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

Ignacio Rodriguez Hurtado*

For latest version click here.

October 20, 2023

Abstract

Many places in the world have a substantial number of criminal groups. The total number of criminal groups is likely a key determinant of crime in these locations. I study the impact of the number of Mexican Drug Trafficking Organizations (DTOs) on homicides and school dropout in Mexican municipalities between 2006 and 2018. 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 selection model delivers interpretable treatment effects for each possible number of DTOs, while conventional two-stage least squares estimators do not. The selection model can be used to examine market structure effects for other licit industries as well. I find that as DTOs increase, there is a significant increase in homicides. Having 8 DTOs, the largest amount observed in my data, causes a 16-fold increase in homicides relative to the overall mean. Moreover, the relationship is increasing and convex, contrary to popular models of crime. Additionally, I find increased DTOs also increase local school dropout in middle school. These effects are concentrated amongst older and male students. I find no effects for younger or female students. The gendered and age-differentiated results suggest a recruitment channel from the DTOs, who rely on adolescent males for their workforce. Consistent with this mechanism, I find that as DTOs increase in number, the number of adolescent crimes with middle-school dropouts as perpetrators increase. Moreover, homicides with adolescent victims increase as well. This paper is the first to show the non-linear impacts of the number of criminal groups on crime, and how these non-linear impacts spill over into local schooling decisions.

JEL Codes: C26, E26, K42, O15

*Ignacio Rodriguez Hurtado, Duke University Department of Economics, ignacio.rh@duke.edu, ignacio-rh.github.io. I am grateful for the continued guidance and advice from Erica Field, Duncan Thomas, Robert Garlick and Adam Rosen in creating this paper. I also thank the Duke University Economics Department for their support and for the opportunity and privilege of completing a PhD degree. I thank Fernanda Sobrino for generously sharing her data. Last, I am grateful to Matthew Pecenco, Micaela Sviatschi, Levin Zhu, John Barry, Greg Burke, Felipe Montaño Campos, Hema Shah and Andres Santos for helpful comments, which greatly improved this draft. All errors are my own.

1 Introduction

Many places in the world have a large number of criminal organizations.[1] In these locations the number of criminal organizations is likely a key parameter determining violence and local outcomes. This study's setting, Mexico, serves as a prime illustration. Mexico has several large Drug Trafficking Organizations (DTOs) competing amongst each other.[2] Many qualitative accounts argue the growth in DTOs' number and presence is a major reason for increased violence and crime (Medel and Thoumi, 2014). 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). Naturally, the increase in violence and crime burdens surrounding communities, hindering economic growth and human development.

In this paper I examine the effects of **the number of DTOs in a Mexican municipality on two main outcomes: school dropout and homicides.** First, I examine school dropout to document if increased DTO presence affects schooling choices. The direction of the effect is unclear beforehand. On the one hand, DTOs may increase school dropout. The violence they cause may disrupt schooling activities and hinder travel to school, a *disruptive* channel. DTOs may also affect dropout through a *recruitment* channel, as these criminal organizations rely on adolescent males for their workforce (Carvalho and Soares, 2016). On the other hand, DTOs may also cause school dropout to decrease. For example, households may consider schools to be safer relative to outside employment and encourage children to stay longer in school. The overall direction of the effect is therefore ambiguous beforehand.

Furthermore, if DTOs increase dropout, then teasing out the channel is key in order to make accurate policy recommendations. If dropout increases due to disruption, then

[1] This paper's data tracks more than 60 groups in Mexico, while Colombia has more than 25 criminal groups (Colombiano, 2023). For more local examples, Cape Town South Africa counts with more than 90 gangs (Dziewanski, 2020), and Chicago has more than 50 (NBC, 2020).

[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.

increasing policing resources around schools or ensuring safe travel corridors for students are reasonable policies. If instead dropout is driven by recruitment, then policies reducing the opportunity cost of schooling, like conditional cash transfers, are more promising. Under this scenario, investing in educational programs not only raises human capital, but would also reduce DTO recruitment, their capabilities and ultimately violence. Indeed, a recent article in *Science* argues reducing DTO recruitment would be more effective than traditional policing to reduce violence in Mexico (Prieto-Curiel et al., 2023).

Apart from schooling, I examine homicides as a primary outcome to document how the number of DTOs shapes violence and crime. Empirically estimating the relationship between the number of DTOs and homicides is important for policy and theoretical reasons. From a policy perspective, the relationship of violence to number of DTOs can be informative to better target scarce policing resources. For example, a realistic policy may be to deploy policing resources to discourage an additional DTO from entering or remove an existing incumbent.[3] The shape of the violence-to-DTO relation can determine this policy effectiveness. For example, if the violence-to-DTO relationship is convex, preventing DTO entry in locations with many DTOs would lead to disproportionate reductions in homicides. Note that to answer this question, a simple binary comparison of any DTO versus none is uninformative. Instead, we require the full relationship, or dose response function, of violence on the number of DTOs. Furthermore, we need to flexibly estimate this function to allow for non-linearities, since these non-linearities are precisely what can influence policy effectiveness.

Estimating the violence-to-DTO relationship is also of theoretical interest. While many theoretical models predict an increasing, concave relationship between violence and number of criminal organizations (Hirshleifer, 1995), other models can yield a decreasing or non-concave relationship (Gama and Rietzke, 2019). Given the ambiguity of these models, empirically estimating the violence-to-DTO relationship can help adjudicate between existing theoretical models of crime and violence.

However, estimating the effects of the number of DTOs is challenging for two reasons.

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

First, there is a classic selection problem, since DTOs endogenously select where to operate. Indeed, I find DTOs tend to select into richer areas in Mexico. To account for endogenous selection, existing papers examining the impacts of criminal organizations adopt difference in differences approaches (Monteiro et al., 2022; Biderman et al., 2019; Bruhn, 2021). While plausible in other settings, a difference in differences analysis places important restrictions on DTOs' selection patterns, which are unrealistic in my paper's context. A chief concern with a difference in differences analysis are forward looking entry decisions on the part of DTOs. As emphasized in Ghanem et al. (2023), if DTOs have appropriate expectations on how entry impacts their future profits and operations, then a parallel trends assumption would not hold, since DTOs use their expectations to effectively select on the trend of potential outcomes. Given DTOs scale and longevity, I argue DTOs in Mexico likely have a good understanding of the sector and are able to make appropriate assessments about future profitability depending on their entry decisions.

Instead of difference in differences, I leverage instrumental variables in a selection model that accommodates forward looking DTO entry decisions. 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. The logic behind these instruments is simple. The intuition is that DTOs, say Sinaloa Cartel, observes both its own proximity to a target municipality, and also observe its rivals' proximity, say Los Zetas. DTOs then decide to enter based on their own and rivals proximity. Indeed, first stage estimates indicate DTOs behave as strategic rivals: DTOs are more likely to enter when the DTO itself is closer, but are less likely to enter when their competitors are closer. Next, 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. Furthermore, I argue the exclusion restriction holds for these instruments given the limited scope for spillovers across municipalities with and without DTOs.[4]

[4] I exhaustively discuss several possible violations to the exclusion restriction in Section 4.1, page 25, and

Nonetheless, there is an important methodological challenge when using these instruments. As mentioned previously, a binary comparison of any amount of DTOs relative to none would not detect policy-relevant nonlinearities. Instead, to flexibly estimate these nonlinearities, the most appropriate solution is to estimate a separate treatment effect for each possible number of DTO present. However, conventional estimators struggle with multiple treatments. 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 and Sigstad, 2022; Mountjoy, 2022; Kirkeboen et al., 2016). To clarify the intuition behind the problem, 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.[5]

In place of conventional TSLS or difference in differences estimators, I implement a novel selection model to estimate average treatment effects for different "doses" of active DTOs in a Mexican municipality.[6] 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. The role of the instruments is to shift the entry probabilities without directly affecting second stage outcomes, allowing me to estimate the effects of DTO entry on outcomes. For the first stage, **I borrow from the industrial organization literature to model DTO**

present evidence and arguments explaining why these violations are unlikely in my paper's context.

[5]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.

[6]Other treatment effect parameters, like the average treatment effect on the treated (ATT) can also be computed with my model, following Lee and Salanié (2018). I focus instead on average treatment effects (ATEs) of DTO presence for simplicity, and since the ATEs are informative for untreated units, of which there are many. Roughly half of Mexican municipalities never received a DTO during my sample years.

entry as a dynamic firm-entry game, therefore allowing for forward facing entry decisions. Additionally, I lay out which assumptions facilitate this dynamic interpretation and argue they are plausible in my setting. After dynamic entry decisions are made, the DTOs' individual entry decisions then jointly determine the treatment dose, the number of active DTOs in a municipality. The main benefits of the model are twofold. First, my model allows DTOs to be forward looking and overcomes an important challenge to difference in differences designs. Second, it delivers interpretable causal estimates for each dose of treatment, which is required to study the policy relevant non-linearities mentioned earlier. As a result, I trace out the entire dose response function of homicides and school dropout as a function of active DTOs.

I find an increasing and convex relationship between the number of DTOs and homicides. To get a sense of the convex relation, relative to no DTO, the effect of 4 DTOs is 6 times larger than the effect of 1 DTO. In contrast, the effect of 7 DTOs is 22 times larger than the effect of 1 DTO. These results suggest that reducing DTO presence in areas with a large count of DTOs provides disproportionate reductions in homicides, a policy relevant finding. These estimated non-linearities vindicate the selection model, since a simpler linear or binary framework necessarily misses this relationship. Moreover, I find positive and significant effect on homicides for a single, monopolist DTO relative to no DTO. This is an interesting case to study, since monopolist DTOs by definition do not have any existing competitors.

Furthermore, the convex effects disagree with popular models of crime which predict an increasing *concave* relationship (Hirshleifer, 1995). While it is beyond the scope of this paper to comprehensively investigate the discrepancy between my estimates and these theoretical predictions, I offer some speculative reasons. To do so, it is important to understand the possible channels through which the number of DTOs affects homicides. I focus on two. First, there is a competitive effect. As DTOs increase, we expect DTOs to compete more. This is the main channel considered in typical theoretical models. Second, I speculate there is a criminal capital effect with DTO entry, which is typically not present in theoretical models. DTOs in Mexico have extensive criminal know-how and connections with many illicit markets. I argue these connections and knowhow are a form of criminal capital which impacts local crime, and helps explain the discrepancy between my estimates and some models of crime. While I cannot examine these channels separately, my results do point out that a competition mechanism alone cannot explain all the findings. Recall I find an increase in homicides even with a monopolist DTO, when there is clearly no competition between DTOs.

For my other primary outcome, school dropout, I find statistically significant positive effects, and these effects are linear. An additional DTO increases ninth grade dropout rates by 1-2 percentage points. However, the effects are entirely concentrated amongst ninth grade male students. I fail to find an effect for females or younger grades. This pattern of results suggest disruption is not primary driver for the dropout effects, since we would expect at least some effects for younger grades or female students. Indeed, I conduct heterogeneity analysis by previous year's homicides and find that locations with more violence last year do not have more dropout. As another possible mechanism, I examine migration directly with individualized student data for a subset of years and find no migratory effect.

Instead I find direct evidence the dropout effects are driven by the recruitment channel. I complement the main analysis with adolescent crimes as an outcome. I find overall adolescent crimes increase with more DTOs. Even more specifically than that, I also find crimes with middle school dropouts as perpetrators increase. Middle school dropouts are the primary dropout "compliers" affected by increased DTO presence, since I find dropout effects for ninth grade, the last grade in Mexican middle school. In addition to adolescent crimes, I find homicides with male, adolescent victims aged 15-18 increase with more DTOs. These are precisely the age ranges we expect the ninth grade students (typically aged 14) to be at risk for homicide, given previously documented criminal career trajectories (Carvalho and Soares, 2016). The constellation of results therefore suggest ninth grade dropouts are in fact entering criminal activity. As a result, increased DTOs affect school dropout mainly through a recruitment channel, not through disruption. These results highlight the promise of policies changing the opportunity cost of schooling to not only increase human capital achievement, but also reduce violence by reducing DTO recruitment, as highlighted in (Prieto-Curiel et al., 2023).

How do my preferred estimates compare to more conventional models? Difference in differences estimates using OLS systemically underestimate the effects of DTO entrance on homicides, with some estimates being 5 times smaller than my main model's estimate. OLS estimates are also zero for school dropout. I argue DTOs' selection patterns explain this discrepancy. I show descriptively that DTOs select into richer areas of Mexico. The areas DTOs enter in have more police officers and higher education achievement in general. These areas DTOs select into are precisely the areas with most resources to counteract DTOs' negative effects, explaining the discrepancy between OLS and my estimates. Comparing to TSLS, a linear TSLS model also estimates a positive effect of DTOs on homicides, but fails to find an effect for dropout. A saturated version of TSLS which "dummies out" the number of DTOs is largely uninformative, with wide standard errors. However, this version finds a negative effect of dropout for a single DTO. Given the complications with TSLS which I explain in Section 5, it is unsurprising TSLS differs from my main results.

This paper contributes to several strands of literature. First, I contribute to the literature examining the impact of crime on educational choices. To my knowledge, this is the **first paper studying the effects of the number of criminal organizations on educational outcomes. Furthermore, I find evidence specifically in favor of a recruitment channel**. In this aspect, the closest paper to mine is Sviatschi (2022), who also finds a decrease in schooling outcomes in response to an expansion in criminal activity. However, Sviatschi (2022) uses variation in the international price of coca, a labor intensive crop well suited to child cultivation, to show Peruvian school children are more likely to leave school and engage in future crime. I find a similar effect and mechanism through an entirely different source of variation, increased DTO presence. My results therefore build on Sviatschi (2022) by showing the number of criminal organizations can have similar effects to natural resource shocks. Furthermore, a wider literature has examined the effect of violence and conflict on educational attainment in Mexico and other developing countries, mainly through a disruptive channel (Brown and Velásquez, 2017; León, 2012; Lopez Cruz and Pairolero, 2019). Other papers corroborate these findings 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). I complement these papers by showing that increased criminal organizations affect schooling through a recruitment channel.

As a second contribution, **I document crime and violence increases convexly with the number of criminal groups.** Existing papers either do not examine non-linearities between crime and number of criminal groups (Bruhn, 2021; Monteiro et al., 2022; Alcocer, 2023), or rely on event study designs with restrictive assumptions on criminal entry (Sobrino, 2020). In contrast, I leverage an instrumental variables to flexibly estimate potential non linearities. The convex relation is important for policy, since it indicates targeting enforcement in areas with a large number of groups results in disproportionate reductions in violence. The convex relation is also important for theoretical reasons, since the empirical estimates can help adjudicate between competing theoretical models. More broadly, I contribute to the extensive literature studying the drivers of crime in the developing world. Olken and Barron (2009) is a related paper and studies the industrial organization of corruption and bribes in Indonesia. In another strand of literature, 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. I complement these papers by showing the number of criminal groups is an important driver of crime, and by documenting the convex shape between violence and number of DTOs.

A third empirical contribution of my paper is to empirically document that criminal groups are strategic rivals. An important question concerning criminal organizations and DTOs is the nature of their competition. Are they strategic rivals or strategic complements? While it might be tempting to assume they are rivals, this needn't be the case, they may be complements. For example, criminal organizations have a vested interest to degrade government institutions and capacities. As such, it could be the case criminal groups are complementary to each other, as they can exploit agglomeration economies to further weaken the state. However, my first stage estimates indicate that DTOs in fact behave like strategic rivals. I estimate DTOs' probability of entry as a function of competitors' distance and incumbency in a municipality. I find DTOs are more likely to enter if their competitors are far away or a location has fewer existing competitors. To my knowledge, this is the first empirical evidence showing criminal organizations behave like strategic rivals, a critical behvaior to understand when examining criminal competition. There is a related literature examining how criminal organizations interact with the state, but these papers do not empirically investigate how criminal organizations interact with one another (Trejo and Ley, 2018; Acemoglu et al., 2020; Daniele and Dipoppa, 2017; Alesina et al., 2019).

As a final contribution, the econometric model in this paper combines firm entry models from industrial organization (IO) with the modern treatment effect literature to examine the causal impacts of market structure on downstream outcomes. My innovation is to estimate effects of market structure, measured as the number of groups in my application, within a heterogeneous treatment effect framework (Heckman et al., 2006a; Lee and Salanié, 2018; Mogstad et al., 2018), in contrast to existing empirical work.[7] Of course, there exists a rich literature in industrial organization examining the effects of market structure on downstream outcomes, which use different empirical strategies. However, none of these strategies allow for heterogeneous treatment effects. A few papers use instruments combined with TSLS (Olivares and Cachón, 2009; Hackl et al., 2014; Jeanjean and Houngbonon, 2017). These papers ultimately assume a constant treatment effect. Another popular empirical strategy is a structural approach, where the effects of market structure on outcomes are modeled

[7] In essence, my paper is an application of Lee and Salanié (2018), whcih allows for heterogeneous effects, but tailored for dynamic group entry decisions, in line with the industrial organization literature.

according to a structure.[8] These papers then study how entry can endogenously affect outcomes, for instance prices. (Ryan, 2012; Collard-Wexler, 2013; Holmes, 2011; Ciliberto et al., 2021). This approach tends to be computationally challenging and requires correctly modeling the effects of market structure on downstream outcomes.[9] Last a few studies have employed a similar approach to my paper and use a selection correction to account for endogenous market structure (Mazzeo, 2002; Jeanjean and Houngbonon, 2017; Hackl et al., 2014). All of these papers are for static models only, and firms cannot be forward looking. These papers therefore do not address a key limitation of difference in differences designs, since firms are not allowed to be forward looking. Overall, **I provide a tractable way to examine the effects of market structure on downstream outcomes, without having to estimate a full structural model, allowing for dynamic agents, and within a heterogeneous treatment effect framework.**[10]

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.

[8] A typical example is modeling the effect of market structure of prices using a linear or logit demand system.

[9] Of course, the benefit of the structural approach is the evaluation of counterfactual policies, which I do not pursue in this paper.

[10] 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), and results in point estimates instead of parameter bounds.

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).

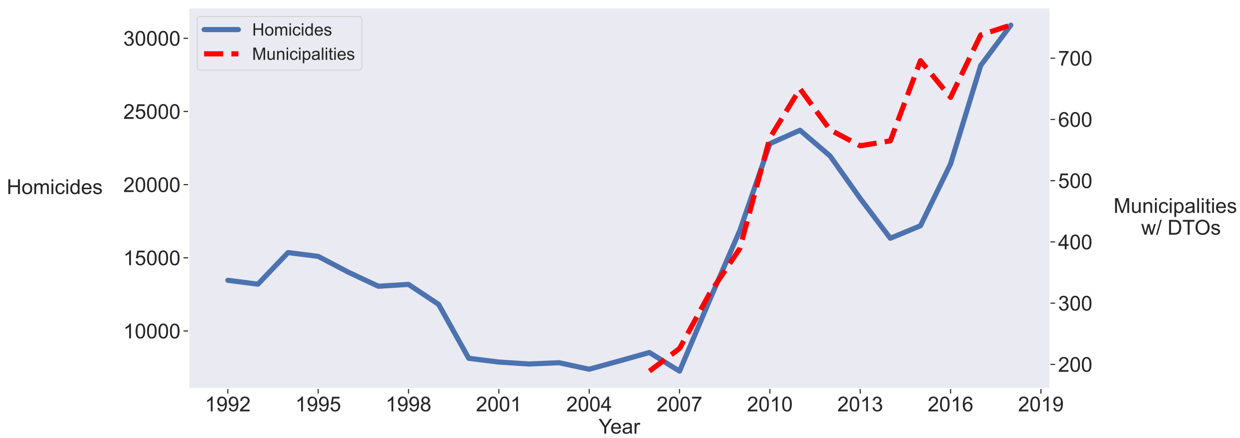
Several studies have by now documented the crackdown during the Calderón administration led to a sustained increase in homicides and drug-related violence. Dell (2015) documents that municipalities narrowly won by PAN experienced increases in violence. The crackdown likely spurred further violence by 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.

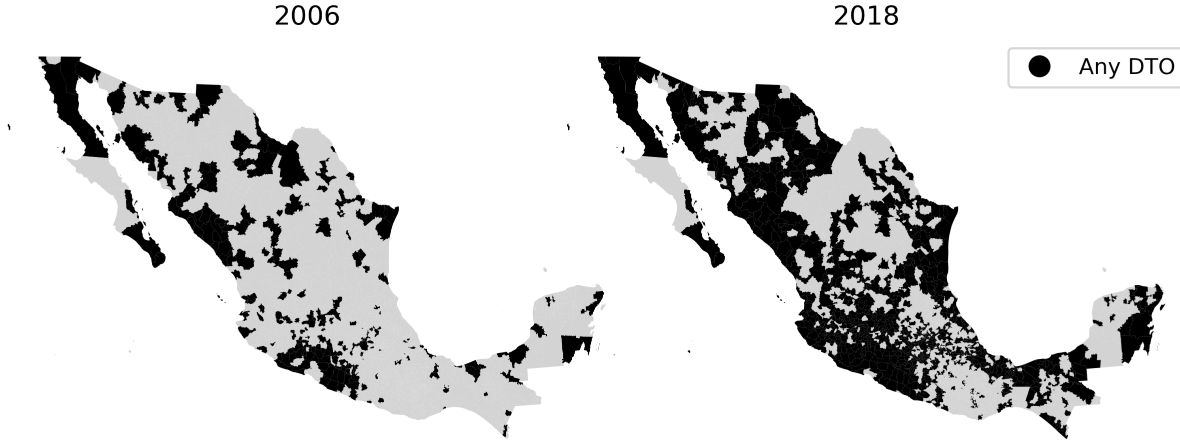
This paper focuses on the eight largest DTOs in Mexico.[11] The selection of these DTOs is based on their mentions in the DEA's National Drug Threat Assessment reports from 2008 to 2018. Notably, these DTOs represent the largest and most significant players in the Mexican drug trade, regularly involved in trafficking substantial quantities of drugs into the United States. Moreover, their operations span vast territories across Mexico. At the start of my study period, 2006, not all of these 8 large DTOs existed and three of them were created after 2006. Table A1 lists the years when the DTOs are observed in my data. The three DTOs splintered off from preexisting organizations.[12]

This fragmentation is a 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

[11] 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*).

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

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.

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. The data in this study bears this out: in addition to the 8 large DTOs, the data tracks 64 smaller DTOs. Among all observations during the study period with any DTO presence (encompassing the 8 large and 64 smaller DTOs), the 64 smaller groups account for roughly 32% of these instances, while the 8 larger DTOs make up the remaining 68%. Last, these smaller DTOs predominantly operate within limited regions and demonstrate minimal expansion into new areas. Consequently, their analysis does not align with the empirical strategy of this paper.

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

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.[13]

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[14], 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 that do not describe presence but instead mention the DTO for an unrelated reason. Therefore a news article is classified as describing a particular DTO's presence in a municipality if it contains a sentence with both the DTO-municipality pair, and the CNN predicts the sentence is describing the presence of the DTO. The version of the dataset I use is at the municipality-year level with a dummy variable whenever any news article claims DTO is present in a municipality-year according to these criteria. The dataset tracks presence for 72 DTOs between 1990 and 2020. I focus on the 8 largest DTOs during 2006 and 2020.

Furthermore for the rest of analysis I use an imputed measure of presence. I 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

[13] 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.

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

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.[15] The school year in Mexico starts in August and finishes in July

[15] 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

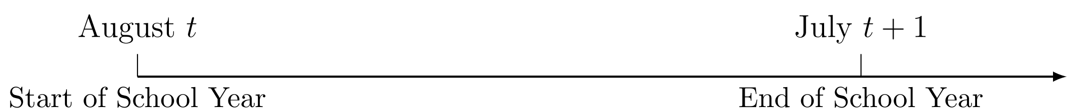
of the following year. To calculate intra-curricular dropout rates for grade 9, I first select schools that reported both at the start and end of the same academic year. I then aggregate enrollment counts for all schools in each municipality. As a result, if a student changes school within the same municipality, they are still captured in my data as being in school.

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

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

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

Figure 3: School Calendar



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

3.3 Other data

I gather other data to supplement the analysis.

Agricultural Revenue Data: I gather municipality specific records of agricultural production and revenue, in nominal Mexican pesos. These are estimates published by Mexico's SIAP office (Servicio de Información Agroalimentaria y Pesquera). SIAP estimates crop

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.

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

Table 1: Descriptive Statistics

	Panel A: Means By Number of Groups				
	0	1	2	3+	Overall
Covariates					
Population (2005)	11,366	23,611	38,742	109,233	44,725
Poverty Rate (2005)	0.74	0.66	0.62	0.55	0.66
Literacy Rate (2005)	0.80	0.84	0.86	0.88	0.83
Elect. Users	3,396	7,144	12,813	36,331	14,373
kWH (millions)	14.60	39.22	71.54	232.66	86.64
kWH per Capita	3,796	5,748	8,619	10,143	6,418
Ag. Revenue	54.01	127.77	170.79	376.38	169.19
Total Police	27	41	73	168	75
Outcomes					
Homicide Rate (per 100,000)	11.80	13.94	17.65	23.01	15.95
9th Grade Enrollment	225.69	460.97	760.59	2,060.12	841.37
9th Grade Dropout Rate	0.03	0.03	0.03	0.03	0.03
9th Grade Dropout Rate (Girls)	0.03	0.03	0.03	0.03	0.03
9th Grade Dropout Rate (Boys)	0.03	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
$\geq$ 1 DTOs	0.14	0.27	0.35	0.39	0.30
$\geq$ 2 DTOs	0.07	0.18	0.25	0.27	0.20
$\geq$ 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.

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.

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.

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

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

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 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, 2022) 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 country. If DTOs are able to successfully manipulate their entire network to systematically target more profitable regions over time, then the distance instrument would fail. In this case, places that are close to DTOs are just fulfilling the DTO's long term plans and are therefore selected either along observable or unobservable margins. In contrast, if DTOs are unable to carry out a cross-regional plan, I argue the distances will largely be a function of the initial DTO placement and unobserved cross-sectional differences in profitability for DTO entry. Therefore, conditional on municipality fixed effects, the distances are plausible independent of potential outcomes.

There are two main reasons why DTOs are unable to carry out multi-year cross regional-plans. First, DTOs are large and decentralized operations. Due to the 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.[16]

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 municipalities 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 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 8.6 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

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

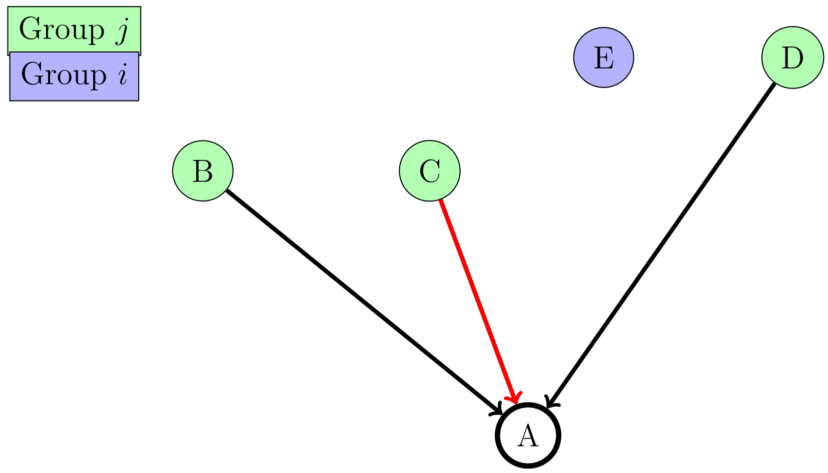
et al. (2023) find a delayed response of many years in migration due to an opium demand shock in Mexico for the same period studied here. Second, I extensively investigate if my main results are driven by migration in Section 8.6.1. I find no evidence that the large effects on dropout are due to to out migration.

Last, we may be concerned that DTOs displace local criminals into surrounding areas. As a result, places nearby a DTO will additionally see an increase in crime due to the displaced criminals. While it is difficult to examine this possibility quantitaviley, qualitive work suggests this is not an important concern. DTOs tend to either co-opt local criminals or engage in a turf battle with local criminals; these qualitive accounts do not highlight the displacement of local criminals (Correa-Cabrera, 2021).

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 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

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.

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.[17] 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 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

[17]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.

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	2.36***	4.41***	2.23***	4.82***	1.04***
	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)
	With Year, Muni. FEs				
Average Distance	0.01	-0.01	-0.01	0.21***	0.04
	(0.01)	(0.02)	(0.02)	(0.04)	(0.05)
	Individual Distances With Year, Muni. FEs				
All Distances Joint F-Stat	7.31	7.97	6.64	45.38***	8.27
P-Value	[0.50]	[0.44]	[0.58]	[0.00]	[0.41]
	Individual Distances With Controls, Year, Muni. FEs				
All Distances Joint F-Stat	8.83	7.41	6.33	42.19***	6.87
P-Value	[0.36]	[0.49]	[0.61]	[0.00]	[0.55]
Outcome Mean	14.69	88291.11	6.04	199202.65	75.03
N	18779	18779	18779	18779	9424

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

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 8.6.6 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:

$$N_{mt} = \alpha_m^1 + \tau_t^1 + W_{mt}\delta^1 + \sum_{g=1}^{G} \beta_g^1 Z_{gmt-1} + \beta_g^2 Z_{gmt-1}^2 + \beta_g^3 Z_{gmt-1}^3 + u_{mt}^1 \tag{4}$$

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

Where Y_{mt} is either the count of homicides or 9th grade dropout rate, W_{mt} is as in Equations 2 and 3, with the addition of agricultural revenue. The controls include the lagged presence indicators for each of the 8 DTOs, the total active DTO count for previous year, the IHS of agricultural revenue and the yearly log of population counts as imputed by the census. 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. For the instrument, I control for a cubic polynomial of group distances, to mimick the main model used later in the paper. I find the cubic provides a better fit for the first stage later on.

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.[18] 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 9.93 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 -0.0085.

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, since the effective F-stat is 11.46. Indeed a simpler model with simple linear distances, and no cubic controls, yields an even larger F-stat of 29.88. However I include the cubic controls to harmonize the TSLS results with the main model results later on.

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

[18] 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
# DTOs	9.93*	-0.0085
	(5.99)	(0.0065)
Outcome Mean (levels)	8.03	0.03
# Instruments	24	24
N	30619	30619
Effective F-stat	11.46	11.46
Sargan-Hansen Test	44.40	61.02
Sargan-Hansen Test p-val	0.00	0.00
Controls	Yes	Yes

Notes: Unit of observation is a municipality-year. "# DTOs" 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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.[19] Moreover, even with the strong monotonicity assumption, the results in Angrist and Imbens (1995) would imply that the coefficient in Table 3 is weighted average across different "doses" of DTO group count (N_{mt}) and different complier groups. The TSLS estimate may for instance be placing more weight on the transition between 1 DTO versus

[19]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}

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.

However 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, 2022; 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 $0...G$.[20] For an outcome Y_{mt}, I denote the vector of potential outcomes as

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

[20]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

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

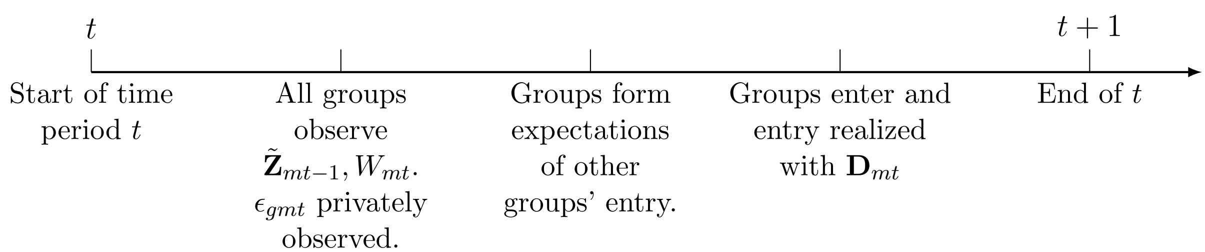
6.2 First-Stage Model

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, 2007a; Aradillas-López, 2020). Importantly, groups are allowed to be forward looking and make decisions based on dynamic considerations.

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 $(\tilde{\mathbf{Z}}_{mt-1}, W_{mt})$ which is common knowledge to all groups. In my application $\tilde{\mathbf{Z}}_{mt-1} = \{\tilde{Z}_{gmt-1}\}_{g=1}^{G}$ are the group-specific distances for the previous year, which is common knowledge to all groups. I bold $\tilde{\mathbf{Z}}$ to emphasize it's a vector across the G 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 $\mathbf{D}_{mt-1} = \{D_{gmt-1}\}_{g=1}^{G}$, 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 $W_{mt} = (R_m, L_t, X_{mt}, \mathbf{D}_{mt-1})$. Again, I bold $\mathbf{D}_{mt-1}$ since it is a vector across the G groups.

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 $\tilde{\mathbf{Z}}_{mt-1}, W_{mt}, \epsilon_{gmt}$, 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 $\epsilon_{gmt} \mid W_{mt}$ *is i.i.d. across time, groups and municipalities.*

FS2 Groups make entry decisions independently across municipalities.

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

Once groups observe the covariates $\tilde{\mathbf{Z}}_{mt-1}, W_{mt}$, 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 (Threshold Crossing Representation)**.**
Each group g's entry decision, D_{gmt} *can be represented as*

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

Where v_{gm} *is a possibly nonlinear function*

I make this high level assumption in the main text to stress what my model requires, a threshold crossing representation in the first stage, alongside Assumption 1. Any entry game that micro-founds this reduced form entry decision is compatible with my estimation. But what kind of entry games are consistent with this representation? A dynamic Markov game with private information shocks is one such representation, and it provides sufficient conditions for Assumption 2. Additional assumptions are needed for this game, and I clarify the structure in Appendix Section D. The key assumption of this section is a stationarity assumption. The covariates $(\tilde{\mathbf{Z}}_{mt-1}, W_{mt})$ evolve according to a stable pattern over time, one which the DTOs know and can anticipate. While this might seem unrealistic for my time period, which includes substantial DTO expansion, I check the sensitivity of my results to this stationarity assumption in Section 8.6.5. I find the results are robust to allowing for non-stationarity.

Furthermore, notice I allow the decision rule to explicitly vary by municipality m, with v_{gm}. This is to explicitly allow certain forms of multiple equilibria. Entry games may have multiple equilibria, where the same set of observables $(\tilde{\mathbf{Z}}_{mt-1}, W_{mt})$ lead to different DTO entry decisions depending on the equilibrium played. With Assumption 2, I assume there may be multiple equilibria across municipalities m. However, within the same municipality m, I exclude different equilibria played over time. That is, DTOs play the same equilibrium across time within the same municipality. This seems like a reasonable restriction. Furthermore, it is compatible with the Markov game outlined in Appendix Section D.

Nonetheless, some games are ruled out by Assumption 2. Complete information games like in (Bresnahan and Reiss, 1991; Ciliberto and Tamer, 2009) are ruled out. This is because in complete information games, groups decide on entry not only on their own first stage error ϵ_{gmt} but also on their rivals' errors ϵ_{-gmt}. As mentioned earlier, given the clandestine nature of criminal groups, they likely do not have high quality on their rivals' operations and profits, so excluding complete information games seems reasonable. Similarly, games where the ϵ_{gmt} shocks are correlated amongst the G groups are excluded, Assumption FS1 needs to be satisfied. If these correlations are present, then the ϵ_{gmt} will generally enter the v_{gm} and we won't have a threshold crossing rule. However, notice that I include municipality dummies in W_{mt} and can account for municipality specific unobservable variables in estimation. Controlling for these municipality specific unobservables effectively allows for a degree of correlation in the errors, one that I can control for. I explain how I accomplish this in Section 7.1.

6.2.1 Re-expressing Entry Decisions

It is convenient to re-express the entry decision in Assumption 2 as follows:

$$\begin{aligned} D_{gmt} &= 1[v_{gm}(\tilde{\mathbf{Z}}_{mt-1}, W_{mt}) > \epsilon_{gmt}] \\ &= 1[F_{\epsilon|W}(v_{gm}(\tilde{\mathbf{Z}}_{mt-1}, W_{mt})|W_{mt}) > F_{\epsilon|W}(\epsilon_{gmt}|W_{mt})] \\ &= 1[Q_{gm}(\tilde{\mathbf{Z}}_{mt-1}, W_{mt}) > e_{gmt}] \end{aligned} \tag{9}$$

Where F_ϵ denotes the CDF for ϵ, $Q_{gm}(\tilde{\mathbf{Z}}_{mt-1}, W_{mt}) := F_{\epsilon|W}(v_{gm}(\tilde{\mathbf{Z}}_{mt-1}, W_{mt})$ and $e_{gmt} := F_{\epsilon|W}(\epsilon_{gmt}|W_{mt})]$. This transformation is innocuous and it eases the subsequent exposition, as in (Heckman et al., 2006b). First, notice that $Q_{gm}(\tilde{\mathbf{Z}}_{mt-1}, W_{mt}) = P[D_{gmt} = 1 \mid \mathbf{Z}_{mt-1}, W_{mt}]$, hence we can directly interpret Q_{gm} as the probability g enters.[21] Furthermore, e_{gmt} is uniformly distributed on [0,1] conditional on $(\mathbf{Z}_{mt-1}, W_{mt})$, and e_{gmt} still inherits independence w.r.t. $\mathbf{Z}_{mt-1}$ from Assumption FS3. The benefit of this transformation is we can characterize entry decisions with the probability Q_{gm} along the [0,1] interval.

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 $\tilde{\mathbf{Z}}_{mt-1}$ variables do not enter the potential outcomes equation. In other words, lagged group distances do not affect current outcomes directly. The current outcomes in this paper are homicides and school dropout. I presented arguments for this exclusion restriction in Section 4.1, page 25. 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 $U_{mt}(n) \perp \tilde{\mathbf{Z}}_{mt-1} \mid W_{mt}$

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

[21] $P[D_{gmt} = 1 \mid \mathbf{Z}_{mt-1}, W_{mt}] = P[v_{gm}(\tilde{\mathbf{Z}}_{mt-1}, W_{mt}) > \epsilon_{gmt} \mid \mathbf{Z}_{mt-1}, W_{mt}] = P[v_{gm}(\tilde{\mathbf{Z}}_{mt-1}, W_{mt}) > \epsilon_{gmt} \mid W_{mt}] = F_{\epsilon|W}(v_{gm}(\tilde{\mathbf{Z}}_{mt-1}, W_{mt})|W_{mt}) = Q_{gm}(\tilde{\mathbf{Z}}_{mt-1}, W_{mt})$. The first equality imposes Assumption 2. The second equality employs Assumption FS3. The third equality re-expresses the probability with the CDF of $\epsilon|W$. The last equality follows by the definition of Q.

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, $U(n)$. I presented evidence for this exogeneity in Table 2, Section 4.2. 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}. That is, the shape of this expectation is the same for all values of covariates W_{mt}.

It is worth discussing the implications of Assumption SS2 for $\mathbf{D}_{mt-1}$, which is the vector of past entry decisions and part of W_{mt}. Assumption SS2 is not saying $\mathbf{D}_{mt-1}$ is uncorrelated with $U_{mt}(n)$. Rather, its saying that once we control for the first stage errors $\mathbf{e}_{mt}$, $\mathbf{U}_{mt}(n)$ is mean independent of $\mathbf{D}_{mt-1}$. 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 $\mathbf{Q_m}(\tilde{\mathbf{Z}}_{mt-1}, W_{mt})$ the vector of entry probabilities for each group, $\mathbf{Q_m}(\tilde{\mathbf{Z}}_{mt}, W_{mt}) = \{Q_{1m}(\tilde{\mathbf{Z}}_{mt}, W_{mt}), Q_{2m}(\tilde{\mathbf{Z}}_{mt}, W_{mt})\}$. The mean regression of Y_{mt} on W_{mt} and the first stage entry probabilities, $\mathbf{Q_m}(\tilde{\mathbf{Z}}_{mt}, W_{mt})$ for the subset of observations with one group operating, $N_{mt} = 1$ is given by:

$$\begin{aligned}
&E[Y_{mt}|\mathbf{Q_m}(\tilde{\mathbf{Z}}_{mt}, W_{mt}) = q, W_{mt}, N_{mt} = 1] = \alpha_1 + \psi(W_{mt}) \\
&+ E[U(1)|q_1 > e_1, q_2 < e_2]P[q_1 > e_1, q_2 < e_2|\{q_1 > e_1, q_2 < e_2\} \cup \{q_1 < e_1, q_2 > e_2\}, W_{mt}] \\
&+ E[U(1)|q_1 < e_1, q_2 > e_2]P[q_1 < e_1, q_2 > e_2|\{q_1 > e_1, q_2 < e_2\} \cup \{q_1 < e_1, q_2 > e_2\}, W_{mt}]
\end{aligned} \tag{11}$$

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.[22] This representation exploits the exclusion restriction, Assumption 3, and the first stage model of entry.[23]

The probability terms are estimated using the first stage model. For instance, the term $P[q_1 < e_1, q_2 > e_2 | \{q_1 > e_1, q_2 < e_2\} \cup \{q_1 < e_1, q_2 > e_2\}]$ is just the conditional probability that group 1 enters but 2 does not, conditional on one of the two groups entering.[24] 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 $\mathbf{Q_m}(\tilde{\mathbf{Z}}_{mt}, W_{mt})$. **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 set

$$\mathcal{E}_g(d_g, q_g) := \begin{cases} [0, q_g) & \text{if } d_g = 1 \\ (q_g, 1] & \text{if } d_g = 0 \end{cases} \tag{12}$$

Recall entry occurs if $q_g > e_g$ and there is no entry if $q_g < e_g$. The set $\mathcal{E}_g(d_g, q_g)$ is the region where group g's first stage error e_g must lie given an entry decision d_g, and given g's probability of entry q_g. The set $\mathcal{E}_g(d_g, q_g)$ is tailored for a specific player. We can expand this concept and consider the region implied for an entire entry vector $\mathbf{d} = \{d_g\}_{g=1}^G$. Similarly

[22] Proofs are found in the appendix

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

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

let $\mathbf{q} = \{q_g\}_{g=1}^G$. The symbols $\mathbf{q}, \mathbf{d}$ are in bold to emphasize that they are a vector for all G groups. Define the following set:

$$\mathcal{E}_{\mathbf{d}}(\mathbf{d}, \mathbf{q}) := \bigcap_g \mathcal{E}_g(d_g, q_g) \tag{13}$$

$\mathcal{E}_{\mathbf{d}}(\mathbf{d}, \mathbf{q})$ defines the region where the vector of errors $\mathbf{e} = \{e_g\}_{g=1}^G$ must lie given entries and probabilities $\mathbf{d}, \mathbf{q}$. This set pertains to a single entry vector $\mathbf{d}$. I study the effects of group counts in my paper. Hence define a set for a particular group count n:

$$\mathcal{E}_n(n, \mathbf{q}) := \bigcup_{\mathbf{d} \in \mathcal{N}(n)} \mathcal{E}_{\mathbf{d}}(\mathbf{d}, \mathbf{q}) \tag{14}$$

Where the set $\mathcal{N}(n) := \{d \in \mathcal{D} | \sum d_g = n\}$ contains all the entry configurations $\mathbf{d}$ that yield a group count n. $\mathcal{E}_n(n, \mathbf{q})$ is simply the region where all first stage errors $\mathbf{e}$ can lie such that the final group count is n. For example, with two players, $\mathcal{E}_n(1, \mathbf{q}) = \{q_1 > e_1, q_2 < e_2\} \cup \{q_1 < e_1, q_2 > e_2\}$. The regions have simple representations due to the threshold crossing for entry in Assumption 2 and the independence between groups' first stage errors by Assumption FS1.

Now again consider the mean regression of Y_{mt} on W_{mt} and the first stage entry probabilities, $\mathbf{Q}_{\mathbf{m}}(\tilde{\mathbf{Z}}_{mt-1}, W_{mt})$, limited to observations with $N_{mt} = n$:

$$\begin{aligned} &E[Y_{mt} | \mathbf{Q}_{\mathbf{m}}(\tilde{\mathbf{Z}}_{mt-1}, W_{mt}) = \mathbf{q}, W_{mt}, N_{mt} = n] \\ &= \alpha_n + \psi(W_{mt}) + \sum_{d \in \mathcal{N}(n)} E[U_{mt}(n) | \mathbf{e} \in \mathcal{E}_{\mathbf{d}}(\mathbf{d}, \mathbf{q})] P[\mathbf{e} \in \mathcal{E}_{\mathbf{d}}(\mathbf{d}, \mathbf{q}) | \mathbf{e} \in \mathcal{E}_n(n, \mathbf{q}), W_{mt}] \end{aligned} \tag{15}$$

Similar to the two player case, the only unknown quantity is $E[U_{mt}(n) | \mathbf{e} \in \mathcal{E}_{\mathbf{d}}(\mathbf{d}, \mathbf{q})]$. This expression is just a more complicated version of traditional selection correction terms, like in Heckman (1979). This term is more complicated since we condition on a vector $\mathbf{e}$ which is truncated along its many dimensions according to $\mathbf{d}, \mathbf{q}$, instead of being truncated along a single dimension like in the binary treatment case. However, these complications are not conceptual and only create a few additional computational difficulties. I explain in detail how I handle the selection correction terms in Section 7.4.

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 fixed effect for each municipality. For example, consider the fixed effect model $D_{gmt} = 1[\phi_m + \delta\tilde{\mathbf{Z}}_{mt-1} + \gamma W_{mt} > \epsilon_{gmt}]$. The effect ϕ_m is an additive fixed effect, which varies by municipality.

Modeling time invariant heterogeneity as an additive fixed effect is broadly inconsistent with incomplete information entry games, as explained in Aguirregabiria and Mira (2019). The reason is we must also allow the covariates' effects to vary by municipality as well. In other words, a more flexible model like $D_{gmt} = 1[\phi_m + \delta_m\tilde{\mathbf{Z}}_{mt-1} + \gamma_m W_{mt} > \epsilon_{gmt}]$ with municipality-specific coefficients (δ_m, γ_m) is more appropriate. There are two reasons behind the municipality-specific coefficients (δ_m, γ_m), as explained in Aguirregabiria and Mira (2019). First, entry games typically do not have a unique equilibrium, and instead can have multiple possible equilibria across municipalities. With multiple equilibria, DTOs have different parameters for entry across municipalities. As a result, the same value of observable covariates $(\tilde{\mathbf{Z}}_{mt-1}, W_{mt})$ will have different effects across the equilibria. Under this scenario, a model with municipality specific coefficients (δ_m, γ_m), is therefore more desirable. The second reason behind municipality specific coefficients is more subtle. Even if groups are playing the same equilibrium across municipalities, any time-invariant differences in flow profits across municipalities will require municipality specific coefficients. This point is more nuanced than the point concerning multiple equilibria. I refer readers to Aguirregabiria and Mira (2019) which explains it in better detail.[25]

In summary, to make the estimation robust to multiple equilibria and time-invariant effects, a more flexible model with municipality specific coefficients, like $D_{gmt} = 1[\phi_m + \delta_m \tilde{\mathbf{Z}}_{mt-1} + \gamma_m W_{mt} > \epsilon_{gmt}]$ is desirable. Of course this model is demanding and is infeasible in my setting or with datasets with similar time-horizons.

Instead, to control for time-invariant differences in the first stage, I employ the group fixed effect estimation strategy (Bonhomme et al., 2022; Saggio, 2012). The substance of the procedure involve using k-means clustering (Steinley, 2006) to group similar municipalities into C groups, with $C < M$, where M is the number of municipalities. This strategy reduces the dimensionality of the M fixed effects into a lower C dimensional grouping, by clustering together municipalities that are similar according to covariates specified by the researcher. 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.*

The grouped fixed effect framework is a tractable procedure that solves the complications of multiple equlibria and time-invariant effects in entry game models. I allow for *interactions* between the municipality types, given by $C(R_m)$, and the $(\tilde{\mathbf{Z}}_{mt-1}, W_{mt})$ covariates. This model is given by $D_{gmt} = 1[\phi_c + \delta_c \tilde{\mathbf{Z}}_{mt-1} + \gamma_c W_{mt} > \epsilon_{gmt}]$ where the $(\phi_c, \delta_c, \gamma_c)$ now vary by the municipality type $C(R_m)$, instead of municipality. This model is therefore a middle-ground between models that force (δ, γ) to be constant across municipalities and model that allows a separate coefficient for each municipality. Allowing for these interactions is a simple and tractable way to check the sensitivity of the results to multiple equilibria.

In the grouped fixed effect framework, the researcher only needs to specify the number of clusters C and which covariates to use in the clustering algorithm. Following Bonhomme et al. (2022) I use the average number of years each municipality is observed with a group

[25] Another reason to avoid including municipality specific intercepts is the the incidental parameters problem (Neyman and Scott, 1948), which makes estimating additive effects difficult in binary choice models. Intuitively, the non-linearity of typical binary choice problems means the inconsistency in estimating ϕ_m spills over into the structural parameters (δ, γ), even if these are the same across municipalities.

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.

My baseline model includes 5 clusters and excludes interactions between the covariates and municipality types. I show in Section 8.6.4 the results are robust to allowing for interactions between municipality types and the covariates. As a result, the results are robust to allowing for multiple equilibria across municipalities. Nonetheless, researchers in other settings can use the previous framework to check the sensitivity of their own results. Also, in Section 8.6.6 I show the results are robust to increasing the number of clusters from 5.

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.

$$
\begin{aligned}
D_{gmt} = 1[&\underbrace{\xi}_{\text{Constant}} + \underbrace{X_{mt}\beta_1^1}_{\text{Population, Agri. Controls}} + \underbrace{\phi_{0g}^1 1[G = g]}_{\text{DTO Fixed Effect}} + \underbrace{L_t\phi_1^1 + \sum_{c=1}^{C} 1[C(R_m) = c] \times \phi_{2c}^1}_{\text{Time and Municipality Type Fixed Effects}} \\
&+ \underbrace{\gamma_1^1 D_{gmt-1} + \gamma_2^1 N_{mt-1}}_{\text{Incumbency Controls}} + \\
&\underbrace{\gamma_{3c}^1 \tilde{Z}_{gmt-1} + \gamma_4^1 \tilde{Z}_{gmt-1}^2 + \gamma_5 \tilde{Z}_{gmt-1}^3 + \gamma_6 \overline{\tilde{Z}}_{-gmt-1} + \gamma_7^1 \overline{\tilde{Z}^2}_{-gmt-1} + \gamma_8^1 \overline{\tilde{Z}^3}_{-gmt-1}}_{\text{Cubic Polynomial Instruments}} > \epsilon_{gmt}]
\end{aligned} \tag{16}
$$

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. N_{mt-1} is the lagged group count. $\tilde{Z}_{gmt-1}$ is the log of group g's distance to m for the preceding year; $\tilde{Z}^2_{gmt-1}, \tilde{Z}^3_{gmt-1}$ is its square and cube. $\overline{\tilde{Z}}_{-gmt-1}$ is the average of the log distance for all other groups $j \neq g$; $\overline{\tilde{Z}^2}_{-gmt-1}, \overline{\tilde{Z}^3}_{-gmt-1}$ is the average of the square and cubed logged distances.

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 first stage models entry decisions for all the DTOs. However, the DTOs' collective decisions will jointly determine the municipality-year's treatment dose at the municipality-year level, the number of active DTOs. Using the previous first stage model I estimate the entry probabilities for all players and collect these in $Q_{mt}(Z)$. I parametrize the potential outcomes to be of the form:

$$\begin{aligned} Y_{mt}(n) = \; & \underbrace{\alpha_n}_{\text{Treatment Effect}} + \underbrace{X_{mt}\beta_1^2}_{\text{Population, Agri. Controls}} \\ & + \underbrace{L_t\phi_1^2 + R_m\phi_2^2}_{\text{Time and Muni. FE}} + \underbrace{\sum_{g=1}^{G} \gamma_{1g}^2 D_{mtg-1} + \gamma_2^2 N_{mt-1}}_{\text{Incumbency Controls}} + U_{mt}(n) \end{aligned} \tag{17}$$

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 17 has M fixed effects captured in $R_m\phi_2^2$, while Equation 16 only has C fixed effects for the C 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 8.6.6 the robustness of my findings to the specification in Equation 17. 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 $E[U_{mt}(n)|\mathbf{e} \in \mathcal{E}_{\mathbf{d}}(\mathbf{d},\mathbf{q})]P[\mathbf{e} \in \mathcal{E}_{\mathbf{d}}(\mathbf{d},\mathbf{q})|\mathbf{e} \in \mathcal{E}_n(n,\mathbf{q}), W_{mt}]$ terms in equation 15. 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:

$$E[U_{mt}(n) - U_{mt}(0)|\mathbf{e} \in \mathcal{E}_{\mathbf{d}}(\mathbf{d},\mathbf{q})]P[\mathbf{e} \in \mathcal{E}_{\mathbf{d}}(\mathbf{d},\mathbf{q})|W_{mt}] = \sum_{g=1}^{G}(\sigma_{ng} - \sigma_{0g})\lambda(d_g, q_g) \tag{18}$$

where d_g, q_g are the g^{th} components of an entry vector $d \in \mathcal{D}$ and the vector of entry probabilities q, and λ is a known function.[26] As we can see the normality assumption yields

---

[26]

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

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.

a convenient functional form.

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 $Q(\tilde{\mathbf{Z}}_{mt-1}, W_{mt})$ and covariates W_{mt} This yields the following estimating equation:

$$
\begin{aligned}
E[Y_{mt}|\mathbf{Q_m}(\tilde{\mathbf{Z}}_{mt-1}, W_{mt}) = \mathbf{q}, W_{mt}] = & \underbrace{\alpha_0}_{\text{Intercept}} + \underbrace{X_{mt}\beta_1^2}_{\text{Population, Agri. Controls}} \\
& + \underbrace{\phi_1^2 L_t + \phi_2^2 R_m}_{\text{Muni. Time FEs}} + \underbrace{\sum_{g=1}^{G} \gamma_{1g}^2 D_{mtg-1} + \gamma_2^2 N_{mt-1}}_{\text{Incumbency Controls}} \\
& + \sum_{n=0}^{G} \Big\{ \underbrace{(\alpha_n - \alpha_0) P[N_{mt} = n|\mathbf{Q_m}(\tilde{\mathbf{Z}}_{mt-1}, W_{mt}), W_{mt}]}_{\text{Treatment Effect}} + \underbrace{\sum_{g=1}^{G} (\sigma_{ng} - \sigma_{0g}) \sum_{d \in \mathcal{N}(n)} \lambda(d_g, q_g)}_{\text{Control Function}} \Big\}
\end{aligned}
\tag{19}
$$

Notice that equation 19 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 $P[N_{mt} = n|\mathbf{Q}_m(\tilde{\mathbf{Z}}_{mt-1}, W_{mt}), W_{mt}]$, the probability of observing a particular group count n. Given the individual groups first stage probabilities and the independent errors between groups, this term is a simple function of the individual group probabilities.[27] Second the probabilities q are used to construct the control function $\lambda(d_g, q_g)$.

I explore the sensitivity of my results to the normality assumption in Section 8.6.3. I broadly find that results are similar to a global polynomial control function.

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 16. I then use these probabilities to construct the appropriate terms in equation 19 and run OLS.

[27]For instance with two groups, $P[N_{mt} = 1] = q_1(1 - q_2) + (1 - q_1)q_2$

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 guaranteed to be placed in only one treatment dose. As a result they are poor control units when trying to estimate treatment effects. I drop 1575 observations with this restriction.[28] Table A7 provides descriptives for these observations. 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.[29]

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

[28] 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.

[29] 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 but still have non-degenerate probability for at least one treatment dose. These observations are informative for *some* but not *all* of the treatment doses, and it is therefore useful to keep them in the sample.

municipalities over others across simulation draws. This mimics a classical bootstrap, where municipalities that are sampled more effectively receive more weight during estimation. The main benefit of this procedure is no observations are ever dropped, which is important for the fixed effect OLS in the second step.

8 Results

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 16 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 0.01 percentage point (pp) increase in the likelihood of entry, while a 1% decrease in the average of rival distances decreases entry by 0.01 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 60.8 pp more likely to remain

Table 4: First Stage Average Partial Effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Own Distance	-0.010***	-0.010***	-0.010***	-0.010***	-0.009***	-0.011***	-0.010***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Avg. Rival Distance	0.011***	0.007***	0.010***	0.011***		0.011***	0.011***
	(0.001)	(0.001)	(0.001)	(0.001)		(0.001)	(0.001)
Min. Rival Distance					0.002		
					(0.002)		
Lag Presence	0.608***	0.656***	0.613***	0.608***	0.621***	0.605***	0.634***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)	(0.141)
Lag # DTOs	-0.005***	0.010***	-0.005***	-0.005***	-0.005***		-0.002
	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)		(0.001)
1 DTO						-0.002	
						(0.002)	
2 DTOs						-0.005*	
						(0.003)	
3 DTOs						-0.014***	
						(0.003)	
4 DTOs						-0.013***	
						(0.003)	
5 DTOs						-0.019***	
						(0.003)	
6 DTOs						-0.024***	
						(0.003)	
7 DTOs						-0.027***	
						(0.004)	
8 DTOs						-0.032***	
						(0.006)	
IHS(Ag. Value)	0.000**	0.001***	0.000*	0.001**	0.001**	0.000**	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log(Population)	0.004***	0.010***	0.004***	0.004***	0.004***	0.004***	0.004***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Muni. Type FE	✓		✓	✓	✓	✓	✓
Muni. Type Interaction							✓
Max Power	3	3	2	4	3	3	3
BIC	41863	46385	41898	41861	42009	41845	41197
AIC	42213	46694	42227	42231	42359	42267	41877
Log-Likelihood	-20897	-23162	-20917	-20894	-20970	-20882	-20533
Model d.f.	33	29	31	35	33	40	65
N	219070	219070	219070	219070	219070	219070	219070

Notes: Average partial effects (APEs) of a probit model. 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. For variables that are interacted with municipality type, the APEs are averaged across the interactions. All estimates include controls for log population, IHS of agricultural revenue, DTO and year fixed effects. Models use 5 municipality types. Column 1 estimates the pooled first stage equation 16. Column 2 excludes municipality type fixed effects. Column 3 uses a quadratic polynomial for the distances, Column 4 uses a fourth order polynomial. Column 5 uses the minimum rival distance instead of the rival. Column 6 dummies out the lag number of DTOs. Column 7 includes municipality type interactions between the distances variables, lag presence and lag # DTOs. "Muni. Type FE" indicates if regression includes municipality type fixed effects. "Muni. type interactions" indicates if the municipality types are interacted with the distance and incumbency variables. Max power shows the polynomial used for 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. Unit of observation is a DTO-municipality-year. Sample years are 2006-2018. Standard errors from 1000 bootstrap replications.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

in a municipality if it was there before. Lag DTO count also has a significant coefficient of -0.005. That is, an additional DTO incumbent reduces the probability of entry by 0.5 percentage points. Column 6 estimates a similar model but dummying out the value of lagged incumbents. The coefficients indicate there is no substantial non-linearity in the number of incumbents, and the linear model offers a good approximation. Overall, the effects of the incumbency variables are again consistent with DTOs behaving as strategic rivals.

Column 2 repeats the Equation 16 but with no indicators for municipality type. Reassuringly the distance instruments still have the same direction and significance. However, the sign for the lag # DTOs is now positive, a counter intuitive result which is inconsistent with the sign of the distance instruments. This highlights the importance of controlling for time-invariant differences across municipalities. Columns 3 and 4 repeat the main equation using square and fourth order polynomials, and the estimates are essentially identical. Column 5 uses the minimum of rivals' distance instead of the average. The coefficient for the minimum is still positive but statistically insignificant and much smaller than the value for the average. Moreover its Bayesian and Akaike Information Criterion (BIC,AIC) values are larger than in column 1, suggesting the minimum is a worse fit.

Finally, column 7 repeats the model in column 1, but allows for interactions between the instrument and incumbency variables, with the municipality types. As explained in Section 7.1, these interactions allow a tractable and simple way to model multiple equilibria. Again the significance and sign of the distance instruments are unchanged, but the effect of lag incumbents is now insignificant. The model in column 7 offers a better fit than column 1 according to BIC and AIC. However, the added flexibility results in noisier second stage estimates, as I explore in Section 8.6.4. As a result, I opt for the main model in column 1 and exclude these interactions.

To my knowledge these estimates are the first empirical evidence demonstrating DTOs within Mexico are strategic rivals. Their rivalry is not obvious a priori. DTOs may instead have been complementary. For example, existing DTOs may degrade government resources and public institutions, thereby making criminal activity easier. Under this scenario, DTOs may actually benefit from rivals' presence and decide to enter where other DTOs already operate. My estimates indicate this is not the case, and that instead we should treat these criminal groups as rivals.

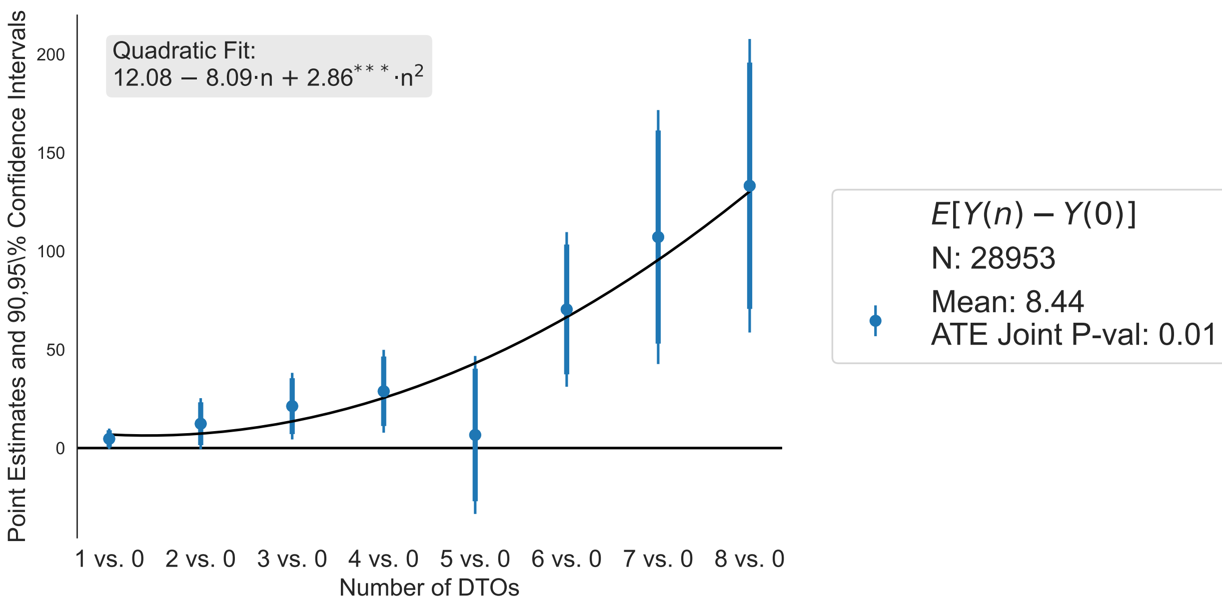
8.2 Homicide Estimates

I first examine homicide outcomes. I show that homicides increase with more DTOs, and this relationship is convex. This convex relationship is important for policy since it suggests reducing DTO presence is more effective in places with many DTOs. Furthermore, the results suggests that even a marginal change of a 1 DTO reduction can yield substantive reductions in violence. It is unnecessary to eliminate all DTOs in order to reduce violence; small changes in DTO count can have important effects. These results contrast with many stylized models of crime, which predict that an additional entrant should have little to no effect with a high number of criminal groups (Hirshleifer, 1995). These theoretical models predict a concave relationship, while my empirical results suggest a convex one. As a result, the theoretical models predict an additional entrant should lead to small changes in violence, while I find the opposite, and an additional DTO can lead to large changes in violence.

The results of the main model are in Figure 7. As in Section 5 I examine the count of homicides as the main outcome, and I control for log population, IHS of agricultural revenue, year and municipality dummies in the specifications. Each point is an estimated coefficient along with its 90% and 95% confidence interval resulting from the 1000 bootstrap replications. All the coefficients show the ATE relative to having zero DTOs, that is $E[Y(n)-Y(0)]$. Additionally, I overlay a quadratic fit to the estimate ATEs to statistically examine its non-linearity.

We observe a clear upwards trend in the count of homicides: Homicides tend to increase as more DTOs enter. Furthermore the relationship is increasing and convex. For example, the reduction in homicides from 4 DTOs to 3 is 7.56 additional homicides, while from 7 to 6 DTOs it is 36.76, almost 5 times larger than the decrease from 4 DTOs to 3. The quadratic fit confirms the non-linear relationship. The quadratic term is positive and statistically significant at conventional levels, providing statistical evidence for a non-linear relationship. Appendix Table A8 additionally examines a cubic relationship. A joint test of the non-linear terms also rejects a null effect. Overall there is strong evidence for an increasing, convex relationship.

Figure 7: Main Homicide Results



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 is the number of homicides. Controls include log population, the IHS of agricultural revenue, and municipality-year effects. The black line is a quadratic fit of the Average Treatment Effects. The quadratic fit equation is in the top left corner. 90 and 95 % confidence intervals shown. Standard errors from 1000 bootstrap replications.

How to interpret these effects? The regressor of interest is the number of DTOs in a municipality. An increased number of DTOs bundles together several potential mechanisms. I offer two mechanisms for interpretation. First, we expect more competition between DTOs in the activities they engage in. This competition may be price competition in local retail sales, labor competition for workers, or increased competition in extortions and kidnappings. Second, in addition to competition, we expect there to be an increase in criminal capital. DTOs in Mexico have extensive criminal know-how and connections with many illicit markets. I argue these connections and knowhow are a form of criminal capital which impacts the level of crime, and is another avenue explaining the expansion in violence and crime in the homicide estimates. Therefore, the entry of additional DTOs not only represents increased crime due to competition, but also by an increase in criminal capital.

In fact, the results for a single DTO suggest there must be an additional channel apart from competition. Having a single DTO versus none adds an additional 4.7 homicides, and this effect is statistically significant at the 10% level.[30] Though only marginally significant this positive effect indicates there must be a channel apart from competition driving the results. There clearly is no increase in competition when there is a single DTO, yet I find an increase in homicides. I interpret an increase in criminal capital as a likely explanation, but there may be other explanations as well.

The result for a monopolist contrasts previous results in the literature. The case of a criminal monopoly has been studied previously in Brazil and Mexico (Biderman et al., 2019; Sobrino, 2020), finding either no effect or even a *decrease* in violence after the monopolist's entry. These findings are rationalized as a monopolist criminal organization imposing its own version of law and order, something it can't do when competing with rival criminal groups. However, my results point that this is not the case, and that even a monopolist criminal group can increase violence and crime.

Several reasons could explain the discrepancy between my results and (Biderman et al., 2019; Sobrino, 2020). First, this set of papers use a difference in differences design, which assumes there are no time-varying unobservables affecting DTO entry and homicides. In contrast, my instrumental variable approach allows for time-varying unobservables, so long as the exclusion and independence conditions hold. Second, panel designs typically deliver Average Treatment Effects on the Treated (ATT), not the ATEs I focus on. If DTOs enter precisely when they expect there to be fewer homicides after entry, due to time-varying factors, that could also explain the lower estimated effects for the previous panel studies relative to my main estimates. I leave this for future work and only offer speculative reasons on the discrepancy for now.

[30]The confidence interval for 1 DTO is difficult to observe in the graph. Estimates are also reported in column 1 of Table A5.

8.2.1 Comparison to Linear Model

Given the non-linear relationship, to what extent would a linear approximation deviate from my main estimates? To examine this discrepancy, I estimate a competing model which is linear in the number of DTOs. Specifically, I model potential outcomes as:

$$Y_{mt}(n) = \alpha + \delta \times n + \psi(W_{mt}) + U_{mt}(n) \tag{20}$$

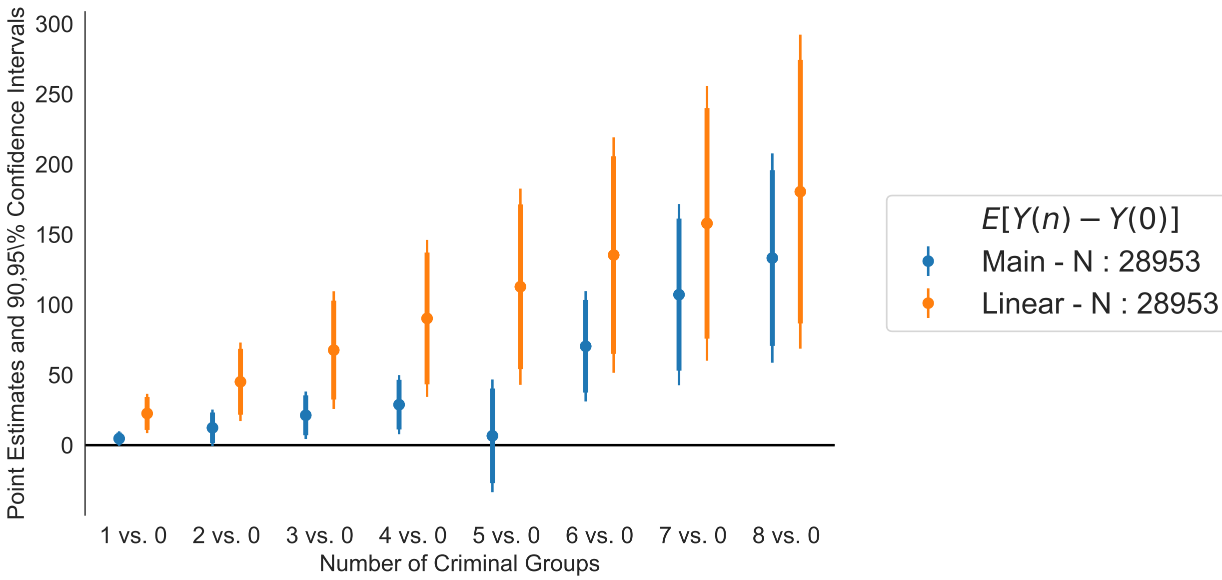
where instead of allowing for a separate intercept for each n, I impose a linear relationship with common intercept α. I then reestimate the model, but with this linear relationship.

Results can be see in Figure 8. We see the linear relationship consistently over-estimates the impact of an additional DTO. These differences are large. The effect with 4 DTOs relative to 0 is 28.8 in my main model. The corresponding linear estimate is 90.24, almost three times larger. Estimates are also reported in Table A9 for clarity. Additionally, we observe that the standard errors are also wider in the linear model in comparison to the main estimates. Overall, I find a linear model is a poor fit for the non-linear relationship with homicides as an outcome.

The linear model is not always a poor fit. As discussed in the next Section 8.3, the results for school dropout are linear, and as a result the linear potential outcomes model is a good fit. The same comparison between the linear model and the main model is seen in Figure A9. As can be seen, the linear model is an excellent approximation for school dropout, and has slightly smaller standard errors.

I conclude the linear model is a poor approximation for homicides because the homicide relationship is highly non-linear. Furthermore, the linear model necessarily masks the relevant non-linearities documented earlier. As a result, the linear model not only provides a poor approximation, but ignores the disproportionate policy gains to reducing DTO presence with many existing DTOs.

Figure 8: Homicides - Comparison to Linear Model



Notes: This figure plots Average Treatment Effect estimates, $E[Y(n) - Y(0)]$, using the normal model outlined in sections 6 and 7. I compare the main estimates to a linear model of potential outcomes. Controls include log population, the IHS of agricultural revenue, and municipality-year effects. 90 and 95 % confidence intervals shown. Standard errors from 1000 bootstrap replications.

8.3 School Dropout Estimates

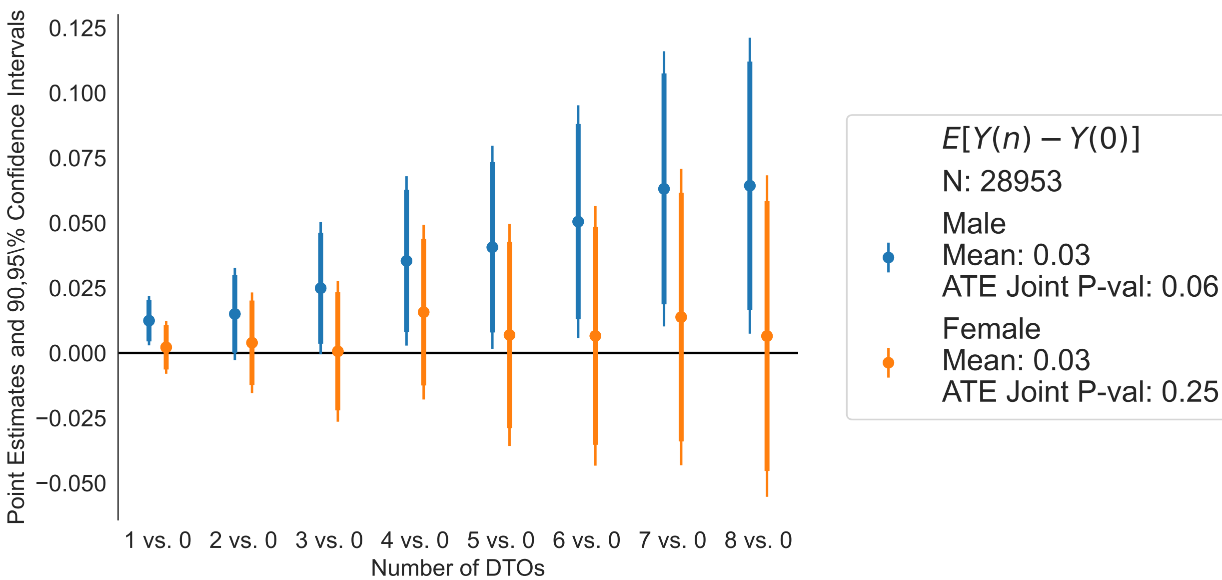
I now examine effects on school dropout. The results for homicide indicate an important expansion in criminal activity. Increased crime can affect schooling by a variety of channels. I focus on three possible channels in this section. First, crime can affect schooling by being through a *disruption effect*. Increased crime can make it more dangerous to travel to school, disrupting school days with violent events or increasing teacher turnover due to safety issues. Second, households may decide to *migrate* as a result of the violence. Another channel is a *labor demand effect*. Increased DTO presence may affect dropout by increasing labor demand for potential criminals, typically young men. In this section, I explore these channels by examining grade by grade dropout results, decomposed by gender. Overall the effects are concentrated amongst male students in 9th grade, and I find no effect for females or younger grades. Put together, the gendered and age-differentiated results suggest neither disruption of school activities nor migration are a primary mechanism, and that DTO recruitment may instead be the primary channel.

Ninth grade results for school dropout are in Figure 9. I separate effects for male and female students in the ninth grade.[31] Dropout rates are calculated as the dropout between the start and end of the school year, divided by enrollment at the beginning of the year. As we can see, an increase in DTO presence increases school dropout for male students but does not affect female students. Furthermore this relationship is largely linear, with each additional DTO causing an increase of 1-2 percentage points in school dropout. Moreover, while there is overlap in the standard errors for male and female students, the joint test testing if all the estimates are zero rejects for male students with a p-value of 0.06 and fails to reject for female students, with a p-value of 0.25. Overall, the estimates suggest important effects for male students, but do not provide evidence for effects for female students.

To help contextualize the 9th grade results, I examine effects for other grades as well. I examine effects on grades 1-9, grouping together grades 1-5 for parsimony. Grade 6 is the final grade of elementary school in Mexico and is an important transition grade. To

[31] Figure A2 shows results combining gender.

Figure 9: 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 for panel (a) is 9th grade dropout rates for all genders. Panel (b) displays effects for ninth grade dropout rates for males and females. Dropout rates is calculated based on the dropout between the start and end of school, as a fraction of start of school enrolment. 90 and 95 % confidence intervals shown. Standard errors from 1000 bootstrap replications.

ease exposition, I show the estimated effects for having "Any" DTO. Specifically, I modify potential outcomes to be

$$Y_{mt}(n) = \begin{cases} \alpha_0 + \psi(W_{mt}) + U_{mt}(n) & n = 0 \\ \alpha_* + \psi(W_{mt}) + U_{mt}(n) & n > 0 \end{cases} \tag{21}$$

Where α_* is the effect of having any non-zero amount of DTOs, and the controls $\psi(W_{mt})$ are as in the main Equation 17. I do this for parsimony, in order to reduce the number of parameters displayed in the graph. Nonetheless, Figure A3 shows estimates where I do not collapse the treatment effects. The joint F-tests of the unrestricted model agree with the simplified version presented in the main text below.

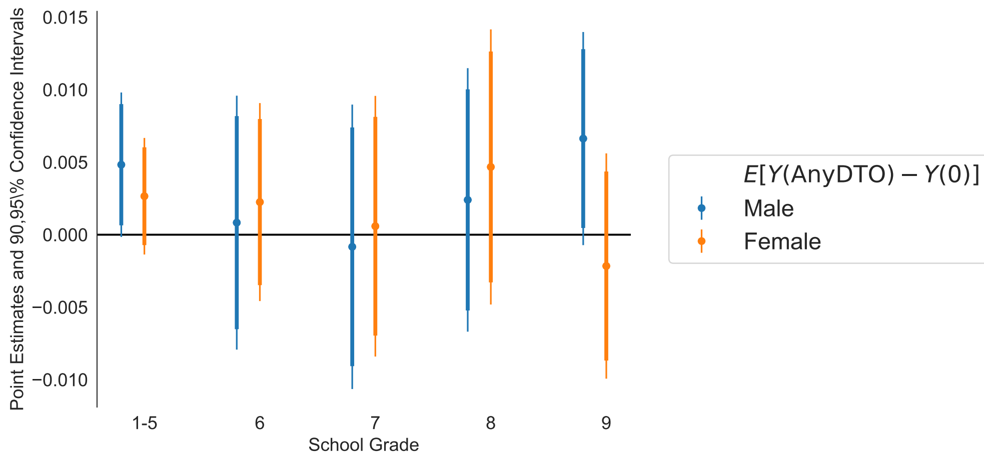
Results are in the top panel of Figure 10. The x-axis shows the grades. Reassuringly, we still find a positive effect for males in 9th grade, and a null effect for females in 9th grade as well, repeating the previous 9th grade results. There is only one statistically significant effect for males in younger grades, for grades 1-5. I fail to find any significant effects for grades 6, 7 and 8. Overall, there do not seem to be important responses to DTO presence in grades below ninth grade.

Panel (b) of the figure plots average dropout rates. As we can see, there is less than 1% dropout in elementary school, between grades 1-6. Grade 7 has a large increase in dropout, reaching almost 5% for males and 3% for females. Grade 7 is the first grade of middle school in Mexico, and explains why there is an increase in dropout, since students adjust to the new phase of schooling. Nonetheless it is interesting to note that there no effects for DTO presence in grade 7, precisely when mean dropout is largest.

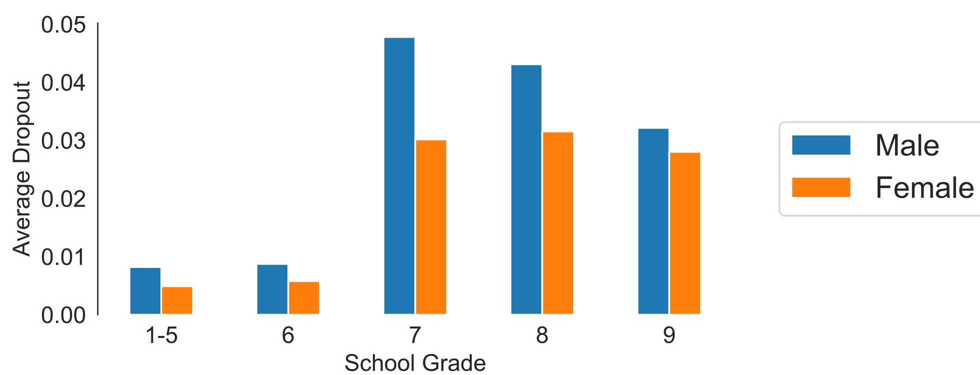
Overall, the grade-by grade results indicate that disruption and migration are an unlikely explanations for the 9th grade dropout results. If criminal disruption were a primary driver, we would expect effects for both genders and across grades. Similarly, if migration were a concern, we would expect to find effects in younger grades as well. However, there are only significant effects for 9th grade males and a single positive effect for grade 1-5. In particular, I fail to find effects for grade 7, when there is the largest amount of dropout

Figure 10: Dropout Effects for Any DTO By Grade

(a) Dropout Effects for Any DTO by Grade



(b) Average Dropout by Grade



Notes: Panel (a) plots Average Treatment Effects for having any positive number of DTOs. The outcomes are dropout for male and female students for each grade grade in the x-axis. Panel (b) plots mean dropout rates by grade. Unit of observation muncipality-year. Sample years are 2006-2018. 90 and 95 % confidence intervals shown. Standard errors from 1000 bootstrap replications.

for males on average. Last, while the grade-by-grade patterns suggest migration is not a primary concern, I examine migration directly using individualized data in Section 8.6.1. Again I find migration does not drive my results.

8.3.1 Heterogeneity by Violence

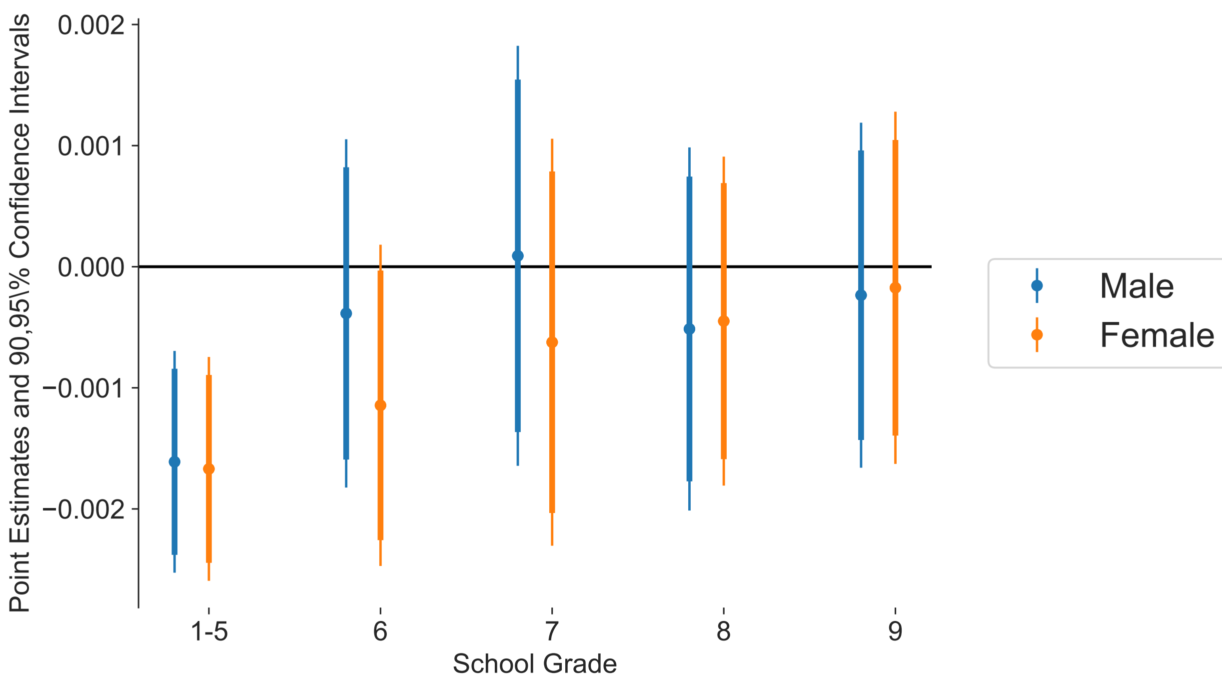
To further examine migration and disruption as possible channels, I examine heterogeneity by the previous year's homicide count. If migration and disruptions are important channels, then we expect last year's homicides to inform migratory decisions. I hypothesize more violent locations should induce outmigration. Similarly, more violent locations should suffer from more disruptive criminal activity. To examine these effects, I again change potential outcomes to include an interaction effect with the square root of last years homicides:

$$Y_{mt}(n) = \begin{cases} \alpha_0 + \beta_0 \sqrt{\text{Homicides}_{mt-1}} \psi(W_{mt}) + U_{mt}(n) & n = 0 \\ \alpha_* + \beta_* \sqrt{\text{Homicides}_{mt-1}} + \psi(W_{mt}) + U_{mt}(n) & n > 0 \end{cases} \tag{22}$$

I employ the square-root due to the long right tail in muncipal homicides within Mexico. Table A3 also uses the Inverse Hyperbolic Sine and the raw count of homicides. The IHS results show a similar pattern, and unsurpirsingly the raw homicide counts are significantly noisier and uninformative. As a result, I focus on the square root coefficients β_* below.

Surprisingly, I find no effect for violence in grades 6-9, though there is a marginally significant effect for 6th grade females. Moreover, I find *negative* effects for grades 1-5. This indicates that, in places with more violence in the previous year, dropout decreases for elementary school. This is consistent with families keeping their young children longer in school during more violent times, a natural response if schools are considered to be safe spaces. Regardless of this last interpretation, these patterns do not support disruption or migration as important channels for the 9th grade effects.

Figure 11: Dropout Effects for Any DTO. Heterogeneity by Square Root of Lagged Homicides



Notes: This figure plots the coefficients for the square root of lagged homicides interacted with the probability of having any DTO. The outcomes are dropout for males and females. 90 and 95 % confidence intervals shown. Standard errors from 1000 bootstrap replications.

8.4 Criminal Labor Demand

The grade by grade and gendered results so far suggest disruption or migration are not primary drivers of the dropout effects observed for ninth grade males. Another possible channel is increased labor demand for criminal activities on the part of DTOs. I examine this channel directly below. I examine two indicators suggestive of criminal labor demand: Crimes with adolescent age offenders and homicides with adolescent victims. Both indicators increase with the number of active DTOs.

For adolescent crimes, the arrest data indicates if the perpetrator was a middle school dropout. This is a highly relevant measure for my study, since I mainly find effects for dropout in the ninth grade, the last grade of middle school for Mexico. Criminals that are middle school dropouts therefore contain the students affected by DTO presence in the previous results in Figure 9. As a result, the previous results in Figure 9 document that dropout increases for ninth grade males. In this section I show if crimes with these same dropouts also increase. Important to note, this data is only available for the years 2014-2018 and hence we expect the results to be slightly noisier.

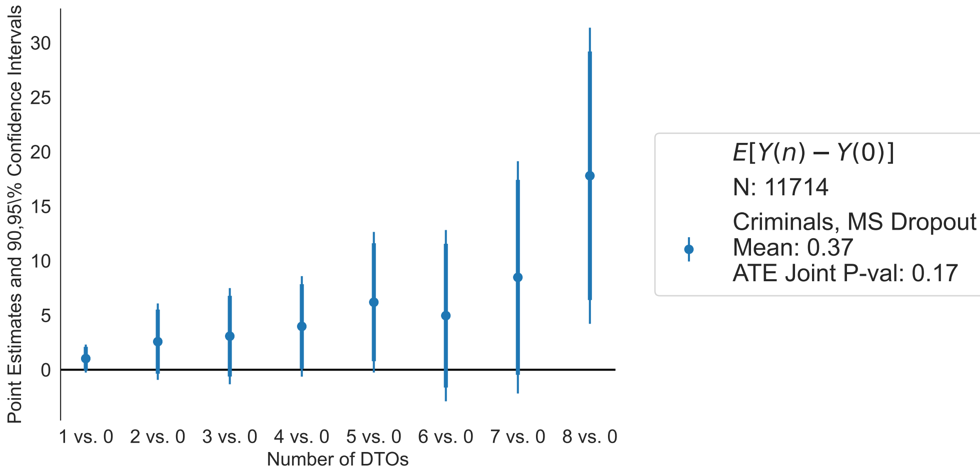
For adolescent homicides, I examine homicides with a 15-18 year old victim. I focus on this age group since it is after ninth grade in Mexico. Ninth grade students are typically 14 years old in Mexico. Prior work has shown (Carvalho and Soares, 2016) there is a career ladder for young criminals, typically beginning with low-stakes responsibilities and progressing in importance and risk after a few years. As a result, we expect there to be no detectable effect for 14 year old homicides, but do expect an effect for 15-18 year olds. I examine homicides with 12-14 year old victims in Appendix Figure A5 and indeed find no effect.

Results can be seen in the top and bottom panels of Figure 12. Here I replicate the main model of Equation 10, which estimates a separate effects for each number of DTO present. The top panel shows results for adolescent crimes with middle school dropout as perpetrators. The bottom panel shows homicides with 15-18 year old victims.

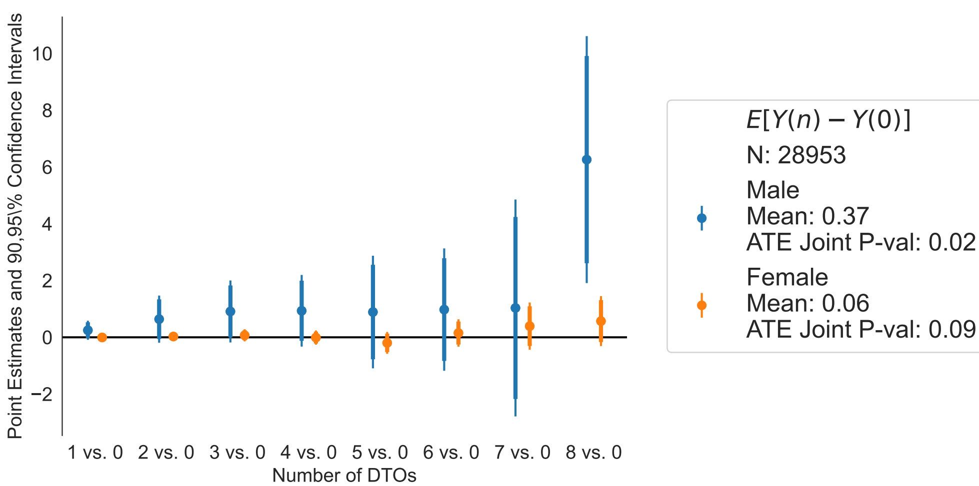
We find a positive effect of DTOs for both outcomes. In particular, crimes with middle

Figure 12: Adolescent Crimes and Homicides

(a) Adolescent Crimes (Middle School Dropouts)



(b) Homicides, Victims Aged 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 the number of crimes with an adolescent perpetrator who also dropped out of middle school, for years 2014-2018. Panel (b) displays outcomes for homicides with a 15-18 year old victim, by gender, and for years 2006-2018. 90 and 95 % confidence intervals shown. Standard errors from 1000 bootstrap replications.

school dropouts have an increasing pattern, with the largest increase occurring with 8 DTOs. The figure suggest adolescent crimes increase with DTOs. However, the joint F-test fails to reject with a p-value of 0.17. This is due to two reasons. First, the limited years with this outcome's data, only from 2014-2018. Second, adolescent crimes with middle school dropouts are relatively rare, with an overall mean of only 0.37 crimes. Appendix Figure A4 instead examines all adolescent crime, and finds a similar pattern. However this outcome is almost twice as likely, with a mean of 0.63 adolescent crimes, and as a result the joint null is confidently rejected at conventional levels.

The bottom panel of Figure 12 examines homicides with 15-18 year olds, by gender. Though most of the male coefficients are insignificant, they are all positive and the joint-test rejects the null of no effect with a p-value of 0.02. The joint test for females is marginally significant with a p-value of 0.09, but all the estimated coefficients are much smaller in magnitude and close to zero.

Overall, these findings suggest that crimes with middle school dropout adolescents increase with DTOs, in line with the previous middle school dropout effects. Though naturally a harder channel to test due to data constraints, the criminal labor demand channel seems to be the primary explanation behind the ninth grade dropout effects documented earlier.

8.5 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 ??. 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. 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.

8.6 Additional Results and Robustness Checks

I conduct several robustness checks to validate the main findings. I also compare my estimates to more conventional OLS estimates.

8.6.1 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 A7. 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.

8.6.2 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.

8.6.3 Sensitivity to Normality

My main model assumes a joint normal distribution for the errors, which yields a convenient functional form. As a robustness check to this assumption, I instead employ a global polynomial model as a control function.[32] I examine quadratic and cubic versions of the control functions. These estimates are significantly noisier than the normal model results, likely due to the increased correlation between the probability of observing a certain group count N in equation 19 and these polynomial terms. As a result, I collapse together potential outcomes for $n > K$ where K is a cutoff. I collapse effects for 3,4 and 5 DTOs or more.

Results can be seen in Figures A13-A15. As we can see the estimates get noisier as we increase the number of possible treated states. Nonetheless, the joint F-test for the ATEs agree with the normal model in all versions. Furthermore the effects are qualitatively similar, especially when I restrict the number of treatment effects to only 3. Overall, I find the normal model largely agrees with the polynomial results, but it is much more precise.

8.6.4 Sensitivity to Multiple Equilibria

To investigate the sensitivity of my results to multiple equilibria in the entry, I interact the covariates in the model with the grouped fixed effects. Specifically, I estimate the first stage as:

$$\begin{aligned} D_{gmt} = 1[\xi + X_{mt}\beta^1_{1c} + \phi^1_{0g}1[G = g] + L_t\phi^1_1 + \sum_{c=1}^{C} 1[C(R_m) = c] \times \phi^1_{2c} \\ + \gamma^1_{1c} D_{gmt-1} + \gamma^1_{2c} N_{mt-1} + \\ \gamma^1_{3c}\tilde{Z}_{gmt-1} + \gamma^1_{4c}\tilde{Z}^2_{gmt-1} + \gamma_{5c}\tilde{Z}^3_{gmt-1} + \gamma_{6c}\overline{\tilde{Z}}_{-gmt-1} + \gamma^1_{7c}\overline{\tilde{Z}^2}_{-gmt-1} + \gamma^1_{8c}\overline{\tilde{Z}^3}_{-gmt-1} > \epsilon_{gmt}] \end{aligned} \tag{23}$$

Where now all the coefficients on observable variables vary by municipality type c, by including interactions with municipality type. As such, this function can vary across municipalities

[32]Specifically, I model the unobservable expectation with $E[U(n)|\mathbf{e} = e] = \sum_{g=1}^{G}\sum_{p=1}^{P}\phi_{pg}e^p_g$ for a maximum power P.

m, mimicking the representation of Assumption 2. I modify the second stage equation as well to include these interactions:

$$
\begin{aligned}
E[Y_{mt}|\mathbf{Q_m}(\tilde{\mathbf{Z}}_{mt-1}, W_{mt}) = \mathbf{q}, W_{mt}] &= \alpha_0 + X_{mt}\beta_{1c}^2 \\
&+ \phi_1^2 L_t + \phi_2^2 R_m + \sum_{g=1}^{G} \gamma_{1gc}^2 D_{mtg-1} + \gamma_{2c}^2 N_{mt-1} \\
&+ \sum_{n=0}^{G} \Big\{ (\alpha_n - \alpha_0) P[N_{mt} = n | \mathbf{Q_m}(\tilde{\mathbf{Z}}_{mt-1}, W_{mt}), W_{mt}] + \sum_{g=1}^{G} (\sigma_{ng} - \sigma_{0g}) \sum_{d \in \mathcal{N}(n)} \lambda(d_g, q_g) \Big\}
\end{aligned} \tag{24}
$$

Where again the coefficients on the controls can vary by municipality type c. Results can be seen in Figure A11. As we can see the results are largely similar, although including the interactions does add more noise to the estimates. I also consider a middle ground between interacting all the covariates and interacting none of the covarites with the municipality type indicators. In this middle ground I only interact the instruments in the first stage model. Again the results are similar. I conclude the results are robust to multiple equilibria across municipalities, and accounting for these differences only adds noise to the main estimates.

8.6.5 Sensitivity to Non-Stationarity

I also assess the sensitivity of my results to a non stationary environment. As mentioned in Section 6.2 and Appendix Section D, the dynamic Markov game which can yield the reduced form entry in Assumption 2 requires a stationary environment. That is, groups have a good sense of how the covariates $(\mathbf{Z}_{mt}, W_{mt})$ evolve over time, and can form expectations on the future. My environment may not be stationary, given the rapid expansion of DTO presence during my sample years and that some DTOs did not exist at the beginning of my study period.

I estimate a similar model to the main one, except I allow for the coefficients to vary by different time periods. I divide my sample years into 3 time periods: 2006-2008,2009 and 2010-2018. I chose these time periods as they correspond to different periods with different amounts of active DTOs, as can be seen in Table A1. Different numbers of active DTOs provide a natural division of different possible environments across time.

Results can be seen in Figure A12. The results for homicides are comparable but systematically lower than the main estimates. Importantly, the joint test for any effect rejects with a p-value of 0.07. For male dropout in 9th grade, the estimates are slightly larger than the main estimates, but are similar in magnitude and statistical significance.

8.6.6 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 17.[33]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 agricultural output controls, and in column 4 I keep agricultural controls but drop population controls. Both the homicide and dropout effects are largely unchanged. Last, column 5 drops includes controls for the number of small DTOs in a municipality. These are DTOs tracked in my data, but which are not the 8 large DTOs studied in this paper. These DTOs are regional and have limited scope. The dropout results remain very similar. The homicide results are also similar but with overall smaller estimated magnitudes. Importantly, the joint-test rejects the null of no effects in all specifications.

Table A6 rerun the main specification but with 10, 20 and 50 municipality types respec-

[33]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.

tively. The dropout effects are again unchanged. The results on homicides become slightly smaller and the effects with less than 5 DTOs become insignficant. However th effects with 5 DTOs or more remains similar. Here as well, the joint test rejects the null of no effects for all 4 columns.

Overall we can see that the effects are quite robust and insensitive to the different specifications.

8.6.7 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. 1575 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 A7. Furthermore the dropped observations are concentrated at the beginning of the sample period, before the significant expansion of DTOs' in Mexico.

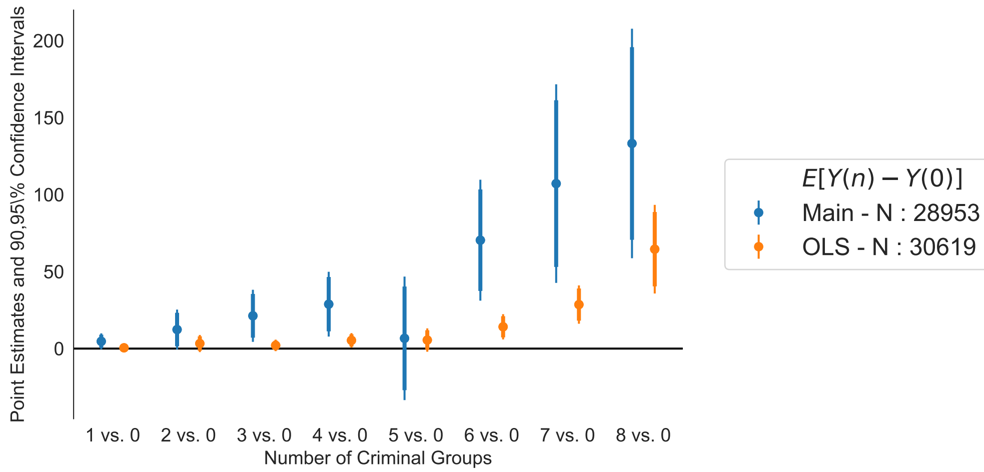
I assess the robustness of my main findings by including the entire sample. Results are displayed in Appendix Figure A8. As can be seen the estimates are virtually identical with or without the sample restriction for homicides and dropout. The estimates are slightly smaller in magnitude for dropout when I include all observations, and the joint test is insignificant with a p-value of 0.29. However the estimates are indistinguishable from the main specification, with overlapping confidence intervals.

8.6.8 Comparison to OLS estimates

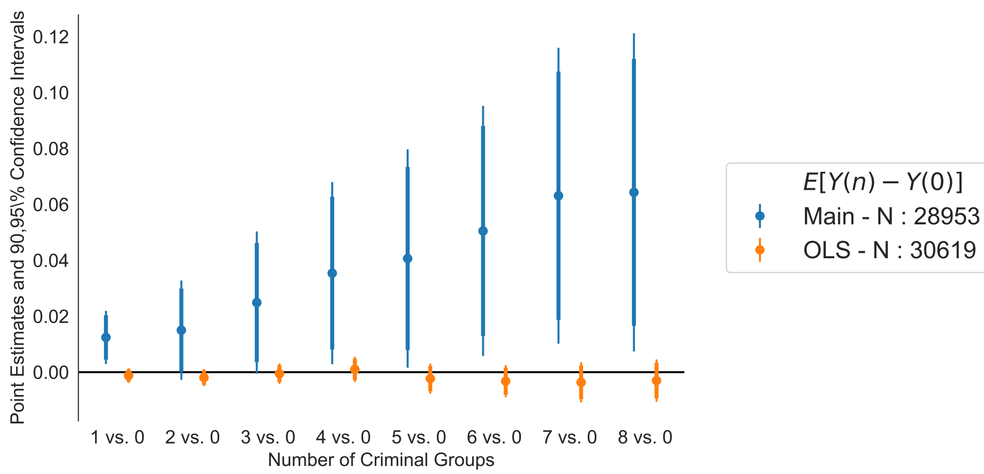
An OLS difference in differences specification provides a natural comparison to my main estimates. I show the estimates in Figure 13, also reported in Table A9. This figure reports my main estimates for homicides and male dropout, compared to a panel OLS model. In the OLS model, I dummy out the number of DTOs as the main regressors of interest. The panel OLS models have the same controls as the main estimates and also control for municipality and year fixed effects. The OLS model is therefore a difference in differences specification.

Figure 13: Comparison to OLS Estimates

(a) Homicides



(b) Male 9th Grade Dropout



Notes: Sample years are 2006-18. Panel (a) displays effects with homicides as outcomes, Panel (b) for 9th grade dropout for males. 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. Both models control for log population, IHS of agricultural production, dummy indicators for lag presence for the 8 DTO groups, and the lag DTO count. Main estimates display standard errors from 1000 bootstrap replications, OLS estimates display clustered standard errors by municipality. Standard errors from 1000 bootstrap replications.

As can be seen, the OLS estimates for homicides in the top panel are systematically smaller than the main estimates, though they are more precise. Interestingly, the OLS estimates also display the same convex shape. However the discrepancies are large and economically meaningful. For example, the OLS estimates of the effect of 4 DTOs relative to none are more than 5 times smaller than the main estimates (5.33 versus 28.84). Moreover the OLS estimates find no effect for a monopolist DTO, while I find a marginally significant positive effect. On the other hand the effects on dropout are all precisely estimated zeros, and therefore also underestimate the impacts of DTO entry.

What could explain the discrepancy in the estimates? I speculate it is the nature of selection. Table 1 establishes DTOs select to operate in relatively wealthier areas of Mexico. These well-endowed areas are precisely the areas with most resources to counteract the negative effects of DTOs. Table 1 shows they have more police offers, and their schools may also receive more investment. The fact that OLS underestimates the effects for both outcomes corroborates this view. However, establishing the exact reason for the discrepancy between my main estimates and OLS estimates is beyond the scope of this paper.

Table A9 also compares the main estimates to TSLS estimates. Column 4 produces results from the linear model in Section 5. Column 5 produces results form a "dummied" out model, where the TSLS endogenous variables is the dummied out DTO count. Column 6 produces results for the same dummied out model, but grouping together the effects for 5-8 DTOs. The linear TSLS model detects a positive effect for homicides, but no effect for school dropout. The dummied out models are noisy too imprecise to yield meaningful estimates. Nonetheless, the coefficient for one DTO is statistically significant for male 9th grade dropout.

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. Crimes with middle school dropout perpetrators also increase.

In contrast, the increases for homicides have an increasing convex relation in the number of 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 which can increase the production of crime, as a type 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

Acemoglu, D., G. De Feo, and G. D. De Luca (2020). Weak states: Causes and consequences of the Sicilian Mafia. *The Review of Economic Studies 87*(2), 537–581.

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 (2007a). Sequential Estimation of Dynamic Discrete Games. *Econometrica 75*(1), 1–53.

Aguirregabiria, V. and P. Mira (2007b). Sequential estimation of dynamic discrete games. *Econometrica : journal of the Econometric Society 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.

Alesina, A., S. Piccolo, and P. Pinotti (2019). Organized crime, violence, and politics. *The Review of Economic Studies 86*(2), 457–499.

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.

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

Bresnahan, T. F. and P. C. Reiss (1991, October). Entry and competition in concentrated markets. *Journal of Political Economy 99*(5), 977–1009.

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., C. Murry, and E. Tamer (2021, November). Market structure and competition in airline markets. *Journal of Political Economy 129*(11), 2995–3038.

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.

Colombiano, E. (2023). Cinco megacarteles y 23 narcobandas: así es el mapa criminal en Colombia. https://www.elcolombiano.com/colombia/cuales-son-los-grupos-armados-del-pais-y-como-estan-organizados-eln-y-clan-del-golfo-JK21184195.

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. and G. Dipoppa (2017, October). Mafia, elections and violence against politicians. *Journal of Public Economics 154*, 10–33.

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.

Dziewanski, D. (2020, July). It's hard to leave a Cape Town gang. But these men's stories show that it's possible. http://theconversation.com/its-hard-to-leave-a-cape-town-gang-but-these-mens-stories-show-that-its-possible-141208.

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.

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.

Ghanem, D., P. H. C. Sant'Anna, and K. Wüthrich (2023, June). Selection and parallel trends.

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

Hackl, F., M. E. Kummer, R. Winter-Ebmer, and C. Zulehner (2014, May). Market structure and market performance in E-commerce. *European Economic Review 68*, 199–218.

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.

Holmes, T. J. (2011). The diffusion of Wal-Mart and economies of density. *Econometrica : journal of the Econometric Society 79*(1), 253–302.

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

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.

Jeanjean, F. and G. V. Houngbonon (2017, March). Market structure and investment in the mobile industry. *Information Economics and Policy 38*, 12–22.

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., E. Fagundes, M. Carvalho, and R. C. Gomes (2022). Criminal enterprises: Evidence from rio de janeiro.

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. (2022). 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.

NBC (2020). Chicago's Top Cop Breaks Down Number of Gangs in City, How They're Structured – NBC Chicago. https://www.nbcchicago.com/news/local/chicagos-top-cop-breaks-down-number-of-gangs-in-city-how-theyre-structured/2309434/.

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.

Olivares, M. and G. P. Cachón (2009). Competing Retailers and Inventory: An Empirical Investigation of General Motors 7 Dealerships in Isolated U.S. Markets. *Management Science 55*(9), 1586–1604.

Olken, B. A. and P. Barron (2009, June). The simple economics of extortion: Evidence from trucking in aceh. *Journal of Political Economy 117*(3), 417–452.

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.

Prieto-Curiel, R., G. M. Campedelli, and A. Hope (2023, September). Reducing cartel recruitment is the only way to lower violence in Mexico. *Science 381*(6664), 1312–1316.

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.

Trejo, G. and S. Ley (2018). Why Did Drug Cartels Go to War in Mexico? Subnational Party Alternation, the Breakdown of Criminal Protection, and the Onset of Large-Scale Violence. *Comparative Political Studies 51*(7), 900–937.

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	Total DTOs
2006	✓	-	-	✓	✓	✓	✓	-	5
2007	✓	-	-	✓	✓	✓	✓	-	5
2008	✓	-	-	✓	✓	✓	✓	-	5
2009	✓	✓	-	✓	✓	✓	✓	-	6
2010	✓	✓	✓	✓	✓	✓	✓	✓	8
2011	✓	✓	✓	✓	✓	✓	✓	✓	8
2012	✓	✓	✓	✓	✓	✓	✓	✓	8
2013	✓	✓	✓	✓	✓	✓	✓	✓	8
2014	✓	✓	✓	✓	✓	✓	✓	✓	8
2015	✓	✓	✓	✓	✓	✓	✓	✓	8
2016	✓	✓	✓	✓	✓	✓	✓	✓	8
2017	✓	✓	✓	✓	✓	✓	✓	✓	8
2018	✓	✓	✓	✓	✓	✓	✓	✓	8

DTO Data on Lagged Distance Instrument by Year

Year	BLO	CABT	CDG	CDS	CJ	CJNG	FM	LZ	Total DTOs
2006	✓	-	✓	✓	✓	-	✓	-	5
2007	✓	-	✓	✓	✓	-	✓	-	5
2008	✓	-	✓	✓	✓	-	✓	-	5
2009	✓	-	✓	✓	✓	-	✓	-	5
2010	✓	-	✓	✓	✓	✓	✓	-	6
2011	✓	✓	✓	✓	✓	✓	✓	✓	8
2012	✓	✓	✓	✓	✓	✓	✓	✓	8
2013	✓	✓	✓	✓	✓	✓	✓	✓	8
2014	✓	✓	✓	✓	✓	✓	✓	✓	8
2015	✓	✓	✓	✓	✓	✓	✓	✓	8
2016	✓	✓	✓	✓	✓	✓	✓	✓	8
2017	✓	✓	✓	✓	✓	✓	✓	✓	8
2018	✓	✓	✓	✓	✓	✓	✓	✓	8

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

Table A2: TSLS Sensitivity to Controls

	(1) Homicides	(2) Homicides	(3) Homicides	(4) Dropout 9th	(5) Dropout 9th	(6) Dropout 9th
Group Count	8.97*	7.97***	8.99*	-0.011*	-0.0038**	-0.011*
	(4.98)	(1.64)	(5.00)	(0.0065)	(0.0018)	(0.0065)
Outcome Mean (levels)	8.03	8.03	8.03	0.03	0.03	0.03
# Instruments	8	8	8	8	8	8
N	30619.00	30619.00	30619.00	30619.00	30619.00	30619.00
Effective F-stat	29.45	60.48	29.44	29.45	60.48	29.44
Effective F-stat Cutoff 10%	13.24	14.71	13.25	13.24	14.71	13.25
Sargan-Hansen Test	20.56	18.76	20.57	26.56	28.30	26.43
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 A3: Dropout Effects for Any DTO. Alternative Homicide Transformations for Heterogeneity by Lagged Homicides

	Grade 1-5	Grade 6	Grade 7	Grade 8	Grade 9
	Panel A: Male 9th Grade Dropout Rate				
$\sqrt{\text{Lag Homicides}}$	-0.0016***	-0.0004	0.0001	-0.0005	-0.0002
	(0.0005)	(0.0007)	(0.0009)	(0.0008)	(0.0007)
IHS(Lag Homicides)	-0.0024***	-0.0010	0.0005	-0.0002	-0.0005
	(0.0006)	(0.0011)	(0.0014)	(0.0011)	(0.0011)
Lag Homicides (1000s)	-0.0176	-0.0031	-0.0164	-0.0189	-0.0033
	(0.0307)	(0.0317)	(0.0267)	(0.0242)	(0.0222)
	Panel B: Female 9th Grade Dropout Rate				
$\sqrt{\text{Lag Homicides}}$	-0.0017***	-0.0011*	-0.0006	-0.0004	-0.0002
	(0.0005)	(0.0007)	(0.0009)	(0.0007)	(0.0007)
IHS(Lag Homicides)	-0.0024***	-0.0014	-0.0003	-0.0007	-0.0002
	(0.0006)	(0.0010)	(0.0014)	(0.0010)	(0.0011)
Lag Homicides (1000s)	-0.0164	-0.0126	-0.0095	-0.0091	-0.0101
	(0.0313)	(0.0368)	(0.0263)	(0.0197)	(0.0217)
N	24634	24634	24634	24634	24634

Notes: Unit of observation is municipality-year for years 2006-2018. Standard errors from 1000 bootstrap replications.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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	4.70*	4.40*	0.27	4.11*	3.07*
	(2.65)	(2.40)	(0.28)	(2.39)	(1.73)
2 vs. 0	12.36*	11.70*	0.61	10.92*	8.10*
	(6.62)	(6.04)	(0.66)	(6.08)	(4.66)
3 vs. 0	21.28**	19.64**	1.56*	18.03**	13.21**
	(8.63)	(7.93)	(0.82)	(7.75)	(5.89)
4 vs. 0	28.84***	27.02***	1.67	24.27**	17.90**
	(10.73)	(9.86)	(1.02)	(9.49)	(7.21)
5 vs. 0	6.65	7.57	-0.80	6.01	9.87
	(20.47)	(18.50)	(2.12)	(17.00)	(11.34)
6 vs. 0	70.41***	63.78***	6.44***	55.95***	31.15***
	(20.04)	(18.30)	(2.02)	(17.31)	(12.02)
7 vs. 0	107.17***	98.59***	8.62**	86.99***	58.78***
	(32.91)	(29.50)	(3.77)	(28.08)	(19.50)
8 vs. 0	133.23***	120.17***	12.34***	109.05***	81.14***
	(38.03)	(34.49)	(3.74)	(32.47)	(22.17)
Mean	8.44	7.56	0.86	6.4	4.84
N	28953.0	28953.0	28953.0	28953.0	28953.0

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)
	Panel A: Male 9th Grade Dropout Rate				
1 vs. 0	0.01**	0.01**	0.01**	0.01**	0.01**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
2 vs. 0	0.02*	0.01	0.01	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
3 vs. 0	0.02*	0.02*	0.02*	0.02	0.02*
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
4 vs. 0	0.04**	0.03*	0.03**	0.03*	0.03**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
5 vs. 0	0.04**	0.04*	0.04*	0.03*	0.04*
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
6 vs. 0	0.05**	0.05**	0.05**	0.04*	0.05**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
7 vs. 0	0.06**	0.06**	0.06**	0.05**	0.05*
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
8 vs. 0	0.06**	0.06**	0.06**	0.06*	0.06*
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
	Panel B: Homicides				
1 vs. 0	4.70*	4.66*	4.90*	3.45	2.52
	(2.65)	(2.74)	(2.73)	(2.68)	(3.17)
2 vs. 0	12.36*	12.28*	12.66*	9.08	6.61
	(6.62)	(6.90)	(6.70)	(5.74)	(7.41)
3 vs. 0	21.28**	20.85**	21.75**	15.51*	11.26
	(8.63)	(8.83)	(8.71)	(8.13)	(9.67)
4 vs. 0	28.84***	28.88***	29.31***	22.18**	13.79
	(10.73)	(10.82)	(10.90)	(10.59)	(12.07)
5 vs. 0	6.65	6.84	5.37	0.08	-5.49
	(20.47)	(21.05)	(20.27)	(20.35)	(23.44)
6 vs. 0	70.41***	70.46***	69.38***	63.06***	47.85*
	(20.04)	(20.15)	(20.25)	(20.62)	(27.41)
7 vs. 0	107.17***	106.39***	108.37***	85.87***	72.92**
	(32.91)	(32.93)	(32.13)	(33.08)	(35.71)
8 vs. 0	133.23***	133.15***	136.30***	120.69**	62.47*
	(38.03)	(38.06)	(35.51)	(52.17)	(33.16)
# Muni. Types	5.0	5.0	5.0	5.0	5.0
Log Pop. Control	✓	✓	✓	-	✓
IHS Ag. Control	✓	✓	-	✓	✓
Dist.$_t$ Controls	-	✓	-	-	-
Small DTOs	-	-	-	-	✓
ATE Joint p-val	0.01	0.01	0.01	0.03	0.03
N	28953.0	28953.0	28959.0	29340.0	28947.0

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: Sensitivity to GFE Groups

	(1)	(2)	(3)	(4)
	Panel A: Male 9th Grade Dropout Rate			
1 vs. 0	0.012**	0.012**	0.013***	0.015***
	(0.005)	(0.005)	(0.005)	(0.005)
2 vs. 0	0.015*	0.010	0.011	0.013
	(0.009)	(0.009)	(0.009)	(0.010)
3 vs. 0	0.025*	0.018	0.022*	0.022*
	(0.013)	(0.012)	(0.013)	(0.013)
4 vs. 0	0.035**	0.030*	0.032**	0.036**
	(0.017)	(0.016)	(0.017)	(0.017)
5 vs. 0	0.041**	0.032*	0.032*	0.038*
	(0.020)	(0.019)	(0.019)	(0.021)
6 vs. 0	0.051**	0.037*	0.046**	0.046*
	(0.023)	(0.022)	(0.021)	(0.024)
7 vs. 0	0.063**	0.046*	0.052**	0.048*
	(0.027)	(0.025)	(0.026)	(0.028)
8 vs. 0	0.064**	0.051*	0.055*	0.054*
	(0.029)	(0.027)	(0.028)	(0.030)
	Panel B: Homicides			
1 vs. 0	4.70*	-0.68	-3.00	0.49
	(2.65)	(2.68)	(3.90)	(2.99)
2 vs. 0	12.36*	2.37	-1.98	3.46
	(6.62)	(5.67)	(7.26)	(5.72)
3 vs. 0	21.28**	9.39	-1.24	6.97
	(8.63)	(7.66)	(10.38)	(8.38)
4 vs. 0	28.84***	10.07	4.16	15.23
	(10.73)	(9.92)	(13.60)	(10.12)
5 vs. 0	6.65	-16.22	3.73	25.15
	(20.47)	(20.46)	(21.59)	(16.86)
6 vs. 0	70.41***	41.58**	5.25	35.53
	(20.04)	(17.20)	(25.97)	(24.92)
7 vs. 0	107.17***	78.89***	73.23**	80.82***
	(32.91)	(27.73)	(29.12)	(26.83)
8 vs. 0	133.23***	99.89***	100.87***	90.25***
	(38.03)	(28.63)	(31.20)	(26.77)
# Muni. Types	5.0	10.0	20.0	50.0
Log Pop. Control	✓	✓	✓	✓
IHS Ag. Control	✓	✓	✓	✓
ATE Joint p-val	0.01	0.0	0.0	0.0
N	28953.0	27639.0	27105.0	23951.0

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 A7: Descriptive Statistics by Treatment Risk

	At Risk	Not At Risk	Overall
Covariates			
Population (2005)	67,124.80	7,186.08	44,725.82
Poverty Rate (2005)	0.61	0.75	0.66
Literacy Rate (2005)	0.86	0.79	0.83
Elect. Users	14,373.24	-	14,373.24
kWH (millions)	86.64	-	86.64
kWH per Capita	6,418.51	-	6,418.51
Ag. Revenue	177.47	16.56	169.19
Total Police	75.03	-	75.03
Year	2,012.29	2,006.52	2,012
Outcomes			
Homicide Rate (per 100,000)	16.27	10.03	15.95
9th Grade Enrollment	875.81	206.11	841.37
9th Grade Dropout Rate	0.03	0.03	0.03
9th Grade Dropout Rate (Girls)	0.03	0.02	0.03
9th Grade Dropout Rate (Boys)	0.03	0.02	0.03
Treatments and Instruments			
Number of DTOs	0.85	0.01	0.81
$\geq$ 1 DTOs	0.31	0	0.30
Average Dist.$_{t-1}$ (km)	71.44	172.04	76.44
# Municipalities	2435	951	3386
N	30080	1575	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 16. 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.

Table A8: Testing for Non-linearities in Homicide Treatment Effects

	(1)	(2)	(3)
Intercept	-13.63	12.08	-9.36
	(10.13)	(12.34)	(23.05)
DTOs Linear	13.34***	-8.09	18.86
	(3.07)	(8.60)	(26.06)
DTOs Quadratic		2.86***	-5.24
		(1.11)	(7.48)
DTOs Cubic			0.66
			(0.60)
Polynomial Terms Wald Test P-Value	.	0.01	0.02
N	8.0	8.0	8.0

Notes: This table shows coefficients for regressions of the estimated Average Treatment Effects on polynomial terms of the number of DTOs. All the regressions are variance weighted according to the average treatment effects standard errors. Sample size is 8 since there are only 8 treatment effects. "Polynomial Terms Wald Test P-Value" reports the P-value for a Wald test on all the non-linear terms.

Table A9: Comparison to Linear Model, OLS and TSLS Results

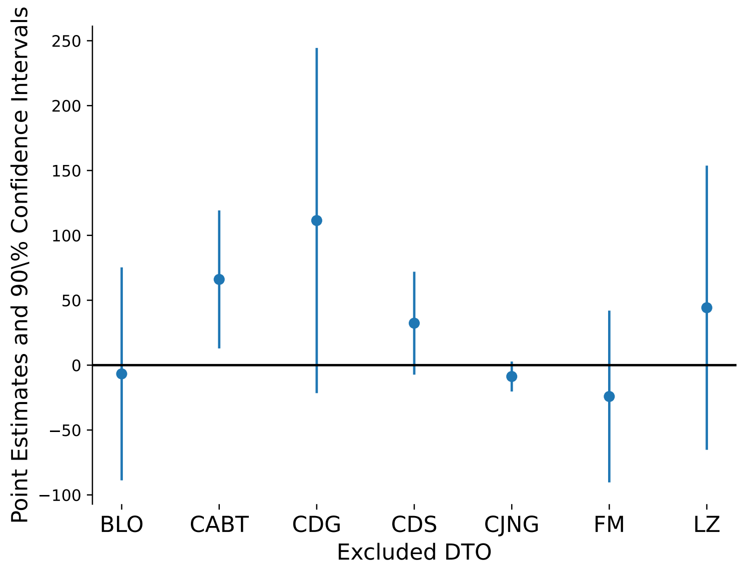
	(1) Main	(2) Linear	(3) OLS	(4) TSLS - Linear	(5) TSLS - Dummied	(6) TSLS - Dummied 5 DTOs
	Panel A: Male 9th Grade Dropout Rate					
1 vs. 0	0.012**	0.008**	-0.001	-0.010	-0.127**	-0.102**
	(0.005)	(0.004)	(0.001)	(0.008)	(0.051)	(0.042)
2 vs. 0	0.015*	0.015**	-0.002	-0.021	0.054	0.072
	(0.009)	(0.007)	(0.002)	(0.017)	(0.070)	(0.057)
3 vs. 0	0.025*	0.023**	-0.001	-0.031	-0.014	0.022
	(0.013)	(0.011)	(0.002)	(0.025)	(0.060)	(0.044)
4 vs. 0	0.035**	0.031**	0.001	-0.042	-0.232*	-0.138*
	(0.017)	(0.014)	(0.002)	(0.033)	(0.127)	(0.082)
5 vs. 0	0.041**	0.038**	-0.002	-0.052	0.066	-0.017
	(0.020)	(0.018)	(0.003)	(0.042)	(0.116)	(0.087)
6 vs. 0	0.051**	0.046**	-0.003	-0.063	-0.255	
	(0.023)	(0.021)	(0.003)	(0.050)	(0.246)	
7 vs. 0	0.063**	0.054**	-0.004	-0.073	-0.216	
	(0.027)	(0.025)	(0.004)	(0.058)	(0.297)	
8 vs. 0	0.064**	0.062**	-0.003	-0.084	-0.273	
	(0.029)	(0.028)	(0.004)	(0.066)	(0.240)	
	Panel B: Homicides					
1 vs. 0	4.70*	22.56***	0.50	9.93*	-40.02	-30.40
	(2.65)	(7.13)	(1.07)	(5.99)	(37.49)	(22.89)
2 vs. 0	12.36*	45.12***	3.28	19.87*	21.86	24.08
	(6.62)	(14.26)	(2.84)	(11.99)	(40.43)	(31.72)
3 vs. 0	21.28**	67.68***	2.07	29.80*	61.62	46.81
	(8.63)	(21.39)	(1.90)	(17.98)	(49.48)	(31.92)
4 vs. 0	28.84***	90.24***	5.33**	39.74*	128.80	103.51
	(10.73)	(28.52)	(2.41)	(23.97)	(100.13)	(62.97)
5 vs. 0	6.65	112.80***	5.54	49.67*	41.03	-3.61
	(20.47)	(35.65)	(3.87)	(29.97)	(96.83)	(53.97)
6 vs. 0	70.41***	135.36***	14.15***	59.60*	-312.69	
	(20.04)	(42.78)	(4.19)	(35.96)	(376.36)	
7 vs. 0	107.17***	157.92***	28.59***	69.54*	372.48	
	(32.91)	(49.90)	(6.36)	(41.96)	(346.79)	
8 vs. 0	133.23***	180.48***	64.57***	79.47*	-0.57	
	(38.03)	(57.03)	(14.69)	(47.95)	(175.24)	
N	28953	28953	30619	30619	30619	30619

Notes: This table shows estimated Average Treatment Effects. The first column is the main model used in the paper. Column 2 is a linear model using Equation 20 to model potential outcomes. Column 3 are OLS estimates. Column 4 are results from a linear TSLS model. Column 5 are results from the same linear TSLS model, but dummying out the endogenous variable. Column 6 limits the dummies so that 5 DTOs includes 5 DTOs or more, for increased power. Columns 1 and 2 use 1000 bootstrap replications for inference, column 3-6 uses municipality clustered standard errors. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

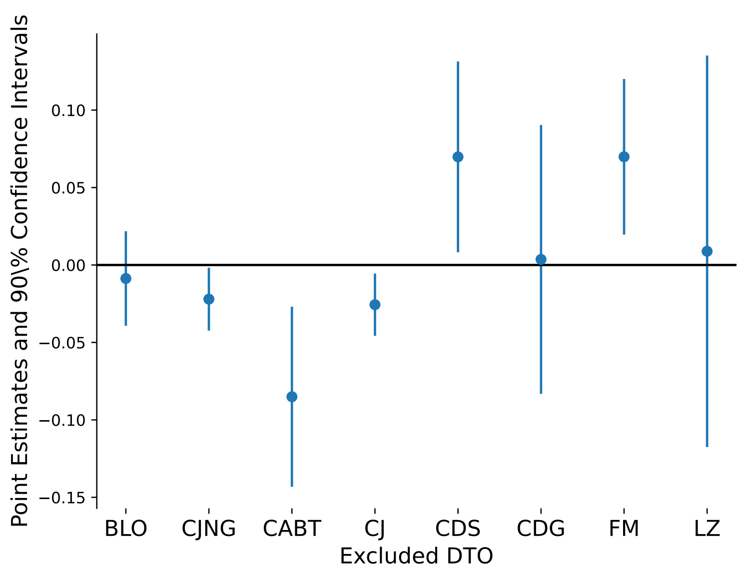
A.2 Figures

Figure A1: One at a Time TSLS estimates

(a) Homicides

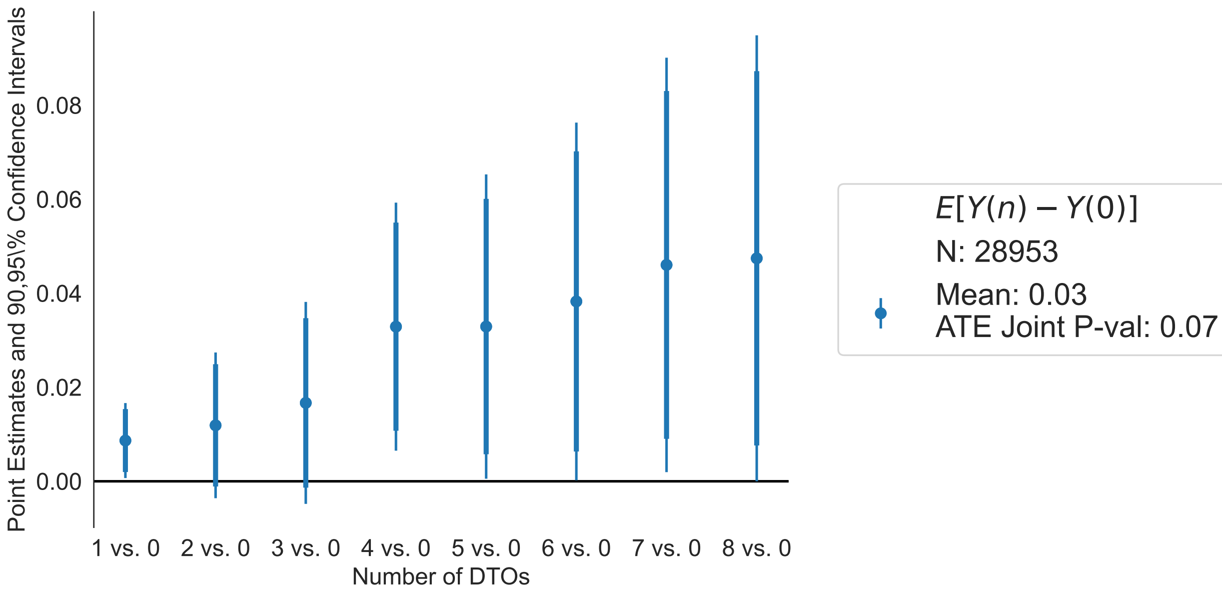


(b) 9th Grade Dropout



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.

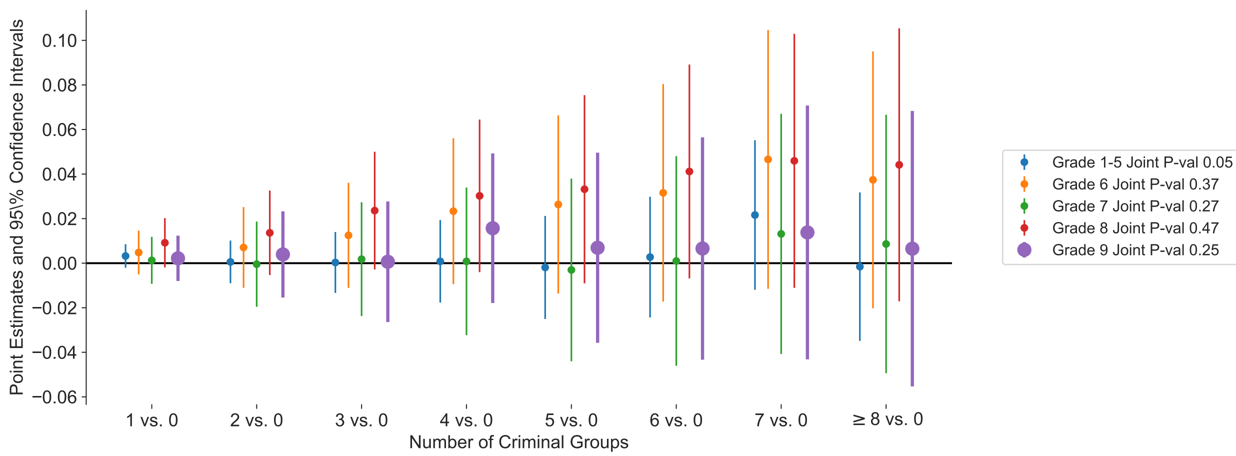
Figure A2: 9th Grade Dropout Rates All Genders



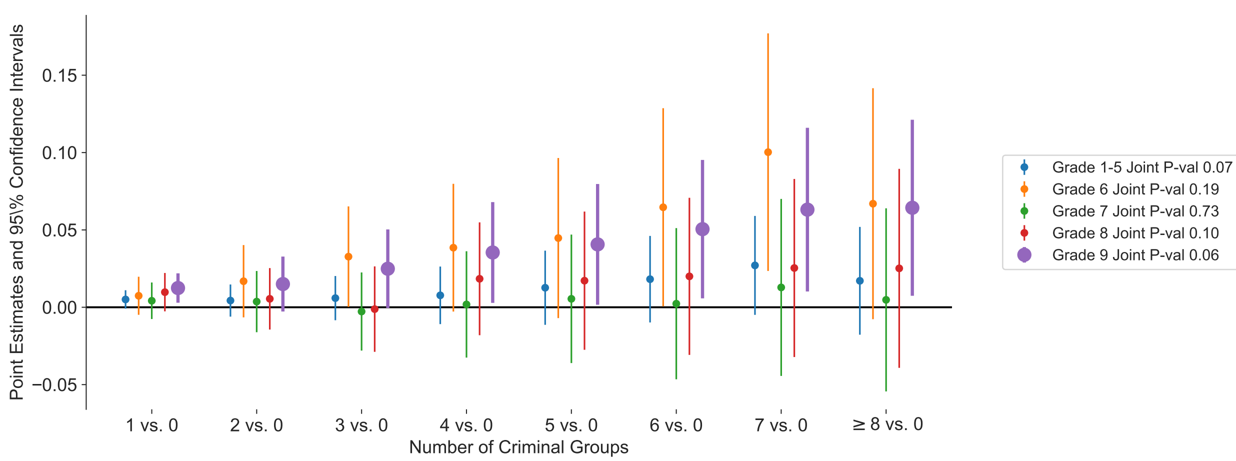
Notes: This figure plots Average Treatment Effect estimates, $E[Y(n) - Y(0)]$, using the normal model outlined in sections 6 and 7. Effects for 9th grade dropout rates, including both male and female students. 90 and 95 % confidence intervals shown. Standard errors from 1000 bootstrap replications.

Figure A3: Dropout by Grade and Gender

(a) Dropout for Female Students

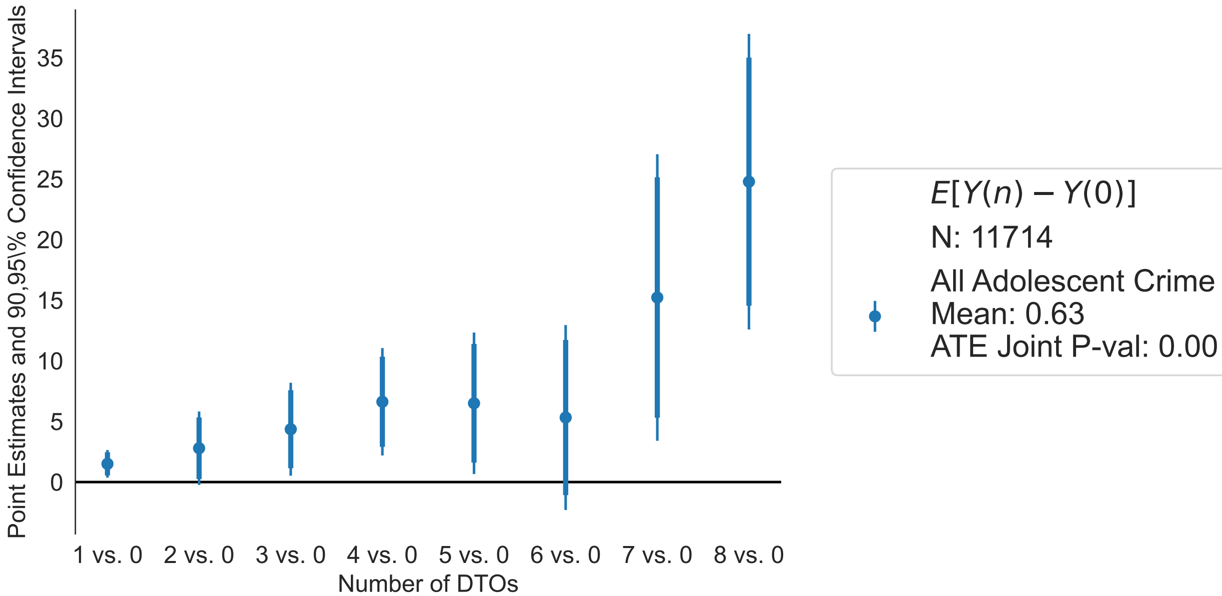


(b) Dropout for Male Students



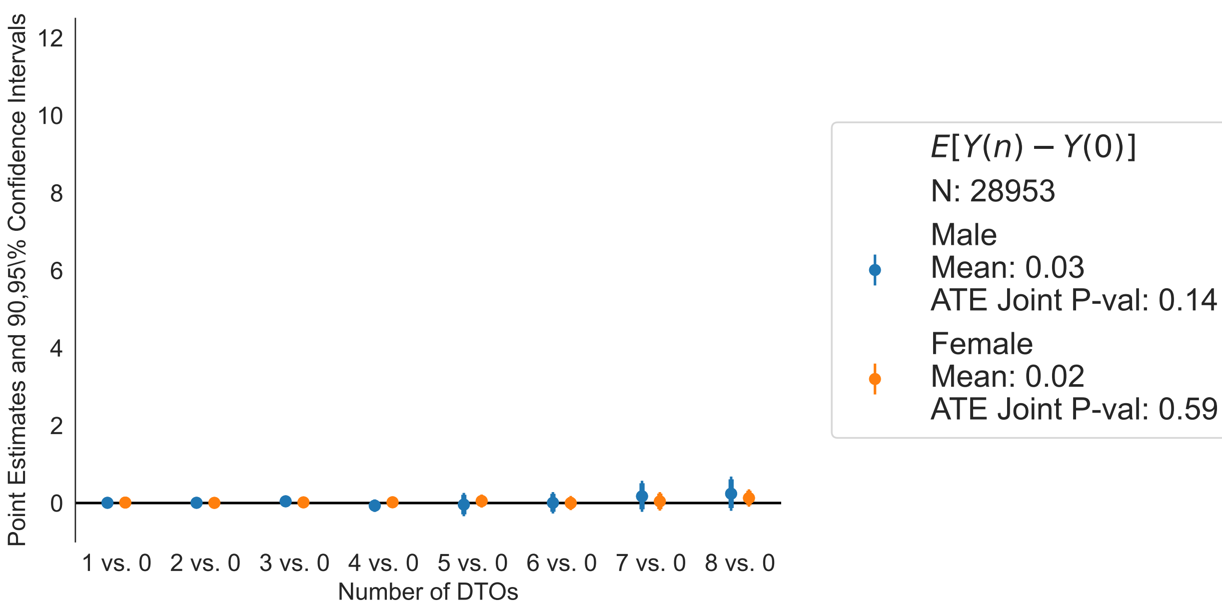
Notes: Standard errors from 1000 bootstrap replications.

Figure A4: Homicides, Victims Aged 12-14



Notes: This figure plots Average Treatment Effect estimates, $E[Y(n) - Y(0)]$, using the normal model outlined in sections 6 and 7. Outcome is all crimes with adolescent perpetrators. Sample years 2014-2018. 90 and 95 % confidence intervals shown. Standard errors from 1000 bootstrap replications.

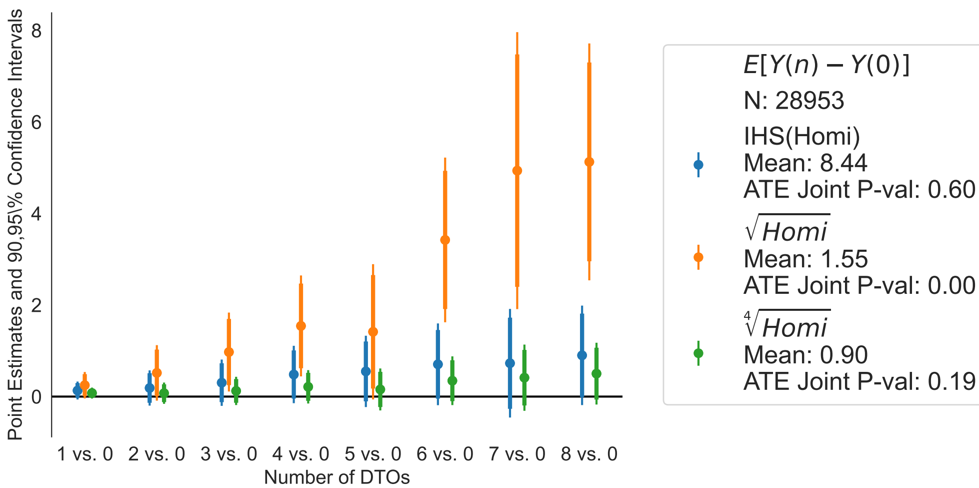
Figure A5: Homicides, Victims Aged 12-14



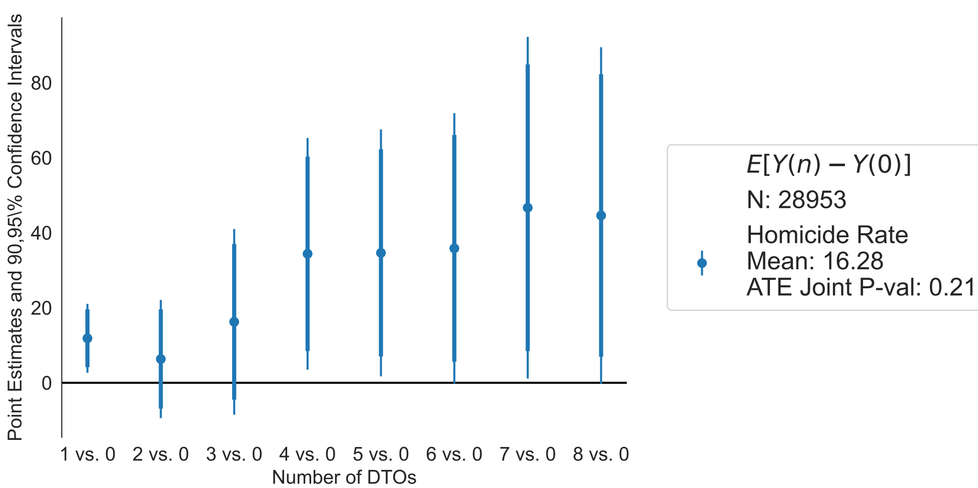
Notes: This figure plots Average Treatment Effect estimates, $E[Y(n) - Y(0)]$, using the normal model outlined in sections 6 and 7. Effects for homicides with 12-14 year old victims. 90 and 95 % confidence intervals shown. Standard errors from 1000 bootstrap replications.

Figure A6: Homicide Results

(a) Alternative Transformations for Homicides

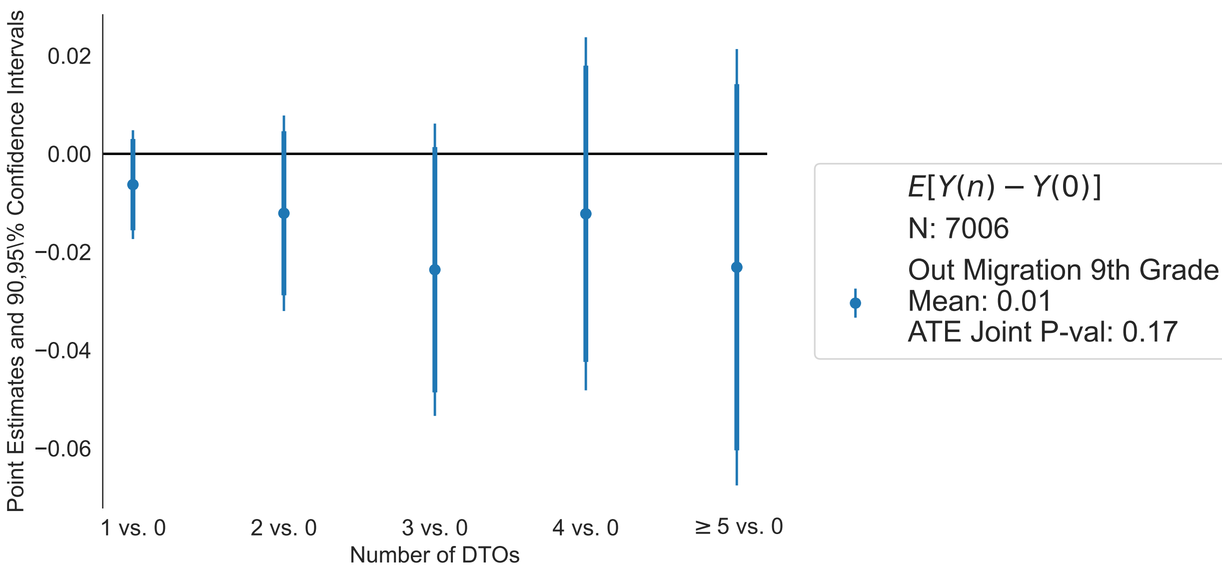


(b) Homicide Rates



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.

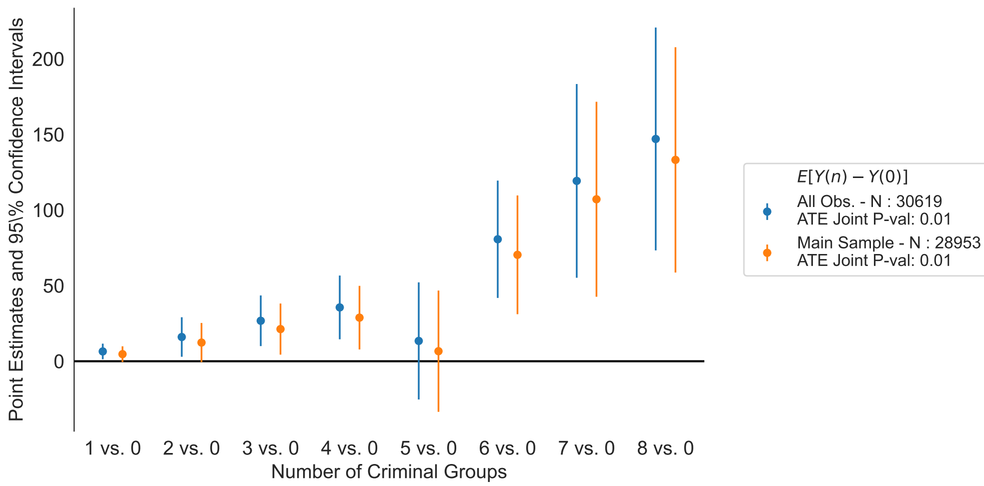
Figure A7: ENLACE Out Migration Results



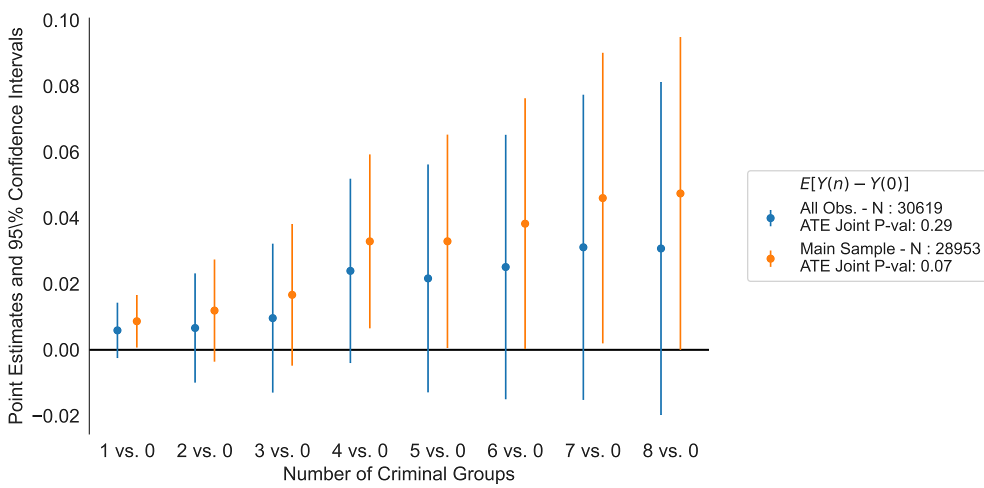
Notes: This figure displays estimates using outmigration as measured by ENLACE test data. Sample years are 2009-12. I limit the maximum number of DTOs to 3 to gain precision. Standard errors from 1000 bootstrap replications.

Figure A8: Comparison of Main Sample Results to Full Sample Results

(a) Homicides



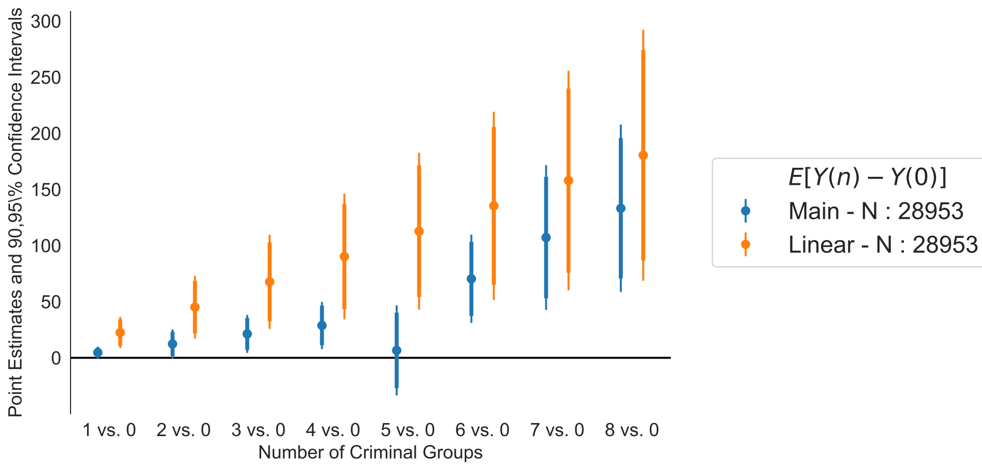
(b) 9th Grade Dropout Both Sexes



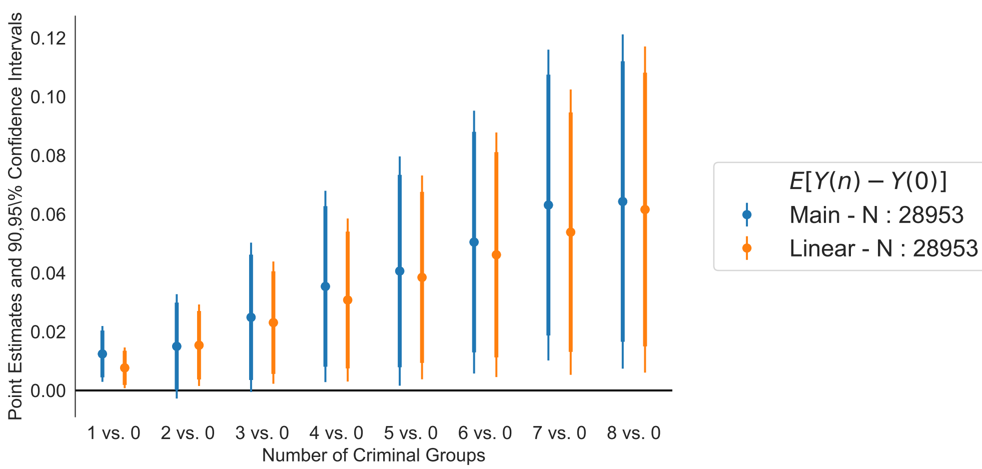
Notes: Unit of observation is municipality-year. Sample years are 2006-18. Panel (a) displays effects with homicides as outcomes, Panel (b) for 9th grade dropout for both males and females. This figure replicates the main estimates and compares them to estimates that include all observations. The main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

Figure A9: Comparison to Linear Potential Outcome Model

(a) Homicides



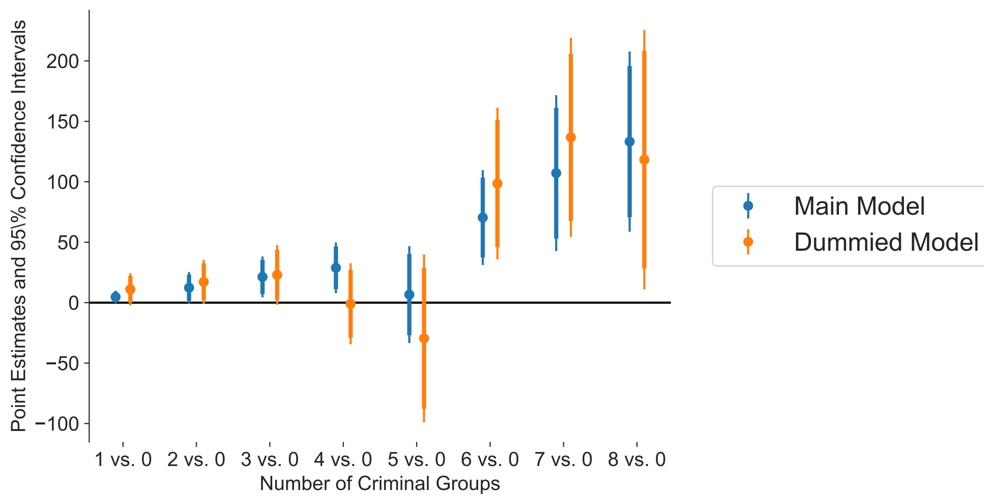
(b) Male 9th Grade Dropout



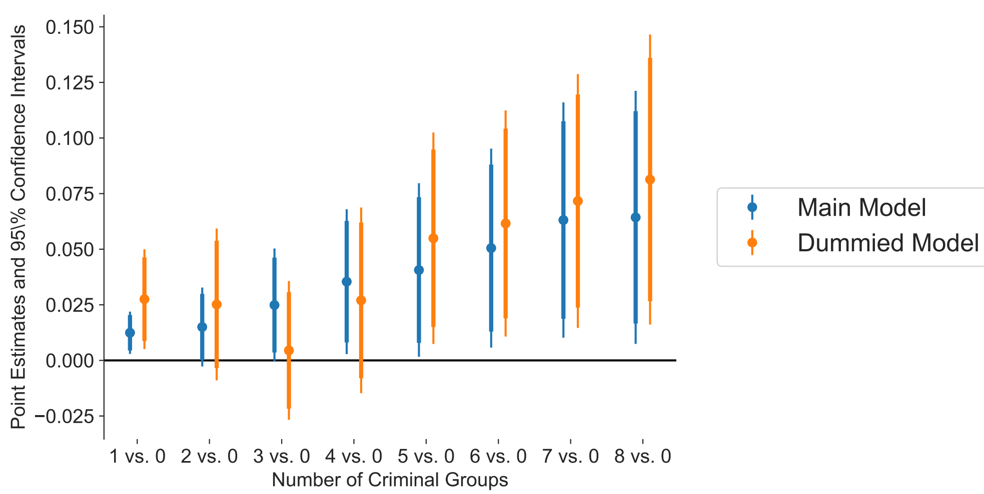
Notes: Sample years are 2006-18. Panel (a) displays effects with homicides as outcomes, Panel (b) for 9th grade dropout for males. This figure replicates the main estimates and compares them to similar estimates from a linear model. The linear model replicates the main estimation procudure, but models potential outcomes as linear function of the number of DTOs, instead of allowing for a separate intercept for each number of DTOs. Both models control for log population, IHS of agricultural production, dummy indicators for lag presence for the 8 DTO groups, and the lag DTO count. Standard errors from 1000 bootstrap replications.

Figure A10: Comparison to Different First Stage Models

(a) Homicides



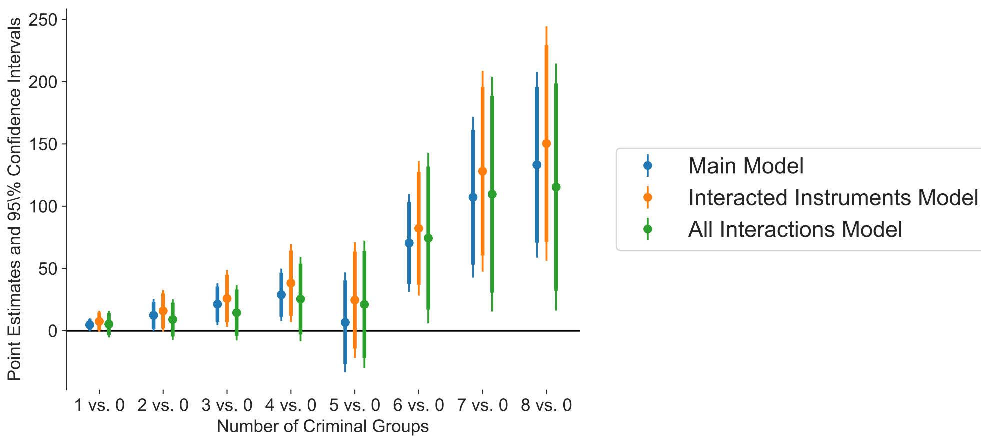
(b) Male 9th Grade Dropout



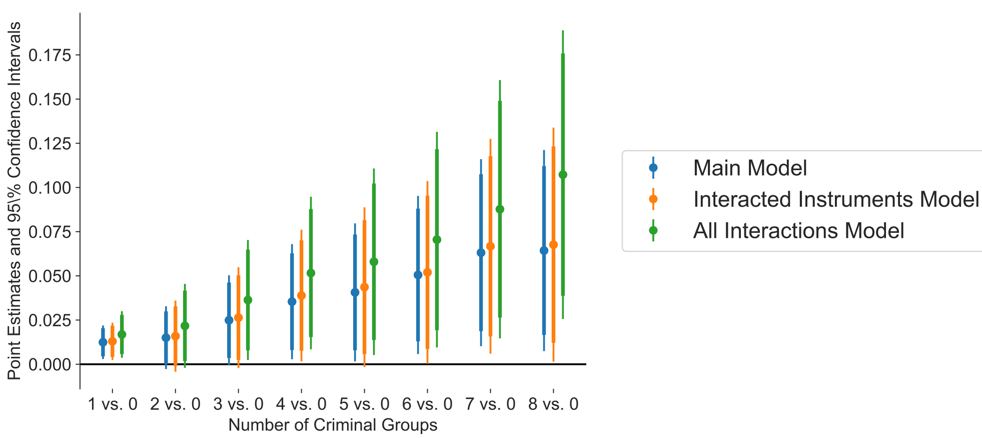
Notes: Sample years are 2006-18. Standard errors from 1000 bootstrap replications.

Figure A11: Robustness to Multiple Equilibria in First Stage

(a) Homicides



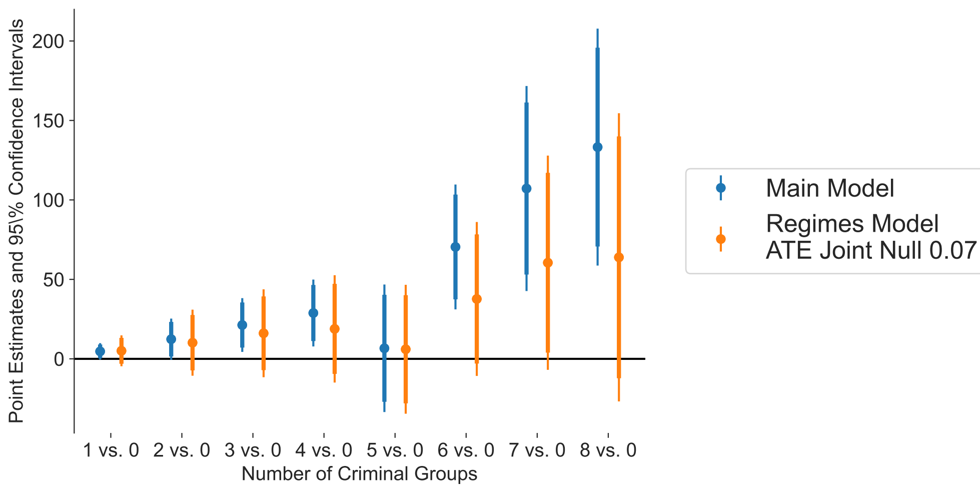
(b) Male 9th Grade Dropout



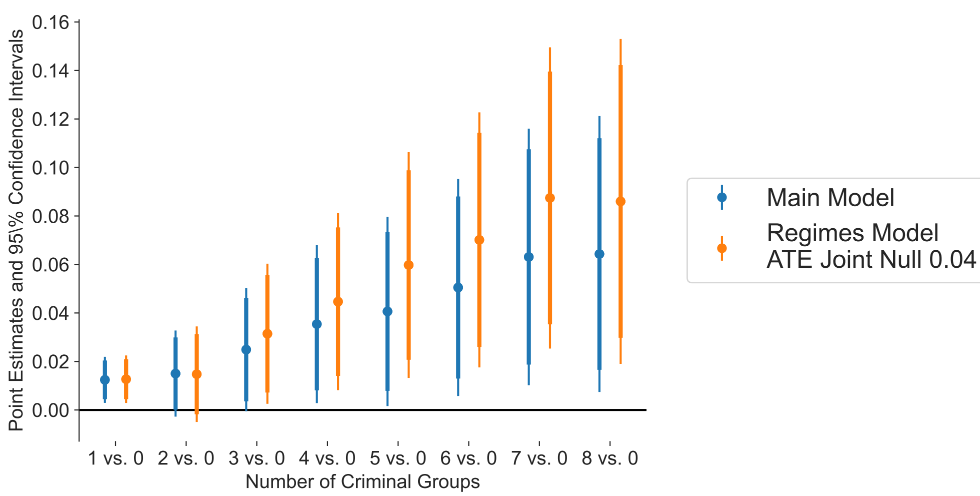
Notes: Sample years are 2006-18. Standard errors from 1000 bootstrap replications.

Figure A12: Robustness to Non-stationarity

(a) Homicides



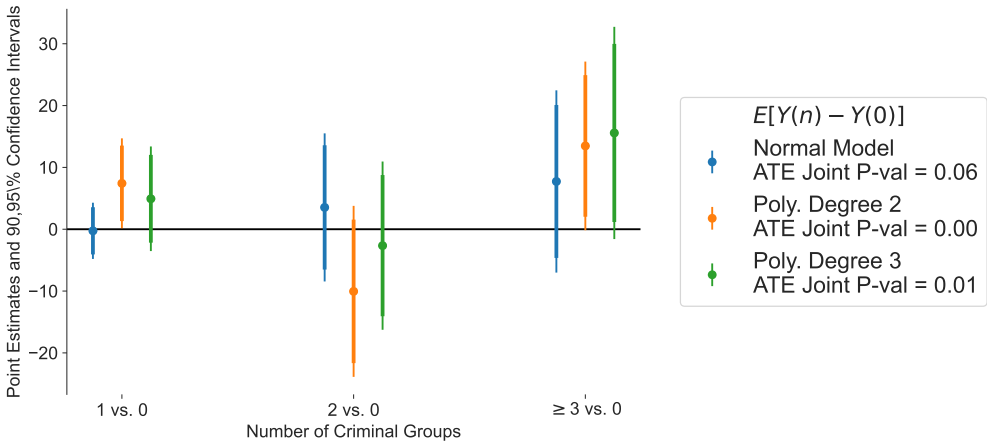
(b) Male 9th Grade Dropout



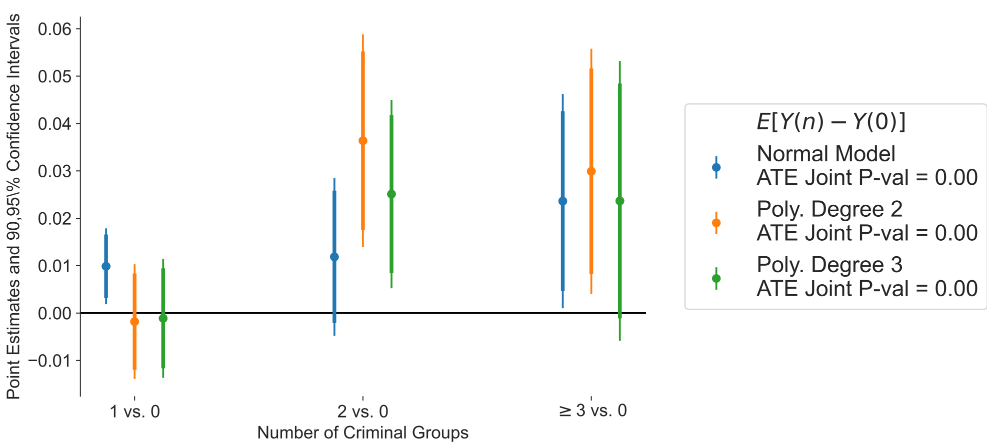
Notes: Sample years are 2006-18. Standard errors from 1000 bootstrap replications.

Figure A13: Comparison to Polynomial Controls, Maximum 3 DTOs

(a) Homicides



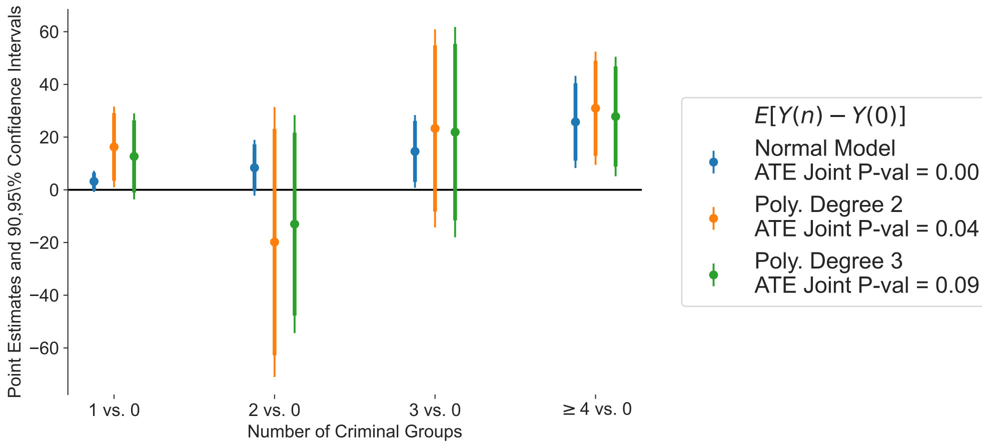
(b) Male 9th Grade Dropout



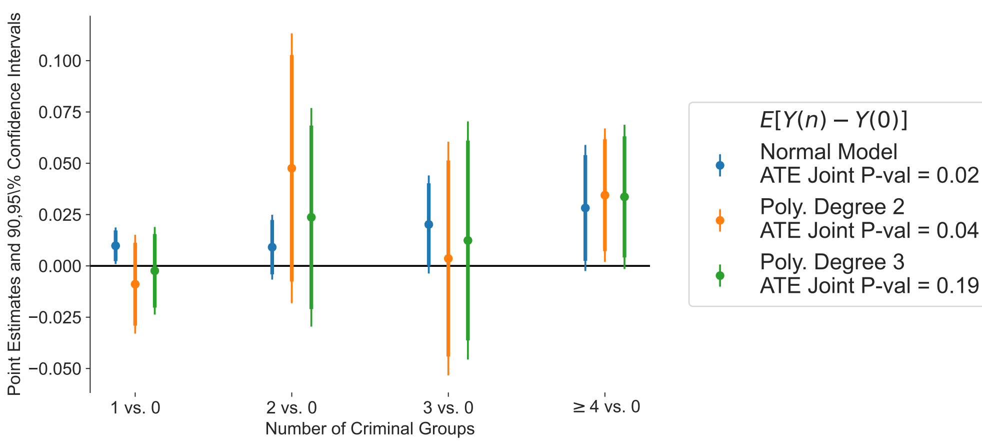
Notes: Sample years are 2006-18. Standard errors from 1000 bootstrap replications.

Figure A14: Comparison to Polynomial Controls, Maximum 4 DTOs

(a) Homicides



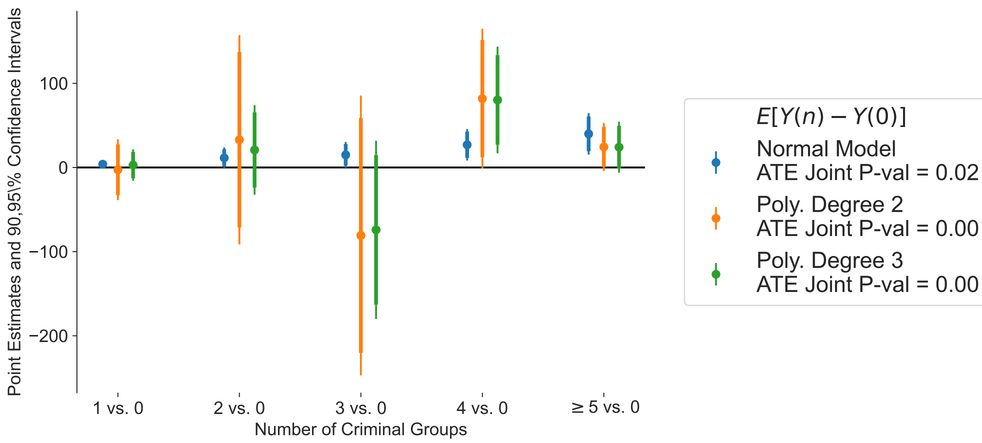
(b) Male 9th Grade Dropout



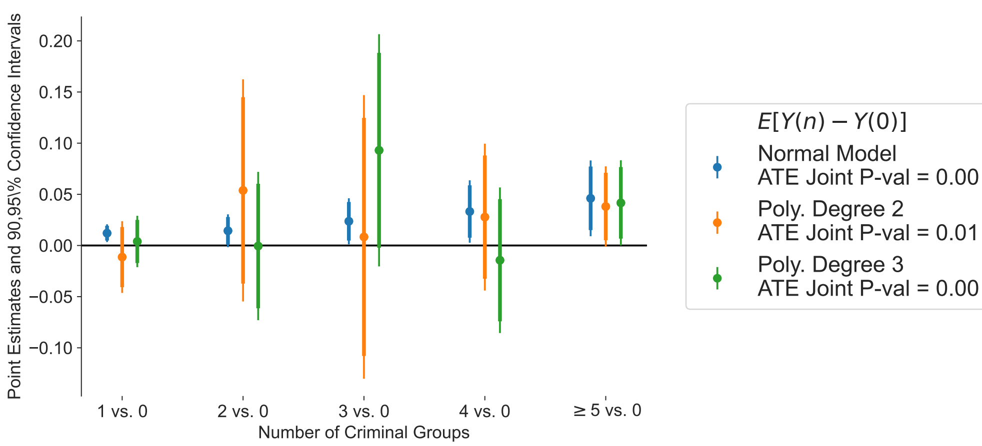
Notes: Sample years are 2006-18. Standard errors from 1000 bootstrap replications.

Figure A15: Comparison to Polynomial Controls, Maximum 5 DTOs

(a) Homicides

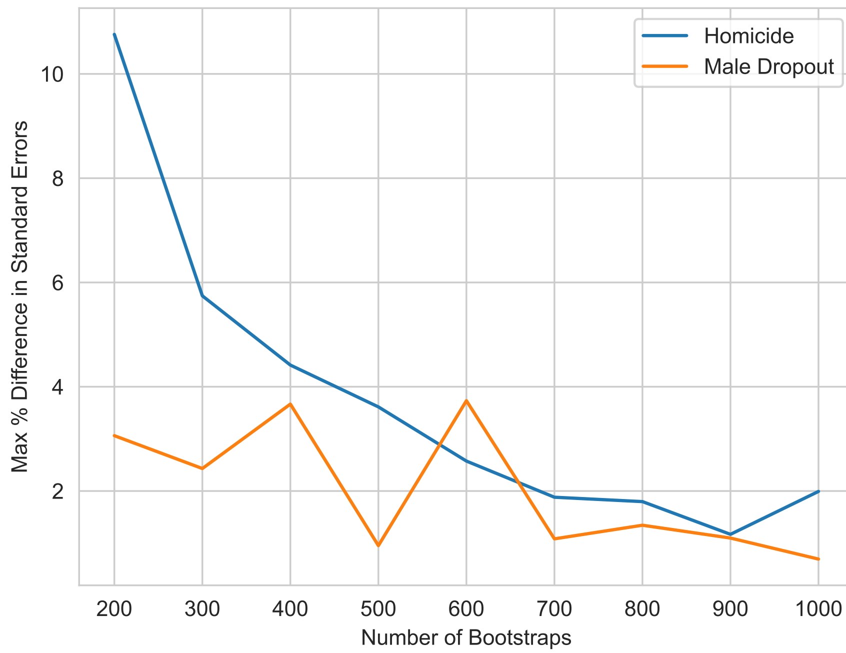


(b) Male 9th Grade Dropout



Notes: Sample years are 2006-18. Standard errors from 1000 bootstrap replications.

Figure A16: Standard Errors With Fewer Bootstraps



Notes: This figure examines how the standard errors evolve with the number of bootstraps. I first compute standard errors based on 100 bootstraps, focusing on the estimates for the 8 average treatment effects. I then calculate standard errors using 100 more bootstraps, shown on the x-axis. I compare the standard errors to the errors using 100 fewer bootstraps, calculate the percentage difference, and take the maximum discrepency across the 8 coefficients. The figure shows results for homicides and male dropout outcomes. With 1000 bootstraps, the standard errors vary less than 2% with 100 extra bootstraps.

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:

$$\text{Homicides}_{mt} = \alpha_m + \tau_t + \delta \text{Log Population}_{mt} + \sum_{i=1}^{8} \beta_i 1[N_{mt} = i] + u_{mt} \tag{25}$$

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	0.783	0.705
1	0.089	0.098
2	0.055	0.073
3	0.030	0.045
4	0.018	0.028
5	0.011	0.018
6	0.006	0.012
7	0.004	0.010
8	0.004	0.010

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	69.44	(63.41)	87.03	(73.43)
CABT	68.74	(70.1)	80.68	(80.19)
CDG	58.30	(53.57)	77.58	(66.3)
CDS	51.63	(47.31)	66.52	(53.67)
CJ	105.74	(74.28)	155.90	(121.35)
CJNG	67.54	(71.29)	78.63	(76.0)
FM	97.18	(122.98)	131.86	(154.39)
LZ	34.76	(30.68)	45.31	(39.82)
Average	76.44	(57.37)	98.95	(65.54)

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	30959.0	30959.0

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 25. 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	0.01**	0.04***
	(0.00)	(0.01)
2 vs. 0	0.01	-0.01
	(0.01)	(0.04)
3 vs. 0	0.02	0.04
	(0.01)	(0.03)
$\geq$ 4 vs. 0	0.03**	0.05***
	(0.01)	(0.02)
	Panel B: Homicides	
1 vs. 0	3.18	30.79**
	(2.05)	(13.84)
2 vs. 0	8.37	-31.85
	(5.40)	(32.92)
3 vs. 0	14.57**	40.66
	(7.04)	(42.31)
$\geq$ 4 vs. 0	25.76***	74.11***
	(8.92)	(17.83)
N	27860.0	28635.0
Mean	5.31	6.98

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 one of the displayed treatment doses, 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 15 and the estimation equation 19.

$$
\begin{aligned}
&E[Y_{mt}|\mathbf{Q_m}(\tilde{\mathbf{Z}}_{mt-1}, W_{mt}) = q, W_{mt}, N_{mt}(q) = n] \\
&\qquad = E[Y_{mt}(n)|\mathbf{Q_m}(\tilde{\mathbf{Z}}_{mt-1}, W_{mt}) = q, W_{mt}, N_{mt}(q) = n] \\
&\qquad = \alpha_n + \psi(W_{mt}) + E[U_{mt}(n)|\mathbf{Q_m}(\tilde{\mathbf{Z}}_{mt-1}, W_{mt}) = q, W_{mt}, N_{mt}(q) = n] \\
&\qquad = \alpha_n + \psi(W_{mt}) + E[U_{mt}(n)|W_{mt}, N_{mt}(q) = n] && \text{(Assumptions SS1 FS3)} \\
&\qquad = \alpha_n + \psi(W_{mt}) + E[U_{mt}(n)|N_{mt}(q) = n] && \text{(Assumption SS2)} \\
&\qquad = \alpha_n + \psi(W_{mt}) + \sum_{d \in \mathcal{N}(n)} E[U_{mt}(n)|D_{mt}(q) = d, N_{mt}(q) = n] P[D_{mt}(q) = d|N_{mt}(q) = n] \\
&\qquad = \alpha_n + \psi(W_{mt}) + \sum_{d \in \mathcal{N}(n)} E[U_{mt}(n)|D_{mt}(q) = d] P[D_{mt}(q) = d|N_{mt}(q) = n]
\end{aligned}
$$

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

$$
\begin{aligned}
&E[Y_{mt}|\mathbf{Q_m}(\tilde{\mathbf{Z}}_{mt-1}, W_{mt}) = q, W_{mt}] \\
&\quad = Y_{mt}(0) + \sum_{n=1}^{G} E[Y_{mt}(n) - Y_{mt}(0)|\mathbf{Q_m}(\tilde{\mathbf{Z}}_{mt-1}, W_{mt}) = q, W_{mt}] \\
&\quad = \alpha_0 + \psi(W_{mt}) + E[U_{mt}(0)|\mathbf{Q_m}(\tilde{\mathbf{Z}}_{mt-1}, W_{mt}) = q, W_{mt}] \\
&\quad + \sum_{n=1}^{G} P[N_{mt}(q) = n|\mathbf{Q_m}(\tilde{\mathbf{Z}}_{mt-1}, W_{mt}) = q, W_{mt}]\{\alpha_n - \alpha_0\} \\
&\quad + P[N_{mt}(q) = n|\mathbf{Q_m}(\tilde{\mathbf{Z}}_{mt-1}, W_{mt}) = q, W_{mt}]E[U_{mt}(n) - U_{mt}(0)|\mathbf{Q_m}(\tilde{\mathbf{Z}}_{mt-1}, W_{mt}) = q, W_{mt}, N_{mt}(q) = \\
&\qquad \text{(Iterate Expectations)} \\
&\quad = \alpha_0 + \psi(W_{mt}) + \sum_{n=1}^{G} P[N_{mt}(q) = n|W_{mt}]\{\alpha_n - \alpha_0\} \\
&\quad + P[N_{mt}(q) = n|W_{mt}]E[U_{mt}(n) - U_{mt}(0)|\mathbf{Q_m}(\tilde{\mathbf{Z}}_{mt-1}, W_{mt}) = q, W_{mt}, N_{mt}(q) = n] \\
&\qquad \text{(Assumptions SS1 FS3 and } E[U_{mt}(0)] = 0) \\
&\quad = \alpha_0 + \psi(W_{mt}) + \sum_{n=1}^{G} P[N_{mt}(q) = n]\{\alpha_n - \alpha_0\} \\
&\quad + P[N_{mt}(q) = n|W_{mt}] \sum_{d \in \mathcal{N}(n)} E[U_{mt}(n)|D_{mt}(q) = d]P[D_{mt}(q) = d|N_{mt}(q) = n, W_{mt}] \\
&\quad = \alpha_0 + \psi(W_{mt}) + \sum_{n=1}^{G} P[N_{mt}(q) = n|W_{mt}]\{\alpha_n - \alpha_0\} \\
&\quad + P[N_{mt}(q) = n|W_{mt}] \sum_{d \in \mathcal{N}(n)} E[U_{mt}(n)|D_{mt}(q) = d]\frac{P[D_{mt}(q) = d \cap N_{mt}(q) = n|W_{mt}]}{P[N_{mt}(q) = n|W_{mt}]} \\
&\quad = \alpha_0 + \psi(W_{mt}) + \sum_{n=1}^{G} P[N_{mt}(q) = n|W_{mt}]\{\alpha_n - \alpha_0\} \\
&\quad + P[N_{mt}(q) = n|W_{mt}] \sum_{d \in \mathcal{N}(n)} E[U_{mt}(n)|D_{mt}(q) = d]\frac{P[D_{mt}(q) = d|W_{mt}]}{P[N_{mt}(q) = n|W_{mt}]} \\
&\qquad (D_{mt}(q) = d \text{ necessarily yields n}) \\
&\quad = \alpha_0 + \psi(W_{mt}) + \sum_{n=1}^{G} P[N_{mt}(q) = n|W_{mt}]\{\alpha_n - \alpha_0\} \\
&\quad + \sum_{d \in \mathcal{N}(n)} E[U_{mt}(n)|D_{mt}(q) = d]P[D_{mt}(q) = d|W_{mt}]
\end{aligned}
$$

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 Dynamic Entry Game Assumptions

The key assumption for the control function approach used in this paper is Assumption 2, which states DTOs entry decisions admit a threshold crossing representation. In this section, I describe sufficient conditions that yield this threshold crossing model. In particular, I describe a Markov game where DTOs are allowed to be forward-looking, yet we still obtain threshold crossing representation in Assumption 2. In this section, I closely follow the expositions in (Aguirregabiria and Mira, 2007b; Aguirregabiria et al., 2021).

D.1 Assumptions and Notation

First I define notation. Recall there are $1...T$ time periods, $1...M$ municipalities $1...G$ DTOs. At every time period t, DTOs decide simultaneously which municipalities to enter. Group g's entry decision for m, t is given by D_{gmt}. Each DTO chooses D_{gmt} optimally, according to the criteria I define next.

At the beginning of each period t, firms profits are affected by observable covariates $(\tilde{\mathbf{Z}}_{mt-1}, W_{mt})$ and a vector of unobservables $\tilde{\epsilon}_{gmt}$. In line with Aguirregabiria and Mira (2007b), $\tilde{\epsilon}_{gmt}$ contains two components, one with no entry and another with entry, represented with $\tilde{\epsilon}_{gmt} = (\tilde{\epsilon}_{gmt}(0), \tilde{\epsilon}_{gmt}(1))$. Recall $\tilde{\mathbf{Z}}_{mt-1} = \{\tilde{Z}_{gmt-1}\}_{g=1}^{G}$ is the vector of all DTOs group for the last year. W_{mt} contains 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 collected in the vector $W_{mt} = (R_m, L_t, X_{mt}, \{D_{gmt-1}\}_{g=1}^{g})$. $(\tilde{\mathbf{Z}}_{mt-1}, W_{mt})$ are common knowledge to all groups at the beginning of t. Additionally, let $\tilde{\epsilon}_{mt} = \{\tilde{\epsilon}_{gmt}\}_{g=1}^{G}$ be the vector of all groups unobservables. I assume $(\tilde{\mathbf{Z}}_{mt-1}, W_{mt}, \tilde{\epsilon}_{mt})$ follows a controlled Markov process with transition probability $p(\tilde{\mathbf{Z}}_{mt-1}, W_{mt+1}, \tilde{\epsilon}_{mt+1} | \tilde{\mathbf{Z}}_{mt-1}, W_{mt}, \tilde{\epsilon}_{mt}, \mathbf{D}_{mt})$. $\mathbf{D}_{mt}$ is the vector of all groups' entry decisions, $\mathbf{D}_{mt} = \{D_{gmt}\}_{g=1}^{G}$

Let $\tilde{\Pi}_g(\mathbf{D}_{mt}, \tilde{\mathbf{Z}}_{mt-1}, W_{mt}, \epsilon_{gmt})$ be group g's flow profits. Notice flow profits for group g can be affected by the entire vector of entry $\mathbf{D}_{mt}$, allowing that DTO competition can affect profits. Since DTOs are allowed to be forward looking, their goal is to choose entry decisions D_{gmt} to maximize expected discounted profits:

$$E\left\{\sum_{s=t}^{\infty}\beta_g^{s-t}\Pi_g(\mathbf{D}_{ms},\tilde{\mathbf{Z}}_{ms-1},W_{ms},\tilde{\epsilon}_{gms})\mid\tilde{\mathbf{Z}}_{mt-1},W_{mt},\tilde{\epsilon}_{gmt}\right\} \tag{26}$$

Where $\beta_g \in (0,1)$ is a discount factor. The primitives of the model are the profit functions $\{\Pi_g\}_{g=1}^{G}$, the transition probability $p(\cdot|\cdot)$ and the discount factor β_g. Notice that the DTO groups do not treat $\tilde{Z}_{gmt}$ as a control variable to maximize profits. They take the distances as given. This is in line with assumption FS3 and with the decentralized nature of DTOs. I assume DTOs are not solving the full problem of choosing a network and internalizing the benefits due to $\tilde{Z}_{gmt}$. Instead, they take these distances as given.

I now add the following assumptions on the model primitives:

Assumption 4 (Additively Separable Flow Profit)**.**
Private information appears additively separably in the flow profit function. Additionally, $\tilde{\mathbf{Z}}_{mt-1}$ only affects group g's profits through $\tilde{Z}_{gmt-1}$ so rival distances are irrelevant for profits. Furthermore this component is also additively separable. Put together:
$\tilde{\Pi}_g(\mathbf{D}_{mt},\tilde{\mathbf{Z}}_{mt-1},W_{mt},\epsilon_{gmt}) = \Pi_g(\mathbf{D}_{mt},W_{mt}) + c(\tilde{Z}_{gmt-1}) + \tilde{\epsilon}_{gmt}(D_{gmt})$*, where* Π_g *is a real-valued function and* $\tilde{\epsilon}_{gmt}(D_{gmt})$ *is the D'th component of* $\tilde{\epsilon}_{gmt}$

Assumption 5 (Conditional Independence)**.**
Transition probability $p(\cdot|\cdot)$ factors as
$p(\tilde{\mathbf{Z}}_{mt-1},W_{mt+1},\tilde{\epsilon}_{mt+1}|\tilde{\mathbf{Z}}_{mt-1},W_{mt},\tilde{\epsilon}_{mt},\mathbf{D}_{mt}) = p(\tilde{\epsilon}_{mt+1}|W_{mt})p(\tilde{\mathbf{Z}}_{mt-1},W_{mt+1}|\tilde{\mathbf{Z}}_{mt-1},W_{mt},\mathbf{D}_{mt})$.
That is, given the groups decisions at t, private information variables $\tilde{\epsilon}_{mt}$ do not affect the transitions of the observed covariates, and $\tilde{\epsilon}_{mt}$ are iid across municipalities and time conditional on W_{mt}

Assumption 6 (Normalization of Flow Profits)**.**
If group g does not enter, their observable flow profits are normalized to zero,
$\Pi_g(0,W_{mt}) + c(\tilde{Z}_{gmt-1}) = 0$.

Assumption 7 (Known Discount Factor)**.**

β_g is known to the researcher.

Under the assumptions, and those in Assumption 1, the entry game is identified. That is, $\{\Pi_g\}_{g=1}^{G}$ and the transition probabilities $p(\cdot|\cdot)$ are identified.

Assumption 4 serves three purposes. **First**, it separates out the unobservable component of profits as an additively separable term, $\tilde{\epsilon}_{gmt}(D_{gmt})$. Without this separability, there is no hope to attain a threshold crossing representation. **Second**, it assumes a certain type of exclusion restriction. The rival group distances for g, which are in the vector $\tilde{\mathbf{Z}}_{mt-1}$, do not affect group g's flow profits. That is, only $\tilde{Z}_{gmt-1}$ matters for group g's flow profits. Pesendorfer and Schmidt-Dengler (2008) show an exclusion restriction of this kind is needed for identification - flow profits have to depend on group-specific variables that in turn do not affect competitors' flow profits. In principle, groups' incumbency indicators D_{gmt-1} could also fulfill this role, however I focus on the instruments $\tilde{Z}_{gmt-1}$ given their importance in my application. **Third**, the separate term $c(\tilde{Z}_{gmt-1})$ captures the idea that $\tilde{Z}_{gmt-1}$ matters insofar as it changes fixed operating costs for DTOs. It does not interact with the terms in $\Pi_g(\mathbf{D}_{mt}, W_{mt})$ to alter flow profits. This is in line with the role the instrument $\tilde{Z}_{gmt-1}$ fulfills in my analysis: it alters the probability of a group entering, but it does not affect their operations or other choices.

Assumption 5 places structure on the evolution of the variables. First, it sets $\tilde{\epsilon}_{mt} \mid W_{mt}$ to be i.i.d. across municipalities and time periods, as in assumption FS1. This aspect is redundant given my Assumption FS1, but this representation is common in dynamic discrete choice problems, hence I keep it for familiarity. The more substantive implication of Assumption 5 is that once we control for groups actions D_{mt}, the unobserved variables $\tilde{\epsilon}_{mt}$ do not affect the evolution of observable variables. That is, $\tilde{\epsilon}_{mt}$ is only allowed to influence group decisions at time t, and it does not affect future time periods.

D.2 Markov Strategies

This game has a Markov structure and I assume groups play Markov strategies. That is, if $(\tilde{\mathbf{Z}}_{mt-1}, W_{mt}, \tilde{\epsilon}_{mt}) = (\tilde{\mathbf{Z}}_{ms-1}, W_{ms}, \tilde{\epsilon}_{ms})$ then group g's decisions will be identical in t, s. This is an important restriction. Within each municipality, groups must play the same equilibrium over time. However, groups can play different equilibria across municipalities. Let $\sigma = \{\sigma_g(\tilde{\mathbf{Z}}_{mt-1}, W_{mt}, \tilde{\epsilon}_{gmt})\}_{g=1}^{G}$ be a set of of strategy functions or decision rules for each group. Note that the strategies depend on all groups distances $\tilde{\mathbf{Z}}_{mt-1}$, seemingly at odds with Assumption 4. However there is no inconsistency here. While flow profits depend on DTO specific distances, DTOs observe their rivals distances and can base their strategy on rivals' distances. Associated with these strategy functions σ are conditional choice probabilities for entry, i.e.

$$P_g^{\sigma}(D_{gmt} \mid \tilde{\mathbf{Z}}_{mt-1}, W_{mt}) := P[\sigma_g(\tilde{\mathbf{Z}}_{mt-1}, W_{mt}) = D_{gmt}] \tag{27}$$

Let $\pi_g^{\sigma}(D_{gmt}, \tilde{\mathbf{Z}}_{mt-1}, W_{mt})$ be g's expected profit if it chooses D_{gmt} and rival groups play according to σ. By the independence of the error terms across groups, we have

$$\begin{aligned} &\pi_g^{\sigma}(D_{gmt}, \tilde{\mathbf{Z}}_{mt-1}, W_{mt}) = \\ &\sum_{D_{-gmt} \in \mathcal{D}^{G-1}} \left(\prod_{j \neq g} P_j^{\sigma}(D_{-gmt}[j] \mid \tilde{\mathbf{Z}}_{mt-1}, W_{mt}) \right) \{\Pi_g(D_{gmt}, D_{-gmt}, W_{mt}) + c(\tilde{Z}_{gmt-1})\} \end{aligned} \tag{28}$$

where D_{-gmt} is $G-1$ vector the the decisions of all groups except g and where $D_{-gmt}[j]$ is the j'th component of the vector. This expression is simply the expected flow profit of D_{gmt} if rival groups follow strategies σ.

With this notation in hand, we can rewrite lifetime profits in Equation 26 using Bellman's

Principle:

$$
\begin{aligned}
&\tilde{V}_g^{\sigma}(\tilde{\mathbf{Z}}_{mt-1}, W_{mt}, \tilde{\epsilon}_{gmt}) = \\
&\max_{D_{gmt}\in\{0,1\}} \Big\{ \pi_g^{\sigma}(D_{gmt}, \tilde{\mathbf{Z}}_{mt-1}, W_{mt}) + \tilde{\epsilon}_{gmt}(D_{gmt}) \\
&+\beta_g \int\int \tilde{V}_g^{\sigma}(\tilde{\mathbf{Z}}_{mt-1}, W_{mt+1}, \tilde{\epsilon}_{gmt+1}) f_g^{\sigma}(\tilde{\mathbf{Z}}_{mt-1}, W_{mt+1} \mid \tilde{\mathbf{Z}}_{mt-1}, W_{mt}, D_{gmt}) d\tilde{\mathbf{Z}}_{mt-1} dW_{mt+1} \Big\}
\end{aligned} \tag{29}
$$

where $f_g^{\sigma}(\tilde{\mathbf{Z}}_{mt-1}, W_{mt+1} \mid \tilde{\mathbf{Z}}_{mt-1}, W_{mt}, D_{gmt})$ is the transition probability of $\tilde{\mathbf{Z}}_{mt-1}, W_{mt+1}$ conditional on group g choosing D_{gmt} and all other groups behaving according to σ. As an additional piece of notation, consider the integrated version of Equation 29 with respect to $\tilde{\epsilon}$, $\tilde{V}_g^{\sigma}(\tilde{\mathbf{Z}}_{mt-1}, W_{mt}) = \int \tilde{V}_g^{\sigma}(\tilde{\mathbf{Z}}_{mt-1}, W_{mt}, \tilde{\epsilon}_{gmt}) f(\tilde{\epsilon}_{gmt} \mid W_{mt}) d\tilde{\epsilon}_{gmt}$. These simple representations are due to Assumption 5, which rules out any effect of $\tilde{\epsilon}_{gmt}$ on future transitions.

We are now in a position to express the entry probabilities in the threshold crossing representation of Assumption 2. Note that since $D_{gmt} \in \{0,1\}$ entry will occur if the lifetime profits of entering are larger than the lifetime profits of not entering. We can rewrite the lifetime profits of choosing $D_{gmt} \in \{0,1\}$ as

$$
\pi_g^{\sigma}(D_{gmt}, \tilde{\mathbf{Z}}_{mt-1}, W_{mt}) + \tilde{\epsilon}_{gmt}(D_{gmt}) + \beta_g E\left\{ \tilde{V}_g^{\sigma}(\tilde{\mathbf{Z}}_{mt-1}, W_{mt+1}) \mid D_{gmt}, \tilde{\mathbf{Z}}_{mt-1}, W_{mt} \right\} \tag{30}
$$

$$
= \underbrace{\pi_g^{\sigma}(D_{gmt}, \tilde{\mathbf{Z}}_{mt-1}, W_{mt}) + \beta_g E\left\{ \tilde{V}_g^{\sigma}(\tilde{\mathbf{Z}}_{mt-1}, W_{mt+1}) \mid D_{gmt}, \tilde{\mathbf{Z}}_{mt-1}, W_{mt} \right\}}_{\text{choice-specific value functions}:=\tilde{v}_g^{\sigma}(D_{gmt}, \tilde{\mathbf{Z}}_{mt-1}, W_{mt})} + \tilde{\epsilon}_{gmt}(D_{gmt}) \tag{31}
$$

$$
= \tilde{v}_g^{\sigma}(D_{gmt}, \tilde{\mathbf{Z}}_{mt-1}, W_{mt}) + \tilde{\epsilon}_{gmt}(D_{gmt}) \tag{32}
$$

Where $\tilde{v}_g^{\sigma}(D_{gmt}, \tilde{\mathbf{Z}}_{mt-1}, W_{mt})$ is simply the observable lifetime profits of choosing D_{gmt} given strategies σ, which is commonly called the *choice specific value function*. Using this notation, the decision rule for entry is given by:

$$D_{gmt} = 1\left[\tilde{v}_g^\sigma(1, \tilde{\mathbf{Z}}_{mt-1}, W_{mt}) + \tilde{\epsilon}_{gmt}(1) > \tilde{v}_g^\sigma(0, \tilde{\mathbf{Z}}_{mt-1}, W_{mt}) + \tilde{\epsilon}_{gmt}(0)\right] \tag{33}$$

$$D_{gmt} = 1\left[\underbrace{\tilde{v}_g^\sigma(1, \tilde{\mathbf{Z}}_{mt-1}, W_{mt}) - \tilde{v}_g^\sigma(0, \tilde{\mathbf{Z}}_{mt-1}, W_{mt})}_{v_g^\sigma(\tilde{\mathbf{Z}}_{mt-1}, W_{mt})} > \underbrace{\tilde{\epsilon}_{gmt}(0) - \tilde{\epsilon}_{gmt}(1)}_{\epsilon_{gmt}}\right] \tag{34}$$

$$D_{gmt} = 1\left[v_g^\sigma(\tilde{\mathbf{Z}}_{mt-1}, W_{mt}) > \epsilon_{gmt}\right] \tag{35}$$

Which is essentially the representation in Assumption 2. Assumption 2 additionally allows v_g to vary by m with v_{gm}. This is technically redundant given that W_{mt} contains municipality indicators and can already vary by m. However, Assumption 2 makes this more explicit and stresses that groups can have different decision rules across m. Nonetheless, we are still missing the equilibrium concept, which I explain next

D.3 Equilibrium Concept

Given the previous notation, we can define equilibrium behavior. So far, the strategies σ have been arbitrary. The following definition characterizes groups' strategies as best-responses to one another.

Definition D.1. A stationary Markov Perfect Equilibrium (MPE) in this game is a set of strategy functions σ^* such that for any group and value of covariates

$$\sigma_g^*(\tilde{\mathbf{Z}}_{mt-1}, W_{mt}, \tilde{\epsilon}_{gmt}) = \arg \max_{D_{gmt}\in\{0,1\}} \left\{\tilde{v}_g^\sigma(D_{gmt}, \tilde{\mathbf{Z}}_{mt-1}, W_{mt}) + \tilde{\epsilon}_{gmt}(D_{gmt})\right\} \tag{36}$$

That is, σ_g^* is a best response given other groups' strategies in σ^*. In turn, equilibrium entry decisions are given by

$$D_{gmt} = 1\left[v_g^{\sigma^*}(\tilde{\mathbf{Z}}_{mt-1}, W_{mt}) > \epsilon_{gmt}\right] \tag{37}$$

where I used the representation in Equation 35.

Overall, the equilibrium strategy supposes a stable Markov environment where the actions and observables $(\tilde{\mathbf{Z}}_{mt-1}, W_{mt})$ are in a stationary process. This might be too strong an assumption in my setting. For instance, some of the DTO groups came were formed during my sample years (listed in Table A1). As I explain in Section 8.6.5, I investigate how sensitive my results are to significant differences in DTO behavior across time. I find that the results are qualitatively similar when we allow DTO entry decisions to be different for different time periods. This is reassuring, since it suggests the main estimates, which assume a stationary environment, are a good enough approximation for practical purposes.

Furthermore, notice the ancillary role played by the discount factor β_g in the analysis. Entry game models assume this factor is known and common to all groups to identify the underlying parameters of the flow profits $\tilde{\Pi}_g$. However, since my estimation routine does not compute the underlying structural parameters of profit, the exact value of β_g is largely irrelevant since it does not change the reduced form entry probability given by $v_g^{\sigma^*}(\tilde{\mathbf{Z}}_{mt-1}, W_{mt})$. This reduced form is simply a function of observed covariates $(\tilde{\mathbf{Z}}_{mt-1}, W_{mt})$. Additionally since my first stage model does allow for DTO specific effects, I can allow the β_g to vary by group as well. Overall, this implies the groups can be myopic with $\beta_g = 0$ or value the future with $\beta_g \approx 1$ and my empirical strategy is still sound.

D.4 Municipality and Time Fixed Effects

Finally, one note on the municipality and time dummies in W_{mt}. Because of the expectations DTOs take before entering, accounting for municipality or time specific effects is not trivial. Section 7.1 explains how I account for municipality fixed effects in greater detail. However, my main specification includes time fixed effects, as in Equation 16. Formally, these time effects need to be rationalized within the previous model. The time effects are complicated since they may potentially inform DTOs' expectations about the future. For instance, if the time shocks are correlated, then a positive time effect in the current period would alter expectations for the next period, which in turn would complicate estimation. As a result, to be consistent with the previous dynamic model, I assume these time effects are independent and identically distributed.

Though strong, this assumption seems like a good compromise for two reasons. First, as discussed before, the results are overall not sensitive to allowing for different entry patterns across time, as shown in Section 8.6.5. Second, the main reason to include the time dummies is to transparently include them in the second stage as well. My goal is to purge the second stage estimate of any common shocks which affected all the municipalities in each year. Including time effects in the first and second stage is the simplest way to achieve that goal.