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

Ignacio Rodriguez Hurtado*

Preliminary, please do not circulate. For latest version click <u>here</u>.

September 9, 2023

Abstract

Crime across the world hinders health, community development and economic growth. Criminal market structure is likely a key determinant. I study the impact of criminal market structure on homicides and school dropout in Mexico between 2006 and 2018. I measure criminal market structure with the number of active Mexican Drug Trafficking Organizations (DTOs) in Mexican municipalities. For identification, I develop a novel selection model using distance-based instrumental variables to estimate the causal effects. The instruments are based on DTOs' distance to their existing base of operations. My framework delivers interpretable treatment effects for each possible number of active DTOs, while conventional two-stage least squares estimators do not. Although my focus is on criminal market structure, the same framework can be used to examine other licit industries as well. I find more DTOs lead to more homicides, with the relationship being increasing and S-Shaped in the number of DTOs, while school dropout also increases, but with a linear relationship. Furthermore the effects on dropout are driven by older, male students, which suggests DTOs are pulling students into criminal activities. This is the first paper to study the non-linear effects of criminal market structure on both crime and education in an instrumental variables framework.

1 Introduction

Crime is costly. The Institute for Economics and Peace (2015) estimates 1.9 trillion USD was spent containing crime worldwide during 2014. This large figure understates the true cost of crime, since it doesn't account for the indirect effects of homicide, theft and kidnapping on the economy. This paper studies how criminal market structure shapes crime and local outcomes.[1] Theoretically, criminal market structure is a key parameter in models of violence and crime, which typically predict violence and crime rise with the number of criminal organizations (Hirshleifer, 1995; Gama and Rietzke, 2019). Popular accounts agree with these predictions. Several news reports highlight violent turf battles between criminals in places such as Latin America, South Africa and even North American cities like Chicago, severely burdening surrounding communities (Guardian, 2016; Post, 2022; de Greef, 2019).

This study's setting, Mexico, serves as a prime illustration of the costs imposed by criminal organizations. Mexico has several large Drug Trafficking Organizations (DTOs)[2] competing amongst each other. As a result Mexico is in a period of extreme violence with a homicide rate of 29 homicides per 100,000 inhabitants in 2018, almost six times larger than the US homicide rate (INEGI, 2020; FBI, 2023). The violence is so egregious Mexico had more homicides in 2015 than Afghanistan or Iraq (Insight Crime, 2016). Many qualitative accounts argue the growth in DTOs' footprint and reach is a major reason for the increase in violence (Medel and Thoumi, 2014; Snyder and Duran-Martinez, 2009).

My paper quantifies the effect of the number of active DTOs, a crucial feature of criminal market structure, on two outcomes: homicides and school dropout. I find that increased DTOs increases both homicides and school dropout. I study homicides since it garners substantial policy attention and is better measured relative to other indicators of crime. I study school dropout since it is an important indicator for human capital

---

[1] I define criminal market structure as the collection of active criminal organizations operating in a geographic area. This definition of market structure is similar to Ciliberto and Tamer (2009) for the airline industry.

[2] 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.

formation. I largely focus on ninth grade dropout, the last year of middle school in Mexico. However, I examine younger grades and no effects throughout elementary school or the first year of middle school.[3] As I explain later, these patterns are consistent with older students being pulled into criminal activities by the DTOs.

Furthermore, most models of crime predict violence to be *non-linear* in the number of criminal groups (Hirshleifer, 1995; Dal Bó and Dal Bó, 2011).[4] Yet, few empirical studies have investigated the relationship between the number of criminal actors and crime in order to estimate these non-linearities. These non-linearities are of policy interest since they can help target scarce policing resources to reap the most reductions in violence. To the extent that government forces can remove DTOs from Mexican municipalities by employing policing resources, a policy that removes a DTO may be more effective based on the non-linearities between crime and number of actors.[5] For example, if this relationship is increasing and convex, then removing DTOs will be more effective in places with a high number of DTOs.

However, estimating these non-linear effects is challenging for two reasons. First, DTOs endogenously select where to operate, which motivates an instrumental variables approach. For identification, I leverage the sudden and large expansion in DTO presence between 2006 and 2018. I use DTO's one year lagged distance to their existing operations as instrumental variables, and I show DTOs expanded in areas close to their existing operations. I show that after controlling for municipality fixed effects, lagged distance is uncorrelated with several indicators of economic performance or the number of police officers in municipalities. This suggests one year lagged distance is indeed a viable instrument for DTO presence.

The second challenge for estimating the non-linear effects is methodological. Even with valid instruments, conventional Two Stage Least Squares (TSLS) typically cannot deliver interpretable causal estimates for multi-valued treatments (Heckman and Urzúa, 2010; Bhuller

---

[3]I do not study high school dropout because high school was not universal or compulsory throughout my study period. In 2006 there are more than 500 municipalities without any high schools. High school became compulsory and universal in 2012, but even still there are roughly 200 municipalities in my data without any high schools in 2018.

[4]The intuition is that, as the number of groups grows, the market power of each group falls. In the limit, this market power falls to zero, but is positive for any given number of organizations. This dynamic is analogous to a market with Cournot duopoly transitioning to perfect competition.

[5]The Mexican government regularly targets DTOs to reduce their scope of operations (Jones, 2016).

and Sigstad, 2022; Mountjoy, 2022a; Kirkeboen et al., 2016). To clarify the intuition behind this challenge, consider the Sinaloa Cartel and one of its rivals, Los Zetas. If Sinaloa Cartel is nearby a municipality, it is more likely to enter. Yet Los Zetas also observes Sinaloa's proximity and Los Zetas may itself be less inclined to enter in order to avoid competition. In aggregate, a reduction in Sinaloa Cartel's proximity will induce some municipalities to receive more active DTOs (as Sinaloa enters), while other municipalities will see *fewer* DTOs (as Los Zetas does not enter). A shift in the instrument therefore leads to an unclear treatment dose, and TSLS fails to make the correct counterfactual comparison. As a result, multi-valued TSLS estimates typically result in a combination of different treatment effects, complicating their interpretation.[6]

Instead of TSLS, **I develop a novel selection model to estimate treatment effects** for different "doses" of active DTOs. The model is estimated with a simple Heckman (1979) two-step approach, where I use first-stage probabilities to selection correct the second stage regression. For the first stage, I borrow from industrial organization to model DTO entry as a firm-entry game. The DTOs' individual entry decisions then 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, which is required to study the non-linearities predicted by theory. Finally, I trace out the entire dose response function of homicides and school dropout as a function of active DTOs.

Using the previous framework, I estimate important non-linearities for homicides. I find that increased DTOs increases homicides and this relationship has an increasing S-shape with an inflection at 6 DTOs. The largest reduction in homicides resulting from removing one DTO happens from 6 to 5 DTOs with 65 fewer homicides. The next largest reduction in homicides occurs from 7 to 6 DTOs with 36 fewer homicides. On the whole, these patterns are incongruous with predictions from simple models of crime and violence, which instead predict the curve should be concave with a horizontal asymptote (Hirshleifer, 1995). To my

---

[6]For example, suppose one "dummies out" the number of active DTOs in a TSLS regression, and there is a maximum of 2 DTOs. The coefficient for 1 DTO will be a combination of the effect of 1 DTO vs 0 DTOs, along with the effect of 1 DTO relative to 2 DTOs. This is true even if the TSLS is just or over-identified.

knowledge, this is the first paper that empirically documents these non-linear patterns.

In contrast, the effects on ninth grade dropout are linear, with an additional DTO increasing dropout by 2-3 percentage points. The effects on school dropout are all positive and are driven by older, male students. For male students, all the effects are positive and significant, while for female students the effects are estimated to be zero, though with wide confidence intervals. Furthermore, the effects are more muted for 8th grade students and estimated to be zero for elementary school students. The gendered and age-differentiated effects are consistent with at least some students dropping out to be employed by the DTOs. Consistent with this, I find that homicides for 15 to 18 year old males also increase.

This is the first paper to study the non-linear effects of criminal market structure on both crime and education in an instrumental variables framework. Bruhn (2021) and Sobrino (2020) are similar papers and also study the impacts of more criminal organizations on crime by employing panel studies in Chicago and Mexico. They also find that more criminal organizations lead to more crime and violence. I complement these papers in two ways. First, in addition to homicides, I also study the effects of DTOs on education. Second I allow DTOs to select entry based not only on time-invariant factors, as the previous panel studies do, but also based on time-varying unobservables by leveraging an instrumental variables framework. This framework then allows me to estimate the nonlinear effects of market structure on crime, which the previous two papers do not explicitly estimate.

My finding that DTOs increase homicides contributes to the extensive literature studying the drivers of crime 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. Calderón et al. (2015) and Lindo and Padilla-Romo (2018) expand on these findings and document that government captures of prominent DTO leaders causes an increase in violence, as the weakened DTO attracts more competition. Alcocer (2023)

documents oil windfalls attracted DTOs in Mexico and increase homicides. I complement these papers by showing the effects of market structure on crime and by documenting its non-linear relationship.

I also contribute to the literature examining the impact of crime on educational choices. To my knowledge, this is the first paper studying the effects of criminal market structure on educational outcomes. My contribution is to find that increased DTO presence increases school dropout, mainly for older and male students. Moreover, I find this relationship is linear in the number of DTOs, with no significant non-linearities. In similar papers, several authors have found violence decreases educational attainment in Mexico and elsewhere (Brown and Velásquez, 2017; León, 2012; Lopez Cruz and Pairolero, 2019). Other papers corroborate this finding by showing that hyper-local 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 closely related paper Sviatschi (2022) documents that booms in illegal coca planting induces drop out and incarceration later in life in Peru. I complement this paper by showing criminal market structure is an important factor for schooling choices as well.

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., 2006a; Lee and Salanié, 2018; 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 legal, 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). A closely related paper is Balat and Han (2022) which derives informative bounds based on shape restrictions for a static complete information entry game. In contrast, my framework models the first stage as a private information entry game involving different

groups, which is a more common model in IO (Aguirregabiria et al., 2021). Furthermore, I allow for dynamic games with forward looking agents, multiple equilibria and do not require to estimate any structural parameters for the underlying profit functions.

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 and shows descriptive statistics. Section 4 describes the empirical strategy and shows evidence in support of the proposed instruments. Section 5 provides preliminary results TSLS results and argues TSLS is an inappropriate strategy in my setting. Section 6 presents the main model used for estimation while Section 7 describes estimation details. Section 8 discusses the results and Section 9 concludes.

2　The Mexican War on Drugs and Drug Trafficking Organizations

I first provide context on the Mexican War on Drugs and Mexico's DTOs. I describe the sudden and unexpected expansion of DTOs across Mexico beginning in 2006 due to a radical shift in government policy. Furthermore I provide background on the 8 large DTOs I study. A major feature of these DTOs is their decentralized nature, with their internals actors being largely unconnected amongst each other. **DTOs' decentralized structure** hinders their ability to execute a **long-term, cross-regional expansion plan**, which is important for the validity of my proposed instrument.

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) 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, 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 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.[7] The selection of these DTOs is based on their mentions in the DEA's National Drug Threat Assessment reports from

---

[7]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*).

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.[8]

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.

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 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

---

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

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.

3   Data Sources and Descriptives

I now explain the sources of data used and present basic summary statistics. I show **DTOs select into richer and more developed areas within Mexico.** 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.[9]

3.1   Drug Trafficking Organization Presence

The data source on DTO presence comes from Sobrino (2020). The data agrees with reports from the US Drug Enforcement Agency[10], is consistent with similar datasets that manually collect news reports on DTO presence, and Sobrino (2020) shows reports on DTO presence are uncorrelated with the assassination of journalists in Mexico. Data on DTO presence 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

---

[9]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.

[10]These reports do not systematically cover the sample years in my paper.

that do not describe presence but instead mention the DTO for an unrelated reason. There-fore 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 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 there was not a news worthy event to mark the DTOs as present in the data. For each DTO I impute between 200 and 800 observations as being present, which represents 1-2% of the observations in my sample. The average length of the gap I fill is roughly 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 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 since it is a simple, reliable measure of dropout. This is dropout that happens between the start of the academic year and the end of the same school year.[11] 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.

3.3   Other data

I gather other data to supplement the analysis.

---

[11]Measuring *inter*-curricular dropout, from one school year to the next, is more complicated since students can change schools between school years or repeat grades. The repetition of grades complicates the dropout measure. Additionally, the fact that students can change schools complicates studying dropout for grade 7, since students change from elementary school to middle school and it is difficult to keep the same set of schools for comparison. The measure I use, intra-curricular, dropout avoids these issues.

Figure 3: School Calendar

August t July $t+1$

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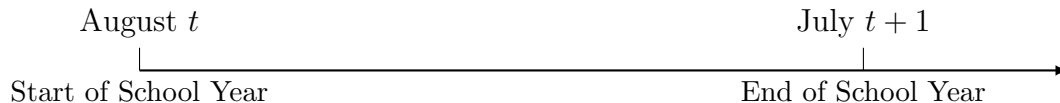
Notes: Figure describes the end and start of school year s relative to calendar year t. The end of the school year in Mexico happens in July, and begins again in August. 9th grade dropout for calendar year t is calcluced relative to enrollment at the start of the school year. I then compare how many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

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, and I use it as a control variable in the analysis.

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. I use this data to check if the instrument is correlated with economic activity.

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.*

3.4 Descriptive Statistics

Table 1 presents descriptive statistics for my sample, which documents that DTOs select into richer and more populous areas. This can be seen in the top panel, which 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. Furthermore, traditional panel data methods may fail in this setting if DTOs select entry based on a time-varying unobservable. I show later this is indeed the case, as the time-varying measure of distance used as an instrument predicts DTO entry.

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 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.

Table 1: Descriptive Statistics

| | Panel A: Means By Number of Groups | | | | |
	0	1	2	3+	Overall
Covariates					
Population (2005)	10,787	23,206	38,742	109,233	43,432
Poverty Rate (2005)	0.75	0.66	0.62	0.55	0.67
Literacy Rate (2005)	0.80	0.84	0.86	0.88	0.83
Elect. Users	3,335	7,089	12,780	36,326	14,218
kWH (millions)	14.31	38.90	71.34	232.62	85.68
kWH per Capita	3,739	5,719	8,600	10,142	6,360
Ag. Revenue	52.98	126.85	170.53	376.33	167.39
Total Police	27	41	73	168	74
Outcomes					
Homicide Rate (per 100,000)	11.94	14.04	17.62	23.01	15.98
9th Grade Enrollment	221.37	457.67	759.15	2,059.89	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	0.27	0.70	1.03	1.12	0.81
≥ 1 DTOs	0.14	0.27	0.35	0.39	0.30
≥ 2 DTOs	0.07	0.18	0.25	0.27	0.20
≥ 3 DTOs	0.03	0.11	0.16	0.18	0.12
Instruments					
Average Dist.$_{t-1}$ (km)	127.13	78.17	57.98	50.98	76.44
Dist.$_{-1}$ BLO (km)	120.38	59.76	52.50	51.20	69.44
Dist.$_{-1}$ CABT (km)	.	128.04	66.62	55.50	68.74
Dist.$_{-1}$ CDG (km)	82.13	58.56	48.71	47.42	58.30
Dist.$_{-1}$ CDS (km)	89.61	46.90	40.47	35.07	51.63
Dist.$_{-1}$ CJ (km)	154.75	103.55	86.37	85.16	105.74
Dist.$_{-1}$ CJNG (km)	.	113.87	73.30	40.04	67.54
Dist.$_{-1}$ FM (km)	188.79	90.83	61.87	59.71	97.18
Dist.$_{-1}$ LZ (km)	.	41.00	34.03	33.74	34.76

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, reflecting the rapid expansion of DTOs in my study period.

4   Identification Strategy and Instrument Diagnostics

This section lays out my paper's identification strategy, which is based on DTOs distance to their existing operations. I showed in Table 1 there is a selection problems: DTOs select into richer areas to operate. As a result, I pursue an instrumental variables strategy. I argue that **conditional on fixed effects, DTOs' one year lagged distances are plausible instruments to study the effects of DTO presence.** After presenting arguments for my identification strategy, I provide empirical evidence that DTOs' lagged distances are uncorrelated with several economic characteristics of Mexican municipalities.

4.1   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 Yucatán peninsula in the east of the country as well. By 2018 we can see it expanded heavily nearby its main center of operations in 2006 and also expanded in Yucatán.

The basic logic of my instrument is that groups' distance to their existing operations is an important predictor of the number of active DTOs in a municipality, the dimension of market structure I study. The distances should affect market structure along two channels. First, a group's own proximity will make them more likely to enter. If Sinaloa Cartel is closer, Sinaloa Cartel is more likely to enter. The second channel is through a competitive effect. If Sinaloa Cartel's *rival's* are closer, Sinaloa Cartel will be *less* likely to enter since they anticipate more competition. My first stage Table 4 corroborates these patterns: Groups are more likely to enter when they are nearby, and less likely to enter if its rivals are nearby. Combining these two channels, DTOs' distances to different municipalities is an important

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

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

predictor of the number of active DTOs in a municipality

My identification argument is that DTO's distances to their existing operations is a plausible instrument to identify the effects of DTO presence and group count. The argument is this past proximity should have little to no effect on current outcomes, except for the changes 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). In my setting, I use each group's one-year lagged distance to their operations as instruments. Additionally I strengthen my identification argument by controlling for municipality 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 must affect outcomes only through their effect on DTO presence. I present arguments for both these assumptions 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 coun-

try. 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 clandestine nature of criminal organizations difficulties, information and orders cannot flow freely from the top to the bottom of the organization (Von Lampe, 2015). 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.[12]

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 nearby municiaplities with no DTOs. Moreover these spillovers need to be dynamic since I'm using last year's distances as instruments for current DTO group count. Therefore, it is not enough for there to be a spillover at time t. The spillover's effect also needs to persist into $t + 1$ for the exclusion restriction to be violated.

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 year, therefore affecting current outcomes directly. However

---

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

the decentralized nature of DTOs rules this possibilty out since it makes sharing resources across the DTO network unlikely. Moreover in the robustness checks of Section 8.4 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 main results are driven by migration in Section 8.4.1. 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, or affect economic outcomes and therefore school dropout. In contrast, my strategy based on past proximity can alleviate some of these concerns.

Figure 5 displays how I calculate distances. All distances are based on straight-line distances from the largest population center in each municipality. In this example, I calculate

Figure 5: Calculating Group-Specific Distances to Municipalities
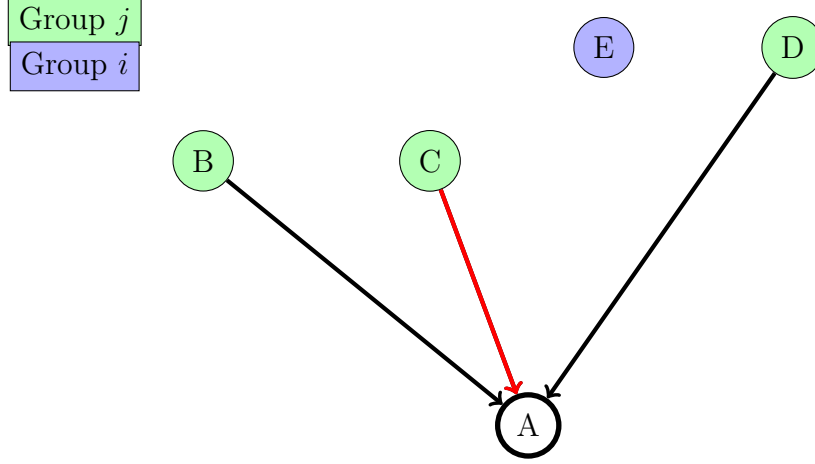
Notes: The figure illustrates the distance calculation for group j for municipality A at a particular year t. For each municipality, I consider its largest population center to calculate distances. I first calculate all pairwise distances from municipality A to all other municipalities where group j 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.

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.

4.2   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. Otherwise, this would be evidence against the instrument's independence of potential outcomes. To test this, I run the following fixed effect regression

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

I log all distances 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 yearly log of population counts as imputed by the census, which is not lagged. 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 electricity 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.[13] 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. 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.

To complement this analysis, I run a similar regression which controls separately for each

---

[13]Only agricultural revenue and total police officers have zero-valued covariates. Agricultural revenue has 107 observations with zero and total police officers has 394 observations with zero.

23

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. 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 in column 5 is reported biannually for 2012, 2014, 2016 and 2018. The row labeled Outcome mean displays the untransformed means of the outcomes, in levels.

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 preceding 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 DTOs in the preceding year.

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

groups' distance:

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

Where Z_{gmt-1} are the individual DTO distances based on its previous year network. The third panel estimates equation 3, without controlling for W_{mt}. Here instead of showing a single coefficient, I display the F statistic and associated p-value testing 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, not other observable covariates. Indeed, adding W_{mt} 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 evidenced by the electricity covariates, or places with less police enforcement, as evidenced by column 5. This is reassuring for the proposed empirical strategy. Furthermore, Section 8.4.3 conducts a robustness check to assess the sensitivity to controlling for agricultural revenue. The main results are robust to controlling for agricultural revenue.

5 Two Stage Least Squares and Its Challenges

In this section I present preliminary two stage least squares (TSLS) analysis. I show there is no weak instrument problem and present evidence for treatment effect heterogeneity in my setting. I discuss how treatment effect heterogeneity complicates TSLS analysis for multi-valued treatments, which is required to flexibility study the non-linear effects of DTO count. **I conclude TSLS is not an appropriate strategy to answer my research question.**

5.1 Two Stage Least Squares Results

The TSLS regressions I run are of the form:

$$N_{mt} = \alpha_m^1 + \tau_t^1 + W_{mt}\delta^1 + \sum_{g=1}^{G} \beta_g^1 Z_{gmt-1} + u_{mt}^1 \tag{4}$$

$$Y_{mt} = \alpha_m^2 + \tau_t^2 + W_{mt}\delta^2 + \gamma^2 N_{mt} + u_{mt}^2 \tag{5}$$

Where Y_{mt} is either the count of homicides or 9th grade dropout rate, W_{mt} is as [14] in Equations 2 and 3 with the additional inclusion of the IHS of municipal agricultural output, 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.[15] 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.

Furthermore, I compute the Sargan-Hansen J-statistic, which uses the over-identified

---

[14]These were: the lagged presence indicators for each of the 8 DTOs, the inverse hyperbolic of the total active DTO count for previous year and the yearly log of population counts as imputed by the census.

[15]I view this imputation as an undesirable yet necessary feature of the TSLS analysis. In this and other settings, groups will naturally come into and out of existence. TSLS requires an awkward imputation to handle this. In contrast, my main method of Section 6 handles this differently. There, I simply set the group's probability of entry to be zero if they are non-existent.

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 DTOs in the preceding year. All models include year and municipality fixed effects.

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

model to test the validity of the TSLS estimates with constant effects. For both the homicide and dropout outcomes I confidently reject the constant treatment effects model. This is suggestive of heterogeneous effects for the γ^2 coefficient in Equation 5, as argued in Rose and Shem-Tov (2021). 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 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 directly 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.[16] 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

---

[16]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}

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 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 intuition is that the TSLS coefficient will make the wrong counterfactual comparison. With a binary treatment, a shift in the instrument can only place an observation into one counterfactual treatment status. In contrast, with multiple treatment doses, a shift in the instruments can place a unit into several possible counterfactual doses, and as a result **TSLS coefficients provide uninterpretable estimates of effects.** 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, including 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 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

Having shown evidence for the instrument's validity, and discussed the shortcomings of TSLS in my setting, this section lays out the econometric framework I use to estimate treatment effects. Estimation is a simple two step procedure. In the **first step** I estimate the DTOs' probability of entry, in a model consistent with IO models of firm entry. These probabilities of entry are consistent with the entry models used in the IO literature. In **the second step**, I run a regression of my outcomes on the probability of treatment status and a control function term, like in Heckman (1979). The instruments facilitate the construction of the control function.

6.1 Model Setup

We observe $1...M$ municipalities over $1...T$ time periods. There are $1...G$ DTOs that can enter or exit each municipality in each time period. The *group count* is the sum of active DTOs operating in a single municipality. I am interested in estimating different treatment effects for each possible group count, which ranges from $0...G$.[17] For an outcome Y_{mt}, I denote the vector of potential outcomes as

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

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

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

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

6.2 First-Stage Model

I model the first stage as a private information group entry game. In this game, groups observe a vector of covariates which are common knowledge, and also privately observe a group-specific profitability shock which other groups do not observe. However groups know the distribution of rival groups' shocks. Using this distribution, groups form expectations on rivals' actions based on the common knowledge covariates. This group entry game is consistent with the structural models of entry widely used in the industrial 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. Entry is not permanent, and groups can exit municipalities after entry. 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 unexpected manpower losses in a turf battle, unexpected profits from a kidnapping 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
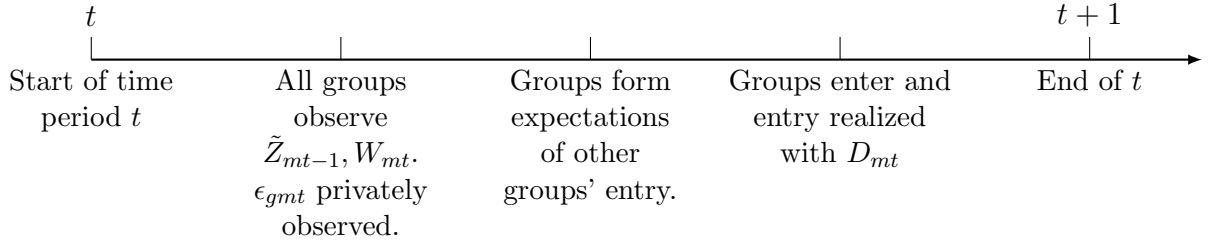
The private information framework is especially appealing in the context of the criminal sector. Criminal organizations strive to be clandestine and hide from competitors and the government. Furthermore, reports on competitor's actions are likely hindered by the decentralized nature of criminal organizations. These organizations benefit from a loose chain of command where underlings cannot freely report to more important actors in the organization, since keeping them separated mitigates the damage done by an informant to the government or a rival (Von Lampe, 2015). Taken together, it is realistic to think that criminal organizations have accurate perceptions of their own profitability, yet are unaware of their competitor's conditions.

I place the following standard 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 competitors' 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.[18] This assumption is standard, follows the IO literature and has been extensively studied (Aguirregabiria and Mira, 2007; Collard-Wexler, 2013). Under mild conditions, $v_{gm}(\cdot)$ is identified for each group (Pesendorfer and Schmidt-Dengler, 2008; Blevins, 2014).[19]

6.2.1 Normalizing Probabilities

It is convenient to normalize the probabilities of entry. This is a standard transformation in the literature and imposes no further assumptions (Heckman et al., 2006b).

$$v_{gm}(\tilde{Z}_{mt-1}, W_{mt}) > \epsilon_{gmt}$$
$$F_\epsilon[v_{gm}(\tilde{Z}_{mt-1}, W_{mt})] > F_\epsilon(\epsilon_{gmt}) \tag{9}$$
$$Q_{gm}(\tilde{Z}_{mt-1}, W_{mt}) > e_{gmt}$$

---

[18]Appendix Section D explains what kinds of entry games are consistent with the model used for estimation.
[19]With forward looking groups, sufficient conditions for identification of v_g would be: 1) The distribution of ϵ and the discount factor is known 2) Group g's per period profits do not depend on the other groups' incumbency choices, $D_{mt-1,-g}$ where $-g$ excludes group g. 3) 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.

6.3 Second-stage

I model the potential outcomes as:

$$Y_{mt}(n) = \alpha_n + \psi(W_{mt}) + U_{mt}(n) \tag{10}$$

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

$$E[Y_{mt}(n) - Y_{mt}(0)] = E[\alpha_n - \alpha_0]$$

Hence by estimating α_n we can obtain the treatment effects of interest.

Note that equation 10 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 $E[Y_{mt}(n)|W_{mt}, \mathbf{e}_{mt}] = \alpha_n + \psi(W_{mt}) + E[U_{mt}(n)|\mathbf{e}_{mt}]$

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 the vector of past entry decisions and 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 by implicitly changing entry probabilities for time t. The same logic applies to other covariates in W_{mt}.

6.4 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:

$$E[Y_{mt}|Q_m(\tilde{Z}_{mt}, W_{mt}) = q, W_{mt}, N_{mt} = 1] = \alpha_1 + \psi(W_{mt})$$
$$+ E[U(1)|q_1 > e_1, q_2 < e_2]P[q_1 > e_1, q_2 < e_2|\{q_1 > e_1, q_2 < e_2\} \cup \{q_1 < e_1, q_2 > e_2\}, W_{mt}]$$
$$+ E[U(1)|q_1 < e_1, q_2 > e_2]P[q_1 < e_1, q_2 > e_2|\{q_1 > e_1, q_2 < e_2\} \cup \{q_1 < e_1, q_2 > e_2\}, W_{mt}]$$

$$(11)$$

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.[20] This representation exploits the exclusion restriction, Assumption 3, and the first stage model of entry.[21]

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

---

[20]Proofs are found in the appendix

[21]Recall entry occurs if $e_g < q_g$, using the normalization in Equation 9.

that group 1 enters but 2 does not, conditional on one of the two groups entering.[22] ψ can 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 11 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 $E[U(1)|q_1 > e_1, q_2 < e_2], E[U(1)|q_1 < e_1, q_2 > e_2]$.

6.5 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:

$$1[D_{mt}(q) = d] := 1[\{e_{gmt}\}_{g=1}^G \in \{[0,1]^G | D_{mt} = d, Q_{mt}(Z) = q\}] \tag{12}$$

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

---

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

who enter given entry probabilities q. That is

$$N_{mt}(q) := \sum_{g=1}^{G} D_{gmt}(q) \tag{13}$$

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, $Q_m(\tilde{Z}_{mt}, W_{mt})$:

$$
\begin{aligned}
E[Y_{mt}|Q_m(\tilde{Z}_{mt}, W_{mt}) &= q, W_{mt}, N_{mt}(q) = n] \\
&= \alpha_n + \psi(W_{mt}) + \sum_{d \in \mathcal{N}(n)} E[U_{mt}(n)|D_{mt}(q) = d]P[D_{mt}(q) = d|N_{mt}(q) = n]
\end{aligned} \tag{14}
$$

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 α_n, similar to a Heckman (1979) style procedure.

7 Identification and Estimation

After explaining the conceptual framework to estimate treatment effects, 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. For example, consider the fixed effect model $D_{gmt} = 1[\phi_m + \delta\tilde{Z}_{mt} + \gamma W_{mt} > \epsilon_{gmt}]$. The effect ϕ_m is an additive effect.

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.[23] Rather, there will be interactions between the other covariates $(\tilde{Z}_{mt}, W_{mt})$ and the municipality fixed effect.

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 $(\tilde{Z}_{mt}, W_{mt})$ 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. Again there will be interactions between the municipality fixed effects .

Third, 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 (Bonhomme et al., 2022; Saggio, 2012). The grouped fixed effect framework is a tractable procedure that solves the last three problems. The first and second problems are solved by allowing *interactions* between the municipality fixed effect and the $(\tilde{Z}_{mt}, W_{mt})$ covariates. First, these interactions account for time-invariant differences while still being coherent with the non-linearities of the entry game model. Second, multiple equilibria are transparently accounted for. The interactions allow for the effects of $(\tilde{Z}_{mt}, W_{mt})$ to vary across m, much like the function v_{gm} can vary across m. Third, the group fixed effect procedure avoids the incidental parameter problem of Neyman and Scott (1948) by reducing the dimensionality of the unobserved heterogeneity.

The substance of the procedure involve using k-means clustering (Steinley, 2006) to group municipalities into C groups, with $C < M$, where M is the number of municipalities. 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

---

[23]Section D.1 explains this in more detail.

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, and this informs the time-invariant differences across municipalities. 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. Intuitevly, all that is needed for multiple equilibria is to allow the effects of $(\tilde{Z}_{mt}, W_{mt})$ to vary by municipality type, so that their effects can reflect the different underlying equilibria. 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

This section describes the functional forms used to estimate each DTO's probability of entry. Entry decisions are made at the DTO, municipality and year level. I pool all of the G groups

entry decisions and estimate the following probit for the first stage.

$$D_{gmt} = \underbrace{\xi}_{\text{Constant}} + \underbrace{X_{mt}\beta_1^1}_{\text{Controls}} + \underbrace{\phi_{0g}^1 1[G=g]}_{\text{DTO Fixed Effect}} + \underbrace{L_t\phi_1^1 + \sum_{c=1}^{C} 1[C(R_m) = c] \times \phi_{2c}^1}_{\text{Time and Municipality Type Fixed Effects}}$$

$$+ \underbrace{\sum_{c=1}^{C} 1[C(R_m) = c] \times \left[\gamma_{1c}^1 D_{gmt-1} + \gamma_{2c}^1 \times IHS(N_{mt-1}) \right]}_{\text{Municipality Type Specific Controls}} \qquad (15)$$

$$\underbrace{\sum_{c=1}^{C} 1[C(R_m) = c] \times \left[\gamma_{3c}^1 \tilde{Z}_{gmt-1} + \gamma_{4c}^1 \tilde{Z}_{gmt-1}^2 + \gamma_{5c}^1 \overline{\tilde{Z}}_{-gmt-1} + \gamma_{6c}^1 \overline{\tilde{Z}^2}_{-gmt-1} \right]}_{\text{Municipality Type Specific Instruments}} + \epsilon_g$$

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.[24]

7.3 Second-Stage Estimation

I now explain the functional forms used in the second stage regression. In contrast to the first stage which is at the DTO-municipality-year level, the second stage is at the municipality-year level. The reason is that each DTO's decision will jointly determine the municipality-year's treatment dose, the number of active DTOs. Using the previous first stage model I

---

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

estimate the entry probabilities for all players and collect these in $Q_{mt}(Z)$. I parametrize the potential outcomes to be of the form:

$$
\begin{aligned}
Y_{mt}(n) = &\underbrace{\alpha_n}_{\text{Treatment Effect}} + \underbrace{X_{mt}\beta_1^2}_{\text{Controls}} + \underbrace{L_t\phi_1^2 + R_m\phi_2^2}_{\text{Time and Muni. FE}} + \\
&\underbrace{\sum_{c=1}^{C} 1[C(R_m) = c] \times \left[\sum_{g=1}^{G} \gamma_{1g}^2 D_{mtg-1} + \gamma_{2c}^2 IHS(N_{mt-1})\right]}_{\text{Municipality Type Specific Controls}} + U_{mt}(n)
\end{aligned}
\tag{16}
$$

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. Note additionally that I do not restrict the municipality fixed effects in the second stage to use the same clusters from the grouped fixed effects procedure. Here, each municipality gets its own fixed effect. Equation 16 has M fixed effects captured in $\phi_2^2 R_m$, while Equaltion 15 only has K fixed effects for the K clusters given by $C(R_m)$. This is because the model is linear and these additional fixed effects do not cause substantive problems. This approach follows Cornelissen et al. (2018) in modeling fixed effects for this kind of estimation. Finally, I check in Section 8.4.3 the robustness of my findings to the specification in Equation 16. The results are robust to alternative versions of this model.

7.4 Control Function

I next show the functional form for the control function and the final estimating equation. The control function is given by the $E[U(n) - U(0)|D(q) = d]P[D(q) = d]$ terms in equation 14. As a benchmark, I assume a normal model as in (Heckman, 1979; Kline and Walters, 2016; Dahl, 2002). The normal model is familiar to many 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:

$$E[U_{mt}(n) - U_{mt}(0)|D_{mt}(q) = d]P[D_{mt}(q) = d] = \sum_{g=1}^{G} (\sigma_{ng} - \sigma_{0g})\lambda(d_g, q_g) \qquad (17)$$

where d_g, q_g are the gth components of an entry vector $d \in \mathcal{D}$ and the vector of entry probabilities q, and λ is a known function.[25] 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}

---

[25]

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

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:

$$E[Y_{mt}|Q(\tilde{Z}_{mt}, W_{mt}) = q, W_{mt}] = \underbrace{\alpha_0}_{\text{Intercept}} + \underbrace{X_{mt}\beta_1^2 + \sum_{g=1}^{G}\beta_{2g}^2 D_{mtg-1}}_{\text{Controls}}$$

$$+ \underbrace{\phi_1^2 L_t + \phi_2^2 R_m}_{\text{Muni. Time FEs}} + \underbrace{\sum_{c=1}^{C}\gamma_{1c}^2 1[C(R_m) = c] \times IHS(N_{mt-1})}_{\text{Municipality Type Specific Controls}}$$

$$+ \sum_{n=0}^{G}\left\{\underbrace{(\alpha_n - \alpha_0)P[N(q) = n|W_{mt}]}_{\text{Treatment Effect}} + \underbrace{\sum_{g=1}^{G}(\sigma_{ng} - \sigma_{0g})\sum_{d \in \mathcal{N}(n)}\lambda(d_g, q_g)}_{\text{Control Function}}\right\}$$

$$(18)$$

Notice that equation 18 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 $P[N(q) = n|W_{mt}]$. Given the individual groups first stage probabilities and the independent errors between groups, this term is a simple function of the individual group probabilities.[26] 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 15. I then use these probabilities to construct the appropriate terms in equation 18 and run OLS.

7.5   Inference and Sample Restrictions

Additionally, as a sample restriction, I drop observations where the unit has a probability of 1 for one of the possible treatment doses, the group counts. These observations are, according to my model, not at risk to receiving any of the counterfactual group counts considered. They therefore are not informative of *any* counterfactual treatment dose, since they are

---

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

guaranteed to be placed in only one treatment dose. They are therefore poor control units when trying to estimate treatment effects. I drop 3134 observations with this restriction.[27] These are mainly small municipalities at the beginning of the sample period, that did not have any DTO and were far away from any DTO. These observations are not informative of counterfactual outcomes with a non-zero group count.[28]

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.

---

[27]To operationalize this procedure, I compute the estimated probability for each possible group count, ranging from 0 DTOs to 8 DTOs. I round this probability to two digits, and exclude observations where one group count probability is equal to 1.

[28]These types of restrictions are commonplace in applied work. Panel methods effectively drop observations with no within-group variation. With binary treatments, propensity score matching drops observations with probability 1 or 0 for treatment. In my setting, after dropping probability 1 observations, I keep observations that may have probability 0 for some treatment doses. These observations are informative for *some* but not *all* of the treatment doses, and it is therefore useful to keep them in the sample.

8 Results

With the econometric model in hand, I next show first stage and second stage results for homicides and outcomes. I document that increased in DTOs increases homicides and school dropout, largely for male and older students. Furthermore, I find supporting evidence for the hypothesis that DTOs actively hire students dropping out.

8.1 First Stage

Table 4 displays results of Equation 15 in column 1, estimated with probit. The results show DTOs behave as if they are strategic substitutes, with closer rival DTOs discouraging entry. Table 4 shows other variations of the main equation 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% decrease in a DTO's own distance leads to a 1 percentage point (pp) increase in the likelihood of entry, while a 1% decrease 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 15 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 grouped fixed effect strategy matters for correctly

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 15. 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 15. 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.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

modeling the entry decisions, and that unobserved heterogeneity matter.

Both the Bayesian and Akaike Information Criterion (BIC,AIC), widely used measures to evaluate model fit, 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. The results are shown in Figure 7. Each point is an estimated coefficient 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 $E[Y(n)-Y(0)]$.

In the top panel we observe a clear upwards trend in the count of homicides: homicides tend to increase as more DTOs enter. The biggest jump in treatment effects happens with 6 DTOs, with an estimate coefficient of 117 extra homicides which is more than double of any of the previous coefficients. The estimates stabilize thereafter with 150 extra homicides with 7 or 8 DTOs. Although the standard errors are wide, this increasing "S" shape display important non-linearities and suggest reducing DTO presence would have the largest impact with 6 DTOs.

The coefficients for 1 to 2 DTOs are insignificant, but the point estimates still reveal important increases in homicides. The coefficient for 1 DTOs is 14.11, for 2 DTOs is 25.58. The overall mean in the estimation sample is 8.7 homicides per municipality-year. Hence the coefficients for having 1 and 2 DTOs translates to doubling and tripling the average number of homicides in the estimation sample. Even if the estimates are insignificant, the point estimates suggest an important increase in homicides with a monopolist DTO. This contrasts previous studies which find more precise null effects or even negative effects for monopolist criminal organizations (Sobrino, 2020; Biderman et al., 2019)

47

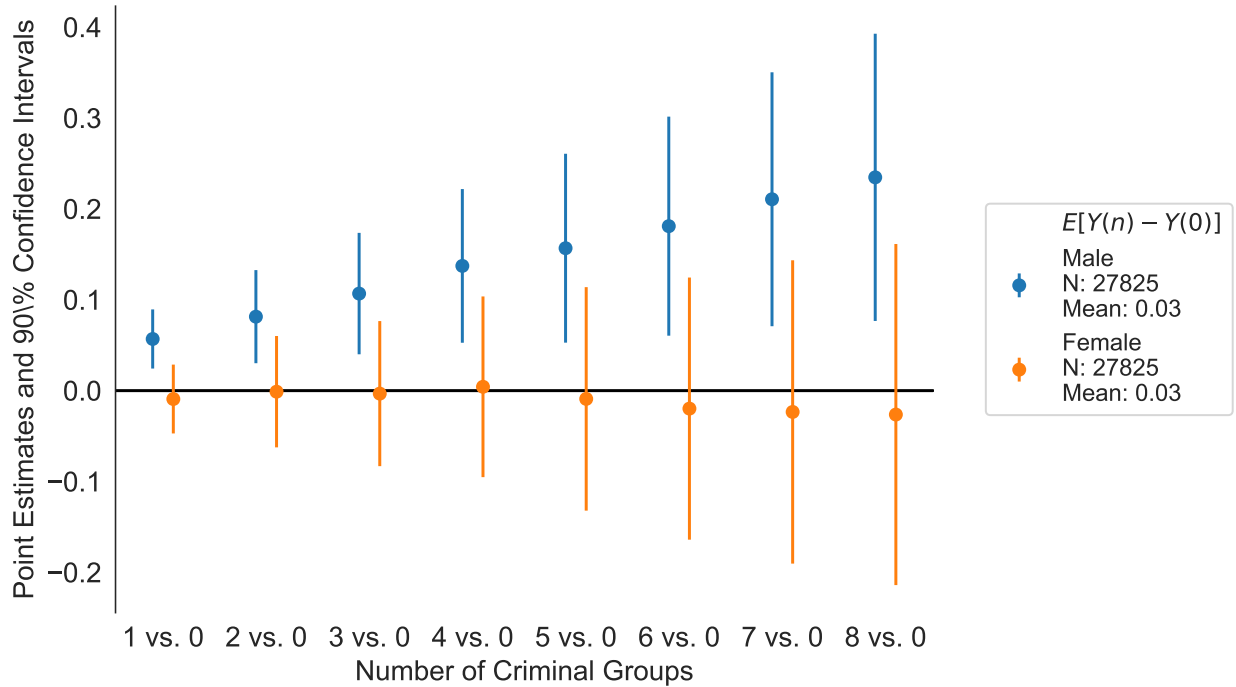Figure 7: Homicide and Dropout ATE Estimates

(a) Homicides
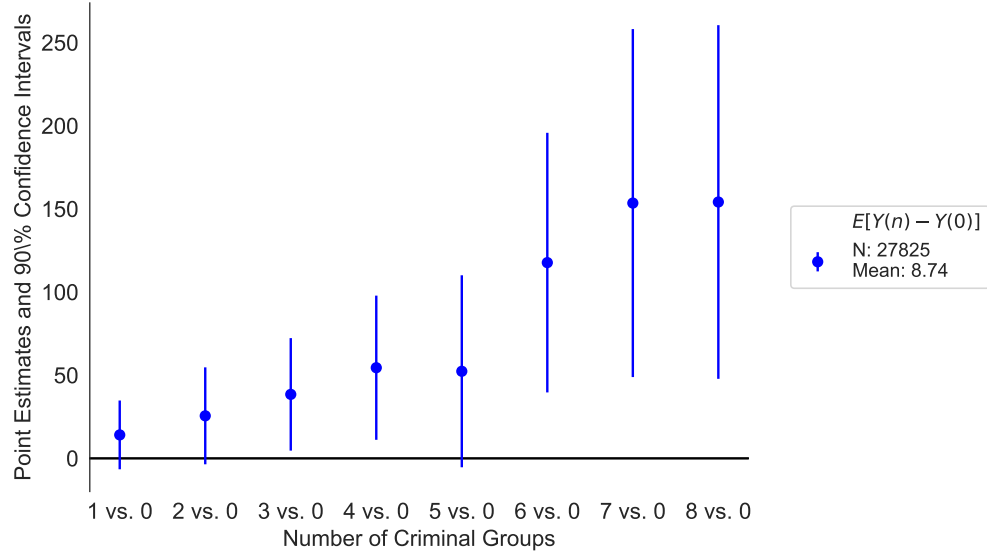
(b) Dropout 9th Grade
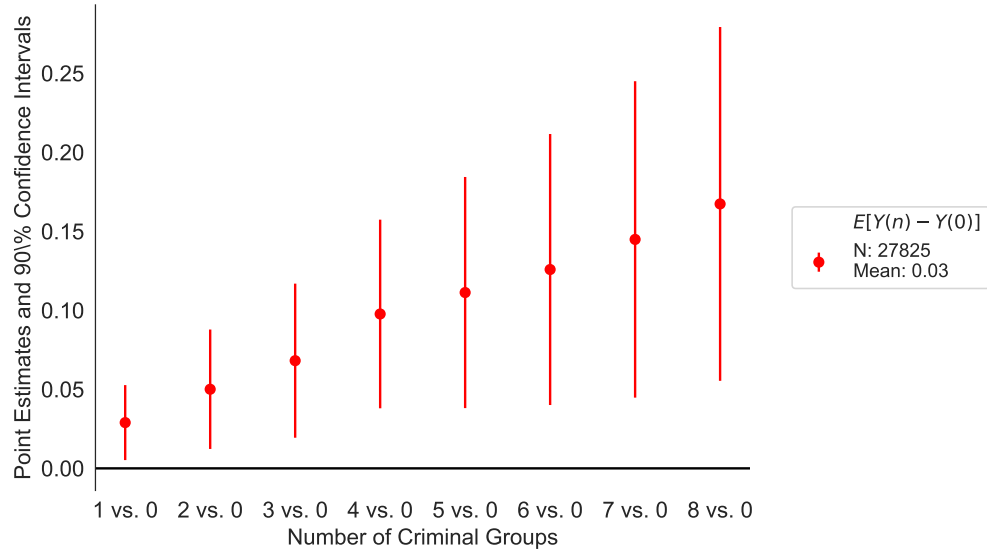
Notes: This figure plots Average Treatment Effect estimates, $E[Y(n)-Y(0)]$, using the normal model outlined in sections 6 and 7. 90% confidence intervals shown. Standard errors from 1000 bootstrap replications.

In contrast, the effects are linear and more pronounced for 9th grade dropout. Each coefficient is statistically significant at the 90% confidence level. 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.[29]

Figure 8: 9th Grade Dropout by Gender

Notes: This figure plots Average Treatment Effect estimates, $E[Y(n) - Y(0)]$, using the normal model outlined in sections 6 and 7. The outcomes are ninth grade dropout for males and females. 90% confidence intervals shown. Standard errors from 1000 bootstrap replications.

Figure 8 examines how these effects vary by the gender of the students. The figure displays

_____

[29]I do not study high school dropout in this paper because high school was not universal or compulsory throughout my study period. In 2006 there are more than 500 municipalities without any high schools. High school became compulsory and universal in 2012, but even still there are roughly 200 municipalities in my data without any high schools in 2018.

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 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. Moreover, in Section 8.4.1 I examine migration in closer detail and also find that it does not drive the dropout effects.

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 dropouts. 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 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

Figure 9: Homicides Ages 12-18

(a) Homicides Ages 12-14
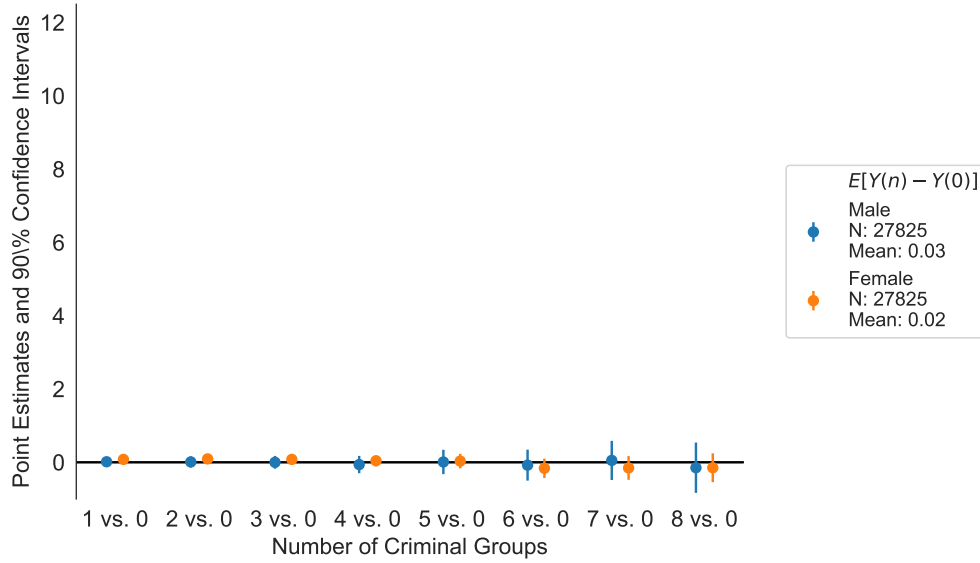
(b) Homicides Ages 15-18
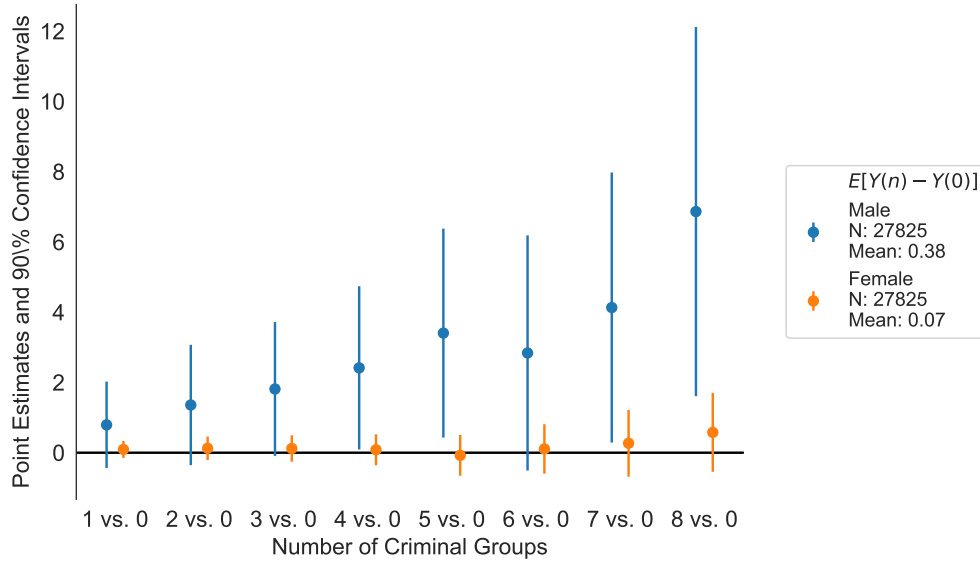
Notes: This figure plots Average Treatment Effect estimates, $E[Y(n) - Y(0)]$, using the normal model outlined in sections 6 and 7. The outcomes for panel (a) are homicides with a 12-14 year old victim, by gender. Panel (b) displays outcomes for homicides with a 15-18 year old victim, by gender. 90% confidence intervals shown. Standard errors from 1000 bootstrap replications.

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 1-5, 6-7, 8 and 9. I group grades 1-5 and 6-7 into two separate categories. 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 are also small and insignificant. Overall, I do not find evidence that females are affected by DTO entry, although the standard errors are large.

In contrast I find positive significant effects for male students. First, we again see small and mostly insignificant effect for grades 1-5. This is encouraging, since it again suggests migration is not driving the response to DTO presence. Similarly, grades 6 and 7 also display small and insignificant coefficients.

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.

Figure 10: Dropout by Grade and Gender

(a) Dropout for Female Students
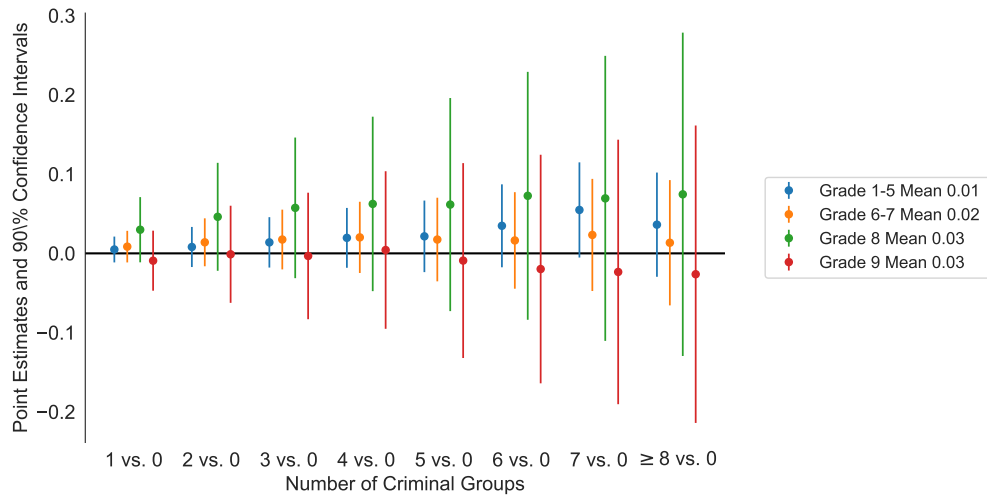
(b) Dropout for Male Students
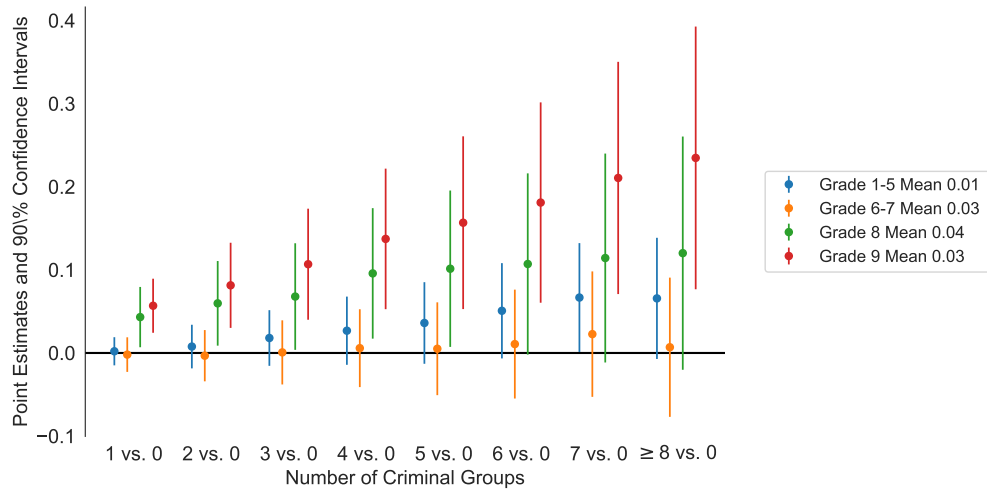
Notes: Standard errors from 1000 bootstrap replications.

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 Robustness Checks

I conduct several robustness checks to validate the main findings.

8.4.1 Migratory Response

As I argued before, the gendered effects preclude migration to be the main driver of the dropout results. It is unlikely households selectively choose to migrate, a costly decision, based on their children's gender. Nonetheless, in this section I use individualized test data to directly test if migration is an important mechanism.

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.

8.4.2 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.

8.4.3 Specification Checks

My identification strategy relies on one year lagged distance satisfying an exclusion restriction. 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 16.[30]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

---

[30]As in Carneiro et al. (2011), I do not include the current distances in the first stage since these would not be in the information set used by DTOs to decide entry.

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.

8.4.4 Robustness to Sample Restriction

As explained in Section 7.5, I drop observations that have probability 1 for at least one of the possible treatment states. I do so because these observations are not at risk for *any* treatment status and therefore are not informative. 3134 observations are dropped. These observations are typically from small municipalities with no DTOs, and are additionally far from any DTO, as can bee seen in Appendix Table A6. Furthermore the dropped observations are concentrated at the beginning of the sample period, before the significant expansion of DTOs' in Mexico.

I assess the robustness of my main findings by including the entire sample. Results are displayed in Appendix Figure A4. As can be seen the estimates are virtually identical with or without the sample restriction.

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 linear effects for for school dropout. The effects are concentrated in older and male students, which suggests at least some students are being pulled into criminal activity. Corroborating this hypothesis, I showed homicides with 15-18 year old males victims also increase with more DTOs. There is no corresponding effects for female students.

In contrast, the increases for homicides have an increasing S-shape in the number of DTOs, with an inflection point at 6 DTOs. Overall, the patterns I find for homicides are inconsistent with the predictions of models of crime Hirshleifer (1995), which typically predict an increasing concave function with a horizontal asymptote. I conjecture this is likely due to the simplifying assumptions needed for theoretical analysis. For instance, these models typically assume groups have no capital they bring when entering. It is plausible to think that the DTOs in Mexico have important criminal know-how and reputations which can increase the production of crime as a form of criminal capital. In this scenario, when an additional group enters they wouldn't only diminish the market power of the existing groups. An additional entrant would also increase the stock of criminal capital, therefore leading to more violence and crime. Nonetheless, this is speculation on my part and would require future theoretical research.

In answering this research question I developed a novel selection framework to estimate treatment effects for questions regarding market structure. 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. For instance, treatment effects could vary both by the current number of DTOs and last year's number as well. 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*.

Biderman, C., J. M. P. De Mello, R. S. De Lima, and A. Schneider (2019, September). Pax Monopolista and Crime: The Case of the Emergence of the Primeiro Comando da Capital in São Paulo. *Journal of Quantitative Criminology 35*(3), 573–605.

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.

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

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.

de Greef, K. (2019, August). As Gang Murders Surge, South Africa Sends Army to Cape Town, and the City Cheers. *The New York Times*.

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.

FBI (2023). Table 1. https://ucr.fbi.gov/crime-in-the-u.s/2018/crime-in-the-u.s.-2018/tables/table-1/table-1.xls.

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.

Guardian, T. (2016, September). Hobos gang trial puts deadly turf war ravaging chicago front and center.

Heckman, J., S. Urzua, and E. Vytlacil (2006a). 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.

Heckman, J. J., S. Urzua, and E. Vytlacil (2006b, August). Understanding instrumental variables in models with essential heterogeneity. *Review of Economics and Statistics 88*(3), 389–432.

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?

INEGI (2020). Homicidios a nivel nacional. Serie anual de 1990 a 2019.

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/.

Institute for Economics and Peace (2015). Global peace index.

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.

Jones, N. P. (2016). *Mexico's Illicit Drug Networks and the State Reaction*. Georgetown University Press.

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.

Lee, S. and B. Salanié (2018, November). Identifying effects of multivalued treatments. *Econometrica : journal of the Econometric Society 86*(6), 1939–1963.

León, G. (2012). Civil Conflict and Human Capital Accumulation: The Long-term Effects of Political Violence in Perú. *The Journal of Human Resources 47*(4), 991–1022.

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.

Lopez Cruz, I. and N. A. Pairolero (2019). Criminal recruitment or victimization: The effects of the mexican drug war on education. *Available at SSRN 3480335*.

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.

Post, M. D. (2022, September). Mexican cartels are increasingly harming civilian communities in their turf battles.

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.

Von Lampe, K. (2015). *Organized Crime: Analyzing Illegal Activities, Criminal Structures, and Extra-Legal Governance*. Sage Publications.

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

	(1) All Homi.	(2) Male Homi.	(3) Female Homi.	(4) Gun Homi.	(5) Potentially DTO Related Homi.
1 vs. 0	14.11	12.05	0.63	14.78	10.61
	(12.57)	(11.50)	(1.23)	(11.84)	(9.33)
2 vs. 0	25.58	22.46	1.06	25.37	18.11
	(17.72)	(16.24)	(1.70)	(16.66)	(13.15)
3 vs. 0	38.48*	34.30*	2.09	36.33*	25.46*
	(20.57)	(18.84)	(1.90)	(19.08)	(14.75)
4 vs. 0	54.51**	49.95**	2.46	49.71**	34.93*
	(26.36)	(24.11)	(2.41)	(23.97)	(17.92)
5 vs. 0	52.37	49.19	1.06	47.15	37.69*
	(35.10)	(32.09)	(3.32)	(31.13)	(22.39)
6 vs. 0	117.71**	107.35**	8.26*	98.32**	55.69**
	(47.46)	(43.06)	(4.56)	(40.47)	(26.19)
7 vs. 0	153.52**	141.26**	9.98	128.66**	83.32**
	(63.61)	(57.20)	(6.56)	(53.78)	(33.93)
8 vs. 0	154.16**	141.51**	10.10	128.16**	88.52**
	(64.65)	(58.75)	(6.36)	(55.13)	(36.03)
Mean	8.74	7.83	0.89	6.63	5.01
N	27825.0	27825.0	27825.0	27825.0	27825.0

Notes: Unit of observation is municipality-year for years 2006-2018. Average Treatment Effect estimates, $E[Y(n) - Y(0)]$, using the normal model outlined in sections 6 and 7. The outcomes differ across columns. Column 1 uses all homicides. Column 2 examines homicides for males, column 3 for females. Column 4 examines homicides with a firearm, Column5 examines homicides for males aged 15-39, an age group very likely to be a part of criminal activity as in Calderón et al. (2015). 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*	0.03**	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**	0.05**	0.06***	0.05**
	(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**	0.07**	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***	0.09**	0.10***	0.10***	0.12***	0.11***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
5 vs. 0	0.11**	0.12***	0.10**	0.11**	0.11***	0.14***	0.13***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)
6 vs. 0	0.13**	0.13***	0.12**	0.12**	0.13***	0.16***	0.17***
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)
7 vs. 0	0.14**	0.15***	0.14**	0.14**	0.16***	0.19***	0.20***
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)
8 vs. 0	0.17**	0.18***	0.16**	0.17**	0.18***	0.22***	0.22***
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.08)
	Panel B: Homicides						
1 vs. 0	14.11	13.62	8.65	12.74	12.44	5.28	5.27
	(12.57)	(11.93)	(7.59)	(11.56)	(8.09)	(8.41)	(6.04)
2 vs. 0	25.58	24.92	14.31	23.52	19.61*	12.37	9.53
	(17.72)	(16.76)	(10.54)	(16.30)	(11.16)	(11.93)	(8.68)
3 vs. 0	38.48*	37.54**	17.79	36.14*	24.38*	23.36*	11.53
	(20.57)	(18.85)	(13.62)	(18.57)	(13.92)	(13.73)	(10.64)
4 vs. 0	54.51**	53.97**	24.12	51.96**	32.54	31.32*	12.84
	(26.36)	(24.08)	(20.27)	(23.91)	(20.35)	(18.69)	(16.35)
5 vs. 0	52.37	51.20	11.08	49.05	21.12	21.78	9.91
	(35.10)	(32.87)	(31.39)	(32.50)	(31.34)	(28.34)	(24.04)
6 vs. 0	117.71**	115.98***	55.21	113.93**	68.48	81.79**	34.30
	(47.46)	(44.96)	(45.82)	(44.66)	(44.95)	(36.73)	(48.38)
7 vs. 0	153.52**	150.91**	61.40	150.90***	75.06	112.66**	40.33
	(63.61)	(59.27)	(63.74)	(58.54)	(62.87)	(46.30)	(56.56)
8 vs. 0	154.16**	152.39**	45.17	153.97**	56.23	81.47	-23.56
	(64.65)	(61.00)	(66.36)	(60.81)	(64.99)	(50.22)	(58.26)
# Muni. Types	5	5	5	5	5	10	20
Log Pop. Control	✓	✓	-	✓	-	✓	✓
IHS Ag. Control	✓	✓	-	-	✓	✓	✓
Dist.$_t$ Controls	-	✓	-	-	-	-	-
N	27825	27825	27889	27829	27978	26856	26271

Notes: Average Treatment Effect estimates, $E[Y(n) - Y(0)]$, using the normal model outlined in sections 6 and 7. The sample sizes change across columns because the underlying first stage estimates also change, and observations with probability 1 are excluded as explained in Section 7.5. Standard errors from 1000 bootstrap replications.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A6: Descriptive Statistics by Treatment Risk

	At Risk	Not At Risk	Overall
Covariates			
Population (2005)	82,719	12,827	43,432
Poverty Rate (2005)	0.56	0.75	0.67
Literacy Rate (2005)	0.88	0.79	0.83
Elect. Users	14,231	1,686	14,218
kWH (millions)	85.72	48.15	85.68
kWH per Capita	6,259	103,953	6,360
Ag. Revenue	182.08	36.90	167.39
Total Police	74	21	74
Year	2,012	2,006	2,012
Outcomes			
Homicide Rate (per 100,000)	16.61	10.40	15.98
9th Grade Enrollment	903.90	196.84	832.32
9th Grade Dropout Rate	0.03	0.03	0.03
9th Grade Dropout Rate (Girls)	0.03	0.02	0.02
9th Grade Dropout Rate (Boys)	0.03	0.02	0.03
Treatments and Instruments			
Number of DTOs	0.90	0	0.81
≥ 1 DTOs	0.33	0	0.30
Average Dist.$_{t-1}$ (km)	65.85	172.87	76.44
# Municipalities	2435	1346	3781
N	28521	3134	31655

 Notes: Unit of observation is municipality-year. This table displays averages for different variables. Observations are categorized according to whether they are at risk for a treatment state or not. An observation is at risk if it has a probability less than 1 for all 9 possible DTO group counts, including 0. Observations that have probability 1 for a DTO group count are not at risk, since the treatment status is degenerate. I calculate the probability of treatment status using to the probit in equation 15. I round the probability for each treatment status to two digits to determine which observations have degenerate probability. In practice the only observations that are not at risk are not at risk to have any DTO presence.

 The row labeled "Year" displays the average year for the municipality-year observations at risk and not at risk.
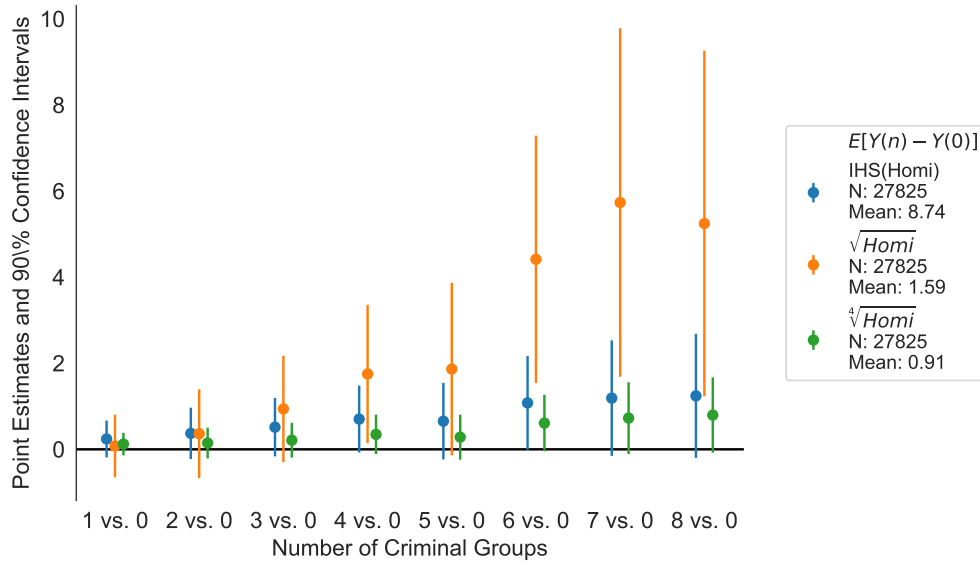
A.2 Figures

Figure A1: One at a Time TSLS estimates

(a) Homicides


(b) 9th Grade Dropout
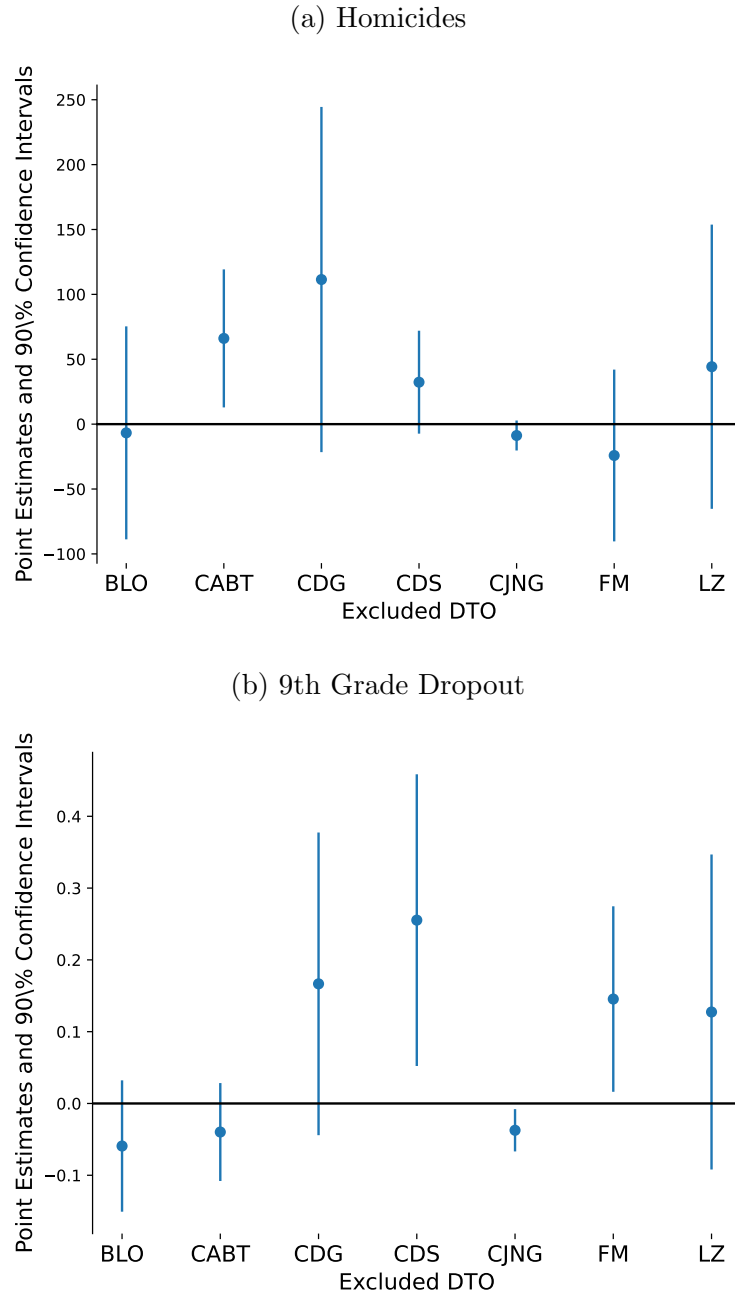
Notes: The figure captures effects along different complier margins, captured for each DTO group instrument. The unit of observation is municiaplity-year, sample years 2006-18. These figures display TSLS estimates using one DTO's distance as the instrument, with all the other distances included as controls. The distances are the one-year lagged distances, after applying the log transformation. The endogenous variable is the total number of active groups in municipality m at time t, N_{mt}. The group used as an instrument is labelled in the X axis. Cártel Juárez is excluded due to low variation; it is the DTO with the least expansion. I only exclude this group for this analysis, but include it in all other analysis. 90% confidence intervals shown, with standard errors clustered by municipality.

Figure A2: Homicide Results

### (a) Alternative Transformations for Homicides


(b) Homicide Rates
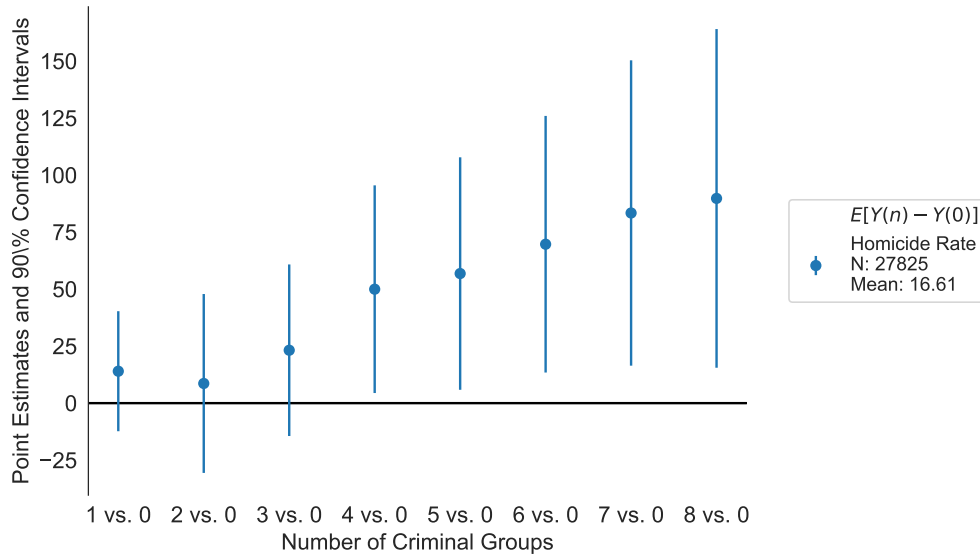
Notes: These figures display effects using alternative transformations for the homicide outcome. Panel A displays results using the inverse hyperbolic sine "IHS(Homi)", the square root transformation $\sqrt{Homi}$ and the quartic root transformation $\sqrt[4]{Homi}$. Panel B display results using homcide rates per 100,000 population. Standard errors from 1000 bootstrap replications.

Figure A3: ENLACE Out Migration Results
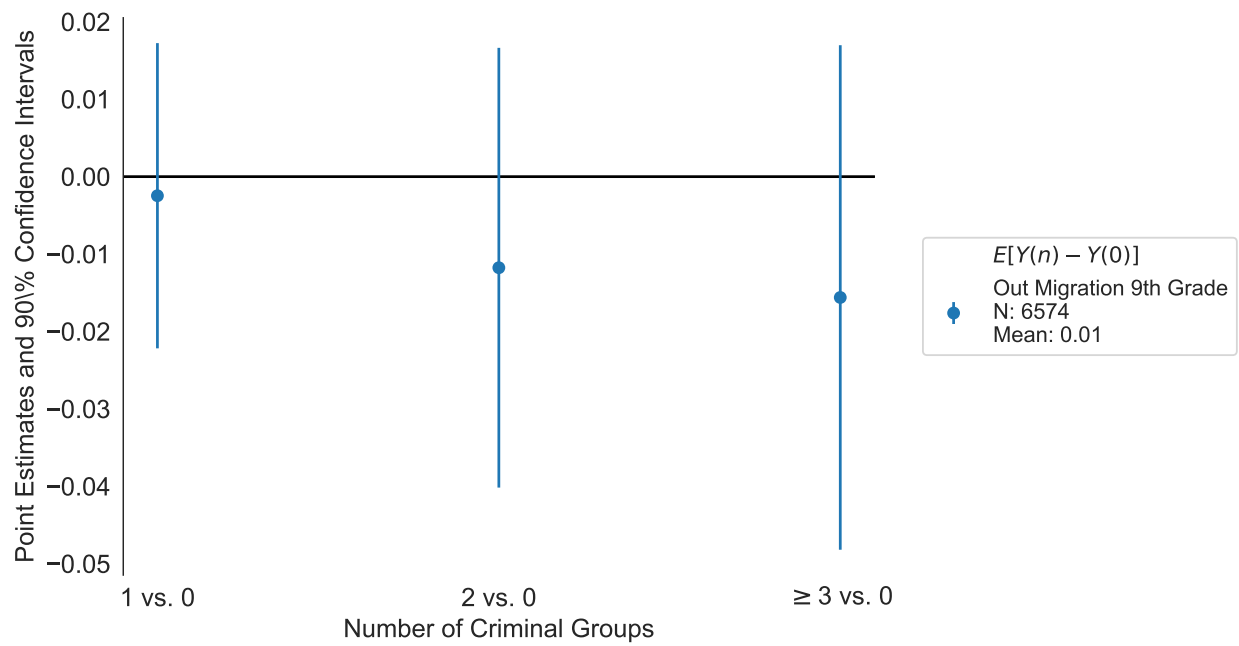
Notes: This figure displays estimates using outmigration as measured by ENLACE test data. Sample years are 2009-12. I limit the maximum number of DTOs to 3 to gain precision. Standard errors from 1000 bootstrap replications.

Figure A4: Comparison of Main Sample Results to Full Sample Results
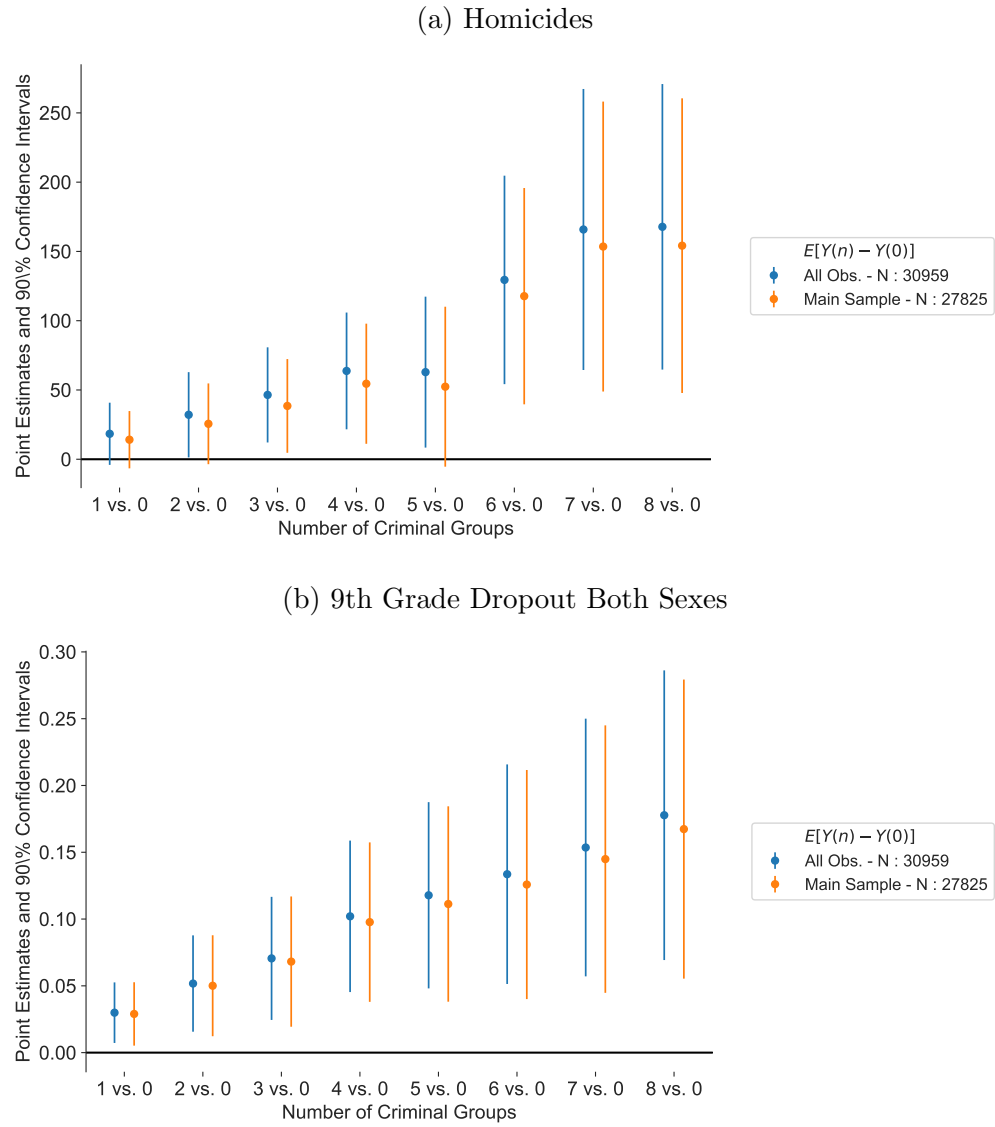
(a) Homicides
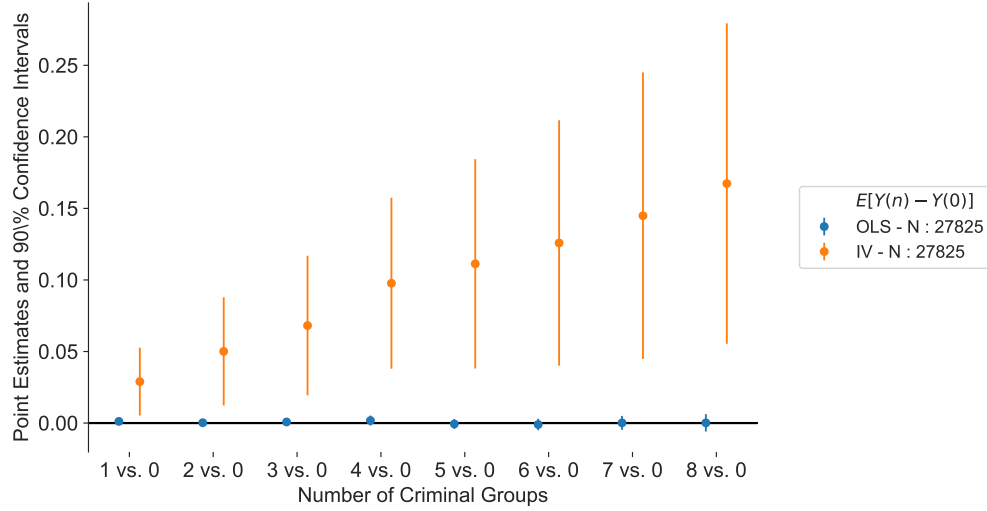

(b) 9th Grade Dropout Both Sexes

Notes: Unit of observation is municipality-year. Sample years are 2006-18. Panel (a) displays effects with homicides as outcomes, Panel (b) for 9th grade dropout for both males and females. This figure replicates the main estimates and compares them to estimates that include all observations. The main sample drop observations with probability 1 of having no DTO. "All Obs." instead keeps all observations. Standard errors from 1000 bootstrap replications.

Figure A5: Comparison to OLS Estimates

(a) Homicides
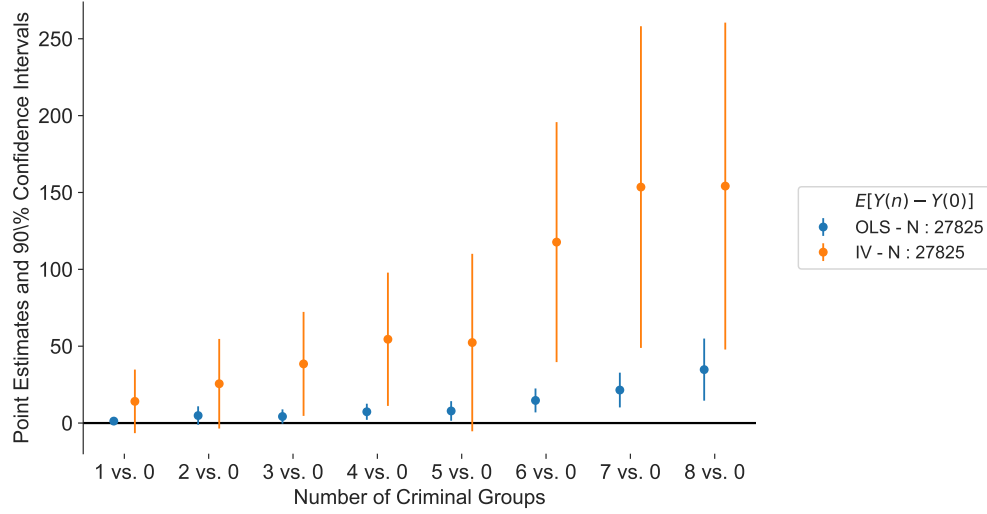
(b) 9th Grade Dropout Both Sexes

Notes: Sample years are 2006-18. Panel (a) displays effects with homicides as outcomes, Panel (b) for 9th grade dropout for both males and females. This figure replicates the main estimates and compares them to OLS estimates using the same sample and controls. The OLS estimates use the observed number of DTOs as the regressors of interest. Standard errors from 1000 bootstrap replications.

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:

$$\text{Homicides}_{mt} = \alpha_m + \tau_t + \delta \text{Log Population}_{mt} + \sum_{i=1}^{8} \beta_i 1[N_{mt} = i] + u_{mt} \tag{19}$$

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	% Obs. Imputed
BLO	486	2.22	0.02
CJNG	565	1.99	0.02
CABT	309	1.80	0.01
CJ	261	2.34	0.01
CDS	741	2.41	0.02
CDG	585	2.58	0.02
FM	534	2.18	0.02
LZ	696	2.10	0.02
Any Imputation	2657	1.96	0.08

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 19. 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   Proofs for Econometric Section

This section derives Equation 14 and the estimation equation 18.

$$E[Y_{mt}|Q_m(\tilde{Z}_{mt}, W_{mt}) = q, W_{mt}, N_{mt}(q) = n]$$

$$= E[Y_{mt}(n)|Q_m(\tilde{Z}_{mt}, W_{mt}) = q, W_{mt}, N_{mt}(q) = n]$$

$$= \alpha_n + \psi(W_{mt}) + E[U_{mt}(n)|Q_m(\tilde{Z}_{mt}, W_{mt}) = q, W_{mt}, N_{mt}(q) = n]$$

$$= \alpha_n + \psi(W_{mt}) + E[U_{mt}(n)|W_{mt}, N_{mt}(q) = n] \qquad \text{(Assumptions SS1 FS3)}$$

$$= \alpha_n + \psi(W_{mt}) + E[U_{mt}(n)|N_{mt}(q) = n] \qquad \text{(Assumption SS2)}$$

$$= \alpha_n + \psi(W_{mt}) + \sum_{d \in \mathcal{N}(n)} E[U_{mt}(n)|D_{mt}(q) = d, N_{mt}(q) = n]P[D_{mt}(q) = d|N_{mt}(q) = n]$$

$$= \alpha_n + \psi(W_{mt}) + \sum_{d \in \mathcal{N}(n)} E[U_{mt}(n)|D_{mt}(q) = d]P[D_{mt}(q) = d|N_{mt}(q) = n]$$

To get the final estimating equation, start with observed outcomes and keep the α_n, ψ notation

$$E[Y_{mt}|Q_m(\tilde{Z}_{mt}, W_{mt}) = q, W_{mt}]$$

$$= Y_{mt}(0) + \sum_{n=1}^{G} E[Y_{mt}(n) - Y_{mt}(0)|Q_m(\tilde{Z}_{mt}, W_{mt}) = q, W_{mt}]$$

$$= \alpha_0 + \psi(W_{mt}) + E[U_{mt}(0)|Q_m(\tilde{Z}_{mt}, W_{mt}) = q, W_{mt}]$$

$$+ \sum_{n=1}^{G} P[N_{mt}(q) = n|Q_m(\tilde{Z}_{mt}, W_{mt}) = q, W_{mt}]\{\alpha_n - \alpha_0\}$$

$$+ P[N_{mt}(q) = n|Q_m(\tilde{Z}_{mt}, W_{mt}) = q, W_{mt}]E[U_{mt}(n) - U_{mt}(0)|Q_m(\tilde{Z}_{mt}, W_{mt}) = q, W_{mt}, N_{mt}(q) = n]$$

$$\text{(Iterate Expectations)}$$

$$= \alpha_0 + \psi(W_{mt}) + \sum_{n=1}^{G} P[N_{mt}(q) = n|W_{mt}]\{\alpha_n - \alpha_0\}$$

$$+ P[N_{mt}(q) = n|W_{mt}]E[U_{mt}(n) - U_{mt}(0)|Q_m(\tilde{Z}_{mt}, W_{mt}) = q, W_{mt}, N_{mt}(q) = n]$$

$$\text{(Assumptions SS1 FS3 and } E[U_{mt}(0)] = 0)$$

$$= \alpha_0 + \psi(W_{mt}) + \sum_{n=1}^{G} P[N_{mt}(q) = n]\{\alpha_n - \alpha_0\}$$

$$+ P[N_{mt}(q) = n|W_{mt}] \sum_{d \in \mathcal{N}(n)} E[U_{mt}(n)|D_{mt}(q) = d]P[D_{mt}(q) = d|N_{mt}(q) = n, W_{mt}]$$

$$= \alpha_0 + \psi(W_{mt}) + \sum_{n=1}^{G} P[N_{mt}(q) = n|W_{mt}]\{\alpha_n - \alpha_0\}$$

$$+ P[N_{mt}(q) = n|W_{mt}] \sum_{d \in \mathcal{N}(n)} E[U_{mt}(n)|D_{mt}(q) = d]\frac{P[D_{mt}(q) = d \cap N_{mt}(q) = n|W_{mt}]}{P[N_{mt}(q) = n|W_{mt}]}$$

$$= \alpha_0 + \psi(W_{mt}) + \sum_{n=1}^{G} P[N_{mt}(q) = n|W_{mt}]\{\alpha_n - \alpha_0\}$$

$$+ P[N_{mt}(q) = n|W_{mt}] \sum_{d \in \mathcal{N}(n)} E[U_{mt}(n)|D_{mt}(q) = d]\frac{P[D_{mt}(q) = d|W_{mt}]}{P[N_{mt}(q) = n|W_{mt}]}$$

$$\text{(} D_{mt}(q) = d \text{ necessarily yields n)}$$

$$= \alpha_0 + \psi(W_{mt}) + \sum_{n=1}^{G} P[N_{mt}(q) = n|W_{mt}]\{\alpha_n - \alpha_0\}$$

$$+ \sum_{d \in \mathcal{N}(n)} E[U_{mt}(n)|D_{mt}(q) = d]P[D_{mt}(q) = d|W_{mt}]$$

After plugging in the functional forms for ψ, and imposing normality for the control function term (Tallis, 1961), we obtain the final estimating equation:

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

D   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 [31] 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 $\tilde{Z}_{mt}$[32] 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

---

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

[32] 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 $Z_{mt} = \{Z_{gmt}\}_{g=1}^{G}$.

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:

$$\pi_g(\tilde{Z}_{mt-1}, X_{mt}, D_{mt}, D_{mt-1}) = r_g(X_{mt}, D_{gmt}, D_{-gmt}) - \rho_g(\tilde{Z}_{gmt-1}) - (1 - D_{gmt-1})\delta + \epsilon_{gmt} \quad (20)$$

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.

D.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. This complicates traditional fixed effects analysis. Let ϕ_m represent a time invariant unobservable for municipality m's profitability. Consider the following linear parametrization of flow profits

$$\pi_g(\cdot) = \phi_m + \beta_1 X_{mt} - \beta_2 \log \left(\sum_{j \neq g} D_{gmt} + 1 \right) - \beta_3 \tilde{Z}_{gmt} - (1 - D_{gmt-1})\delta + \epsilon_{gmt} \qquad (21)$$

where $r_g(\cdot) = \phi_m + \beta_1 X_{mt} - \beta_2 \log \left(\sum_{j \neq g} D_{gmt} + 1 \right)$, $\rho_g(\cdot) = \beta_3 \tilde{Z}_{gmt}$. $\log \left[\sum_{j \neq g} D_{gmt} + 1 \right]$ 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:

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

$$\beta_3 \tilde{Z}_{gmt} - (1 - D_{gmt-1})\delta + \epsilon_{gmt}$$

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

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

will typically be a non-linear function in $\phi_m, \tilde{Z}_{mt-1}, X_{mt}, D_{mt-1}$. 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 $Z_{mt} = (\phi_m, \tilde{Z}_{mt-1}, X_{mt}, D_{mt-1})$. Furthermore, the original linear time-invariant effects, ϕ_m, also affect this non-linear function. As a result, in a private information entry game,

additive time-invariant differences in flow profits still result in non-linear functions, due to the expectation term.