Criminal Competition:
How market structure determines violence in illegal sectors[‡]

Davide Zufacchi
University College London, IFS[§]

Isaia Sales
Università Sour Orsola Benincasa

Salvatore Leotta
Direzione Centrale Servizi Antidroga

Guido Persico
Polizia Scientifica

Luciano Lombardi
Polizia di Stato

February 2025

Preliminary version, please do not circulate

Abstract

This paper examines how market structure determines incentives for violence in the retail drug sector. Using geolocalized intelligence data, we document that areas with larger gangs experience lower levels of violence. To explain this regularity, we develop a quantitative model of an oligopoly market in which gangs decide whether to engange in conflict, trading off the expected additional profits from a higher market share in the area with the cost deriving from a higher presence of police. We estimate the model by combining newly collected data on quality and quantity of drugs seized, fights between gangs, and different measures of police attention. Counterfactual scenarios show that, due to an heterogenous police reaction, larger gangs face a diseconomy of scale in the fighting process, generating the negative correlation we observe in the data. The optimal police reaction function is then determined.

JEL Codes: D74, K14, K42, L13

[‡]We are grateful for suggestions and support to Maurizio Agricola, Alfedo Fabbrocini, Libera D'Angelo, Alessandro Giuliano, and Giovanni Leuci. We are indebted to Eliana La Ferrara, Joao Granja, Imran Rasul, Tommaso Sonno, and Gabriel Ulyssea for the guidance and support throughout this project. We are grateful for their valuable comments and suggestions to Jeremy Boccanfuso, Javier Boncompte, Anne Brockmeyer, Marco Castelluccio, Mathieu Couttenier, Magdalena Dominguez, Alessandro Gavazza, Rachel Griffith, Jonas Hjort, Lars Nesheim, Aureo de Paula, Nicola Persico, Lavinia Piemontese, Vincenzo Scrutinio, and Liyang Sun. We are grateful for the comments of participants in seminars at the University College London, and the Institute of Fiscal Studies. We wish to thank the Laboratory for Effective Anti-Poverty Policies (LEAP - Bocconi) for funding. Luca Matarazzo provided excellent research assistance. All errors remain our own.

[§]Corresponding author: Department of Economics, University College London and IFS, davide.zufacchi.20@ucl.ac.uk

1 Introduction

Criminal groups generated an estimated $3.6-$4.8 trillion in revenues during 2015/2016, approximately 7% of global GDP. Illegal drug trafficking alone accounts for more than $500 billion in annual revenues, a figure larger than the global smartphone industry's total revenue.[1] This massive illegal market creates powerful incentives for violence, with drug trafficking linked to approximately half of all homicides in the European Union, imposing substantial societal costs.[2] As a result, studying the functioning of this market and the violence it generates has always been a classical area of investigation for economists and other social scientists (Schelling, 1971, Levitt & Venkatesh, 2000, Becker et al., 2006). This usually involves studying both the strategic interaction among gangs in the market and their relationship with law enforcement. However, the covert nature of the market, and the consequential data limitations, have so far limited our ability to study these interactions.

This paper examines the relationship between market structure and violence leveraging a unique dataset on the retail drug market in Naples province, Italy. The horizontal structure with multiple independent gangs makes it an ideal setting to explore this relationship. Using geolocalized intelligence data on gang presence, we document a negative correlation between gang size and violence.[3] To explore the mechanisms behind this correlation, we develop a quantitative model of an oligopoly market where gangs decide whether to engage in conflict, balancing the expected additional profits from a higher market share against the cost of increased police presence. We estimate this model using newly collected data on drug seizures, inter-gang conflicts, and various measures of police activity. Our findings indicate that larger gangs encounter a diseconomy of scale in conflict dynamics due to a stronger police reaction, which explains the negative correlation observed in the data.

This study draws on an original dataset on the retail drug market in Naples province from 2015 to 2022, constructed by combining intelligence, administrative, and survey data from multiple sources. The primary data source is a set of matrices showing gangs presence in this market

[1] Source for total revenues: estimate by Global Initiative against Transnational Organized Crime (Global Initiative). Source for estimate drug trafficking revenues: Kar and Spanjers (2017). The global revenue in the "Smartphones" segment of the consumer electronics market was $450 billion in 2019. Source: Statista.

[2] Source for the fact on drug related homicides: Eurojust on europol data. EU cost of responding to illegal drugs was approximate 34 billion euros in 2013 (EMCDDA, 2013), with the direct cost of their homicides alone estimated to be between 500 million euros and 5 billions, depeding on the statistical value of life (authors calculations on data from EU - Policy deparment, 2013 on gang related homices in EU).

[3] This negative relationship has been hypothetised by mulitiple sholars in all the sciences studying criminal groups. From Schelling (1971) in Economics, to Reuter (1985) in Criminology, Volkov (1999) in Antropology, Catino (2019) in Sociology.

across areas and years. This is enriched with a comprehensive list of fights between gangs from investigations, including information about the involved parties, year, contested area, and the winner. To measure violence and police activity, we use intelligence data on homicides, attempted homicides, and arrests for different crimes. To study fighting incentives in the retail drug market, we use confidential administrative data on drug seizures from the antidrug authority, containing geolocalized information on approximately 14,000 seizures during the sample period, along with a random sample of chemical analyses of the seized drugs to measure their quality. The dataset is completed with intelligence data on average retail and wholesale prices, together with survey data on consumer and area characteristics. Throught the paper, we measure the dimension of a gang with the number of areas in which it is active (gang *scale*), and define a gang to be large if its scale is higher than 3–the 90th percentile.[4]

The data reveals three key patterns in the relationship between market structure and violence. First, violence increases with the number of active gangs in an area. Second, consistent with classical theoretical conflict literature (e.g., Morgenthau, 1973; Reed, 2003), we find an inverse-U shaped relationship between violence and the standard deviation of gang size in the area. Violence is reduced both when power in an area is balanced–low standard deviation, making conflict outcomes too unpredictable–and when it is extremely unbalanced–high standard deviation, making conflict outcomes too predictable. Third, there is a negative relationship between violence and the average size of gangs in the area. The probability of homicide is particularly low in areas with no gangs. It then increases dramatically, reaching up to 35% in areas with only small active gangs (where the mean scale is equal to 1), before beginning to decrease, reaching its lowest point at 20% in the fifth quintile of the average scale distribution. These patterns holds across different measures of violence (probability of homicide, probability of gang-related homicides, probability of fight, number of homicides, number of homicides per capita), and persists when considering the maximum scale of gang present in the area, rather than its mean, to control for variations in the number of active gangs. This evidence suggests that areas with larger gangs experience lower levels of violence.

We then explore the variation driving this correlation. First, we examine its robustness to the inclusion of area and year fixed effects, to account for unobserved differences across areas and systematic selection in gang presence.[5] Motivated by evidence previously described, we control

[4] We use the terms scale, size, dimension independetly in this paper. We are aware that when we talk about "size" of firms, we usually refer to the number of employees, revenues, or other type of measures. However, the number of areas is the only measure of size we can directly observe from the data. Moreover, this measure is often correlated to more common ones of firm size which, unfortunately, are unobserved.

[5] An "area" is a city in the province of Naples, and a district in the city of Naples. Naples has 30 districts.

for the number of active gangs in the area and the standard deviation of their scale. The negative correlation remains robust to these tests. We then investigate how variation in gang size, due to the entry of new gangs, creates marginal changes in violence using an event-study design. The entry of a new gang in the area is associated with an approximately 11% increase in the probability of homicide during the first period, compared to the pre-entry period, and a 20% increase in the second period. This average effect masks substantial heterogeneity based on gang size. The increase in violence is more pronounced when a small gang enters the area (15% and 24%). Conversely, the entry of a large gang is associated with a decrease in the probability of homicide of approximately 15% in both periods. As before, this evidence is tobust to different measures of violence (probability of homicide, number of homicides, probability of gang-related homicides, and probability of fights). Therefore, we conclude that gangs of different sizes exhibit different behaviors, potentially reflecting variations in their propensity for violence.

To understand the underlying mechanisms, we study fighting incentives in the retail drug market (their primary activity, accounting for more than 50% of estimated revenues–Eurispes, 2005). We present three facts about this market. First, consumers value drug quality and dislike police presence. Second, larger gangs have a higher probability of winning fights. Third, police response to violence varies with the scale of gangs involved. Specifically, police respond more strongly to fights involving larger gangs. These facts suggest several potential mechanisms behind the negative correlation between gang size and violence: (i) the *heterogeneous police response* may generate diseconomies of scale for large gangs in the fighting process; (ii) *violence spillovers* may contribute to higher fighting costs for gangs present in multiple areas[6]; and (iii) the higher probability of victory of larger gangs may *deter* smaller ones from initiating fights with them.

To test these potential mechanisms we build a quantitative model of the retail drug market. Uninformed consumers decide where to search for drugs, considering transport costs, signals on drug quality distribution, and average police presence in the area (fact 1 above). Gangs sell vertically differentiated drugs and are heterogeneous in their initial scale. They solve a static two-stage game. In the second stage, conditional on the realized market structure, gangs simultaneously choose drug quality to maximize profits, given average prices, production function, and demand. In the first stage, gangs decide whether, where, and against whom to fight. Building on the seminal paper by Donohue and Levitt (1998), we construct a step fighting mechanism with an observable component–the fighting ability–and an unobservable component, which introduces randomness in fight outcomes. Consistent with the presented evidence, we allow gangs of dif-

[6]We find that violence generated by fights, along with the resulting police attention, spread in all the areas in which the involved gangs are present, not only in the area interested by the fight.

ferent sizes to have heterogeneous fighting abilities, which influence their probability of winning fights (fact 2). When deciding whether to fight, gangs trade off the expected additional profits from gaining a larger market share against costs stemming from increased police presence. In line with fact 3, we allow for heterogeneous police reactions, which affect how police respond to fights involving gangs of different scales.

We use newly collected intelligence and administrative data on the retail drug market to estimate the model in four steps. First, we estimate the elasticity of police attention to fights with an event-study design. To identify the true change in police attention, rather than a change in the underlying illegal activity, we use arrests by the local police–which is likely independent from drug or gang dynamics–as a measure of police attention. To deal with the endogeneity of fighting choice, we use an IV: releases of gang members from prisons. Second, we estimate demand parameters by exploiting variation in seized drug quantities across areas, estimating the across-nests log-linear demand function. The main parameter in this second step is the consumer disutility from police presence, which will then determine the market cost of fights. This parameter is particularly challenging to identify, because any change in police attention has two effects on the amount of drug seized: (i) the direct effect on drugs sold in the area–the elasticity we want to identify, and (ii) the indirect effect through selection into sample due to seizures. To separate the two, we use an idea similar to the Heckman (1990) selection model: we use the observed change in drug-unrelated arrests to *control* for selection. Third, we estimate the unobserved marginal costs with simulated method of moments, using data on average drug quality. Finally, we estimate fighting parameters via simulated method of moments, using data on fighting outcome probabilities.

Fights involving at least one large gang increase police attention by approximately 250%, compared to only 53% for fights between small gangs. We also find positive spillover effects across areas, with fights elsewhere increasing local police attention by 33% when involving gangs present in both locations. The demand estimation confirms consumers' preference for quality (0.32% demand increase for 1% drug purity increase) and shows substantial price elasticity (-2.5%), comparable to other studies on illegal drugs (Jacobi & Sovinsky, 2016). Crucially, police presence significantly impacts demand, with a 1% increase in police attention reducing demand by 4%–equivalent to a 1.67% price increase. Consequently, fights involving large gangs reduce demand by 14%, compared to 6% for small gang fights. Our fighting parameters indicate large gangs have a 70% higher fighting ability than smaller ones, matching the increased winning probability of 8.4%. In line with most fights ending in draws (71%), we estimate a large step of the fighting process and standard deviation of the fighting shock, determining an high unpredictability of the fighting process.

The model demonstrates a strong fit, with robust positive correlations between observed data and model predictions for both drug quantity seized and average quality. Moreover, the fighting propensities simulated by the model closely match those observed in the data. In the model, larger gangs have a 61% lower fighting propensity compared to small ones; in the data, this difference is 71%. Counterfactual analysis reveals that neither homogeneous fighting ability (which brings this difference from 61% to 27%) nor zero violence spillover (reducing the gap only to -49%) are able to fully explain the large gangs' lower fighting propensity. Instead, equalizing police reaction reverses the pattern entirely (+35% fights for large gangs), identifying the heterogenous police reaction and its associated demand effects as the primary mechanism behind the negative correlation between gang size and violence. Building on this insight regarding the importance of police reaction in shaping gangs' fighting incentives, the optimal police reaction function is determined.

This paper is related to several strand of the literature. First, it contributes to literature on the "organization" of crime, studying internal organization of criminal groups (e.g. Mastrobuoni & Patacchini, 2012), expansion in legal markets (e.g. Gambetta & Reuter, 2017; Le Moglie & Sorrenti, 2020; Mirenda et al., 2022), structure of illegal markets (e.g. Schelling, 1967; Becker, 1968; Polo, 1995; Gambetta, 1996; Levitt & Miles, 2006; Levitt, 2017; Lessing & Willis, 2019; Blattman, Duncan, Lessing, & Tobón, 2021), with a special focus on the functioning of the retail drug market (e.g. Levitt & Venkatesh, 2000; Galenianos et al., 2012; Galenianos & Gavazza, 2017), and its relationship with violence incentives (e.g. Dobkin & Nicosia, 2009; Dell, 2015; Battiston et al., 2022). The main contribution to this literature is to investigate violent competition between gangs in the retail drug market. In particular, we show how market structure shapes violent incentives and competition in this illegal market and highlight the role of police reaction in determining these incentives.

Second, it contributes to the industrial organization literature on competition in vertically differenciated retail markets (from Hotelling, 1929; Bresnahan & Reiss, 1991; Berry et al., 1993 to Aguirregabiria & Mira, 2007; Jia, 2008; Ciliberto & Tamer, 2009; Houde, 2009; Holmes, 2011; Aguirregabiria & Vicentini, 2016; Miller & Weinberg, 2017; Crawford et al., 2019; Murry & Zhou, 2020; Ciliberto et al., 2021). The main contribution to this literature is methodological: introducing, modelling, and estimating a model of targeted negative actions (in this case, fighting). This type of models could be applied in marketing for negative advertising (e.g. Grossman & Shapiro, 1984; Bass et al., 2005; Bostanci et al., 2021; Iyer & Singh, 2022), political confrontation (e.g. Coate, 2004; Schultz, 2007; Barton et al., 2016), and legal competition (e.g. Boncompte & Zufacchi, 2025).

To conclude, this paper also relates to the much broader literature examining the impact of

crime on various outcomes, including electoral results (e.g. Acemoglu, Robinson, & Santos, 2013; Alesina, Piccolo, & Pinotti, 2019), development (e.g. Pinotti, 2011, 2015; Melnikov et al., 2020; Fenizia & Saggio, 2024), firms (e.g. Barone & Narciso, 2015; Le Moglie & Sorrenti, 2020; Mirenda et al., 2022), and public sector (e.g.Daniele & Marani, 2011; Acconcia et al., 2014; Di Cataldo & Mastrorocco, 2022). In particular, we explore the relationship betweek market structure and gangs' violence, highligting the role of police reaction function in shaping it.

The paper is structures as follows. Section 2 introduces the setting of the study and the data. In section 3, we present evidence on the relationship between market structure and violence. Section 4 document key facts about the retail drug market and build the structure of the model. Section 5 describes the estimation procedure, discusses identification of the model parameters, and summarizes the results of the estimation. These are used to construct counterfactual scenarios in Section 6. Section 7 discuss the optimal police reaction. Section 8 concludes.

2 Setting and Data

2.1 Setting

Organized crime has a long history in the province of Naples, with some evidence of gang behavior going back to the 17th century, and the first written proof of its existence dating back to the 1800s (Barbagallo, 2014; Sales, 2022). The historian Marc Monnier describes these gangs as "a kind of schismatic police" that controlled Naples' prisons, markets, and illegal gambling houses (Monnier, 1862). These organizations experienced a significant evolution in the post-World War II period, transforming from a local phenomenon into an international criminal network involved in cigarette smuggling, drug trafficking, and other illicit activities (Sales, 1994, 2022). These new opportunities, along with business from the post-earthquake reconstruction following the 1980 disaster, provided the gangs with immense financial resources. At the same time, they created incentives for heightened competition for dominance among rival groups, leading to the first great gang fight in 1980 (Marrazzo, 2005), which resulted in approximately 1,000 homicides in six years.

From the 1990s to today, organized crime in Naples has maintained its horizontal structure with multiple independent gangs (Brancaccio & Martone, 2019). These clans are usually smaller than other Italian mafia groups and are characterized by lower internal organization and more intense conflicts (Catino, 2014). These characteristics make most of these gangs similar to some in North and South America (Dugato et al., 2020). This criminal ecosystem is structured around

two dominant cartels. While these two cartels try to maintain control over distinct macro-areas of metropolitan Naples, they are mainly composed by a complex network of small satellite groups that exercise tight territorial control over specific neighborhoods and districts. Thereofore, this dichotomous structure is marked by constant instability and fluid dynamics. Even within areas of exclusive cartel control, precarious equilibriums persist due to frictions between gangs–both across and within cartels–and the emergence of new criminal groups (for example of these frictions, please refer to reports from DIA, 2015 to DIA, 2022). As a result, this ecosystem is extremely violent, with mafia-related homicides accounting for 60-70% of all intentional homicides in its territories, significantly exceeding the national average of 10-15% (Catino, 2014). The scale of this "system" is massive, with activities in numerous countries (DIA, 2020), and thousands of affiliates divided into more than 100 gangs.

Today, these gangs engage in a wide range of criminal enterprises, with their core activities including drug trafficking, extortion, infiltration of public contracts, arms trafficking, counterfeiting, usury, protection rackets, prostitution, illegal waste disposal, and money laundering (Barbagallo, 2014). Estimated revenues from all these activities amount to approximately 9 billion euros annually (Transcrime, 2013). Despite this wide diversification, drug trafficking remains their main area of activity, accounting for more than 50% of estimated revenues (Eurispes, 2005). This is consistent with other criminal groups worldwide, as the drug business is often recognized as the most lucrative market for criminal organizations (Reuter, 2014).

Neapolitan gangs involved in the drug market tend to control the retail market by enlisting low-level criminal labor. Direct management of drug dealers creates an efficient distribution system and generally adheres to precise territorial boundaries with other gangs (Becucci, 2004). The exact production and distribution process of drugs varies dramatically among different gangs. However, it develops around some common grounds. International brokers, allied with the previously mentioned cartels, or operating independently, usually import raw drugs from other continents (mostly from Africa and South America - DCSA, 2024). When the broker arrives with the delivery, each gang takes its share, then cuts, packages, and prepares the drug for the retail market (Gribaudi et al., 2009). Local drug dealers then sell the drugs on the street, or, more recently, online. The prevailing method for retail drug sales consists of *piazze di spaccio* (literally, drug dealing squares), which are well defined and organized areas dedicated to the sales of illegal drugs, and their control has often fueled fights between gangs, resulting in numerous casualties (Catino, 2019). The management of dealers is very heterogeneous: the majority receive a fixed salary, some receive daily payment based on sales, while others operate as independent entrepreneurs. As well as paying salaries, a gang's common fund is often used to pay for assistance

and legal advice, the "tools of the trade" (weapons, cars, phones, etc.), and health and housing emergencies that members may face (Catino, 2019).

As an example of how well organized retail drug distribution is, here we cite one of these dealers (Ferrillo, M., *Spacciatore o spacciato?* Torino: Gruppo Abele (unpublished 2000, p. 39), from Becucci, 2004):

[...] street dealers are called to meet in a large apartment where drugs are distributed along with instructions. Everybody is assigned a precise spot to stand and somebody warns the dealer right away in case of danger when the police are on the way.

As a result, we should think about the retail drug market in the province of Naples as *competitive*, extremely *well-organized*, and *territorial*.

2.2 Italian enforcement system

Before turning to the actual data, it is worth discussing in few words the organization of the Italian law enforcement system. This system operates through four main police forces, each with distinct responsibilities and jurisdictions. The *Polizia di Stato* (State Police), under the Ministry of Interior, primarily handles urban policing, immigration, highway patrol, and railway security, operating mainly in cities and organized in provincial departments. Therefore, the whole Naples province will have only one State Police division. The *Arma dei Carabinieri*, a military force with police duties under both the Ministry of Defense and Interior, serves as a national gendarmerie force, particularly active in rural areas and small towns. The *Guardia di Finanza*, under the Ministry of Economy and Finance, specializes in financial crime, tax evasion, smuggling, and international drug trafficking. Finally, the *Polizia Locale* (Municipal Police), managed by individual municipalities, focuses on local law enforcement, traffic control, and urban regulations enforcement within specific city boundaries. These forces collaborate through multiple inter-agency organizations. For example, the Direzione Investigativa Antimafia (DIA), coordinates anti-mafia operations across all police forces, while the Direzione Centrale per i Servizi Antidroga (DCSA) specializes in coordinating drug-related investigations and operations among the various police forces, facilitating information sharing and joint operations against drug trafficking. We use data coming from all these agencies, often together, sometimes separately.

2.3 Data

To study the relationship betwen market structure and violence in the retail drug market, this paper builds a comprehensive dataset on this illegal sector by combining intelligence, administra-

tive, and survey data from different sources.

An important consideration is the data's geographical granularity. All administrative data, by their nature, are at the city-year level. For example, we observe the number of arrests for each city in the province annually, resulting in a single value for the entire city of Naples–a limitation inherent to the data. In contrast, intelligence data are collected at a finer level: city-year for locations outside Naples, and district-year within Naples city itself, which comprises 30 districts. This granularity allows deeper geographical analysis, enabling us to observe metrics like homicides or gang presence for each district in Naples city annually. Notably, most choice variables in the model come from administrative data, while state variables (primarily the matrix of gang presence) derive from intelligence sources. Consequently, we estimate the model at the city-year level and simulate it at the city-year level for Naples province and the district-year level for Naples city.

That being said, our sample is composed by all the cities in the province of Naples (and district within Naples city when available) for the period between 2015 and 2022 included.

Gangs presence The key data for studying the retail drug market is a set of matrices showing gang activity in this market across areas and years. We construct these matrices from intelligence data in two steps. First, using DIA reports (e.g. DIA, 2020), we constructed an initial 0-1 matrix indicating whether a certain gang was mentioned in the report, in a given area, in the semester under study. This reporting could appear in a map (e.g., Figure A1), or in the text (e.g., Figure A2 indicates a gang's entry into a new district). Here it is important to mention our definition of a gang, which we maintain consistently throughout data construction and empirical analysis. Defining gang boundaries is inevitably challenging, perhaps more so than for firms. We define a gang as a group of individuals active in illegal sectors (such as illegal drug trafficking) under a single name, which is clearly *not* a subgroup of another gang. This definition parallels Holmstrom and Tirole (1989) definition of the firm. While we acknowledge this approach's limitations, the absence of a comprehensive and well-defined gang list, this was, to our knowledge, the only feasible approach. The matrix constructed this way has two main limitations: it is not specific to retail drug market activities, and it records mentions in reports rather than actual presence. Therefore, in the second step, we submitted these matrices to the Neapolitan police to "correct them". By combining these matrices with their knowledge and intelligence data, they changed them to reflect, as accurately as possible, the retail drug market in each year (the shift from semester to annual data stems from the ability to identify gang presence in given areas more precisely). This process results in 8 matrices, one per sample year, reflecting different gangs' retail drug market

activity across cities–outside Naples city–and districts–within Naples city. Descriptive evidence aligning with expectations and literature (e.g., Figures 3 and 4), along with event-study variations consistent with plausible changes (e.g., Figure 6), provides reassurance about the quality of these matrices'.

Throughout this paper, we will define a gang as *large* if its scale (i.e., the number of areas in which it is present in the same year) exceeds 3. This classification corresponds to 10% of all gangs in the sample.

Fight matrices The data collection process for the fighting matrices follows the same approach. First, we compiled a list of fights from DIA reports, documenting the involved parties, year, contested area, and the winner. For each conflict, we noted whether the report mentioned a friction–*frizione*–or an actual fight–*guerra*–between the two gangs or groups of gangs. We then submitted these preliminary lists to the Neapolitan police for verification and correction. Using the finalized list, we constructed the fighting matrices (incorporating both frictions and fights).

Homicides To measure violence, we use intelligence data on the list of homicides and attempted homicides in the province of Naples from 2015 to 2022. For each violent event, we have information on the victim, where the event happened, and when it happened. By using these information, we also identified whether the event was gang related or not by looking at online gang mentions associated with the event.

Arrests To measure police attention, we use the administrative data on the number of arrests. In particular, for each city and year in the sample, we observe the number of arrests for different drug related crimes and conspiracy (from the antidrug authority).

Drug seizures To measure the retail drug market's size, we use administrative data on drug seizures in Naples province from the Italian antidrug authority (Direzione Centrale Servizi Antidroga). Enforcement agencies must report relevant information to the antidrug authority whenever they seize illegal substances. For all approximately 14,000 seizures during the sample period, we have data on city, date, enforcement authority, seizure type (territorial or maritime; we retain only territorial), type of illegal substance, and quantity seized in kilograms. This paper focuses on three main illegal substances: marijuana, heroin, and cocaine. Thus, we included only seizures of these drugs.[7] Since we are interested in the *amount* of drugs seized, we retain only seizures with

[7] Specifically, we included seizures classified as: "cocaine", "crack cocaine", "liquid cocaine" for cocaine; "heroin"

non-missing and non-zero quantities (quantities can be zero when the amount seized is negligible in kilograms). We then calculate, for each city, year, and drug type, the total amount of illegal drugs seized, the average quantity seized, and the number of seizures.

Chemical analysis For a subsample of the previous dataset, we also have information on the chemical analysis of the seized drugs–the percentage of pure substance in the sample. When a penal litigation is opened after a seizure (which happens whenever the defendant does not plead guilty immediately), the judge should require chemical analysis from the forensic police as a preliminary step for the litigation. Indeed, Italian law requires a minimum percentage of pure substance to be present in the seized sample for the crime of drug dealing/trafficking to be established. Therefore, the forensic police performs the relevant chemical analysis, creates a folder with the results, and submits it to the judge. We construct two anonymized random samples of these analyses in three steps. First, we randomly draw 567 analyses from all the relevant folders (province of Naples, from 2015 to 2022 included). Each of these folders contains information on the identity of the defendant and the percentage of pure substance in the sample. Second, these analyses are merged internally with a private version of the drug seizures dataset containing information on the identity of the person receiving the seizures. To conclude, the resulting dataset is anonymized and shared with the research team. Given the high incidence of seizures in the city of Naples, 86% of chemical analyses in this first completely random sample are from the capital city. To increase the number of cities covered, in the second sample we stratify on the city-year-drug level, trying to draw few analyses from each stratum and, instead, to cover as many strata as possible. These efforts result in 660 chemical analyses, widely distributed across cities, years, and drugs. For each analysis in the two samples, we observe all the information from the seizure (city, year, quantity) and the percentage of pure drug. Therefore, we calculate, for each city, year, and drug, the within dispersion and average drug purity.

Average prices Each semester the Italian antidrug authority requires the three main enforcement authorities (*Polizia di Stato*, *Arma dei Carabinieri*, and *Guardia di Finanza*) to fill an official form with the average retail and wholesale prices, the average purity, and average quantity for each type of drug. We collect all these forms, and calculate the average retail and wholesale price for each year and type of drug.[8]

for heroin; "hashish", "marijuana", and "cannabis plants" for marijuana.

[8]The heroin wholesale price was missing for the period from 2019 to 2021. We impute it using the average between the two extreme available periods: 2018 and 2022.

Other data We complete our set of data with additional sources, such as ISTAT data on mortality and population, survey data on drug consumption, survey data on city characteristics (including the activity of municipal police), and data on distances between geographical units. See Appendix D for a detailed description of these data.

3 Descriptive Evidence on Market Structure and Violence

In this section, we present evidence on the relationship between market structure and violence in the retail drug market in the province of Naples. First, we provide summary statistics regarding the characteristics of the gangs and their presence over time in various areas. Second, we offer descriptive evidence on the correlation between violence and market characteristics in each area and year. Third, we show how variations in average gang size–due to the entry of new gangs–creates marginal changes in violence in the area.

3.1 Summary statistics and gangs' characteristics

Figure A3 illustrates the number of active gangs across the province over time (shown in blue), along with the average number of gangs per area (shown in red). Over the sample period, the total number of active gangs in the province has increased significantly, from approximately 75 in 2015 to 85 in 2022. Notably, there was a sharp rise in the number of gangs following the Covid outbreak, continuing until the end of the restrictions from 2019 to 2021. The average number of gangs per area is consistently higher than one in each period, indicating the presence of multiple gangs in each area. As the number of gangs increases, so does the average number of gangs per area.

Not only are multiple gangs present in the same area, but the same gang can also operate in multiple areas. Figure A4 illustrates the density of our primary measure of gang size: the number of areas in which a gang is present at a given point in time, referred to as the gang *scale*. Approximately 60% of gangs operate within a single area. However, the distribution exhibits a long right tail, indicating the existence of gangs that are active in multiple areas within the same year. Specifically, around 20% of gangs are active in two areas, while the remaining 20% are active in more than two areas.

Importantly, even though the same gang can operate in multiple areas, there appears to be no spatial correlation in their presence. Figure A5 presents the distribution of the Moran's I index

calculated for each gang. For gangs present in only one area, this index is trivially equal to zero. The distribution of the index is centered around 0, with a slight positive tail; however, the overall values remain far from the 0.5 threshold typically associated with spatial autocorrelation.

Figures 1 and 2 jointly illustrate the variation that we exploit in this paper. If gangs were to systematically select areas–resulting in the observation of only areas with large gangs or only those with small gangs–or if the distribution of gang scale remained unchanged over time, there would be insufficient variation in gang scale for analysis. However, this is not the case in the province of Naples, making it an optimal setting to explore the relationship between market structure and violence. In Figure 1, we display the cross-sectional variation we exploit in this study. Each point in the scatterplot represents an area, plotted along two dimensions: the number of gangs present in that area and the average gang scale. The color of each dot indicates the maximum scale of the gang in the area. A significant insight from this graph is the absence of systematic sorting. There are areas with a very low number of gangs (or even monopolies) that host gangs of all sizes–some monopolies have small gangs while others have large gangs. There are areas with large gangs that have both many and few gangs, as well as areas that feature numerous small gangs and areas with several large gangs. In other words, there is considerable cross-sectional variation in gang scale and the number of gangs present across different areas. Figure 2 illustrates the variation over time. Specifically, each dot represents an area, positioned along two dimensions: the average gang scale in that area in 2015 and in 2022. The red line represents the 45-degree line; in the absence of temporal variation, we would expect all dots to lie on this line. However, this is not the case. For example, there are areas that had very large gangs in 2015 but no gangs in 2022, and vice versa.

To gain a better understanding of the temporal variation in the data, Figure A7 presents the transition matrix illustrating the changes in the number of gangs present in the areas over time. Two equilibriums appear to be relatively stable: areas with no gangs and areas with a very large number of gangs (more than 5 present simultaneously). In contrast, intermediate equilibriums seem to be quite unstable. For instance, the data indicate a 24% probability of transitioning from 3 gangs today to 2 gangs tomorrow.

Table A1 presents summary statistics on violence and police responses. Gang presence is widespread, with an average prevalence of 78%. Fights between gangs are relatively common, with a 12% probability of a fight occurring in an area within a given year. These areas are known for their high levels of violence, averaging 0.62 attempted homicides and 0.32 concluded homicides per year. Moreover, there is significant heterogeneity, with the maximum number of attempted homicides reaching 36 and concluded homicides totaling 9. The average likelihood of a homicide occurring in a specific area in a given year is 20%. Figure A6 illustrates the time trend of

violence. The number of homicides remains fairly stable over time, potentially showing a slight decrease over the sample period. In contrast, attempted homicides exhibit a noticeable spike during the Covid-19 period. Therefore, in our main analysis, we will concentrate on homicides. Police responses are also substantial, with an average of 47 arrests for conspiracy and 157 arrests for drug-related activities, as shown in Table A1. Local police, consistent with their role in the enforcement system in Italy, exhibit a much lower average number of arrests.

Table A2 presents summary statistics on fights between gangs. In our sample, we observe a total of 100 fights, with the contested area identified in 93 cases. In 35 of these fights, a winner exists and is known. Fights typically involve two *groups* of gangs, with an average of 3.25 gangs participating; the minimum number of gangs involved is, reassuringly, 2, while the maximum is 14. We do not observe any conflicts for a contested market in which neither of the two gangs was present at the time of the fight. Instead, in the majority of cases (80 out of 93), both gangs were active in the market for which they were fighting. Finally, most fights (72%) last only one year, resulting in an average fight duration of 1.49 years.

Table 1 provides summary statistics on average quality and average prices for each drug type. Figure A12 illustrates the distribution of observed drug quality–measured as the percentage of pure substance–for the three types of drugs considered in this paper. There is significant heterogeneity across drugs, with an average purity of 62% for cocaine, 22% for heroin, and 14% for marijuana. Moreover, there is substantial variation within each category of drug. For instance, we observe doses of cocaine ranging from 2% to 98% pure substance, indicating that we have data covering the entire spectrum of drug quality.

3.2 Market structure and violence

Having examined the characteristics of the retail drug market in the province of Naples, we now shift our focus to the correlation between violence and market structure. Specifically, we investigate how the probability of homicide varies with the number and composition of gangs present in an area.

Figure 3 presents the average probability of homicide along with the 95% confidence interval for areas with no gangs (0) and separately for the five quintiles of the distribution of the number of gangs present in the area (0-20, 20-40, 40-60, 60-80, 80-100). As expected, the probability of homicide is very low in areas with no active gangs. It increases significantly to approximately 20% in areas where a single gang holds a monopoly, and subsequently rises linearly with the number of active gangs, reaching a notable 40% in areas within the fifth quintile of active gangs. In summary, as the number of active gangs in an area increases, so does the probability of homicide.

Table 1: Summary statistics drugs

	Mean	SD	Min	Max
Quality cocaine	61.76	20.21	2	98
Average retail price cocaine	77.81	11.21	67.5	95
Average wholesale price cocaine	38.25	2.60	35	42
Quality heroin	21.19	12.3	0.63	76.5
Average retail price heroin	63.75	10.61	45	75
Average wholesale price heroin	25.19	5.66	15	32
Quality marijuana	13.78	6.96	0.2	58
Average retail price marijuana	11.19	2.23	9	15
Average wholesale price marijuana	2.58	0.57	2	3

Figure 4 illustrates how the probability of homicides varies with the standard deviation of the active gangs' scale. Consistent with findings in the conflict literature (e.g. Morgenthau, 1973; Reed, 2003), we observe an inverse U-shaped relationship between these two dimensions. In areas where the active gangs are very similar in size (whether small or large), the probability of homicide is relatively low. As the differences in gang sizes increase, so does the level of violence in the area. Between 1 and 1.7 standard deviations of gang scale, the probability of homicide reaches its peak before quickly declining. At the other extreme, where there is a significant inequality in the distribution of active gangs, the probability of homicide falls to its lowest point. Thus, violence is reduced when power in an area is balanced–making conflict outcomes too unpredictable–and when it is extremely unbalanced–making conflict outcomes too predictable. This is a classic result from the theoretical conflict literature, which is confirmed in this context.

In Figure 5, we examine how the probability of homicide changes with the average scale of active gangs in an area. Similar to the previous figure, the probability of homicide is very low in areas with no gangs. It increases dramatically, reaching up to 35% in areas where only very small active gangs are present (with a mean scale of 1). Interestingly, as the average scale of active gangs rises, the probability of homicide gradually decreases, reaching its lowest point at 20% in the fifth quintile of the distribution. In other words, there is a 15% difference in the probability of homicide between areas with very small gangs and those with large active gangs. This suggests that the size of the active gangs in an area may have a potential negative relationship with the level of violence generated in that area.

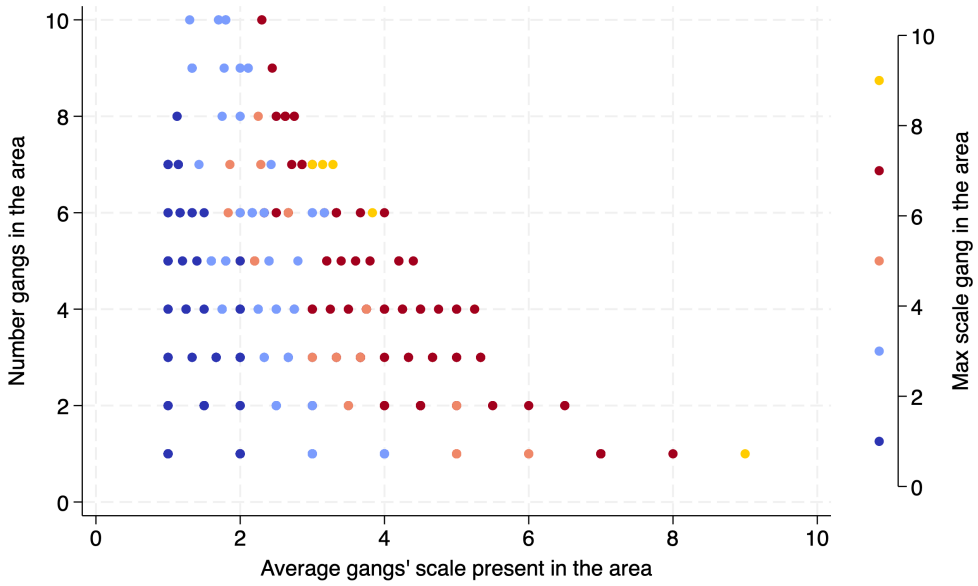


Figure 1: Cross section variation in gang scale

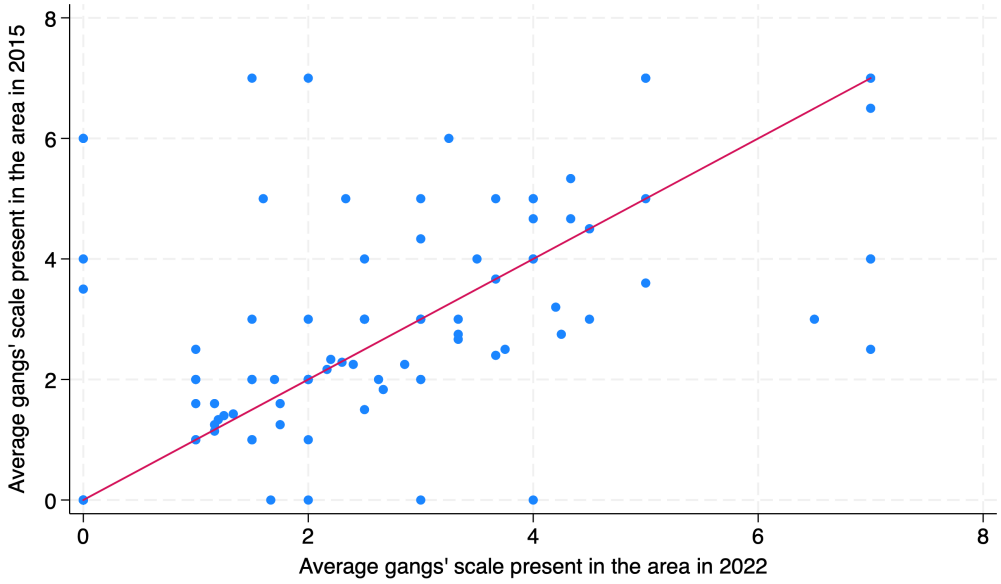


Figure 2: Time variation in gang scale

To further investigate this possibility, Figure A8 examines how the probability of homicide varies with the maximum scale of the active gang in the area, allowing us to clean for variation in the number of active gangs. The conclusions remain unchanged: we observe a very low probability of homicide in areas with no active gangs, followed by a sharp increase in areas where the

largest gang is small (with a maximum scale of 1). Subsequently, there is a consistent decline in the probability of homicide, with roughly a 10% difference in areas that have a large active gang (with scales ranging from 4 to 9).

This descriptive evidence may be influenced by both violence directly perpetrated by gangs and unrelated violence in the area, as we cannot distinguish–at least with the probability of homicide–whether a homicide is gang-related or not. Consequently, we are unable to differentiate between active violence and the gangs' ability to *control* the territory. To address this concern, in Figure A9, we replicate the previous four figures using only gang-related homicides, identified by mentions of gangs associated with the homicides on the web. Reassuringly, the probability of gang-related homicides is always exactly zero in areas with no active gangs. All prior conclusions remain robust when focusing on this more specific outcome. In Figure A10, we shift our attention to the probability of observing a fight in the area, which serves as a direct observation of active violence by gangs. Once again, it is encouraging to note that the probability of conflict is exactly zero in areas without active gangs. All previous conclusions still hold: (1) the probability of conflict increases with the number of active gangs in the area; (2) the relationship between conflict probability and the standard deviation of gangs' scale follows an inverse U-shape; (3) the probability of conflict decreases as the average scale of active gangs increases; and (4) the probability of conflict decreases with the maximum scale of the active gang in the area.

3.3 Impacts of gangs' entry

Are larger gangs less violent? The previous descriptive evidence suggests that areas with larger gangs experience lower levels of violence. Moreover, by simply counting the number of fights for each gang in each year, large gangs are 14% less likely to be involved in a fight.[9] However, without additional evidence, it is difficult to attribute this finding to gang behavior rather than to differences across areas and systematic selection regarding gang presence.

The first step is to account for unobserved differences across areas and years. Table A3 presents the correlations between various measures of violence (probability of homicide, probability of fight, number of homicides, and number of homicides per capita) and the average scale of active gangs. In columns 1 and 4, we control for both area and year fixed effects. Consistent with previous findings, all measures indicate that an increase in average gang scale is associated with a decrease in violence. Specifically, a one unit increase in average gang scale–with respect to an average of 1– is associated with a 10% reduction in the probability of homicide, 17% decrease in

[9]Notice that large gangs are present, by definition, in more areas. This is associated with an higher number of fighting possibilities, which makes this figure even more striking.

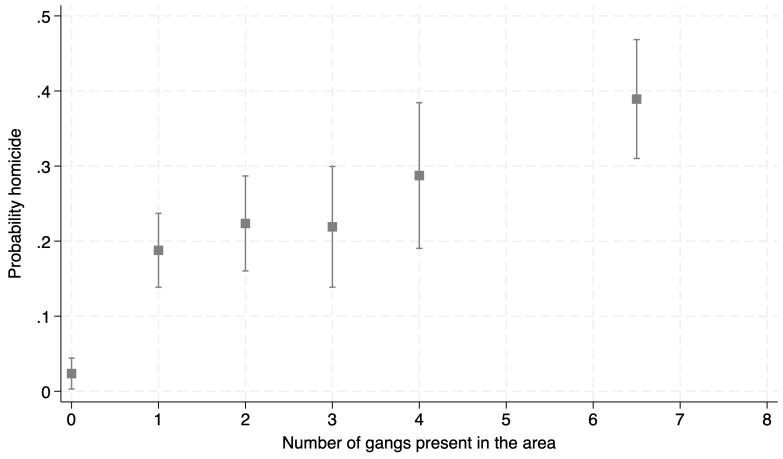


Figure 3: Number gangs

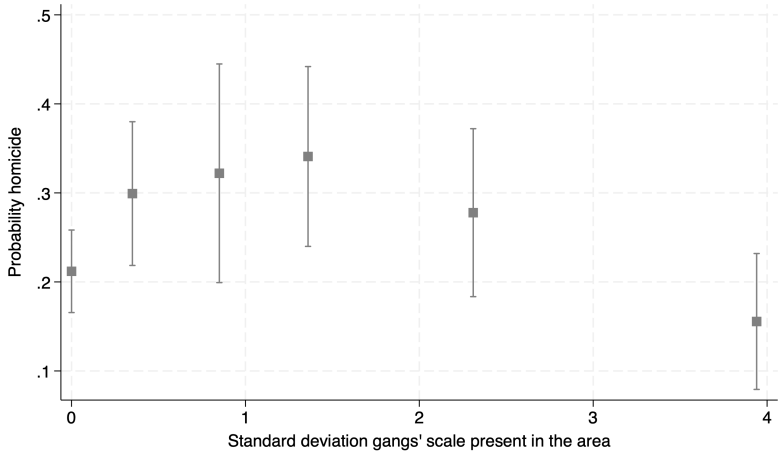


Figure 4: SD scale gangs

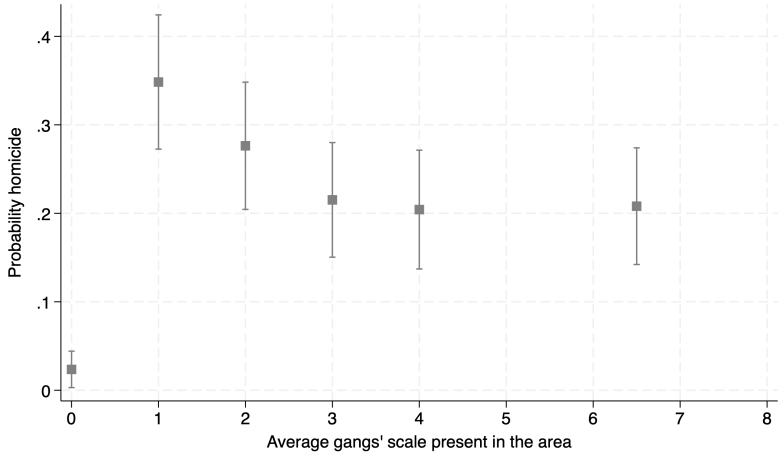


Figure 5: Mean scale gangs

the probability of conflict, 18% decline in the total number of homicides, and a 33% reduction in the number of homicides per capita. The results remain quantitatively similar when controlling for the number of active gangs in the area (columns 2 and 5) and the standard deviation of active gangs in the area (columns 3 and 6). Consequently, our descriptive evidence cannot be explained by fixed characteristics of the areas and years.

To better understand the factors driving this correlation, we need to focus on the residual variation in the average scale of gangs after controlling for area and year fixed effects. This variation can arise from two scenarios: gangs increasing or decreasing their scale over time, or gangs entering or exiting the area. We focus on the second type of variation and explore the heterogeneous effects of gang entry on violence. To do so, we estimate the following event-study design:

$$\begin{aligned} V_{yt} =& \alpha + \sum_{j=-4}^{-2} \beta_j (\text{Lag j}) + \sum_{k=0}^{4} \beta_k (\text{Lead k}) + \epsilon_{yt} \\ (\text{Lag j})_{yt} &= 1[t = E_y - j] \quad \text{for } j \in \{1, \ldots, -4\} \\ (\text{Lead k})_{yt} &= 1[t = E_y + k] \quad \text{for } k \in \{1, \ldots, 4\} \end{aligned} \tag{1}$$

where V_{yt} is the violence measure in area y at time t, E_y is a dummy variable equal to 1 when a new gang enters the area in y. Period -1 serves as the omitted category. To prevent contamination from later entries, we restrict our analysis to the most recent entry in the area. To explore the heterogeneity of gang entry, we will estimate equation 1 separately for generic entries, entries of small gangs (with a scale of 3 or lower), and entries of large gangs (with a scale greater than 3).

Figure 6 displays the classic event-study graph for three different models: generic entry (in grey), entry of a small gang (in red), and entry of a large gang (in blue). Consistent with expectations, the entry of a new gang into an area disrupts the equilibrium, leading to an increase in violence. Specifically, the entry of a new gang is associated with an approximately 11% increase in the probability of homicide in the area during the first period, compared to the omitted category, and a 20% increase in the second period. After two years, this effect gradually declines toward zero. This average effect conceals substantial heterogeneity based on gang size. Notably, the increase is more pronounced when a small gang enters the area, with increases of approximately 15% and 24% in the first and second periods, respectively. Conversely, the entry of a large gang is associated with a decrease in the probability of homicide, approximately -15% in both the first and second periods compared to the reference category. In both cases, these effects dissipate two years after entry.

The key assumption of this research design is that the gangs' choice of when and where to

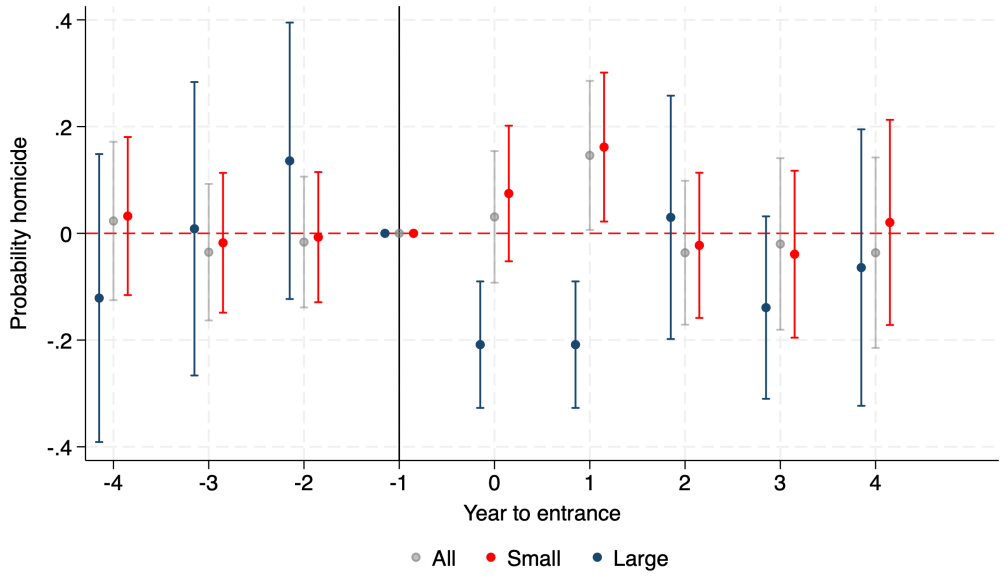


Figure 6: Event study - probability homicide

enter is uncorrelated with residual violence. The absence of pre-trends–though there appears to be a slight pre-trend in the model for large gang entry, more on this follows–is reassuring about the strength of the identification assumption. Figure A11 shows the "raw data" version of the event study shown in Figure 6. In particular, we plot the average probability of homicide in the different years with respect to the entry of a gang, small gang, and large gang. As expected, results are in line with the event-study approach. However, given that here we don't have a reference category, we can dig deeper on the subtle pre-trend for large gangs. This is composed by two factors: a slight increase in probability of homicide in -2 and a decrease in the reference category (-1). Therefore, this could be partially explained by an error in the entry measurement.

The coarse nature of the temporal unit of observation (year) may raise some concerns about the robustness of this exercise. It's possible that unobserved factors within a given year could simultaneously influence both the entry of gangs and the level of violence in an area. For these results to hold, such unobserved factors would need to have heterogeneous effects based on gang scale, which, while unlikely, remains a possibility. Thus, we should be cautious when interpreting these findings as causal. Nonetheless, it is reasonable to conclude that gangs of different sizes appear to behave differently, potentially reflecting differences in their propensity for violence.

Table A4 evaluates the robustness of these findings across different measures of violence. Specifically, column (1) shows the event study estimates for the probability of homicide, column

(2) for the number of homicides, column (3) for the probability of gang-related homicides, and column (4) for the probability of observing a fight in the area. The table provides separate event studies for all gangs entering an area, small gangs entering, and large gangs entering. For the first two categories (all gangs and small gangs), we observe an increase in all violence measures during the first two periods following entry. Pre-trends are absent for the first three measures. For the probability of observing a fight, pre-entry coefficients are statistically different from zero, though they show an opposite trend compared to the post-entry period. For large gang entries, results align with the heterogeneity observed in figure 6. Following the entry of a large gang, there is a reduction in all measures of violence in the first two periods post-entry, and a zero effect on fighting incidence, compared to the positive effect observed with smaller gangs.

4 The retail drug market

In the previous section, we provided evidence supporting the idea that larger gangs tend to be less violent. To understand why this is the case, we explore gangs' incentives for fighting. In this section, we outline some key facts about the retail drug market and develop a quantitative fighting model.

We present three facts about this market. First, consumers seem to value drug quality and dislike police presence. Second, larger gangs have a higher probability of winning fights. Third, police reaction to violence is heterogeneously influenced by gang size. Specifically, the police tend to respond more strongly to fights involving large gangs.

We then present a model of the retail drug market where gangs can engage in fights to shut down a competitor's selling point. The model's key mechanisms are informed by the evidence presented in the first part of the section. Uninformed consumers decide where to search for drugs, considering transport costs, signals on the distribution of drug quality. Gangs maximize profits by setting drug quality and deciding whether to engage in fights, and against whom. When deciding whether to fight, gangs weigh the expected additional profits from gaining a larger market share against the costs stemming from increased police presence. Consistent with the presented evidence, we allow gangs of different sizes to have heterogeneous fighting abilities, which influence their probability of winning the fight, and heterogeneous police reactions, which affect how police respond to fights involving gangs of different scales.

4.1 Three facts

Here we present three facts about the retail drug market, which will guide the structure of the quantitative model.

FACT 1: Consumers value drug quality and dislike police presence.

The large heterogeneity in drug quality presented in section 3 suggests that products in the market, even when conditioned on drug type, are inherently (vertically) differentiated. Now, the question arises: do consumers perceive this differentiation, and how do they react to it? Figure A13 presents the scatter plot between the average quantity seized[10] and average drug quality, standardized by type. The data shows a upward trend in quantity with increasing quality: areas selling purer drugs tend to have a higher average quantity per seizure. This suggests consumer sorting, indicating that consumers may prioritize quality when selecting where to purchase drugs.[11] Hence, the evidence suggests that consumers recognize quality differences and respond accordingly, implying that they positively value drug quality. Nevertheless, the correlation is quite weak, and the correlation coefficient is not statistically different from zero, implying a noisy estimate. This is consistent with a scenario of imperfect information about drug quality, which is in line with the existing literature (Galenianos et al. (2012) and Galenianos and Gavazza (2017)). Turning to police presence, figure A14 presents the scatter plot between the average quantity seized and the number of municipal police arrests per capita (per 1k)–a measure of police attention independent from drug activities. This correlation seems to be positive, but this is driven by two outliers with very high arrest rates. When cleaning from these outliers in Figure A15, we discover a strongly negative correlation, indicating that in areas wigh higher police attention tend to have a lower average quantity per seizure. This suggests that consumers may negatively value police attention.

FACT 2: Larger gangs have a higher probability of winning fights.

Table A5 presents the differences in the probability of the fighting outcomes for large gangs compared to small gangs. This is calculted with a simple probit regression on the dummy variabile indicating the fighting outcome on a dummy variable that indicating the size of the gang. The

[10] This is the number of grams in each seizure. Focusing on this measure allows us to ignore seizure selection at this stage of the paper, while still providing some evidence on the quantity of drugs found in each seizure. We will tackle this in further stages of the paper.

[11] An alternative explanation may be that, as gangs sell more drugs, they become *better* at cutting drugs. But if this would be the case, and consumers do not value drug quality, then we would observe exactly the opposite.

sample is restricted to fights involving only two gangs (which accounts for 70 out of 100 fights). Perhaps unsurprisingly, large gangs have approximately three times the probability of winning fights compared to small ones. This increased probability is primarily driven by a reduced likelihood of draws, meaning that the fights have fewer inconclusive outcomes during the period under consideration. In other words, when engaging in a fight, a large gang is three times more likely to successfully push its competitor out of the area within the year.

This economy of scale in the fighting mechanism may arise from various factors, including increased manpower availability and potentially greater financial resources. For the purposes of this paper, we acknowledge its existance and incorporate a heterogeneous fighting process in our quantitative model. However, it is interesting to examine the levels of violence generated in different fights. To do this, we estimate an event-study model similar to the one described in equation 1, analyzing how the number of homicides changes with the onset of a fight, differentiating between fights involving only small gangs and those that also include large gangs. The results are summarized in Figure A16. Reassuringly, the data quality is confirmed by a spike in the number of homicides during the year a fight begins. This effect persists for up to two years following the start of the fight. When we differentiate between fights involving only small gangs and those involving larger gangs, we observe that the effect is present in both cases; however, it is significantly larger in the latter scenario. Thus, fights involving larger gangs result in greater levels of violence. Pre-trends are absent in all models. This increase in violence is not confined solely to the contended area; we also observe spillover effects in other areas where the gang operates. Figure A17 replicates the previous analysis, illustrating the impact on the number of homicides in both the contended area and other areas where the gangs are present. We will consider these spillover effects when discussing the potential mechanisms underlying our results and their implications for the model.

FACT 3: Police reaction to fights is heterogenous by gang size.

Figure 7 presents how the number of local police arrests, one possibile measure of police attention, changes on the onset of a fight. The year in which a fight starts, we observe a large increase in the number of drug related arrests. When dividing between fights among small gangs and involving a large gang also, the difference is massive. A fight between small gangs generates an increase in 10 arrests in the area in the first year, with respect to the previous period; a fight involving a large gang 25. This means that police respond differently to fights involving gangs of different scale. Therefore, we will allow for an heterogenous police reaction in the model.

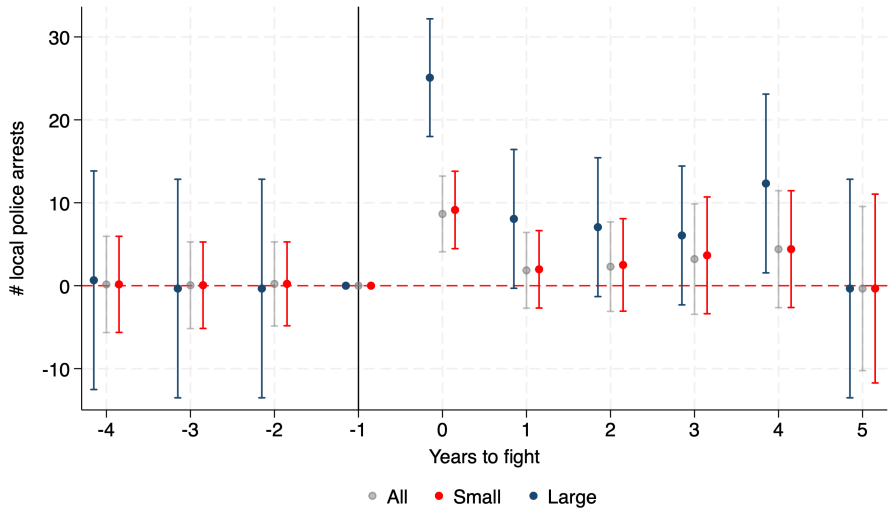


Figure 7: Event study - number of local police arrests

The results remain consistent when considering alternative measures of police attention, such as drug related arrests (illustrated in figure A19). Importantly, examining different measures of police attention reveals substantial insights into the persistence of the effects over time. In the case of drug-related arrests, we observe a persistent effect. However, when we focus on other potential measures of police attention, such as arrests made by local police, the effect is significantly less persistent. This discrepancy may be attributable to changes in the underlying criminal activity rather than variations in police attention itself. We will leverage this intuition extensively in section 5.

4.2 Potential mechanisms

These facts suggest various potential mechanisms behind the negative correlation between gang size and violence. First, the *heterogeneous police reaction function*. Fact 3 highlights how police appear to respond more strongly to fights involving larger gangs, potentially leading to higher fighting costs for these gangs, and ultimately determining a diseconomy of scale in fighting.

Second, as shown in Figure A17, violence generated by fights, along with the resulting police attention (as seen in Figure A18), spread in all the areas in which the involved gangs are present. Therefore, violence and attention *spillovers* may create heterogeneity in the costs of fighting, contributing to the observed correlation. However, the presence of larger gangs in multiple areas could also reduce the cost of fighting. As long as attention spillover is less than complete, larger

gangs can diversify revenues from less affected areas, potentially reducing the impact on their total revenues.

Finally, considering Fact 2, a *deterrence* mechanism. Smaller gangs, recognizing their lower probability of winning fights against larger gangs, may choose to avoid confrontation.

To test these potential mechanisms we develop a quantitative fighting model.

4.3 Model - Setup

The economy consists of a finite set of areas (indexed $y = 1, 2, \ldots, Y$) and features a discrete number of gangs $j = 1, 2, \ldots, J$, and consumers $i = 1, 2, \ldots, I$, of different drugs $v = 1, 2, \ldots, V$ over various periods $t = 1, 2, \ldots, T$.

Gangs Gangs are heterogenous in starting scale and produce vertically differentiated drugs. They decide: (i) drug quality, given prices–quality competition– and (ii) whether, and against whom, to fight—criminal competition.

Consumers Consumers have homogeneous preferences and imperfect information over the quality distribution. This is in line with the literature on drug quality and demand (Galenianos et al. (2012) and Galenianos and Gavazza (2017)). However, differently from this literature, here we want to model the one year drug-consumer relationship. Hence, we will not have repeated search, but one single supply decision, made under some degree of uncertainty. There is a fixed amount of consumers for each type of drug. For each drug, in each period, they choose where to buy drugs and from which gang. Hence, they choose a bundle (j, y). The outside option is not to buy any drug. Demand has a nested logit structure (similar to Goldberg, 1995) with imperfect information.

Police Police, in the benchmark model, is essentially a function that, given the fighting matrix of the period, allocates attention across the different areas.

Time The model is static, which is justified by the need to maintain tractability in computing an equilibrium with many gangs and locations. Consequently, in each period, gangs observe their exogenous starting scale[12] and make optimal decisions based on it.

[12] The initial exogenous starting scale is reconstructed from the data.

Space Although in reality gangs might operate more than one dealing point in each area, we assume that they can have at most one point in each market (similar to most of the literature on spatial competition in retail sectors; e.g. Aguirregabiria & Vicentini, 2016).

Denote by $\mathcal{Y}_{jt} = \{y \mid y_{jt} = 1\}$ the set of areas in which gang j is active at time t. To conclude, define the matrix $\mathbf{Y}$ where $y_{yj} = 1$ if gang j has a dealing point in y. Given the absence of fixed costs associated with entering different types of drug markets, we can safely assume that, on the extensive margin, if a gang is present in an area, it can produce all types of drugs.[13] This significantly simplifies the state space, as we need to consider only one network matrix per period instead of V matrices. That being said, if selling a specific drug in an area is not profitable, the gang can always adjust on the intensive margin and set the quantity to zero.

4.4 Model - Demand

Utility Consumer i, living in area y_i, of drug v, at time t, decide whether to buy drugs from gang j in area y, or the outside option of not to buy any drugs $y = 0$. The conditional indirect utility that consumer i receives from buying drugs v from j in y is:

$$u_{ijyvt} = \beta_v x_{jvty} + \alpha_v p_{vt} + \mu_v O_{yt} + \xi_{yvt} + d_v g(||z_i - z_y||) + \epsilon_{ijvt} + (1-\rho)\epsilon_{iyvt} \quad (2)$$

where p_{vt} is the average price charged for drug v at t, ξ_{yv} are unobserved, to the econometrician, but observed by suppliers and consumers, characteristics of the drug v supplied in area y, and $g(||z_i - z_y||) = g_{iy}$ is the distance between the areas. x_{jvty} is the inverse of percentage of pure substance of the drug v supplied by gang j at t in y. All consumers have the same ranking over x_{jvty} and ξ_{yv}, which reflect the vertical differenciation across products. O_{yt} is a measure of police attention in the area. Hence, in line with fact 1, we allow consumers to have preferences over quality–β_v–and police presence–μ_v. If this last parameter is negative, as suggested by the preliminary evidence presented in section 4.1, this generates a cost of fights for gangs (in a similar vein with respect to Levitt & Venkatesh, 2000). When fighting, gangs actract the attention of police in the area, as seen in section 3, discouraging consumers. ϵ_{ijvt} are consumer specific taste shocks for drug v supplied by j in t, and, similarly, ϵ_{iyvt} are consumer specific taste shocks for drug v in y in t. I assume tha both shocks are standard type-I extreme value distributed.

Define the nests to be $g = 0, 1, ...Y \ \forall v$ where $g = 0$ denotes the outside option (assume

[13] This assumption is justified by observations that all drugs are seized in the areas. Figures A20 and A21 show that areas with cocaine seizures typically also see seizures of the other two drugs. Therefore, we do not observe specialization in drug trafficking by area, but rather a scenario in which all drugs are sold in the same area, consistent with our assumption.

$g_{i0} = 0 \ \forall i$), and its other values are the different areas in the province of Naples. Therefore, parameter ρ, the nesting parameter, governs the switching between the areas. We impose perfect substuitability across nests ($\rho = 0$). Notice that, in this framework, consumers are not allowed to switch type of drugs.

As standard in the literature, denote by $\delta_{jyvt} = \beta_v x_{jvty} + \alpha_v p_{vt} + \mu_v O_{yt} + \xi_{yvt}$ the mean consumer valuation.

Information sets Consumers do not observe perfectly drug quality at the begining of their decision problem. When choosing across nests (the areas) the consumers has a limited information set. In particular, they know which grup is present in the area $\mathcal{J}_y$, its area specific taste shock ϵ_{iyvt}, the unobserved characteristics of the drug v supplied in y ξ_{yvt}, police attention O_{yt}, and the average drug quality in the area $\bar{x}_{ytv}$. Once arrived in the area, the consumer observe all the present suppliers, and therefore learns about the true quality distribution in the area. In other words, x_{jytv} becomes known $\forall j \in \mathcal{J}_y$. This knowledge structure allows us to introduce uncertainty in drug quality, without assuming perfect unobservability and repeated intereaction, which would complicate substantially the demand side of the model. The information structure of the game can be represented in the game tree in figure A22. From the perspective of the consumer at the beginning of the decision problem, we can model the unobserved quality as an additive separable noisy process: $x_{jyvt} = \bar{x}_{vty} + \zeta_{jtvy}$. We impose some structure on this uncertainty, assuming that $\zeta_{jtvy} \sim N(0, \sigma_v^{\zeta})$ and $\epsilon_{jyvt} \perp\!\!\!\perp \zeta_{jtvy}$.

Consumer problem Define $d_{(y,j)}$ to be the indicator of the consumer decision; $d_{(y,j)} = 1$ if the consumer chooses the bundle (y, j) and 0 otherwise. The consumer problem may be written in the following way:

$$\begin{aligned} \max_{d_{(y,j)}} \textstyle\sum_{(y,j)} d_{(y,j)} \mathbb{E} u_{ijyvt} \quad & \forall v, t \\ \text{s.t.} \textstyle\sum_{(y,j)} d_{(y,j)} = 1 & \end{aligned}$$

Since in this model an individual is defined by product specific shocks $(\epsilon_{ijvt}, \epsilon_{iyvt})$ and by its location y_i, this implictly defines the set of individual attributes that lead to the purchase of good (j, y) from y_i. For ease of exposition, we can think about this problem as if it was composed by two steps. First, the across nest decision, where the consumers choose where to buy drugs. Second, the within nest decision, where consumers, after having observed the full distribution of drug quality in the area, choose from which gang to buy drugs.

Within nests The within nest decision problem is fairly simple. The consumer has complete information, and given the consumer specific taste shock, chooses the gang in the area which maximize its utility. Given that the only difference between consumers is their location, and in this problem they have already chosen the optimal area, the share of consumers choosing a specific gang will not depend on i. Appendix A.1 provides the conditional demand at its complete derivation.

Across nests Instead, when choosing across nests (areas and outside option), the consumer face uncertainty about drug quality in the area. In particular, the consumer does not observe the whole distribution of drug quality in the different areas. Therefore, he chooses the area which gives him the highest *expected* utility. By exploiting the properties of the standard type-I extreme value distributed consumer specific taste shock ϵ_{iyvt}, and the structure imposed on the drug quality, uncertainty, we derive the across nests demand function (please refer to appendix A.1 for additional details):

$$s_{yvt|y_i} = \frac{\exp(\bar{\delta}_{yvt} + d_v g_{iy} + \sqrt{\varsigma \sigma_g^2 + 1} log(N_{yt}))}{1 + \sum_{\tilde{y} \in \mathcal{Y}} \exp(\bar{\delta}_{\tilde{y}vt} + d_v g_{i\tilde{y}} + \sqrt{\varsigma \sigma_g^2 + 1} log(N_{\tilde{y}t}))}$$

where $\bar{\delta}_{yvt} = \bar{x}_{vty}\beta_v + \alpha_v p_{vt} + \mu_v O_{yt} + \xi_{yvt}$ is the average mean consumer valuation in the area, and N_y is the number of gangs active in the area. One thing is worth mentioning from this expression. All else equal, the nest share is increasing in the number of gangs active in the area. This effect is driven by the uncertainty in drug quality. The higher the number of gang in the area, the more ample will be the ex-post choice of drugs.

Consumer sorting across locations plays a key role in demand estimation, given the nature of the data. We use consumer sorting across, rather than within, nests to estimate demand parameters. To do so, we need to bring these unconditional demand for nests in an reduced form model we can estimate in the data. By writing the unconditional demand for drugs in nest y as weighted sum of the conditional demands, taking the logarithm of these shares, and exploiting the properties of the logarithm (appendix A.1 shows steps), we can rewrite these uncondtional demand functions in a log-lin form:

$$log(s_{yvt}) = \bar{x}_{vty}\beta_v + \alpha_v p_{vt} + \mu_v O_{yt} + \zeta_v log(N_{yt}) + \mu_t + \mu_y + \mu_v + \xi_{yvt} \tag{3}$$

which we then can bring directly to the data for demand estimation.

Unconditional demand By combining the conditional demand within nests, with the across

nest demand, we obtain the unconditional demand for the bundle (y, j) (please refer to appendix A.1).

4.5 Model - Supply

Production Gangs produce drugs using pure drug r and cutting substances s according to a Leontief production function:

$$q_{jyvt} = min\{\frac{s_{jyvt}}{1 - x_{jyvt}}, \frac{r_{jyvt}}{x_{jyvt}}\} \tag{4}$$

x_{jyvt} is the inverse of the percentage of pure drug in the dose, and it is a choice variable of gangs. This production function is very intuitive: to produce $q_{jyvt} = 1$ gram of drug, the gang needs x_{jyvt} grams of pure substance and $1 - x_{jyvt}$ grams of cutting substance. To conclude, we introduce an unobserved marginal cost θ. This could represent dealers costs, organizational ones, or equipments. Profits from dealing point y and drug v are: $p_{vt}q_{jyvt} - k_{vt}s_{jyvt} - c_{vt}r_{jyvt} - \theta q_{jyvt}$.

Firm problem As mentioned in the setup of the model, gangs solve a two-stage game:

1. In the first stage they decide whether, where, and against whom to fight. In other words, they choose the optimal fighting matrix $\mathbf{F}_{jt}$. The combination of these matrices, and the realization of the fighting outcomes will determine the network matrix of the period $\mathbf{Y}_t$;

2. Conditional on fighting outcomes, gangs to maximize profits simultaneously, given average prices, production function, and demand.

Second stage In the second stage gangs maximize profits simultaneously by choosing quantity, labour, raw substance, and drug quality given prices, the network matrix, demand, and production function. Details in appendix A.2. Inverting the production function, and writing the problem with an implicit demand function, the profit maximization problem can be written as:

$$\max_{x_{jyvt} \forall y \in \mathcal{Y}_{jt}} \quad \psi(\mathbf{F}_j) \sum_{y \in \mathcal{Y}_{jvt}} \sum_{v} [p_{vt}q_{jyvt}(x_{jyvt}) - q_{jyvt}(x_{jyvt})[k_{vt} + (c_{vt} - k_{vt})x_{jyvt} + \theta]] \tag{5}$$

In other words, gangs maximize profits by choosing the optimal degree of drug quality. They do so by solving the tradeoff between the additional cost of an additional unit of pure substance in the dose, with the positive effect on quality, and therefore demand. Hence, we model the retail drug market as classical example of *vertically differenciated* goods market. where firms differ

over quality, and consumers are able to rank their options from best to worst using an objective measurement. $\psi(\mathbf{F}_j)$ is a shrinkage coefficients for fights, such that $\psi(\max \mathbf{F}_j = 1) = \psi$ and $\psi(\max \mathbf{F}_j = 0) = 1$. In other words, profits in fights states are shrinkade by a ψ factor. This may represent other costs of fights, such as arrests, operational costs, etc. As a result, a fight has two distinct effects on profits: (i) the negative effect on demand, and (ii) the shrinkage of profits. Notice that, by modelling the second cost in this multiplicative way, it does not affect the optimal profit maximization decisions in the first stage. Moreover, it is scale neutral, thus not affecting the different fighting propensity, the object under study.

The system of best response equations can be written as:

$$\mathbf{x} - \Lambda(\mathbf{x})^{-1} q(\mathbf{x}) = \tfrac{\mathbf{k}}{\mathbf{c}-\mathbf{k}}(\mathbf{p} - \boldsymbol{\theta})$$

where $\Lambda(\mathbf{x})$ is the $JVT \times JVT$ matrix of quality derivatives (following Berry et al., 1993; Berry, 1994). An equilibrium of this game is a the vector $\mathbf{x}^*$ that solve this fixed-point mapping. Given our assumptions, the mappings $q(\mathbf{x})$ and $\Lambda(\mathbf{x})$ are continuously differentiable. These conditions are necessary and sufficient for this equilibrium to exists and be unique (as outlined in Mizuno, 2003).

4.6 Model - The fighting game

The core part of the model is the fighting game. In this game, gangs must decide whether to fight, against whom to fight, and for which area. The choice set of the game is $\mathcal{F}_{jt} = \{(y, o) = 0, 1 : \forall (y, o) \in \mathcal{Y}_{jt} \times \mathcal{J} - j\}$, where $\mathcal{Y}_{jt} = \{y \mid y_{jt} = 1\}$ is the set of areas in which gang j is active at time t. In other words, we restrict gangs to fight only for areas in which they are present in the period (which accounts for 86% of fights in the data). This assumption substantially reduces the dimensionality of the choice space. We do not restrict the number of fights a gang can engage in during one period. However, we do not model any interaction between fights within the period. In other words, as will become clear when discussing the fighting mechanism, the probability of a gang winning a fight is independent of the number of fights in which the gang is involved in the period. While we acknowledge this assumption is problematic, it allows us to avoid strategic interaction among gangs which would substantially complicate the game. Moreover, it enables us to treat fights among multiple gangs (which represent approximately 40% of fights in the sample) as independent fights between single gangs. Finally, we allow fights to end with either a definitive outcome (one gang wins and the other loses) or in a draw. This allows us to accommodate fights lasting more than one period as repeated fights with draws. The choice set can be represed by the set of $JY \times JY$ matrices $\{\mathbf{F}_{jt} \ \forall j, t\}$ where $f_{yo} = 1$ if gang j at time t decide to fight

against gang o for y. With this notation, we are implictly setting $f_{yo} = 0 \; \forall y \notin \mathcal{Y}_{jt}$, and, clearly $f_{yj} = 0$ always. Given the similarity between this fighting game and a classical static entry model, following Reiss & Bresnahan, 1990, Berry, 1992, and Mazzeo, 2002, we do not consider mixed strategy equilibria.

Throughout this subsection, we will compare gangs profits–determined in the second stage of the game–across different possible gangs fighting decisions and states of the world. To easy exposition, let's define an implicit profit function:

$$\Pi(\boldsymbol{\Upsilon}, \boldsymbol{\omega}) = \pi^* : \mathbb{R}^Y \times \Omega^{YJJ \times YJJ} \to \mathbb{R}$$

where $\boldsymbol{\Upsilon} = [\theta, \mathbf{p}, \mathbf{c}, w, \mu, \alpha, \beta, \xi, \psi]$ is the vector of exogenous profit shifters, and $\boldsymbol{\omega}$ is the matrix of possible possible states of the world with elements:

$$\omega_{yjo} \in \Omega = \{\text{WNF,WF,P,D,NF,L}\}$$

which indicate, respectively, win without fighting, win with fighting, peace, draw, non fight, and loose.

The rest of this subsection is structured as follows: (i) we discuss the fighting mechanism; (ii) we show how police reacts to fights; (iii) we present a simplifyied two gangs game; (iv) we extend this simple game to the multi-gang context.

Fighting mechanism Building on the seminal paper by Donohue and Levitt (1998), each gang j is characterized by a fighting ability ϕ_j, which is common knowledge among all gangs. Each fight has a fight-specific unobservable (to everyone) shock ϵ_{jot}. This unobservable component can be thought of as randomess in the fight outcomes. The ϵ's are assumed to be independently and identically normally distributed with mean 0 and variance σ_ϵ. Hence, σ_ϵ measures the unpredictability of the fighting outcome. Notice that, consistently with fact 2, we allow the fighting ability–and therefore the fighting outcome probability–to be heterogenous across gangs, and hence gangs' scale. Given that the observable fighting ability is observed by all the players, as all the other components of profits, this is a game of complete information. One can in principle allow for incomplete information by making the fighting ability private information.

Suppose a fight between a gang j and $-j$ at time t. Fighting mechanism is a step function:

$$\begin{cases} \phi_j + \epsilon_{jot} - \phi_{-j} > \Phi & j \text{ wins} \\ \Phi \geq \phi_j + \epsilon_{jot} - \phi_{-j} > -\Phi & \text{draw} \\ -\Phi \geq \phi_j + \epsilon_{jot} - \phi_{-j} & j \text{ loses} \end{cases} \tag{6}$$

where Φ is the step parameter. As mentioned before, here there is no mention of other fights. In other words, the fighing ability, and therefore the probabilty of the different fighting outcomes is independent from the number of other fights. This is an important assumption which simplifies substantially the transition from the two gangs game to the multi-gang one, since it allows us to treat fights as independent, and abstract from coalitions.

Police reaction As mentioned previously, in the bechmark model, police is essentially a function, the *police reaction function*:

$$O_{yt} = \bar{o}_y + f(\lambda_m \mathbf{F}_{jty}, \delta \mathbf{F}_{jt-y}) \tag{7}$$

where $\bar{o}_y$ is the average police attention in y, $\mathbf{F}_{jty}$ is the matrix of fights of gang j in y at t, and $\mathbf{F}_{jt-y}$ the matrix of fights of gang j *not* in y at t. Consistently with fact 3, we allow for an heterogenous police reaction λ_m to fights of gangs of different scale. Moreover, in line with the evidence presented in the previous section, we allow for attention spillovers in other areas in which the same gang is present through δ.

A simple fighting game For simplicity, we start with a simple fighting game between two gangs j and o in an area y. In this simplified game, we compare the following set of profits:

$$\{\Pi(\boldsymbol{\Upsilon}, WNF) = \pi_{WNF}; \Pi(\boldsymbol{\Upsilon}, WF) = \pi_{WF}; \Pi(\boldsymbol{\Upsilon}, P) = \pi_P; \Pi(\boldsymbol{\Upsilon}, D) = \pi_D; \Pi(\boldsymbol{\Upsilon}, NF) = \pi_{NF}; \Pi(\boldsymbol{\Upsilon}, L) = \pi_L\}$$

Each year t is divided in four periods. In period t0, before the actual start of the game, all gangs learn about the observable fighting ability of all the others. Then, in period t1, they repeatedely play the sequencial fighting game. Once a set of optimal fighiting matrices is determined, in period t2 shocks are revealed and fighting outcomes realized. Therefore, at period t2, the new network matrix of the period $\mathbf{Y}_t$ is determined. In period t3 gangs compete over quality given the network matrix (second stage of the gang problem), and profits are realized in period t4.

In period 1, the two gangs repeatedely play a sequencial fighting game. Each day of the year, nature gives a random opportunity to fight to one of the gangs. The chosen gang can choose whether to fight (F) or not (NF). If NF is chosen, then they will get peace profits π_P. If, instead, F is chosen, it is the turn of the other gang, which can choose whether to fight back (F) or not (NF). If choose not to fight, then it exits the area, determining the win-non fight profit for j, π_{WNF}, and the non fight profit for o, π_{NF}. If instead, o chooses to respond, then both get the expected profit from fighting, which is:

$$\mathbb{E}\pi_F = P(\phi_j + \epsilon_{jot} - \phi_o > \Phi)\pi_{WF} + P(\phi_j + \epsilon_{jot} - \phi_o < -\Phi)\pi_L + (1 - P(\phi_j + \epsilon_{jot} - \phi_o > \Phi) - P(\phi_j + \epsilon_{jot} - \phi_o < -\Phi))\pi_D$$

for j, and similar by changing the probabilties and relevant profits for o. Given the sequencial nature of the game, the appropriate solution concept is the Subgame Perfect Nash Equilibrium (SPNE). By repeatedly playing this sequencial game, the possible yearly equilibria emerge as self absorbing state.

Before turning to the SPNE with uncertainty, let us examine the case of no uncertainty $\sigma_\epsilon = 0$. Appendix A.3 presents all the SPNEs for this game. In this case, we have six possible SPNEs, depending on parameter values and the distribution of second-stage profits. These six possible SPNEs map into three scenarios: (i) j exits, (ii) o exits, (iii) multiple equilibria. Specifically, j exits whenever o will win the fight, or they will draw but o can credibly sustain the threat of fighting. The opposite holds for the scenario in which o exits. Finally, we have multiple equilibria when they will draw and nobody can credibly threaten to fight (a scenario in which the first to get the opportunity to fight wins the area) or both can credibly threaten to fight (a scenario in which the first to get the opportunity to fight exits the area). Importantly, in all these scenarios, a unique outcome emerges: no fighting. This result is in line with the classical literature on conflict with no uncertainty.

In appendix A.4 we present SPNEs for the game with uncertainty. When we allow for $\sigma_\epsilon > 0$, nine SPNEs emerge. These correspond to 5 possible scenarios: (i) j exits, (ii) o exits, (iii) multiple equilibria, (iv) nothing happens, (v) fight. Specifically, j exits whenever expected profits from fighting are lower than those from not fighting, and o can credibly threaten to fight. The opposite holds for the scenario in which o exits. Multiple equilibria appear when expected profits from fighting are lower than those from not fighting for both gangs. In this case, no gang can credibly threaten to fight, and whoever gets the first opportunity to fight wins the area. Scenario (iv) occurs–nothing happens–when expected profits are higher than profits from not fighting for both gangs, but never higher than profits from peace. In this case, both gangs can credibly threaten to fight, so neither chooses to fight. Finally, a fight happens whenever both gangs have incentive to fight as long as the other fights (expected profits higher than profits from not fighting), and at least one of them has expected profits from fighting higher than peace profits. In this case both gangs can credibly threaten to fight, and given that fighting is optimal for at least one of them, eventually a fight will occur. So, with uncertainty, two outcomes may occur, depending on the profit distribution: fighting and not fighting. In other words, some degree of uncertainty is necessary to sustain a fighting equilibrium.

To shed light on the probability of the fighting outcome happening, let's dig deeper into the

choice probabilities predicted by the model. Figure 8 reports the choices determined by the game in the space of expected profits of the two gangs.

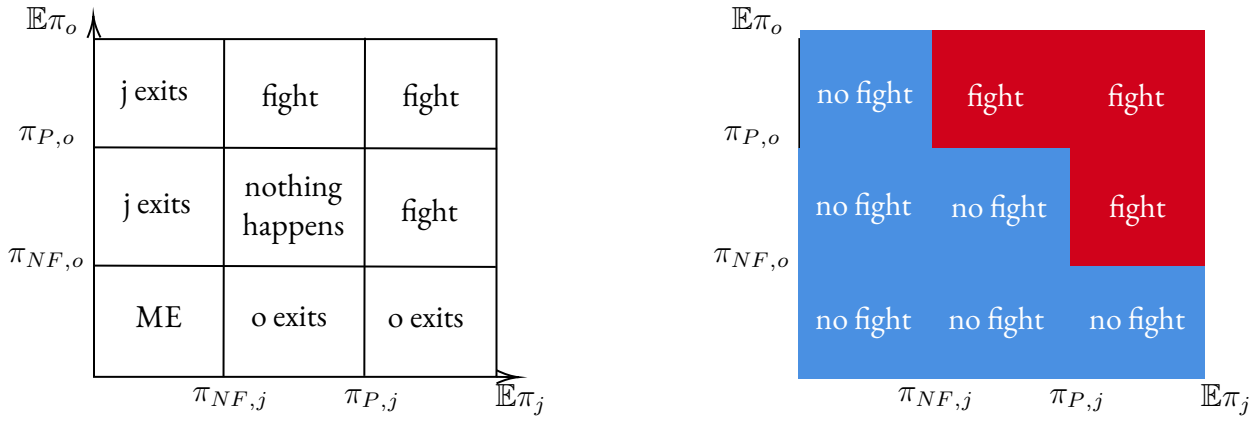


Figure 8: Fighting outcomes

The choice probabilities predicted by the model are:

$$\begin{cases} Pr(NF,NF) = Pr(\pi_j^{NF} \leq E\pi_j \leq \pi_j^p \ \& \ \pi_o^{NF} \leq E\pi_o \leq \pi_o^p), \\ Pr(F,F) = Pr(\pi_j^{NF} \leq \pi_j^p \leq E\pi_j \ \& \ \pi_o^{NF} \leq E\pi_o \leq \pi_o^p; \\ \quad \pi_j^{NF} \leq E\pi_j \leq \pi_j^p \ \& \ \pi_o^{NF} \leq \pi_o^p \leq E\pi_o; \pi_j^{NF} \leq \pi_j^p \leq E\pi_j \ \& \ \pi_o^{NF} \leq \pi_o^p \leq E\pi_o), \\ Pr(NF,F) = Pr(E\pi_j \leq \pi_j^{NF} \leq \pi_j^p \ \& \ \pi_o^{NF} \leq E\pi_o \leq \pi_o^p; E\pi_j \leq \pi_j^{NF} \leq \pi_j^p \ \& \ \pi_o^{NF} \leq \pi_o^p \leq E\pi_o) + \\ \quad \int T(NF,F) 1[(E\pi_j, E\pi_o) \in R] dF_{E\pi_j, E\pi_o}, \\ Pr(F,NF) = Pr(\pi_j^{NF} \leq E\pi_j \leq \pi_j^p \ \& \ E\pi_o \leq \pi_o^{NF} \leq \pi_o^p; \pi_j^{NF} \leq \pi_j^p \leq E\pi_j \ \& \ E\pi_o \leq \pi_o^{NF} \leq \pi_o^p) + \\ \quad \int (1 - T(NF,F)) 1[(E\pi_j, E\pi_o) \in R] dF_{E\pi_j, E\pi_o} \end{cases} \tag{8}$$

where

$$R = \{E\pi_j, E\pi_o | E\pi_j \leq \pi_j^{NF} \leq \pi_j^p \ \& \ E\pi_o \leq \pi_o^{NF} \leq \pi_o^p\}$$

The first two equalities in 8 are simple. For example, the model predicts *(NF,NF)* uniquely if and only if the expected profits belong to the central quadrant. Equalities three and four are more complicated. Both probabilities consists of the case when *(NF,F)*, or *(F,NF)*, are the unique equilibrium of the game (top left quadrants and bottom right ones), and olso when *(NF,F)*, or *(F,NF)*, are a *potentially observable outcome* of the game and it is the outcome that was "selected". Following Ciliberto & Tamer, 2009, define $T(NF,F)$ the selection mechanism function. This

function "splits" the ME quadrant in the two possible equilibrium, determining the choice probabilities. The problem is that this function is unkown to the econometrician, and can potentially be different over gangs, areas, and time. In other words, if we were interested in the composition of the market, this simple fighting game would have multiple equilbria.

However, we are not interested in the composition of the market, but only in whether a fight occurs or not, and this outcome is predicted uniquely by the model; this argument is very similar to the insight on the prediction in the *number*, rather than the identity, of entrants by Reiss and Bresnahan (1990) and Berry (1992). This is clear if one looks at the second matrix of figure 8. Outcome probabilities predicted by the model are summarized in section A.5. Given uniqueness of the second stage equilibrium, the model predicts *uniquely* whether a fight is going to happen or not, conditional on the observables and parameters. In other words, the fighting game has a unique equilibrium in the fighting outcome, which is what we are interested in.

This paper investigates how market structure determine fights in the retail drug market. In particular, we are interested in how the fighting propensity changes with the scale of gangs in the area. Given the derived outcome probabilities, this translates into: $\frac{\triangledown Pr(F_{jo}=1)}{\triangledown m_j}$. Theoretically, what is the sign of this derivative? Computing this derivative explicitly is a challenging exercise. However, we can achieve the same result by exploring how the outcome changes in the expected profit space. We do so in Figure A24. Consider starting from a non-fighting equilibrium, as in panel A. Suppose now that, all else remaining constant, we counterfactually increase gang j's scale. This has six distinct consequences on the model's primitives. Let us examine their effects on the fighting probability separately. The first obvious consequence is an increase in the probability of winning the fight, which increases j's expected profits from fighting. As shown in Figure A24 B, this can lead to a switch into fighting equilibrium. Second, the decreased probability of the opponent winning the fight reduces their expected profits from fighting (*deterrence* effect), which could again lead to a switch in the equilibrium (panel C). Third, non-zero spillover of police attention causes a decrease in profits across all areas where the gang is active. Higher scale and more markets lead to lower expected profits from fighting (panel D). Fourth, conversely, by being active in multiple areas, as long as the spillover coefficient is strictly less than one, higher scale ensures revenue diversification, which increases the expected profits from fighting (panel E) and could potentially switch the equilibrium. Fifth, if police reaction is stronger for larger gangs (as suggested by fact 3), higher scale implies lower expected profits from fighting (panel F), which could again switch the equilibrium. Taking these factors together, theoretically, the sign of the derivative we are interested in is a priori *ambiguous*, as it combines different mechanisms with opposing signs. Ultimately, it depends on the combination and strength of these mechanisms,

and therefore on demand elasticities, police reaction coefficients, and spillover effects. Different parameter sets may yield different signs of the derivative. Therefore, to answer this research question, we must estimate these parameters and determine which sign emerges from the estimated parameter set. We will then counterfactually vary parameter values to study how different mechanisms contribute to the sign of this derivative.

The extended game We now extend back our game to a multi-gang and multi-area setting. Let's define a fighting outcome function:

$$P(\mathbf{m}, \mathbf{F}, \omega) = p : [0,1]^{J \times T \times Y \times J} \times \mathbb{R}^J \times \Omega \to \mathbb{R}$$

which, for each set of gang scale, fighting decisions, and possible state reports the probability of that state happening. Then, gang j in the extended game solve:

$$V(\boldsymbol{\Upsilon}, \mathbf{F}_{-j}) = \max_{\mathbf{F}_j} \sum_{\omega \in \Omega} P(\mathbf{m}, \mathbf{F}_j, \mathbf{F}_{-j}, \omega) \Pi(\boldsymbol{\Upsilon}, \mathbf{F}_j, \mathbf{F}_{-j}, \omega) \tag{9}$$

We can think about this extended fighting game as a collection of *connected* 2 gangs 1 area games. They are connected because decisions in one simple game has consequences on the optimal decisions of the others. The existance of multiple equilibria should not be an issue for identification of the parameters. Nevertheless, it may still be an undersiderable feature when turning to simulations and results. Characterizing all the equilibria may be very computational intensive, and iterative procedures may converge to one possible equilibrium, with impossibility of discussing general implications for all the possible equilbria.

Hence, we need to study the sources of multiplicity. Consider the 2-1 game between j and o in y. We are interested in the sign of $\frac{\triangledown Pr(F_{joy}=1)}{\triangledown F_{j'o'y'}}$. If these derivatives would be all equal to zeros, then we would have a single equilibrium. Unfortunately, this is not the case. Given our assumptions on the absence of strategic interactions in fighting decisions (no economy/diseconomy of scale in the number of fights), the only connection between the simple 2-1 games is in the (expected) profits. Formally, $\frac{\triangledown P(\mathbf{m},\mathbf{F}_j,\mathbf{F}_{-j},\omega)_{joy}}{\triangledown F_{j'o'y'}} = 0$ but $\frac{\triangledown \Pi(\boldsymbol{\Upsilon},\mathbf{F}_j,\mathbf{F}_{-j},\omega)_{joy}}{\triangledown F_{j'o'y'}}$ potentially not. In particular, fights by other gangs and/or in other areas, generate changes in police attention and potential number of gangs in the areas, modifying consumer decisions, profits in all the possible states, and therefore fighting incentives.

However, although it is definitively possible that $\frac{\triangledown \Pi(\boldsymbol{\Upsilon},\mathbf{F}_j,\mathbf{F}_{-j},\omega)_{joy}}{\triangledown F_{j'o'y'}}$ is different from zero, we are not really interested in the sign of this derifative per se. Instead, our goal is to determine:

$$\frac{\triangledown Pr(F_{joy}=1)}{\triangledown F_{j'o'y'}} = f\left(\frac{\triangledown \mathbb{E}\pi_{joy}}{\triangledown F_{j'o'y'}} - \frac{\triangledown \pi^{NF}_{joy}}{\triangledown F_{j'o'y'}}, \frac{\triangledown \mathbb{E}\pi_{joy}}{\triangledown F_{j'o'y'}} - \frac{\triangledown \pi^{p}_{joy}}{\triangledown F_{j'o'y'}}\right)$$

As a result, it is possible that these derivatives are sufficiently close to zero even though the profit ones are not. Indeed, if the change in the profits in the different states of the world is approximately the same, then other fights, potentially done in other areas and/or by other gangs, have no impact on the 2-1 game under consideration. In this case, we would have a single equilibrium also in this extended game, given that all the 2-1 games would actually turn out to be independent from each other. For the moment, assume that this is the case. Given independence between games, we have a single equilibrium, which allows us to compute the optimal gang decisions in parallel, significantly reducing time required for the simulation of the model.

4.7 Equilibrium and selection

Given a set of starting network matrix $\{\mathbf{Y}_t\}$, the joint distribution of consumer demographic groups and locations $F(D_i, y_i)$, gang scale $\{m_{jt}\}$, and areas characteristics $\{\xi_{yvt}, \bar{o}_y, g_{yy'}\}$, the equilibrium is defined as the set of fighting matrices $\{\mathbf{F}_t\}$, and gang quality choices $\{x_{jyvt}\}$ that, given the vector of parameters $[p_t, c_t, k_t, \theta_j, \alpha_v, \beta_v, d_v, \mu_v, \psi, \lambda, \phi, \delta, \sigma_\epsilon, \sigma_\psi, \Phi]$, satisfy the following (independently) for each period t:

1. Quality choices solve the profit maximization problem 5;
2. The fight matrix solve the fighting game 9.

Data Before turning to identification and estimation, we think it is important to discuss briefly again the data to underline the bridge between the model and the data. Table A6 summarizes all the available data, the corresponding role in the model, and its source. We have data on the network matrices (both ex-ante and realized, since we can re-construct the ex-ante matrices by reverting the fights), observed average drug quality, prices, and other characteristics of the areas. However, we do not directly observe drug quantities and police attention. We have some measures of these two unobserved variables, which are inherently selected, as discussed in detail in section 4.1. Hence, the final step is to *model* this selection.

Selection Drug quantity and police attention are unobserved ($q_{yvt} \to q^*_{yvt} \quad O_{yt} \to O^*_{yt}$). In the data, we have different measures of police attention: the outcomes of police activity. Suppose I_{yt} is the incidence of an illegal activity i. Then we can assume that arrests for a given activity i at time t in area y, are a function of police attention O^*_{yt} and the incidence of the legal activity I_{yt}:

$$A^i_{yt} = a(I_{yt}, O^*_{yt})$$

For example, the arrests function for drug related crimes could be specified as $A^d_{yt} = a(I^d_{yt}, O^*_{yt})$. To conclude, following a similar reasoning, we assume a selection equation:

$$Q_{vyt} = s(q^*_{vyt}, O^*_{yt})$$

5 Identification and Estimation

To quantify the strenght of the different mechanisms contributing to the relationship between market structure and fighting probability, we structurally estimate the model. As a preliminary step, a subset of parameters are directly taken from the data $\{p_t, c_t, k_t, \sigma_\psi\}$. The estimation is carried out in four steps. First, spillover and police reaction function parameters $\{\bar{o}_y, \delta, \lambda_m\}$ are identified through changes in arrests for drug-unrelated crimes due to exogenously predicted fights. Estimation combines an event-study design with IV. Second, $\{\alpha_v, \beta_v, d_v, \mu_v\}$ are identified through the correlations between changes in seized drug quantities and changes in observed quality choices, average prices, and police attention–controlling for seizure selection. Estimation procedure is an exploits the linearity of the across areas unconditional demand 3. Third, given demand parameters, we use data on average observed quality to identify and estimate the unobserved marginal cost $\{\theta\}$ via simulated method of moments. Finally, given fighting choices, we use data on fighting outcomes and decisions and their correlation with gang scale to estimate fighting parameters $\{\psi, \phi_j, \Phi, \sigma_\epsilon\}$ via simulated method of moments.

Model simplification We make a few simplifying assumptions to bring the model to the data. First, we reduce the degree of heterogeneity across gangs. In particular, we reduce the degree of heterogeneity to two type of gangs: small and large, as in the reduced form part of the paper. As a result, we only have two parameters of fighting ability (ϕ_s, ϕ_l). We then normalize $\phi_s = 1$, so that ϕ_l should be interepreted in relative terms.

Second, given the limited sample size, we reduce the number of demand parameters by assuming heterogenous preferences across drugs: $\{\alpha, \beta, d, \mu\}$. This is clearly a strong assumption. However, the limited sample size limit our power in investigating heterogenous consumer preferences.

Third, we need to specify a functional form for the police reaction function. Also in this case, the dimensionality of the gang heterogeneity is reduced to the dichotomy between small and large gangs. For simplicity, we assume a linear police reaction function:

$$O^*_{yt} = \bar{o}_y + \lambda_s F_{sty} + \lambda_l F_{lty} + \delta F_{tj-y} \times 1[j \in y] + u_{yt}$$

where F_{sty} is equal to one if a fight between small gangs happened at time t in y, F_{lty} equal to one if this fight involve at least one large gang, and the product $F_{tj-y} \times 1[j \in y]$ when a fight invoving a gang active in y happened at t in another area $-y$.

To conclude, we need to specify a functional form for the arrest function and the selection equation. We assume the following functional forms:

$$A^i_{yt} = I_{yt}(e^{\gamma O^*_{yt} + \epsilon_{yt}} - 1) \quad Q_{vyt} = q^*_{vty}(e^{\gamma O^*_{yt} + \epsilon_{vyt}} - 1) \tag{10}$$

where γ is a parameter indicating the effectiviness of police attention, and ϵ_{yt}, ϵ_{ytv} are two shocks distributed around 0. Notice that we have specified the same functional form for both equations, since it is reasonable to assume that the effect of police attention on drug seized and arrests is the same. This functional form simplifies substantially the mathematical proofs of identification, but it is also reasonable. When police attention is zero in an area $O^*_{yt} = 0$, then the expected number of arrests/drug seized is indeed zero. The number of arrests/drug seized is increasing in police attention (freely, with elasticity depending on the parameter γ), plus some idiosincratic shock. Notice that we can rewrite these equations in arrests/drug probabilities as follows: e.g. $\frac{Q_{vyt}}{q^*_{vty}} = e^{\gamma O^*_{yt} + \epsilon_{vyt}} - 1$).[14] Building on this, due to normalization of police attention ($\sum_y O^*_{yt} = 1$), and reasonable restrictions on the behaviour of the two equations ($0 < e^{\gamma + \epsilon_{vyt}} - 1 < 1$), we can derive some bounds on the effectiveness parameter $0 < \gamma < 0.3$, which will be useful in identification.

Preliminary step As a preliminary step for the estimation, a subset of parameters are directly taken from the data $\{p_t, c_t, k_t, \sigma_\psi\}$. In particular, average retail and wholesale prices are taken from the average official forms given to the antidrug authority by the enforcement authorities. The average price of cutting substances is set at 0.01 euros per gram. Drug quality dispersion within areas is directly estimated from the data, taking the average standard deviation in the percentage of pure substance within areas in the same year.

First Step: Police reaction function The first step is to estimate the police reaction function. The problem here is that we don't observe police attention directly, but only through some indirect measures: the number of arrests for different crimes. To recover the parameters of the police reaction function, we explore how the number of arrests change with fights. By taking the

[14] We will use the following approximation substantially in what follows: $log(e^x - 1) \approx x - 0.7 \quad \forall x \in [0, 1]$

derivative of the arrest function:

$$\frac{\nabla A^i_{yt}}{\nabla \mathbf{F}} = \frac{\nabla a(I_{yt}, O^*_{yt})}{\nabla I_{yt}} \times \frac{\nabla I_{yt}}{\nabla \mathbf{F}} + \frac{\nabla a(I_{yt}, O^*_{yt})}{\nabla O^*_{yt}} \times \frac{\nabla O^*_{yt}}{\nabla \mathbf{F}}$$

Changes in arrests for crime i following fights (the data) is the weighted sum of two (unobserved) components: how the crime incidence changes with fights $\frac{\nabla I_{yt}}{\nabla \mathbf{F}}$ (the first component), and the parameter we are interested in, the change in police attention $\frac{\nabla O^*_{yt}}{\nabla \mathbf{F}}$ (the second one). As a result, an identification concern arises here, since without assumptions or additional data, we cannot differenciate the two through changes in the number of arrests. To solve this problem, we make the following identification assumption:

Assumption A. *$\exists \hat{I}$ such that* $\frac{\nabla \hat{I}_{yt}}{\nabla \mathbf{F}} = 0$

What we are assuming here is the existance of one type of crime which is completely unresponsive to fights. Given the nature of the municipal police, described in section 2.3, the nature of illegal activities they prosecute is most likely indepenent to drug or gang dynamics, therefore likely satisfying the uncorrelation assumption. Given this assumption we identify changes in police attention due to fights through changes in drug-unrelated arrests (i.e. the first part of the derivatives is equal to zero, allowing the identification of the second component). Another identification issue we face in estimating reaction parameters is the endogeneity of fighting choices. Whether to fight or not, especially given the fighting game previously described, is a choice made by gangs, after having potentially observed unobserved shocks to police activity in the different areas. To deal witht this source of endogeneity, we use an IV: releases of gang's members from prisons. Anectodally, it is a well established fact that releases of gang's members from prisons, expecially if important members in the hierarchy of the gang, is often connected with fight outbreaks, given the distruption in the equilibrium. Therefore, in a first stage regression, we use this event to predict "exogenous" fight outbreaks–i.e. those outbreaks spurred by exits from prisons $\hat{F}_{yt}$. Then, we use these fights to identify and estimate reaction parameters with an event-study design. Appendix A.6 shows identification of police reaction function based on assumption A and a parallel trend assumption. Given bounds on police effectiviness, we are able to identify a lower-bound for the police reaction parameters. However, the ratio between reactions to fights of gangs of different scales, which is what we are mostly interested in, is point identified. Given

identification, we use a two step procedure for estimation:

$$\begin{aligned}
&\text{FS:} \quad F_{ytj} = \beta Z_{tj} + \mu_y + \mu_t + \mu_j + u_{ytj} \\
&\text{SS:} \quad A^i_{yt} = \alpha + \sum_{j=-4}^{-2} \lambda_j (\text{Lag j}) + \sum_{k=0}^{4} \lambda_k (\text{Lead k}) + \mu_y + \mu_t + \epsilon_{yt} \\
&\quad (\text{Lag j})_{yt} = 1[t = \hat{F}_y - j] \quad \text{for } j \in \{1, \dots, -4\} \\
&\quad (\text{Lead k})_{yt} = 1[t = \hat{F}_y + k] \quad \text{for } k \in \{1, \dots, 4\}
\end{aligned} \tag{11}$$

where Z_{tj} is a dummy variable indicating the realease of a gang j member from prison at time t, and $\hat{F}_y$ is the predicted fight from the first stage. This is similar to what done in section 4.1 for Figures A17 and 7, but with the predicted rather than true fights. The parameter of interest is λ_0, estimated separately for fights only involving small gangs (λ_s), fights involving large gangs (λ_l), and fights in other areas involving gangs present in y (for δ). Results are summarized in Table A7. Although it is inherently impossible to test the parallel trend assumption, the absence of pre-trends provides reassurance about the validity of our approach.

Second Step: Demand Parameters The main parameter of interest here is the consumer dis-utility from police presence μ. Naively, we would like to estimate it through changes in drug quantities with police attention, or fights. However, we do not observe the total quatity of drug sold in an area, but only the seized one. Let's explore how this changes with fights:

$$\frac{\nabla Q_{vyt}}{\nabla \mathbf{F}} = \frac{\nabla s(q^*_{vyt}(O^*_{yt}), O^*_{yt})}{\nabla O^*_{vyt}} \times \frac{\nabla O^*_{vyt}}{\nabla \mathbf{F}} + \frac{\nabla s(q^*_{vyt}(O^*_{yt}), O^*_{yt})}{\nabla q^*_{vyt}} \times \frac{\nabla q^*_{vyt}(O^*_{yt})}{\nabla O^*_{vyt}} \times \frac{\nabla O^*_{vyt}}{\nabla \mathbf{F}}$$

Therefore, changes in drug seized at the fight outbreak is the composition of two effects: (i) the direct effect of the fight on the quantity sold, and (ii) the change in selection/seizures due to the fight. Our parameter of interest is $\frac{\nabla q^*_{vyt}(O^*_{yt})}{\nabla O^*_{vyt}} = \mu$. Hence, without a method to isolate the selection effect, this parameter is not identified. To solve this problem, we use the observed change in drug-unrelated arrests to *control* for selection. Comparing how seized drug changes with fights, for similar values of drug-unrelated arrests, we isolate the negative effect of fights on drug incidence. This idea is very similar to the Heckman, 1990 selection model. To do so, in addition to the previous assumption A, we need to assume that selection is homogenous across the two crimes. Formally:

Assumption B. $\frac{\nabla a(\hat{I}_{yt}, O^*_{yt})}{\nabla O^*_{yt}} = \frac{\nabla s(q^*_{vyt}(O^*_{yt}), O^*_{yt})}{\nabla O^*_{vyt}}$

If this is the case, we can write:

$$\frac{\nabla Q_{vyt}}{\nabla \mathbf{F}} - \frac{\nabla \hat{A^i_{yt}}}{\nabla \mathbf{F}} = \frac{\nabla s(q^*_{vyt}(O^*_{yt}), O^*_{yt})}{\nabla q^*_{vyt}} \times \frac{\nabla q^*_{vyt}(O^*_{yt})}{\nabla O^*_{vyt}} \times \frac{\nabla O^*_{vyt}}{\nabla \mathbf{F}}$$

which isolate the parameter of interest. Appendix A.7 shows identification under assumption A, B, and a parallel trend assumption. As in the Heckman, 1990 selection model, estimation follows a two step approach. First, we estimate selection exploiting the correlation between drug seized and the number of arrests by the local police. Second, we estimate the log-linear demand *controlling* for selection:

$$\begin{aligned} &\text{1 step:} \quad log(Q_{yvt}) = \eta log(\hat{A^i_{yt}}) + u_{yvt} \\ &\text{2 step:} \quad log(Q_{yvt}) = \beta \bar{x}_{yvt} + \alpha p_{vt} + \zeta log(N_{yt}) + \mu F_{yt} + \nu \hat{\eta} log(\hat{A^i_{yt}}) + f_y + f_v + f_t + \xi_{yvt} \end{aligned} \tag{12}$$

Where F_{yt} is a dummy variable equal to one for the lead 0 and lead 1 of the first fight in the area. To see why chose to include also the first period after the outbreak in the estimation, let's explore the dynamic effect of fights on drug seizures. Figure A25 presents the event study graph.[15] There is no effect on the quantity of drug seized at the outbreak of a fight. There is seems to be some negative effect the period after the outbreak, but it is very small and not statistically different from zero. However, these results are the composition of the selection effect and the true effect on the drug market. Figure A26 replicates the event study controlling for selection. Here there is a clear negative trend after the fight outbreak, with the lead 1 coefficient being seizable and statistically different from zero. In other words, once we control for selection, we discover the negative effect of (increased) police attention on drug dealing. Moreover, this effect seems to be larger the period after the fight outbreak, which indicates a possible dynamic nature of this effect, justifying our modelling choice in the demand specification.

Parameter d is then recovered from the estimated fixed effect $\hat{f}_y$. Results of demand estimation are summarized in table A8.

Intermediate Step: Simulation parameters The last two steps of the estimation procedure require to simulate the model. To do so, we need to recover the unobserved taste preferences ξ_{yvt}. We assume that they follow a normal distribution $\xi_{yvt} \sim N(\bar{\xi}, \sigma_\xi)$ in the population. Then, we estimate residuals from the across nests log-lin demand function, and then compute the average

[15] In this figure, as well as in figure A26 we plot the event study, not controlling for quality, price, and number of active gangs to explore the effect on the whole sample; figures A27 and A28 controls for price and number of active gangs. Results are very similar. Quality is included as control in the main regression table A8.

and variance of the residuals to estimate the moments of the distribution. Given these moments, we simulate the unobserved taste preferences.

Third Step: Production Parameters We estimate the unobserved marginal cost θ with simulated method of moments, using data on average drug quality. To do so, we use the unconditional demand equation (appendix A.1) and the gang quality FOC. Simulation procedure detailed in appendix A.8.

The estimation targets the average drug quality. Given demand elasticities and shares, the average drug quality identify θ. This becomes clear when taking the average of the quality FOC across areas, time, drugs, and gangs. Figure A29 shows the value of the objective function (the absolute difference between the average observed quality and the predicted one) over the grid for θ. The objective function is clearly concave, with a minimum around -7.75.

Fourth Step: Fighting Parameters To conclude, fighting parameters $\{\psi, \phi_l, \Phi, \sigma_\epsilon\}$ are estimated via simulated method of moments, using data on fighting outcomes and decisions. In particular, we simulate the fighing process, described in section 4.6, of realized fights to match the realized outcomes. Then, we simulate the first stage to predict the optimal number of fights.

The estimation targets the following moments: increased probability of winning the fight if large, average probability of the fight ending in a draw, the variance in fighting outcomes, and the total number of fights. The increased probability of winning the fight for large gangs (similar to fact 2, table A5) identifies the higher fighting ability of large gangs ϕ_l. Given the scale of the gangs involved in the fight, the average probability of the fight ending in a draw identifies the scale of the fighting process Φ. Given scale and step, higher variance in fight outcomes indicates greater unpredictability in the fighting mechanism. Therefore, the variance in the probability of a gang winning a fight identifies σ_ϵ. Finally, the number of optimal fights identifies the shrinkage factor of profits under fights ψ. Notice that the identification of the two costs of fights (μ and ψ), comes from two separate type of variations. In particular, we identify μ using variations in quantity seized and ψ with the optimal number of fights. This allows us to separately identify these two different costs stemming from fights.

Estimation results Table 2 shows the results of the estimation for police reaction, demand, production, fighting process, as well as the other parameters. Here, we discuss the magnitude of the estimated parameters.

In line with the evidence presented in fact 3, Figure 6, we find a heterogeneous police reaction

Table 2: Estimated parameters

	Parameter	Estimate	Std Error	Elasticity
Panel A: Police reaction				
Police reaction (small)	λ_s	0.023	0.020	53%
Police reaction (large)	λ_l	0.053	0.030	250%
Spillover coefficient	δ	0.020	0.016	33%
Panel B: Demand				
(inverse of) Quality	β	-0.000391	0.00084	-0.32%
Price	α	-0.000507	0.00224	2.5%
Police dis-utility	μ	-1.5132	1.4130	-4%
Trasportation cost	d	-0.0006	0.0006	-2.4%
Average unobserved taste preference	$\bar{\xi}$	0		
SD unobserved taste preference	σ_ξ	0.0730		
Panel C: Production				
Unobserved marginal cost	θ	-7.7		
Panel D: Fighting				
Fighting ability large	ϕ_l	1.7		
Threshold fight	Φ	1.4		
Dispersion shock	σ_ϵ	0.63		
Shrinkage factor	ψ	0.87		
Panel E: Others				
Dispersion quality	σ_ζ	11.065	1.750	

Notes:

function. Specifically, the elasticity of police attention to fights is higher when the fight involves at least one large gang. Notice that coefficients in table 2 are those in table A7 divided by 0.3–the maximum possible value of police effectiviness parameter γ, hence being a lower bound. For additional details, please refer to section A.6. To put this difference in perspective, an average fight involving only small gangs increases police attention by approximately 53%–(0.023-0.015)/0.015, the average share of arrests by local police. However, when the fight involves at least one large gang, this figure rises to approximately 250%. Consistent with the evidence presented in Figures A17 and A18, we find a positive spillover of 0.020. This means that an average fight in another area, involving gangs present in the area, increases police attention by approximately 33%.

Panel B of Table 2 presents estimated parameters for the demand function. Consistent with fact 1, we find a positive coefficient for quality preference. Notice that this is the coefficient for the *inverse* of quality. The magnitude is quite small, with a 1% increase in percentage of pure substance being associated with a 0.32% increase in demand. Price elasticity is negative and approximately 2.5%–a value very close to other papers estimating demand for illegal drugs (Jacobi & Sovinsky, 2016 estimate a -2.1% price elasticity). As expected, higher average distance of the area is associated with decreased quantity sold. Specifically, a 1% increase in km is associated with approximately 2.4% decrease in demand. To conclude, we turn to the main parameter of interest:

the cost of police attention. Notice that this is the coefficient estimated in table A8 divided by the average police reaction. This elasticity is quite large, with a 1% increase in police attention being associated with approximately a 4% decrease in demand–approximately equivalent to a 1.67% increase in price. As a result, a fight involving a small gang is associate with a 6% decrease in demand, while a fight involving a large gang with a 14%.

Finally, in Panel D we examine parameters governing the fighting process. In line with fact 2, the estimated fighting ability of large gangs is 70% higher than that of smaller gangs, associated with a matched 8.4% increase in the probability of winning a fight. The estimated threshold fight is 1.4; importantly, this exceeds the largest difference between fighting abilities (0.7). This aligns with the matched 71% probability of a fight ending in a draw. The estimated fighting shock is quite high–0.63–consistent with high unpredictability in fighting outcomes. To conclude, the estimated shrinkage factor is equal to 0.87, implying a 13% reduction in profits–unrelated to demand variations–due to fights.

Model fit Figures A30 and A31 test the model fit by comparing, across areas and years, (i) the observed drug quantity seized with that predicted by the model, and (ii) the observed average quality with that predicted by the model. The model fit is good, with a strong positive correlation in both outcomes. In other words, in areas/years/drugs in which we observe an higher seized quantity and quality, the model predicts higher quantities sold and higher average drug purity. The correlation coeffiicent is statistically different from zero at any conventional level. Therefore, this high correlation provides reassurance about the model's ability to capture retail drug market dynamics. We can thus proceed to fights predictions and counterfactuals.

6 Counterfactuals

Table 3 shows the fighting propensity, computed separately for small and large gangs in 2022. in the data, as simulated by the model, and across the different counterfactual scenarios.[16] This is:

$$p_j = \frac{\sum_{oy} F_{joy}}{\sum_{oy} |\mathcal{Y}_o \cap \mathcal{Y}_j|}$$

In the data, we observe 16 fights in 2022. Consistent with the evidence presented in section 3, the fighting propensity is significantly lower for larger gangs with respect to smaller ones (-71%). This result remains unchanged when excluding wholesalers from the computation of fights, as defined in section 2.

[16] The computational time to simulate the model is quite high. Therefore, for the moment, we simulate only the last period in the sample.

Table 3: Fights

	Small	Large	Perc. change
Panl A: Data			
Including wholesalers (2022)	0.0159	0.0046	-71%
Excluding wholesalers (2022)	0.0159	0.0046	-71%
Panel B: Simulation			
Simulation model (2022)	0.0187	0.0072	-61%
Panel C: Counterfactuals			
Homogeous fighiting ability (2022)	0.0211	0.0155	-27%
No attention spillover (2022)	0.0200	0.0101	-49%
Homogenous police reaction (2022)	0.0187	0.0254	+35%

Notes:

Panel B presents the fights simulated by the model in 2022. The fighting propensities simulated by the model are very similar to the ones observed in the data. Therefore, our results indicate a -61% fighting propensity for large gangs with respect to small ones. In other words, the results from the model simulation align with what we observe in the data: a lower fighting propensity among large gangs.

Finally, we can investigate the mechanism behind these results by running counterfactual simulations. Panel C presents the results of three simulations. First, we assume homogeneous fighting ability. Formally, we set $\phi_l = \phi_s = 1$. This change has two distinct effects. On the one hand, it decreases the probability of large gangs winning fights against small ones, which should decrease the expected benefits of fighting for these gangs. On the other hand, it also decreases the probability of smaller gangs losing against large ones, which increases the expected profit of fighting for smaller gangs. If the mechanism behind our results was a *deterrence* one, where smaller gangs choose to avoid confrontation with larger ones, then this counterfactual scenario should mute this effect, and we would observe a higher fighting propensity among larger gangs. However, this is not the case. Instead, under this counterfactual scenario, we do observe a lower percentage difference (from -61% to -27%), but this is still lower than zero. In other words, also when gangs have homogeneous fighting ability, and thus the same probability of winning fights, large gangs still have a lower fighting propensity.

Second, we turn to the violence *spillover*. The existence of violence spillovers (figures A17, A18) may increase the cost of fighting for larger gangs, defined in the paper as gangs operating in

multiple areas. To test this potential mechanism, we counterfactually set the violence spillover to zero ($\delta = 0$) and simulate the model. In line with expectations, we observe an increase in the fighiting propensity predicted by the model for both categories of gangs. Moreover, this increase is higher for larger gangs, consistent with our theoretical reasoning. Nevertheless, this is not sufficient to fully explain the regularity: we still observe a lower fighting propensity (-49%) involving larger gangs. As a result, although violence spillover may be a potential mechanism behind the negative correlation between gang size and violence, it is probably not the main one.

To conclude, we examine the counterfactual on the *heterogeneous police reaction function*. Formally, we set $\lambda_l = \lambda_s = 0.023$. In other words, in this scenario we are counterfactually equalizing the police attention attracted–and therefore the resulting effect on demand–by fights between large and small gangs. The predicted difference in fighting propensity is now reversed: in this counterfactual scenario we observe a +35% higher number of fights involving larger gangs than smaller ones. Hence, when fights between gangs of different scale attract the same amount of attention, larger gangs tend to fight more, most likely because of their higher probability of winning conflicts. Therefore, we can conclude that the heterogeneous police reaction function is an important mechanism behind the negative correlation we observe in the data between gang scale and violence. Because a fight involving larger gangs attracts more attention, larger gangs face a diseconomy of scale in the fighting mechanism, which in turn substantially increases the costs associated with fights for this category of gangs.

7 Optimal police reaction

Given the importance of police reaction to determe fighting incentives of gangs, one question arises naturally: what is the optimal reaction function? What is the optimal set of reaction parameters $\{\lambda_l, \lambda_s\}$? In other words, while previously we presented police as an exogenous function, we now endogenize it, modelling the decision behind the reaction parameters.

To talk about optimality, we first need to specify a police objective function. One clear police objective is to minimize fights, and therefore violence. Hence, the optimal police reaction function is determined by:

$$\min_{\lambda_s, \lambda_l} \sum_{yj} F_{yj} \quad \text{s.t.} \quad \lambda_s + \lambda_l \leq 1$$

However, also alternative objective functions could be considered. Say, for example, that the police objective is to reduce fights and, at the same time, reduce the dimension of gangs. Then,

the optimal reaction function would be:

$$\min_{\lambda_s,\lambda_l} \alpha \sum_{ytj} F_{ytj} + (1-\alpha) \sum_{j} m_j \quad \text{s.t.} \quad \lambda_s + \lambda_l \leq 1$$

where α indicates the weight of the first objective with respect to the other. Alternatively, the police may would like to minimize the size of the overall retail drug market. In this case, the optimal reaction function would be:

$$\min_{\lambda_s,\lambda_l} \alpha \sum_{ytj} F_{ytj} + (1-\alpha) \sum_{yjv} Q_{yvj} \quad \text{s.t.} \quad \lambda_s + \lambda_l \leq 1$$

For all these possibile objective functions, the optimal reaction is determined by simulating the model with different set of reaction coefficients, and explore how the objective function changes.

8 Conclusions

This paper examines the relationship between market structure and violence on the retail drug market in Naples province, Italy. To do so, we leverage a unique dataset on gang presence across areas, inter-gang conflicts, intelligence on homicides and police activities, and detailed information on approximately 14,000 drug seizures including chemical analyses of drug quality.

The data reveals three key patterns in the relationship between market structure and violence. First, violence increases with the number of active gangs in an area. Second, we find an inverse-U shaped relationship between violence and the standard deviation of gang size in the area. Third, there is a negative relationship between violence and the average size of gangs in the area. To further explore the possibility of a lower fighting propensity of large gangs, we investigate how variation in gang size, due to the entry of new gangs, creates marginal changes in area violence using an event-study design. Our results indicate a lower violence propensity of large gangs.

To understand the underlying mechanisms, we develop a quantitative model of an oligopoly market in which gangs decide whether to engange in conflict or not, trading off the expected additional profits from a higher market share in the area with the cost deriving from a higher presence of police. Counterfactual scenarios show that, due to an heterogenous police reaction, larger gangs face a diseconomy of scale in the fighting process, generating the negative correlation we observe in the data. The optimal police reaction function is then determined.

References

Acconcia, A., Corsetti, G., & Simonelli, S. (2014). Mafia and public spending: Evidence on the fiscal multiplier from a quasi-experiment. *American Economic Review*, *104*(7), 2185–2209.

Acemoglu, D., Robinson, J. A., & Santos, R. J. (2013). The monopoly of violence: Evidence from colombia. *Journal of the European Economic Association*, *11*(suppl_1), 5–44.

Aguirregabiria, V., & Mira, P. (2007). Sequential estimation of dynamic discrete games. *Econometrica*, *75*(1), 1–53.

Aguirregabiria, V., & Vicentini, G. (2016). Dynamic spatial competition between multi-store retailers. *The Journal of Industrial Economics*, *64*(4), 710–754.

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

Barbagallo, F. (2014). *Storia della camorra*. Gius. Laterza & Figli Spa.

Barone, G., & Narciso, G. (2015). Organized crime and business subsidies: Where does the money go? *Journal of Urban Economics*, *86*, 98–110.

Barton, J., Castillo, M., & Petrie, R. (2016). Negative campaigning, fundraising, and voter turnout: A field experiment. *Journal of Economic Behavior & Organization*, *121*, 99–113.

Bass, F. M., Krishnamoorthy, A., Prasad, A., & Sethi, S. P. (2005). Generic and brand advertising strategies in a dynamic duopoly. *Marketing Science*, *24*(4), 556–568.

Battiston, G., Daniele, G., Le Moglie, M., & Pinotti, P. (2022). Fueling organized crime: The mexican war on drugs and oil thefts.

Becker, G. S. (1968). Crime and punishment: An economic approach. *The economic dimensions of crime/Springer*.

Becker, G. S., Murphy, K. M., & Grossman, M. (2006). The market for illegal goods: the case of drugs. *Journal of Political Economy*, *114*(1), 38–60.

Becucci, S. (2004). Old and new actors in the italian drug trade: Ethnic succession or functional specialization? *European Journal on Criminal Policy and Research*, *10*, 257–283.

Berry, S. T. (1992). Estimation of a model of entry in the airline industry. *Econometrica: Journal of the Econometric Society*, 889–917.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 242–262.

Berry, S. T., Levinsohn, J. A., & Pakes, A. (1993). *Automobile prices in market equilibrium: Part i and ii.* National Bureau of Economic Research Cambridge, Mass., USA.

Blattman, C., Duncan, G., Lessing, B., & Tobón, S. (2021). *Gang rule: Understanding and countering criminal governance* (Tech. Rep.). National Bureau of Economic Research.

Boncompte, J., & Zufacchi, D. (2025). Legal competition.

Bostanci, G., Yildirim, P., & Jerath, K. (2021). Negative advertising and competitive product positioning.

Brancaccio, L., & Martone, V. (2019). The camorras in naples and campania: Business, groups and families. In *Italian mafias today* (pp. 30–46). Edward Elgar Publishing.

Bresnahan, T. F., & Reiss, P. C. (1991). Empirical models of discrete games. *Journal of Econometrics*, *48*(1-2), 57–81.

Catino, M. (2014). How do mafias organize?: Conflict and violence in three mafia organizations. *European Journal of Sociology/Archives Europeennes de Sociologie*, *55*(2), 177–220.

Catino, M. (2019). *Mafia organizations*. Cambridge University Press.

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

Ciliberto, F., & Tamer, E. (2009). Market structure and multiple equilibria in airline markets. *Econometrica*, *77*(6), 1791–1828.

Coate, S. (2004). Political competition with campaign contributions and informative advertising. *Journal of the European Economic Association*, *2*(5), 772–804.

Crawford, G. S., Shcherbakov, O., & Shum, M. (2019). Quality overprovision in cable television markets. *American Economic Review*, *109*(3), 956–995.

Daniele, V., & Marani, U. (2011). Organized crime, the quality of local institutions and fdi in italy: A panel data analysis. *European Journal of Political Economy*, *27*(1), 132–142.

DCSA. (2024). *Relazione annuale direzione centrale servizi antridroga 2024* (Tech. Rep.).

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

DIA. (2015). *Relazione del ministreo dell'interno al parlamento: attività svolta e risultati conseguiti dalla direzione investigativa antimafia* (Tech. Rep.).

DIA. (2020). *Relazione del ministreo dell'interno al parlamento: attività svolta e risultati conseguiti dalla direzione investigativa antimafia* (Tech. Rep.).

DIA. (2022). *Relazione del ministreo dell'interno al parlamento: attività svolta e risultati conseguiti dalla direzione investigativa antimafia* (Tech. Rep.).

Di Cataldo, M., & Mastrorocco, N. (2022). Organized crime, captured politicians, and the allocation of public resources. *The Journal of Law, Economics, and Organization*, *38*(3), 774–839.

Dobkin, C., & Nicosia, N. (2009). The war on drugs: methamphetamine, public health, and crime. *American Economic Review*, *99*(1), 324–349.

Donohue, J. J., & Levitt, S. D. (1998). Guns, violence, and the efficiency of illegal markets. *The American Economic Review*, *88*(2), 463–467.

Dugato, M., Calderoni, F., & Berlusconi, G. (2020). Forecasting organized crime homicides: Risk terrain modeling of camorra violence in naples, italy. *Journal of interpersonal violence*, *35*(19-20), 4013–4039.

EMCDDA. (2013). *European monitoring centre for drugs and drug addiction: Drug report* (Tech. Rep.). European Monitoring Centre for Drugs and Drug Addiction.

Eurispes. (2005). *Rapporto italia 2005* (Tech. Rep.).

Fenizia, A., & Saggio, R. (2024). *Organized crime and economic growth: Evidence from municipalities infiltrated by the mafia* (Tech. Rep.). National Bureau of Economic Research.

Galenianos, M., & Gavazza, A. (2017). A structural model of the retail market for illicit drugs. *American Economic Review*, *107*(3), 858–896.

Galenianos, M., Pacula, R. L., & Persico, N. (2012). A search-theoretic model of the retail market for illicit drugs. *The Review of Economic Studies*, *79*(3), 1239–1269.

Gambetta, D. (1996). *The sicilian mafia: The business of private protection*.

Gambetta, D., & Reuter, P. (2017). Conspiracy among the many: the mafia in legitimate industries. In *Transnational organized crime* (pp. 247–266). Routledge.

Goldberg, P. K. (1995). Product differentiation and oligopoly in international markets: The case of the us automobile industry. *Econometrica: Journal of the Econometric Society*, 891–951.

Gribaudi, M. G., et al. (2009). Clan camorristi a napoli: radicamento locale e traffici internazionali. In *Traffici criminali. camorra, mafie e reti internazionali dell'illegalità* (pp. 187–240). Bollati Boringhieri.

Grossman, G. M., & Shapiro, C. (1984). Informative advertising with differentiated products. *The Review of Economic Studies*, *51*(1), 63–81.

Heckman, J. (1990). Varieties of selection bias. *The American Economic Review*, *80*(2), 313–318.

Holmes, T. J. (2011). The diffusion of wal-mart and economies of density. *Econometrica*, *79*(1), 253–302.

Holmstrom, B. R., & Tirole, J. (1989). The theory of the firm. *Handbook of industrial organization*, *1*, 61–133.

Hotelling, H. (1929). Stability in competition. *The Economic Journal*.

Houde, J.-F. (2009). Spatial differentiation and vertical contracts in retail markets for gasoline. *Available at SSRN 1417506*.

Iyer, G., & Singh, S. (2022). Persuasion contest: Disclosing own and rival information. *Marketing Science*, *41*(4), 682–709.

Jacobi, L., & Sovinsky, M. (2016). Marijuana on main street? estimating demand in markets with limited access. *American Economic Review*, *106*(8), 2009–2045.

Jia, P. (2008). What happens when wal-mart comes to town: An empirical analysis of the discount retailing industry. *Econometrica*, *76*(6), 1263–1316.

Kar, D., & Spanjers, J. (2017). Transnational crime and the developing world. *Global Financial Integrity*, 53–59.

Le Moglie, M., & Sorrenti, G. (2020). Revealing maiia inc. financial crisis, oc, and the birth oi new enterprises. *The Review of Economics and Statistics, pp. 145*.

Lessing, B., & Willis, G. D. (2019). Legitimacy in criminal governance: Managing a drug empire from behind bars. *American Political Science Review*, *113*(2), 584–606.

Levitt, S. D. (2017). The economics of crime. *Journal of Political Economy*, *125*(6), 1920–1925.

Levitt, S. D., & Miles, T. J. (2006). Economic contributions to the understanding of crime. *Annu. Rev. Law Soc. Sci.*, *2*(1), 147–164.

Levitt, S. D., & Venkatesh, S. A. (2000). An economic analysis of a drug-selling gang's finances. *The quarterly journal of economics*, *115*(3), 755–789.

Marrazzo, G. (2005). *Il camorrista*. Tullio Pironti Editore.

Mastrobuoni, G., & Patacchini, E. (2012). Organized crime networks: An application of network analysis techniques to the american mafia. *Review of Network Economics*, *11*(3).

Mazzeo, M. J. (2002). Product choice and oligopoly market structure. *RAND Journal of Economics*, 221–242.

Melnikov, N., Schmidt-Padilla, C., & Sviatschi, M. M. (2020). *Gangs, labor mobility and development* (Tech. Rep.). National Bureau of Economic Research.

Miller, N. H., & Weinberg, M. C. (2017). Understanding the price effects of the millercoors joint venture. *Econometrica*, *85*(6), 1763–1791.

Mirenda, L., Mocetti, S., & Rizzica, L. (2022). The economic effects of mafia: firm level evidence. *American Economic Review*, *112*(8), 2748–2773.

Mizuno, T. (2003). On the existence of a unique price equilibrium for models of product differentiation. *International Journal of Industrial Organization*, *21*(6), 761–793.

Monnier, M. (1862). *La camorra notizie storiche raccolte e documentate*. G. Barbera.

Morgenthau, H. J. (1973). *Politics among nations*.

Murry, C., & Zhou, Y. (2020). Consumer search and automobile dealer colocation. *Management Science*, *66*(5), 1909–1934.

Pinotti, P. (2011). The economic consequences of organized crime: Evidence from southern italy. *Bank of Italy*.

Pinotti, P. (2015). The economic costs of organised crime: Evidence from southern italy. *The Economic Journal*, *125*(586), F203–F232.

Polo, M. (1995). Internal cohesion and competition among criminal organizations. *The economics of organised crime*, 87–115.

Reed, W. (2003). Information, power, and war. *American Political Science Review*, *97*(4), 633–641.

Reiss, P., & Bresnahan, T. (1990). Entry in monopoly markets. *Review of Economic Studies*, *57*(4), 531–53.

Reuter, P. (1985). *The organization of illegal markets: An economic analysis* (Vol. 84) (No. 9). US Department of Justice, National Institute of Justice.

Reuter, P. (2014). Drug markets and organized crime. *The Oxford handbook of organized crime*, 359–381.

Sales, I. (1994). *La camorra, le camorre*. Editori Riuniti.

Sales, I. (2022). *Storia delle camorre: passato e presente*. Rubbettino.

Schelling, T. C. (1967). Economics and criminal enterprise. *The Public Interest*, *7*, 61.

Schelling, T. C. (1971). What is the business of organized crime? *The American Scholar*, 643–652.

Schultz, C. (2007). Strategic campaigns and redistributive politics. *The Economic Journal*, *117*(522), 936–963.

Transcrime, C. (2013). Progetto pon sicurezza 2007–2013. gli investimenti delle mafie. *Rapporto Linea*, *1*.

Volkov, V. (1999). Violent entrepreneurship in post-communist russia. *Europe-Asia Studies*, *51*(5), 741–754.

Appendix

A Proofs and Derivations

A.1 Derivation demand

Consumer problem Consumer solves:

$$\begin{array}{c} \max_{d_{(y,j)}} \sum_{(y,j)} d_{(y,j)} \mathbb{E}u_{ijyvt} \quad \forall v,t \\ \text{s.t.} \sum_{(y,j)} d_{(y,j)} = 1 \end{array}$$

Since in this model an individual is defined by product specific shocks $(\epsilon_{ijvt}, \epsilon_{iyvt})$, this implictly defines the set of individual attributes that lead to the purchase of good (j, y). Formally:

$$A_{jy} = \{(\epsilon_{ijvt}, \epsilon_{iyvt}) | \mathbb{E}u_{ijyvt} \geq \mathbb{E}u_{i\tilde{j}\tilde{y}vt} \forall \tilde{j} \in \mathcal{J} \text{ and } \forall \tilde{y} \in \mathcal{Y}\}$$

Assuming ties occur with zero probability, and the iid nature of the two EVT1 shocks, the market share of the (j, y) bundle is just an integral over the mass of consumers in region A_{jyv}:

$$s_{jyvt} = \int_{A_{jy}} dP(\epsilon_{ijvt}, \epsilon_{iyvt}) =^{iid} \int_{A_{jy}} dP(\epsilon_{ijvt}) P(\epsilon_{iyvt})$$

We will now work to find an explicit expression definying the set A_{jy}. In particular, we will exploit the nested structure of the model. Notice that we can decompose:

$$P(ivt \in A_{jy}) = P_{ivt \to jy} = P_{ivt \to j|y} \times P_{ivt \to y}$$

In the remaining of this section we will derive an expression for these two probabilities.

Within nests Let's start with $P_{ivt \to j|y}$. A consumer in area y chooses the organization which gives him the highest utility. Notice that here the consumer has full information about drug quality in the area. Formally:

$$\begin{array}{c} P_{ivt \to j|y} = P(u_{ijyvt} \geq u_{i\tilde{j}yvt} \, \forall \tilde{j} \in \mathcal{J}_y) = \\ = P(\delta_{jyvt} + d_v g_{iy} + \epsilon_{ijvt} + \epsilon_{ijvt} \geq \delta_{\tilde{j}yvt} + d_v g_{iy} + \epsilon_{i\tilde{j}vt} + \epsilon_{iyvt} \, \forall \tilde{j} \in \mathcal{J}_y) \\ = P(\epsilon_{i\tilde{j}vt} \leq \epsilon_{ijvt} + \delta_{jyvt} - \delta_{\tilde{j}yvt} \, \forall \tilde{j} \in \mathcal{J}_y) \end{array}$$

Since the ϵ_s are indipendent, this set comulative distribution can be rewritten as the product of the individual comulative distributions:

$$P_{ivt \to j|y} | \epsilon_{ijvt} = \prod_{\tilde{j} \in y - j} \exp[-\exp[\epsilon_{ijvt} + \delta_{jyvt} - \delta_{\tilde{j}yvt}]]$$

The problem with the expression is that we don't know ϵ_{ijvt}. So we integrate it away:

$$P_{ivt\to j|y} = \int_{-\infty}^{+\infty} [\prod_{\tilde{j}\in y-j} e^{-e^{-[\epsilon_{ijvt}+\delta_{jyvt}-\delta_{\tilde{j}yvt}]}}] e^{-\epsilon_{ijvt}} e^{e^{-\epsilon_{ijvt}}} d\epsilon_{ijvt}$$

Since $\delta_{jyvt} - \delta_{jyvt} = 0$ we can rewrite the integral as:

$$\begin{aligned} P_{ivt\to j|y} &= \int_{-\infty}^{+\infty} [\prod_{\tilde{j}\in y} e^{-e^{-[\epsilon_{ijvt}+\delta_{jyvt}-\delta_{\tilde{j}yvt}]}}] e^{-\epsilon_{ijvt}} d\epsilon_{ijvt} = \\ &= \int_{-\infty}^{+\infty} [\prod_{\tilde{j}\in y} e^{-e^{-\epsilon_{ijvt}} e^{-[\delta_{ijyvt}-\delta_{i\tilde{j}yvt}]}}] e^{-\epsilon_{ijvt}} d\epsilon_{ijvt} \end{aligned}$$

Define $t = e^{\epsilon_{ijvt}}$ and let's change variable with $d\epsilon_{ijvt} = \frac{1}{t} dt$:

$$\begin{aligned} P_{ivt\to j|y} &= \int_0^{+\infty} e^{-t[\sum_{\tilde{j}\in y} e^{-[\delta_{ijyvt}-\delta_{i\tilde{j}yvt}]}]} dt = \\ &= [-\frac{e^{-t[\sum_{\tilde{j}\in y} e^{-[\delta_{ijyvt}-\delta_{i\tilde{j}yvt}]}]}}{\sum_{\tilde{j}\in y} e^{-[\delta_{ijyvt}-\delta_{i\tilde{j}yvt}]}}]_0^{\infty} = \\ &= 0 + \frac{1}{\sum_{\tilde{j}\in y} e^{-[\delta_{ijyvt}-\delta_{i\tilde{j}yvt}]}} = \frac{\exp(\delta_{jyvt})}{\sum_{\tilde{j}\in y} \exp(\delta_{\tilde{j}yvt})} \end{aligned}$$

Across nests Now we need to compute the second probability: $P_{ivt\to y}$. Here the consumer face the uncertainty about drug quality in the area. Hence, instead of only taking the expectation with respect to its own taste shock, he takes also the expectaction about drug quality uncertainty. Notice that the maximum expected utility from nest/area y is:

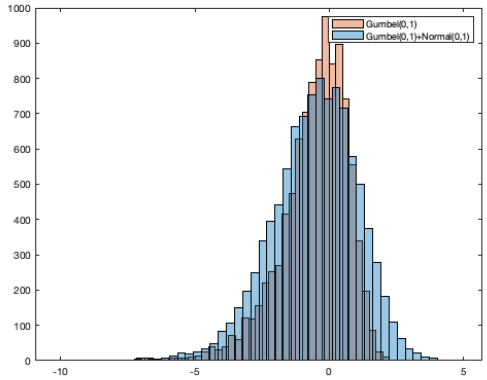
$$\begin{cases} \mathbb{E}_{\epsilon_{jyvt},\zeta_{jtvy}}(\max_{j\in y} u_{ijyvt}) & y \neq 0 \\ 1 & y = 0 \end{cases}$$

where the second case comes from the normalization of expected utility of outside option. Substituting our equation for the indirect utility:

$$\begin{aligned} &\mathbb{E}_{\epsilon_{jyvt},\zeta_{jtvy}}(\max_{j\in y} \delta_{jyvt} + \epsilon_{iyvt} + \epsilon_{ijvt} + d_v g_{iy}) = \\ &= \mathbb{E}_{\epsilon_{jyvt},\zeta_{jtvy}}(\max_{j\in y} x_{jvty}\beta_v + \alpha_v p_{vt} + \mu_v O_{yt} + \xi_{yvt} + \epsilon_{ijvt}) + \epsilon_{iyvt} + d_v g_{iy} = \\ &= \mathbb{E}_{\epsilon_{jyvt},\zeta_{jtvy}}(\max_{j\in y} x_{jvty}\beta_v + \epsilon_{ijvt}) + \mu_v O_{yt} + \alpha_v p_{vt} + \xi_{yvt} + \epsilon_{iyvt} + d_v g_{iy} = \\ &= \mathbb{E}_{\epsilon_{jyvt},\zeta_{jtvy}}(\max_{j\in y} \zeta_{jtvy}\beta_v + \epsilon_{ijvt}) + \bar{x}_{yvt}\beta_v + \mu_v O_{yt} + \alpha_v p_{vt} + \xi_{yvt} + \epsilon_{iyvt} + d_v g_{iy} \end{aligned}$$

Since $\epsilon_{ijvt} \sim \text{ETV1}(0,1)$ and $\zeta_{jtv} \sim N(0, {}^{\zeta}\sigma_g^2)$, then $S_{ijvt} = \zeta_{jtvy}\beta_v + \epsilon_{ijvt}$ is approximately distributed like an ETV1 with some scaling (${}^{\zeta}\sigma_g^2$). This can be seen in the following simulation of the two distributions: Hence we have:

- $\mathbb{E}(\zeta_{jtvy}\beta_v) = \beta_v 0 = 0$
- $\mathbb{E}(\epsilon_{ijvt}) = \gamma$



- $\mathbb{E}(S_{ijvt}) = \gamma$

where γ is the Euler-Mascheroni constant. Moreover, due to the extreme value properties of the ETV1 component we have (approximation of the expectation of the maximum of an ETV1 distribution):

$$\mathbb{E}(\max_{j \in y} S_{ijvt}) \approx \mathbb{E}(S_{ijvt}) + log(N_{yt})\sigma_{S_{ijvt}}$$

where N_y is the number of groups in the area. To conclude, given independence of the two summed RVs, we have:

$$\mathbb{E}(\max_{j \in y} S_{ijvt}) \approx \gamma + \sqrt{\zeta\sigma_g^2 + 1}log(N_{yt})$$

Substituting into the previous problem:

$$\mathbb{E}_{\epsilon_{jyvt},\zeta_{jtvy}}(\max_{j \in y} u_{ijyvt}) \approx \bar{x}_{yvt}\beta_v + \alpha_v p_{vt} + \xi_{yvt} + \epsilon_{iyvt} + d_v g_{iy} + \gamma + \sqrt{\zeta\sigma_g^2 + 1}log(N_{yt})$$

Hence the maximum expected indirect utility is:

$$V_{iyvt} = \bar{\delta}_{yvt} + \epsilon_{iyvt} + d_v g_{iy} + \gamma + \sqrt{\zeta\sigma_g^2 + 1}log(N_{yt})$$

where $\bar{\delta}_{yvt} = \bar{x}_{yvt}\beta_v + \mu_v O_{yt} + \alpha_v p_{vt} + \xi_{yvt}$ is the average consumer valuation in the area. Similarly to the reasoning done before, the consumer chooses the nest which gives him the highest expected utility:

$$P_{ivt \to y} = P(V_{iyvt} \geq V_{i\tilde{y}vt}\ \forall \tilde{y} = 0, 1, ..., Y)$$

Hence:

$$P_{ivt\to y} = P(\bar{\delta}_{yvt} + \epsilon_{iyvt} + d_v g_{iy} + \gamma + \sqrt{\zeta\sigma_g^2 + 1} log(N_y) \geq$$
$$\bar{\delta}_{\tilde{y}vt} + \epsilon_{i\tilde{y}vt} + d_v g_{i\tilde{y}} + \gamma + \sqrt{\zeta\sigma_g^2 + 1} log(N_{\tilde{y}}) \forall \tilde{y} \in \mathcal{Y} =$$
$$= P(\epsilon_{i\tilde{y}vt} \leq \epsilon_{iyvt} + \bar{\delta}_{yvt} - \bar{\delta}_{\tilde{y}vt} + d_v g_{iy} - d_v g_{i\tilde{y}} + \sqrt{\zeta\sigma_g^2 + 1} log(N_{yt}) - \sqrt{\zeta\sigma_g^2 + 1} log(N_{\tilde{y}t}))$$

To simplify notation, let's denote: $M_{yvt} = \bar{\delta}_{yvt} + d_v g_{iy} + \sqrt{\zeta\sigma_g^2 + 1} log(N_{yt})$. Therefore:

$$P_{ivt\to y} | \epsilon_{iyvt} = \prod_{\tilde{y} \in \mathcal{Y}/y} \exp(-\exp(M_{yvt} - M_{\tilde{y}vt}))$$

As before, the problem with the expression is that we don't know ϵ_{iyvt}. So we integrate it away:

$$P_{ivt\to y} = \int_{-\infty}^{+\infty} [\prod_{\tilde{y} \in \mathcal{Y}/y} e^{-e^{-[\epsilon_{iyvt} + M_{yvt} - M_{\tilde{y}vt}]}}] e^{-\epsilon_{iyvt}} e^{e^{-\epsilon_{iyvt}}} d\epsilon_{iyvt}$$

Since $M_{yvt} - M_{yvt} = 0$ we can rewrite the integral as:

$$P_{ivt\to y} = \int_{-\infty}^{+\infty} [\prod_{\tilde{y} \in \mathcal{Y}} e^{-e^{-[\epsilon_{iyvt} + M_{yvt} - M_{\tilde{y}vt}]}}] e^{-\epsilon_{iyvt}} d\epsilon_{iyvt} =$$
$$= \int_{-\infty}^{+\infty} [\prod_{\tilde{y} \in \mathcal{Y}} e^{-e^{-\epsilon_{iyvt}} e^{-[M_{yvt} - M_{\tilde{y}vt}]}}] e^{-\epsilon_{iyvt}} d\epsilon_{iyvt}$$

Define $t = e^{\epsilon_{iyvt}}$ and let's change variable with $d\epsilon_{iyvt} = \frac{1}{t} dt$:

$$P_{ivt\to y} = \int_0^{+\infty} e^{-t[\sum_{\tilde{y} \in \mathcal{Y}} e^{-[M_{yvt} - M_{\tilde{y}vt}]}]} dt = = [-\frac{e^{-t[\sum_{\tilde{y} \in \mathcal{Y}} e^{-[M_{yvt} - M_{\tilde{y}vt}]}]}}{\sum_{\tilde{y} \in \mathcal{Y}} e^{-[M_{yvt} - M_{\tilde{y}vt}]}}]_0^{\infty} =$$
$$= 0 + \frac{1}{\sum_{\tilde{y} \in \mathcal{Y}} e^{-[M_{yvt} - M_{\tilde{y}vt}]}} = \frac{\exp(M_{yvt})}{\sum_{\tilde{y} \in \mathcal{Y}} \exp(M_{\tilde{y}vt})} =$$
$$= \frac{\exp(\bar{\delta}_{yvt} + d_v g_{iy} + \sqrt{\zeta\sigma_g^2 + 1} log(N_{yt}))}{1 + \sum_{\tilde{y} \in \mathcal{Y}} \exp(\bar{\delta}_{\tilde{y}vt} + d_v g_{i\tilde{y}} + \sqrt{\zeta\sigma_g^2 + 1} log(N_{\tilde{y}t}))}$$

Before turning to the unconditional demand, we will now see how to transform this across nest share into an unconditional demand for drugs in area y. The unconditional demand for drugs in nest y can be written as weighted sum of the conditional demands:

$$s_{yvt} = \sum_{y_i} \frac{M_{y_i}}{\sum_y M_y} s_{yvt|y_i} = \sum_{y_i} \frac{M_{y_i}}{\sum_y M_y} \frac{\exp(\bar{\delta}_{yvt} + d_v g_{iy} + \sqrt{\zeta\sigma_g^2 + 1} log(N_{yt}))}{1 + \sum_{\tilde{y} \in \mathcal{Y}} \exp(\bar{\delta}_{\tilde{y}vt} + d_v g_{i\tilde{y}} + \sqrt{\zeta\sigma_g^2 + 1} log(N_{\tilde{y}t}))}$$

Denote by $S_{vt} = 1 + \sum_{\tilde{y} \in \mathcal{Y}} \exp(\bar{\delta}_{\tilde{y}vt} + d_v g_{i\tilde{y}} + \sqrt{\zeta\sigma_g^2 + 1} log(N_{\tilde{y}t}))$, and $Z_{yvt} = \bar{\delta}_{yvt} + \sqrt{\zeta\sigma_g^2 + 1} log(N_{yt})$. Taking logs:

$$log(s_{yvt}) = log(\sum_{y_i} \frac{M_{y_i}}{\sum_y M_y} s_{yvt|y_i}) = log(\sum_{y_i} \frac{M_{y_i}}{\sum_y M_y} \frac{\exp(Z_{yvt} + d_v g_{iy})}{S_{vt}}) =$$
$$= log(\frac{1}{S_{vt}} \sum_{y_i} \frac{M_{y_i}}{\sum_y M_y} exp(Z_{yvt}) exp(d_v g_{iy})) =$$
$$= log(\frac{1}{S_{vt}} exp(Z_{yvt}) \sum_{y_i} \frac{M_{y_i}}{\sum_y M_y} exp(d_v g_{iy})) =$$
$$= -log(S_{vt}) + Z_{yvt} + log(\sum_{y_i} \frac{M_{y_i}}{\sum_y M_y} exp(d_v g_{iy})) =$$
$$= -log(S_{vt}) + \bar{\delta}_{yvt} + \sqrt{\zeta\sigma_g^2 + 1} log(N_{yt}) + log(\sum_{y_i} \frac{M_{y_i}}{\sum_y M_y} exp(d_v g_{iy}))$$

Which, by adding the fixed effects, becomes equation 3.

Unconditional demand Finally, we can put toghether the two probabilities to compute the probability of consumer i choosing the bundle (j, y) for drug v at t:

$$P(ivt \in A_{jy}) = P_{ivt \to jy} = \frac{\exp(\delta_{jyvt})}{\sum_{\tilde{j} \in y} \exp(\delta_{\tilde{j}yvt})} \times \frac{\exp(\bar{\delta}_{yvt} + d_v g_{iy} + \sqrt{\zeta \sigma_g^2 + 1} log(N_{yt}))}{1 + \sum_{\tilde{y} \in \mathcal{Y}} \exp(\bar{\delta}_{\tilde{y}vt} + d_v g_{i\tilde{y}} + \sqrt{\zeta \sigma_g^2 + 1} log(N_{\tilde{y}t}))}$$

Unfortunately, due to the uncertainty process, here we cannot simplify as in the standard BLP case, and demand is represented by this set of individual probabilities:

$$P_{ivt \to jy} = \frac{\exp(\delta_{jyvt}) \exp(\bar{\delta}_{yvt} + d_v g_{iy} + \sqrt{\zeta \sigma_g^2 + 1} log(N_{yt}))}{\sum_{\tilde{j} \in y} \exp(\delta_{\tilde{j}yvt}) [1 + \sum_{\tilde{y} \in \mathcal{Y}} \exp(\bar{\delta}_{\tilde{y}vt} + d_v g_{i\tilde{y}} + \sqrt{\zeta \sigma_g^2 + 1} log(N_{\tilde{y}t}))]} \tag{13}$$

To conclude, we need to compute the market shares. Returning to our first derivations we have:

$$s_{jyvt} = \int_{A_{jy}} dP(\epsilon_{ijvt}) P(\epsilon_{iyvt}) \approx$$
$$\sum_{y_i} \frac{M_{y_i}}{\sum_y M_y} \frac{\exp(\delta_{jyvt}) \exp(\bar{\delta}_{yvt} + d_v g_{iy} + \sqrt{\zeta \sigma_g^2 + 1} log(N_{yt}))}{\sum_{\tilde{j} \in y} \exp(\delta_{\tilde{j}yvt}) [1 + \sum_{\tilde{y} \in \mathcal{Y}} \exp(\bar{\delta}_{\tilde{y}vt} + d_v g_{i\tilde{y}} + \sqrt{\zeta \sigma_g^2 + 1} log(N_{\tilde{y}t}))]}$$

These individual buying probabilities can be used to compute the aggregate demanded quantities.

Denote by D_v the average annual amount of drug bought by an average consumer of drug v. Then:

$$q_{jyvt} \approx D_v \sum_{y_i} M_{y_i} \frac{\exp(\delta_{jyvt}) \exp(\bar{\delta}_{yvt} + d_v g_{iy} + \sqrt{\zeta \sigma_g^2 + 1} log(N_{yt}))}{\sum_{\tilde{j} \in y} \exp(\delta_{\tilde{j}yvt}) [1 + \sum_{\tilde{y} \in \mathcal{Y}} \exp(\bar{\delta}_{\tilde{y}vt} + d_v g_{i\tilde{y}} + \sqrt{\zeta \sigma_g^2 + 1} log(N_{\tilde{y}t}))]}$$

Hence, aggregate quantity demanded in area y, time t, of drug v is:

$$q_{yvt} \approx \sum_{j \in \mathcal{J}_y} D_v \sum_{y_i} M_{y_i} \frac{\exp(\delta_{jyvt}) \exp(\bar{\delta}_{yvt} + d_v g_{iy} + \sqrt{\zeta \sigma_g^2 + 1} log(N_{yt}))}{\sum_{\tilde{j} \in y} \exp(\delta_{\tilde{j}yvt}) [1 + \sum_{\tilde{y} \in \mathcal{Y}} \exp(\bar{\delta}_{\tilde{y}vt} + d_v g_{i\tilde{y}} + \sqrt{\zeta \sigma_g^2 + 1} log(N_{\tilde{y}t}))]}$$

A.2 Second stage

Profit maximization problem Suppose a network structure $\mathbf{Y}_t$ determined in the first stage. Profits of gang j in t are:

$$\sum_{y \in \mathcal{Y}_{jt}} \sum_v [p_{vt} q_{jyvt} - q_{jyvt} [s_{jyvt} k_{vt} + r_{jyvt} c_{vt} + \theta]]$$

Each gang maximize this by choosing its best-responce vector of quality and dealers given demand and production function:

$$\max_{q_{jyvt},\ell_{jyvt},r_{jyvt},x_{jyvt}\forall y\in\mathcal{Y}_{jt}} \quad \sum_{y\in\mathcal{Y}_{jt}}\sum_{v}[p_{vt}q_{jyvt} - q_{jyvt}[s_{jyvt}k_{vt} + r_{jyvt}c_{vt} + \theta]]$$

$$\text{s.t.} \quad q_{jyvt} = min\{\frac{s_{jyvt}}{1-x_{jyvt}}, \frac{r_{jyvt}}{x_{jyvt}}\}\ \forall y \in \mathcal{Y}_{jt}$$

$$q_{jyvt} \approx D_v \sum_{y_i} M_{y_i} \frac{\exp(\delta_{jyvt})\exp(\bar{\delta}_{yvt} + d_v g_{iy} + \sqrt{\zeta\sigma_g^2 + 1}log(N_{yt}))}{\sum_{\tilde{j}\in y}\exp(\delta_{\tilde{j}yvt})[1 + \sum_{\tilde{y}\in\mathcal{Y}}\exp(\bar{\delta}_{\tilde{y}vt} + d_v g_{i\tilde{y}} + \sqrt{\zeta\sigma_g^2 + 1}log(N_{\tilde{y}t}$$

Now let's simplify the problem. Invert the Leontief production function is very easy. Indeed, by its definition, at the optimum we would have:

$$q_{jyvt} = \frac{s_{jyvt}}{1-x_{jyvt}} = \frac{r_{jyvt}}{x_{jyvt}}$$

Optimal input quantities are then $s^*_{jyvt} = (1 - x_{jyvt})q_{jyvt}$ and $r^*_{jyvt} = q_{jyvt}x_{jyvt}$.
Hence, the maximization problem becomes:

$$\max_{q_{jyvt},x_{jyvt}\forall y\in\mathcal{Y}_{jt}} \quad \sum_{y\in\mathcal{Y}_{jt}}\sum_{v}[p_{vt}q_{jyvt} - q_{jyvt}[k_{vt} + (c_{vt} - k_{vt})x_{jyvt} + \theta]]$$

$$\text{s.t.} \quad q_{jyvt} \approx D_v \sum_{y_i} M_{y_i} \frac{\exp(\delta_{jyvt})\exp(\bar{\delta}_{yvt} + d_v g_{iy} + \sqrt{\zeta\sigma_g^2 + 1}log(N_{yt}))}{\sum_{\tilde{j}\in y}\exp(\delta_{\tilde{j}yvt})[1 + \sum_{\tilde{y}\in\mathcal{Y}}\exp(\bar{\delta}_{\tilde{y}vt} + d_v g_{i\tilde{y}} + \sqrt{\zeta\sigma_g^2 + 1}log(N_{\tilde{y}t}))]}$$

Using an implcit notation for demand:

$$\max_{x_{jyvt}\forall y\in\mathcal{Y}_{jt}} \quad \sum_{y\in\mathcal{Y}_{jt}}\sum_{v}[p_{vt}q_{jyvt}(x_{jyvt}) - q_{jyvt}(x_{jyvt})[k_{vt} + (c_{vt} - k_{vt})x_{jyvt} + \theta]]$$

Best responce system The best response of organization j for drug v at time t can be characterized by the first-order condition for each quality x_{jyvt}:

$$\frac{\nabla q_{jyvt}}{\nabla x_{jyvt}}[p_{tv} - k_{vt} - (c_{vt} - k_{vt})x_{jyvt} - \theta] - (c_{vt} - k_{vt})q_{jyvt} + \sum_{\tilde{y}\in\mathcal{Y}_{jt}/y}\frac{\nabla q_{j\tilde{y}vt}}{\nabla x_{jyvt}}[p_{tv} - k_{vt} - (c_{vt} - k_{vt})x_{j\tilde{y}vt} - \theta] = 0 \quad \forall v, j$$

The first term is the gain from an increased demand resulting from an increase in drug quality. The second term is the associated increase in costs. The third term is the quantity effect of x_{jyv} on all the other dealing points of gang j. This last term is zero for an organization with only one dealing point. In our demand system, dealing points of the same gang are gross substitutes (i.e. $\frac{\nabla q_{jy'vt}}{\nabla x_{jyvt}} > 0$) and therefore the third term is always positive. Notice that here we don't have any spillover between drug types because we assumed that consumers buy inelastically the preferred type of drug (if any).

Alternatively, denoting marginal costs as $mc_{jyvt} = k_{vt} + (c_{vt} - k_{vt})x_{jyvt} + \theta$, we can rewrite the quality FOC as:

$$\frac{\nabla q_{jyvt}}{\nabla x_{jyvt}}[p_{vt} - mc_{jyvt}] - c_{vt}q_{jyvt} + \sum_{\tilde{y}\in\mathcal{Y}_{jt}/y} \frac{\nabla q_{j\tilde{y}vt}}{\nabla x_{jyvt}}[p_{vt} - mc_{jyvt}] = 0 \quad \forall v, j$$

Notice that each organization has Y quality first order conditions that can be collectively rewritten in matrix form as follows, for each drug type v and time t:

$$-c_{vt}\begin{bmatrix} q_{j1vt} \\ q_{j2vt} \\ \vdots \\ q_{jYvt} \end{bmatrix}_{Y\times 1} + \begin{bmatrix} \frac{\nabla q_{j1vt}}{\nabla x_{j1vt}}\Big|_{1\in\mathcal{Y}_{jt}} & \cdots & \frac{\nabla q_{jYvt}}{\nabla x_{j1vt}}\Big|_{Y\in\mathcal{Y}_{jt}} \\ \vdots & \vdots & \vdots \\ \frac{\nabla q_{j1vt}}{\nabla x_{jYvt}}\Big|_{1\in\mathcal{Y}_{jt}} & \cdots & \frac{\nabla q_{jYvt}}{\nabla x_{jYvt}}\Big|_{Y\in\mathcal{Y}_{jt}} \end{bmatrix}_{Y\times Y} [p_{vt} - \begin{bmatrix} \text{mc}_{j1vt} \\ \text{mc}_{j2vt} \\ \vdots \\ \text{mc}_{jYvt} \end{bmatrix}_{Y\times Y}] = 0 \quad \forall v, t$$

Following Berry et al., 1993; Berry, 1994, we define a square matrix Λ_{jvt} of dimensions $Y \times Y$ with elements:

$$\Lambda_{jvt}^{y\tilde{y}} = \begin{cases} \frac{\nabla q_{j\tilde{y}vt}}{\nabla x_{iyvt}} & \text{if } \tilde{y} \in \mathcal{Y}_{jt} \\ 0 & \text{otherwise} \end{cases}$$

We can write the entire system of best-response equations for organization j, v in vector form as:

$$-c_{vt}q_{jvt}(\mathbf{x}_{jvt}) + \Lambda_{jvt}(\mathbf{x}_{jvt})[p_{vt} - \mathbf{mc}_{jvt}] = 0 \quad \forall v, t$$

Rearranging and substituting marginal costs:

$$\mathbf{x}_{jvt} = \frac{k_{vt}}{c_{vt}-k_{vt}}(p_{vt} - \theta) - \Lambda_{jvt}^{-1}(\mathbf{x}_{jvt})q_{jvt}(\mathbf{x}_{jvt}) \quad \forall v, t$$

Stacking all the organizations, drugs, and time toghether we have the fixed-mapping described in body:

$$\mathbf{x} - \Lambda(\mathbf{x})^{-1}q(\mathbf{x}) = \frac{\mathbf{k}}{\mathbf{c}-\mathbf{k}}(\mathbf{p} - \boldsymbol{\theta})$$

Existance and uniqueness of the equilibrium

A.3 Fighting game with no uncertainty

$$\begin{cases} \phi_j > \Phi + \phi_{-j} & 0.5(F,NF) + 0.5(F,NF) \rightarrow \text{-j exits} \rightarrow F_{j,-j} = 0 \\ -\Phi < \phi_j - \phi_{-j} \leq \Phi \;\&\; \pi_j^{NF} > \pi_j^d \;\&\; \pi_{-j}^{NF} > \pi_{-j}^d & 0.5(F,NF) + 0.5(NF,F) \rightarrow \textbf{ME} \rightarrow F_{j,-j} = 0 \\ -\Phi < \phi_j - \phi_{-j} \leq \Phi \;\&\; \pi_j^{NF} > \pi_j^d \;\&\; \pi_{-j}^{NF} \leq \pi_{-j}^d & 0.5(NF,F) + 0.5(NF,F) \rightarrow \text{j exits} \rightarrow F_{j,-j} = 0 \\ -\Phi < \phi_j - \phi_{-j} \leq \Phi \;\&\; \pi_j^{NF} \leq \pi_j^d \;\&\; \pi_{-j}^{NF} > \pi_{-j}^d & 0.5(F,NF) + 0.5(F,NF) \rightarrow \text{-j exits} \rightarrow F_{j,-j} = 0 \\ -\Phi < \phi_j - \phi_{-j} \leq \Phi \;\&\; \pi_j^{NF} \leq \pi_j^d \;\&\; \pi_{-j}^{NF} \leq \pi_{-j}^d & 0.5(NF,F) + 0.5(F,NF) \rightarrow \textbf{ME} \rightarrow F_{j,-j} = 0 \\ \phi_{-j} > \Phi + \phi_j & 0.5(NF,F) + 0.5(NF,F) \rightarrow \text{j exits} \rightarrow F_{j,-j} = 0 \end{cases}$$

A.4 Fighting game with uncertainty

$$\begin{cases} E\pi_j \leq \pi_j^{NF} \leq \pi_j^p \;\&\; E\pi_{-j} \leq \pi_{-j}^{NF} \leq \pi_{-j}^p & 0.5(F,NF) + 0.5(NF,F) \rightarrow \textbf{ME} \rightarrow F_{j,-j} = 0 \\ \pi_j^{NF} \leq E\pi_j \leq \pi_j^p \;\&\; E\pi_{-j} \leq \pi_{-j}^{NF} \leq \pi_{-j}^p & 0.5(F,NF) + 0.5(F,NF) \rightarrow \text{-j exits} \rightarrow F_{j,-j} = 0 \\ \pi_j^{NF} \leq \pi_j^p \leq E\pi_j \;\&\; E\pi_{-j} \leq \pi_{-j}^{NF} \leq \pi_{-j}^p & 0.5(F,NF) + 0.5(F,NF) \rightarrow \text{-j exits} \rightarrow F_{j,-j} = 0 \\ E\pi_j \leq \pi_j^{NF} \leq \pi_j^p \;\&\; \pi_{-j}^{NF} \leq E\pi_{-j} \leq \pi_{-j}^p & 0.5(NF,F) + 0.5(NF,F) \rightarrow \text{j exits} \rightarrow F_{j,-j} = 0 \\ \pi_j^{NF} \leq E\pi_j \leq \pi_j^p \;\&\; \pi_{-j}^{NF} \leq E\pi_{-j} \leq \pi_{-j}^p & 0.5(NF,F) + 0.5(F,NF) \rightarrow \text{nothing happens} \rightarrow F_{j,-j} = 0 \\ \pi_j^{NF} \leq \pi_j^p \leq E\pi_j \;\&\; \pi_{-j}^{NF} \leq E\pi_{-j} \leq \pi_{-j}^p & 0.5(F,F) + 0.5(F,NF) \rightarrow \text{war} \rightarrow F_{j,-j} = 1 \\ E\pi_j \leq \pi_j^{NF} \leq \pi_j^p \;\&\; \pi_{-j}^{NF} \leq \pi_{-j}^p \leq E\pi_{-j} & 0.5(NF,F) + 0.5(NF,F) \rightarrow \text{j exits} \rightarrow F_{j,-j} = 0 \\ \pi_j^{NF} \leq E\pi_j \leq \pi_j^p \;\&\; \pi_{-j}^{NF} \leq \pi_{-j}^p \leq E\pi_{-j} & 0.5(NF,F) + 0.5(F,F) \rightarrow \text{war} \rightarrow F_{j,-j} = 1 \\ \pi_j^{NF} \leq \pi_j^p \leq E\pi_j \;\&\; \pi_{-j}^{NF} \leq \pi_{-j}^p \leq E\pi_{-j} & 0.5(F,F) + 0.5(F,F) \rightarrow \text{war} \rightarrow F_{j,-j} = 1 \end{cases}$$

A.5 Multiple equilibria simplified game

$$\begin{cases} Pr(F_{jo} = 1) = Pr(\pi_j^{NF} \leq \pi_j^p \leq E\pi_j \;\&\; \pi_o^{NF} \leq E\pi_o \leq \pi_o^p; \\ \qquad \pi_j^{NF} \leq E\pi_j \leq \pi_j^p \;\&\; \pi_o^{NF} \leq \pi_o^p \leq E\pi_o; \pi_j^{NF} \leq \pi_j^p \leq E\pi_j \;\&\; \pi_o^{NF} \leq \pi_o^p \leq E\pi_o), \\ Pr(F_{jo} = 0) = Pr(\pi_j^{NF} \leq E\pi_j \leq \pi_j^p \;\&\; \pi_o^{NF} \leq E\pi_o \leq \pi_o^p; \\ \qquad E\pi_j \leq \pi_j^{NF} \leq \pi_j^p \;\&\; \pi_o^{NF} \leq E\pi_o \leq \pi_o^p; E\pi_j \leq \pi_j^{NF} \leq \pi_j^p \;\&\; \pi_o^{NF} \leq \pi_o^p \leq E\pi_o; \\ \qquad \pi_j^{NF} \leq E\pi_j \leq \pi_j^p \;\&\; E\pi_o \leq \pi_o^{NF} \leq \pi_o^p; \pi_j^{NF} \leq \pi_j^p \leq E\pi_j \;\&\; E\pi_o \leq \pi_o^{NF} \leq \pi_o^p; \\ \qquad E\pi_j \leq \pi_j^{NF} \leq \pi_j^p \;\&\; E\pi_o \leq \pi_o^{NF} \leq \pi_o^p) \end{cases} \tag{14}$$

A.6 Identification and estimation police reaction function

Let's take the arrest function: $A^i_{yt} = I_{yt}(e^{\gamma O^*_{yt} + \epsilon_{yt}} - 1)$. By taking the logarithm and appoximating $log(e^x - 1) \approx x - 0.7 \quad \forall x \in [0, 1]$, we have:

$$log(A^i_{yt}) - log(I_{yt}) = \gamma O^*_{yt} + \epsilon_{yt} - 0.7$$

Let's define the Δ_p, where $p = s, l, -y$ operator to be:

$$\Delta_p x_y = x|F_{yt} = 1 - x|F_{y,t-1} = 0$$

This is the "event-study" operator. It will be used extensively in the identification sections. By applying this operator on the previous equation:

$$\Delta_p log(A^i_y) - \Delta_p log(I_y) = \gamma \Delta_p O^*_y + \Delta_p \epsilon_y$$

By applying the same operator on the police reaction function:

$$\Delta_p O^*_y = \lambda_p + \Delta_p u_y$$

where $\lambda_p = \lambda_s, \lambda_l, \delta$. By plugging one into the other:

$$\Delta_p log(A^i_y) - \Delta_p log(I_y) = \gamma \lambda_p + \Delta_p \epsilon_y + \gamma \Delta_p u_y$$

Let's now make the following identification assumptions:

- Uncorrelated crime: $\exists \hat{I}$ such that $\frac{\triangledown \hat{I}_{yt}}{\triangledown \mathbf{F}} = 0$
- Parallel trend 1: $\mathbb{E}(\Delta_p \epsilon_y) = 0$
- Parallel trend 2: $\mathbb{E}(\Delta_p u_y) = 0$

Then, we have:

$$\mathbb{E}(\Delta_p log(\hat{A^i_y})) = \gamma \lambda_p$$

Which give us the following identification equations:

$$\begin{cases} \lambda_s = \frac{1}{\gamma} \Delta_s log(\hat{A^i_y}) > \Delta_s log(\hat{A^i_y}) \\ \lambda_l = \frac{1}{\gamma} \Delta_l log(\hat{A^i_y}) > \Delta_l log(\hat{A^i_y}) \\ \delta = \frac{1}{\gamma} \Delta_{-y} log(\hat{A^i_y}) > \Delta_{-y} log(\hat{A^i_y}) \end{cases}$$

Which underlines the identification of the lower bound through the event study design, given $0 < \gamma < 0.3$. Moreover, notice that the ratio between the reaction parameters is indeed identified. To conclude, notice that we can recover: $\mathbb{E}(log(\hat{A^i_{yt}})|F_p = 0 \forall p) = \bar{o}'_y$ where $\bar{o}'_y = \mathbb{E}(log(I_{yt})|F_p = 0 \forall p) + \gamma \bar{o}_y - 0.7$. Hence, we are unable to identify the average police presence, but a function of it. However, given that we are interested in the changes of this variable over areas, the identified parameter is good enough, since its variation over areas is the same as the variation in average police attention.

A.7 Identification and estimation demand

As in the previous step, we can write arrests function and selection equation as:

$$log(A^i_{yt}) - log(I_{yt}) = \gamma O^*_{yt} + \epsilon_{yt} - 0.7$$
$$log(Q_{vyt}) - log(q^*_{yvt}) = \gamma O^*_{yt} + \epsilon_{vyt} - 0.7$$

That being said, let's take the across nest linear demand equation, and apply the usual Δ operator:

$$\Delta log(q^*_{yv}) = \beta \Delta \bar{x}_{yv} + \zeta \Delta log(N_y) + \mu \Delta O^*_y + \Delta \mu_{vt} + \Delta \xi_{yv}$$

Equivalently:

$$\Delta log(A^i_y) - \Delta log(I_y) = \gamma \Delta O^*_y + \Delta \epsilon_y$$
$$\Delta log(Q_{vy}) - \Delta log(q^*_{yv}) = \gamma \Delta O^*_y + \Delta \epsilon_{vy}$$

Substituting the selection equation in the demand:

$$\Delta log(Q_{vy}) - \gamma \Delta O^*_y - \Delta \epsilon_{vy} = \beta \Delta \bar{x}_{yv} + \zeta \Delta log(N_y) + \mu \Delta O^*_y + \Delta \mu_{vt} + \Delta \xi_{yv}$$

To conclude, let's substitute the arrest equation:

$$\Delta log(Q_{vy}) - \Delta log(A^i_y) + \Delta \epsilon_y - \Delta \epsilon_{vy} = \beta \Delta \bar{x}_{yv} + \zeta \Delta log(N_y) + \mu \Delta O^*_y + \Delta \mu_{vt} + \Delta \xi_{yv}$$

As one cna notice, the "selection" effect of police attention ($\gamma \Delta O^*_y$), being equal among the two measures, goes away, allowing identification of μ. To conclude, notice that $\Delta O^*_y = \lambda$. Therefore:

$$\Delta log(Q_{vy}) - \Delta log(A^i_y) + \Delta \epsilon_y - \Delta \epsilon_{vy} = \beta \Delta \bar{x}_{yv} + \zeta \Delta log(N_y) + \mu \lambda + \Delta \mu_{vt} + \Delta \xi_{yv}$$

Let's now make the following identification assumptions:

- Uncorrelated crime: $\exists \hat{I}$ such that $\frac{\nabla \hat{I}_{yt}}{\nabla \mathbf{F}} = 0$

- Parallel trend 1: $\mathbb{E}(\Delta\epsilon_y) = 0$
- Parallel trend 2: $\mathbb{E}(\Delta\epsilon_{yv}) = 0$
- Parallel trend 3: $\mathbb{E}(\Delta\xi_{yv}) = 0$

Then, we have:

$$\mathbb{E}(\Delta log(Q_{vy}) - \Delta log(A^i_y)) = \beta\Delta\bar{x}_{yv} + \zeta\Delta log(N_y) + \mu\lambda$$

Hence, our demand parameters β, ζ, μ are identified. This identification procedure directly links with an event-study estimation strategy:

$$\begin{aligned} log(Q_{yvt}) - log(A^i_{yt}) =& \alpha + \beta\bar{x}_{yvt} + \alpha p_{vt} + \zeta log(N_{yt}) + \sum_{j=-4}^{-2} \tilde{\mu}_j(\text{Lag j}) + \sum_{k=0}^{4} \tilde{\mu}_k(\text{Lead k}) + \mu_y + \mu_v + \mu_t + \epsilon \\ (\text{Lag j})_{yt} &= 1[t = F_y - j] \quad \text{for } j \in \{1, \ldots, -4\} \\ (\text{Lead k})_{yt} &= 1[t = F_y + k] \quad \text{for } k \in \{1, \ldots, 4\} \end{aligned} \tag{15}$$

where E_y is a dummy variable equal to 1 with the outbreak of the first fight in the area. We then recover our parameter of interest as: $\mu = \hat{\tilde{\mu}}_0/\hat{\tilde{\lambda}}$. To conclude, we use the estimated area fixed effect:

$$\hat{\mu}_y = d \times log(\textstyle\sum_{y_i} \frac{M_{y_i}}{\sum_y M_y} exp(g_{iy})) + u_y$$

to get an estimate for the transportation cost $\hat{d}$.

A.8 Simulation algorithm

Guess $[\tilde{\theta}]$

while $||\hat{m}_{jvyt} - \tilde{m}_{jvyt}|| > \epsilon$ **do**

  Guess $\tilde{x}_{jvyt}$

  while $||\hat{x}_{jvyt} - \tilde{x}_{jvyt}|| > \epsilon$ **do**

    Step 1 : Simulate market shares: $\breve{s}_{jyvt}$

    Step 2 : Simulate semi-elasticities

$$\breve{e}_{j\tilde{j}y\tilde{y}tv} = \frac{\nabla \breve{s}_{\tilde{j}\tilde{y}tv}}{\nabla x_{jytv}} = \begin{cases} -\beta x_{jvyt}(1 - \breve{s}_{\tilde{j}\tilde{y}tv}) & \tilde{y} = y,\ \tilde{j} = j \\ \beta x_{jvyt}\breve{s}_{\tilde{j}\tilde{y}tv} & \tilde{y} \neq y,\ \tilde{j} = j \\ 0 & \tilde{j} \neq j \end{cases}$$

    Step 3 : Construct matrix:

$$\Lambda_{jv}^{y\tilde{y}} = \begin{cases} \frac{\nabla q_{j\tilde{y}v}}{\nabla x_{iyv}} & \text{if } \tilde{y} \in \mathcal{Y}_j \\ 0 & \text{otherwise} \end{cases}$$

    Step 4 : Estimate predicted quality using quality FOC

$$\underline{x}_{vt} = \frac{k_t}{c_t - k_t}(p_{vt} - \tilde{\theta}) - \underline{\Lambda}_{vt}{}^{-1}(\underline{x}_{vt})\underline{q}_{vt}(\underline{x}_{vt}) \quad \forall v, t$$

  end

  Step 5 : Compute moment

$$\tilde{m}_{yvt} = \frac{1}{T}\frac{1}{N_y}\sum_{yt} \breve{x}_{tvy}$$

end

B Additional figures

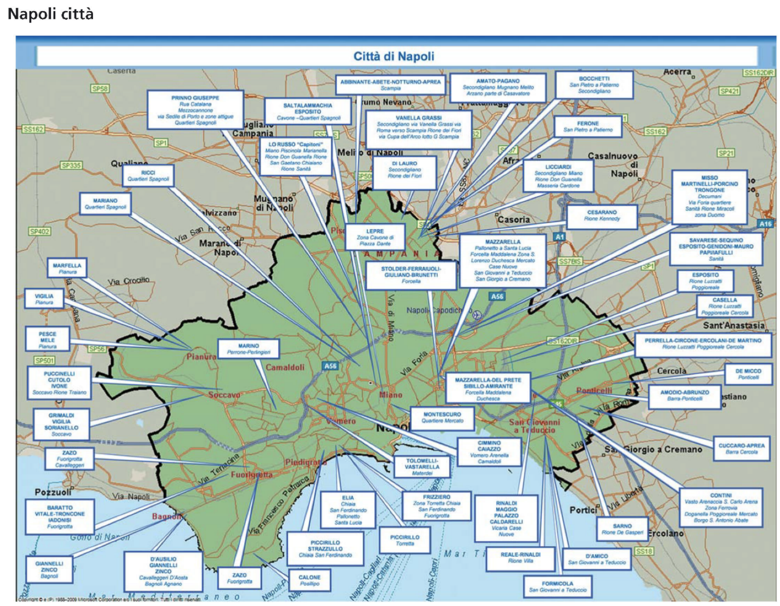
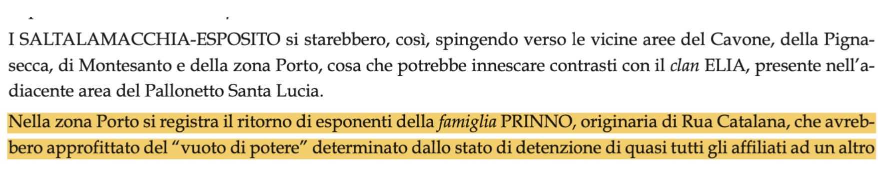


Figure A1: DIA report - map



I SALTALAMACCHIA-ESPOSITO si starebbero, così, spingendo verso le vicine aree del Cavone, della Pignasecca, di Montesanto e della zona Porto, cosa che potrebbe innescare contrasti con il *clan* ELIA, presente nell'adiacente area del Pallonetto Santa Lucia.

Nella zona Porto si registra il ritorno di esponenti della *famiglia* PRINNO, originaria di Rua Catalana, che avrebbero approfittato del "vuoto di potere" determinato dallo stato di detenzione di quasi tutti gli affiliati ad un altro

Figure A2: DIA report - text

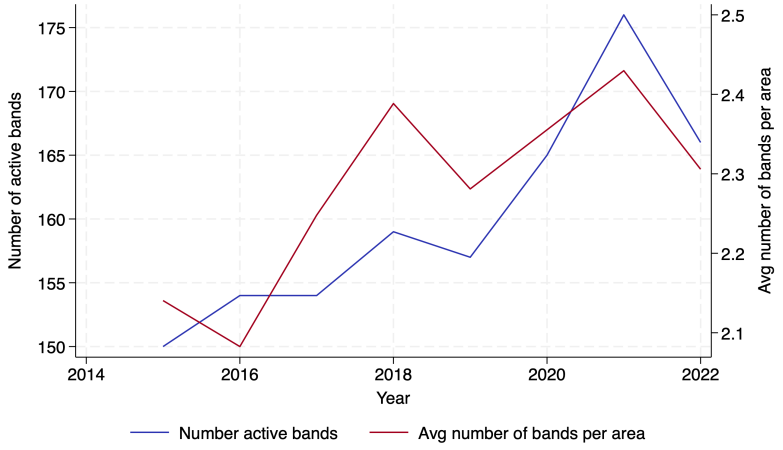


Figure A3: Number of active gangs

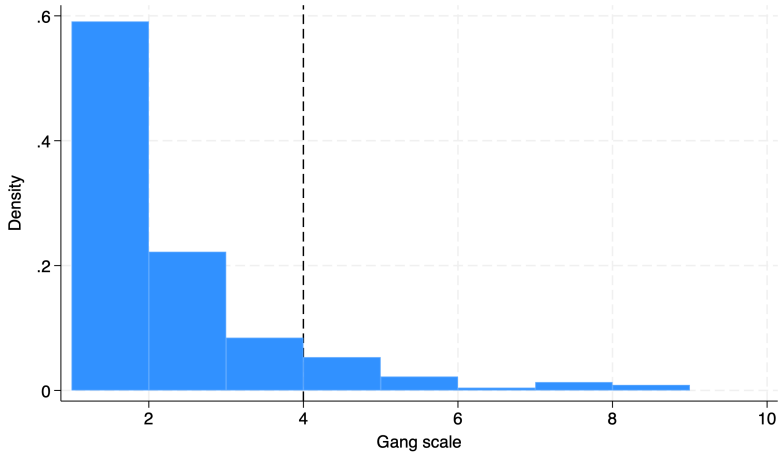


Figure A4: Density gang scale

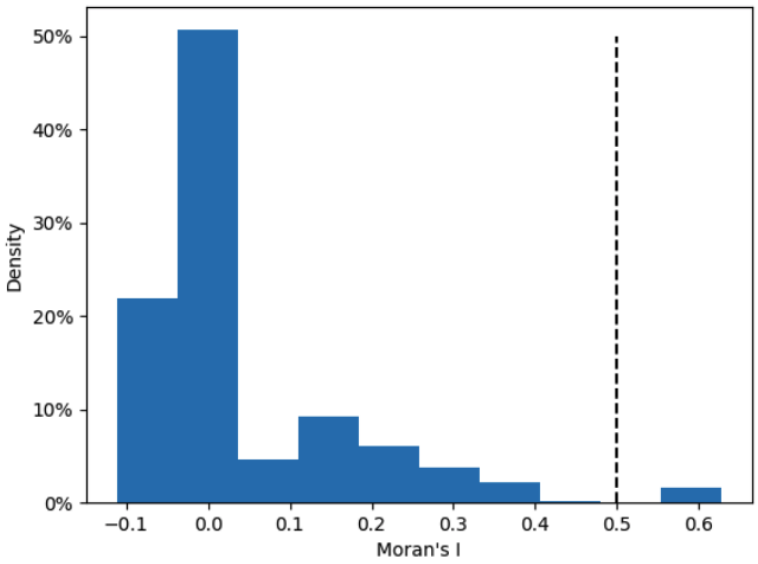


Figure A5: Moran's I

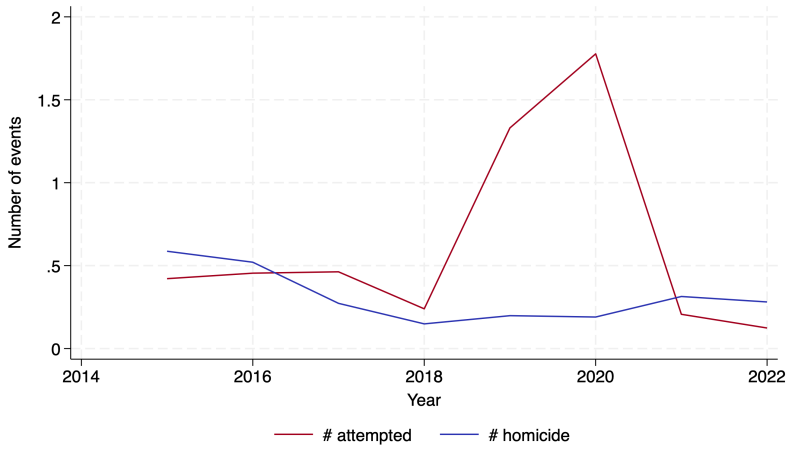


Figure A6: Time trend violence

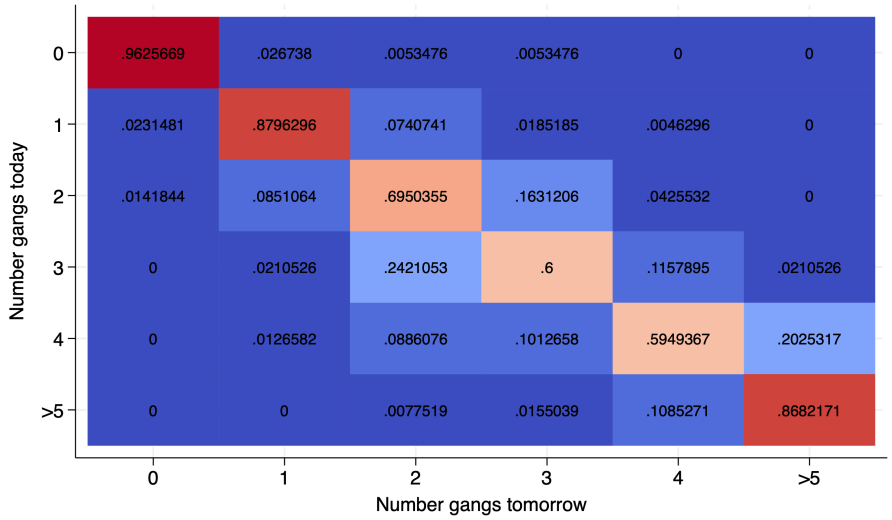


Figure A7: Transition matrix - number of gangs in the area

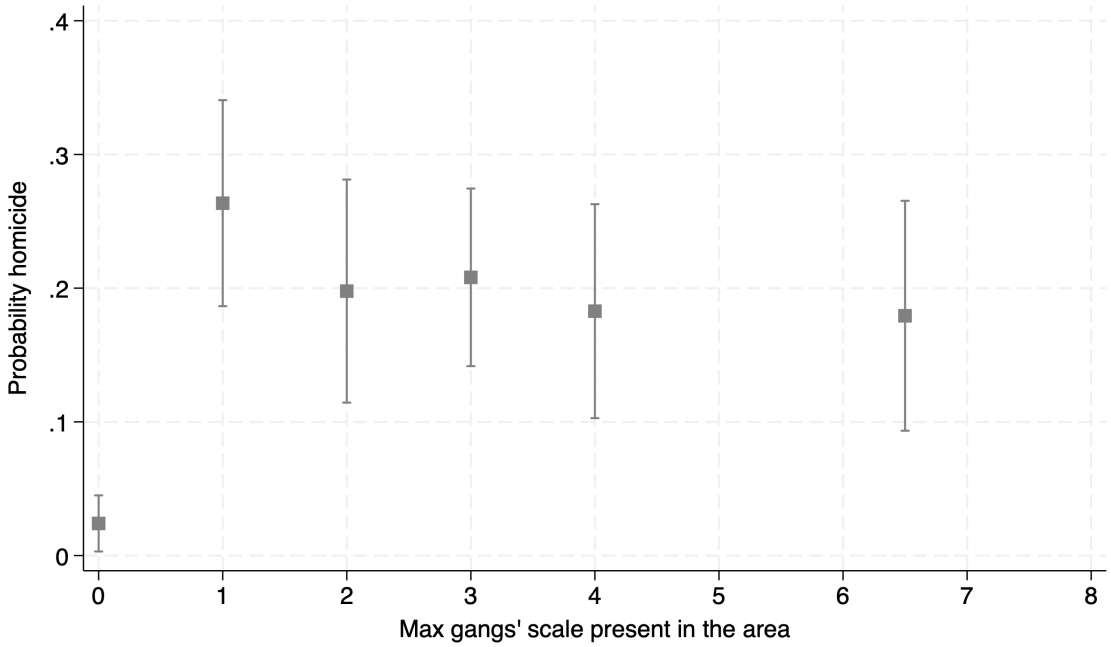


Figure A8: Max scale gangs

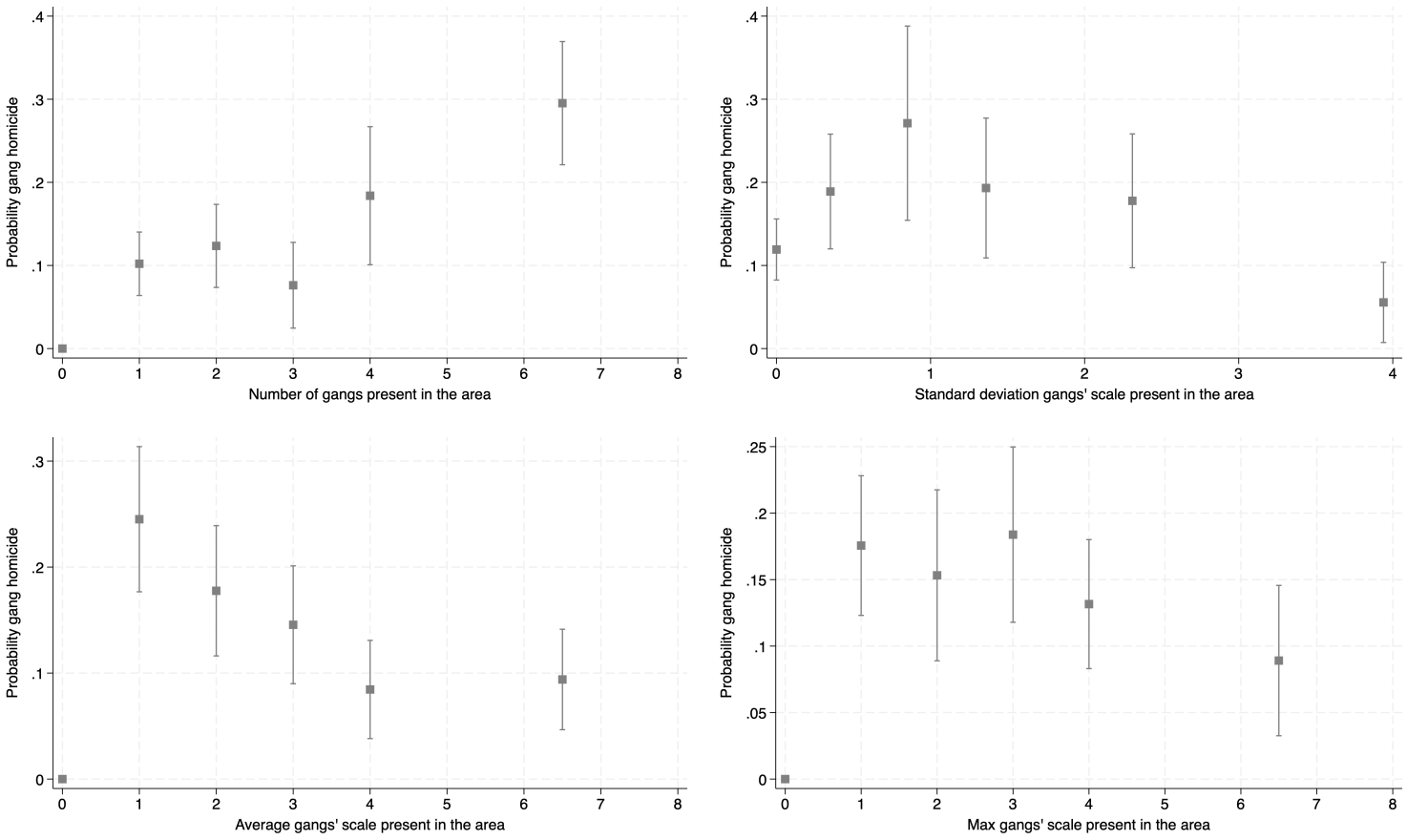


Figure A9: Market structure and probability gang homicide

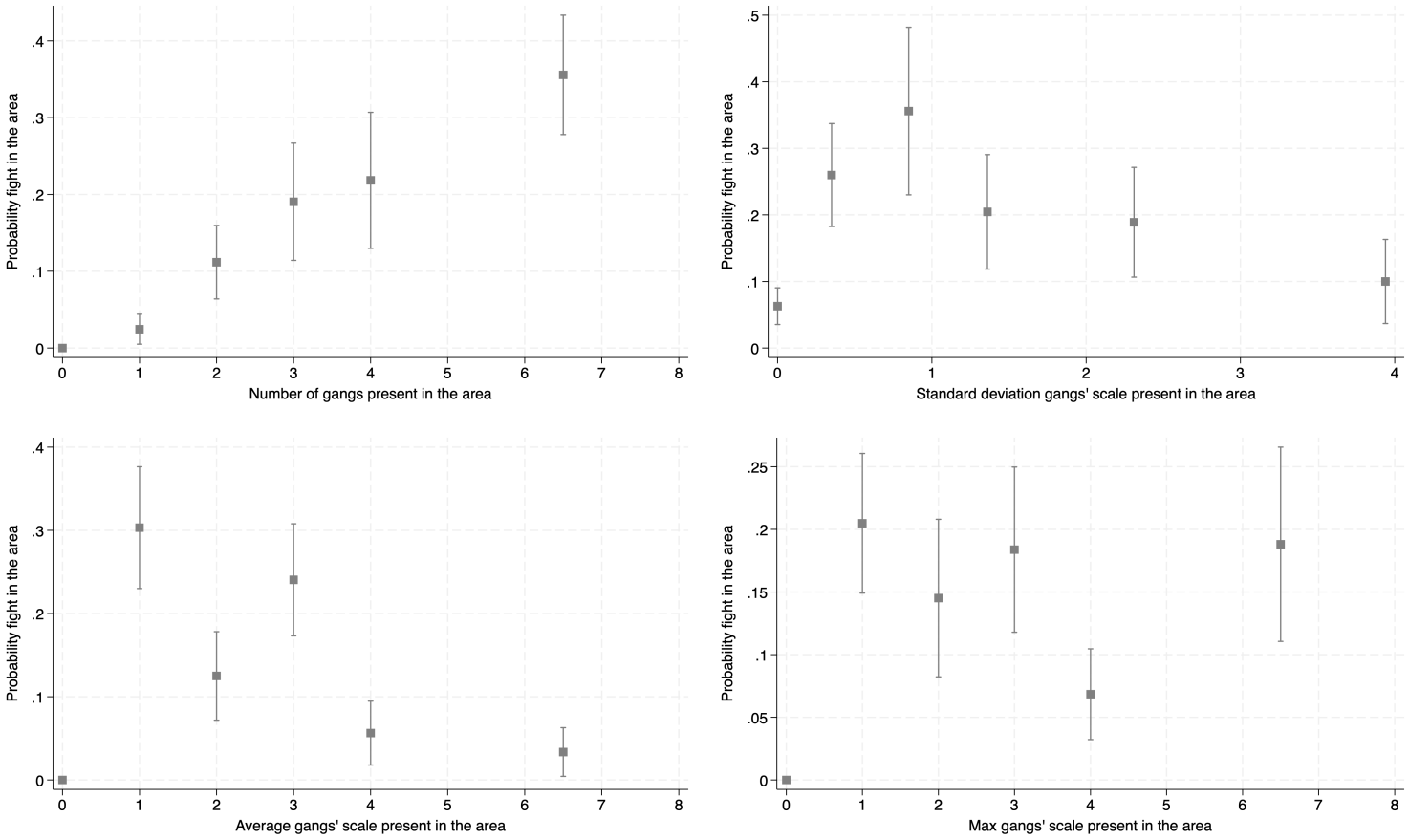


Figure A10: Market structure and probability war

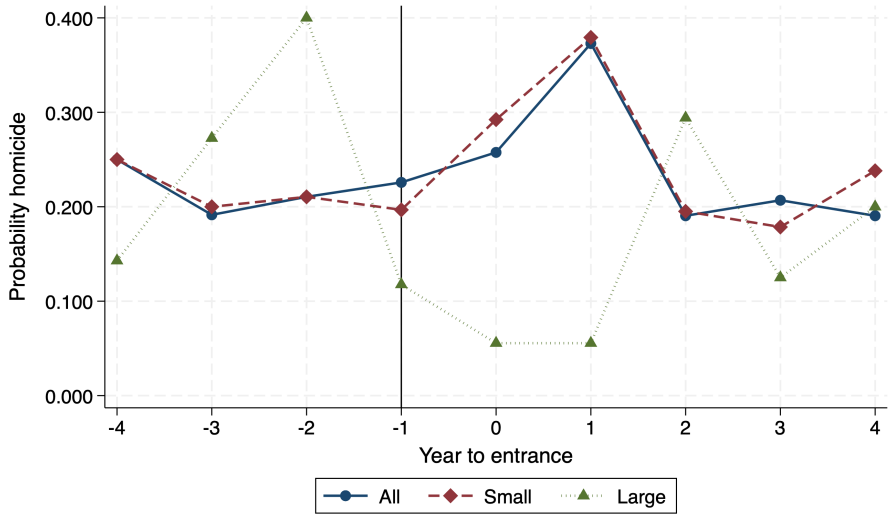


Figure A11: Event study - probability homicide, raw data

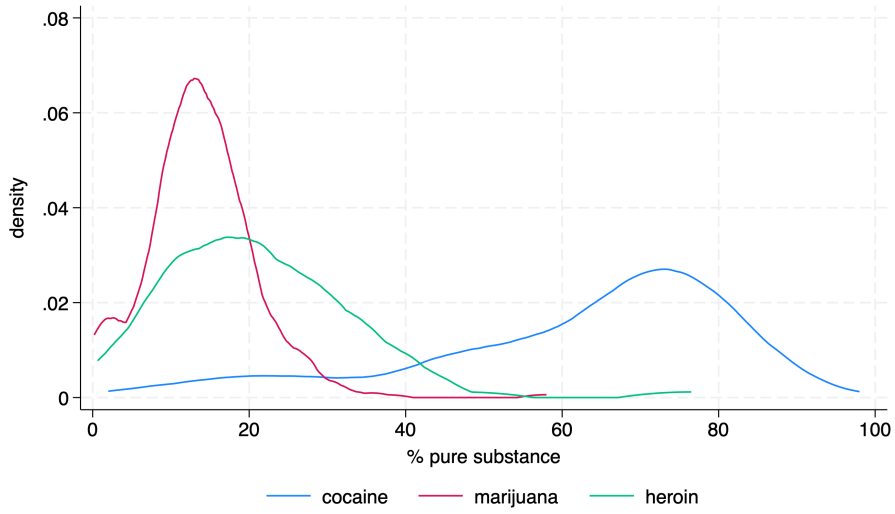


Figure A12: Distribution drug quality

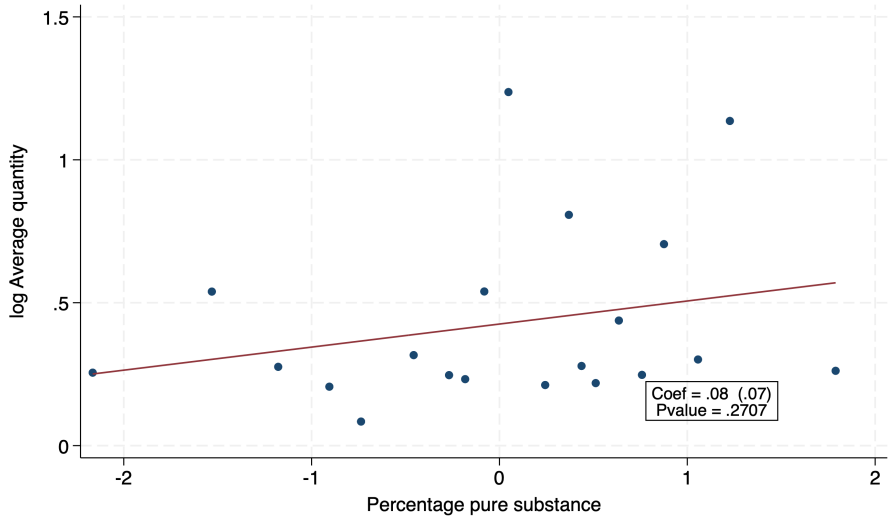


Figure A13: Correlation average quantity and quality drugs

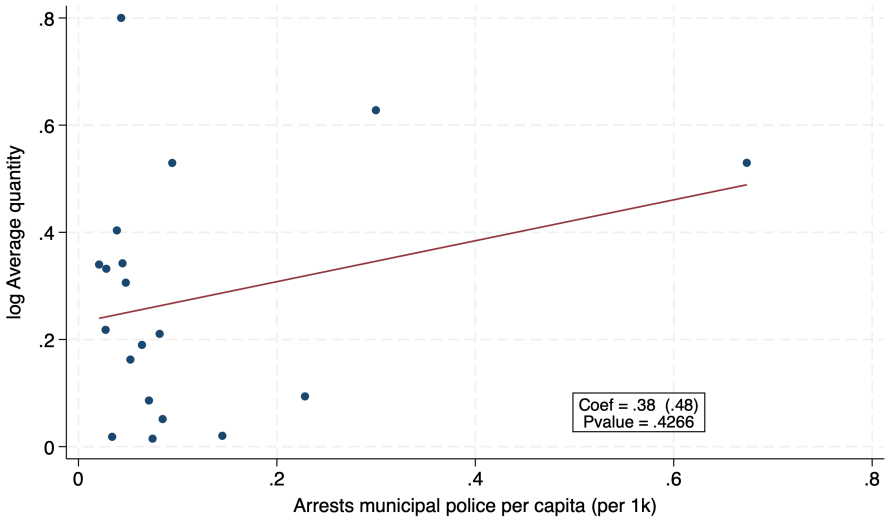


Figure A14: Correlation average quantity and municipal police arrests

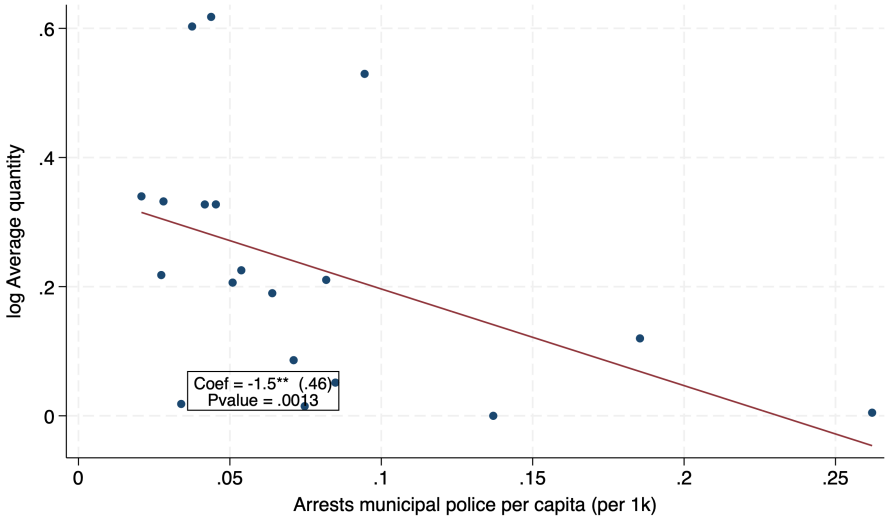


Figure A15: Correlation average quantity and municipal police arrests - no outliers

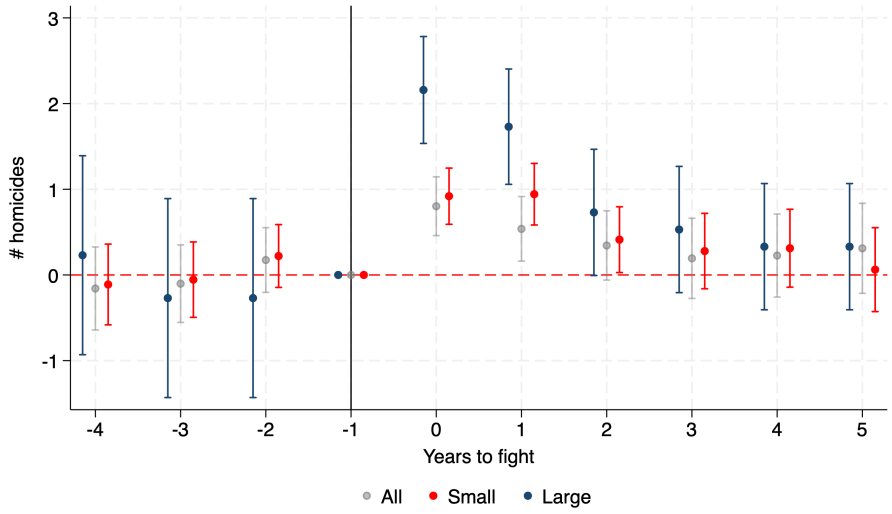


Figure A16: Event study - number of homicides

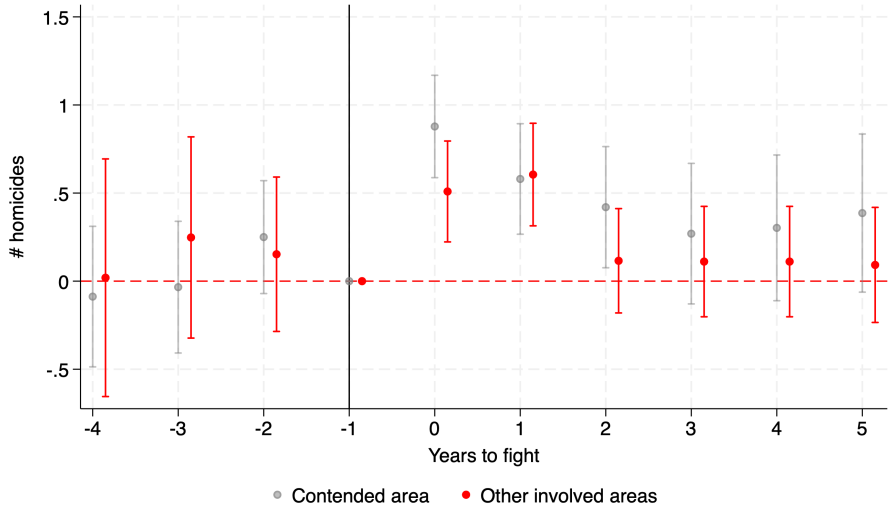


Figure A17: Event study - spillover violence

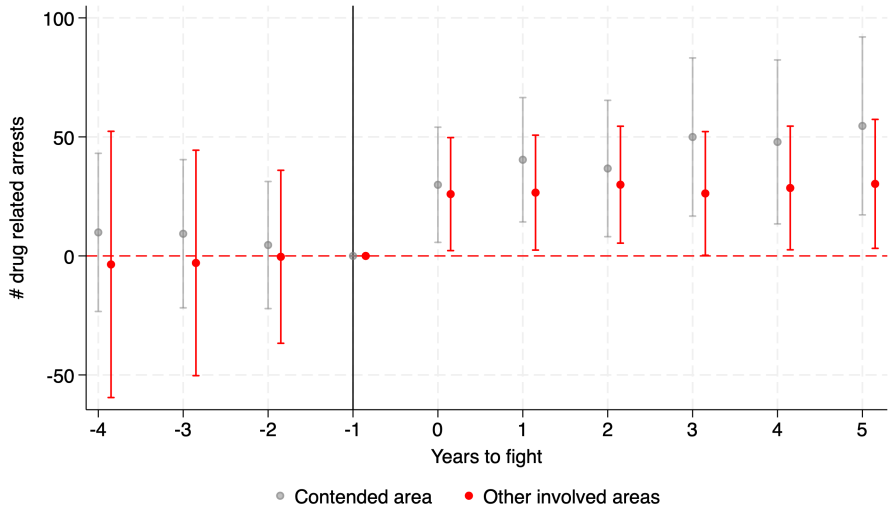


Figure A18: Event study - spillover arrests

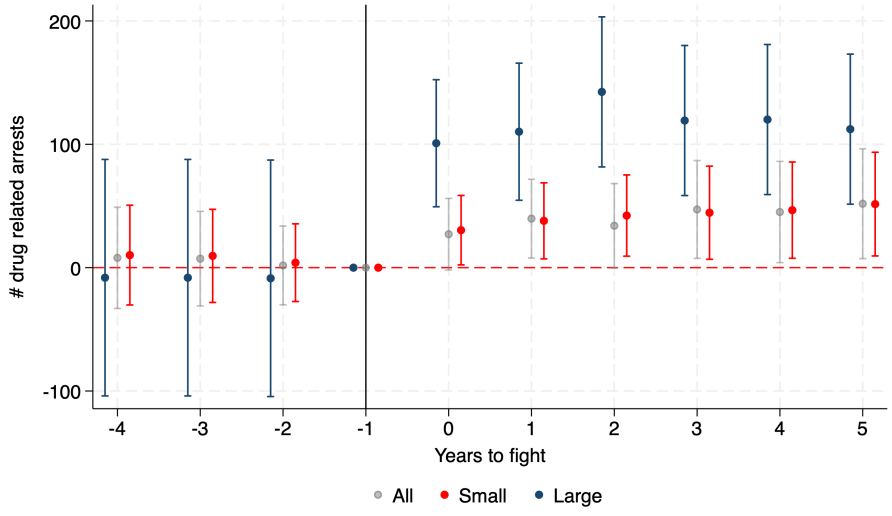


Figure A19: Event study - number drug related arrests

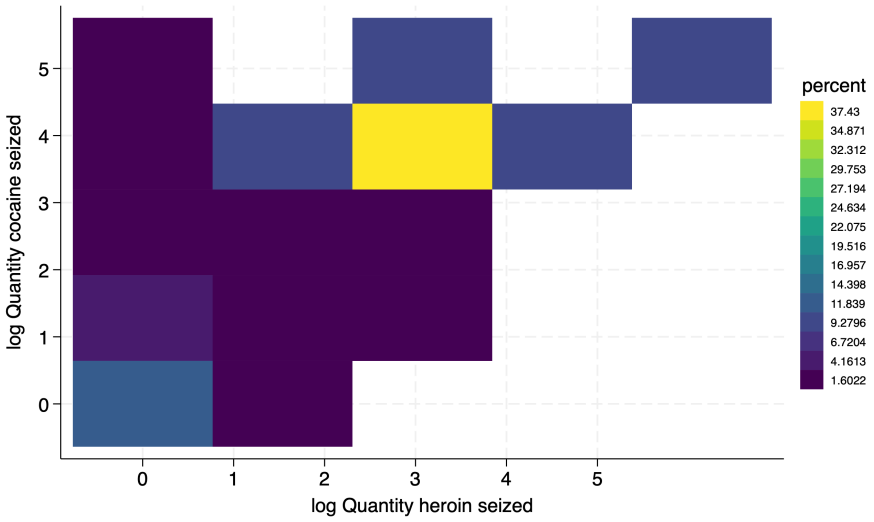


Figure A20: Heatplot - seizures cocaine and heroin

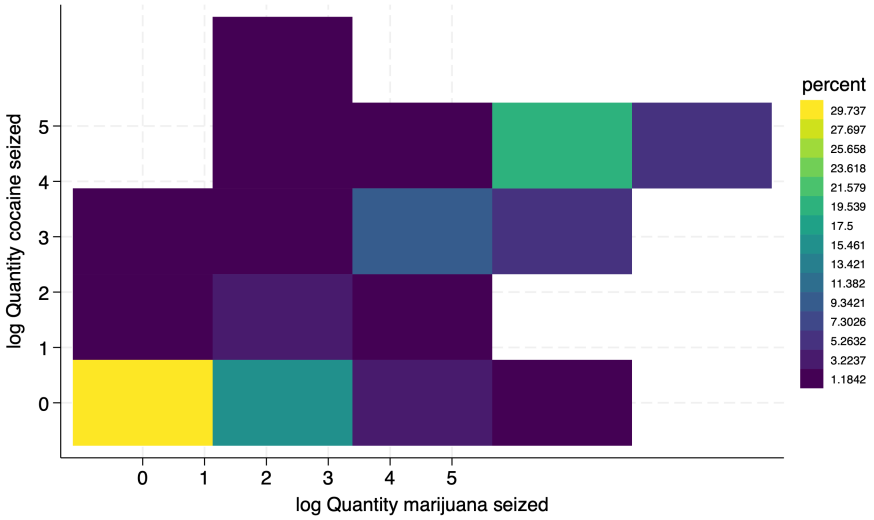


Figure A21: Heatplot - seizures cocaine and marijuana

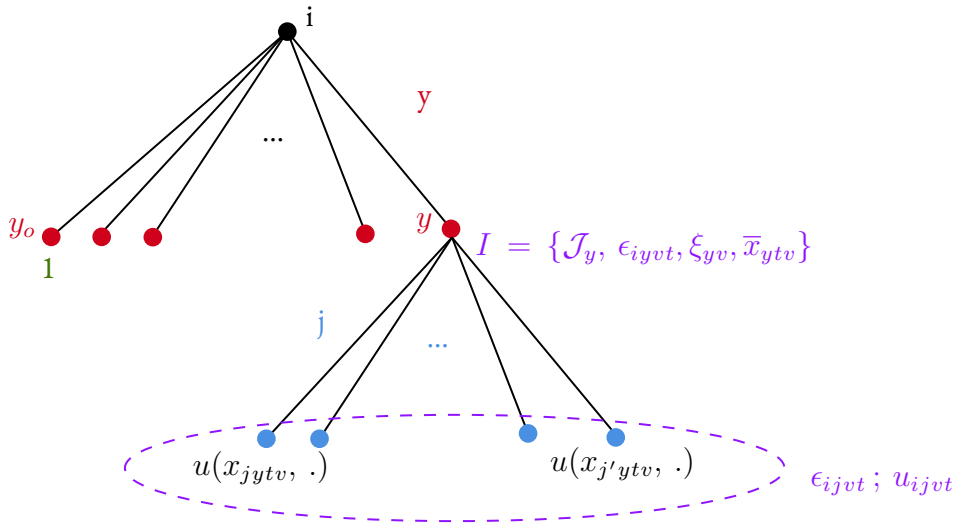


Figure A22: Consumer information sets

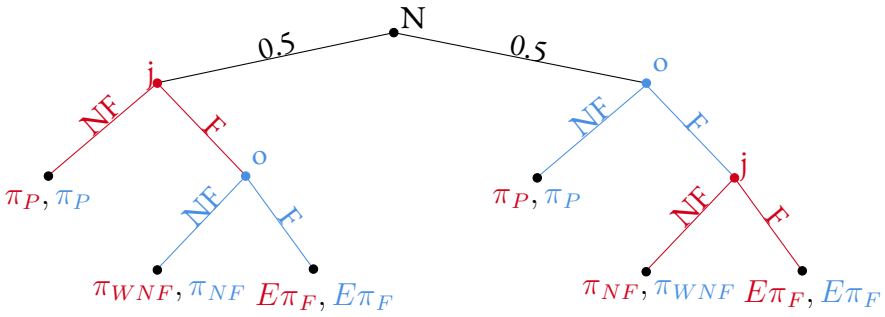


Figure A23: Fighting game

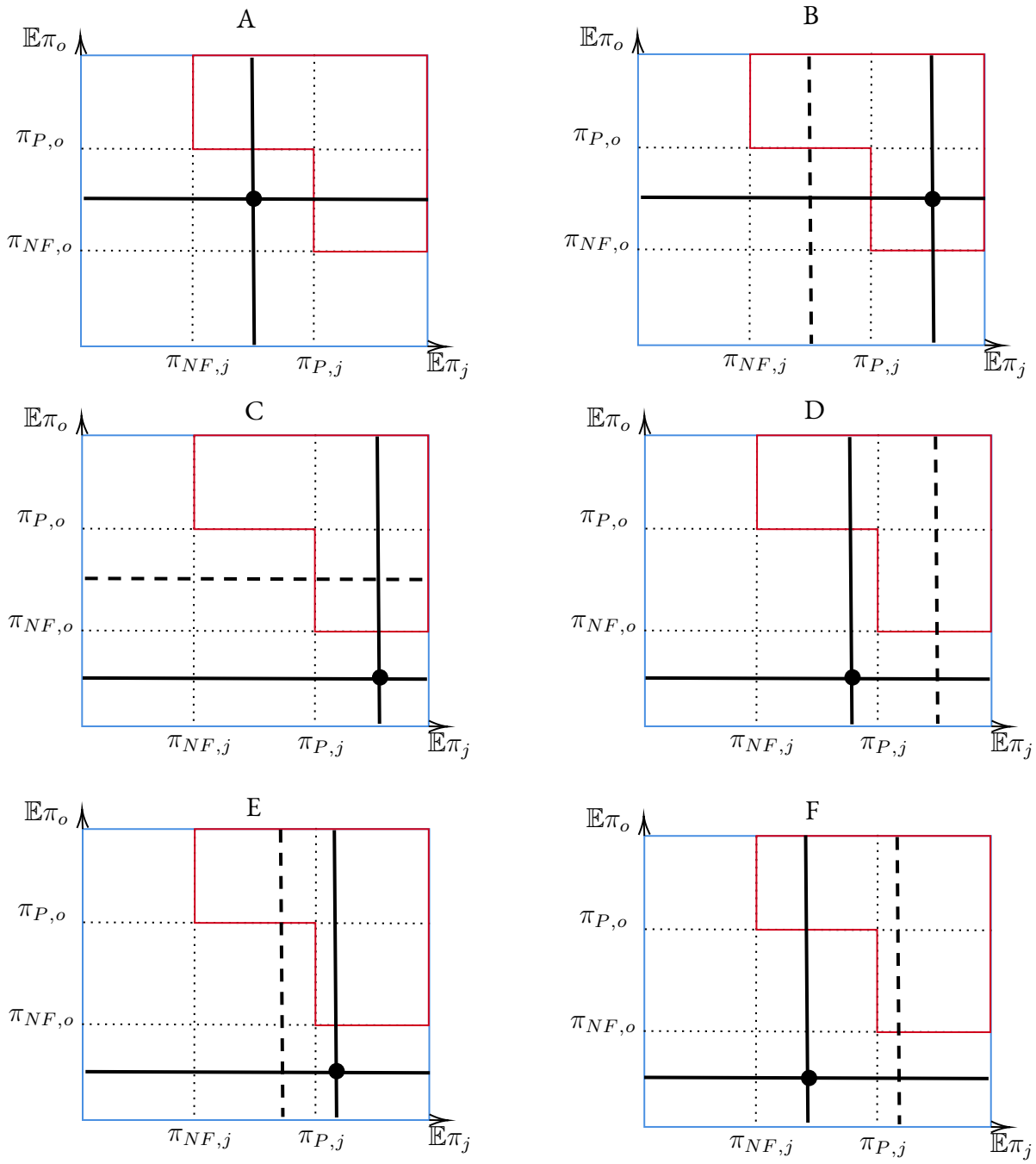


Figure A24: Theoretical results

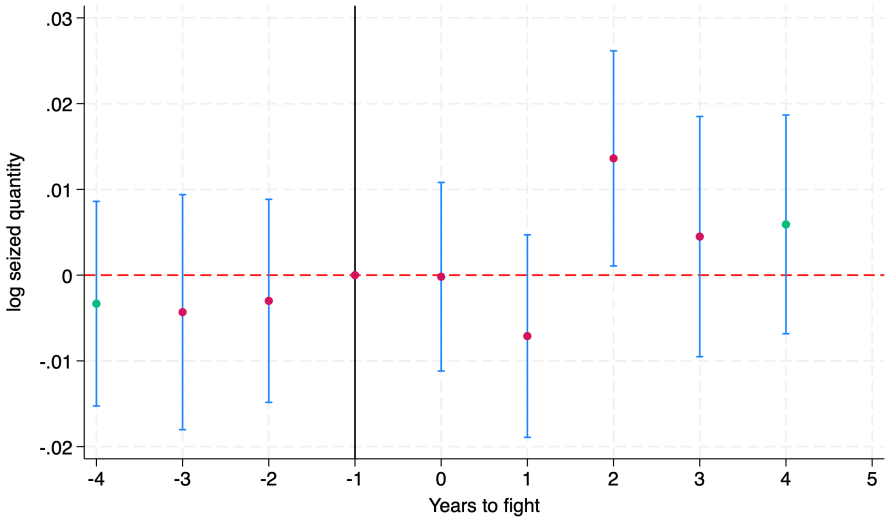


Figure A25: Event study - seized quantity

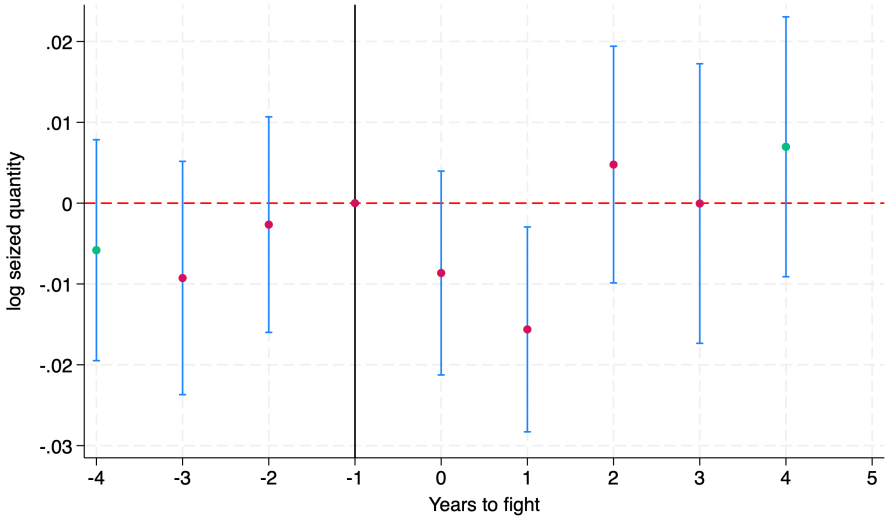


Figure A26: Event study - seized quantity, controlling for selection

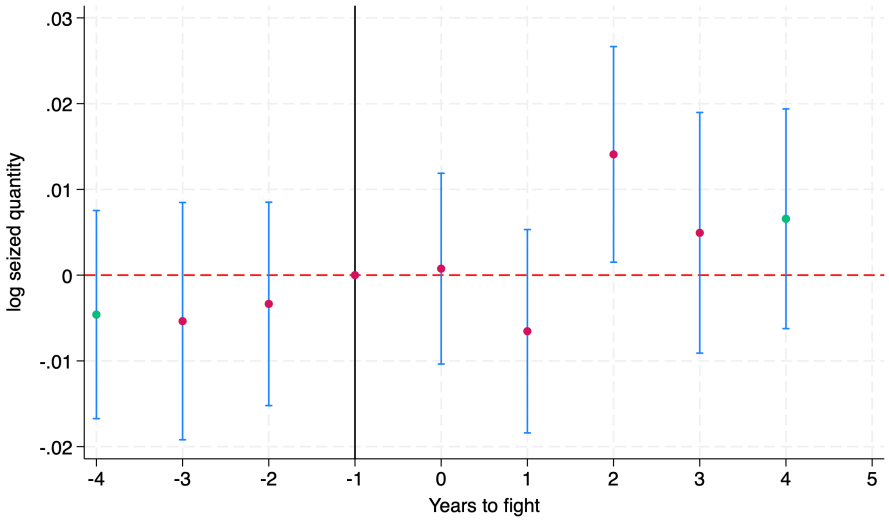


Figure A27: Event study - seized quantity, robustness

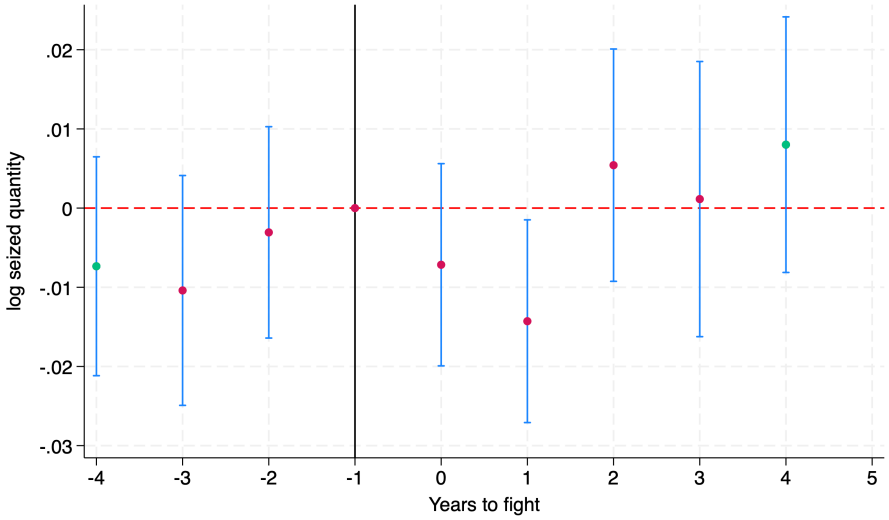


Figure A28: Event study - seized quantity, controlling for selection, robustness

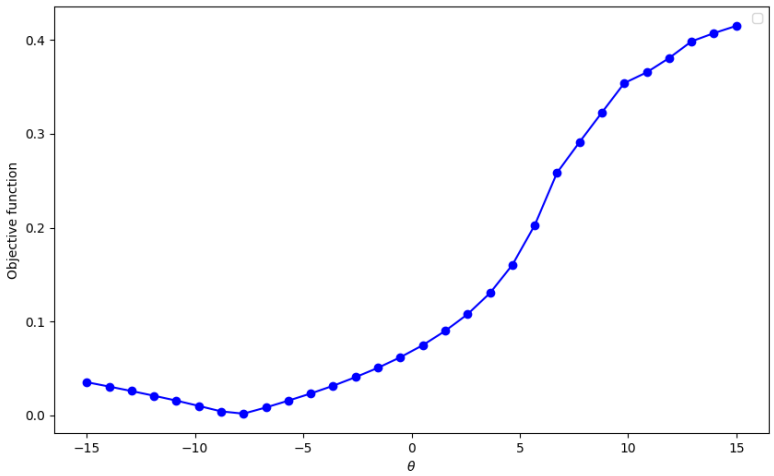


Figure A29: SMM unobserved productivity - objective function over grid

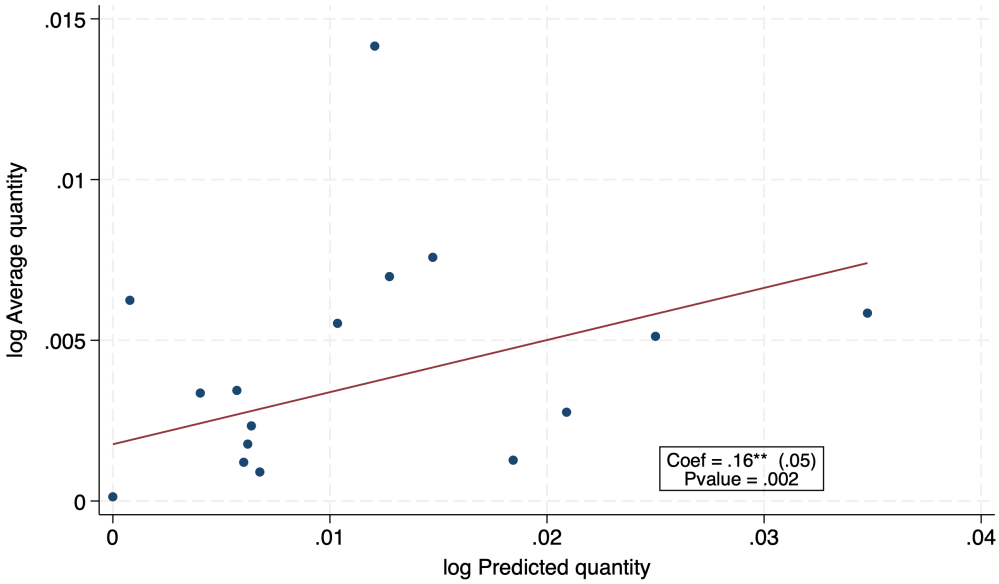


Figure A30: Model fit - quantities

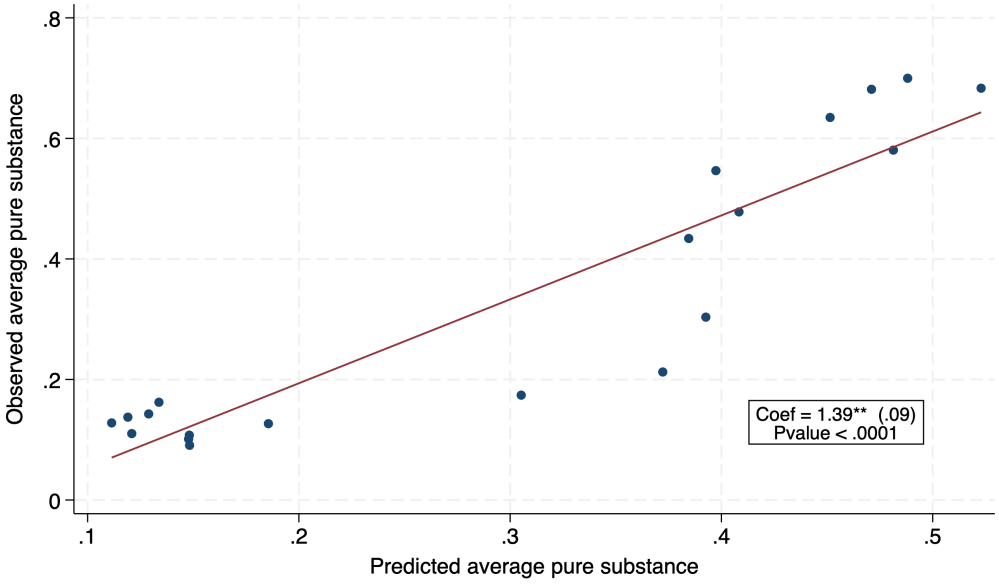


Figure A31: Model fit - quality

C Additional tables

Table A1: Summary statistics area-year

	Mean	SD	Min	Max	N
Ever presence	0.780992	0.413788	0	1	968
War here	0.120868	0.326142	0	1	968
Number attempted homicides	0.627066	2.348704	0	36	968
Number homicides	0.314050	0.798479	0	9	968
Pr. attempted homicides	0.317149	0.465606	0	1	968
Pr. homicides	0.201446	0.401288	0	1	968
Number arrests conspiracy	47.845041	85.531653	0	242	968
Number arrests drugs	157.314050	264.076899	0	706	968
Number arrests local police	15.752475	37.440674	0	178	707

Table A2: Summary statistics fights

# fights	100
# fights with known contended market	93
# fights with (known) winner	35
Mean number of gangs involved	3.25
Min number of gangs involved	2
Max number of gangs involved	14
# fights in areas with no gang present	0
# fights in areas with one gang present	13
# fights in areas with both gang present	80
Average number of periods in fight	1.49
Percentage fights lasting one period	72%

Table A3: Average scale and violence

Dep. Variable: Probability	(1) Homicide	(2) Homicide	(3) Homicide	(4) Fight	(5) Fight	(6) Fight
Average scale	-0.0341**	-0.0315*	-0.0413	-0.0445**	-0.0257	-0.0693**
	(0.0171)	(0.0171)	(0.0489)	(0.0173)	(0.0171)	(0.0340)
Number gangs		0.0112			0.0812***	
		(0.0223)			(0.0188)	
Sd scale			-0.0253			0.00848
			(0.0469)			(0.0378)
Observations	753	753	503	753	753	503
R-squared	0.257	0.257	0.255	0.365	0.385	0.361
Area FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Mean Dep | Average scale = 1	0.302	0.302	0.302	0.226	0.226	0.226
Dep. Variable: Number	Homicide	Homicide	Homicide	Per 1k pl	Per 1k pl	Per 1k pl
Average scale	-0.0837**	-0.0825**	-0.0452	-0.00327**	-0.00310**	-0.00367
	(0.0354)	(0.0407)	(0.0840)	(0.00142)	(0.00143)	(0.00302)
Number gangs		0.00529			0.000732	
		(0.0654)			(0.000995)	
Sd scale			-0.0229			0.00151
			(0.0777)			(0.00276)
Observations	753	753	503	753	753	503
R-squared	0.351	0.351	0.361	0.251	0.251	0.209
Area FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Mean Dep | Average scale = 1	0.453	0.453	0.453	0.009	0.009	0.009

Table A4: Event study

	All				Small				Large			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Dep. Variable:	Pr. Homicide	# Homicide	Pr. Gang Homicide	Fight	Pr. Homicide	# Homicide	Pr. Gang Homicide	Fight	Pr. Homicide	# Homicide	Pr. Gang Homicide	Fight
Lag -4	0.079	0.118	0.059	0.175***	0.083	0.125	0.064	0.174***	-0.061	0.824	0.022	-0.119***
	(0.067)	(0.106)	(0.054)	(0.064)	(0.066)	(0.106)	(0.054)	(0.064)	(0.135)	(1.069)	(0.134)	(0.019)
Lag- 3	0.024	0.108	0.057	0.121**	0.039	0.133	0.070	0.131**	0.026	-0.010	-0.043	-0.043
	(0.056)	(0.113)	(0.050)	(0.052)	(0.059)	(0.199)	(0.052)	(0.055)	(0.120)	(0.176)	(0.076)	(0.077)
Lag -2	0.061	0.090	0.068	0.093**	0.055	0.090	0.077	0.080*	0.184	0.126	0.102	0.046
	(0.056)	(0.082)	(0.047)	(0.044)	(0.056)	(0.083)	(0.048)	(0.043)	(0.118)	(0.150)	(0.100)	(0.091)
Lag -1	-	-	-	-	-	-	-	-	-	-	-	-
	-	-	-	-	-	-	-	-	-	-	-	-
Lead 0	0.109*	0.194*	0.074	0.342***	0.148**	0.253**	0.112**	0.347***	-0.148**	-0.207*	-0.064	0.102
	(0.058)	(0.101)	(0.047)	(0.061)	(0.060)	(0.105)	(0.051)	(0.062)	(0.058)	(0.118)	(0.057)	(0.100)
Lead 1	0.225***	0.376***	0.177***	0.202***	0.235***	0.393***	0.186***	0.205***	-0.148**	-0.262***	-0.119***	-0.064
	(0.067)	(0.124)	(0.058)	(0.057)	(0.067)	(0.125)	(0.058)	(0.058)	(0.058)	(0.071)	(0.016)	(0.056)
Lead 2	0.042	0.109	0.018	0.043	0.050	0.124	0.025	0.044	0.089	0.034	0.056	0.056
	(0.067)	(0.129)	(0.048)	(0.047)	(0.065)	(0.131)	(0.049)	(0.048)	(0.113)	(0.151)	(0.095)	(0.095)
Lead 3	0.058	0.007	0.060	-0.016	0.034	-0.014	0.035	-0.052***	-0.079	-0.130	0.005	0.005
	(0.078)	(0.081)	(0.064)	(0.037)	(0.076)	(0.079)	(0.059)	(0.014)	(0.085)	(0.139)	(0.084)	(0.085)
Lead 4	0.033	0.027	-0.077***	-0.005	0.093	0.092	-0.024	0.042	- 0.022	-0.136	-0.119***	-0.029
	(0.084)	(0.109)	(0.014)	(0.047)	(0.095)	(0.118)	(0.049)	(0.066)	(0.119)	(0.126)	(0.016)	(0.089)
Observations	968	968	968	968	968	968	968	968	968	968	968	968
R-squared	0.043	0.071	0.049	0.104	0.047	0.074	0.053	0.107	0.017	0.018	0.016	0.013

Table A5: Probability outcomes fights

	(1)	(2)	(3)
Fighting outcome	Win	Draw	Loose
Large	0.776*	-0.306	-0.0190
	(0.416)	(0.418)	(0.494)
Observations	70	70	70
Mean Dep | Large = 0	0.262	0.672	0.115

Table A6: Data and model

Data	Model	Source
Network matrix	$\mathbf{Y}_t \ \forall t$	Inteligence data - Neapolitan Police
Matrix of fights	$\mathbf{F}_t \ \forall t$	Inteligence data - Neapolitan Police
# of arrests for drug dealing and related crimes	Measure $O_{ty} \ \forall y, t$	Administrative data - DCSA
Geolicalized seized drug type, quantity and time	Measure $q_{yvt} \ \forall y, t, v$	Administrative data - DCSA
Geolicalized quality of drug	$x_{vty} \ \forall y, t, v$	DCSA + Scientific Police
Average retail drug price over time and type	$p_{vt} \ \forall t, v$	Inteligence data - DCSA
Average wholesale drug price over time and type	$c_{vt} \ \forall t, v$	Inteligence data - DCSA
Distances between centroids of all markets	$g_{yy'} \ \forall y, y'$	Google maps API
Joint distribution of consumer groups and locations	$F(D_i, y_i)$	ISTAT + IPSAD survey

Table A7: Police reaction function

	(1) Spillover	(2) Small	(3) Large	(4) Spillover	(5) Small	(6) Large
Panel A: Second Stage						
Lag -4	-0.004	-0.003	-0.005	0.000	0.000	0.001
	(0.013)	(0.012)	(0.020)	(0.006)	(0.006)	(0.009)
Lag -3	-0.001	0.003	-0.005	0.002	0.005	0.001
	(0.016)	(0.016)	(0.024)	(0.006)	(0.006)	(0.009)
Lag -2	-0.004	-0.001	-0.005	0.000	0.001	0.000
	(0.14)	(0.016)	(0.018)	(0.005)	(0.006)	(0.008)
Lag -1	-	-	-	-	-	-
	-	-	-	-	-	-
Lead 0	0.027*	0.038**	0.069***	0.006	0.007	0.016*
	(0.014)	(0.016)	(0.021)	(0.005)	(0.006)	(0.009)
Lead 1	0.018	0.038**	0.034*	-0.003	-0.006	-0.001
	(0.015)	(0.018)	(0.017)	(0.005)	(0.007)	(0.008)
Lead 2	0.026	0.037**	0.079***	-0.003	-0.006	-0.003
	(0.016)	(0.018)	(0.017)	(0.006)	(0.007)	(0.011)
Lead 3	0.062**	0.091***	0.155***	-0.012	-0.019*	-0.017
	(0.025)	(0.029)	(0.036)	(0.008)	(0.010)	(0.013)
Lead 4	0.070***	0.139***	0.113***	-0.007	-0.016*	-0.002
	(0.018)	(0.024)	(0.021)	(0.007)	(0.009)	(0.010)
Observations	383	383	383	379	379	379
R2	0.065	0.122	0.149	0.942	0.943	0.943
City FE	No	No	No	Yes	Yes	Yes
Year FE	No	No	No	Yes	Yes	Yes
Mean Dep | Lag -1	0.003	0.005	0.005	0.003	0.005	0.005
Panel B: First Stage						
Prison Exit	0.079**	0.079**	0.079**	0.079**	0.079**	0.079**
	(0.034)	(0.034)	(0.034)	(0.034)	(0.034)	(0.034)
Observations	2190	2190	2190	2190	2190	2190
R2	0.348	0.348	0.348	0.348	0.348 0.348	0.348
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Band FE	Yes	Yes	Yes	Yes	Yes	Yes
Mean Dep | Exit = 0	0.128	0.128	0.128	0.128	0.128	0.128

Table A8: Demand Estimation

	(1)	(2)	(3)	(4)
Panel A: Second Stage				
Outbreak fight (Lead 0 and Lead 1)	-0.00566	-0.0122***	-0.0112**	-0.0575*
	(0.00388)	(0.00440)	(0.00453)	(0.0325)
Retail price			-0.0000155	-0.000507
			(0.000281)	(0.00224)
log Number active bands			-0.00836	0.0839
			(0.00877)	(0.0948)
(inverse of) Average quality				-0.000391
				(0.000836)
Selection		-0.157	-0.163	0.259
		(0.120)	(0.120)	(0.580)
Observations	1584	1149	1149	75
R2	0.517	0.582	0.582	0.814
City FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Drug FE	Yes	Yes	Yes	Yes
Mean Dep	0.013	0.013	0.013	0.013
Panel B: First Stage				
log Arrests local police		0.593***	0.593***	0.593***
		(0.018)	(0.018)	(0.018)
Observations		1149	1149	1149
R2		0.461	0.461	0.461
Panel C: Distance				
Distance				-0.0006
				(0.0006)
Observations				288
R2				0.003

D Other data

Mortality To measure overdoses, we use administrative data on mortality from ISTAT. For each city, in each year, we observe the number of deaths for each cause in a very granular list (ICD classification). In line with the guidelines of the ICD classification, we define a death to be overdose with codes: "X42", "X62", "F11", "F12", "F13", "F14", "F15", "F16", "F17", "F18", "F19", "P04", "P96", "R78", "T36", "T37", "T38", "T39", "T40", "T41", "T42", "T43", "T44", "T45", "T46", "T47", "T48", "T49", "T50", "Z79". We then compute the number of overdoses per each city and year.

Population To measure population size and demographics, we use administrative data on population from ISTAT. For each city and year, we observe the number of people living in the area by sex (Male or Female), age group (0-9, 9-24, 25-49, 50-64, >64), and educational attainment (no degree, primary school, lower secondary, upper secondary, bachelor, master or above). From these classifications, we create nine demographic groups using the intersections of three dummy variables: (i) male, (ii) young (age lower than 49), (iii) uneducated (lower than lower secondary educational attainment). We then compute the number of people for each demographic group, in each city. To project the number of people living in Naples across different districts within the city, we use the 2011 census of Naples city. For each district, we observe the percentage of people in the previously mentioned demographic groups. In particular, we directly observe from the census the percentage of people living in each district for the sex-education classification and age classification separately. We then estimate the percentage of people living in the district by multiplying these probabilities. This assumes no correlation between these two classifications, which is certainly an approximation. However, without an explicit relationship between the two, it is challenging to estimate the percentage in the intersection of the two. To conclude, we multiply this probability by the total number of people living in the city of Naples in each demographic group.

Distances To measure the distance between units, we used the Google maps API. In particular, we compute the average car travel time–independent from traffic conditions–between centroids of the different cities/districts, both in kilometers and seconds. We will use the distance in time.

Consumption survey To estimate the different propensity of drug consumption for differ-

ent types of drugs and demographic groups, we use the publicly available aggregate statistics for the whole country from the IPSAD survey. This is an Italian survey following the guidelines provided by the EMCDDA (European Monitoring Centre for Drug and Drug Abuse) in Lisbon and allows the country to fulfill its information requirements regarding the first of the five epidemiological indicators proposed by the Council of Europe. The sample is drawn from the national population (through civil registry lists), aged between 15-84 years, and since 2005 has had a sample size (approximately 85,000 subjects) sufficient to allow prevalence estimates at the regional level. As public data, they provide average consumption propensity of different types of drugs across different demographic groups.

City survey To measure different characteristics of the cities in the sample, as well as the activity of the municipal police, we use a city survey from OpenCivitas. This is a portal for accessing local authorities' information, a transparency initiative promoted by the Ministry of Economy and Finance and SOSE. Each year–apart from 2020, due to Covid disruption–they send a survey to all cities in Italy. This survey collects balance sheet values, as well as different information on the services provided by the city. Importantly for this paper, they also collect information on the resources allocated to the municipal police as well as the outcomes of their activity (e.g., number of arrests).