The Dose Response of Criminal Competition: Effects on Homicides and School Dropout

Ignacio Rodriguez Hurtado*

Preliminary, please do not circulate

For latest version click here

August 17, 2023

Abstract

I examine the effects of Mexican Drug Trafficking Organizations (DTOs) on homicides and school dropout between 2006 and 2018. I find that criminal competition, proxied by the number of DTOs in Mexican municipalities, leads to increases in both homicides and school dropout. Having 5 DTOs or more result in 52 additional homicides (from a mean of 6) and the relationship appears to be increasing with an S-shape. 5 DTOs or more increase school dropout by 11 percentage points (from a mean of 3 percentage points) and the relationship seems linear in the number of DTOs. Furthermore the effects on dropout are driven entirely by male students. For identification, I employ instrumental variables to identify treatment effects. I develop and estimate a novel selection model using these instruments to obtain causal estimates of interest. The selection model is designed to answer questions concerning market structure more broadly, which I adapt to estimate the effects of criminal competition in Mexico.

*Duke University, Department of Economics. I am grateful for the continued support and advice from Erica Field, Duncan Thomas, Robert Garlick and Adam Rosen in creating this paper. I also thank Hema Shah and Andres Santos for in depth discussions.

1 Introduction

Throughout Latin America, Africa and even in North American cities like Chicago, criminals engage in violent turf wars to capture illicit revenue streams. These revenue streams are large: Criminal revenues were estimated to be 3.6% of global GDP in 2009 (UNODC, 2011) and between 1.6 - 2.2 trillion USD in 2017 (GFI, 2017). This study's setting, Mexico, serves as a prime illustration of the violent competition between criminal organizations. Mexico has several violent and powerful Drug Trafficking Organizations (DTOs) competing for these illicit revenues. As a result Mexico is in a period of extreme violence with more than 30,000 homicides in 2018 and a homicide count which occasionally eclipsed that of war-torn Afghanistan or Iraq since 2006 (Insight Crime, 2016). By many accounts competition between DTOs is a major reason for the increase in violence (Medel and Thoumi, 2014; Snyder and Duran-Martinez, 2009).

This paper quantifies the effects of this criminal competition on two important outcomes: homicides and school dropout. Homicides reflect the level of violence and crime prevalent in different regions. I first document that increased criminal competition leads to an increase in violence and crime. Additionally I assess the impact of criminal competition on school dropout rates, recognizing the significance of human capital accumulation. I find that increased criminal competition increases school dropout. The repercussions of criminal competition therefore extend beyond the criminal world and have broader societal consequences.

Most models of crime predict violence to be non-linear in the number of competitors (Hirshleifer, 1995; Dal Bó and Dal Bó, 2011), and these non-linearities are of policy interest. As a basic example we expect the effect of having a monopolist DTO should be different than having a duopoly, simply because the monopolist has no competitors. Beyond this simple case, the relationship between number of DTOs and violence may be nonlinear. For instance, if there are increasing returns to criminal activity, then we might expect an increasing convex relationship between the number of DTOs and crime. The increasing returns would yield disproportionately more criminal activity with a higher number of DTOs. These non-linearities are important for policy-makers. Documenting these non-linearities can help deploy scarce policing resources to reap the most reduction in violence. For instance, if the effects on homicides are convex, then reducing DTO presence is going to have disproportionate effects in areas with a high number of DTOs.

To capture these non-linearities I estimate separate treatment effects for different “doses” of criminal competition, proxied by the number of active DTOs within a municipality. I find that increased DTOs increases homicides: having 5 DTOs or more causes 52 additional homicides on average, while having a monopolist DTO causes 13 additional homicides relative to no DTO, though the latter effect is imprecise. Moreover, this relationship has an increasing S-shape. Homicides increase in a convex fashion up until 6 DTOs and flatten out for 7 and 8 DTOs, though the standard errors are wide. This is the first paper that I know of that finds these non-linear patterns.[1]

The effects on school dropouts are all positive and are driven by older, male students. Having 5 DTOs or more causes 9th grade drop out to increase by 11 percentage points, while having a single DTO increases dropout by 3 percentage points, relative to no DTO. However these effects appear to linearly increase in the number of DTOs; I fail to find evidence of important non-linearities. For male students all the effects are positive and signficant, while for female students the effects are estimated to be zero, though with wide confidence intervals. The gendered and age-differentiated effects are consistent with at least some students dropping out to be employed by the DTOs. Indeed I find that homicides for 15 to 18 year old males also increase. This is consistent with some of the school dropouts entering criminal activity.

Naturally, DTOs are not randomly assigned. For identification, I employ instrumental variables based on DTOs' past proximity to different parts of Mexico. I leverage the sudden and large expansion in DTO presence between 2006 and 2018 to document that DTOs tended to expand in areas close to their existing operations. I argue past proximity provides useful variation and can serve as a plausible instrument for DTO presence. Moreover, I show that

[1]In similar papers, Bruhn (2021) and Sobrino (2020) also estimate the effects of the number of criminal organizations. They also find that crime increases with more groups, but do not document the non-linear patterns I find here.

after controlling for municipality fixed effects, past proximity is not correlated with several indicators of economic performance or the number of police officers in municipalities. This suggests past proximity is indeed a viable instrument for DTO presence.

I incorporate the instruments in a novel selection model to estimate treatment effects. This selection model is designed to estimate causal effects for questions related to market structure.[2] I borrow from industrial organization (IO) models of firm entry to examine DTO entry in Mexico: The first stage is a firm-entry game and the DTOs' individual entry decisions jointly determine the treatment dose, the number of active DTOs. The main benefit of the model is that it delivers interpretable causal estimates for each dose of treatment. This is in contrast to conventional Two Stage Least Squares, which typically cannot deliver interpretable causal estimates for multi-valued treatments with treatment effect heterogeneity (Heckman and Urzúa, 2010; Bhuller and Sigstad, 2022; Mountjoy, 2022a; Kirkeboen et al., 2016). The model is estimated with a simple Heckman (1979) style approach.

This paper contributes to three strands of literature. The first is the literature studying the drivers of crime, primarily in the developing world. Several authors have documented resource windfalls increase crime and violence (Berman et al., 2017; Dube and Vargas, 2013; Wright, 2015; McGuirk and Burke, 2020), while others show trade induced unemployment also increases criminal activity (Dix-Carneiro et al., 2018; Dell et al., 2019). In the context of Mexico, Dell (2015) documents increased government enforcement led to increases in homicides and violence. A closely related literature looks at the effects of criminal organizations specifically. Sobrino (2020) and Alcocer (2023) document that opium poppy and oil windfalls attracted DTOs in Mexico and increase homicides. On the other hand, Calderón et al. (2015) and Lindo and Padilla-Romo (2018) document that government captures of prominent DTO leaders causes an increase in violence, as the weakened DTO attracts more competition. Bruhn (2021) is a closely related paper, and finds that the entry of gangs in Chicago leads to increases in various measures of crime. I complement this extensive literature by finding that all else constant, increased DTO presence causes more homicides. I

[2] Similar to Ciliberto and Tamer (2009), I define market structure as the collection of active firms in a given market

also provide a rich characterization of these effects by tracing out the dose-response curve of homicides as a function of DTOs.

The second is the impact of crime on educational choices. Several papers show that violent events generally decrease school performance as measured by test scores (Chang and Padilla-Romo, 2022; Jarillo et al., 2016; Monteiro and Rocha, 2017). In a similar paper Sviatschi (2022) documents that booms in illegal coca planting induces drop out and incarceration later in life in Peru. I find that increased criminal competition by itself can increase school dropout, mainly for older male students. I further find suggestive evidence that this dropout does induce some teenagers to be pulled into crime, as I find that homicides for 15-18 year old males also increases.

Last, a major innovation of this paper is to tie together the vast IO literature on firm entry (Ryan, 2012; Aguirregabiria and Mira, 2007; Collard-Wexler, 2013; Mazzeo, 2002) with the modern treatment effect literature (Heckman et al., 2006; Mountjoy, 2022b; Mogstad et al., 2018). The econometric model developed in this paper can be used to obtain causal estimates of the impact of market structure on generic downstream outcomes. Other industries could be studied as well, not only the criminal market. For example the model could be used to estimate the effects of export manufacturing firms on school dropout, like the setting in Atkin (2016). Overall, the model is similar to recent papers estimating treatment effects for non-standard settings (Walters, 2018; Kline and Walters, 2016; Rose and Shem-Tov, 2021). My framework models the first stage as a private information entry game involving different groups, which is a commonly used model in IO (Aguirregabiria et al., 2021). Furthermore, it allows for dynamic games with forward looking agents, can accommodate multiple equilibria and does not require to estimate any structural parameters. A closely related paper is Balat and Han (2022) which derives informative bounds based on shape restrictions, while my framework produces point estimates instead and does not impose shape restrictions.

The rest of this paper is organized as follows. Section 2 provides context on the Mexican War on Drugs and its DTOs. Section 3 describes the data sources used. Section 4 provides descriptive statistics and describes the empirical strategy. Section 5 provides preliminary results, including Two-Stage Least Squares results. Section 6 exposes the main model used for estimation and Section 7 details the estimation procedure. Section 8 discusses the results and Section 9 concludes.

2 The Mexican War on Drugs and Drug Trafficking Organizations

2.1 Mexican War on Drugs

The Mexican War on Drugs refers to the ongoing conflict between the Mexican government and powerful Drug Trafficking Organizations (DTOs)[3] that intensified in the 2000s. The roots of this war can be traced back to the 1980s when the United States began implementing stricter drug policies, leading to the relocation of major drug trafficking routes from the Caribbean to Mexico (Medel and Thoumi, 2014). As a result, between the 1980s and 2000s Mexican DTOs established their dominance over the profitable drug trade into the US.

In spite of the increase of drug trafficking, the period between the 1980s and the early 2000s was a relatively peaceful time in Mexico. Several scholars indicate that a tacit understanding between the Mexican government and DTOs fostered a relatively peaceful environment (Snyder and Duran-Martinez, 2009; Shirk and Wallman, 2015). The hegemonic political actor during this time was the PRI (*Partido Revolucionario Institucional*) party in Mexico, which ruled Mexico uninterruptedly for much of the 20th century. The PRI party had an accommodating relationship with DTOs: in exchange for bribes, the government did not harass the DTOs (Medel and Thoumi, 2014).

This panorama changed in the 2000s, when the PRI party lost power for the first time. In the 2000 election to the opposition PAN party (*Partido Acción Nacional*) gained power, led by President Vicente Fox. The pact between the government and DTOs became untenable

[3] Mexican DTOs are commonly called cartels. This is a misnomer because these organizations do not collude to raise prices of the goods they offer. Throughout this paper I refer to them as Drug Trafficking Organizations (DTOs) instead.

as new political actors gained power and previous relationships between government officials and DTOs were eroded (Snyder and Duran-Martinez, 2009). The Fox regime adopted a confrontational approach towards the DTOs, even capturing a prominent DTO leader, and as a result drug-related violence began to increase (Medel and Thoumi, 2014).

The year 2006 marked a turning point in the struggle between the new government and DTOs. President Calderón of the PAN party replaced the Fox administration and promptly instituted a drug trafficking crackdown of unprecedented scale. The Calderón administration deployed 45,000 troops to combat smuggling and between 2006 and 2008 alone 184 smugglers were extradited to the United States (Medel and Thoumi, 2014).

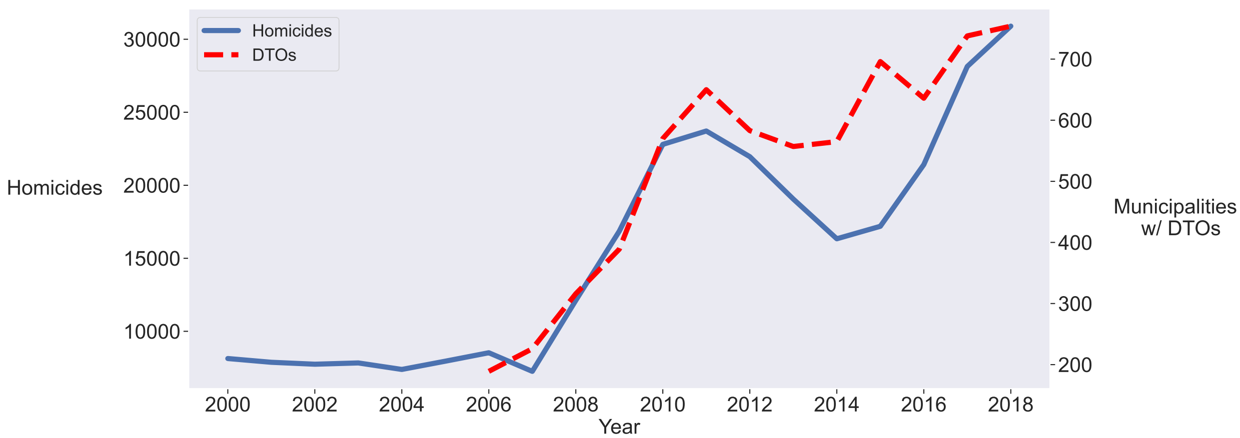
Several studies have by now documented the crackdown during the Calderón administration led to a sustained increase in homicides and drug-related violence. Dell (2015) documents that municipalities narrowly won by PAN experienced increases in violence. The crackdown likely spurred further violence by weakening incumbents and encouraging new entrants. Other studies have shown that when a top DTO kingpin is captured, homicides tend to increase. This spike in violence is attributed to the organization's weakening after government intervention, which often leads to heightened competition in the affected area (Shirk and Wallman, 2015; Lindo and Padilla-Romo, 2018; Calderón et al., 2015).

Violence did not stop after the Calderón administration. In 2012, the PRI party regained control of the presidency with the Peña Nieto. The new administration tried a less militarized approach and increased the federal police force by 35000. Nonetheless, within a year the president had mobilized military troops to the state of Michoacán (Medel and Thoumi, 2014). Despite the attempted change in policy, the administration continued with many of the Calderón crackdown policies.

These dynamics can clearly be seen in Figure 1. The solid line shows the number of homicides, with axis on the left. There is a dramatic rise in homicides after 2006, which decreased slightly during 2012. However the increase resumed and by 2018 homicides had more than tripled relative to the 2000 baseline. Simultaneously DTOs expanded their footprint across Mexico. The dashed line of Figure 1 counts the number of municipalities with at least one DTO in Mexico using the right axis. The number of municipalities with some DTO presence increased form roughly 200 in 2005 to more than 700 in 2018.

Due to the tremendous rise in violence during the 2006 crackdown, I limit the sample years of this paper to be between 2006 and 2018.

Figure 1: Homicides and DTO Presence



Notes: This figure plots the number of homicides by year in Mexico in the blue solid line, displayed on the left axis. The red dashed line displays the number of municipalities with at least one of the 8 large DTOs studied in this paper, displayed on the right y-axis. The 8 DTOs are mentioned in Section 2.2.

2.2 Mexico's Drug Trafficking Organizations

Mexican DTOs have evolved into large transnational operations engaged in a diverse range of activities. Apart from drug trafficking, they have expanded their operations to include oil theft (Franco-Vivanco et al., 2023; Alcocer, 2023), extortion and kidnapping, among others (Correa-Cabrera, 2021). Their operational scale and scope resemble that of multi-product, multi-national firms in the legitimate business world.

Furthermore DTOs differ in their practices, just like legitimate businesses operating within the same industry. For example, while the Sinaloa Cartel focuses primarily on drug trafficking as its core revenue stream, Los Zetas have carved a niche for themselves in the realm of extortion, and they are often regarded as employing more ruthless methods compared to other DTOs (Correa-Cabrera, 2021).

A pivotal aspect that sets DTOs apart from other businesses is their decentralized nature, which shapes their operations and behavior significantly. In general, criminal organizations exhibit low capacity to give top-down orders and effectively supervise lower-ranking members (Pereyra, 2012; Natarajan, 2006). For example, Benson and Decker (2010) interview arrested drug traffickers and finds that actors within these organizations operate independently and are often disconnected from other crucial actors within the same organization.

The lack of coordination within DTOs has two important implications for this paper's empirical strategy. First, the lack of coordination complicates the establishment and continuity of inter-DTO alliances. The lack of coordination hampers DTO's ability to negotiate and enforce agreements between organizations (Pereyra, 2012). Second, the decentralized organizational structure of DTOs limits their capacity to execute extensive multi-year expansion plans within Mexico. With no centralized command structure, DTOs are curtailed in their ability to orchestrate coordinated efforts across Mexico.

This paper focuses on the eight largest DTOs in Mexico.[4] The selection of these DTOs is based on their mentions in the DEA's National Drug Threat Assessment reports from 2008 to 2018. Notably, these DTOs represent the largest and most significant players in the Mexican drug trade, regularly involved in trafficking substantial quantities of drugs into the United States. Moreover, their operations span vast territories across Mexico. At the start of my study period, 2006, not all of these 8 large DTOs existed and three of them were created after 2006. Table A1 lists the years when the DTOs are observed in my data. The three DTOs splintered off from preexisting organizations.[5]

This fragmentation is a common feature of Mexican DTOs. In addition to the creation of three large DTOs after 2006, the paper's study period also witnessed the emergence of several smaller DTOs. The phenomenon of fragmentation has garnered attention from various scholars (Pereyra, 2012; Medel and Thoumi, 2014). However, this study does not

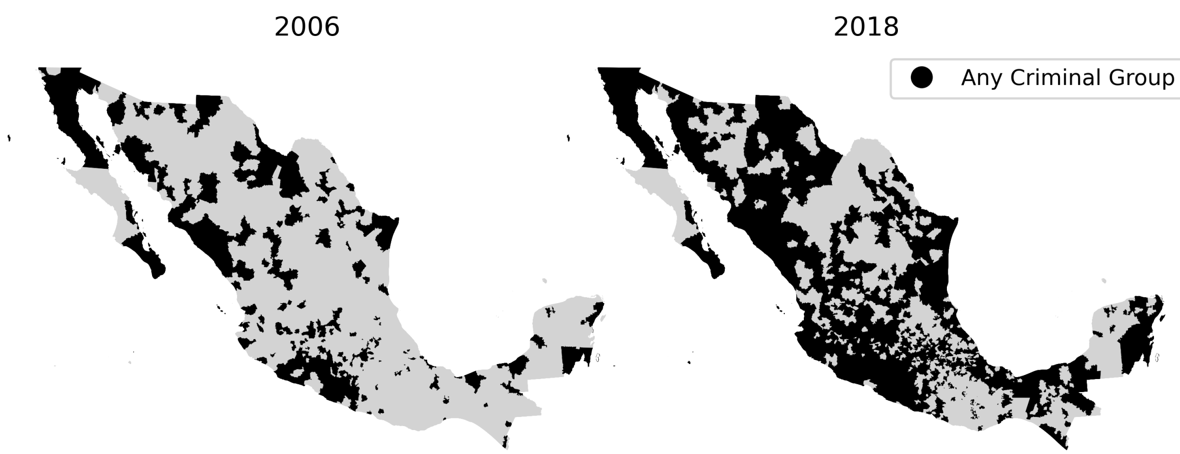
[4]These groups are: Sinaloa Cartel (*Cártel de Sinaloa*), *Los Zetas*, Juárez Cartel (*Cártel de Ju'arez*), New Generation Jalisco Cartel (*Cártel Jalisco Nueva Generación*), Beltran-Leyva Organization, Gulf Cartel (*Cártel del Golfo*), Michoacan Family (*Familia Michoacana*), Knights Templar (*Caballeros Templarios*).

[5]*Los Zetas* splintered from *Cártel del Golfo*. *Caballeros Templarios* splintered from *Familia Michoacana* and *Cártel Jalisco Nueva Generación* emerged from the defunct Milenio Cartel.

concentrate on these smaller groups for three reasons. First, their relatively small size limits their significance, as they lack access to the highly lucrative US drug trade (Pereyra, 2012). Second, the fragmented DTOs operate on a much smaller scale compared to the 8 larger DTOs. That data in this study bears this out: in addition to the 8 large DTOs, the data tracks 64 smaller DTOs. Among all observations during the study period with any DTO presence (encompassing the 8 large and 64 smaller DTOs), the 64 smaller groups account for roughly 32% of these instances, while the 8 larger DTOs make up the remaining 68%. Last, these smaller DTOs predominantly operate within limited regions and demonstrate minimal expansion into new areas. Consequently, their analysis does not align with the empirical strategy of this paper.

Figure 2 displays the expansion of the 8 large DTOs between 2006 and 2018. As we can see there was a significant of DTO presence during this time period. The 8 large DTOs cover significant portions of the country, speaking to the large scale of the DTOs.

Figure 2: Any DTO Presence, 2006 and 2018



Notes: This figure plots municipalities with any of the 8 large DTOs studied in this paper in black. The gray area are municipalities with no DTO presence. The left panel plots presence in 2008, the right panel plots presence in 2018.

3 Data Sources

I gather data from a variety of sources for my analysis. I use the data to construct a municipality-year panel in Mexico between the years of 2006 and 2018 for 2435 municipalities.[6]

3.1 Drug Trafficking Organization Presence

The data source on DTO presence comes from Sobrino (2020). It is collected from Google by crawling through news articles for matches of criminal group names and municipality mentions across various years. The crawler searches for articles that explicitly mention both the DTO and the municipality within the same sentence. As a next step, a convolutional neural network (CNN) was trained on 5000 sentences to further distinguish sentences that describe a cartel as being present in the municipality versus sentence that do not describe presence but instead mention the DTO for an unrelated reason. Therefore a news article is classified as describing a particular DTO's presence in a municipality if it contains a sentence with both the DTO-municipality pair, and the CNN predicts the sentence is describing the presence of the DTO. The version of the dataset I use is at the municipality-year level with a dummy variable whenever any news article claims DTO is present in a municipality-year according to these criteria. The dataset tracks presence for 72 DTOs between 1990 and 2020. I focus on the 8 largest DTOs during 2006 and 2020.

Furthermore for the rest of analysis I use an imputed measure of presence. I smooth over the presence indicators and fill in the gaps between two time periods when a DTO is marked as being within a municipality. My claim is that these gaps likely still had DTO presence, and its just there was not a news worthy enough event to mark them as present in the data. For each DTO I impute between 200 and 800 observations as being present, and the average length of the gap I fill is around 2 years. On the whole, the main results are not sensitive to this imputation. However the imputation allows for a richer characterization of the effects of

[6] I group together all the municipalities corresponding to Ciudad de México into a single unit. For municipalities that changed or were created after 2005, I group them according to 2005 boundaries.

DTO count, since the imputed measures naturally has more non-zero DTO counts. I discuss these results more in detail in Appendix Section B.

3.2 Data for Main Outcomes

3.2.1 Homicides

Information on homicides comes from mortality records provided by Mexico's census office INEGI. INEGI's annual database details all registered deaths in Mexico. It contains basic information on the victim such as gender, age of the deceased, occupation, municipality of residence, and municipality of homicide. Using the correct mortality codes, I keep all deaths that are deemed to be homicides.

3.2.2 Middle School Drop Out

Data on dropouts come from Mexico's 911 School Registry information (*Estadística del Sistema Educativo Mexicano Formato 911*). These data were provided by *Aprender Con Evidencia*, a Mexican educational non-profit. These are administrative forms every primary, middle school and high school have to complete in Mexico. The 911 registries have detailed enrollment information and they track enrollment by grade, sex and age. They separate out newly enrolled students versus repeaters for each grade. The registry has information for the start and end of each school year.

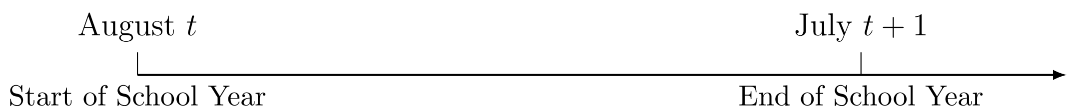
In this paper I focus on intra-curricular dropout. This is dropout that happens between the start of the academic year and the end of the same school year. The school year in Mexico starts in August and finishes in July of the following year. To calculate intra-curricular dropout rates for grade 9, I first select schools that reported both at the start and end of the same academic year. I then aggregate enrollment counts for all schools in each municipality. As a result, if a student changes school within the same municipality, they are still captured in my data as being in school.

The total number of students that dropout for grade 9 is the difference between enrollment at the end of the school year relative to the start of the school year. I normalize this difference by the start of year enrollment, as in the below formula:

$$\text{Dropout}^9_{mt} = \frac{\text{Enrolled Start of Year}^9_{mt} - \text{Enrolled End of Year}^9_{mt}}{\text{Enrolled Start of Year}^9_{mt}} \tag{1}$$

I repeat this calculation broken out by gender as well, where I use gender specific enrollment counts. I also repeat this for other grades. Dropout for time t corresponds to intra-curricular dropout for the school year beginning at t. For instance, dropout in the 2014-2015 school year is marked as dropout for the calendar year 2014 as summarized in Figure 3.

Figure 3: School Calendar



Notes: Figure describes the end and start of school year s relative to calendar year t. The end of the school year in Mexico happens in July, and begins again in August. 9th grade dropout for calendar year t is calculcated relative to enrollment at the start of the school year. I then compare how many students were enrolled at the end of the same school year during the calendar year $t+1$.

3.3 Other data

I gather other data to supplement the analysis.

Agricultural Revenue Data: I gather municipality specific records of agricultural production and revenue, in nominal Mexican pesos. These are estimates published by Mexico's SIAP office (Servicio de Información Agroalimentaria y Pesquera). SIAP estimates crop production for a variety of crops for each municipality, and estimates the revenue brought by these crops. This data is available for the entire sample period, 2006-2018

Population Counts I use Mexico's census counts for 2005, 2010, 2015 and 2020 to estimate population counts by municipality. I linearly interpolate the population counts for the years between the census. I use population counts as a control variable in the analysis.

Electricity Information Mexico's CFE (*Comisión Federal de Electricidad*) publishes information on the amount of registered households and total electricity consumption for each municipality at the yearly level. This data is only available from 2010-2018.

Police Officer Information. Mexico's Census office, INEGI, carries out a biannual census on municipal public services, which has information on the number of police officers employed across municipalities. This data is available for 2012, 2014, 2016 and 2018. I use this data to check if the instrument is correlated with higher levels of policing. Due to the limited availability, it is not used for the main analysis.

ENLACE Test data: in supplemental analysis, I use individual test score data for the ENLACE test between 2008 and 2013. ENLACE was a universal, obligatory test administered to Mexican students during this time period. I use this data to validate my results on dropout since the data effectively tracks students' enrollment over time. The data was provided by *Aprender con Evidencia.*

4 Descriptive Statistics and Identification Strategy

4.1 Descriptive Statistics

The main econometric challenge for my research question is a classic selection problem: DTOs are not randomly assigned and instead concentrate in selected areas. This can be seen in Table 1. The top panel splits municipalities according to the maximum number of DTOs every observed between 2006 and 2018. We observe a clear pattern: areas with more population, higher literacy rates and less poverty attract more DTOs. Places that receive many DTOs are also more violent: municipalities that at some point had 3 or more DTOs have a homicide rate of 22.97, while municipalities that never had a DTO have a homicide rate of 11.64. There is a similar pattern with 9th grade dropout: places with more DTOs have a slightly higher dropout rate of 0.06 relative to 0.04 for places with no DTOs. Disentangling the causal effect of DTO concentration is complicated by this selection problem. The difference in homicides and dropout may be driven by the substantive economic differences between places with no DTOs relative to places with more DTOs.

The bottom panel of Table 1 shows summary statistics of the rapid expansion of DTO presence between 2006 and 2018. We observe important changes in the average number of

Table 1: Descriptive Statistics

	Panel A: Means By Number of Groups				
	0	1	2	3+	Overall
Covariates					
Population (2005)	10,787	25,413	30,074	98,972	43,432
Poverty Rate (2005)	0.75	0.67	0.63	0.56	0.67
Literacy Rate (2005)	0.80	0.84	0.85	0.88	0.83
Elect. Users	3,335	7,833	9,380	32,993	14,218
kWH (millions)	14.31	37.23	58.47	209.14	85.68
kWH per Capita	3,739	5,224	8,581	9,668	6,360
Ag. Revenue	52.98	120.81	155.47	346.70	167.39
Total Police	27	44	59	153	74
Outcomes					
Homicide Rate (per 100,000)	11.94	13.27	15.08	22.77	15.98
9th Grade Enrollment	221.37	493.02	587.57	1,873.87	832.32
9th Grade Dropout Rate	0.03	0.03	0.03	0.03	0.03
9th Grade Dropout Rate (Girls)	0.02	0.03	0.03	0.03	0.02
9th Grade Dropout Rate (Boys)	0.02	0.03	0.03	0.03	0.03
	Panel B: Dynamics of Criminal Group Presence				
	2006-2008	2009-2011	2012-2014	2015-2018	Overall
Treatments					
Number of DTOs	0.37	0.92	1.52	1.54	1.12
$\geq$ 1 DTOs	0.14	0.28	0.36	0.39	0.30
$\geq$ 2 DTOs	0.10	0.22	0.32	0.33	0.25
$\geq$ 3 DTOs	0.05	0.14	0.23	0.23	0.17
Instruments					
Average Dist.$_{-1}$ (km)	126.06	77.45	57.57	50.64	75.81
Dist.$_{-1}$ BLO (km)	119.59	59.75	52.50	51.11	69.21
Dist.$_{-1}$ CABT (km)	.	127.80	66.30	55.37	68.52
Dist.$_{-1}$ CDG (km)	81.32	57.59	48.20	46.71	57.54
Dist.$_{-1}$ CDS (km)	88.77	46.39	39.87	34.94	51.13
Dist.$_{-1}$ CJ (km)	152.55	102.00	85.55	84.50	104.46
Dist.$_{-1}$ CJNG (km)	.	113.20	72.62	39.59	66.94
Dist.$_{-1}$ FM (km)	188.04	90.17	61.81	59.43	96.72
Dist.$_{-1}$ LZ (km)	.	40.43	33.71	33.51	34.45

Notes: This table displays averages for different variables. The top panel display averages splitting observations according to the highest number of DTOs observed between 2006-2018 for each municipality. The bottom panel splits observations according to different time periods.

DTOs during the sample period. Between 2006 and 2008 the average number of DTOs for a given municipality-year was 0.27. The average increases to 1.12 DTOs for 2015-2018. This change is driven both by the extensive and intensive margin of DTO group count. Between 2006 and 2008 14% of municipality-years have at least 1 DTO and 3% of observations have 3 or more DTOs. Between 2015 and 2018 these proportions increase to 39% and 18% respectively.

4.2 Identification Strategy

To address the selection problem I leverage the rapid expansion of DTOs across Mexico between 2006 and 2018. During this time period DTOs demonstrated a preference for past proximity: they were more likely to expand to areas nearby their existing operations. As an example, consider the expansion of the *Familia Michoacana* between 2006 and 2018, as displayed in Figure 4. *Familia Michoacana* originate in the state of Michoacán in the south-west of the country and had a large presence in the southwestern region in 2006. Furthermore, in 2006 it was present in the Yucatan peninsula in the east of the country as well. By 2018 we can see *Familia Michoacana* expanded heavily nearby its main center of operations in 2006 and also expanded in Yucatan.

Figure 4: Familia Michoacana (FM) Presence 2006 and 2018

Notes: This table plots the presence of *Familia Michoacana* for 2006 and 2018.

My identification argument is that DTO's preference for past proximity provides useful variation to identify the effects of DTO presence and group count. Insofar as DTOs enter municipalities solely because they are nearby existing operations, the proximity to the different DTOs may be a useful instrument for their presence. I use past proximity to instrument current DTO presence, using the one-year lagged distances as instruments. The argument is that places that are nearby to already existing DTO operations are at risk for DTO entry, and that this past proximity should have little to no effect on current outcomes, except for the increase in DTO entry. Distance instruments have been commonly used in other settings, for example in studying returns to schooling Carneiro et al. (2011); Mountjoy (2022b) and also in the IO literature studying firm entry Aguirregabiria and Magesan (2020); Ellickson et al. (2013). Additionally I strengthen my identification argument by controlling for municipality and time fixed effects. That is, once we account for permanent differences in the cross-section, municipalities that are further from DTOs should be similar to places that are nearby DTOs, both along observable and unobservable margins.

The proposed instrument will only be valid if it is independent of potential outcomes and satisfies an exclusion restriction. Distances can affect outcomes only through their effect on DTO presence. I present arguments for this case next.

I argue the distance instruments satisfy the independence condition because DTOs are unable to carry out multi-year and cross-regional plans targeting different areas of the country. If DTOs are able to successfully manipulate their entire network to systematically target more profitable regions over time, then the distance instrument would fail. In this case, places that are close to DTOs are just fulfilling the DTO's long term plans and are therefore selected either along observable or unobservable margins. In contrast, if DTOs are unable to carry out a cross-regional plan, I argue the distances will largely be a function of the initial DTO placement and unobserved cross-sectional differences in profitability for DTO entry. Therefore, conditional on municipality fixed effects, the distances are plausible independent of potential outcomes.

There are two main reasons why DTOs are unable to carry out multi-year cross regional-plans. First, DTOs are large and decentralized operations. Due to the inherent difficulties of managing crime at scale, information and orders cannot flow freely from the top to the bottom of the organization. As a result, even if the top level of management would like to execute a cross-regional plan, the execution would be complicated by the relative lack of oversight on lieutenants. Second, the competition between DTOs would severely hinder a successful execution of such a multi-year plan. DTOs operate in a chaotic environment where their competitors actions can be unpredictable, further complicating the execution of a multi-year and cross-regional plan. [7]

Regarding the exclusion restriction, the main concern we should have are spillovers, since the distance instrument may simply be capturing a spillover from municipalities with DTOs to those without. Moreover these spillovers need to be dynamic since I'm using last year's distances as instruments for current DTO group count.

The decentralized nature of the DTOs limits the scope of such a spillover. A natural concern is that last year's distance is correlated with contemporary distance. If DTOs shift resources and labor across their network then clearly the exclusion restriction would not hold: last year's distances would mean resources can be allocated to areas with more intense fighting in the current period, therefore affecting current outcomes directly. However the decentralized nature of DTOs rules this possibilty out since it makes sharing resources across the DTO network unlikely. Moreover in robustness checks I control for current distances as well and my main results are unaffected.

Another spillover to be concerned about is migration; past proximity to DTOs may spur out migration as inhabitants anticipate DTO entry and a subsequent rise in violence. This out-migration could then lead to increased school dropout. I claim a migratory response is unlikely for two reasons. First migration is a costly decision and it is odd to think inhabitants are so forward looking and sensitive to the possibility of a rise in violence that they abandon their homes before a rise in violence even occurs. Prior work corroborates this view: Daniele et al. (2023) find a delayed response of many years in migration due to an opium demand shock in Mexico for the same period studied here. Second, I extensively investigate if my

[7] The adage "No plan survives contact with the enemy" captures this idea.

main results are driven by migration. I find no evidence that the large effects on dropout are due to to out migration.

More generally, the distance instrument should be more credible than other approaches commonly used in the literature. For instance, many researchers exploit variation in the price of natural resources to document that natural resource windfalls increase conflict and crime (Berman et al., 2017; Dube and Vargas, 2013; Wright, 2015; McGuirk and Burke, 2020). These price shocks are excellent sources of identification to study the effects of natural resource windfalls on crime. However, in my setting these shocks would likely be poor instruments for DTO presence. While a natural resource windfall would likely attract Mexican DTOs, it is difficult to attribute the subsequent change in violence or other outcomes solely to increased DTO presence. The windfall itself would likely cause direct increases in violence. For example, the windfall may allow incumbent DTOs to hire more foot soldiers without further DTOs entering as well. In contrast, my strategy based on past proximity can alleviate some of these concerns.

Figure 5 displays how I calculate distances. In this example, I calculate the distance for municipality A for group j. For each year, I calculate the distance from A to all other municipalities where group j is present. I then take the shortest distance (highlighted in red) and call that group j's distance to municipality A. I repeat this for all municipalities, years and groups. I then take the one year lag and use these distances as my instruments.

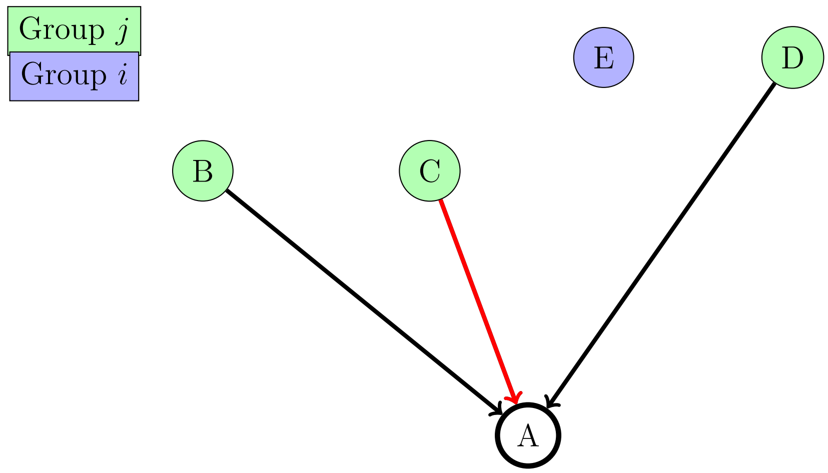
5 Instrument Diagnostics and Preliminary Results

5.1 Diagnostics

I first assess the credibility of my instrumentation strategy using observable covariates. If my identification strategy is sound, we expect the one year lagged distances to be uncorrelated with important municipality features. I run the following fixed effect regressions

$$Y_{mt} = \alpha_m + \tau_t + \beta \overline{Z}_{mt-1} + u_{mt} \tag{2}$$

Figure 5: Calculating Group-Specific Distances to Municipalities



Notes: Example calculation of distance calculation. The figure illustrates the distance calculation for group j for municipality A at a particular year t. I first calculate all pairwise distances from municipality A to all other municipalities where group j is present at t. I then set the minimum distance, depicted in red, as the distance of municipality A for group j for year t. I repeat the calculation for all groups, municipalities and years.

$$Y_{mt} = \alpha_m + \tau_t + \delta W_{mt} + \sum_{g=1}^{G} \beta_g Z_{gmt-1} + u_{mt} \tag{3}$$

Z_{gmt-1} are the individual DTO distances based on its previous year network. I log all distances Z_{gmt-1} for ease of interpretation. $\overline{Z}_{mt-1}$ is the average log distance across all DTOs g for municipality m at time $t-1$. The terms $\sum_{g=1}^{G} \beta_g Z_{gmt-1}$ are instead the individual DTO distances. W_{mt} are controls typically used in IO models of entry. These include lagged presence indicators for each of the 8 DTOs, the inverse hyperbolic of the total active DTO count for previous year and the log of population counts as imputed by the census. I include these controls to be consistent with the econometric model developed in Section 6.

Y_{mt} are different measures of municipality characteristics. If my identification strategy is correct, we expect the DTO distances to be uncorrelated with some of the observable municipality characteristics I've gathered. These are the number of residential electricity users, the kWH electricity consumption, kWH electircity consumption per capita, agricultural revenue and the total number of police officers. I apply the inverse hyperbolic transformation to these variables because of their skewed distributions and because some of them have observations with zero.[8] For simplicity, I focus on the years 2011 until 2018, since most of these covariates are only available after 2010 and 2011 is the first year where all the groups have a lagged distance variable.

The results are shown in Table 2. For reference the first panel shows results of equation 2 with no fixed effects, while the second panel shows the same regression with fixed effects. We clearly see that the fixed effects eliminate most of the correlations in the first panel. The first panel shows that DTOs networks are on average further away from economically active areas. For instance, column 1 indicates that a 10% increase in the average distance corresponds to a 23% increase in the number of electrical users. increase in the number of electric users. In contrast, the second panel has economically small and mostly insignificant coefficients. The only statistically significant coefficient is for the IHS of agricultural revenue, which is significant at the 1% level. Nonetheless, its magnitude is relatively small. For instance, a 10% increase in the average distance across all DTOs corresponds to a 2.1% increase in agricultural output.

The third panel estimates instead equation 3 without controlling for W_{mt}. Here instead of showing a single coefficient, I display the F statistic and associated p-value that all distance coefficients are zero. We find the same statistical patterns as the top two panels. The last panel adds the W_{mt} controls. The third panel is indistinguishable from the fourth panel, which is reassuring: my identification argument relies on controlling for the appropriate fixed effects. Adding further controls does not change the relationship between my instruments and these covariates. I include the controls to be consistent with the model developed in Section 6.

Overall the instrument seem to be uncorrelated with several important observables. After controlling for fixed effects, DTOs do not shift their network to target economically prosperous areas, as evidence by the electricity covariates, or places with less police enforcement, as evidenced by column 5. This is reassuring for the proposed empirical strategy. Furthermore, I include robustness checks where I check the sensitivity to controlling for agricultural

[8] Only Agricultural revenue and total police officers have zero-values covariates. Agricultural revenue has 107 observations with 0 and total police officers has 394 observations with zero.

Table 2: Instrument Balance

	(1) IHS(Elect. Users)	(2) IHS(kWH)	(3) IHS(kWH per Capita)	(4) IHS(Ag. Revenue)	(5) IHS(Total Police)
	No Year, Muni. FEs				
Average Distance	2.33***	4.36***	2.21***	4.78***	1.03***
	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)
	With Year, Muni. FEs				
Average Distance	0.01	-0.01	-0.01	0.21***	0.05
	(0.01)	(0.02)	(0.02)	(0.04)	(0.05)
	Individual Distances With Year, Muni. FEs				
All Distances Joint F-Stat	8.02	10.08	7.46	48.11***	10.27
P-Value	[0.43]	[0.26]	[0.49]	[0.00]	[0.25]
	Individual Distances With Controls, Year, Muni. FEs				
All Distances Joint F-Stat	8.72	9.28	7.22	45.54***	9.10
P-Value	[0.37]	[0.32]	[0.51]	[0.00]	[0.33]
Outcome Mean	14.22	85350.13	5.88	192665.40	73.06
N	19430	19430	19430	19480	9740

Notes: Unit of observation is a municipality-year. Column headings indicate the outcome used in the regression. IHS stands for the inverse hyperbolic sine transformation. Elect. Users refers to the number of residential users reported by municipality and year. kWH refers to kilowatt-hours consumed in each municipality. kWH per Capita is kWH consumption divided by the number of residential users. Ag. Revenue refers to the amount of of agricultural revenue in nominal Mexican pesos. Total Police refers to the total number of officers dedicated to public security.

The row labeled Outcome mean displays the untransformed means of the outcomes, in levels. Column 1 reports average residential users in thousands, Column 2 reports average kWH consumption in thousands, column 3 reports average kWH consumption in thousands per capita, column 4 reports average agricultural revenue in thousands of Mexican pesos, in nominal terms. Column 5 reports the average number of police officers. The row labeled N displays the number of observations included in each regression.

Sample years are 2011-2018. Columns 1-4 have annual data for this period. Total Police is reported biannually for 2012, 2014, 2016 and 2018.

The panel labeled "No Year Muni. FEs" reports estimates of β from equation 3 without the municipality and year effects. β is the coefficient on average logged distance for all criminal groups in the lagged year. The panel labeled "With Year Muni. FEs" reports estimates of β from equation 3. The panel labeled "Individual Distances With Year Muni. FEs" reports the F-statistic and associated value for the null hypothesis that $\beta_g = 0 \forall g$ in equation 3. β_g correspond to the coefficients on each group g's logged distance for the lagged year. "Individual Distances With Controls, Year Muni. FEs" reports F-statistics for models that include controls and fixed effects. The controls are dummy variables for the presence of each of the 8 criminal groups for the preceding year and the inverse hyperbolic sine of the total number of active criminal groups in the preceding year.

Standard errors in parenthesis clustered by municipality. P-values for the F-tests reported in brackets.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

revenue. The main results are robust to controlling for agricultural revenue.

5.2 Two Stage Least Squares Results

Next I examine preliminary two stage least squares (TSLS) regressions. Specifically I run the model:

$$N_{mt} = \alpha^1_m + \tau^1_t + \delta^1 W_{mt} + \sum_{g=1}^{G} \beta^1_g Z_{gmt-1} + u^1_{mt} \tag{4}$$

$$Y_{mt} = \alpha^2_m + \tau^2_t + \delta^2 W_{mt} + \gamma^2 N_{mt} + u^2_{mt} \tag{5}$$

Where Y_{mt} is either the count of homicidies or 9th grade dropout rate, W_{mt} is as before, and N_{mt} is the total number of active DTOs in municipality m at time t. I refer to this variable as the *group count* moving forward. Equation 4 is the first stage for equation 5, where the group count is the endogenous variable.

Results can be seen in Table 3. I use all the sample years 2006-2018. Some groups are created after 2006, so I impute distances to be zero for the years where the DTO does not have a lagged distance variable. This simple linear model finds a positive effect of group count on homicides at the 10% significance level. An additional group results on average in 9.14 extra homicides. This represents a large increase in homicides. The average level of homicides in my sample is 7.95, so one additional DTO implies a substantial increase in the number of homicides relative to the overall mean. There is no statistically significant effect on dropout rates and the estimated coefficient is -0.014.

In addition to the point estimates, I computed the Olea and Pflueger (2013) statistic for weak instruments, as suggested in Andrews et al. (2019). We can see there is no weak instrument problem in my setting. The effective F-stat is 29.88 well above the threshold of 10% bias relative to OLS. The distance instruments strongly predict DTO group count.

In addition, I compute the Sargan-Hansen J-statistic. For both the homicidal and dropout outcomes we confidently reject the over-identification test of constant treatment effects. This is suggestive of heterogeneous effects for the γ^2 coefficient in Equation 5. Heterogeneity is especially likely given the nature of the instrument. For example, if the Sinaloa Cartel is nearby they are more likely to enter. The Sinaloa Cartel may have different a different effect on homicides than another DTO, e.g. the Gulf Cartel. This may be due to different man power capacities or different tactical strategies. In other words, the compliers shifted by particular DTO's distance have a different response that compliers shifted by another DTO's distance. Informally, this is the same intuition analyzed in more detail in Mogstad et al. (2021). The different instruments will naturally have different complier groups.

I explore this heterogeneity in Appendix Figure A1. This figure redoes the TSLS estimation in Equations 5 4, but changes the instruments so it only uses one DTO's distance at a time. The other DTO distances are included as controls in the first and second stage regression. The TSLS model is therefore just identified, with only one distance being used as an instrument. The estimates are naturally noisier since I use less information to predict DTO group count. Nonetheless we can observe important differences between the point estimates. For instance the effect on dropout is much larger using only Sinaloa Cartel's (CDS) distance instead of Beltran Leyva Organization's (BLO) distance. These estimates suggest we should expect effect heterogeneity to matter when using the DTO distances.

The evidence of treatment effect heterogeneity complicates the task of interpreting the results in Table 3 as a causal estimate. We could appeal to the ordered treatment model in Angrist and Imbens (1995) in order to interpret the coefficient on group count. However this would require a particularly strong monotonicity assumption which is unrealistic in this setting.[9] Moreover, even with the strong monotonicity assumption, the results in Angrist and Imbens (1995) would imply that the coefficient in Table 3 is weighted average across different "doses" of DTO group count (N_{mt}) and different complier groups. The TSLS estimate may for instance be placing more weight on the transition between 1 DTO versus 0 DTOs relative to the transition between 3 DTOs versus 2. It is difficult to know which transition is driving

[9] In particular, we would need to assume a DTO j's distance has a monotonic effect on group count N_{mt}. This would imply that e.g. a closer distance for j would never decrease N_{mt}. This precludes strategic effects where j's closer distance never intimidates competitor's from entering, thereby reducing N_{mt}

Table 3: Two Stage Least Squares Results

	(1) Homicides	(2) Dropout 9th
Group Count	9.14*	-0.014
	(4.93)	(0.0088)
Outcome Mean (levels)	7.95	0.03
# Instruments	8	8
N	30957	30957
Effective F-stat	29.88	29.88
Effective F-stat Cutoff 10%	13.26	13.26
Sargan-Hansen Test	20.45	36.31
Sargan-Hansen Test p-val	0.00	0.00
Controls	Yes	Yes

Notes: Unit of observation is a municipality-year. Group count is the number of active DTOs at year t. Column headings indicate the outcomes. The outcomes for the table are the number of homicides, "Homicides" and dropout rates for grade 9 as a fraction relative to enrollment at the start of grade 9 of the same academic year, "Dropout 9th".

Outcome Mean (levels) reports the average number of homicides in (1), and the average drop out rate in (2). # Instruments reports the number of instruments used, these are the group specific lagged distances, in logs, for the 8 criminal groups tracked. N reports the sample size for each regression. "Effective F-stat" reports the Montiel Pflueger effective F-statistic (Olea and Pflueger, 2013). "Effective F-stat Cutoff 10%" reports the critical value for the null hypothesis that the worst case bias of TSLS exceeds 10% of the worst case bias for OLS, with a 5% confidence. "Sargan-Hansen Test' and "Sargan-Hansen Test p-val"' reports the Sargan–Hansen test statistic the p-value. Controls indicates that the regressions include controls. The controls are the inverse hyperbolic sine of municipal agricultural revenue, dummy variables for the presence of each of the 8 criminal groups for the preceding year and the inverse hyperbolic sine of the total number of active criminal groups in the preceding year. All models include year and municipality fixed effects.

Standard errors in parenthesis clustered by municipality. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

the results in Table 3. Furthermore, the complier groups for the different doses are not the same; complier observations for one dose of N_{mt} can be different than complier observations for another dose level of N_{mt}. Therefore in contrast to classical LATE interpretations, the coefficients in Table 3 doesn't hold the complier group constant across different doses of N_{mt}. As a result, the coefficient in Table 3 would be difficult to interpret under the ordered treatment model in Angrist and Imbens (1995).

Additionally, we expect there to be important non-linearities in the treatment effects. These non-linearities are of interest in themselves. The treatment effect of going from a single, monopolist DTO to no DTOs should be much different than going from a DTO duopoly to a DTO monopoly. These non-linearities are important for policy makers and can help guide the deployment of scarce policing resources. For instance, if the largest increase in violence occurs from 1 DTO to 2 DTOs, that can help target areas with a duopoly. Averaging over these non-linearities like in Table 3 is inadequate. Ideally we would like to estimate them directly.

Unfortunately a growing literature has documented that TSLS is ill-suited to estimate effects in scenarios with multiple treatment states and heterogeneous treatment effects (Heckman and Urzúa, 2010; Bhuller and Sigstad, 2022; Mountjoy, 2022a; Kirkeboen et al., 2016). That is exactly the situation we are confronted in this paper. A strategy where I "dummy out" the endogenous variable N_{mt} would lead to uninterpretable estimates, even if the TSLS is just or over-identified. The problem is the TSLS coefficients would be contaminated. For example, the coefficient on the dummy for $N_{mt} = 1$ would not capture the treatment effect of a single DTO to no DTOs. Instead the coefficient will be a mixture other treatment effects as well and include the effect of 2 DTOs relative to 1, or 3 DTOs relative to 1. TSLS is not a good estimation strategy to estimate treatment effects in this setting.

In order to estimate treatment effects, I develop a selection model similar to Kline and Walters (2016); Walters (2018); Rose and Shem-Tov (2021). This selection model will allow me to estimate interpretable treatment effects. Moreover the selection model is consistent with the industrial organization literature studying firm entry and market structure. I develop this model in the next section.

6 Econometric Model

We observe $1...M$ municipalities over $1...T$ time periods. There are $1...G$ criminal groups that can enter or exit each municipality in each time period.[10] The *group count* is the sum of active groups operating in a single municipality. I am interested in estimating different treatment effects for each possible group count, which ranges from $0...G$.[11]. For an outcome Y_{mt}, I denote the vector of potential outcomes as

$$\{Y_{mt}(0), Y_{mt}(1), ..., Y_{mt}(G)\} \tag{6}$$

These potential outcomes span all possible group counts $0...G$. The goal is to estimate different treatment effects of these potential outcomes. Denote the potential outcome for a particular group count n as $Y(n)$. In this paper, I focus on the the Average Treatment Effects (ATEs) of different sizes n relative to no groups operating, $n = 0$. These ATEs are given by

$$E[Y_{mt}(n) - Y_{mt}(0)] \text{ for } n > 0 \tag{7}$$

Observed group counts are denoted by N_{mt}. Observed outcomes are given by

$$Y_{mt} = Y_{mt}(0) + \sum_{n=1}^{G} 1[N_{mt} = n][Y_{mt}(n) - Y_{mt}(0)] \tag{8}$$

6.1 First-Stage Model

I pursue an instrumental variables strategy to correct for endogenous group counts N_{mt}. To do so, I model the first stage as a private information group entry game. This group entry game is consistent with the structural models of entry employed by the industrial

[10] To ease exposition, here I assume the G groups are fixed across time. The case with differing potential entrants across time needs to elaborated elsewhere.

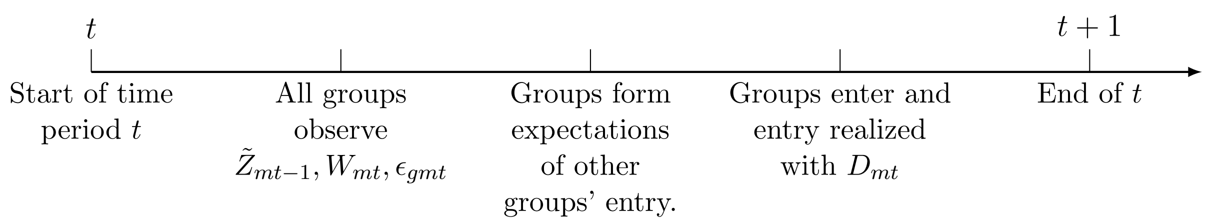
[11] For instance, with two groups, the possible group counts are $\{0, 1, 2\}$

organization literature (Collard-Wexler, 2013; Ryan, 2012; Arcidiacono and Miller, 2011; Aguirregabiria and Mira, 2007; Aradillas-López, 2020).

The environment of the entry game is as follows. For each of the t periods and m municipalities, each group g decides whether to operate or not. Each group g observes a vector of covariates $Z_{mt} = (\tilde{Z}_{mt-1}, W_{mt})$ which is common knowledge to all groups. In my application $\tilde{Z}_{mt-1} = \{\tilde{Z}_{gmt-1}\}_{g=1}^{G}$ are the group specific distances for the previous year, which is common knowledge to all groups. These will be the instruments I use to identify the causal effects of group count in the second stage. Groups also observe the other covariates W_{mt}. These are municipal agricultural output at time t and the log population at time t collected in the vector X_{mt}, incumbency indicators for all groups, (D_{gmt-1}), time dummies (L_t) and municipality dummies (R_m) to control for time-invariant municipality characteristics. For convenience these covariates are included in the vector $W_{mt} = (R_m, L_t, X_{mt}, \{D_{gmt-1}\}_{g=1}^{g})$.

Groups also observe a private profit shock ϵ_{gmt}. Each ϵ_{gmt} is known only to each group g. While each realization ϵ_{gmt} is unknown to the other groups $g' \neq g$, the distribution of ϵ_{gmt} is known to all players. In the context of my paper, the ϵ_{gmt} could represent demand shocks for the different services provided by the criminal groups, or differing managerial ability for the group's local lieutenants at different times. Once groups observe $\tilde{Z}_{mt}, W_{mt}, \epsilon_{gmt}$, they decide on their own entry D_{gmt}. The sequence of the game is summarized in Figure 6

Figure 6: Timeline of Entry Decisions



I place the following assumptions on the profitability shocks and entry decisions:

Assumption 1 (First Stage Assumptions)**.**

FS1 The ϵ_{gmt} are i.i.d. across time, groups and municipalities.

FS2 Groups make entry decisions independently across municipalities.

FS3 $\epsilon_{gmt} \perp \tilde{Z}_{mt-1} \mid W_{mt}, \forall g$

Once groups observe the covariates $\tilde{Z}_{mt-1}, W_{mt}$, they form expectations on competitor's actions and decide whether to enter. The groups' expectations are formed over the distribution of ϵ_g. I make the following high level assumption:

Assumption 2 (Additively Separable Entry Decisions)**.**
Each group g's entry decision, D_{gmt} can be represented as

$$D_{gmt} = 1[v_{gm}(\tilde{Z}_{mt-1}, W_{mt}) > \epsilon_{gmt}]$$

Where v_{gm} is a possibly nonlinear function

This high level assumption nests both static and dynamic games with myopic or forward looking agents.[12] This set up follows most applications of dynamic games found in the literature (Aguirregabiria and Mira, 2007; Collard-Wexler, 2013) and has been extensively studied. Under mild conditions, $v_{gm}(\cdot)$ is identified for each group Pesendorfer and Schmidt-Dengler (2008); Blevins (2014).[13]

6.1.1 Normalizing Probabilities

It is convenient to normalize the probabilities of entry as follows:

$$\begin{aligned} v_{gm}(\tilde{Z}_{mt-1}, W_{mt}) &> \epsilon_{gmt} \\ F_\epsilon[v_{gm}(\tilde{Z}_{mt-1}, W_{mt})] &> F_\epsilon(\epsilon_{gmt}) \\ Q_{gm}(\tilde{Z}_{mt-1}, W_{mt}) &> e_{gmt} \end{aligned}$$

[12] Appendix Section C explains what kinds of entry games are consistent with the model used for estimation.
[13] With forward looking groups, sufficient conditions for identification of v_g would be:

- The distribution of ϵ and the discount factor is known
- Group g's per period profits do not depend on the other groups' incumbency choices, $D_{mt-1,-g}$ where $-g$ excludes group g.
- Per period profits with no entry are normalized to zero.

Where F_ϵ denote the CDF for ϵ. This transformation is innocuous, and it has the benefit that e_{gmt} is now normalized to be uniform [0,1], and the observable utility term, v_{gm} can now be interpreted directly as a probability, which I call Q_{gm}. Q_{gm} is the probability that g enters in m, and it plays a similar role to the propensity score in the treatment effect literature.

6.2 Second-stage

I model the potential outcomes as:

$$Y_{mt}(n) = \psi^n(W_{mt}) + U_{mt}(n) \tag{9}$$

$\psi^n(W_{mt})$ represent mean potential outcomes given W_{mt}. With this representation, the Average Treatment Effects are given by:

$$E[Y_{mt}(n) - Y_{mt}(0)] = E[\psi^n(W_{mt}) - \psi^0(W_{mt})]$$

Where the average is taken over W_{mt}. Hence by estimating ψ^n we can obtain the treatment effects of interest.

Note that equation 9 assumes an exclusion restriction: the $\tilde{Z}$ variables do not enter the potential outcomes equation. In other words, lagged group distances do not affect current outcomes directly. In addition to this exclusion restriction, I employ a mean independence assumption to allow for identification of causal effects.

Assumption 3 (Second Stage Assumptions)**.**

SS1 $U_{mt}(n) \perp \tilde{Z}_{mt-1} \mid W_{mt}$

SS2 $E[Y_{mt}(n)|W_{mt}, \mathbf{e}_{mt}] = \psi^n(W_{mt}) + E[U_{mt}(n)|\mathbf{e}_{mt}]$

Where $\mathbf{e}_{mt} = \{e_{gmt}\}_{g=1}^{G}$

Assumption SS1 is an exogeneity assumption typical in instrumental variable models. I assume once we control for W_{mt}, the distance instrument is independent of the unobserved component of potential outcomes, $U(n)$. Assumption SS2 is the same assumption employed in Mogstad et al. (2018); Brinch et al. (2017). The main substance of this assumption is that $E[U_{mt}(n)|\mathbf{e}_{mt}]$ does not depend on any of the covariates W_{mt}.

It is worth discussing the implications of Assumption SS2 for D_{mt-1}, which is part of W_{mt}. Assumption SS2 is not saying D_{mt-1} is uncorrelated with $U_{mt}(n)$. Rather, its saying that once we control for the first stage errors e_{mt}, $U_{mt}(n)$ is mean independent of D_{mt-1}. As a result, this model assumes past entry affects observed outcomes at time t only through two ways: 1) directly through ψ^n and 2) by implicitly changing entry probabilities for time t. The same logic applies to other covariates in W_{mt}.

6.3 Two Group Example

To see how these assumptions attain identification, consider a simple example with only two groups, $G = 2$. Denote by $Q(\tilde{Z}_{mt}, W_{mt})$ the vector of entry probabilities for each group, $Q_m(\tilde{Z}_{mt}, W_{mt}) = \{Q_{1m}(\tilde{Z}_{mt}, W_{mt}), Q_{2m}(\tilde{Z}_{mt}, W_{mt})\}$. The mean regression of Y_{mt} on W_{mt} and the first stage entry probabilities, $Q_m(\tilde{Z}_{mt}, W_{mt})$ for the subset of observations with one group operating, $N_{mt} = 1$ is given by:

$$
\begin{aligned}
&E[Y_{mt}|Q_m(\tilde{Z}_{mt}, W_{mt}) = q, W_{mt}, N_{mt} = 1] = \psi^n(W_{mt}) \\
&+ E[U(1)|q_1 > e_1, q_2 < e_2]P[q_1 > e_1, q_2 < e_2|\{q_1 > e_1, q_2 < e_2\} \cup \{q_1 < e_1, q_2 > e_2\}] \\
&+ E[U(1)|q_1 < e_1, q_2 > e_2]P[q_1 < e_1, q_2 > e_2|\{q_1 > e_1, q_2 < e_2\} \cup \{q_1 < e_1, q_2 > e_2\}]
\end{aligned}
\tag{10}
$$

Where q_1, q_2 are the estimated entry probabilities for group 1 and group 2, respectively, and e_1, e_2 are the normalized first stage errors. This representation exploits the exclusion restriction, Assumption 3, and the first stage model of entry. [14]

The probability terms are estimated using the first stage model. For instance, the term $P[q_1 < e_1, q_2 > e_2|\{q_1 > e_1, q_2 < e_2\} \cup \{q_1 < e_1, q_2 > e_2\}]$ is just the conditional probability that group 1 enters but 2 does not, conditional on one of the two groups entering.[15] ψ^n can

[14] Recall entry occurs if $e_g < q_g$

[15] This probability is just $\frac{q_1(1-q_2)}{q_1(1-q_2)+(1-q_1)q_2}$

be estimated after imposing an appropriate functional form. As a result, we only need to control for the unknown variables

$$E[U(1)|q_1 > e_1, q_2 < e_2], E[U(1)|q_1 < e_1, q_2 > e_2]$$

which boils down to a control function approach. I estimate the parameters of the outcome equation 10 using a two-step control function approach, similar to a classic Heckman model (Heckman, 1979). I first estimate the entry probabilities for each of the groups and collect $Q_m(\tilde{Z}_{mt}, W_{mt})$. I then run a second step regression using these probability terms and control for $E[U(1)|q_1 > e_1, q_2 < e_2], E[U(1)|q_1 < e_1, q_2 > e_2]$.

6.4 Generalization to G Groups

Generalizing to the case with G groups is straightforward. First, I establish some notation for clarity. Define the random variable $D_{mt}(q)$ with support $\mathcal{D} = \{0,1\}^G$. $D_{mt}(q)$ spans all possible entry configurations and depends on the G players' entry probabilities $q = \{q_g\}_{g=1}^G$. For a particular value $d \in \mathcal{D}$ set:

$$1[D_{mt}(q) = d] := 1[\{e_{gmt}\}_{g=1}^G \in \{[0,1]^G | D_{mt} = d, Q_{mt}(Z) = q\}] \tag{11}$$

That is, $D_{mt}(q) = d$ implies the vector of first stage errors $\{e_{gmt}\}_{g=1}^G$ is in the region of $[0,1]^G$ where given the entry probabilities $Q_{mt}(Z) = q$, the entry vector d can be realized. For example, in the earlier two player example, the event $1[D_{mt}(q) = (1,0)]$ corresponds to $1[q_1 > e_{1mt} \cap q_2 < e_{2mt}]$, i.e. $e_{1mt} \times e_{2mt} \in [0, q_1] \times [q_2, 1]$. $D_{mt}(q) = d$ is simply a compact way to denote this region for more than two players.

Similarly, define $N(q)$ as the random variable representing the total number of groups who enter given entry probabilities q. That is

$$N_{mt}(q) := \sum_{g=1}^{G} D_{gmt}(q) \tag{12}$$

Where $D_{gmt}(q)$ is the g component of $D_{mt}(q)$. Now again consider the mean regression of Y_{mt} on W_{mt} and the first stage entry probabilities, $Q_m(\tilde{Z}_{mt}, W_{mt})$:

$$\begin{aligned} &E[Y_{mt}|Q_m(\tilde{Z}_{mt}, W_{mt}) = q, W_{mt}, N_{mt}(q) = n] \\ &\qquad = \psi^n(W_{mt}) + \sum_{d \in \mathcal{N}(n)} E[U_{mt}(n)|D_{mt}(q) = d] P[D_{mt}(q) = d|N_{mt}(q) = n] \end{aligned} \tag{13}$$

Where the set $\mathcal{N}(n) := \{d \in \mathcal{D} | \sum d_g = n\}$ contains all the entry configurations d that yield a group count n. Similar to the two player case, the only unknown quantity is $E[U_{mt}(n)|D_{mt}(q) = d]$. Classic control function estimators will then deliver estimates of ψ^n, similar to a Heckman (1979) style procedure.

7 Identification and Estimation

In this section I explain the functional forms I use to identify and estimate the model. I also explain how I account for multiple equilibria across municipalities, and how I conduct inference.

7.1 Time Invariant Effects and Multiple Equilibria

Municipalities are different due to time-invariant characteristics, and conditioning on these differences is central to my identification argument. Typically, most empirical studies model this unobserved heterogeneity with a simple additive effect for each municipality.[16]

In this setting modeling this heterogeneity as an additive effect is unsatisfactory for three main reasons. First, additive effects are difficult to motivate in an entry game model Aguirregabiria and Mira (2019). Even if the groups' underlying flow profits have an additive effect by municipality, the non-linearities induced by the entry game ultimately mean the function v_{gm} will not contain an additive effect for each m.[17] As a result, a simple additive effect is fundamentally at odds with my first stage model of entry.

[16] For example, consider the fixed effect model $D_{gmt} = \phi_m + \delta \tilde{Z} + \epsilon_{gmt}$. The effect ϕ_m is an additive effect.
[17] Section C.1 explains this in more detail.

Second, these private information games can exhibit multiple equilibria (Aguirregabiria and Mira, 2019; Aradillas-López, 2020), even if agents are not forward looking. Effectively multiple equilibria imply that the function v_{gm} varies across m. As a result, a change in $\tilde{Z}_{mt}$ may have a different effect on the entry probabilities, depending on which municipality (and equilibrium) is considered. A simple additive effect would fail to account for this possibility.

Last, estimating additive effects is notoriously difficult in binary choice models due to the incidental parameters problem (Neyman and Scott, 1948). As a result, even if an additive effect were appropriate in my setting, estimating it would not be trivial.

To control for time-invariant differences in the first stage, I instead employ the group fixed effect estimation strategy developed in Bonhomme et al. (2022); Saggio (2012). The grouped fixed effect framework is a tractable procedure that allows for a richer model than a simple additive effect, and avoids the incidental parameter problem of Neyman and Scott (1948).

The substance of the procedure involve using k-means clustering (Steinley, 2006) to group municipalities into C groups, with $C < M$. Therefore this strategy reduces the dimensionality of the M fixed effects into a lower C dimensional grouping. The researcher only needs to specify the number of clusters C and which covariates to use in the clustering algorithm. My baseline model includes 5 clusters, though the results are robust to increasing the clusters.

Following Bonhomme et al. (2022) I use the average number of years each municipality is observed with a group count of n, that is $\frac{1}{T}1[N_{mt} = n]$. I create this variable for $n = 0...8$, since the maximum group count is 8. As a result, the clustering algorithm groups municipalities that had similar histories of group count. Places with many DTOs will be clustered with other municipalities with many DTOs as well. Intuitively, these municipalities should be particularly profitable, since they attract so many entrants. The same logic applies for municipalities with fewer entrants; those municipalities should be less profitable to enter. I denote $C(R_m)$ as a function which maps which of the C groups each municipality m belongs to. R_m is a vector of dummies indicating the observation's municipality. As a short-hand, I refer to the $C(R_m)$ groups as *municipality types*.

An important benefit of this approach is it allows for models of multiple equilibria. In particular, my main specification includes interactions of $C(R_m)$ with some of the covariates in $\tilde{Z}_{mt}, W_{mt}$. As a result, my estimates are robust to both time-invariant differences across municipalities and to the presence of multiple equilibria.

7.2 First-Stage Estimation

I pool all of the G groups entry decisions and estimate the following probit for the first stage.

$$\begin{aligned} D_{gmt} = \xi + \underbrace{\beta_1^1 X_{mt}}_{\text{Controls}} + \underbrace{\phi_{0g}^1 1[G=g]}_{\text{DTO Fixed Effect}} + \underbrace{\phi_1^1 L_t + \sum_{c=1}^{C} 1[C(R_m)=c] \times \phi_{2c}^1}_{\text{Time and Municipality Type Fixed Effects}} \\ + \underbrace{\sum_{c=1}^{C} 1[C(R_m)=c] \times \left[\gamma_{1c}^1 D_{gmt-1} + \gamma_{2c}^1 \times IHS(N_{mt-1})\right]}_{\text{Municipality Type Specific Controls}} \\ \sum_{c=1}^{C} 1[C(R_m)=c] \times \underbrace{\left[\gamma_{3c}^1 \tilde{Z}_{gmt-1} + \gamma_{4c}^1 \tilde{Z}_{gmt-1}^2 + \gamma_{5c} \overline{\tilde{Z}}_{-gmt-1} + \gamma_{6c}^1 \overline{\tilde{Z}^2}_{-gmt-1}\right]}_{\text{Municipality Type Specific Instruments}} + \epsilon_g \end{aligned} \tag{14}$$

The unit of observation is a DTO $\times$ municipality $\times$ year. X_{mt} is the IHS of agricultural value and log population at time t. D_{gmt-1} is group g's incumbency status. $\phi_1^1 L_t$ are time effects and $\sum_{c=1}^{C} 1[C(R_m) = c] \times \phi_{2c}^1$ are municipality type fixed effects given by $C(R_m)$, which supplant traditional municipality fixed effects. $IHS(N_{mt-1})$ is the inverse hyperbolic sine of lagged group count. $\tilde{Z}_{gmt-1}$ is the log of group g's distance to m for the preceding year; $\tilde{Z}_{gmt-1}^2$ is its square. $\overline{\tilde{Z}}_{-gmt-1}$ is the average of the log distance for all other groups $j \neq g$; $\overline{\tilde{Z}^2}_{-gmt-1}$ is the average of the square distances.

Importantly, I allow for the instruments, lagged group count and lagged presence to vary according to municipality type. As mentioned before, this accommodates both multiple equilibria and and time-invariant heterogeneity in the entry game. Moreover information-criteria approaches support including interactions by market types.[18]

7.3 Second-Stage Estimation

Using the previous first stage model I estimate the entry probabilities for all players and collect these in $Q_{mt}(Z)$.

I parametrize the potential outcomes to be of the form:

$$Y_{mt}(n) = \underbrace{\alpha_n}_{\text{Treatment Effect}} + \underbrace{\beta_1^2 X_{mt}}_{\text{Controls}} + \underbrace{\phi_1^2 L_t + \phi_2^2 R_m}_{\text{Time and Muni. FE}} + \underbrace{\sum_{c=1}^{C} 1[C(R_m) = c] \times \left[\sum_{g=1}^{G} \gamma_{1g}^2 D_{mtg-1} + \gamma_{2c}^2 IHS(N_{mt-1}) \right]}_{\text{Municipality Type Specific Controls}} + U_{mt}(n) \tag{15}$$

This baseline specification allows for a separate average mean effect for each potential outcome (given by α_n) but restricts there to be no heterogeneity in the other covariates. All the variables present in the first stage are present in the second stage, except for the instruments. In particular notice that the all G of the incumbency indicators are included as separate variables in the second stage, since these were included in the first stage.[19] Note additionally that I do not restrict the municipality fixed effects in the second stage. Equation 15 has M fixed effects captured in $\phi_2^2 R_m$. This is because the model is linear and these additional fixed effects do not cause substantive problems. This follows Cornelissen et al. (2018) in modeling fixed effects for this kind of estimation.

7.4 Control Function

The last step is to control for the $E[U(n) - U(0)|D(q) = d]P[D(q) = d]$ terms in equation 13. As a benchmark, I assume a normal model as in Heckman (1979); Kline and Walters

[18] In results available upon request, a LASSO based approach for this probit includes all the terms included in this regression.

[19] For this reason I opted to not interact the incumbency indicator by municipality type. With 5 municipality types and 8 groups, this would have implied 40 additional covariates in the second stage.

(2016); Dahl (2002). The normal model is familiar to most readers and therefore provides a useful benchmark.

I assume the vector $(\{U(n)\}_{n=0}^{G}, \{\epsilon_g\}_{g=1}^{G})$ follows a normal with parameters $N(0,\Sigma)$. The variances of $\{\epsilon_g\}$ are normalized to 1, and are independent across g according to Assumption FS1. Let σ_{ng} be the covariance between $U(n)$ and ϵ_g. Using multivariate truncated normal results from Tallis (1961) and the independence between ϵ_g, we obtain:

$$E[U_{mt}(n) - U_{mt}(0)|D_{mt}(q) = d]P[D_{mt}(q) = d] = \sum_{g=1}^{G}(\sigma_{ng} - \sigma_{0g})\lambda(d_g, q_g) \tag{16}$$

where d_g, q_g are the g$^{\text{th}}$ components of an entry vector $d \in \mathcal{D}$ and the vector of entry probabilities q, and λ is a known function.[20] The normality assumption provides a convenient functional form to control for $E[U_{mt}(n) - U_{mt}(0)|D_{mt}(q) = d]P[D_{mt}(q) = d]$.

To create the final estimating equation I exploit that observed outcomes are given by $Y_{mt} = Y_{mt}(0) + \sum_{n=1}^{G}\{Y_{mt}(n) - Y_{mt}(0)\}1[N_{mt} = n]$, where N_{mt} is the observed group count. Now consider the regression of Y_{mt} on entry probabilities $Q(\tilde{Z}_{mt}, W_{mt})$ and covariates W_{mt}

20

$$\lambda(d_g, q_g) = \begin{cases} -\phi(\Phi^{-1}(q_g))\prod_{j\neq g} h(q_j, d_j) \text{ if } d_g = 1 \\ \phi(\Phi^{-1}(q_g))\prod_{j\neq g} h(q_j, d_j) \text{ if } d_g = 0 \end{cases}$$

Where ϕ, Φ^{-1} are the pdf and the inverse cdf of the standard normal distribution. $h()$ is given by

$$h(d_j, q_j) = \begin{cases} q_j \text{ if } d_j = 1 \\ 1 - q_j \text{ if } d_j = 0 \end{cases}$$

So $\prod_{j\neq g} h(q_j, d_j)$ is the probability that all players $\neq g$ behave according to the entry vector d.

This yields the following estimating equation:

$$
\begin{aligned}
E[Y_{mt}|Q(Z)=q,W] = \beta_1^2 X_{mt} + \sum_{g=1}^{G} \beta_{2g}^2 D_{mtg-1} \\
+ \phi_1^2 L_t + \phi_2^2 R_m + \sum_{c=1}^{C} \gamma_{1c}^2 1[C(R_m)=c] \times IHS(N_{mt-1}) \\
+ \sum_{n=0}^{G} \left\{ (\alpha_n - \alpha_0) P[N(q)=n] + \sum_{g=1}^{G} (\sigma_{ng} - \sigma_{0g}) \sum_{d \in \mathcal{N}(n)} \lambda(d_g, q_g) \right\}
\end{aligned}
\tag{17}
$$

Notice that equation 17 is linear in the unknown parameters and can therefore be estimated with simple OLS. Notice as well the first stage probabilities preform two important roles. First they allow calculation of $P[N(q) = n]$. Given the individual groups first stage probabilities and the independent errors between groups, this term is a simple function of the individual group probabilities.[21] Second the probabilities q are used to construct the control function $\lambda(d_g, q_g)$.

The estimation procedure is summarized as follows. I first apply k-means clustering to group municipalities. I then first estimate individual group entry probabilities using the probit in equation 14. I then use these probabilities to construct the appropriate terms in equation 17 and run OLS.

7.5 Inference and Sample Restrictions

In order to gain precision, I pool together observations with 5 or more DTOs into the same treatment category. There are relatively few observations with 5 or more DTOs. Only 3.2% of the observations have 5 or more DTOs. The individual coefficients for these large group counts are necessarily more noisy and less informative. [22]

Additionally, as a sample restriction, I drop observations where the unit has a probability of 1 for one of the possible group counts. These observations are, according to my model,

[21]For instance with two groups, $P[N(q) = 1] = q_1(1 - q_2) + (1 - q_1)q_2$

[22]I show in Appendix Figure A4 that the results are robust to allowing for more treatment states than my benchmark 5.

not at risk to receiving any of the counterfactual group counts considered. They therefore are poor control units when trying to estimate treatment effects. I drop 4168 observations with this restriction. These are mainly small municipalities that are never observed with any DTO. As a result these observations are not informative of counter factual outcomes with a non-zero group count.

As mentioned previously, some of the DTO groups did not exist for the entirety of my sample period, 2006-2018. I handle these cases by setting the DTO's probability of entry to zero, since the DTO could not have possibly entered. This additionally means that their contribution to the control function term is also zero.

Last, inference is complicated due to the k-means clustering and the first stage probability estimation used in the second stage. To facilitate inference I apply a block Bayesian bootstrap (Rubin, 1981; Hull, 2022) for inference. For each municipality, I simulate 1000 weights from a Dirichlet distribution with parameter 1. Each simulated weight is applied for all the time periods corresponding to a single municipality. For one simulation draw, I rerun the entire estimation using the draw as a weight for each municipality: I estimate the clusters with k-means, I estimate the first stage probabilities with probit, and then I estimate the second stage equation with OLS. The Dirichlet weights randomly place more weight on some municipalities over others across simulation draws. This mimics a classical bootstrap, where municipalities that are sampled more effectively receive more weight during estimation. The main benefit of this procedure is no observations are ever dropped, which is important for the fixed effect OLS in the second step.

8 Results

8.1 First Stage

Table 4 displays results of Equation 14 in column 1, estimated with probit. Other variations of the same equation are displayed in the other columns as robustness checks. The table displays estimates of average partial effects (APEs) for the different parameters, so we can interpret the magnitudes in probability space. The row "Own Distance" displays APEs of the effect of a DTO's own distance on the probability it enters. The row "Avg. Rival" distance instead is the the APE for the average of the rival groups' distances. These estimates have the anticipated signs: the effect of a DTO's own distance is negative, indicating the further away the DTO was, the less likely they were to enter. In contrast, the effect of rival's distance is positive; the further away the rivals are, the more likely a DTO is to enter. In column 1, a 1% increase in a DTO's own distance leads to a 1 percentage point (pp) increase in the likelihood of entry, while a 1% increase in the average of rival distances decreases entry by 1 pp. This pattern is consistent with DTOs behaving as strategic substitutes.

Lag Presence has a positive effect on DTO presence, as we would expect. Incumbency strongly predicts future presence; in column 1 a DTO is 61.6 pp more likely to remain in a municipality if it was there before. Lag DTO count has a negative though insignificant coefficient of -0.002. This again is consistent with DTOs behaving as strategic substitutes.

Columns 2 and 3 show versions of Equation 14 with cubed and linear terms for the distances only. We can see the estimates are almost identical. Column 4 excludes all controls for municipality type. In this specification the effect of Lag DTO count has a positive estimate of 0.024. This suggests that the groupped fixed effect strategy matters for correctly modeling the entry decisions, and that unobserved heterogeneity matter.

Both BIC and the AIC agree the main model in column 1 provides a good fit. In particular, the model in column 5 which excludes municipality type interactions is not favored by the BIC: excluding market type interactions yields a BIC of 41388, but including these interactions yields a much lower BIC of 40551.

8.2 Second Stage Estimates

I first examine outcomes on homicides and dropout rates for 9th grade. As in Section 5 I examine the count of homicides. Dropout is measured as the fraction of students who dropout in 9th grade, relative to the how many students were enrolled at the start of the academic year. The results are shown in Figure 7. Each point is an estimated coefficient

Table 4: First Stage Probit Results

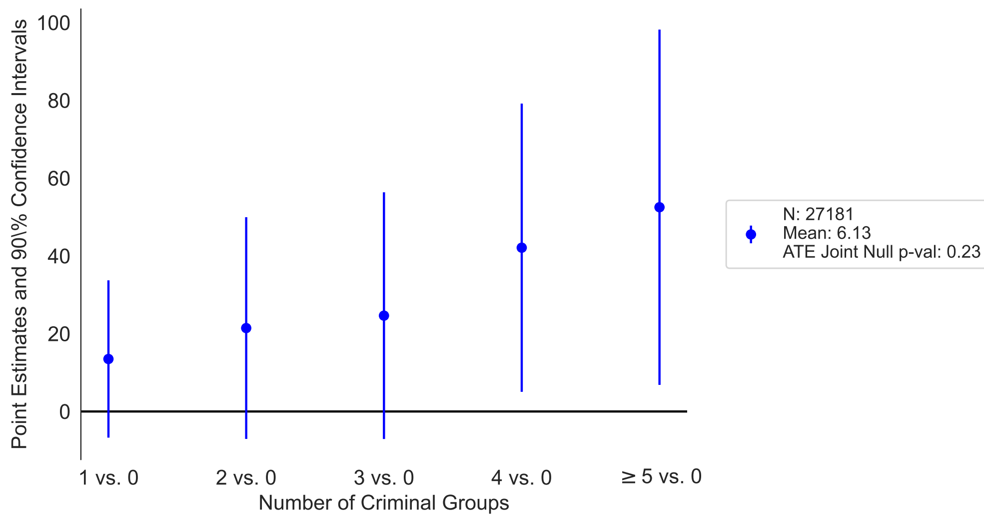
	(1)	(2)	(3)	(4)	(5)
IHS(Ag. Value)	0.000	0.000	0.000	0.001***	0.001**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log(Pop.)	0.004***	0.004***	0.004***	0.009***	0.004***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Lag Presence	0.616***	0.617***	0.655***	0.625***	0.603***
	(0.145)	(0.147)	(0.152)	(0.007)	(0.007)
IHS(Lag DTO Count)	-0.002	-0.001	-0.003**	0.024***	-0.008***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Own Distance	-0.010***	-0.011***	-0.007***	-0.010***	-0.010***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Avg. Rival Distance	0.010***	0.011***	0.006***	0.007***	0.010***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Muni. Type FE	✓	✓	✓		✓
Muni. Type Interaction	✓	✓	✓		
Avg. Dist.	✓	✓	✓	✓	✓
Max Power	2	3	1	2	2
BIC	40551	40526	40834	46059	41388
AIC	41489	41671	41566	46348	41769
Log-Likelihood	-20185	-20152	-20346	-23002	-20657
Model d.f.	90	110	70	27	36
N	221497	221497	221497	221497	221497

Notes: Average partial effects (APEs) of a probit model. For variables that are interacted with municipality type, the APEs are averaged across the interactions.
The estimation stacks entry decisions for all 8 DTO groups across municipality and years. If a DTO hasn't been formed in a certain year, all the observations for that DTO-year are dropped.
All estimates include DTO and year fixed effects. Columns that use municipality type controls have 5 municipality types.
Column 1 estimates the pooled first stage equation 14. Columns 2 adds a cube term to both the own and avg. rival distance. Column 3 drops both the quadratic and cubed distance terms. Column 4 excludes Municipality Type fixed effects and the interactions. Column 5 Includes Municipality Type fixed effects but excludes the interactions.
Muni. Type FE indicates if regression includes municipality type fixed effects. Muni. type interactions indicates if the municipality types are interacted with the variables in equation 14. Max power shows the highest polynomial power used in the distance terms. BIC is the Bayesian Information Criterion, AIC is the Akaike Information Criterion. Model d.f. is the number of parameters in the model, N are the number of observations.
Unit of observation is a DTO-municipality-year. Sample years are 2006-2018.
Standard errors from 1000 bootstrap replications.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

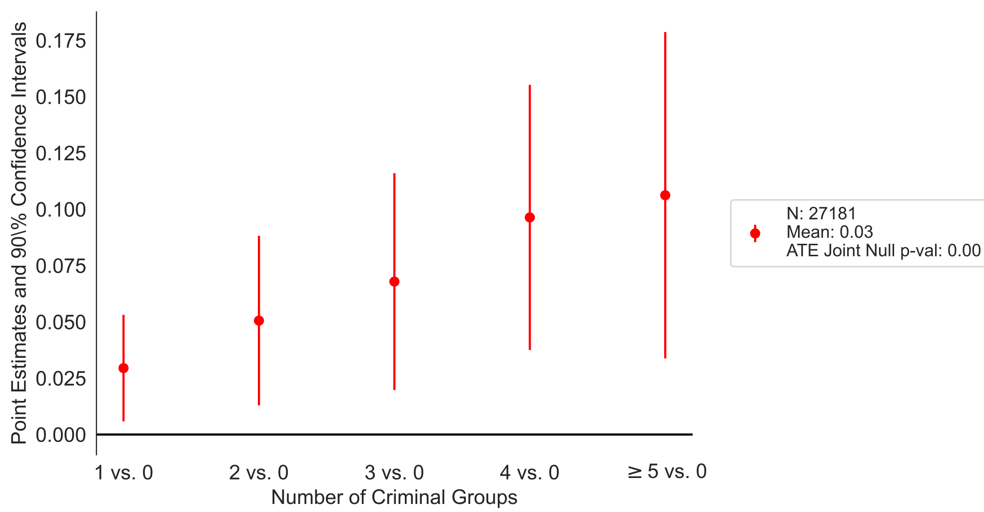
along with its 90% confidence interval resulting from the 1000 bootstrap replications. All the coefficients show the ATE relative to having zero DTOs, that is $E[Y(n) - Y(0)]$.

Figure 7: Main Results

(a) Homicides



(b) Dropout 9th Grade



Notes: Standard errors from 1000 bootstrap replications.

In the top panel we observe a clear upwards trend in the count of homicides : homicides tend to increase as more DTOs enter. Moving from 5 DTOs to 4 DTOs results in a larger increase in homicides than moving from 2 DTOs to only 1 DTO. The coefficients on both 4 and 5 DTOs is significant at the 90% confidence level. The coefficient for 4 DTOs is 33.4 additional homicides and for $\geq$5 DTOs it is 47.2 extra homicides. There are large effects relative to the overall mean of 5.98 homicides in the estimation sample. I interpret the increase in homicides to reflect increased violence between criminals.

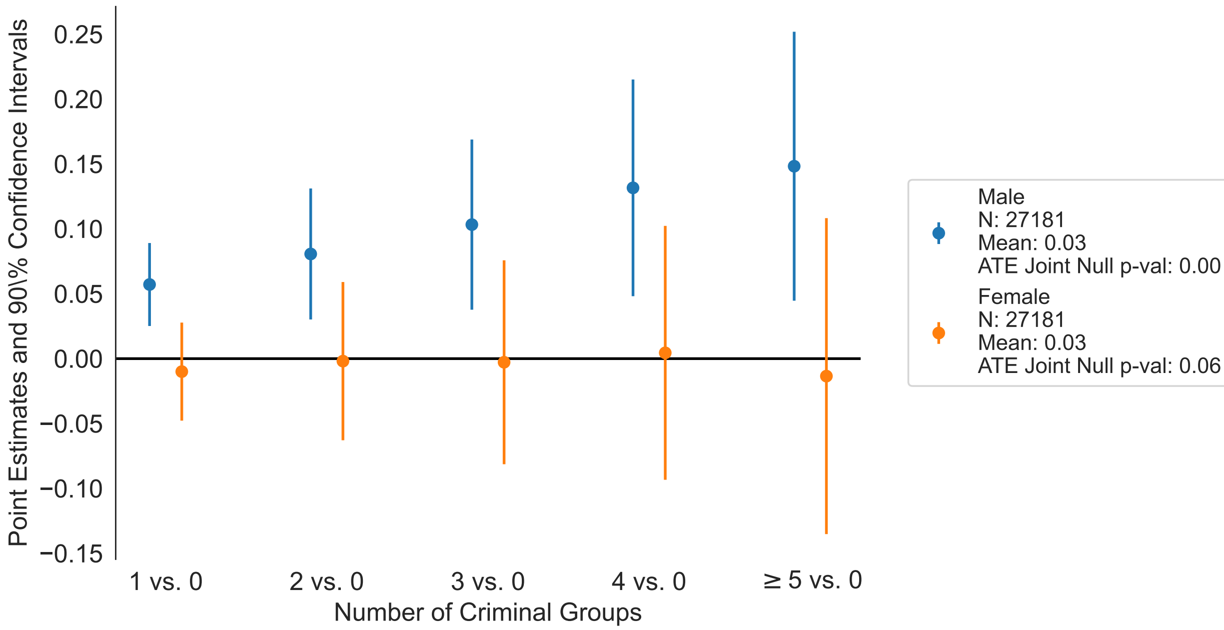
The coefficients for 1 to 3 DTOs are insignificant, but the point estimates still reveal important increases in homicides. The coefficient for 1 DTOs is 2.6, for 2 DTOs is 9.8 and for 3 DTOs is 15.9 additional homicides. The overall mean in the estimation sample is 6 homicides per municipality-year. Hence the coefficients for having 2 and 3 DTOs translates to doubling and tripling the average number of homicides in the estimation sample. Even if the estimates are noisy, the point estimates suggest an important increase in homicides with having 2 DTOs relative to only a single DTO. Overall, I also reject the null hypothesis that all the coefficients are zero. The p-value for this joint test is 0.01.

In contrast, the effects are linear and more pronounced for 9th grade dropout. Each coefficient is statistically significant at the 90% confidence level. I confidently reject the null that all the coefficients are zero, with an effective p-value of 0. Moreover the point-estimates are quite large. Having 5 or more DTOs relative to zero increases 9th grade dropout by more than 10 percentage points. This is relative to an overall mean of 0.03, or 3%. Although large, these effects are not implausible. The implied dropout rate with this treatment effect would be roughly 15%. Dropout during grade 10, the first year of high school in Mexico, is 25% (INEE, 2015). As a result, these results suggest some students are accelerating their dropout decisions for the first year of high school.

Figure 8 examines how these effects vary by the gender of the students. The figure displays effects separately for boys and girls. We can see boys are predominately affected. The point estimates for girls are always insignificant, and much close to zero than for boys. The effects for boys are statistically significant, positive and large. Although there is overlap between the two sets of estimates, the estimates indicate that boys are predominately affected.Male dropout drives the overall 9th grade dropout observed earlier.

The gendered effects also indicate the effects are likely not driven by migration. The

Figure 8: 9th Grade Dropout by Gender



dropout outcomes count a student that moved to a different municipality as dropping out, when in fact the student may enroll in a different municipality. As a result, the positive effects may just be a migration response. However, the gendered response makes this migration story unlikely. I claim that such a migratory response would have to be severely gendered to explain the pattern of results I observe. I arrive to the same conclusion later in the paper by examining earlier grades and using other data sources.

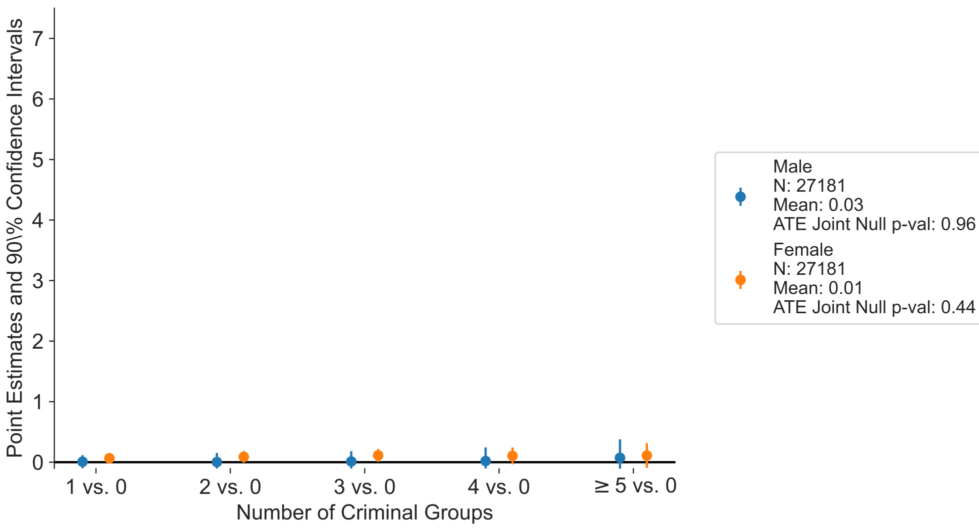
Because criminal activity is male-dominated (Carvalho and Soares, 2016), the pattern of results suggests a main driver of the dropout effects are DTOs pulling students out of middle school and employing them in criminal activity. Testing this mechanism is crucial yet difficult; tracking student achievement and criminal outcomes is difficult.

Nonetheless, the homicide data can help us disentangle if DTOs are indeed hiring middle-school dropout. Using demographic information on the homicide victims, I examine homicides for both males and females of two different age groups: 12-14 and 15-18. Teenagers aged 12-14 should still be in middle school, while 15-18 should have exited middle school.

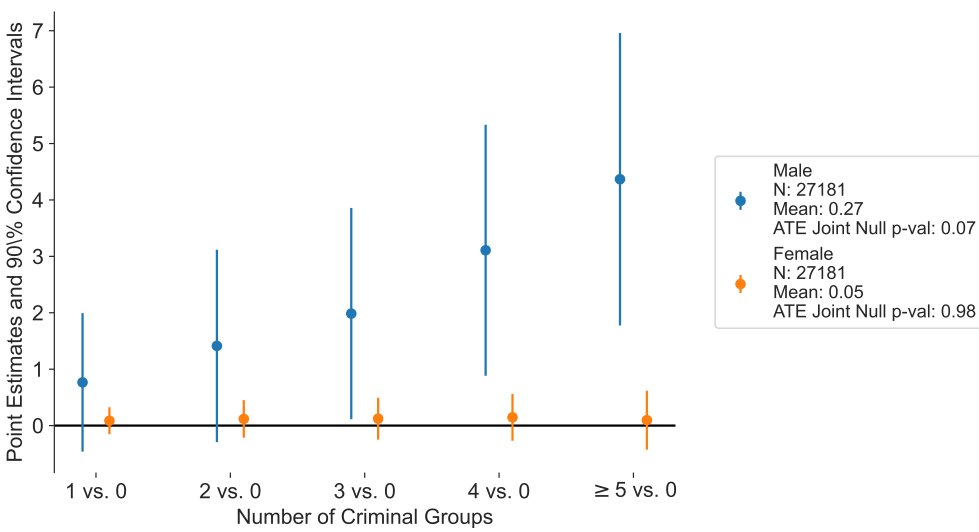
The results in Figure 9 show that 15-18 year old males see a strong increase in homicides

Figure 9: Homicides Ages 12-18

(a) Homicides Ages 12-14



(b) Homicides Ages 15-18



Notes: Standard errors from 1000 bootstrap replications.

relative to the other demographic groups. The top panel displays effects for 12-14 year olds, on the same scale as the bottom panel which displays effects for 15-18 year olds. We can see there is essentially no effect for 12-14 year olds, either female or male. Instead the bottom panel shows an important increase in homicides for 15-18 year old males but no effect for females. This pattern of results is consistent with DTOs attracting young males out of middle school and the dropouts climbing a criminal career ladder. In fact Carvalho and Soares (2016) document that gang members involved in violent conflict (soldiers) typically are more experienced, suggesting soldiers typically started off with a less demanding entry level position in the criminal organization.

8.3 Dropout Effects for Other Grades

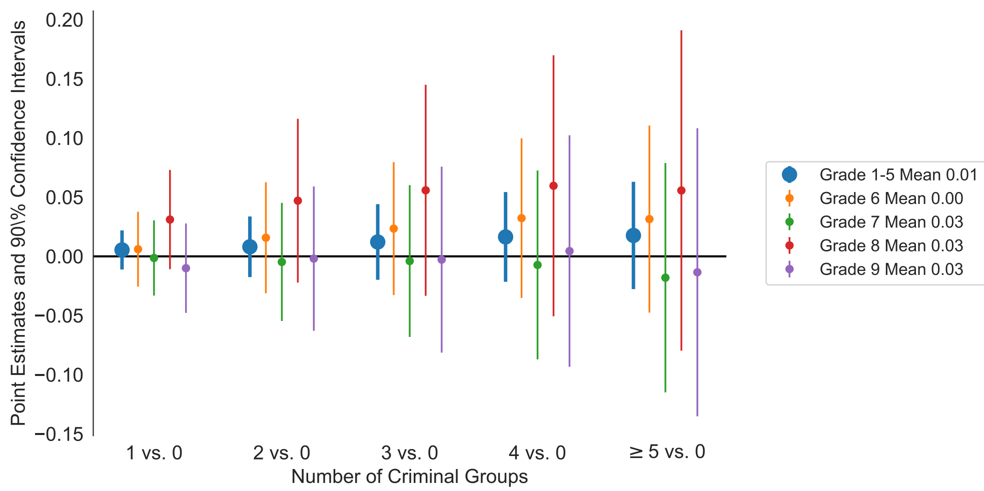
So far the results have focused on 9th grade dropout because it determines middle school completion. However examining other grades is informative for two reasons. First, if dropout is driven by DTO's hiring young males out of school then we expect effects to be more pronounced for older boys. Second, examining younger grades is a helpful placebo test. Earlier grades can help determine if the effects are driven by outmigration. Dropout in early grades in Mexico is virtually non-existent. Therefore, if I find pronounced effects for early grades it likely means the dropout I observe is in reality a migration response.

I reestimate my main specification for grades 6,7,8 and 9. I also group dropout for grades 1-5 into a single outcome; I do so for parsimony and to limit the number of estimates shown on the plots. Results can be seen in Figure 10. The top panel reports effects for females while the bottom for males. The estimates are categorized according to the DTO count on the x-axis. The left-most estimate is for grades 1-5, and the grades increase as you go to the right, ending in grade 9.

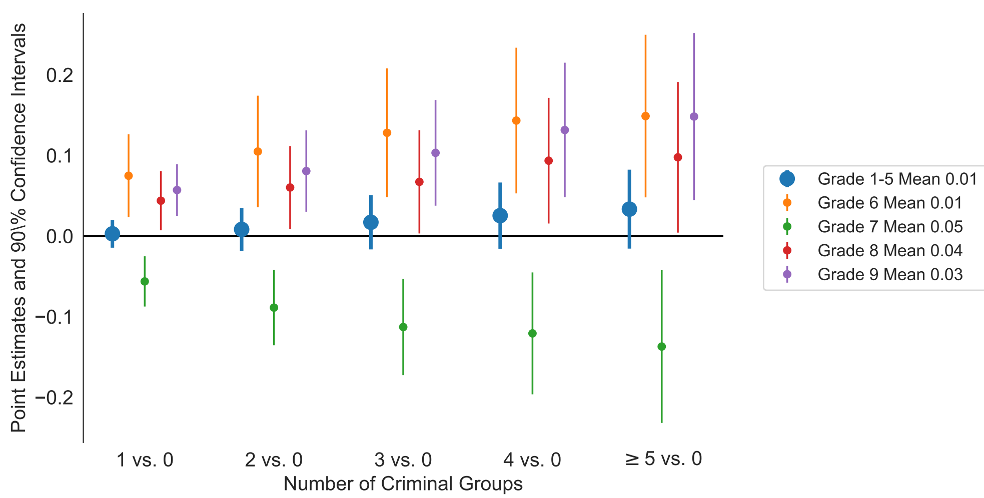
The top panel shows that female students are largely unaffected by DTO presence in grades below 9th grade, though there are large and statistically insignificant effects for grade 8. Moreover the effects for grades 1-5 (presented in thick lines) are also small and insignificant. Overall, I do not find evidence that females are affected by DTO entry, although the

Figure 10: Dropout by Grade and Gender

(a) Dropout for Female Students



(b) Dropout for Male Students



Notes: Standard errors from 1000 bootstrap replications.

standard errors are large.

The effects for males across grades require more nuance to interpret. First, we again see small and insignificant effect for grades 1-5. This is encouraging, since it again suggests migration is not driving the response to DTO presence. However, we see some puzzling results for grades 6 and 7. Grade 6 has large positive effects while grade 7 estimates are negative and of roughly the same magnitude of grade 6.

What is the interpretation of the opposing grade 6 and 7 effects for males? For this we must recall two things. First, grades 6 and 7 are transition grades in Mexico: grade 6 is the last grade of elementary school while grade 7 is the first of middle school. Second the dropout measure I use is intra-curricular dropout, that is dropout between the start and end of the same academic year. The negative effects for grade 7 indicate that there is "negative" dropout, or put differently, there are more students at the end of the academic year than there were at the beginning of grade 7. Similarly the positive effects of grade 6 indicate there is regular dropout, i.e. less students at the end of grade 6 than at the beginning of grade 6. Put together, these estimates suggest male students are pulled out before the completion of elementary school, and are enrolled in middle school after the school year begins. This indicates that DTOs cause a friction for households with male middle schoolers as they transition from elementary to middle school. It is difficult to pin down what friction may be driving the result; one possibility is that parents of male students are more concerned as DTOs presence increase. The increased concern may translate to an earlier culmination of elementary school and a prolonged search for a middle school.

Nonetheless notice that the negative 7th grade results again rule out migration as an important concern: it is odd to think grade 7 specifically would receive a large influx of migrants as more DTOs are present, while simultaneously pushing grade 8 or 9 students to other municipalities.

Grade 8 dropout displays similar patters as grade 9; positive and statistically significant effects. However each coefficient is smaller in magnitude relative to the grade 9 estimates. Naturally the differences between grade 8 and 9 are too small to be statistically significant.

Nonetheless the pattern suggests more pronounced effects for older students, just as we would expect if some of the students dropping out are being employed by the DTOs.

8.4 Non-Linearities in Homicides

Though the main estimates do not seem to display any strong non-linearities, Appendix Figure A4 displays interesting patterns when we consider more than 5 DTOs. The bottom panel reports estimates for 9th grade dropout. These estimates appear to be roughly linear. The top panel displays effects for homicides. We can observe there is a sharp increase in homicides with 6 DTOs, which later tapers of with 7 and 8 DTOs. Although the standard errors are wide, the point estimates suggest an 'S' shape curve with an inflection point at 6 DTOs.

8.5 Robustness Checks

I conduct several robustness checks to validate the main findings.

8.5.1 Migratory Response

I leverage individualized ENLACE test score data to directly measure out-migration over time. Using the individualized test data, I am able to link students over time and determine if they move to another municipality. Results are in Figure A3. This data is available only for the years 2009-2012, leading to a smaller sample and noisier effects. As a result I limit the maximum number of DTOs to be 3 instead of 5. We observe precisely estimated zeros for the effect on outmigration.

The ENLACE test data comes with some caveats. Compliance was not perfect. Although the test was mandatory for all eligible students, only around 85% of eligible students took the ENLACE test each year in my data. The missing 15% is a combination of schools not administering the test and students not taking the exam on test day. As a result, dropout using ENLACE data is larger than in the school record data and the outmigration indicators may underestimate true migration responses for students that moved but did not take the test the next year. Despite these concerns, the individualized ENLACE data does not indicate an important migratory response.

8.5.2 Criminally Related Homicides

I use the homicide data to examine if the murders are criminally related. The homicide data contains demographic information on the victim and the cause of death, though it does not record if the victim was a criminal or a member of a criminal organization. I leverage the demographic information to examine if females have a treatment effect, if the homicides are more likely to involve a firearm, and I examine males aged 15-39 as in Calderón et al. (2015). Calderón et al. (2015) find that this age-gender group is likely to be part of a criminal organization.

Results can be seen in Table A4. We can see homicides only increase for men and not women, homicides involving a gun are also more likely to increase and males aged 15-39 are also likely to be the victim of homicides. Table A4 is consistent with an increase in criminal activity and competition as DTOs enter.

8.5.3 Specification Checks

I conduct several specification checks as well.

My identification strategy relies on one year lagged distance satisfying an exclusion. A natural concern is that lagged distance is correlated with current distance, which in turn affects homicides or school dropout. For instance, DTOs may leverage shorter distances across municipalities to shift resources where competition is most intense. To evaluate this concern, I add current distances as a control to the second stage equation 15. Results can be seen in Table A5 in column 2. We can observe effects for dropout are nearly identical, while effects for both homicides and and 9th grade dropout remain virtually identical. This corroborates the exposition in Section 2.2; current DTO distances likely do not affect outcomes because DTOs operate in a decentralized manner and do not share significant resources from one area to another.

In column 3 of Table A5 I reestimate the base specification without population and agricultural output controls. The dropout effects are largely unchanged while the homicide results become insignificant. This is unsurprising given the importance of population in explaining the total number of homicides. Column 4 includes log population controls and the homicide results are marginally statistically signfiicant. Column 5 drops the population controls but includes the agricultural controls - these estimates are very similar to those that do not control for agriculutral output, suggesting the results are robust to including this control.

The last two columns of Table A5 rerun the main specification but with 10 and 20 municipality types respectively. The dropout effects are again unchanged. The results on homicides become much smaller and statistically insignificant. One explanation as to why the homicide results are sensitive to the number of municipality types could be the sheer number of controls. With individual DTO incumbency controls, increasing municipality types from 5 to 20 represent 120 additional parameters.

Overall we can see that the effects on dropout are quite robust and insensitive to the different specifications. The results on homicides are qualitatively similar across specifications as well, however the statistical significance of the results is more sensitive. Nonetheless the directions and magnitudes of all the estimates are similar and comparable.

9 Conclusion

This paper documents important effects of DTOs on homicides and school dropout during Mexico's drug war. More DTOs raises homicides and increased dropout for older students, with the effects being driven entirely by male students. The effects are roughly linear for school dropout. In contrast the increases for homicides seem to be convex in the number of DTOs. In other words the increase in homicides grow with more DTOs. This suggests policy makers may prioritize areas with more DTOs to reduce violence. My estimates indicate removing one DTO in an area with many DTOs would result in a larger drop in homicides than areas with few DTOs.

Additionally I showed homicides with 15-18 year old males victims also increase with more DTOs. There is no corresponding effects for female students. The localized effects on older, male students suggests at least some of the students that drop out due to DTO presence are being pulled into criminal activity.

In answering this research question I developed a novel selection framework to estimate treatment effects for questions regarding market structure. In doing so I combined elements from the empirical industrial organization literature and the treatment effect literature. A natural avenue for future research is to adapt this paper's methodology to estimate dynamic treatment effects. Doing so would allow researchers to answer questions regarding the *history* of market structure.

References

Aguirregabiria, V., A. Collard-Wexler, and S. P. Ryan (2021, January). Dynamic games in empirical industrial organization. In K. Ho, A. Hortaçsu, and A. Lizzeri (Eds.), *Handbook of Industrial Organization*, Volume 4, pp. 225–343. Elsevier.

Aguirregabiria, V. and A. Magesan (2020). Identification and estimation of dynamic games when players' beliefs are not in equilibrium. *The Review of Economic Studies 87*(2), 582–625.

Aguirregabiria, V. and P. Mira (2007). Sequential Estimation of Dynamic Discrete Games. *Econometrica 75*(1), 1–53.

Aguirregabiria, V. and P. Mira (2019). Identification of games of incomplete information with multiple equilibria and unobserved heterogeneity. *Quantitative Economics 10*(4), 1659–1701.

Alcocer, M. (2023). *Drug Wars, Organized Crime Expansion, and State Capture*. Ph. D. thesis, UC San Diego.

Andrews, I., J. Sock, and L. Sun (2019). "Weak Instruments in IV Regression: Theory and Practice.". *Annual Review of Economics*.

Angrist, J. D. and G. W. Imbens (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American statistical Association 90*(430), 431–442.

Aradillas-López, A. (2020). The econometrics of static games. *Annual Review of Economics 12*, 135–165.

Arcidiacono, P. and R. A. Miller (2011). CONDITIONAL CHOICE PROBABILITY ESTIMATION OF DYNAMIC DISCRETE CHOICE MODELS WITH UNOBSERVED HETEROGENEITY. *Econometrica 79*(6), 1823–1867.

Atkin, D. (2016). Endogenous skill acquisition and export manufacturing in Mexico. *American Economic Review 106*(8), 2046–2085.

Balat, J. F. and S. Han (2022). Multiple treatments with strategic substitutes. *Journal of Econometrics*.

Benson, J. S. and S. H. Decker (2010, March). The organizational structure of international drug smuggling. *Journal of Criminal Justice 38*(2), 130–138.

Berman, N., M. Couttenier, D. Rohner, and M. Thoenig (2017, June). This Mine Is Mine! How Minerals Fuel Conflicts in Africa. *American Economic Review 107*(6), 1564–1610.

Bhuller, M. and H. Sigstad (2022). 2SLS with multiple treatments. *arXiv preprint arXiv:2205.07836*.

Blevins, J. R. (2014). Nonparametric identification of dynamic decision processes with discrete and continuous choices. *Quantitative Economics 5*(3), 531–554.

Bonhomme, S., T. Lamadon, and E. Manresa (2022). Discretizing unobserved heterogeneity. *Econometrica : journal of the Econometric Society 90*(2), 625–643.

Brinch, C. N., M. Mogstad, and M. Wiswall (2017, August). Beyond LATE with a Discrete Instrument. *Journal of Political Economy 125*(4), 985–1039.

Bruhn, J. (2021). Competition in the Black Market: Estimating the Causal Effect of Gangs in Chicago.

Calderón, G., G. Robles, A. Díaz-Cayeros, and B. Magaloni (2015, December). The Beheading of Criminal Organizations and the Dynamics of Violence in Mexico. *Journal of Conflict Resolution 59*(8), 1455–1485.

Carneiro, P., J. J. Heckman, and E. J. Vytlacil (2011, October). Estimating Marginal Returns to Education. *American Economic Review 101*(6), 2754–2781.

Carvalho, L. S. and R. R. Soares (2016). Living on the edge: Youth entry, career and exit in drug-selling gangs. *Journal of Economic Behavior & Organization 121*, 77–98.

Chang, E. and M. Padilla-Romo (2022, July). When Crime Comes to the Neighborhood: Short-Term Shocks to Student Cognition and Secondary Consequences. *Journal of Labor Economics*, 000–000.

Ciliberto, F. and E. Tamer (2009, November). Market structure and multiple equilibria in airline markets. *Econometrica : journal of the Econometric Society 77*(6), 1791–1828.

Collard-Wexler, A. (2013). Demand Fluctuations in the Ready-Mix Concrete Industry. *Econometrica 81*(3), 1003–1037.

Cornelissen, T., C. Dustmann, A. Raute, and U. Schönberg (2018). Who benefits from universal child care? Estimating marginal returns to early child care attendance. *Journal of Political Economy 126*(6), 2356–2409.

Correa-Cabrera, G. (2021). Los Zetas Inc. In *Los Zetas Inc.* University of Texas Press.

Dahl, G. B. (2002). Mobility and the return to education: Testing a Roy model with multiple markets. *Econometrica 70*(6), 2367–2420.

Dal Bó, E. and P. Dal Bó (2011). Workers, warriors, and criminals: Social conflict in general equilibrium. *Journal of the European Economic Association 9*(4), 646–677.

Daniele, G., M. Le Moglie, and F. Masera (2023, January). Pains, guns and moves: The effect of the U.S. opioid epidemic on Mexican migration. *Journal of Development Economics 160*, 102983.

Dell, M. (2015, June). Trafficking networks and the Mexican drug war. *American Economic Review 105*(6), 1738–1779.

Dell, M., B. Feigenberg, and K. Teshima (2019). The Violent Consequences of Trade-Induced Worker Displacement in Mexico. *American Economic Review: Insights 1*(1), 43–58.

Dix-Carneiro, R., R. R. Soares, and G. Ulyssea (2018, October). Economic shocks and crime: Evidence from the Brazilian trade liberalization. *American Economic Journal: Applied Economics 10*(4), 158–195.

Dube, O. and J. F. Vargas (2013, October). Commodity price shocks and civil conflict: Evidence from Colombia. *Review of Economic Studies 80*(4), 1384–1421.

Ellickson, P. B., S. Houghton, and C. Timmins (2013). Estimating network economies in retail chains: A revealed preference approach. *The RAND Journal of Economics 44*(2), 169–193.

Ericson, R. and A. Pakes (1995). Markov-Perfect Industry Dynamics: A Framework for Empirical Work. *The Review of Economic Studies 62*(1), 53–82.

Franco-Vivanco, E., C. B. Martinez-Alvarez, and I. F. Martínez (2023). Oil Theft and Violence in Mexico. *Journal of Politics in Latin America 15*(2), 217–236.

Gama, A. and D. Rietzke (2019). Monotone comparative statics in games with non-monotonic best-replies: Contests and Cournot oligopoly. *Journal of Economic Theory 183*, 823–841.

GFI (2017). Transnational crime and the developing world. *Global Financial Integrity*, 53–59.

Heckman, J., S. Urzua, and E. Vytlacil (2006). Estimation of treatment effects under essential heterogeneity. *Health affairs (Project Hope) 29*(3), 389–432.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the econometric society*, 153–161.

Heckman, J. J. and S. Urzúa (2010, May). Comparing IV with structural models: What simple IV can and cannot identify. *Journal of Econometrics 156*(1), 27–37.

Hirshleifer, J. (1995). Anarchy and its breakdown. *Journal of Political Economy 103*(1), 26–52.

Hull, P. (2022, January). Bayesian Bootstrap Thread.

INEE (2015). ¿Cómo Avanzan los Alumnos en su Trayectoria Escolar?

Insight Crime (2016). Mexico Suffers More Deaths Than War-Torn Iraq, Afghanistan - InSight Crime. https://insightcrime.org/news/brief/mexico-suffers-more-deaths-than-war-torn-iraq-afghanistan/.

Jarillo, B., B. Magaloni, E. Franco, and G. Robles (2016, November). How the Mexican drug war affects kids and schools? Evidence on effects and mechanisms. *International Journal of Educational Development 51*, 135–146.

Kirkeboen, L. J., E. Leuven, and M. Mogstad (2016). Field of study, earnings, and self-selection. *The Quarterly Journal of Economics 131*(3), 1057–1111.

Kline, P. and C. R. Walters (2016, November). Evaluating Public Programs with Close Substitutes: The Case of Head Start*. *The Quarterly Journal of Economics 131*(4), 1795–1848.

Lindo, J. M. and M. Padilla-Romo (2018, March). Kingpin approaches to fighting crime and community violence: Evidence from Mexico's drug war. *Journal of Health Economics 58*, 253–268.

Mazzeo, M. J. (2002). Competitive outcomes in product-differentiated oligopoly. *Review of Economics and Statistics 84*(4), 716–728.

McGuirk, E. and M. Burke (2020). The Economic Origins of Conflict in Africa. *Journal of Political Economy 128*(10), 3940–3997.

Medel, M. and F. Thoumi (2014, October). Mexican Drug "Cartels". In *The Oxford Handbook of Organized Crime*. Oxford University Press.

Mogstad, M., A. Santos, and A. Torgovitsky (2018). Using instrumental variables for inference about policy relevant treatment parameters. *Econometrica : journal of the Econometric Society 86*(5), 1589–1619.

Mogstad, M., A. Torgovitsky, and C. R. Walters (2021, November). The Causal Interpretation of Two-Stage Least Squares with Multiple Instrumental Variables. *American Economic Review 111*(11), 3663–3698.

Monteiro, J. and R. Rocha (2017, May). Drug Battles and School Achievement: Evidence from Rio de Janeiro's Favelas. *The Review of Economics and Statistics 99*(2), 213–228.

Mountjoy, J. (2022a). Community colleges and upward mobility. *American Economic Review 112*(8), 2580–2630.

Mountjoy, J. (2022b, August). Community Colleges and Upward Mobility. *American Economic Review 112*(8), 2580–2630.

Natarajan, M. (2006, June). Understanding the Structure of a Large Heroin Distribution Network: A Quantitative Analysis of Qualitative Data. *Journal of Quantitative Criminology 22*(2), 171–192.

Neyman, J. and E. L. Scott (1948). Consistent estimates based on partially consistent observations. *Econometrica: Journal of the Econometric Society*, 1–32.

Olea, J. L. M. and C. Pflueger (2013). A robust test for weak instruments. *Journal of Business & Economic Statistics 31*(3), 358–369.

Pereyra, G. (2012). México: Violencia criminal y" guerra contra el narcotráfico". *Revista mexicana de sociología 74*(3), 429–460.

Pesendorfer, M. and P. Schmidt-Dengler (2008, July). Asymptotic Least Squares Estimators for Dynamic Games1. *The Review of Economic Studies 75*(3), 901–928.

Rose, E. K. and Y. Shem-Tov (2021). How does incarceration affect reoffending? Estimating the dose-response function. *Journal of Political Economy 129*(12), 3302–3356.

Rubin, D. B. (1981). The Bayesian Bootstrap. *The annals of statistics*, 130–134.

Ryan, S. P. (2012). The Costs of Environmental Regulation in a Concentrated Industry. *Econometrica 80*(3), 1019–1061.

Saggio, R. (2012). Discrete unobserved heterogeneity in discrete choice panel data models. *CEMFI Master Thesis*.

Shirk, D. and J. Wallman (2015). Understanding Mexico's Drug Violence. *The Journal of Conflict Resolution 59*(8), 1348–1376.

Snyder, R. and A. Duran-Martinez (2009, September). Does illegality breed violence? Drug trafficking and state-sponsored protection rackets. *Crime, Law and Social Change 52*(3), 253–273.

Sobrino, F. (2020). Mexican Cartel Wars: Fighting for the U.S. Opioid Market. *Working paper*, 1–63.

Steinley, D. (2006). K-means clustering: A half-century synthesis. *British Journal of Mathematical and Statistical Psychology 59*(1), 1–34.

Sviatschi, M. M. (2022). Making a narco: Childhood exposure to illegal labor markets and criminal life paths. *Econometrica 90*(4), 1835–1878.

Tallis, G. M. (1961). The moment generating function of the truncated multi-normal distribution. *Journal of the Royal Statistical Society: Series B (Methodological) 23*(1), 223–229.

Tullock, G. (2001). Efficient rent seeking. *Efficient rent-seeking: Chronicle of an intellectual quagmire*, 3–16.

UNODC (2011). Estimating illicit financial flows resulting from drug trafficking and other transnational organized crimes. Technical report.

Walters, C. R. (2018). The Demand for Effective Charter Schools. *Journal of Political Economy 126*(6), 2179–2223.

Wright, A. L. (2015, September). Economic Shocks and Rebel Tactics: Evidence from Colombia. *SSRN Electronic Journal*.

A Appendix

A.1 Tables

Table A1: DTO Data on Presence by Year

Year	BLO	CJNG	CABT	CJ	CDS	CDG	FM	LZ
2006	✓	-	-	✓	✓	✓	✓	-
2007	✓	-	-	✓	✓	✓	✓	-
2008	✓	-	-	✓	✓	✓	✓	-
2009	✓	✓	-	✓	✓	✓	✓	-
2010	✓	✓	✓	✓	✓	✓	✓	✓
2011	✓	✓	✓	✓	✓	✓	✓	✓
2012	✓	✓	✓	✓	✓	✓	✓	✓
2013	✓	✓	✓	✓	✓	✓	✓	✓
2014	✓	✓	✓	✓	✓	✓	✓	✓
2015	✓	✓	✓	✓	✓	✓	✓	✓
2016	✓	✓	✓	✓	✓	✓	✓	✓
2017	✓	✓	✓	✓	✓	✓	✓	✓
2018	✓	✓	✓	✓	✓	✓	✓	✓

Table A2: DTO Data on Lagged Distance Instrument by Year

Year	BLO	CJNG	CABT	CJ	CDS	CDG	FM	LZ
2006	✓	-	-	✓	✓	✓	✓	-
2007	✓	-	-	✓	✓	✓	✓	-
2008	✓	-	-	✓	✓	✓	✓	-
2009	✓	-	-	✓	✓	✓	✓	-
2010	✓	✓	-	✓	✓	✓	✓	-
2011	✓	✓	✓	✓	✓	✓	✓	✓
2012	✓	✓	✓	✓	✓	✓	✓	✓
2013	✓	✓	✓	✓	✓	✓	✓	✓
2014	✓	✓	✓	✓	✓	✓	✓	✓
2015	✓	✓	✓	✓	✓	✓	✓	✓
2016	✓	✓	✓	✓	✓	✓	✓	✓
2017	✓	✓	✓	✓	✓	✓	✓	✓
2018	✓	✓	✓	✓	✓	✓	✓	✓

Notes: This table indicates which years has data on presence for the different DTOs, and which years have the 1 year lagged instrument.

Table A3: TSLS Sensitivity to Controls

	(1) Homicides	(2) Homicides	(3) Homicides	(4) Dropout 9th	(5) Dropout 9th	(6) Dropout 9th
Group Count	9.14*	7.98***	9.17*	-0.014	-0.0012	-0.014
	(4.93)	(1.61)	(4.95)	(0.0088)	(0.0023)	(0.0088)
Outcome Mean (levels)	7.95	7.95	7.95	0.03	0.03	0.03
# Instruments	8	8	8	8	8	8
N	30957.00	30957.00	30957.00	30957.00	30957.00	30957.00
Effective F-stat	29.88	62.08	29.88	29.88	62.08	29.88
Effective F-stat Cutoff 10%	13.26	14.73	13.27	13.26	14.73	13.27
Sargan-Hansen Test	20.45	18.64	20.46	36.31	40.63	36.42
Sargan-Hansen Test p-val	0.00	0.01	0.00	0.00	0.00	0.00
Controls	Yes	No	Yes	Yes	No	Yes
With Agri. Control	Yes	No	No	Yes	No	No

Notes: Unit of observation is a municipality-year. # Groups is the number of active criminal groups at year t. Column headings indicate the outcomes. The outcomes for the table are the count of homicides, "Homicides" and dropout rates for grade 9 as a fraction relative to enrollment at the start of grade 9,"Dropout 9th".
Outcome Mean (levels) reports the untransformed average number of homicides in (1), and the average drop out rate in (2). # Instruments reports the number of instruments used, these are the group specific lagged distances, in logs, for the 8 criminal groups tracked. N reports the sample size for each regression. "Effective F-stat" reports the Montiel Pflueger effective F-statistic (Olea and Pflueger, 2013). "Effective F-stat Cutoff 10%" reports the critical value testing that the worst case bias of TSLS exceeds 10% of the worst case bias for OLS, with a 5% confidence. "Sargan-Hansen Test' and "Sargan-Hansen Test p-val"' reports the Sargan–Hansen test statistic the p-value. Controls indicate which regressions include controls. The controls are log population, dummy variables for the presence of each of the 8 criminal groups for the preceding year and the IHS of the total number of active criminal groups in the preceding year. "With Agri. Control" indicates which models include the IHS of municipal agricultural revenue. All models include year and municipality fixed effects.
Standard errors in parenthesis clustered by municipality. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A4: Homicide Type Results

	All Homi.	Male Homi.	Female Homi.	Gun Homi.	Potentially DTO Related Homi.
1 vs. 0	13.50	12.82	0.62	14.38	10.35
	(12.30)	(11.17)	(1.21)	(11.60)	(9.15)
2 vs. 0	21.45	20.45	0.91	22.07	16.82
	(17.33)	(15.79)	(1.67)	(16.34)	(12.93)
3 vs. 0	24.65	23.42	1.12	24.92	21.24
	(19.29)	(17.65)	(1.78)	(18.09)	(14.26)
4 vs. 0	42.12*	39.97*	1.89	39.48*	32.44*
	(22.53)	(20.66)	(2.05)	(21.12)	(16.57)
$\geq$ 5 vs. 0	52.53*	49.80*	2.53	46.37*	41.76**
	(27.78)	(25.46)	(2.59)	(25.32)	(19.21)
Mean	6.13	5.49	0.63	4.61	3.46
N	27181.0	27181.0	27181.0	27181.0	27181.0

Notes:
Standard errors from 1000 bootstrap replications.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A5: Specification Checks

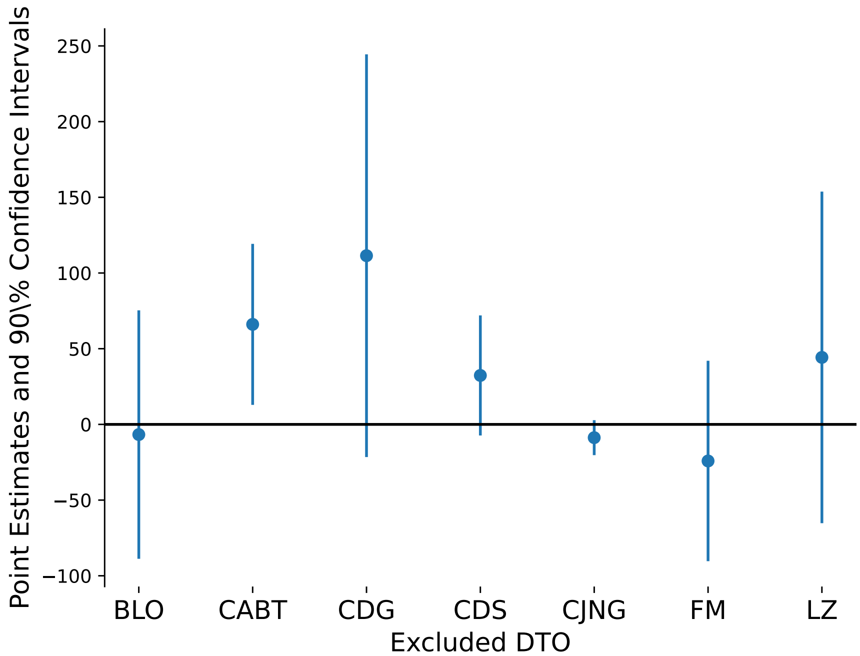
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A: 9th Grade Dropout Rate						
1 vs. 0	0.03**	0.03**	0.02	0.03**	0.02	0.04***	0.03**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
2 vs. 0	0.05**	0.05**	0.04*	0.05**	0.04*	0.06***	0.06**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
3 vs. 0	0.07**	0.07**	0.06**	0.07**	0.06**	0.08***	0.08***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
4 vs. 0	0.10***	0.10***	0.08**	0.10***	0.09**	0.11***	0.11***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)
$\geq$ 5 vs. 0	0.11**	0.11***	0.09**	0.11**	0.10**	0.13***	0.13***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)
	Panel B: Homicides						
1 vs. 0	13.50	14.99	8.13	14.10	10.69	5.46	4.84
	(12.30)	(12.60)	(7.46)	(12.31)	(7.82)	(8.71)	(5.83)
2 vs. 0	21.45	23.52	9.99	22.44	13.47	9.80	7.78
	(17.33)	(17.67)	(10.32)	(17.33)	(10.82)	(12.55)	(8.67)
3 vs. 0	24.65	26.25	3.05	25.74	6.75	12.38	7.56
	(19.29)	(19.34)	(11.94)	(19.26)	(12.50)	(14.13)	(10.45)
4 vs. 0	42.12*	43.32*	10.41	43.30*	15.67	20.11	9.27
	(22.53)	(22.25)	(15.27)	(22.50)	(15.97)	(17.09)	(13.68)
$\geq$ 5 vs. 0	52.53*	53.07*	11.99	54.04*	17.84	18.98	6.17
	(27.78)	(27.18)	(22.02)	(27.72)	(22.82)	(22.37)	(17.84)
# Muni. Types	5	5	5	5	5	10	20
Log Pop. Control	✓	✓	-	✓	-	✓	✓
IHS Ag. Control	✓	✓	-	-	✓	✓	✓
Dist.$_t$ Controls	-	✓	-	-	-	-	-
N	27181	27181	27307	27186	27337	26211	25765

Notes:
Standard errors from 1000 bootstrap replications.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A.2 Figures

Figure A1: One at a Time TSLS estimates

(a) Homicides



(b) 9th Grade Dropout

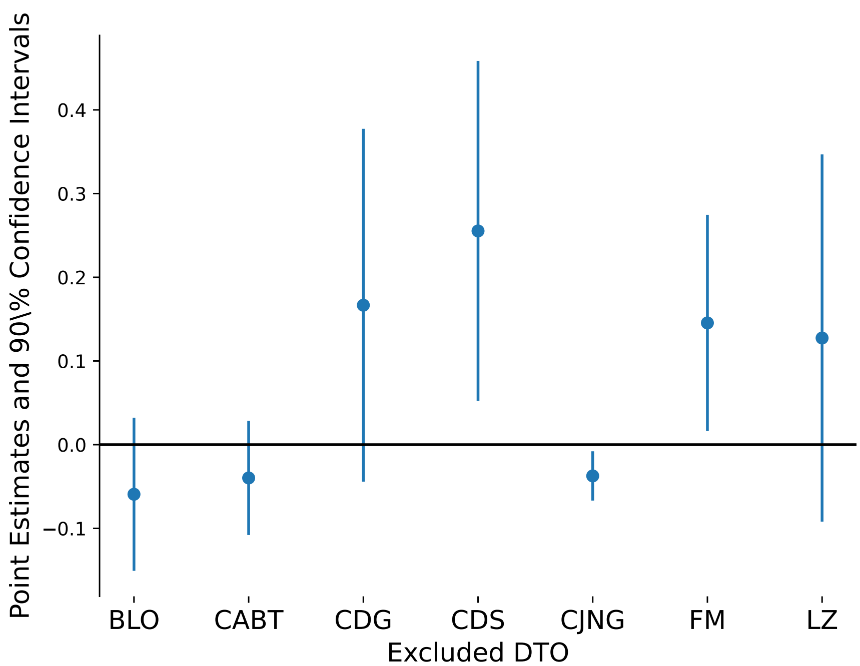
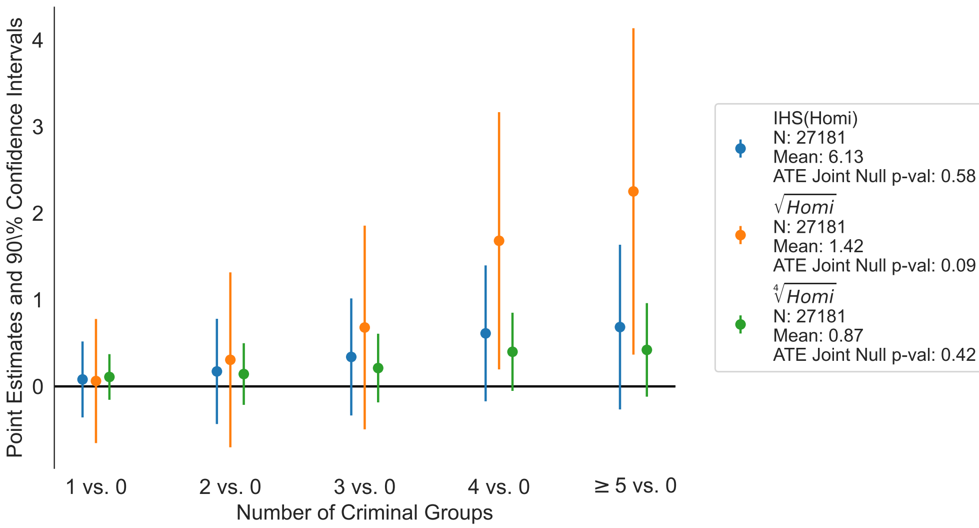


Figure A2: Homicide Results

(a) Alternative Transformations for Homicides



(b) Homicide Rates

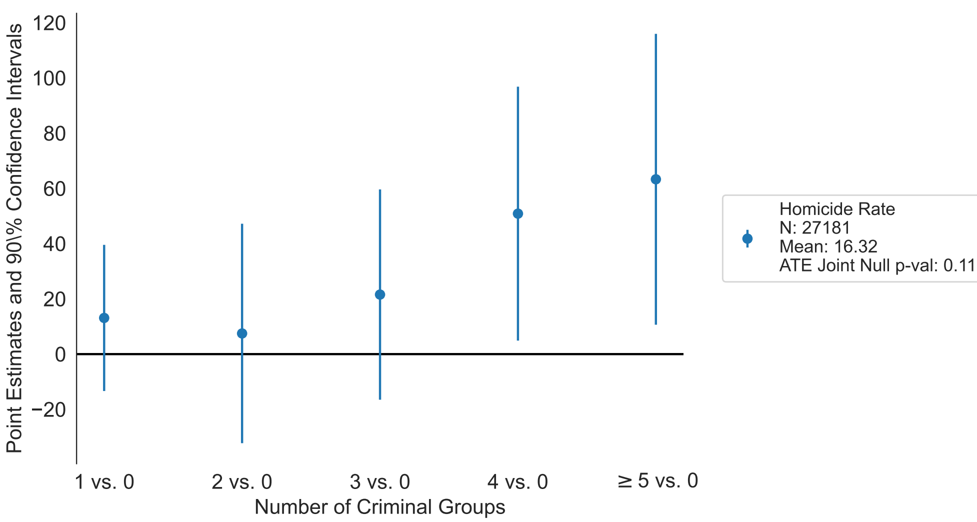


Figure A3: ENLACE Out Migration Results

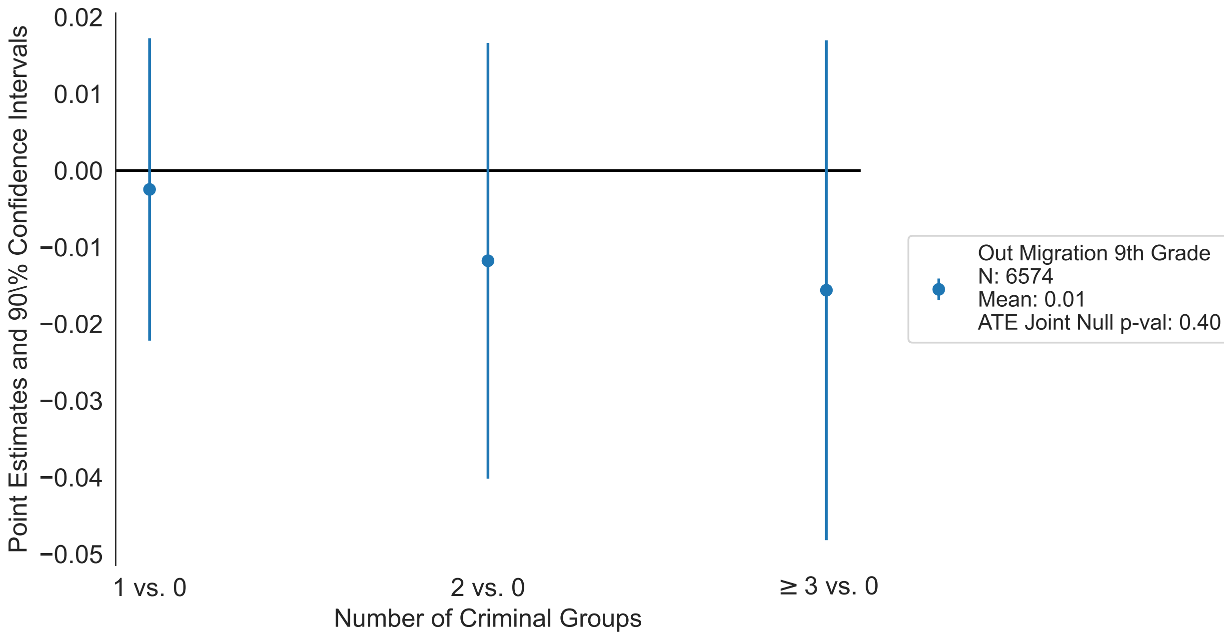
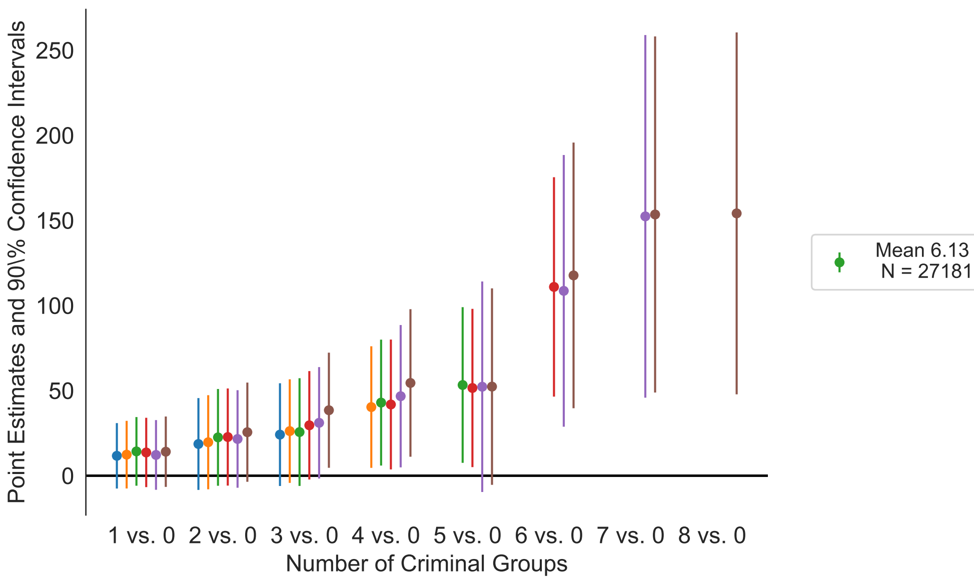
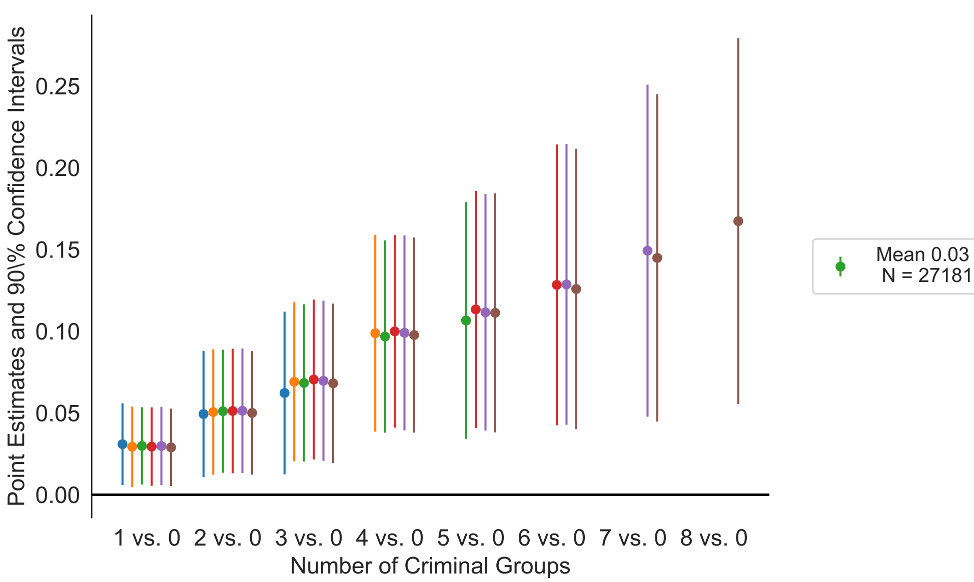


Figure A4: Results with Different Upper Limit on Treatment Effects

(a) Homicides



(b) 9th Grade Dropout Both Sexes



B Presence Imputation

For the analysis in this paper I use an imputed version of DTO presence as recorded in my data. The imputation fills in "gaps" in the presence variables. For each munipality, I fill in the years between two years where a DTO is observed. For example, if Sinaloa Cartel is observed in municipality A in 2010 and is not observed again until 2013, I fill in 2011 and 2012 as having Sinaloa Cartel's presence.

I do this imputation in order to have a better measure of a DTO's network over time, which is crucial for my distance based instrument. A DTO will be marked as absent in my data if there are no relevant news reports concerning that DTO and municipality for a given year. The DTO may still be present in the municipality, due to my data's measurement error. This is especially likely if the gaps we observe are shorter. Table B1 displays the number of imputations and the average length of years for each filled in gap. The average gap is only two years long. Table B2 displays the imputed and non-imputed distrubtion of DTO group counts. Table B3 displays average distances according to the imputation strategy.

To further examine the effect of this imputation, I run the following fixed effects regression:

$$\text{Homicides}_{mt} = \alpha_m + \tau_t + \delta \text{Log Population}_{mt} + \sum_{i=1}^{8} \beta_i 1[N_{mt} = i] + u_{mt} \tag{18}$$

I do this for the imputed and non-imputed versions of N_{mt}. This regression examines if there any drastic differences between the imputed and non-imputed measures. Results are in Table B4 Reassuringly, both measures yield largely the same coefficients for homicides. I take this as evidence that the imputation is largely innocuous.

Finally, I compare my main results with and without this imputation. both the measurement of group counts N_{mt} and the measurement of the instrument Z_{gmt-1}, since the distance instrument depends on where the DTOs are located. Because the non-imputed version has lower DTO group counts and the increased measurement error in the instrument, it is not possible to estimate effects up until 5 DTOs. I limit this estimation to a maximum of 2 DTOs. Results are in Table B5. As we can see the 9th grade dropout results are similar but larger in magnitude with no imputation. The homicide results are also larger and also statistically significant. In summary, the results are qualitatively similar, but the imputation allows me to estimate a richer set of effects.

Table B1: Number of Imputations and Average Length

DTO	# Imputations	Average Imputation Length
BLO	486	2.22
CJNG	564	1.99
CABT	309	1.80
CJ	260	2.35
CDS	739	2.41
CDG	584	2.58
FM	532	2.17
LZ	695	2.09
Total	4169	1.96

Notes: This tables displays the number of imputations for each DTO in the column # Imputations. "Average Imputation Length" displays the average number of years for the imputations.

Table B2: Distribution of DTO Group Count

# DTOs	No Imputation	Imputed
0.00	0.78	0.70
1.00	0.09	0.10
2.00	0.06	0.07
3.00	0.03	0.05
4.00	0.02	0.03
5.00	0.01	0.02
6.00	0.01	0.01
7.00	0.00	0.01
8.00	0.00	0.01

Notes: This tables displays the distribtuion of observations by the DTO group count (# DTOs).

Table B3: Comparison of Distance Instrument by Imputation

DTO	Imputed		Not Imputed	
BLO	69.21	(63.56)	87.01	(73.78)
CABT	68.52	(70.34)	80.57	(80.62)
CDG	57.54	(53.53)	76.89	(66.45)
CDS	51.13	(47.27)	65.92	(53.6)
CJ	104.46	(73.81)	154.45	(121.86)
CJNG	66.94	(71.68)	78.09	(76.46)
FM	96.72	(123.56)	131.54	(155.35)
LZ	34.45	(30.73)	44.82	(39.83)
Average	75.81	(57.37)	98.35	(65.8)

Notes: This table compares the DTO distances by imputation strategy. "Average" refers to the average distance across all 8 DTOs. Standard errors shown in parenthesis.

Table B4: Homicide Regressions by Imputation

	No Imputation	Imputed
1 vs. 0 DTOs	1.43*	1.05
	(0.78)	(0.97)
2 vs. 0 DTOs	4.10*	4.59*
	(2.37)	(2.70)
3 vs. 0 DTOs	6.63***	4.23**
	(1.20)	(1.65)
4 vs. 0 DTOs	10.42***	8.43***
	(1.99)	(2.22)
5 vs. 0 DTOs	11.44**	9.64**
	(5.33)	(3.78)
6 vs. 0 DTOs	26.82***	19.63***
	(4.56)	(3.60)
7 vs. 0 DTOs	46.25***	35.99***
	(9.68)	(6.07)
8 vs. 0 DTOs	106.94***	73.43***
	(25.45)	(15.24)
N	30959	30959

Notes: This table compares OLS regressions of the number of homicides on the number of DTOs, as measured with and without imputation. The first column uses DTO counts as measured with no imputation. The second column includes imputed presence for the 8 large DTOs tracked in this paper. Observations are imputed for each DTO. For all time periods between two years where a DTO is observed within the same municipality, my imputation fills in those periods as if the DTO was present in that interval.

Both regressions include municipality and year fixed effects as in 18. Both regressions control for log population.

Standard errors in parenthesis clustered by municipality. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B5: ATE estimates by Imputation

	(1) Imputed	(2) Not Imputed
	Panel A: 9th Grade Dropout Rate	
1 vs. 0	0.03**	0.07***
	(0.02)	(0.02)
$\geq$ 2 vs. 0	0.04**	0.09***
	(0.02)	(0.02)
	Panel B: Homicides	
1 vs. 0	12.16	-2.45
	(10.32)	(15.59)
$\geq$ 2 vs. 0	21.06	41.11**
	(15.02)	(18.72)
N	25143.0	27537.0
Mean	4.4	5.87

Notes: This table compares ATE estimates of the main model with imputed DTO presence to estimates that do not impute DTO presence. The difference in sample size is due to the different first stage estimates relative to the main model. This leads different observations to have a probability 1 for having 0, 1, or $\geq$ 2 DTOs, which leads them to be excluded by the criterion in Section 7.5.

Standard errors in parenthesis clustered by municipality. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

C An Appropriate Entry Game

This section describes what kind of entry game yields the structure required for the control function approach used in this paper. A simple model in the spirit of the Ericson and Pakes (1995) model [23] helps fix ideas. To start, suppose the G criminal groups are myopic, so the game is not dynamic. Recall the covariates are the excluded instruments $\tilde{Z}_{mt}$[24] and municipality level covariates X_{mt}. $\tilde{Z}_{mt}$ are the distance instruments used in the paper. This paper employs the one year lagged measures of these distances, $\tilde{Z}_{mt-1}$. The covariates X_{mt} contain municipal level information on economic activity. The outcomes of interest, Y_{mt}, are local homicides and local schooling decisions.

After groups enter in time period t, they engage in some form of static competition given group entry and market conditions. In this static competition, groups optimally decide how much output to produce and effort to expend. In the case of criminal groups, their output could be how many illicit activities to engage in or how much effort to expend enforcing contracts. To help fix ideas, I represent each group's *effort* with the variable e_g. e_g is only a function of entry at time t, and the other state variables X_{mt}, D_{mt-1}. Each group's effort e_g decided by all groups will determine the level of homicides at time t, and therefore other outcomes Y_{mt}.

This setting is similar to the Ericson and Pakes (1995) model. However there is an important difference: there are no inter-temporal adjustment cost to increasing or decreasing effort e_g. e_g and therefore Y_{mt} do not depend on previous investment decisions. As a result the model does not need to condition on groups' past investments in m. This is a sensible restriction; observing group level information on criminal capacity or output is difficult.

I summarize these ideas with groups' one-period profit function $\pi_g()$. If group g enters, they receive as flow profit:

$$\pi_g(\tilde{Z}_{mt-1}, X_{mt}, D_{mt}, D_{mt-1}) = r_g(X_{mt}, D_{gmt}, D_{-gmt}) - \rho_g(\tilde{Z}_{gmt-1}) - (1 - D_{gmt-1})\delta + \epsilon_{gmt} \quad (19)$$

[23] Aguirregabiria et al. (2021) provide an excellent review.

[24] For a group g, $\tilde{Z}_{gmt}$ is the shortest distance between each municipality m and any other municipality $m' \neq m$ where g operates, i.e. $D_{gm't} = 1$. I collect these distances in a vector $Z_{mt} = \{Z_{gmt}\}_{g=1}^{G}$.

and profits are normalized to zero if the group does not enter.

$r_g(X_{mt}, D_{gmt}, D_{-gmt})$ represents the static profits received by group g given other groups entry decisions D_{-gmt} and state variables X_{mt}, D_{mt-1}. For example, r_g could be modeled as Cournot or Bertrand competition between groups. The literature studying crime typically models r_g as a rent seeking game in the spirit of Tullock (2001). $\rho_g(\tilde{Z}_{gmt-1})$ represents a fixed operating cost each group g has to pay each period to operate in m. This is a function of $\tilde{Z}_{gmt-1}$; groups that were closer to m in the last period are allowed to have a lower operating cost in m. For instance, groups that are closer may have an easier time enforcing contracts in m due to their nearby reputation, or groups that are closer may have better information on local municipal conditions. $(1 - D_{gmt-1})\delta$ represents an additional costs new entrants (with $D_{gmt-1} = 0$) have to incur to operate in m. This helps capture the intuition that incumbents are more likely to operate in m than new entrants.

The most crucial feature of this simple model is that $\tilde{Z}_{gmt-1}$ does not enter $r()$, and therefore does not affect effort e_g or Y directly. As a result $\tilde{Z}_{gmt-1}$ plays no role in determining the level of group competition. However $\tilde{Z}_{gmt-1}$ does affect $\rho_g()$ and hence the likelihood that g enters. $\tilde{Z}_{gmt-1}$ is therefore a valid instrument in this simple model.

C.1 Nonlinearities induced by the private information game.

The private information game induces non linearities in the entry decisions, even if the flow profits are linear. Consider the following linear parametrization of flow profits

$$\pi_g(\cdot) = \beta_1 X_{mt} - \beta_2 \log\left(\sum_{j \neq g} D_{gmt} + 1\right) - \beta_3 \tilde{Z}_{gmt} - (1 - D_{gmt-1})\delta + \epsilon_{gmt} \tag{20}$$

where $r_g(\cdot) = \beta_1 X_{mt} - \beta_2 \log\left(\sum_{j \neq g} D_{gmt} + 1\right)$, $\rho_g(\cdot) = \beta_3 \tilde{Z}_{gmt}$. $\log\left[\sum_{j \neq g} D_{gmt} + 1\right]$ captures the competitive effects caused by other groups' entry. The above equation describes profits given competitors entry decisions D_{-gmt}, which is unknown at the time of entry.

Instead, groups enter according to expected profits:

$$E\left[\pi|\tilde{Z}_{mt-1}, X_{mt}, D_{mt-1}\right] = \beta_1 X_{mt} - \beta_2 E\left[\log\left(\sum_{j\neq g} D_{gmt} + 1\right)\middle|\tilde{Z}_{mt-1}, X_{mt}, D_{mt-1}\right] - \beta_3 \tilde{Z}_{gmt} - (1 - D_{gmt-1})\delta + \epsilon_{gmt}$$

and groups enter if $E\left[\pi|\tilde{Z}_{mt-1}, X_{mt}, D_{mt-1}\right] > 0$. Importantly, the term

$$E\left[\log\left(\sum_{j\neq g} D_{gmt} + 1\right)\middle|\tilde{Z}_{mt-1}, X_{mt}, D_{mt-1}\right]$$

will typically be a non-linear function in $\tilde{Z}_{mt-1}, X_{mt}, D_{mt-1}$. As a result, even this stylzed example results in the non-linear representation of the main paper, namely

$$v_g(Z_{mt}) > \epsilon_{gmt}$$

with $Z_{mt} = (\tilde{Z}_{mt-1}, X_{mt}, D_{mt-1})$