Formal Employment and Organized Crime: Regression Discontinuity Evidence from Colombia

By: Gaurav Khanna
Carlos Medina
Anant Nyshadham
Jorge Tamayo

Borradores de
ECONOMÍA

No. 1054
2018

tá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Col

Formal Employment and Organized Crime: Regression Discontinuity Evidence from Colombia

Gaurav Khanna[*]

Carlos Medina[†]

Anant Nyshadham[‡]

Jorge Tamayo[§]

Abstract

Canonical models of criminal behavior highlight the importance of economic incentives and employment opportunities in determining crime (Becker, 1968). Yet, there is little causal evidence leveraging individual-level variation in support of these claims. Over a decade, we link administrative micro-data on socio-economic measures with the universe of criminal arrests in Medellin. We test whether increasing the relative costs to formal-sector employment led to more crime. We exploit plausibly exogenous variation in employment around a cutoff in the socio-economic score, below which individuals receive health care if they are not formally employed. Using a regression discontinuity design, we show that the policy had the unintended consequence of inducing a fall in formal-sector employment and a corresponding spike in organized criminal activity. There are no effects on less economically motivated crimes like those of impulse or opportunity. Our results confirm the relationship between formal employment and crime, validating models of criminal activity as a rational economic choice.

Keywords: Gangs, informality, Medellin

JEL Codes: K14, K42, J46

[*] University of California – San Diego, gakhanna@ucsd.edu, gauravkhanna.info

[†] Banco de la Republica de Colombia, cmedindu@banrep.gov.co

[‡] Boston College & NBER, nyshadha@bc.edu, anantnyshadham.com

[§] Harvard University, Harvard Business School, jtamayo@hbs.edu, jorge-tamayo.com

We thank seminar participants at UCSD, Michigan (H2D2), Rosario (Bogota), Colombian Central Bank (Bogota), Minnesota (MWIEDC) for feedback, and Achyuta Adhvaryu, Amanda Agan, Prashant Bharadwaj, Chris Blattman, Gordon Dahl, Gordon Hanson, Mauricio Romero, Mauricio Villamizar-Villegas, Juan Vargas and Jeff Weaver for insightful comments.

Empleo Formal y Crimen Organizado: Evidencia de Regresión Discontinua para Colombia♦

Gaurav Khanna*
Carlos Medina†
Anant Nyshadham‡
Jorge Tamayo§

Resumen

Los modelos canónicos del comportamiento criminal resaltan la importancia de los incentivos económicos y las oportunidades de empleo como determinantes del crimen (Becker, 1968). A pesar de esto, existe poca evidencia causal que soporte estos modelos a nivel de individuos. Nosotros pareamos, a lo largo de una década, información socioeconómica de individuos con el censo de todos los arrestos en Medellín. Nosotros probamos si incrementar el costo relativo del empleo formal conlleva a un incremento del crimen. Se explota una variación exógena en el empleo alrededor de un corte en el puntaje socioeconómico, por debajo del cual los individuos reciben aseguramiento en salud si no están formalmente empleados. Utilizando un diseño de regresión discontinua, mostramos que la política tuvo como consecuencia inducir una reducción en el empleo formal y un correspondiente incremento en la actividad del crimen organizado. No se encuentran efectos en crímenes con una motivación económica menor como aquellos de impulso u oportunidad. Nuestros resultados confirman la relación entre empleo formal y crimen, validando los modelos que explican la actividad criminal como una decisión racional.

Palabras clave: Bandas criminales, informalidad, Medellín

Códigos JEL: K14, K42, J46

♦ La serie Borradores de Economía es una publicación de la Subgerencia de Estudios Económicos del Banco de la República. Los trabajos son de carácter provisional, las opiniones y posibles errores son responsabilidad exclusiva de los autores y sus contenidos no comprometen al Banco de la República ni a su Junta Directiva.
* University of California – San Diego, gakhanna@ucsd.edu, gauravkhanna.info
† Banco de la Republica de Colombia, cmedindu@banrep.gov.co
‡ Boston College & NBER, nyshadha@bc.edu, anantnyshadham.com
§ Harvard University, Harvard Business School, jtamayo@hbs.edu, jorge-tamayo.com

1 Introduction

Classic models of criminal behavior suggest that individuals rationally weigh the expected costs and benefits of engaging in criminal activity (Becker, 1968; Ehrlich, 1973). Here, economic incentives play an important role via alternatives to crime: primarily legitimate employment in the labor market. Understanding the economic decision to engage in crime is important, as reducing crime through incapacitation is ineffective when the elasticity of supply to crime is high (Freeman, 1999). We use administrative data to test the importance of formal employment in reducing the likelihood to engage in crime at the individual level. Exploiting a discontinuity in the cost of formal work in Colombia, we are among the first to establish a causal link between formal employment and participation in organized crime leveraging individual-level variation.

Even though early models of criminal activity are based on individual behavior, data constraints often require us to test these models using aggregate area-based relationships (Agan and Makowsky, 2018; Cornwell and Trumbull, 1994; Entorf, 2000; Foley, 2011; Fougere et al., 2009; Gould et al., 2002; Karin, 2005; Lin, 2008; Machin and Meghir, 2004; Raphael and Winter-Ember, 2001). Area-based relationships, while extremely meaningful, measure a somewhat different association, as unemployment at the regional level reduces the returns to criminal activity (e.g., lowers the resources available to expropriate and is correlated with fewer potential victims in the area (Mustard, 2010)). General equilibrium effects in which a new stock of criminals may crowd-out others, and neighborhood and peer effects both within and across neighborhoods might confound evidence of the relationship between area-based employment and choices to engage in crime (Cullen et al., 2006; Dustmann and Damm, 2014; Ihlanfeldt, 2007; Kling et al., 2005, 2007).[1] Additionally, economic activity and high-income individuals leave areas with high or increasing crime (Cullen et al., 2005; Cullen and Levitt, 1999; Greenbaum and Tita, 2004), further confounding the association between crime and employment observed at the aggregate level. Freeman (1999) claims that such factors make area-based relationships between crime and economic activity "fragile, at best."

Studies that do examine individual-level choices often rely on associations conditional on

[1] Fella and Gallipoli (2014) find that general equilibrium effects explain a substantial portion of the relationship between crime and schooling.

observables (Freeman, 1999; Grogger, 1998; Lochner, 2004), as plausibly exogenous variation is challenging to find. Even though most have an extensive set of individual controls (Gronqvist, 2017), it is difficult to account for time-varying unobservable shocks that would determine both employment and crime. Factors like high discount rates determine both crime and job-search (DellaVigna and Paserman, 2005; Golsteyn et al., 2014), whereas childhood shocks and decisions may affect both adult employment and crime (Doyle, 2008, 2007; Lochner and Moretti, 2004). Reverse causality leads to upward bias as employers are less likely to prefer individuals that may display attributes correlated to criminal behavior (Grogger, 1995; Kling, 2006; Lott, 1992). Lastly, many studies depend on self-reported crime in survey data that may have measurement issues, especially since criminal activity is a rare occurrence (Freeman, 1999).

We overcome each of these issues when examining the relationship between formal employment and criminal activity in Medellin, Colombia. First, we link two sources of administrative data at the individual level: the universe of arrests and the pre-arrest socio-economic characteristics of citizens, overcoming measurement issues that plague self-reported criminal activity and aggregate area-based measures of crime. Next, we exploit quasi-experimental variation in the relative cost of formal-sector employment (or relative benefits to informal employment) derived from a generous healthcare program that requires individuals to be outside the formal sector to be eligible. Rather than associations conditional on observables, we use plausibly exogenous variation to isolate the relationship between employment and crime. Last, our data allow us to distinguish between different types of criminal activity and conduct falsification tests by comparing the impacts on crimes most likely associated with organized criminal enterprises (i.e., gangs) to the impacts on other, more idiosyncratic crimes of impulse and opportunity.

The Colombian government provides subsidized healthcare to all residents that reside within a household that has a socio-economic score (known as the *Sisben* score) below a certain threshold. Formal employment of any member affects the family's eligibility for this extremely generous program, raising the relative benefits to other forms of employment.[2] Using a regression discontinuity design, we find that the policy induced a 3 to 5.4 percentage point lower formal

[2] Eligibility is determined at the family level, with the employment status and incomes of children under the age of 25 living at home also determining eligibility. Accordingly, parents have reason to discourage their children from joining the formal sector to avoid losing access to the subsidized regime.

employment rate at the margin, consistent with estimates from previous studies.[3] This lower formal employment is a combination of fewer youth joining the formal workforce over time, and some dropping out of it.[4]

These same individuals are more likely to be arrested for crimes associated with organized criminal activity. At the RD cutoff we find a roughly 0.45 percentage point rise in gang-related violent crimes, a roughly 0.66 percentage point rise in gang-related property crimes, and a less precisely estimated 0.1 percentage point rise in gang-related drug crimes.[5] Importantly, offenses that are less likely to be associated with organized crime, like rape and the consumption of narcotics, do not show a meaningful increase at the cutoff, allowing us to rule out many alternative theories.[6] At the margin, the generosity of the program raised the opportunity cost of being employed in the formal sector. High-crime environments like Medellin, have an informal market that contains significant opportunities to be employed by organized criminal enterprises (i.e., gangs). Indeed, additional results show that impacts on gang-related criminal arrests are strongest in neighborhoods known to have the highest gang intensity at baseline.

Our contributions lie in validating economic models of criminal behavior (Becker, 1968; Ehrlich, 1973) by providing empirical evidence on the relationship between formal employment and crime that is causal, based on individual-level variation, and leveraging rich administrative data. Such evidence has proven difficult to find in this literature. Our study adds to a small, recent literature presenting empirical evidence of a causal relationship between employment and crime. Dell et al. (2018); Dix-Carneiro et al. (2018) use variation from trade-shocks to present area-based evidence of the relationship between local economic factors and criminal activity. To the best of our knowledge, only two recent papers leverage individual level identifying variation. Blattman and Annan (2015) study how the randomized rehabilitation and work-training of high-risk ex-fighters leads to more legitimate employment and less illicit activity in

[3]When evaluating the effect on the entire country using a different research design, Camacho et al. (2014) find that the program led to a 4 percentage point decrease in formal employment, consistent with the point estimate we obtain using the optimal bandwidth.

[4]Despite the reduction in formal employment, reported incomes are not significantly different at the cutoff suggesting a replacement with informal sources of economic activity.

[5]Additional results in which joint outcomes of non-formal employment *and* criminal arrests are studied confirm that those leaving formal work and those arrested are the same. The results for gang-related drug crimes are significant when we simultaneously measure both non-formal employment and arrests as an outcome.

[6]For instance, insurance may induce risky behavior. Yet, that should increase non-gang arrests as well.

war-stricken Liberia. Pinotti (2017) shows that immigration legalization in Italy led to a drastic reduction in crime, consistent with our hypothesis: formal sector work leads to less crime. Yet, the Italian context is not strictly a test of choice in occupation as without legalization work opportunities for migrants were limited.

Our paper complements these recent studies by leveraging administrative data for the whole population of low-income households and the universe of criminal arrests. In addition, we add to existing evidence by stressing the importance of distinguishing between different types of crime, as some are more likely to be associated with organized criminal enterprises (e.g., homicide and motor vehicle theft) whereas others are more likely to be crimes of impulse, addiction or opportunity (e.g., rape and drug consumption). In doing so, we establish meaningful falsification tests for our mechanisms, and rule out alternative mechanisms that have little to do with occupational choice. Our results indicate that formalizing the workforce can lead to reductions in organized criminal activity. Our magnitudes are similar to the related literature (Pinotti, 2017), as we measure an economically meaningful 21% increase in property crime and 32% increase in violent crime.

Additionally, there are few such studies in the developing world, as many look at the US, the UK or Scandinavian countries (Bhuller et al., 2018; Dustmann and Damm, 2014; Freeman, 1999). In contrast, we study what was at the time of our data one of the most violent cities in the world and a hotbed of organized crime, which has been shown to have particularly detrimental effects on growth and development (Alesina et al., 2017; Pinotti, 2015). More than one in five young men, in our sample, were arrested, and we find that our effects are strongest in neighborhoods that have more opportunities at baseline for joining organized crime.[7] Finally, we highlight an unintended, adverse consequence of a generous policy that contributes to work on the interaction between public sector interventions and crime (Doyle, 2008; Yang, 2008).[8]

[7]Our high crime context may be similar to some parts of the developing world. In the developed world, however, only the US has similar incarceration rates (Kearney et al., 2014).

[8]Related work studies how elected officials may engage in criminal activity (Ferraz and Finan, 2008, 2011; Olken and Pande, 2012), and how multiple prices for public programs lead to distortions (Barnwal, 2018).

2 Background

2.1 Crime in Medellin

Located in the north-western region of Colombia, Medellin is the second largest city after the capital, Bogota. It has strong industrial and financial sectors with approximately 2.3 million people or 5.5% of the Colombian population. The urban zone consists of 249 neighborhoods, divided into 21 (*comunas*), 5 of which are semi-rural townships (*corregimientos*).

Although, Colombian violence has traditionally been high, the emergence of drug cartels in the late 1970s and early 1980s, fueled the emergence of organized crime to support illegal businesses, and guerrilla or paramilitary groups to care for the entire production chain. From the mid 1980s to early 1990s, homicide rates rose rapidly driven by the boom of cartels, paramilitaries, and local gangs. In 2009, Medellin was among the 10 most violent cities of the world (CCSPJP, 2009), placing our analysis among a handful that study motivations behind joining organized crime in high-crime environments. The high homicide rates are a result of fights among urban militias, local gangs, drug cartels, criminal bands, and paramilitaries based in surrounding areas.[9] Many demobilized militias continue to be involved in crimes like extortion and trafficking, given their experience with using guns and avoiding police (Rozema, 2018).

There are two features of the homicide rate that are pertinent for our analysis. First, it is predominantly male. In 2002, the first year of our data, the male homicide rate was 184 per 100,000 whereas the female homicide rate was about 12, less than one-tenth the rate of males. Over the entire sample period (2005-13), 12% of all males (across all age groups) were at some point arrested, while the arrest rate for females was only 1%. Second, youth, between 13 and 26 years, are far more likely to be involved as victims or assailants than other age groups. Approximately, 63% of first arrests are between 13 and 26. Younger individuals are more likely to be engaged in drug trafficking and consumption, whereas slightly older individuals are involved in violent crimes (homicides, extortions, and kidnapping), and the oldest still are involved in property crime. Irrespective of type of crime, however, arrest rates peak within the 13 to 26 age window depicted in Figure A1.

[9] *Operacion Orion*, followed by the demobilization of paramilitary forces led to a sharp decline in homicides, as the military clamped down on urban militias (Medina and Tamayo, 2011).

Anthropological studies and in person interviews show that economic incentives drive young men in Medellin to join organized crime (Baird, 2011). As many respondents highlight, the reason to join crime is mostly "economic" or for a profitable career.[10] Knowing this, paramilitaries and gangs actively recruit idle youth that are *amurrao* (local slang, literally: 'sitting on the wall') and without a formal sector job. An interview with El Mono (*p191*) underlines the recruitment process: "*those guys would hang out around here and be nice to me and say 'come over here, have a bit of money'.*" Having a formal sector job means that one is not "hanging around the neighborhood" when the gangs come recruiting. A desirable outside option would be a job with benefits and social security, yet those with formal sector jobs pay extortion fees to gangs.[11] Indeed, the options are often presented as an occupational choice: "*are you gonna work [for the gang] or do a normal job?*"[12].

Often, however, remunerations for gang-members are higher than jobs for those with similar levels of education (Doyle, 2016). New recruits are employed to run guns (*carritos*), before transitioning to extortion and trafficking. These anecdotes are consistent with our hypothesis: higher costs of formal sector jobs (or better benefits for informal work) discourage youth from joining the formal sector, which in turn leads them to be recruited by gangs.[13]

We restrict our analysis to data on first arrests. Repeat arrests are excluded as time spent under incarceration and the length of sentencing may be endogenous to other characteristics.[14] Indeed, first arrests most closely map to the first decision node between legal and illegal activities. Once captured a criminal career begins, with subsequent decisions to repeat, escalate, or exit the criminal sector based on many factors we do not observe (including prison sentences). Accordingly, subsequent criminal behavior is outside the scope of this study.

For similar reasons, we follow recent studies (Gronqvist, 2017; Kling et al., 2005) in focusing on young men in our analysis. Our primary sample is between 21 and 26 in the last year of our arrest data, or between 13 and 26 for the entire period of study, capturing more than 63%

[10]See *interview with Gato, p264* and *interview with Armando, p197.*

[11]See *interview with El Peludo, p184.*

[12] See *interview with Notes, p193*

[13]During the demobilization of militias in the mid-2000s, many were encouraged to join the formal sector, given identity cards and medical cards (Rozema, 2018). Yet, this disparity in costs across healthcare regimes, discourages formal sector re-integration.

[14]Our results are robust to including repeat arrests.

of first arrests (as shown in Figure A1). Of the individuals arrested more than once during the observation period, 40% are first arrested before the age of 27. At the same time, while incarcerated, individuals would not be able to be arrested for additional crimes and would, therefore, have lower measured propensities to be engaged in new criminal activity. Older individuals may have been arrested in their youth (or currently still be incarcerated) but as our crime data only begins in the early 2000s, we do not have their entire criminal history, and would miss their youth arrest. As such, we exclude older men. Focusing on ages when arrest rates peak reduces these concerns regarding the measurement of criminality, and allow us to emphasize the period when young men first make choices between crime and other jobs in Medellin (Doyle, 2016).

For our sample of young men in the bandwidth of analysis, 21.5% were arrested over the period of study – 11.1% for drug crimes, 5.6% for property crime, and 4.8% for violent crimes. These numbers are high relative to most contexts. Yet, the US has an incarceration rate more than six times the typical OECD nation, where one in ten youths from a low-income family may join a gang, 60% of crimes are committed by offenders under the age of 30, and 72% by males (Kearney et al., 2014). In some regards, our context is similar to not only high-crime regions in many parts of the developing world, but also the US.

2.2 Access to Health Benefits

Prior to 1993, only workers affiliated with the Colombian Institute of Social Insurance were beneficiaries of privately provided health insurance, while uninsured individuals were treated by a network of public hospitals. In 1993, *Law 100* established two tiers of health insurance: the Contributive Regime (CR) and the Subsidized Regime (SR). The CR covers formal workers with a comprehensive set of health services that includes nearly all of the most common illnesses. The SR covers the families of the poorest informal workers and unemployed with a plan that initially covered fewer illnesses than CR, but has since been broadened.[15] Formal workers and their employers fund workers' insurance premiums for coverage by the CR. Between the 1993 reform and 1998, health insurance coverage under both grew from 20% to 60%. In 2005, SR was

[15] In 2008, the Constitutional Court ordered that the basket of health services covered under SR become equal to that of the CR. However, the reform did not come into effect until July 2012

expanded and takeup reached 1.1 million people in Medellin alone. By 2013, 96% of Colombians were covered, with more than half qualifying under SR (Lamprea and Garcia, 2016).

Colombian employers are required by law to enroll all their employees in a Health Promoting Company, which gives them access to health insurance under the CR. Self-employed workers are allowed to enroll in the CR themselves by paying a monthly fixed amount based on a percentage of the monthly minimum wage. Unemployed or inactive individuals (and informal workers) can either get health insurance as the self-employed do through the CR, or apply for access to the SR. Individuals not covered by the CR or the SR, use public hospitals, and are charged fees for both medicines and services.

There are substantial differences across plans in terms of the subsidies received and the payments that affiliates have to pay. Formal sector workers pay 4% of their monthly wage for enrollment in the CR, while the employer pays the other 8.5%.[16] This implies that effectively employees may bear a burden somewhere between 4 and 12.5% of their monthly wage depending on their bargaining power. Formal workers pay 1.5% of their salary to cover informal workers in SR.[17] Over and above this, formal workers have to pay 4% of their wage for their pensions, and also bear other non-wage labor costs like old-age and disability insurance. These costs rose by between 10.5% and 11.5% after the 1993 reforms, with strong evidence that such costs discourage formal sector employment (Kugler and Kugler, 2009). Families enrolled in the CR have to pay co-payments for a variety of medical services and prorated fees. All members in a family eligible for SR, regardless of their relationship to the family head, do not have to pay for prorated fees as they all have free access to a package of services and medications.

To target the SR, roughly 70 percent of the poorest households in the country were interviewed between 1994 and 2003, and a welfare index (*Sisben* score) was calculated using a confidential formula based on respondent characteristics, including incomes and assets, health, education, and housing. Only households with a *Sisben* score below a certain cutoff were eligible to become beneficiaries of the SR. In addition, any household that was formally employed could not become a beneficiary of the SR. Other public programs use the *Sisben* score, but to

[16] Employers' contribution was 8% between 1993 (Law 100) and 2007. On that date it was increased to 8.5% (Law 1122). This contribution was eliminated in 2012 (Law 1607) for incomes up to 10 times minimum wages.

[17] Authorities initially expected the formal sector population to rise and cover costs for SR. But the SR grew faster than the CR population, in part due to the lucrative nature of the SR (Lamprea and Garcia, 2016).

the best of our knowledge, the SR *Sisben* cutoff did not coincide with other major interventions, at the eligibility cutoff of Sisben 1 in the early 2000s.[18] The SR health program is by far the largest and most generous program that has eligibility determined by the *Sisben* score.[19]

That this policy led to a fall in formal-sector employment has been documented in both the academic literature and public discourse. The Minister of Social Protection, in a news article in *Presidencia de la Republica* (February, 2006), claimed that the people's valuation of SR was so high that it discouraged formal employment. Studying the effects on the entire country, Camacho et al. (2014) use individual-level data and control for both region and time fixed effects to show that informal employment increased by 4 percentage points as SR was rolled out across the country. This is a combination of workers dropping out of the formal sector, but also fewer youth joining the formal sector over time (Lamprea and Garcia, 2016). Recognizing these adverse effects on formal employment, the government drastically lowered the costs of being enrolled in CR right at the end of our study period, when Law 1607 was enacted. This led to a significant increase in formal sector employment (Bernal et al., 2017; Fernández and Villar, 2017; Kugler et al., 2017; Morales and Medina, 2017).

Since the *Sisben* score and targeting is at the family level rather than individual level, older family members have reason to discourage youth within the family from joining the formal labor force for fear of losing insurance for the entire household.[20] Large families stay informal in the hope of retaining benefits (Joumard and Londono, 2013).[21] Indeed, Santamaria et al. (2008) find that half of all SR recipients indicated that they would not switch to formal employment as it would mean losing benefits. These effects are not restricted to men, as women's formal-sector participation also decreased in response to SR (Gaviria et al., 2007). Yet, we find that dis-employment effects on men are about four times larger than on women, consistent with the hypothesis that men have a lucrative alternative outside the formal sector: organized crime.

[18]See www.sisben.gov.co/Paginas/Noticias/Puntos-de-corte.aspx for programs by *Sisben 3* cutoff. While the *Sisben* cutoff for SR enrollment may differ across counties, there is only one cutoff for the entirety of Medellin.

[19]The share of the SR in the total budget accounts to nearly 2% of the GDP, while all programs sought to reduce poverty represent less than 0.4% of GDP.

[20]By Article 21, Decree 2353 of 205, the *Sisben* score is determined at the family level. The economic activity of any dependent under the age of 25 will contribute to the *Sisben* score and eligibility for SR, making formal employment a joint family decision.

[21]Similarly interviews in Baird (2011) highlight how being involved in crime can sometimes be a 'family decision' (chapter 6).

3 Data

We combine two sources of data at the individual level using national identification numbers and dates of birth. One source is from successive *Sisben* surveys of the Medellin population for three different years: 2002 (baseline *Sisben I*), 2005 (*Sisben II*) and 2009-2010 (*Sisben III*). The *Sisben* dataset consists of cross sections from censuses of the poor.[22] To create a panel data set, we match household records across the three waves. The second source is the census of individuals arrested between 2002 and 2013 for different types of crimes, whether or not they were convicted, from the Judicial Police Sectional of the National Police Department. 91% of the arrested individuals were apprehended in the act.

Figure 1: Timeline of Data Used

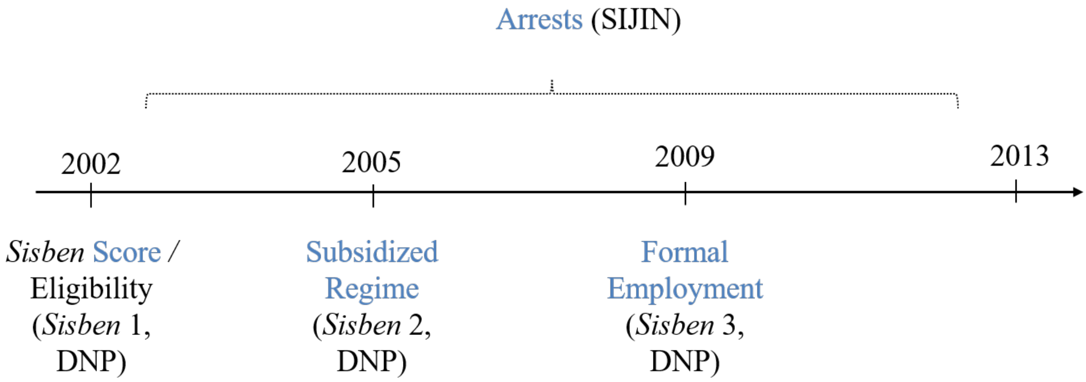


Figure 1 describes the timeline of our data. We use the 2002 *Sisben* as our baseline to create our running variable and predict eligibility for SR.[23] We test for SR enrollment in the 2005 *Sisben*, and for employment status and incomes in the 2009 *Sisben*. We then follow the criminal histories of young men aged 21 to 26 in 2013, between 2005 (after we have a measure of SR enrollment from the second *Sisben*) and 2013. Table 1 presents the 2002 baseline summary statistics of the complete *Sisben* survey and for the subsample of males only.

The arrests data include a detailed description of the person arrested (national identification number and date of birth), the type of crime (e.g., homicide, rape, motor vehicle theft, etc.), the precise article associated with the crime in the penal code, the date of arrest, the location

[22] Municipalities survey a census of people living in the three poorest socioeconomic strata. In low income municipalities, the survey is a census of the whole population, while in larger cities it amounts to 65-80% of the population.

[23] Note that the formula to compute the Sisben score and the eligibility cutoff varies across the Sisben surveys (I, II, and III).

Table 1: Summary Statistics in 2002 (Police data and *Sisben* I Survey)

Variable	*Complete Sample* Mean	*Complete Sample* Std. Dev.	*Males* Mean	*Males* Std. Dev.
Male	0.490	0.500	1.000	0.000
Subsidized Regime	0.319	0.466	0.312	0.463
Contributive Regime	0.228	0.420	0.222	0.416
Age 10-15	0.105	0.306	0.109	0.311
Age 15-20	0.105	0.306	0.110	0.313
Age 20-25	0.089	0.285	0.093	0.290
Age 25-30	0.068	0.251	0.068	0.251
Ever Arrested	0.062	0.242	0.114	0.318
Household Head (HH) Characteristics				
HH-Female	0.387	0.487	0.308	0.462
HH-Employed	0.628	0.483	0.643	0.479
HH-Unemployed	0.106	0.308	0.107	0.309
HH-Married	0.345	0.475	0.377	0.485
HH-Attending School	0.009	0.097	0.008	0.089
HH-Has CR	0.207	0.405	0.207	0.405
HH-Age	43.237	14.302	43.869	14.159
HH-Years of Education	4.542	2.451	4.480	2.454
HH-Owns House	0.314	0.464	0.327	0.469
HH-Sisben Stratum1	0.271	0.444	0.273	0.446
HH-Sisben Stratrum 2	0.620	0.485	0.620	0.485
HH-Sisben score	45.707	9.901	45.716	9.908
Number of members	4.090	1.709	4.215	1.709
N	1,161,446		568,923	

of arrest, and a police generated flag for whether the arresting officer suspected the perpetrator to be gang affiliated. We classify the crimes into three categories – violent, property, and drug crimes – based on the US Bureau of Justice Statistics' classifications in the Sourcebook of Criminal Justice Statistics (BJS, 1994). We further divide our crimes into gang-related and non gang-related based on the prevalence of police-generated flags for gang affiliation among arrests for each type of crime.[24] For instance, this allows us to classify homicides as violent gang-related, and rape or domestic violence as violent non gang-related. In robustness checks, we also use a method that relies on the association between these crimes and high-gang neighborhoods. In this alternative definition, we classify those crimes as gang-related if they are disproportionately

[24]The gang flags exists for about 30% of the data. We classify crimes based on the data that has flag information, and do the final analysis for the entire data.

more likely to list any of these high-gang neighborhoods as a location of arrest.

In Appendix Table A1, we categorize the 25 (of 103) most prevalent crimes under each classification method. These data-driven methods line up with our priors on types of crime: homicides, motor vehicle theft, extortion, kidnapping, break-ins, and the manufacturing, delivery and trafficking of drugs fall under organized crimes. The remaining crimes are often thought of as crimes of impulse or opportunity, like rape, simple assault, and drug consumption.

4 Enrollment in the Subsidized Regime (SR)

As only households in the two lowest levels of *Sisben* I (2002), a score below 47, could qualify for the SR, we compare households on either side of the cutoff to identify the effect of SR eligibility. First, we verify if there is a discontinuity in the probability of SR enrollment at the cutoff. Second, we examine how the likelihood of being in the formal sector changes at the cutoff. Last, we examine the effect on different types of criminal activity.

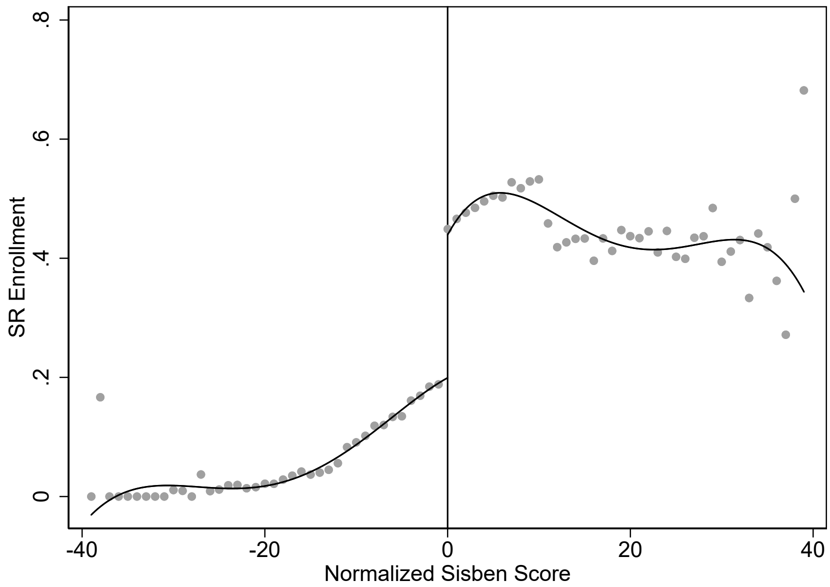
In following RD conventions, we normalize the *Sisben* score so that treated units are individuals with positive values of our new score. Figure 2 presents the first stage: the discontinuity in the probability of SR enrollment using the optimal binning procedure found in Calonico et al. (2014a). The probability of enrollment discontinuously increases by around 26 percentage points.[25] Not all eligible persons enroll in SR, as formal sector jobs may be valuable to some, but enrollment still jumps substantially to 42% at the cutoff.

For two-staged least squares (2SLS) exercises we follow a fuzzy regression discontinuity design, where our running variable is the 2002 *Sisben* score. We use both parametric and non-parametric approaches to estimate the effect of SR eligibility at the cutoff. For the parametric approach we follow Hahn et al. (2001), where we instrument enrollment in the SR with the eligibility indicator $1\left[s_i < 47\right]$, and estimate the following equation as our first stage:

$$SR_{i,n} = \alpha + \alpha_1 1\left[s_{i,n} < 47\right] + X'_{i,n}\alpha_2 + A_i\left(s_{i,n}\right)\alpha_3 + \mu_n + \varepsilon_{i,n} \ ,$$

[25] Around 20% of households that have a high 2002 *Sisben* also avail of SR in 2005, as a fraction of households became eligible under a smaller 1998 *Sisben* survey, and the government allows them to keep their benefits for some time after they graduate out of eligibility.

Figure 2: Discontinuity in the Probability of SR enrollment at Cutoff.



SR enrollment is probability of being enrolled in the subsidized regime in 2005. RD Graph using optimal binning procedure discussed in Calonico et al. (2014a). Normalized Sisben (2002) score on horizontal axis centered around cutoff. Higher values represent low scores (higher poverty).

where A_i is a vector of smooth polynomial functions of the *Sisben* score of each individual, $s_{i,n}$. In robustness checks, we also estimate models conditioning on demographics and whether or not the individual is enrolled in CR. Here $X_{i,n}$ is a vector of demographic characteristics for individual i living in neighborhood n, including an indicator that is equal to one if individual i is enrolled in CR in 2005. μ_n corresponds to neighborhood fixed effects for the 249 neighborhoods.

An important issue in practice is the selection of the smoothing parameter. We use local regressions to estimate the discontinuity in outcomes at the cutoff point. In particular, we estimate local polynomial regressions conducted with a rectangular kernel and employing the optimal data-driven procedure suggested by Calonico et al. (2014b). We use two different optimal bandwidth procedures: the Imbens and Kalyanaraman (2012) method and the Calonico et al. (2014b) bandwidth. The optimal bandwidths from the different procedures lie between 5.5 and 6.2 points, on the 100-point *Sisben* I scale. We present our results for multiple bandwidths to highlight the robust nature of our estimates, varying them from below the optimal bandwidths to larger bandwidths. Specifically, we check for coefficient stability for results spanning these bandwidths ranging between 4 and 10 points around the cutoff. Varying the size of the bandwidth and the polynomial order do not affect the results presented in this section.

Table 2: SR Enrollment at Sisben Cutoff (First Stage)

Variables	Bandwidths:	4	6	10
	Dependent Variable: Enrolled in SR (First Stage)			
Below Sisben Cutoff		0.260***	0.260***	0.269***
		(0.0138)	(0.0132)	(0.0110)
F-stat of IV		354.97	387.97	598.02
Number of observations		181,132	246,974	340,581
Sample mean (in bandwidth)				0.36

Note: Standard errors in parentheses. *** significant at 1%; ** significant at 5%; * significant at 10%. Coefficient of indicator of being below *Sisben* cutoff, with linear controls for 2002 *Sisben* scores that vary flexibly at the cutoff. SR enrollment as measured in the 2005 *Sisben* survey. Standard errors clustered at the *comuna* level.

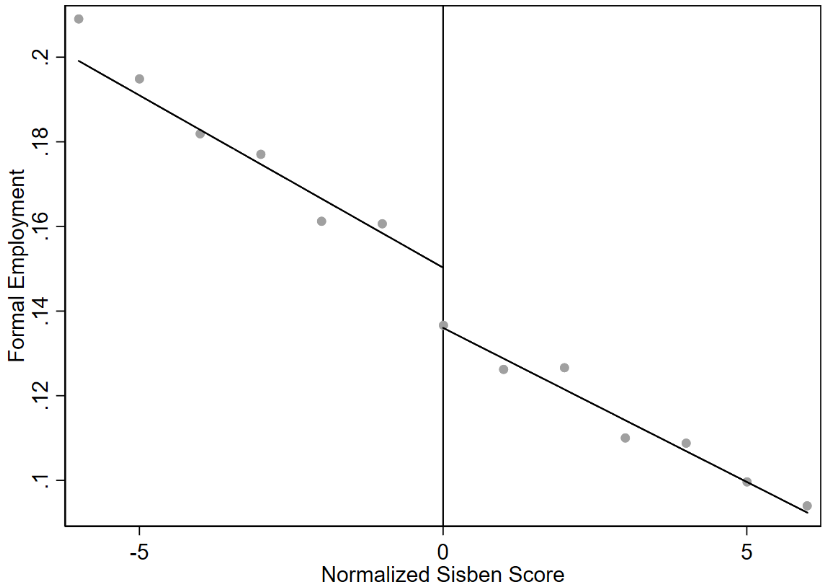
Our first stage results are shown in Table 2, displaying the 26 percentage point increase in SR enrollment shown in Figure 2. As we vary the bandwidths from 4 through 10 the coefficient is stable and both economically and statistically significant. The table also shows that the standard IV F-test suggests a strong instrument, and for our remaining outcomes we conduct two-staged least squares analyses using this is as our first stage.

5 Impacts on Formal Employment and Reported Income

We test the simple hypothesis that the SR conditions disincentivized formal-sector employment and led to an increase in organized-crime activities. We first reproduce a well-established result and show that the program has a negative effect on employment (Camacho et al., 2014; Gaviria et al., 2007; Joumard and Londono, 2013; Santamaria et al., 2008). We exploit the discontinuity in enrollment rates at the cutoff, by using the eligibility indicator as an instrument for enrollment status to identify the effect of SR on formal employment and income. Here $Emp_{i,n}$ is 1 if the individual i from neighborhood n was formally employed and $S\hat{R}_{i,n}$ is the predicted SR enrollment probability from the first stage. In robustness checks we include demographic controls and an indicator for CR enrollment in $X_{i,n}$, and neighborhood fixed effects μ_n.

$$Emp_{i,n} = \beta_0 + \beta_1 S\hat{R}_{i,n} + X'_{i,n}\beta_2 + A_i\left(s_{i,n}\right)\beta_3 + \mu_n + \varepsilon_{i,n} \ ,$$

Figure 3: Discontinuity in Formal Employment (2009).



RD Graph using optimal binning procedure discussed in Calonico et al. (2014a). Formal employment based on measures in 2009 *Sisben* survey. Subsample of males. Normalized Sisben (2002) score on horizontal axis centered around cutoff. Higher values represent low scores (higher poverty).

Figure 3 captures the fall in formal sector employment at the cutoff, where formal employment is defined as a working individual making wage contributions to benefits as measured in the 2009 *Sisben* III survey.[26] In our RD figures we focus on a bandwidth of 6 around the cutoff as it is the Calonico et al. (2014b) optimal bandwidth.

Table 3 presents the results for reported formal employment and incomes in the 2009 *Sisben* survey. The table presents results for the reduced form change at the cutoff, and the two-staged least squares (2SLS) effect of enrolling in SR. These results show that the health insurance program had a negative impact of 4.1 percentage points (when using the optimal bandwidth) on the probability of being employed in the formal sector in 2009.

Lower formal sector employment at the cutoff may be a combination of fewer youth joining the formal sector as they enter working-age, lower transition rates out of informal work, and higher transition probabilities out of formal work at the cutoff. As formal sector employment affects SR enrollment for the entire family, these are often family decisions, where older family members may discourage youth from joining the formal sector (Joumard and Londono, 2013).

[26]While this is a somewhat conservative measure of formal employment, Colombian employees who pay contributions to health insurance have been widely considered by the literature to be formal employees ((See Attanasio et al., 2017; Morales and Medina, 2017). The Sisben does not explicitly ask households whether the worker is in the formal sector, which in any case, the response would be misreported underestimating the formality rate.

Table 3: Reported Formal Employment and Income

Bandwidths:	4	6	10
Panel A: Formal Employment in 2009 (Males)			
Above Cutoff	-0.0147***	-0.0111***	-0.00845***
Reduced Form	(0.00467)	(0.00280)	(0.00217)
Enrolled in SR	-0.0539***	-0.0411***	-0.0301***
2SLS	(0.0166)	(0.0103)	(0.00811)
Number of observations	133,067	180,742	247,886
Sample mean (only males in bandwidth for 2009)			0.14
Panel B: Annual Household Income in 2009 (USD)			
Above Cutoff	-3.837	-3.805	2.347
Reduced Form	(3.100)	(2.295)	(4.008)
Enrolled in SR	-6.481	-3.042	30.49
2SLS	(9.163)	(8.842)	(27.83)
Number of observations	46,797	63,457	87,510
Sample mean (households in bandwidth for 2009)			171.24

Note: Standard errors in parentheses. *** significant at 1%; ** significant at 5%; * significant at 10%. We use the *Sisben* survey of 2009 to construct both outcome variables. Formal employment for males only. The results for women are presented in Table A2. Tables report Two-Staged Least Squares (2SLS) coefficients where the first stage is SR enrollment on being below the Sisben cutoff. Regressions control linearly for the Sisben score, flexibly around the cutoff. We cluster standard errors by *comuna*. Household-level income reported in *pesos* and converted to USD using the average 2009 exchange rate. Sample means for males and households only in bandwidth for 2009.

This effect is larger for men than it is for women (Appendix Table A2), plausibly due to outside options for males in organized crime.

The impact on household-level income is statistically indistinguishable from zero and economically small ($30 per household annually). One caveat is that income is self-reported. In general, respondents may under-report assets and incomes in order to get a lower *Sisben* score. However, as respondents do not know the score formula, perfect manipulation is impossible, and therefore, as we show below, unsurprisingly the density of respondents is smooth around the cutoff. We may expect that incomes from illicit activities are under-reported rather than over-reported, further suggesting that poverty-based desperation is less likely to be driving criminal activity. If anything, the income results suggest that even as workers drop out of the formal sector they find other sources of income. Yet, we wish to be careful in stressing the crudeness of self-reported income measures.

6 Impacts on Crime

Once more, in the second stage for crime outcomes, we use the eligibility indicator as an instrument for enrollment status to identify the effect of SR enrollment on crime. Here $crime_i$ is 1 if the individual i was arrested between 2005 and 2013.

$$Crime_{i,n} = \beta_0 + \beta_1 S\hat{R}_{i,n} + X'_{i,n}\beta_2 + A_i\left(s_{i,n}\right)\beta_3 + \mu_n + \varepsilon_{i,n} \ ,$$

Our main results do not condition on other factors. In robustness checks, we control for various characteristics of the household head in 2002, the baseline year. These controls include an indicator for female-headed households, employment status, years of education, marital status, attendance to any academic institute, year-of-birth fixed effects, socioeconomic strata of the household,[27] home ownership, and neighborhood fixed effects. A literature on neighborhood effects and crime (Cullen et al., 2006; Dustmann and Damm, 2014) highlight the perils of using area-based relationships (like differences in unemployment rates) to study individual-level occupational choice, and re-iterates the strength of our approach.[28] Unsurprisingly, our results are unaffected by the inclusion of neighborhood fixed effects that absorb any neighborhood level characteristics (demographics, amenities, property values and police presence) that may affect crime rates. We cluster standard errors at the *comuna* level.

We present results for violent, property, and drug-related crimes, dividing each group between organized-crime related activities and crimes less likely to be associated with organized criminal entities. As discussed, at the point of every arrest, the police issue a flag if they suspect the arrested individual is involved with a gang or not. We calculate the propensity for being issued this flag for each type of crime, and divide crimes into two groups: high and low-propensity to be organized criminal activity. This data-driven method to groups crimes

[27]Urban areas in Colombia are split into six socioeconomic strata, used by authorities to spatially target social spending to neighborhoods.

[28]There may still be general equilibrium effects of the policy that affect the entire country, but since our variation is not driven by differences across neighborhoods, this is all netted out.

produce meaningful classifications (Table A1).

We hypothesize that organized criminal activities are directly related to our implicit model of occupational choice across legitimate and illegitimate sectors, whereas non-gang-related crimes should be less affected by the opportunity cost of being in the formal sector and hence serve as a useful falsification test. We expect the effects on the latter group to be zero, as crimes of impulse and passion are less directly related to occupational choice.

As we elaborate in a later discussion, over and above a falsification test, the lack of effects on non-gang crimes also allow us to rule out alternative mechanisms. We do not classify crimes based on whether or not they are pecuniary as that captures crimes of desperation and necessity that arise out of poverty. Instead, we posit that the policy induced an *occupational choice* to join a gang, and as such use organized crime as a basis for classification. Alternative mechanisms (such as riskier behavior when having insurance) may have weight if non-gang crimes rose as well, but the lack of effects on non-gang crimes allow us to rule them out.

6.1 Violent Crime

We first start with the probability of engaging in violent criminal activities. Based on the police flags for gang-related activity, organized crimes include homicides, extortion, and kidnapping. Crimes less likely to be associated with an organized entity include domestic violence, rape and injuries. Figure 4 and Table 4 present the results.

Figure 4 shows the jump in violent gang-related crime arrests at the *Sisben* cutoff, concentrating on an optimal bandwidth of 6 points on the 100 point scale. In Table 4 we present the regression discontinuity results varying the bandwidth and specifications. The reduced form results (first row in each panel) show an increase in gang-related violent crime (Panel A), but no corresponding change in less gang-related violent crime (Panel B). Within a bandwidth of 10 points on the *Sisben* scale, these results amount to a 32% increase (or a 0.45 percentage point increase) in violent crime from the mean around the cutoff. While economically meaningful, these magnitudes are similar to work from other contexts (Pinotti, 2017).

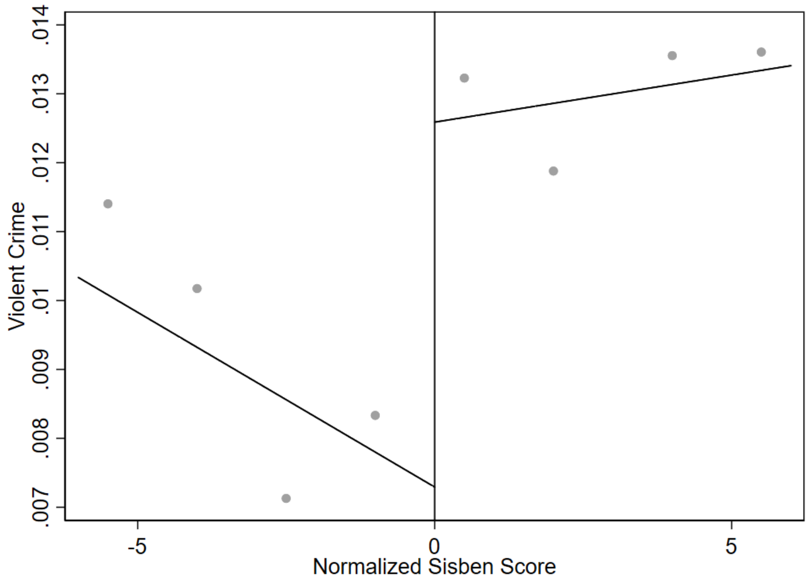
Our 2SLS results (the next two rows of each panel) show an economically and statistically significant increase in the probability of organized-crime related violent arrests for individuals

Table 4: Violent Crimes

Bandwidths:	4	6	10
Panel A: More Gang-Related Violent Crimes			
Above Cutoff	0.00722***	0.00649**	0.00456**
Reduced Form	(0.00236)	(0.00249)	(0.00164)
Enrolled in SR	0.0257***	0.0231***	0.0158***
No Covariates	(0.00873)	(0.00838)	(0.00539)
Enrolled in SR	0.0274***	0.0232**	0.0149**
Including pre-treatment covariates	(0.00950)	(0.00937)	(0.00583)
Number of observations	18,052	24,272	33,027
Sample mean (men 13-26 years old in bandwidth)			0.014
Sample mean for those enrolled in SR and in high-gang comuna			0.020
Panel B: Less Gang-Related Violent Crimes			
Above Cutoff	0.00279	0.000988	-0.000581
Reduced Form	(0.00454)	(0.00304)	(0.00326)
Enrolled in SR	0.00994	0.00349	-0.00201
No Covariates	(0.0158)	(0.0104)	(0.0110)
Enrolled in SR	0.00791	0.00118	-0.00322
Including pre-treatment covariates	(0.0168)	(0.0111)	(0.0125)
Number of observations	18,419	24,768	33,702
Sample mean (men 13-26 years old in bandwidth)			0.034
Sample mean for those enrolled in SR and in high-gang comuna			0.039

Note: Standard errors in parentheses. *** significant at 1%; ** significant at 5%; * significant at 10%. Tables report Two-Staged Least Squares (2SLS) coefficients where the first stage is SR enrollment on being below the Sisben cutoff, where the Sisben score is measured in 2002. We measure crime between 2005 and 2013. Regressions control linearly for the Sisben score, flexibly around the cutoff. We cluster standard errors by *comuna*. We consider only males between 21 to 26 years old in 2013. For regressions that have pre-treatment covariates, we include household characteristics, year of birth fixed effects, neighborhood fixed effects, and an indicator for if the individuals were enrolled in the CR in 2005.

Figure 4: Gang-Related Violent Crimes



RD Graph using optimal binning procedure discussed in Calonico et al. (2014a). Normalized Sisben (2002) score on horizontal axis centered around cutoff. Higher values represent low scores (higher poverty).

enrolled in SR. We don't find any meaningful effect on the probability to be apprehended for non organized-crime related violence. A comparison of the various rows in each panel shows that the estimates are robust to the inclusion of controls, whereas a comparison across columns shows the robustness to the choice of bandwidth.

6.2 Property Crime

In Figure 5 and Table 5 we analyze the effects on property crimes. Based on police flags, we establish that gang-related property crimes include crimes like motor vehicle theft and break-ins to businesses and residences. Crimes like fraud and identify theft are classified as less gang-related. Once again, in the reduced form we see that gang-related property crimes increase, with little change to less gang-related property crimes. This increase constitutes a 21% increase (or a 0.66 percentage point increase) from the mean around the cutoff within a bandwidth of 10 points.

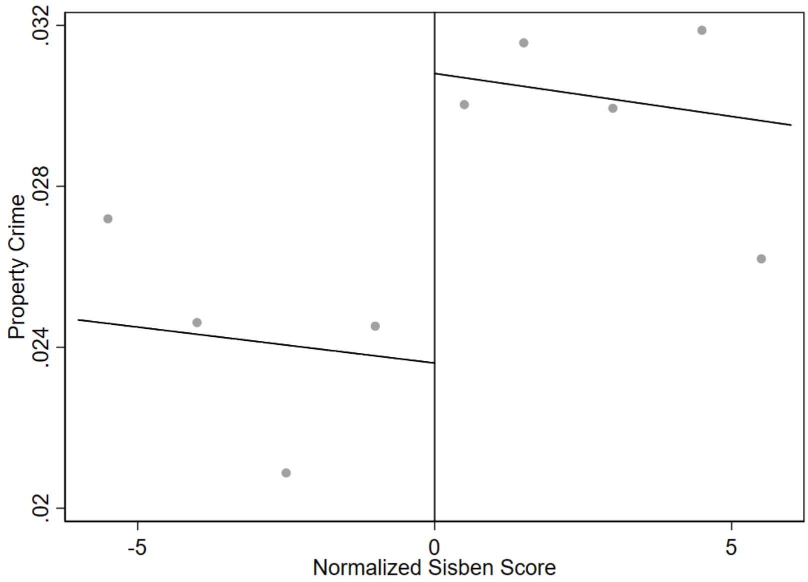
In our 2SLS results we find an economically and statistically significant increase for gang related property crime arrests, and no strong effect for property crimes less associated with organized entities. Once again, our estimates are quite robust to the inclusion of control variables and the choice of the bandwidth, and our magnitudes are economically meaningful.

Table 5: Property Crimes

Bandwidths:	4	6	10
Panel A: More Gang-Related Property Crimes			
Above Cutoff	0.0106**	0.00930**	0.00666*
Reduced Form	(0.00387)	(0.00389)	(0.00350)
Enrolled in SR	0.0380***	0.0331***	0.0232**
No Covariates	(0.0123)	(0.0126)	(0.0113)
Enrolled in SR	0.0408***	0.0341***	0.0240**
Including pre-treatment covariates	(0.0139)	(0.0131)	(0.0108)
Number of observations	18,426	24,740	33,625
Sample mean (men 13-26 years old in bandwidth)			0.032
Sample mean for those enrolled in SR and in high-gang comuna			0.040
Panel B: Less Gang-Related Property Crimes			
Above Cutoff	-0.00263	-0.00217	-0.00205
Reduced Form	(0.00554)	(0.00425)	(0.00336)
Enrolled in SR	-0.00941	-0.00772	-0.00712
No Covariates	(0.0194)	(0.0149)	(0.0112)
Enrolled in SR	-0.0116	-0.00872	-0.00854
Including pre-treatment covariates	(0.0212)	(0.0156)	(0.0119)
Number of observations	18,240	24,523	33,358
Sample mean (men 13-26 years old in bandwidth)			0.024
Sample mean for those enrolled in SR and in high-gang comuna			0.028

Note: Standard errors in parentheses. *** significant at 1%; ** significant at 5%; * significant at 10%. Tables report Two-Staged Least Squares (2SLS) coefficients where the first stage is SR enrollment on being below the Sisben cutoff, where the Sisben score is measured in 2002. We measure crime between 2005 and 2013. Regressions control linearly for the Sisben score, flexibly around the cutoff. We cluster standard errors by *comuna*. We consider only males between 21 to 26 years old in 2013. For regressions that have pre-treatment covariates, we include household characteristics, year of birth fixed effects, neighborhood fixed effects, and an indicator for if the individuals were enrolled in the CR in 2005.

Figure 5: Gang-Related Property Crimes



RD Graph using optimal binning procedure discussed in Calonico et al. (2014a). Normalized Sisben (2002) score on horizontal axis centered around cutoff. Higher values represent low scores (higher poverty).

6.3 Drug Crime

Our last type of crime involves the drug trade in Medellin. We analyze the impact on the probability to engage in drug-related crimes in Figure 6 and Table 6. Organized-crime related drug arrests include the manufacturing, distribution, and trafficking of hard drugs like cocaine and heroin. Drug crimes less likely to be related to organized entities include possession and consumption of drugs, as these are mostly indicative of personal recreational use, along with marijuana-related crimes.

In Figure 6, even as the discontinuity in drug crime arrests is visible, the binned averages suggest a somewhat imprecise relationship. In Table 6 the direction of effects are what we may expect, but our results are not precisely estimated.[29] One possibility for the lack of precision is in the measurement error associated with the classification of such crimes: the difficulty in classifying possession of drugs as consumption or trafficking likely introduces noise. Indeed, offenses related to the trafficking of marijuana are problematic as small amounts of personal possession were made legal during this period. While homicides, assaults and theft produce clear evidence of crimes, encouraging an arrest, drug crimes are often difficult to detect and

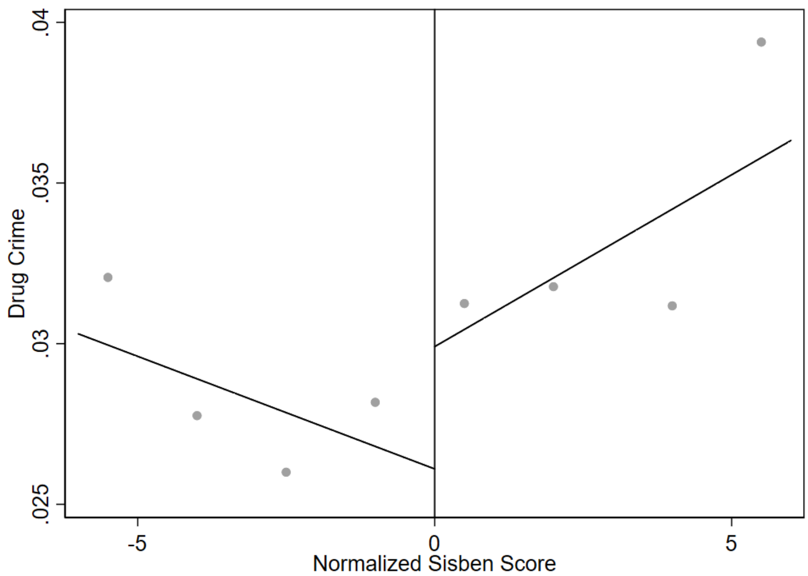
[29] In Appendix Table A7 we explore an alternative specification where we look at arrests conditional on not being in the 2009 formal sector. Here we have enough precision to measure an uptick in drug crimes.

Table 6: Drug Crimes

Bandwidths:	4	6	10
Panel A: More Gang-Related Drug Crimes			
Above Cutoff	0.00799	0.00348	0.00133
Reduced Form	(0.00721)	(0.00492)	(0.00458)
Enrolled in SR	0.0285	0.0124	0.00461
No Covariates	(0.0240)	(0.0169)	(0.0155)
Enrolled in SR	0.0303	0.0135	0.00524
Including pre-treatment covariates	(0.0270)	(0.0180)	(0.0159)
Number of observations	18,463	24,857	33,851
Sample mean (men 13-26 years old in bandwidth)			0.038
Sample mean for those enrolled in SR and in high-gang comuna			0.045
Panel B: Less Gang-Related Drug Crimes			
Above Cutoff	-0.00976	-0.0129	-0.00788
Reduced Form	(0.00774)	(0.00798)	(0.00629)
Enrolled in SR	-0.0348	-0.0458	-0.0274
No Covariates	(0.0280)	(0.0293)	(0.0218)
Enrolled in SR	-0.0385	-0.0501	-0.0277
Including pre-treatment covariates	(0.0299)	(0.0329)	(0.0230)
Number of observations	19,150	25,740	35,104
Sample mean (men 13-26 years old in bandwidth)			0.073
Sample mean for those enrolled in SR and in high-gang comuna			0.088

Note: Standard errors in parentheses. *** significant at 1%; ** significant at 5%; * significant at 10%. TTables report Two-Staged Least Squares (2SLS) coefficients where the first stage is SR enrollment on being below the Sisben cutoff, where the Sisben score is measured in 2002. We measure crime between 2005 and 2013. Regressions control linearly for the Sisben score, flexibly around the cutoff. We cluster standard errors by *comuna*. We consider only males between 21 to 26 years old in 2013. For regressions that have pre-treatment covariates, we include household characteristics, year of birth fixed effects, neighborhood fixed effects, and an indicator for if the individuals were enrolled in the CR in 2005.

Figure 6: Gang-Related Drug Crimes



RD Graph using optimal binning procedure discussed in Calonico et al. (2014a). Normalized Sisben (2002) score on horizontal axis centered around cutoff. Higher values represent low scores (higher poverty).

record.

In sum, our results indicate that the drop in formal employment as a result of the generous benefits for informal workers granted by the SR raised the likelihood of being arrested for gang-related violent and property crimes. The magnitudes of the estimated impacts are also economically meaningful. The pattern of results is similar but imprecise for drug crimes. Importantly, the results also show that non gang-related crimes of each type are not impacted by SR enrollment. In the following section, we investigate whether impacts are strongest in *comunas* that were historically associated with high organized crime activity as further evidence in support of our occupational choice interpretation.

6.4 Heterogeneity by Comuna

Figure A2 shows the spatial distribution of the locations where criminals were arrested in the act between 2005 and 2013, by type of crime. The red circle specifies the downtown of the city. The spatial patterns suggest that there are neighborhood clusters. In our main results we already show specifications that include neighborhood fixed effects, and we cluster errors at spatial levels larger than neighborhoods.[30]

[30] Our results are robust to clustering at smaller spatial levels, like the neighborhood.

The opportunities in a neighborhood affect how easy it is to induce youth into crime (Kling et al., 2005). Here we study the heterogeneity in impacts across neighborhoods by baseline propensities for organized crime. We investigate if *comunas* with a high incidence of gangs demonstrate stronger impacts on gang-related arrests, at the RD cutoff. If the policy induces men to join organized crime, then we may expect that neighborhoods that have more such opportunities would have a larger impact.

Table 7: Heterogeneity by Comuna

Bandwidths:	4	6	10
Panel A: More Gang-Related Violent Crimes			
Enrolled in SR	0.0267***	0.0211**	0.0150***
	(0.00892)	(0.00914)	(0.00538)
Enrolled* Gang Comuna	-0.00152	0.0141***	0.00563
	(0.00464)	(0.00376)	(0.00537)
Panel B: More Gang-Related Property Crimes			
Enrolled in SR	0.0344**	0.0273**	0.0190*
	(0.0134)	(0.0137)	(0.0115)
Enrolled* Gang Comuna	0.0282	0.0364**	0.0258**
	(0.0209)	(0.0167)	(0.0116)
Panel C: More Gang-Related Drug Crimes			
Enrolled in SR	0.0282	0.0131	0.00296
	(0.0248)	(0.0177)	(0.0166)
Enrolled* Gang Comuna	0.000310	-0.00590	0.00690
	(0.0136)	(0.0135)	(0.0129)
Number of observations	18,463	24,857	33,851

Note: Standard errors in parentheses. *** significant at 1%; ** significant at 5%; * significant at 10%. Tables report Two-Staged Least Squares (2SLS) coefficients where the first stage is SR enrollment on being below the Sisben cutoff and an interaction between high-gang comunas and being below the cutoff. The Sisben score is measure in 2002, whereas crime outcomes are measured between 2005 and 2013. Regressions include comuna fixed effects and an interaction between high-gang comunas and indicators for SR enrollment. Regressions control linearly for the Sisben score, flexibly around the cutoff. We consider only males between 21 to 26 years old in 2013. We cluster errors by comuna.

We select the top five *comunas* with the highest number of gang members captured by the police, and create an indicator variable for whether individuals lived in these *comunas* in 2002, our baseline year.[31] We interact this variable with the cutoff to analyze the heterogeneity in effects by area-level gang activity. Table 7 presents the results. The effects on crime are present

[31]The top five are chosen on the criteria of having the most gang-flags as a ratio of total crimes.

in both high and low gang-activity areas, but are larger in areas that have more gang activity, especially for property crimes. This suggests that opportunities present in the neighborhood affect the likelihood of inducement into organized crime at the cutoff.

7 Specification Tests and Robustness

7.1 Density Tests and Balance Tests

Identification relies on the assumption that all other determinants of the outcome vary smoothly at the cutoff. We show that an extensive set of observables display no systematic patterns in discontinuities. In Table A3, we show that baseline characteristics from 2002 (three years before our crime data begins) are balanced. We consider two sets of baseline characteristics: one for household-level socioeconomic variables, and the other for individuals. We report estimates over the range of different bandwidths used in our main analysis: between 4 and 10 points around the cutoff. We find no evidence of systematic discontinuities in covariates at the threshold. In the first row of Table A3, we report a summary measure in which we collapse these variables by taking their first principal component and repeat the same RD analysis that we do for our main results. Once again, there is no detectable difference in this composite measure of baseline characteristics at the cutoff, even for the largest bandwidth of 10.

Additionally, for the empirical strategy to be valid, households must not be able to manipulate their score to cross the cutoff. Work by Camacho and Conover (2011) highlights politically motivated manipulation in certain municipalities in other parts of Colombia where elections were being held. This includes both under-reporting of wealth (not necessarily a threat to our design), but also manipulation of the final score. In other parts of the country the 1998 mayoral elections (years before our sample begins) show evidence of manipulation. We use the raw survey data and the 2002 *Sisben* score only for Medellin, and so are less concerned about any manipulation of the final score.[32] Importantly, our tests of balance in the large set of baseline characteristics of the household are indicative of a lack of systematic manipulation in this context.

[32] The timing of the *Sisben* surveys do not coincide with the Medellin mayoral elections.

Nevertheless, we can test whether there was a discontinuity in the density of scores at the cutoff for the particular context of Medellin after 2002. We do this by following two methods used in the literature: the McCrary (2008) test and a test recently developed by Cattaneo et al. (2017). The Cattaneo et al. (2017) test yields a conventional t-statistic of 0.0489 or a p-value of 0.961, and a robust bias corrected p-value of 0.940, confirming that there is no statistically detectable evidence of manipulation.

Figure A3 shows the distribution of the *Sisben* score for males (non-criminals and criminals) and conducts a McCrary (2008) test. Note that the distribution appears to be smooth with no evidence of bunching before the cutoff between *Sisben* levels 2 and 3 (red line).

7.2 Alternative Crime Classifications, Different Bandwidths, Specifications and Subsamples, and the Nonparametric RD

We conduct a number of robustness checks. First, we re-classify crimes into gang-related and non gang-related groups based on the *location* where these crimes are more likely to occur. We calculate the relative propensity of each crime in each neighborhood. The crimes that have a higher propensity to take place in neighborhoods that were traditionally associated with organized crime are classified as gang-related crimes. These are neighborhoods that also have the highest proportion of gang-related flags associated with them. This 'Neighborhood Classification Method' of crimes produces a list similar to the one in which we use the police generated flags, with minor differences.[33] The lists of the most prevalent crimes by classification method can be found in Appendix Table A1. In Table A4, we re-examine our main results using the alternative classification for gang-related crimes. These results are similar to the main results, with the added statistical significance of drug crimes under some specifications.

Next, we re-examine our main results using the bias-correction methods suggested by Calonico et al. (2014a). In Table A5, we show results that conduct a polynomial bias correction at a larger bias-correction bandwidth (reported in the table). Once again, our results show an economically and statistically significant increase in gang-related violent and property

[33]The big differences among the top 25 crimes is with 'conspiracy to commit murder' that is classified as gang-related under the original classification but not gang-related under the neighborhood definition.

crimes, but the effects on drug crimes are smaller and less precise.

In our main specifications, when looking at a specific type of crime, we exclude arrests from other crimes.[34] In the specification shown in Table A6 we include the other categories along with the non-criminals, and show that are results are similarly robust.

As our story is about both non-formal employment and violent crime, we present a specification in Table A7 that simultaneously captures both. However, since we do not have annual data on formal employment, we use the 2009 *Sisben* to measure formal employment. Here the dependent variable is an indicator that equals one when the individual was not formally employed and arrested for a crime. Our results again show an increase in gang-related criminal activity, with even the drug crimes now being economically and statistically significant. This result allows us to address any concerns that the increase in informality and increase in arrests were independent of each other.

Finally, in Figure A4 we vary the bandwidth through a much wider range – every integer between 2 and 10. Gang-relates violent and property crimes consistently display a positive RD coefficient, whereas drug crimes are not statistically indistinguishable from zero, even as the coefficients are positive and fairly large for smaller bandwidths.

8 Discussion and Alternative Mechanisms

We first reproduce a well-established result: that the design of Colombia's generous health benefits program induced a reduction in formal sector employment (Camacho et al., 2014; Gaviria et al., 2007; Joumard and Londono, 2013; Santamaria et al., 2008). The lower formal sector employment at the cutoff is perhaps an amalgamation of fewer youth joining the formal sector when reaching working age and some even dropping out of the formal sector. This result is almost entirely driven by men rather than women (Table A2), consistent with our hypothesis that organized crime is a lucrative alternative option to formal work for men. Additionally, we find that our increase in arrests is most predominant in neighborhoods that traditionally had greater options for organized crime, consistent with previous results in the literature (Kling et al., 2005).

[34] For instance, when studying violent crimes, we exclude property and drug crimes from the sample altogether.

The simultaneous decrease in formal sector employment and rise in arrests related to organized crime supports a model where a rising opportunity cost of being formally employed induces an occupational choice into a life of crime (Becker, 1968).[35] Given the lack of effects for non gang-related crimes, it is difficult to find alternative models that fit these results.

There are three alternative theories we consider. First, better health benefits at the cutoff may induce one to do riskier tasks, yet it is difficult to support why these riskier tasks do not include non gang-related crimes. Second, formal workers vesting more into the health system may fear losing their jobs if arrested and reduce criminal activities; yet again, this should likely be true for non gang-related crimes. Third, the police may falsely target informal workers even if they are not criminals, but it is unlikely that they would be booked disproportionately under gang crimes. The distinction between gang and non-gang crimes powerfully helps exclude other alternative mechanisms and lends credence to the occupational choice story we posit.

Our magnitudes suggest a 21% rise in property crime and a 32% rise in violent crime. As these are similar to other causal analysis leveraging individual-level variation (Pinotti, 2017), we believe our results to be economically meaningful. Indeed, using the 2009 *Sisben* measures, we show in Table A7 that there was an increase in individuals that were simultaneously both not in the formal sector and arrested.[36]

Yet, these magnitudes should be understood to be context-specific. We study a high-crime environment, similar only to other developing countries and the US. Furthermore, we estimate a Local Average Treatment Effect (LATE) on marginal workers in the neighborhood of an income cutoff. It is plausible that at higher income levels, healthcare is a less important fraction of expenditures, and is less likely to induce such behavior. Finally, it should be noted that the newly induced marginal criminals may be unlike the average criminal along many dimensions, including ease of avoiding arrests, and as such our results may not be widely generalizable for other sub-populations. Nevertheless, since the exogenous probability of getting caught conditional on committing a crime has no reason to be discontinuous at the cutoff, our estimates are unbiased even in the presence of such heterogeneity in criminal "skill."

[35] While we do not discuss in detail specific pathways, anthropological evidence lends credence to active recruitment by gang members of young men that 'hang around' in neighborhoods with idle time, and are not in the formal sector (Baird, 2011).

[36] This result is statistically and economically significant for all gang crimes, including drug crimes.

9 Conclusion

In this paper, we highlight an important fact: disincentivizing formal employment can lead to substantial increases in criminal activity when informal opportunities include employment by organized criminal entities. We evaluate this claim in the context of the high-crime environment of Medellin, Colombia. We first provide strong evidence showing that the criteria behind the generous health benefits policy led to a sharp decrease in formal sector employment. At the margin, the generosity of the policy raised the opportunity cost of formal-sector employment and induced workers to join the informal labor market.

In Medellin, this informal market contains significant opportunities related to organized crime. We follow these same individuals over a decade and show that this decrease in formal sector employment led to an increase in the probability of being arrested for organized-crime related activities. On the other hand, crimes less likely to be associated with criminal economic enterprises, like crimes of impulse or opportunity, show no such impacts at the eligibility threshold, lending credence to the occupational choice mechanism we advance. Together, our simple calculations suggest that as the policy pushed workers out of the formal sector, roughly one-third of these workers were pushed into organized crime. These effects were largest in neighborhoods that had, at baseline, greater opportunities to join organized crime.

Crime deterrence may have limited benefits if the supply elasticity to criminal activity is high (Freeman, 1999). Investigating the decisions behind choosing a life of crime, as we do here, is essential in the fight against crime. Importantly, our work speaks to the determinants of engaging in criminal activity at the individual level, rather than at the aggregate level. The strength of our approach is that we do not need to use area-based variation to identify the relationship between individual employment opportunities and crime. We do this using a unique data set that matches the census of arrests with socio-economic outcomes over a decade, in the context of one of the most violent cities in the world. We find a reliable source of plausibly exogenous variation generated by policy rules, and use a regression discontinuity design to estimate our effects.

We conclude that Colombia's well-intentioned, generous, and broad-based healthcare program had the unintended consequence of amplifying gang activity by way of its distortionary provision rules.[37] The program being popular and important for providing access to healthcare for low income families implies that there is little reason to do away with it. Yet, the rules governing the selection into the program may be distortionary, and as such warrant further examination.

Removing the emphasis on informality (but still maintaining the cutoff) may negate the increase in criminal activity around the cutoff. The costs underlying such a change would be a larger fiscal burden as even low-income formal sector workers would be eligible for SR. The benefits, on the other hand, are far reaching: less crime, less policing and incarceration, and less negative externalities on families and children. This has important welfare implications for the design of many such programs across the developing world which often have far-reaching and under-studied consequences on seemingly unrelated outcomes and behaviors. Our results provide guidance for how impactful improving access to and incentives for formal sector employment can be for deterring criminal activity.

[37] Recognizing some of these adverse effects, policy-makers lowered the costs of CR enrollment at the end of our study period, when Law 1607 was enacted, leading to a significant increase in formal sector employment (Bernal et al., 2017; Kugler et al., 2017; Morales and Medina, 2017).

References

Agan, A. and M. Makowsky (2018). The Minimum Wage, EITC, and Criminal Recidivism. *Working paper*.

Alesina, A., S. Piccolo, and P. Pinottie (2017). Organized Crime, Violence, and Politics. *NBER Working Paper*.

Attanasio, O., A. Guarín, C. Medina, and C. Meghir (2017). Vocational Training for Disadvantaged Youth in Colombia: a Long-Term Follow-up. *American Economic Journal: Applied Economics 9*(2), 131–43.

Baird, A. (2011). Negotiating Pathways to Manhood: Violence Reproduction in Medellin's Periphery. *Thesis, University of Bradford*.

Barnwal, P. (2018). Curbing Leakage in Public Programs: Evidence from India's Direct Benefit Transfer Policy. *Mimeo*.

Becker, G. S. (1968). Crime and Punishment: An Economic Approach. *Journal of Political Economy 76*(2), 169–217.

Bernal, R., M. Eslava, M. Melendez, and A. Pinzon (2017). Switching from Payroll Taxes to Corporate Income Taxes: Firms' Employment and Wages after the Colombian 2012 Tax Reform. *Inter-American Development Bank No. 1268*.

Bhuller, M., G. Dahl, K. Loken, and M. Mogstad (2018). Incarceration, Recidivism and Employment. *NBER Working Paper No. 22648*.

BJS, U. B. o. J. S. (1994). Sourcebook of Criminal Justice Statistics. *University at Albany*.

Blattman, C. and J. Annan (2015). Can Employment Reduce Lawlessness and Rebellion? A Field Experiment with High-Risk Men in a Fragile State. *American Political Science Review 10*(1), 1–17.

Calonico, S., M. Cattaneo, and R. Titiunik (2014a). Robust Data-Driven Inference in the Regression Discontinuity Design. *Stata Journal 14*(4), 909–946.

Calonico, S., M. Cattaneo, and R. Titiunik (2014b). Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica 82*(6), 2295–2326.

Camacho, A. and E. Conover (2011). Manipulation of Social Program Eligibility. *American Economic Journal: Economic Policy 3*(2), 41–65.

Camacho, A., E. Conover, and A. Hoyos (2014). Effects of Colombia's Social Protection System on Workers' Choice between Formal and Informal Employment. *World Bank Economic Review 28*(3), 446–466.

Cattaneo, M., M. Jansson, and X. Ma (2017). Simple Local Regression Distribution Estimators with an Application to Manipulation Testing. *Mimeo Michigan.*

CCSPJP (2009). Consejo Ciudadano para la Seguridad Publica y la Justicia Penal.

Cornwell, C. and W. N. Trumbull (1994). Estimating the Economic Model of Crime with Panel Data. *Review of Economics and Statistics 76*(2), 360–366.

Cullen, J. B., B. A. Jacob, and S. D. Levitt (2005). The Impact of School Choice on Student Outcomes: An Analysis of the Chicago Public Schools. *Journal of Public Economics 89*(6), 729–60.

Cullen, J. B., B. A. Jacob, and S. D. Levitt (2006). The Effect of School Choice on Participants: Evidence from Randomized Lotteries. *Econometrica 74*(5), 1191–1230.

Cullen, J. B. and S. D. Levitt (1999). Crime, Urban Flight, and the Consequences for Cities. *The Review of Economics and Statistics 81*(2), 159–169.

Dell, M., B. Feigenberg, and K. Teshima (2018). The Violent Consequences of Trade-Induced Worker Displacement in Mexico. *American Economic Review: Insights forthcoming*.

DellaVigna, S. and D. M. Paserman (2005). Job Search and Impatience. *Journal of Labor Economics 23*(3), 527–588.

Dix-Carneiro, R., R. Soares, and G. Ulyssea (2018). Economic Shocks and Crime: Evidence from the Brazilian Trade Liberalization. *American Economic Journal: Applied Economics forthcoming*.

Doyle, C. (2016). Explaining Patterns of Urban Violence in Medellin, Colombia. *Laws 5*(3).

Doyle, J. (2008). Child Protection and Adult Crime. *Journal of Political Economy 116*(4), 746–770.

Doyle, J. J. (2007). Child Protection and Child Outcomes: Measuring the Effects of Foster Care. *American Economic Review 97*(5), 1583–1610.

Dustmann, C. and A. P. Damm (2014). Does Growing Up in a High Crime Neighborhood Affect Youth Criminal Behavior? *American Economic Review 104*(6), 1806–1832.

Ehrlich, I. (1973). Participation in Illegitimate Activities: A Theoretical and Empirical Investigation. *Journal of Political Economy 81*(3), 521–565.

Entorf, H. (2000). Socioeconomic and Demographic Factors of Crime in Germany: Evidence from Panel Data of the German States. *International Review of Law and Economics 20*(1), 75–106.

Fella, G. and G. Gallipoli (2014). Education and Crime over the Life Cycle. *The Review of Economic Studies 81*(4), 1484–1517.

Fernández, C. and L. Villar (2017). The Impact of Lowering the Payroll Tax on Informality in Colombia. *Economía 18*(1), 125–155.

Ferraz, C. and F. Finan (2008). Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes. *The Quarterly Journal of Economics 123*(703-745), 2.

Ferraz, C. and F. Finan (2011). Electoral Accountability and Corruption: Evidence from the Audit Reports of Local Governments. *American Economic Review 101*(4), 1274–1311.

Foley, C. F. (2011). Welfare Payments and Crime. *Review of Economics and Statistics 93*(1), 97–112.

Fougere, D., F. Kramarz, and J. Pouget (2009). Youth Unemployment and Crime in France. *Journal of the European Economic Association 7*(5), 909–938.

Freeman, R. B. (1999). The Economics of Crime. *Handbook of Labor Economics 3*(c). edited by O. Ashenfelter and D. Card. Elsevier Science.

Gaviria, A., C. Medina, and C. Mejia (2007). Assessing Health Reform in Colombia: from Theory to Practice. *Economia 7*(1).

Golsteyn, B. H., H. Grnqvist, and L. Lindahl (2014). Adolescent Time Preferences Predict Lifetime Outcomes. *The Economic Journal 124*, F739–F761.

Gould, E. D., B. A. Weinberg, and D. B. Mustard (2002). Crime Rates and Local Labor Market Opportunities in the United States: 1979-1997. *The Review of Economics and Statistics 84*(1), 45–61.

Greenbaum, R. T. and G. E. Tita (2004). The Impact of Violence Surges on Neighborhood Business Activity. *Urban Studies 13*, 2495–2514.

Grogger, J. (1995). The Effect of Arrests on the Employment and Earnings of Young Men. *Quarterly Journal of Economics 110*(1), 51–72.

Grogger, J. (1998). Market Wages and Youth Crime. *Journal of Labor Economics 16*, 756–791.

Gronqvist, H. (2017). Youth Unemployment and Crime: Lessons from Longitudinal Population Records. *Mimeo*.

Hahn, J., P. Todd, and W. van der Klaauw (2001). Identification and Estimation of Treatment Effects with a Regression Discontinuity Design. *Econometrica 69*(1), 201–209.

Ihlanfeldt, K. (2007). Neighborhood Drug Crimes and Young Males Job Accessibility. *The Review of Economics and Statisticsc 89*(1), 151–164.

Imbens, G. W. and K. Kalyanaraman (2012). Optimal Bandwidth Choice for the Regression Discontinuity Estimator. *The Review of Economic Studies 79*(3), 933–959.

Joumard, I. and J. Londono (2013). Income Inequality and Poverty in Colombia: The Role of the Labour Market. *OECD Economics Department Working Papers No. 1036* (1).

Karin, E. (2005). Unemployment and Crime: Is There a Connection? *Scandinavian Journal of Economics 107*(2), 353–373.

Kearney, M. S., B. Harris, E. Jacome, and L. Parker (2014). Ten Economic Facts about Crime and Incarceration in the United States. *The Hamilton Project Policy Memo*. Brookings Institute.

Kling, J. (2006). Incarceration Length, Employment, and Earnings. *American Economic Review*.

Kling, J., J. Ludwig, and L. Katz (2005). Neighborhood Effects on Crime for Female and Male Youth: Evidence from a Randomized Housing Voucher Experiment. *The Quarterly Journal of Economics 120*(1), 87–130.

Kling, J. R., J. B. Liebman, and L. F. Katz (2007). Experimental Analysis of Neighborhood Effects. *Econometrica 75*(1), 83–119.

Kugler, A. and M. Kugler (2009). Labor Market Effects of Payroll Taxes in Developing Countries: Evidence from Colombia. *Economic Development and Cultural Change 57*(2), 335–358.

Kugler, A., M. Kugler, and L. Herrera Prada (2017). Do Payroll Tax Breaks Stimulate Formality? Evidence from Colombia's Reform. *NBER Working Paper No. 23308*.

Lamprea, E. and J. Garcia (2016, December). Closing the Gap Between Formal and Material Health Care Coverage in Colombia. *Health and Human Resources Journal*.

Lin, M.-J. (2008). Does Unemployment Increase Crime? Evidence from U.S. Data 1974-2000. *Journal of Human Resources 43*(2), 413–436.

Lochner, L. (2004). Education, Work, and Crime: A Human Capital Approach. *International Economic Review 45*(3), 811–43.

Lochner, L. and E. Moretti (2004). The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports. *American Economic Review 94*(1).

Lott, J. (1992). An Attempt at Measuring the Total Monetary Penalty from Drug Convictions: the Importance of an Individual's Reputation. *Journal of Legal Studies 29*, 159–187.

Machin, S. and C. Meghir (2004). Crime and Economic Incentives. *Journal of Human Resources 39*(4), 958–979.

McCrary, J. (2008). Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test. *Journal of Econometrics 142*(2), 698–714.

Medina, C. and J. Tamayo (2011). An Assessment of How Urban Crime and Victimization Affects Life Satisfaction. *in Dave, Webb and Eduardo, Wills-Herrera (Eds.), Subjective Well-Being and Security*. Springer, Social Indicators Research Series.

Morales, L. and C. Medina (2017). Assessing the Effect of Payroll Taxes on Formal Employment: The Case of the 2012 Tax Reform in Colombia. *Economia Journal 0*, 75–124. LACEA.

Mustard, D. B. (2010). How Do Labor Markets Affect Crime? New Evidence on an Old Puzzle. *IZA Discussion Paper* (4856).

Olken, B. and R. Pande (2012). Corruption in Developing Countries. *Annual Review of Economics 4*(1), 479–509.

Pinotti, P. (2015). The Economic Costs of Organised Crime: Evidence from Southern Italy. *The Economic Journal 125*, F203–F232.

Pinotti, P. (2017). Clicking on Heaven's Door: The Effect of Immigrant Legalization on Crime. *American Economic Review 107*(1), 138–168.

Raphael, S. and R. Winter-Ember (2001). Identifying the Effect of Unemployment on Crime. *Journal of Law and Economics 44*(1), 259–283.

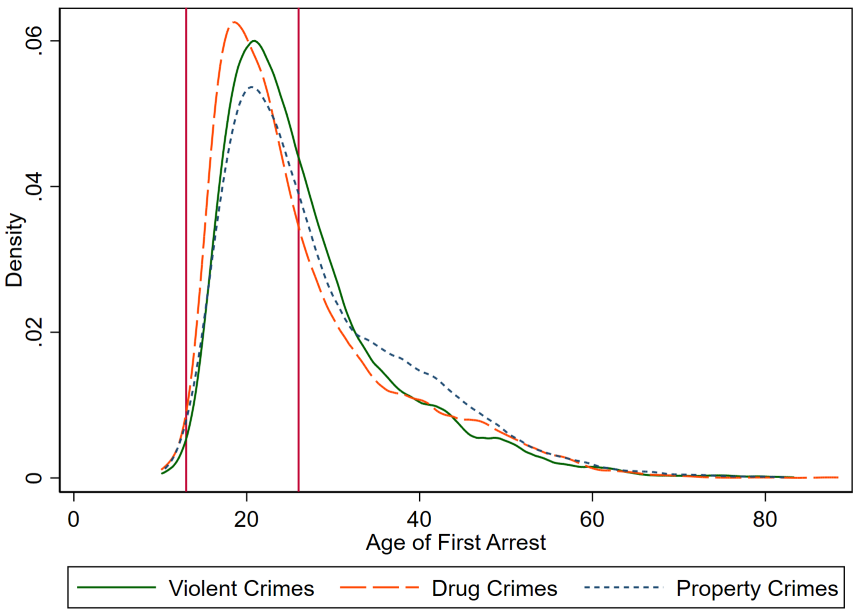
Rozema, R. (2018). Urban DDR-processes: Paramilitaries and Criminal Networks in Medellin, Colombia. *Journal of Latin American Studies 40*, 423–452.

Santamaria, M., F. Garcia, and A. V. Mujica (2008). Los Costos No Laborales y el Mercado Laboral: Impacto de la Reforma de Salud en Colombia. *Working Paper 43. Fedesarrollo, Bogota, Colombia.*

Yang, D. (2008). Can Enforcement Backfire? Crime Displacement in the Context of Customs Reform in the Philippines. *Review of Economics and Statistics 90*(1), 1–14.

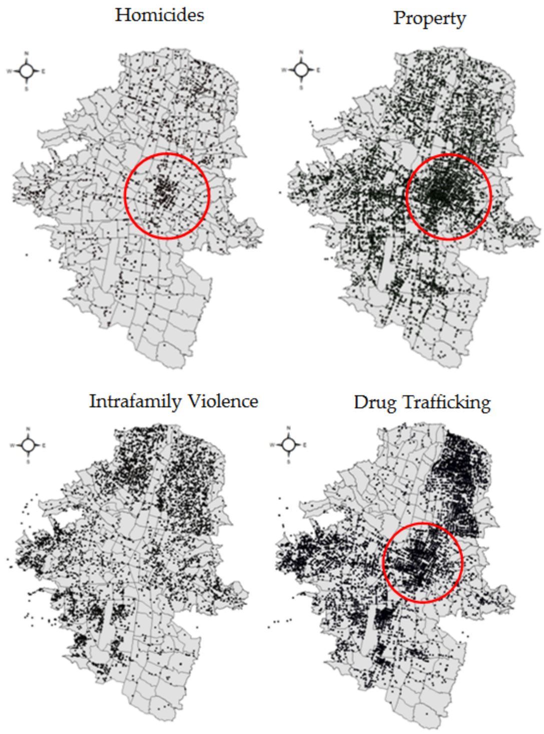
Appendix: Additional Tables and Figures

Figure A1: Distribution of Age at Arrest (Males)



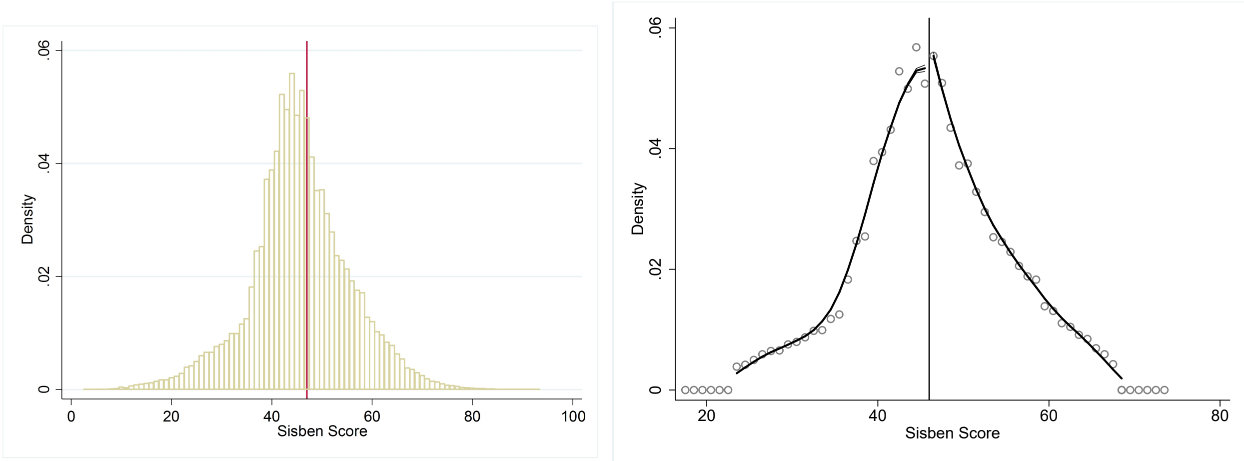
Source: Polica Nacional de Colombia. Vertical lines represent ages 13 and 26.

Figure A2: Location of 'in-the-act' arrests by type of crime, 2005-2013.



Source: Medina and Tamayo (2011) using Polica Nacional de Colombia. Dots indicate arrests. Bold lines are neighborhood boundaries. Red circle is downtown.

Figure A3: Sisben score Distribution (All Males).



Source: *Sisben* survey of 2002. The figure includes all males (i.e. both non-criminals and arrested individuals). The left panel shows the histogram and the right panel conducts a McCrary (2008) test.

Table A1: List of Top 25 Crimes by Data-driven Classifications

Crime	Type	Gang Flags	Neighborhood Method
Drug Consumption / Possession	Drug	No	No
Drug trafficking / Distribution - Marijuana	Drug	No	No
Drug trafficking / Distribution	Drug	Yes	Yes
Drug trafficking / Distribution - Cocaine paste	Drug	Yes	Yes
Drug trafficking / Distribution Heroin	Drug	Yes	Yes
Use of Fake Identification, false document	Property	No	No
Motor vehicle theft (Motorcycles)	Property	No	No
Receiving Bribes (as officials)	Property	No	No
Copyright/Fraud	Property	No	No
Identity Theft	Property	No	No
Fraud	Property	No	No
Theft / Assault	Property	Yes	Yes
Robbery (To Businesses, firms)	Property	Yes	Yes
Property Vandalism	Property	Yes	Yes
Motor Vehicle Theft - Cars	Property	Yes	Yes
Burglary	Property	Yes	Yes
Simple Assault/Battery	Violent	No	No
Rape/Sexual Assault	Violent	No	No
Conspiracy to commit murder	Violent	Yes	No
Homicide	Violent	Yes	Yes
Extortion	Violent	Yes	Yes
Assault / Battery - Against Police	Violent	Yes	Yes
Manufacture, Trafficking Firearms / Weapons	Violent	Yes	Yes
Intimidation and Stalking	Violent	Yes	Yes
Terrorism	Violent	Yes	Yes
Kidnapping	Violent	Yes	Yes

List of top 25 crimes by type and gang classification, out of 103 crimes. The 'Gang Flags' lists whether or not the crime has a high propensity to receive a police reported flag of gang-related at the time of arrest. The 'Neighborhood Method' classifies crimes that have a high propensity to be in neighborhoods that also receive a higher fraction of gang-related flags at the time of arrest.

Table A2: Formal Employment By Gender

Bandwidths:	4	6	10
Panel A: Men Formal Employment in 2009			
Enrolled in SR	-0.0539***	-0.0411***	-0.0301***
	(0.0166)	(0.0103)	(0.00811)
Number of observations	133,067	180,742	247,886
Panel B: Women Formal Employment in 2009			
Enrolled in SR	0.00560	-0.0130*	-0.0169*
	(0.00757)	(0.00786)	(0.00889)
Number of observations	156,942	213,755	292,980

Note: Standard errors in parentheses. *** significant at 1%; ** significant at 5%; * significant at 10%. We use the *Sisben* survey of 2009 to construct formal employment. Tables report Two-Staged Least Squares (2SLS) coefficients where the first stage is SR enrollment on being below the Sisben cutoff. Regressions control linearly for the 2002 Sisben score, flexibly around the cutoff. We cluster standard errors by *comuna*.

Table A3: Baseline (2002) balance tests

Bandwidths:	4	6	10
First Principal Component	-0.115	0.110	0.0796
	(0.118)	(0.0917)	(0.0740)
Years of Education	0.0183	0.00248	-0.0408
	(0.0685)	(0.0606)	(0.0453)
Age	0.0339	0.0255	0.0177
	(0.0469)	(0.0300)	(0.0239)
Age Specific Education Gap	0.00566	0.0159	0.0580
	(0.0613)	(0.0602)	(0.0435)
HH Head Years of Education	-0.0463	-0.0832	-0.105**
	(0.0652)	(0.0592)	(0.0451)
Unemployed	0.0166*	0.0113	0.00971
	(0.00878)	(0.00738)	(0.00726)
Married	0.0161	0.0275	0.0307***
	(0.0228)	(0.0175)	(0.0106)
Employed	-0.00846	-0.0115	-0.0182**
	(0.0129)	(0.0102)	(0.00760)
Attending School	-0.000226	0.000756	0.000217
	(0.00247)	(0.00270)	(0.00221)
Socieconomic Stratum 2	-0.00551	-0.0129	0.00394
	(0.0215)	(0.0144)	(0.0110)
Socieconomic Stratum 1	0.0199	0.0236*	0.00311
	(0.0201)	(0.0126)	(0.00955)
Own House	0.0249	0.0218	0.0192**
	(0.0159)	(0.0127)	(0.00789)
Less than 6 years Olds	0.00783	0.0131	0.0201*
	(0.0104)	(0.0129)	(0.0113)
HH Head Age	-0.000226	0.000756	0.000217
	(0.00247)	(0.00270)	(0.00221)
Observations	181,132	246,974	340,581

Note: Standard errors in parentheses. *** significant at 1%; ** significant at 5%; * significant at 10%. Tables report Two-Staged Least Squares (2SLS) coefficients where the first stage is SR enrollment on being below the Sisben cutoff, where Sisben score is measured in 2002. Regressions control linearly for the Sisben score, flexibly around the cutoff. All variables are measured in 2002. We cluster standard errors by *comuna*. First Principal Component takes the first principal component of all other variables.

Table A4: Neighborhood Classification Method

Bandwidths:	4	6	10
Panel A: Gang-Related Violent Crimes			
Enrolled in SR	0.0171**	0.0121*	0.00891**
	(0.00751)	(0.00684)	(0.00387)
Number of observations	17,995	24,198	32,931
Panel B: Gang-Related Property Crimes			
Enrolled in SR	0.0335**	0.0271**	0.0192*
	(0.0131)	(0.0122)	(0.0107)
Number of observations	18,426	24,740	33,625
Panel C: Gang-Related Drug Crimes			
Enrolled in SR	0.0284*	0.0108	-0.00197
	(0.0163)	(0.0126)	(0.0115)
Number of observations	18,909	25,447	34,661

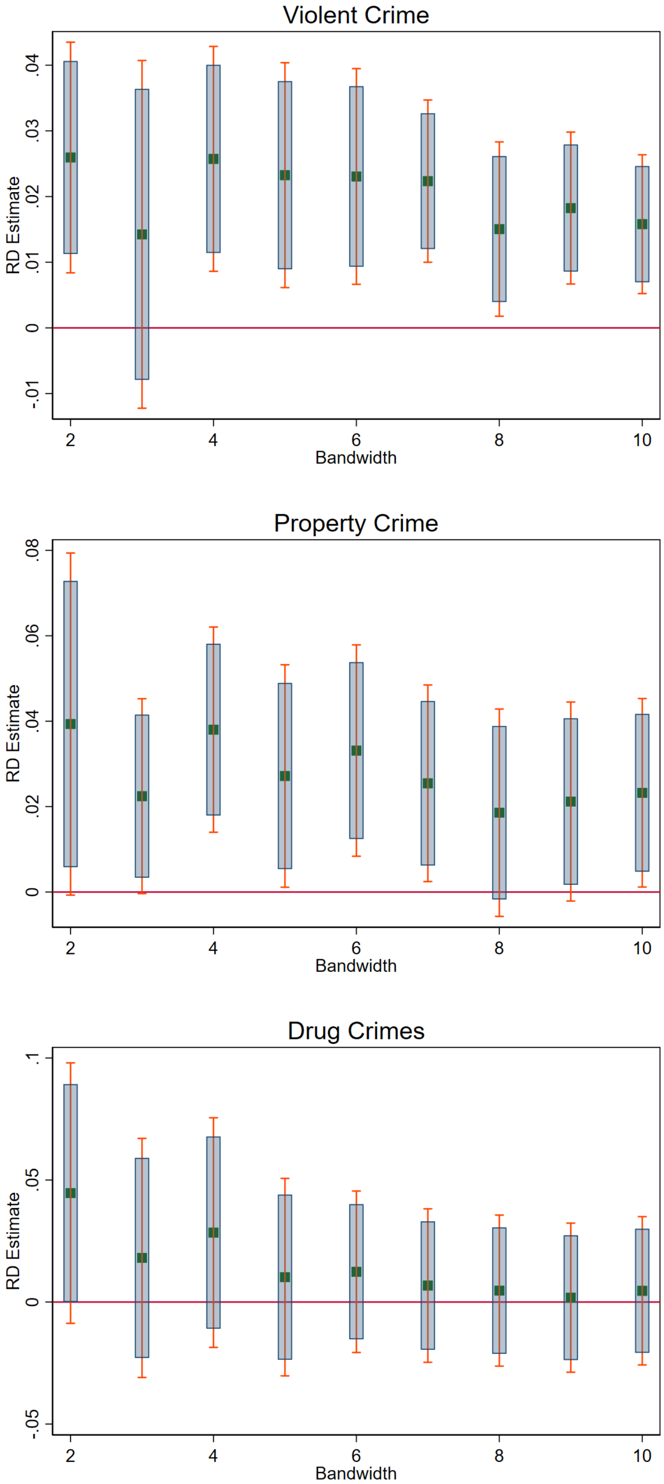
Note: Standard errors in parentheses. *** significant at 1%; ** significant at 5%; * significant at 10%. Tables report Two-Staged Least Squares (2SLS) coefficients where the first stage is SR enrollment on being below the Sisben cutoff, where Sisben score is measured in 2002. Crime data from 2005 to 2013. Results use the neighborhood classification method described in the text to classify crimes. Regressions control linearly for the Sisben score, flexibly around the cutoff. We consider only males between 21 to 26 years old in 2013.

Table A5: Semi-parametric RD with Bias Correction

Type of Crime	Violent	Property	Drug
Enrolled in SR	0.0164	0.02794	0.00768
Standard error	(0.00972)	(0.01636)	(0.02069)
Bias corrected p-value	0.077	0.052	0.57
Bandwidth	5.2	5.8	6.6
Bias correction bandwidth	9.9	8.9	9.6
Number of observations	24,206	26,511	29,102

Note: Results using the Calonico et al. (2014a) CCT method for estimation, where the primary estimation uses a linear functional form and the bias correction uses a quadratic form. Tables report Two-Staged Least Squares (2SLS) coefficients where the first stage is SR enrollment on being below the Sisben cutoff, where Sisben score is measured in 2002. Crime data is from 2005 to 2013. We consider only males between 21 to 26 years old in 2013.

Figure A4: Robustness to Bandwidths (Gang-Related Crime)



Note: Coefficients of RD 2SLS regressions where the first stage is SR Enrollment on being below the Sisben cutoff. Sample of gang-related crimes only. Grey bars indicate 90% confidence intervals. Red lines indicate 95% confidence intervals.

Table A6: Robustness Check: Including Other Crimes in the Sample

Bandwidths:	4	6	10
Panel A: Gang-Related Violent Crimes			
Enrolled in SR	0.0213***	0.0195***	0.0136***
	(0.00718)	(0.00676)	(0.00416)
Number of observations	21,720	29,235	39,877
Panel B: Gang-Related Property Crimes			
Enrolled in SR	0.0320***	0.0289***	0.0204**
	(0.0100)	(0.00998)	(0.00894)
Number of observations	21,720	29,235	39,877
Panel C: Gang-Related Drug Crimes			
Enrolled in SR	0.0238	0.0107	0.00405
	(0.0195)	(0.0138)	(0.0129)
Number of observations	21,720	29,235	39,877

Note: The sample includes other crimes. For instance, when looking at violent gang-crime arrests as the outcome of interest, property crime, drug crime and violent non-gang crime arrests are also in the sample grouped with the people never arrested in this period. Standard errors in parentheses. *** significant at 1%; ** significant at 5%; * significant at 10%. Tables report Two-Staged Least Squares (2SLS) coefficients where the first stage is SR enrollment on being below the Sisben cutoff. Regressions control linearly for the Sisben score, flexibly around the cutoff. We consider only males between 21 to 26 years old in 2013.

Table A7: Simultaneously both Non-formally Employed (in 2009) and Arrested

Bandwidths:	4	6	10
Panel A: Gang-Related Violent Crimes			
Enrolled in SR	0.0104	0.00997**	0.00990**
	(0.00771)	(0.00405)	(0.00436)
Number of observations	12,015	16,023	21,733
Panel B: Gang-Related Property Crimes			
Enrolled in SR	0.0393**	0.0331***	0.0183
	(0.0184)	(0.0122)	(0.0128)
Number of observations	12,244	16,319	22,074
Panel C: Gang-Related Drug Crimes			
Enrolled in SR	0.0207	0.0299***	0.0291**
	(0.0177)	(0.00983)	(0.0126)
Number of observations	12,253	16,368	22,199

Note: The outcome is arrests only for those not formally employed as measured in 2009. We exclude all arrests post 2009. Standard errors in parentheses. *** significant at 1%; ** significant at 5%; * significant at 10%. Tables report Two-Staged Least Squares (2SLS) coefficients where the first stage is SR enrollment on being below the Sisben cutoff, where the Sisben score is measured in 2002. Crime is measured between 2005 and 2009. Regressions control linearly for the Sisben score, flexibly around the cutoff. We consider only males between 21 to 26 years old in 2013.


BANCO DE LA REPÚBLICA
LIBERTAD
COLOMBIA