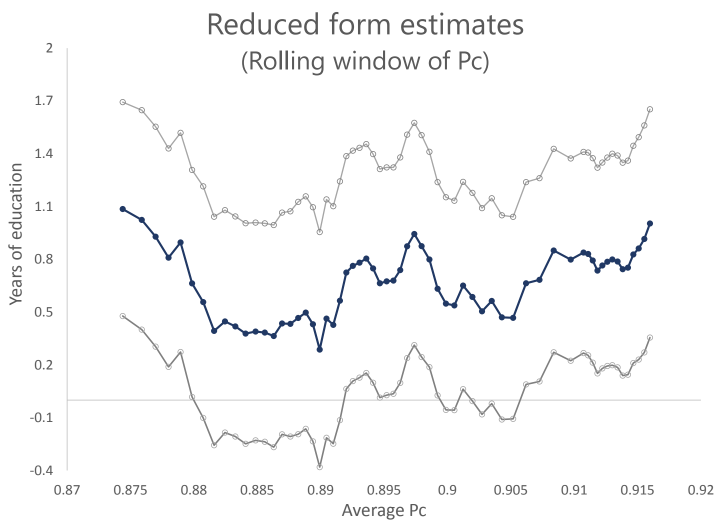


Figure C3: Identification in 2 dimensions

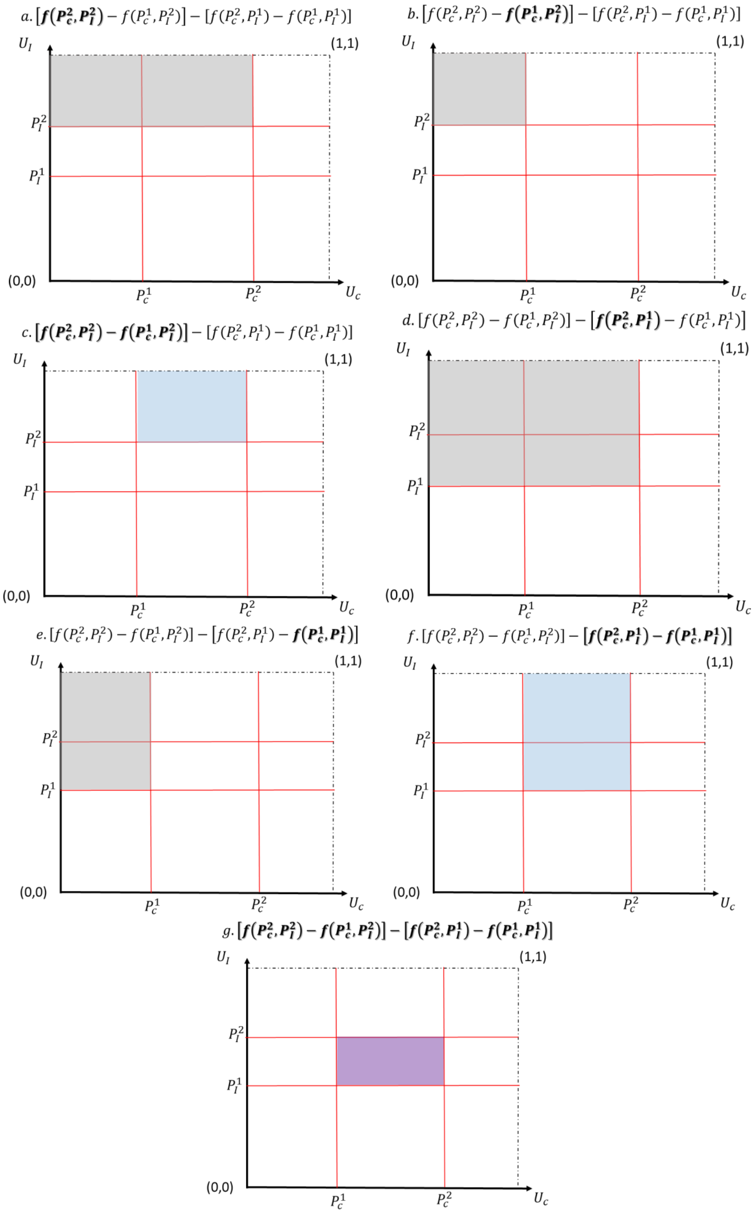


Figure C4: Compliers rectangule

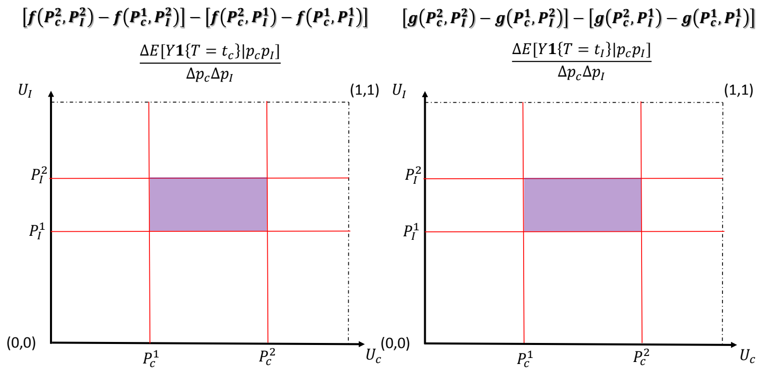


Figure C5: Unconditional MTE