NBER WORKING PAPER SERIES

GANG RULE: UNDERSTANDING AND COUNTERING CRIMINAL GOVERNANCE

Christopher Blattman
Gustavo Duncan
Benjamin Lessing
Santiago Tobón

Working Paper 28458
http://www.nber.org/papers/w28458

NATIONA BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2021, Revised March 2022

For comments and feedback we thank Ana Arjona, Oriana Bandiera, Abhijit Banerjee, Eli Berman, Esther Duflo, Leopoldo Fergusson, Claudio Ferraz, Danilo Friere, Soeren Henn, Macartan Humphreys, Stephen Machin, Martin McGuire, Ted Miguel, Mushfiq Mobarak, Eduardo Montero, Daniel Ortega, Gerard Padró i Miquel, Bob Powell, Gérard Roland, Pieter Serneels, Jake Shapiro, Stergios Skaperdas, Tara Lyn Slough, Carlos Schmidt-Padilla, Rodrigo Soares, Maria Micaela Sviatschi, Lucia Tiscornia, Francesco Trebbi, Juan Vargas, Maarten Voors, Jeremy Weinstein, Austin Wright, participants at several seminars and conferences, and Raul Sánchez de la Sierra who was instrumental in conceptualizing the model. Innovations for Poverty Action coordinated all research activities. For research assistance we thank Bruno Aravena-Maguida, David Cerero, Peter Deffebach, Sebastián Hernández, Sofía Jaramillo, Juan F. Martínez, Juan Pablo Mesa-Mejía, Angie Mondragón, Helena Montoya, José Miguel Pascual, Andrés Preciado, María-Aránzazu Rodríguez-Uribe, Zachary Tausanovitch, Nelson Matta-Colorado and Martín Vanegas. We thank the Secretariat of Security of Medellín for their cooperation, especially the former Secretary of Security Andrés Tobón, as well as Lina Calle, Ana María Corpas and Andrés Machuca. For financial support, we thank the Centro de Estudios sobre Seguridad y Drogas (CESED) of Universidad de los Andes; the Peace and Recovery Program (P&R) at Innovations for Poverty Action (IPA); the PROANTIOQUIA foundation; The National Science Foundation (NSF); the UK Foreign, Commonwealth & Development Office trough the Crime and Violence Initiative at J-PAL; and the Economic Development and Institutions Programme (EDI) funded with UK aid from the UK Government, working in partnership with Oxford Policy Management Limited, University of Namur, Paris School of Economics and Aide á la Décision Économique. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Gang rule: Understanding and Countering Criminal Governance
Christopher Blattman, Gustavo Duncan, Benjamin Lessing, and Santiago Tobón
NBER Working Paper No. 28458
February 2021, Revised March 2022
JEL No. C93,E26,H11,K42,O17

ABSTRACT

Criminal groups govern millions of people worldwide. In Medellín, Colombia, for instance, gangs resolve disputes, police neighborhoods, enforce contracts, and tax businesses in their territories. Why do they rule? Many argue that criminals step into vacuums of order. If so, increasing state security services should crowd out gang rule. But interviews with Medellín gangs suggest this overlooks an indirect incentive to rule: governing protects other illicit businesses, such as drug-selling. We begin with a model of imperfect competition and show how drug profits and the need for civilian loyalty could drive gangs to respond to state competition by intensifying their governance. Empirically, we show this is what happened in Medellín over 32 years. When new borders exogenously increased government services on some blocks, gangs raised their civilian rule in protective response—especially in neighborhoods with larger drug markets. Strategic incentives like these severely complicate efforts to fight organized crime.

Christopher Blattman
Harris School of Public Policy
The University of Chicago
1307 E 60th St
Chicago, IL 60637
and NBER
blattman@uchicago.edu

Gustavo Duncan
Department of Political Science
Universidad EAFIT
Carrera 49 N 7 Sur 50
Medellin, Colombia
gduncan@eafit.edu.co

Benjamin Lessing
Department of Political Science
University of Chicago
5828 S. University Ave
Chicago, IL 60637
blessing@uchicago.edu

Santiago Tobón
Department of Economics
Universidad EAFIT
Carrera 49 N 7 Sur 50
Medellin, Colombia
stobonz@eafit.edu.co

1 Introduction

To protect property, enforce contracts, or resolve disputes, hundreds of millions of people turn to street gangs and mafias rather than the government. Such criminal governance is especially common in Latin America, but today we can find gangs who rule civilians in Italy, the United Kingdom, India, South Africa, and the American prison system (Arias, 2006; Lessing et al., 2019; Lessing, 2020; Melnikov et al., 2021).

Specialists in violence have always emerged to provide security, adjudication, and other protection services in return for taxes and rents. Some have been states, some warlords, and some criminals. Historically, the lines between these actors have been blurry (Tilly, 1985; Olson, 1993; Grossman, 1996; Acemoglu and Robinson, 2006; North et al., 2009; Sánchez De La Sierra, 2020). What's different about criminal governance today is that it has survived the growth of strong states in the modern era. In many countries, the government does not have a monopoly on the legitimate use of force in large swathes of their territory. Instead, residents live under a duopoly of coercion (Skaperdas and Syropoulos, 1997).

This paper asks three questions: How does this duopoly function? What motivates gangs to rule? And can states crowd out criminal governance by ruling better?

The common view is that organized crime fills a vacuum left by weak state presence (Gambetta, 1996; Skaperdas, 2001; Skarbek, 2011). The policy implication is that states can crowd out gang rule by improving the quality and reach of their services, eventually reaching a monopoly on protection and coercion.

Our findings challenge this view. While gang rule sometimes emerges from anarchy, we argue that criminal governance can also be a strategic response to strong states. In particular, gangs and mafias engaged in other illicit businesses, especially drug retailing, have reason to provide public services that not only foster civilian loyalty, but also reduce the demand for state representatives on their turf. If so, efforts to crowd out gangs could backfire.

Our study focuses on a single city: Medellín, Colombia. As the country's second-largest city and commercial heart, Medellín is prosperous, collects extensive tax revenues, and provides considerable public goods and social services to its citizens. Nonetheless, virtually every low- and middle-income neighborhood in the city is occupied by one of more than 350 small gangs called *combos*, and most combos engage in at least some governance activities. These include prohibiting and punishing property crime, settling disputes between neighbors, enforcing community rules, and—in exchange—taxing locals.

Because criminal groups are clandestine and gang rule is poorly understood, a first and major focus of this study is descriptive. We began with large-scale, systematic interviews with criminal organizations. Whereas previous economist–ethnographer collaborations—

such as Levitt and Venkatesh (2000) and Sánchez De la Sierra and Titeca (Sánchez De la Sierra and Titeca)—gained extraordinary access to a single illegal group, we develop sources in many groups at a more moderate depth. Over four years we interviewed dozens of leaders, managers, and foot soldiers in 41 gangs on their organization, operations, and rule.

We then set out to collect systematic data on criminal governance. In 2019 we ran a representative, city-wide survey of roughly 7,000 residents and businesses, focusing on who provides governance services in their neighborhood, who collects taxes and extortion, and who they see as legitimate and why.

These data revealed some unexpected facts. For instance, while the state is the predominant provider of protection on average, the combo is seldom far behind. There is wide variation in the degree of such combo rule, but in a significant minority of neighborhoods the combo is the leading provider. These combos organize governance provision as a business line with dedicated staff. While taxing residents involves a degree of coercion, many payees say they value the services combos offer and see them as legitimate.

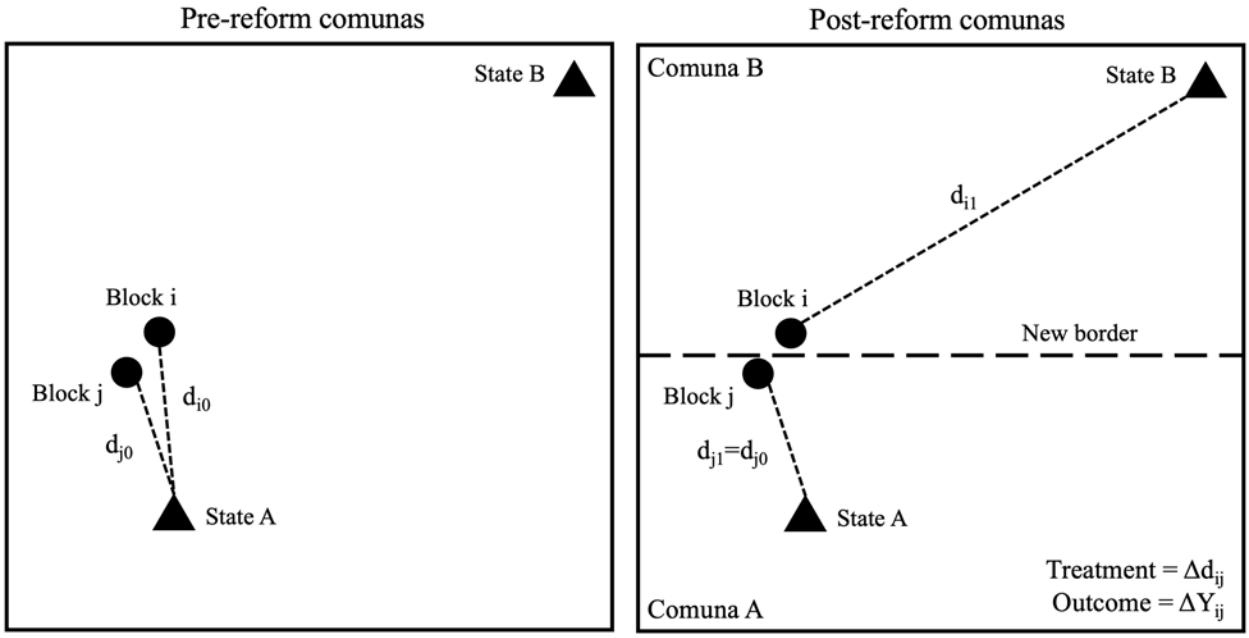
We also learned that many gangs are keen to maintain citizen loyalty and avoid denunciation to the police. Several explained that they rule not for the direct profits, but rather because it protects their other business lines, especially drug sales. Providing neighborhood order reduces the need for routine police patrols and special agents to enter. When police do enter, fostering civilian loyalty makes residents less likely to inform on gang members, and more likely to help the combo evade the authorities.

Thus, while the conventional wisdom suggests gang and state rule are strategic substitutes, our interviews with gang leaders suggest they could also be strategic complements. We illustrate these countervailing responses in a simple model. The strategic substitutes view is consistent with a gang's best response function in models of either imperfect competition or duelling stationary bandits. But if we follow what leaders told us and introduce gang returns to citizen loyalty (to protect drug rents), we show how a gang's best response to increased state presence can flip.

To test this, we study an accidental but decades-long shift in the local presence of state services in Medellín. We see evidence that state-building efforts from the 1990s onward crowded in gang rule over time: gangs began to govern more in areas with closer state presence, especially in the most profitable drug territories.

We take advantage of a natural experiment: a city bill that, in the late 1980s, reorganized Medellín into 16 areas called *comunas*. After introducing the new internal borders, the city spent a decade expanding policing and protection services within them. Importantly, it organized these services within the new comunas. Therefore, street blocks on either side of the new internal borders were generally similar in demographic and economic terms, but

Figure 1: Stylized illustration of the natural experiment



Notes: On the left is a stylized representation of a single pre-1987 comuna, with blocks i and j accessing the state at its closest location (State A). On the right is a representation of the split comuna after 1987, with block i being assigned to access the state at a further location (State B) and block j still accessing the state nearby (State A). See section 5 for details.

from the 1990s onwards, one receives a "distance shock" and is assigned to state protection further away.

Figure 1 illustrates the intuition behind our empirical strategy. Initially, a pair of nearby blocks i and j received their security and dispute resolution services from the same state headquarters, as seen on the left. Once the new border was introduced, as seen on the right, block i was assigned to be supported by a headquarters further away—about 400 meters at the median. In effect, the new borders introduced idiosyncratic increases in distance to police and dispute resolution in one of each pair. We examine the effect of the distance shock, Δd_{ij}, on differences in long-term outcomes within block pairs, ΔY_{ij}.

The key identifying assumption is that no other characteristic (such as economic development) changed discontinuously along the new borders unless it was a consequence of this proximity to state services (Keele and Titiunik, 2015).[1] Consistent with this, we see that the block pairs are balanced according to a wide range of initial characteristics that could confound our estimates, including demographics as well as distance to other social services,

[1]Because of how city services are organized, distance to public security is the only state service where access changes discontinuously at the new borders. All others—schools, health centers, and so on—are not allocated at the communa level, and so the new borders did not change proximity to other arms of the state. Of course, if other public services did shift discontinuously, this would only change the interpretation of the causal effect. As it stands, we can mainly attribute any impacts to public security.

infrastructure, and business agglomerations. To account for any unobserved confounders, we conduct placebo tests of alternate borders and see no evidence of discontinuities there.

We assess impacts three decades later, with our 2019 survey. Not surprisingly, we see large effects of distance on access to and legitimacy of the state. For instance, on blocks assigned to be 400 meters further from the police and municipal dispute resolution agencies (the median change), people reported roughly 11% lower state responsiveness to disputes, disorder, and crime in 2019. This suggests that proximity matters for projecting state power.

How did gangs respond? Three decades after the border introduction, blocks assigned to more distant police and dispute resolution services have *lower* gang rule. For example, streets 400 meters further from the state reported about 18% less combo rule in 2019, and somewhat less combo legitimacy as well. There was little neighborhood gang rule in Medellín before 2000. Hence the results suggest that, from 2000–19, gangs evolved governance services when the state was near.

Why did they do so? The years between 2000 and 2019 also coincide with the expansion of local retail drug sales in Medellín, and with it a dramatic increase in the value of territory and civilian loyalty. Not every street is equally strategic or profitable for drug sales, however. As our theory predicts, we show that the crowding-in of gang governance is greatest in the neighborhoods with above-median value drug markets. The crowding-in is about twice as large in these neighborhoods.

There is another channel by which new borders, and being assigned to faraway security services, could affect gang rule: economic development and migration. If people sorted to the better-policed side of the border, or if businesses grew more in response to security, this could increase the number of transactions, and with it demand for both state and gang rule. Note this would still be a causal effect of Δd_{ij}; it is simply an alternative mechanism to the strategic competition for loyalty. Looking at 2019 levels of development and recent migration, however, we see little evidence that this channel was important.

Finally, we investigate whether measurement error could drive our results. One concern is that people are reluctant to report gang governance when the gang is strong. We show how such systematic under-reporting would generally lead us to underestimate the crowd-in effect. We risk overestimating crowd-in in one highly-specific circumstance: citizens who under-report when the gang is strong and their block is distant from the state, but who report accurately otherwise. A survey experiment and other analyses suggest this is unlikely, and that survey respondents freely report combo activity.

Altogether, these results suggest that gang rule is as much a problem for semi-strong states as for weak ones. Governments powerful enough to prohibit and police illegal markets create incentives for gangs to protect their criminal rents. If, at the same time, the state

is not strong enough to eradicate these activities (and no one has eradicated retail drug markets), then criminal governance may flourish.

These findings also shed light on the divergent styles of criminal rule observed in other countries. For example, Melnikov et al. (2021) show how gangs in San Salvador are repressive and restrict local development. Instead of fostering civilian loyalty through services, they rule through fear and extraction. Notably, the study suggests that Salvadoran gangs rely on extortion and do not have major sources of criminal rents such as retail drugs. If so, this may help explain the relative absence of gang rule and legitimacy.

In contrast, in eastern Congo, Sánchez De La Sierra (2020) shows how roving warlords turn themselves into stationary bandits and deliver security and adjudication when mining rents grow large and can be taxed. While some of the motive to govern comes from the fee collection itself, some of the governing motive may come from the need to protect their rents from other roving bandits and the state.

There are also parallels to a counter-insurgency literature arguing that rebel groups offer justice and welfare services to capture civilian "hearts and minds."[2] This literature argues that military action plus state services can raise state legitimacy and crowd out insurgents (Berman and Matanock, 2015). This raises the question: Why would village insurgents be crowded out when city gangs are not? One possible answer is that rural insurgents have wide areas of operation, where no one town or village is strategic, while urban gangs have small, well-defined territories and are hemmed in by rivals. Gangs have nowhere else to go.

We can see this rural–urban difference among criminal groups as well. Studying the attempted pacification of Rio's favelas, Magaloni et al. (2020) show that gangs resist state incursions more violently in the neighborhoods they rule. In Mexico, however, the drug trafficking organizations have a more diffuse base of operations. Many routes can carry their product to the United States. Thus, crackdowns in one city simply displace traffickers and violence to less aggressive municipalities nearby (Dell, 2015).

Policy-wise, our results suggest governments will need to find other strategies to counter gang rule, perhaps by trying to shape local norms of combo legitimacy, or undermining drug profits. But caution is warranted. Our interviews, model, and empirical results suggest that criminal rents and the need for legitimacy discipline gangs, pushing them to be less extractive and more focused on citizen welfare. Undermining drug profits could result in more ruthless and extractive organizations, which points to a difficult policy trade-off.

This still gives governments clear incentives to improve their presence and protection

[2]Berman et al. (2011, 2013); Crost et al. (2016); Beath et al. (2012); Albertus and Kaplan (2013). This is part of a more widely-studied phenomenon of insurgent governance, when rebels try to foster civilian support not only in a contest for rents but also for control of the state itself (Arjona, 2016; Kasfir, 2015).

services. Not only do citizens value these services, but they also give criminal organizations incentives for restraint. That said, other evidence from Medellín suggests that state building will be challenging on the margin. Blattman, Duncan, Lessing, and Tobon (2021b) examine a 2–year experiment that increased non-police state presence in dozens of small neighborhoods. They find that it is hard to shift citizen impressions of state responsiveness, quality, and legitimacy even with a large increase in street-level bureaucrats. Where they fall short of expectations, and struggle to deliver, the state may even reduce its legitimacy in the short term. This implies that state-building efforts, and attempts to discipline gangs, may need to be long and sustained (like the one in this paper).

This study also engages a broader literature on the economics of organized crime. This literature began with studies of the origins, internal organization, and incentives of these illegal firms and primitive states (Schelling, 1971; Fiorentini and Peltzman, 1997; Konrad and Skaperdas, 1998). More recently, there has been a surge of interest in international organized crime, including the personnel economics and career paths of gang members (Khanna et al., 2019; Sviatschi, 2018; Carvalho and Soares, 2016); studies of market structure and the production of violence (Castillo and Kronick, 2020; Brown et al., 2021; Bueno de Mesquita, 2020); the effects of exogenous supply and demand shocks on competition and violence levels (Castillo et al., 2020; Dube et al., 2016; Sobrino, 2019; Limodio, 2018); the role of prison systems in strengthening organized criminal groups (Lessing and Denyer Willis, 2019; Tobón, 2020); and the historical origins of drug cartels (Murphy and Rossi, 2020). There are also parallels between gangs strategically increasing rule in response to state presence, and a political economy literature on how organized criminals influence elections (De Feo and De Luca, 2017; Alesina et al., 2019; Dal Bó et al., 2006; Acemoglu et al., 2020).

Finally, methodologically, we demonstrate the usefulness of large-sample, primary qualitative data in economics, especially in informal and clandestine settings. The field's beliefs about organized crime are based on possibly unrepresentative groups, especially the Sicilian mafia and one Chicago gang. We try to demonstrate the value of rigorous qualitative methods, primary data collection, and ethnographer–economist collaborations. Many of our most important insights, our theory, and the interpretation of results all come directly from thorough, large-sample, systematically chosen and documented interviews and observation. Multi-method investigations such as this one are essential to understanding, regulating, and reducing complex problems like organized crime.

Table 1: Interviews and focus groups by type of respondent

Respondent type		Total # of participants
Criminal actors	Incarcerated subjects:	
	Active or former combo	39
	Active or former razón	13
	Other illegal organization	3
	Subjects outside prison:	
	Active combo member	22
	Former combo member	16
	Active razón member	4
	Active faction member	1
	Other illegal organization	20
Community	Member	127
	Leader	17
	Shopkeeper	10
Experts	Criminal group experts	9
	Other	1
Public servants	City officials	21
	Police (active or former)	15
	Prosecutor (active or former)	4
Total		322

2 Data and methods

2.1 Qualitative interviews

Most information on criminal markets and organizations comes from secondary sources, such as judicial proceedings or police investigations.[3] These are largely unavailable in Colombia, and so we conducted primary interviews to collect information on illicit markets, group organization, business operations, internal organization and performance incentives, career paths, inter-group relations, civilian governance, and violent conflicts.

Over five years we interviewed 118 criminal leaders and members across 41 groups. These 41 include 28 combos as well as 13 higher-level, mafia-like organizations called *razones*, discussed below. Table 1 lists respondents by type. Our highest-ranking sources are deputies to the most powerful crime bosses in the city. Most are lower in the hierarchy.

This is a convenience sample of criminal actors who were willing to speak. We conducted

[3]Some prominent examples include case studies of the Sicilian mafia (Gambetta, 1996), New York mafia (Reuter, 1983), pirates (Leeson, 2007), and Brazilian and American prison gangs (Skarbek, 2014; Lessing, 2017). Some notable examples of primary sources include Levitt and Venkatesh (2000) on a defunct Chicago gang, Sánchez De la Sierra and Titeca (Sánchez De la Sierra and Titeca) on corrupt Congolese traffic police, Sanchez-Jankowski (1991) who was a participant observer in a large number of US gangs, and Lessing and Denyer Willis (2019) who obtained internal records of Brazil's Primeiro Comando da Capital.

roughly half the interviews in prisons, typically in a wing reserved for high- and middle-ranking criminals. Most continue to run their group's activities while imprisoned. In the beginning, prison wardens announced that anyone who would like to meet with university professors could meet us at a particular room and time. Following this, our sources might or might not continue to make appointments. Some referred us to additional sources. Because the prison affords little privacy, most interviews took place in public areas, out of earshot from most inmates or prison guards, but in public view, usually for roughly 1–2 hours.

Eventually, we developed criminal contacts outside of prison. In contrast to the self-assured, accessible, and surprisingly candid leaders in prison, initially we found it difficult to speak to outside members, especially lower-ranking ones. Besides being more vulnerable (some are fugitives), they also often seemed to lack the experience, power, and confidence to feel safe speaking with academics. To improve access, we hired the city's main organized-crime journalist as a consultant, to provide introductions as well as conduct his own interviews and analysis. We also hired a government gang outreach worker (himself a former gang member and former prison gang leader), who became a full-time research associate. He conducted structured interviews with criminal subjects in his personal network.

We believe our subjects spoke to us for several reasons. In prison, our interviews offered subjects a respite from routine and a chance to exhibit their expertise and insights. Interviews also posed little risk, since most subjects had already been prosecuted for the criminal activities they described. They were generally flattered by academic attention, and many harbored hopes of being the subject of books. Finally, some leaders remarked that the government underestimated their strength, that this interfered with bargaining, and that we could resolve this as we seemed to have a more accurate understanding of the situation.[4]

Finally, we also interviewed local crime experts, members of the Metropolitan Police and the Attorney General's office, and also obtained confidential internal law-enforcement reports. Our research assistants conducted 153 interviews with community leaders and members in around 108 neighborhoods, mainly on the subject of citizen interactions with organized criminals and use of their services and governing activities. We also returned during the coronavirus pandemic (and city-wide lockdown) to interview criminal group members and community leaders about gang governance during the crisis.

[4]Some leaders explained that with the end of Colombia's civil conflict, they expected the government to turn to organized criminal groups with renewed intensity. They hoped for a "peace process" that involves *sometimiento*—submission to justice and a surrender of some of their gains in exchange for a path to exit. One problem, they told us, is that the government does not recognize the true strength of organized crime. In effect, some criminal leaders viewed our study as reducing asymmetric information and facilitating negotiation. We accepted the possibility of playing this role only because we felt it was small, and because on balance it should reduce the possibility of state-gang violence.

Ethics and human subjects protections We had several strategies for maintaining trust, safety, and confidentiality of criminal group members. Above all, we were transparent about our research aims, that we were speaking to other groups and the government, and that we advise the civilian government (but not the criminal justice system). We made every effort to preserve anonymity and confidentiality, while advising subjects in consent scripts of the potential limits to our ability to do so. With prison populations, we also took great efforts to ensure that our interviewees faced no pressure to speak to us. (It is worth noting, however, that our subjects were generally shrewd and powerful businessmen who in many respects are in control of their decisions and lives in the prison, if not the prison itself.)

We consulted extensively with the University of Chicago and Universidad EAFIT human subjects committees, and we obtained written support and assurances of noninterference from the Mayor, the head of the National Prison Authority, and the Colombian Minister of Justice. We also consulted with multiple journalists who specialize in organized crime, who related that they had never been asked by the criminal justice system to betray sources or materials. For this reason, in practice, we believed that speaking to us carried minimal risks to the leaders. Nonetheless, our consent scripts explicitly highlighted those risks.

Qualitative methods Amidst the large volume of primary data collection in economics, informal qualitative research has become widespread. Though they seldom discuss it explicitly, today economists probably do more interviews and observation than ever before.

One challenge with this informal qualitative turn, however, is that it is often confined to small and select samples, is not systematically documented or analyzed, and hence is subject to common biases of selection, recall, and salience. Just as there have been high returns to rigorous quantitative methods, there are surely advantages to better qualitative work, especially in informal economies, corrupt politics, armed groups, and crime.[5]

To do so, we formed a collaboration between two economists and two ethnographers, also employing a journalist and a former criminal group member that we trained as a qualitative researcher. We developed semi-structured interview guides, and adjusted them to investigate hypotheses as we developed them. We recorded and transcribed interviews when possible. When not possible, especially in prison, we took notes and wrote them up formally after each interview. We also sought to verify our observations with multiple sources. For most topics we discuss we have 2–3 sources between gang members and experts.

To organize the vast number of interviews, we created a private encrypted wiki we call *WikiCombo*. A collaborative wiki was a good fit for the networked, non-linear nature of

[5]There is an extensive anthropology and sociology literature that does this, but much of this work rejects conventional economic theory. Notable exceptions include a number of qualitative investigations of informal (not illicit) markets and communities (e.g., Bliss and Stern, 1982; Ensminger, 1996; Venkatesh, 2006).

the data, especially when collected by several contributors. We uploaded and encrypted all primary and secondary sources. We created inter-linked pages on key research themes, individual combos and people, neighborhoods, events such as conflicts, and so forth. Factual claims are linked directly to original transcripts. Every text change and its contributor is tracked and is reversible. Finally, the wiki is an ongoing collaborative tool and sources consented to have their anonymous comments shared with other researchers.

2.2 Survey data and measurement strategies

In 2019 we surveyed nearly 7,000 residents and businesses on the degree of state and combo rule, the perceived legitimacy of both, and levels of taxation and payments to combos. The survey was representative of all 223 low- and middle-income neighborhoods in Medellín, plus nine neighborhoods bordering Medellín in other municipalities (see Figure 2). We randomly sampled 2,347 of the city's 14,600 blocks, stratified by neighborhood, then randomly sampled approximately two households and one business on each block.

Addressing measurement error Naturally, one should be concerned that citizens may misreport gang activities. They may feel uncomfortable talking to outsiders or embarrassed to admit the role of the combo. If so, then our data would underestimate the role and legitimacy of the combo. Several pieces of evidence suggest that survey respondents answered questions freely and truthfully, however.

First, combos are a routine part of everyday life, and in both qualitative interviews and surveys we found that most people spoke freely, at least when interviewed in private. Thus we conducted all surveys anonymously, alone, and indoors. We also refined survey questions after dozens of qualitative interviews, fine-tuning language, questions, and approach to elicit truthful answers.

Second, as we discuss in Section 5, we do not see evidence of the most worrisome kinds of under-reporting. For instance, we look at whether residents from places with higher levels of gang governance leave a larger share of questions unanswered. We find no evidence of such situation. We provide more details on measurement error in Section 5 and Appendix D.

3 Descriptive analysis

Gangs are generally clandestine organizations with obscure operations and motives. In order to develop theories of gang organization and rule, to understand their relationship to the state, and to evaluate programs, we must first establish some basic facts. While case studies

of gangs and gang rule abound, we know of no prior systematic data collection on a large sample of comparable gangs. This descriptive analysis shapes the theory and the empirical results to follow, and is an equally important part of the analysis.

3.1 The state

Medellín has 2.6 million people, with almost 4 million in the metro area. Per capita annual income is roughly $11,500, adjusted for purchasing power parity. The city is divided into 16 urban *comunas* plus 5 peri-urban *corregimientos*. The comunas are formally divided into 269 neighborhoods or *barrios*. Each barrio has an elected community council to manage various aspects of community affairs.

Medellín also has a well-organized, professional bureaucracy with high fiscal capacity and broad-based public services. With a huge commercial sector, the city has ample revenues. Two organizations are responsible for order: the police and the Secretariat of Security.

The Secretariat is a large civilian organization with thousands of staff. It sits directly beneath the Mayor and is the city's primary organization for setting security policy and investing in security infrastructure. Roughly 2,500 civilian staff provide numerous services to residents, including responding to various emergencies and street disorder, directly resolving community disputes and domestic violence, and regulating the use of public space.

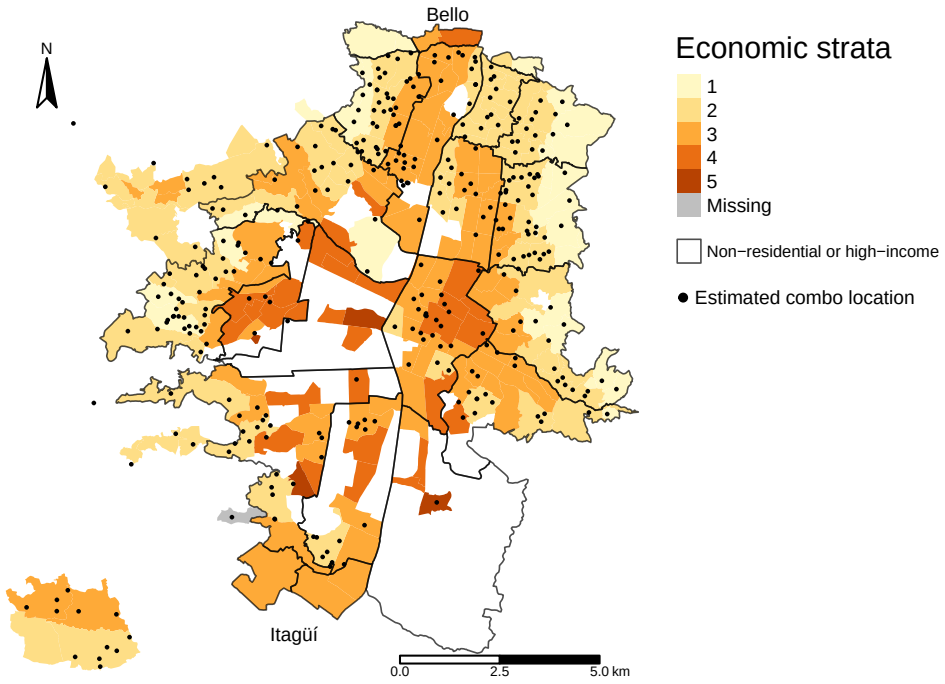
The Metropolitan Police are independent from the city government; they are part of the National Police, a branch of the Defense Ministry. While there are common charges of corruption and poor responsiveness, the National Police is fairly professionalized, particularly in comparison with other Latin American countries. There are 280 officers per 100,000 people in Medellín, similar to cities like Los Angeles. That said, street cops are greatly outnumbered: our data suggest that there are roughly a dozen combo members for every officer.

3.2 Combo organization and operations

Virtually every low- and middle-income residential neighborhood has a local combo. Our 2019 combo census identified 380 active combos—354 in Medellín and the rest in the wider metropolitan area (Blattman, Duncan, Lessing, and Tobon, 2021a). We do not have detailed borders for each combo, but Figure 2 plots our assessment of an intersection or other known location for each combo.

We have detailed organizational data on 12 combos. Almost all of these have a core of 15 to 40 permanent, salaried members (one has nearly 100, but this is unusual). The combo territories we observe are sometimes no more than a few square blocks, but borders are usually long-standing, well-defined, and known to most locals. Combos also tend to be

Figure 2: Combo census: Estimated locations, with barrio income level



long-lived. Many have been present for decades in some form, as younger generations take over from older ones. While there are of course changes in the size, territory, names, and even existence of some combos, most in our sample have been relatively stable over the last decade at least, and our broader interviews suggest this is true of most combos.

Combo revenues come from four main sources. Practically every combo has a local monopoly on retail drug sales in their neighborhood, which occur at defined locales known as *plazas de vicio*. This is typically their most profitable activity. A large number also charge a security fee to at least some residents and businesses, typically in return for protection services (discussed at length below). About a third also engage in a local loan-sharking practice known as *gota a gota* ("drop by drop"), according to the survey. Finally, many combos collect debts for a fee, and also manage, regulate, or participate in local consumer goods markets, such as cooking gas, *arepas*, and eggs.

A companion paper describes the personnel economics and market structure of the combos (Blattman, Duncan, Lessing, and Tobon, 2021a).[6] Members tend to be poor, uneducated young men aged 15–35. Most were born, grew up, and still live in the neighborhood they control. Even low-ranking combo members tend to be well-paid, earning a salary equal to the median salary in the city. They are headed by a leader called a *coordinador*. In general,

[6]See also Martin (2012) and Giraldo et al. (2014).

combos are organized internally by product or service line. Combo members are paid a salary for their role. Other business lines, such as consumer goods sales and loan-sharking, are often given to individual combo members as a personal, local monopoly to operate and exploit. These are key sources of entrepreneurial side income.

Most combos are small and autonomous; horizontal integration across neighborhoods is rare. In our companion paper, we argue that while such integration increases monopoly rents and mitigates inter-combo conflict, it also exposes combos to prosecution and requires managerial capacities that many combos lack.

Finally, Medellín's combos form the base of a pyramid of criminal organization. Above them are roughly 17 mafia-like groups sometimes called *razones*. Razones are typically the wholesale suppliers of drugs to the combos' street retail operations. Most combos have a longstanding business and military alliance with a razón. A small number of combos are vertically integrated into their razón. For the most part, however, combos are small autonomous firms with a long-term relational contract with the razón as their supplier.

3.3 Combo governance

A central component of governance is protection: order, security, and property rights enforcement. Many of Medellín's combos provide these services, often on a private, fee-for-service basis. Examples include dispute resolution, informal contract enforcement, recovering stolen items, and private security for stores, vehicles, and other property. This is a relatively recent phenomenon, starting at the beginning of the 2000s.[7]

The gangs also provide less excludable, public forms of protection, including regulating fights, disorderly conduct, and drug consumption on their blocks. At least one combo even installed security cameras for a time. Often they provide these services in exchange for weekly security fees, a kind of taxation. Even these more public goods, however, are partially excludable. For instance, combos often focus their public protection on blocks where they already have many private customers. Some provide hot-lines to those paying security fees.

Of course, governance also includes material public goods such as infrastructure, as well as collective decision-making and coordination. Our interviews and surveys found that combos rarely offer such services. Infrastructure is provided almost solely by the state, while informal

[7]During the 1980s and early 90s, Medellín gangs—led by the infamous Pablo Escobar—were engaged in open war with the Colombian government. Between 1985 and 2000 more than 66,000 people were murdered in the city (in a population of 2 million). War with the government was followed by inter-gang wars that ended in the early 2000s. Only in this "peaceful" period did gang governance emerge, along with the turn from violence to the extraction of illicit rents in local markets (e.g., Martin, 2012). In our companion paper, we document how this is characteristic of internal governance and conflict regulation by gangs (Blattman, Duncan, Lessing, and Tobon, 2021a).

leaders and elected neighborhood councils manage most local collective decisions. Instead, combos tend to specialize in services that are at least partially excludable, and those that benefit from coercive power. In the remainder of the paper, we use "governance" as shorthand for this set of protection services in which both gang and state participate.

Levels and variation of governance To measure governance, we asked residents how frequently each actor responded to 17 common disputes and forms of disorder (12 from residents and 5 from business-owners). We identified these through our qualitative work. Table 2 reports scaled responses, where 0 = Never, 0.33 = Occasionally, 0.66 = Frequent, 1 = Always. We create average indexes of *State* and *Combo governance* (0 to 1), as well as the difference between them, *Relative state governance*, which can vary from –1 to 1.

The average response for any service by either provider was seldom greater than 0.5, suggesting that for residents, neither the state nor the combo are reliably responsive to disputes and disturbances. In absolute terms, combos were most responsive to unpaid debts, property crimes (car thefts, home robberies, muggings, etc.), and public disturbances (threats and fights on the street). Combo involvement was somewhat lower for inter-neighbor disputes such as noise complaints and property infringements. In relative terms, combo response was generally lower than the state's, but higher in five situations: muggings and theft prevention, business and household debt collection, and street fights.

These averages conceal much variation across neighborhoods. Figure 3 maps relative state governance by barrio. In some, the combo is the dominant provider of protection. In others the state is dominant. Note, however, that high levels of state governance do not imply combos are absent. Nearly every neighborhood has a strong combo presence, running drug corners and other operations. Generally, the combos have chosen not to sell private protection in those neighborhoods.

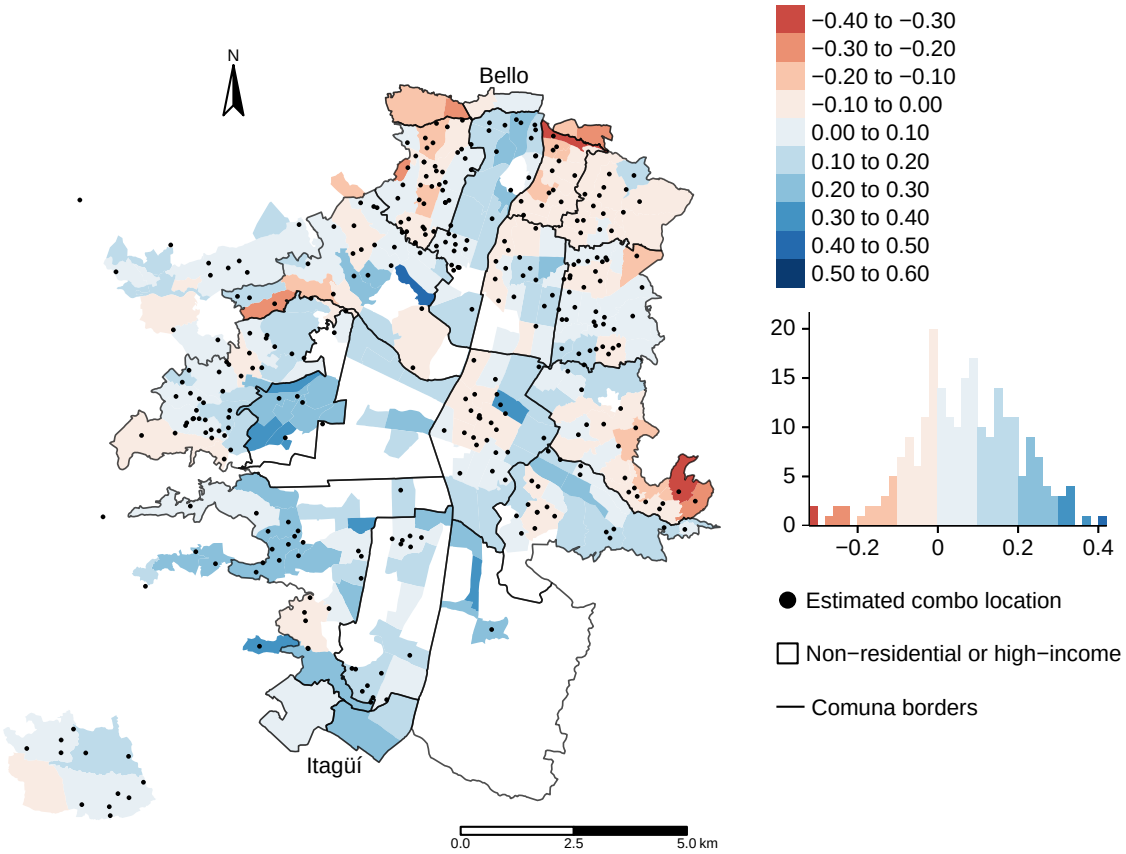
Legitimacy We also asked residents (but not business-owners) about combo and state legitimacy: how much residents trust each actor; whether actors were fair; whether residents were satisfied with each actor; and whether residents thought their neighbors trust and are satisfied with each actor. We averaged these responses into unit indexes for state and combo legitimacy. Table 2 reports barrio averages. On average, residents rate their trust and satisfaction of the combo lower than the state, although the difference is not always large. Not surprisingly, there is a tight correspondence between combo governance and legitimacy. Figure 4 illustrates this, plotting combo legitimacy against combo governance.

Table 2: State and combo governance and legitimacy, barrio survey averages, 2019

	Frequency/Rate (0-1 Scale)				Relative State Governance
	State		Combo		
	Estimate	SD	Estimate	SD	Difference
	(1)	(2)	(3)	(4)	(5)
Governance Index	0.41	0.27	0.33	0.29	0.08
How often they intervene when:					
HH: Someone is making noise	0.42	0.38	0.19	0.30	0.23
HH: Home improvements affect neighbors	0.41	0.37	0.24	0.33	0.16
HH: There is domestic violence	0.50	0.37	0.33	0.37	0.16
Biz: Someone disturbs a business	0.50	0.38	0.36	0.38	0.14
HH: Two drunks fight on the street	0.52	0.36	0.39	0.37	0.13
Biz: You have to react to a robbery	0.53	0.37	0.39	0.39	0.13
Biz: It is necessary to prevent a theft	0.46	0.36	0.37	0.39	0.09
Biz: Businesses in this sector are robbed	0.43	0.39	0.35	0.38	0.07
HH: A car or motorbike is stolen	0.47	0.37	0.41	0.38	0.05
HH: People smoking marijuana near children	0.30	0.36	0.25	0.36	0.05
HH: You have to react to a robbery	0.46	0.36	0.44	0.38	0.02
HH: Someone is threatening someone else	0.41	0.36	0.40	0.37	0.01
HH: Someone is mugged on the street	0.39	0.36	0.40	0.38	-0.01
HH: It is necessary to prevent a theft	0.39	0.36	0.41	0.38	-0.02
HH: Kids fight on the street	0.28	0.35	0.31	0.36	-0.03
Biz: Someone does not want to pay a debt	0.18	0.31	0.24	0.35	-0.06
HH: Someone refuses to pay a big debt	0.21	0.31	0.36	0.37	-0.16
Legitimacy Index	0.57	0.21	0.43	0.28	0.14
When solving problems in the neighborhood:					
How much do you trust the...	0.57	0.30	0.36	0.36	0.21
How fair is the...	0.55	0.27	0.41	0.35	0.14
How much do your neighbors trust the...	0.57	0.28	0.47	0.36	0.10
How would your neighbors trust the...	0.59	0.23	0.50	0.29	0.09
How do you rate the...	0.60	0.22	0.51	0.28	0.09

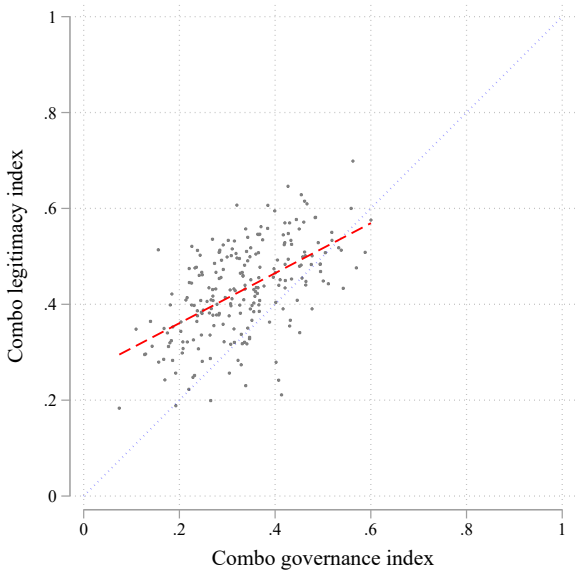
Notes: Different governance questions were asked of household (HH) and business (Biz) respondents. Only households answered legitimacy questions. The survey is representative of Medellín's 224 low- and middle-income barrios, with 20–25 respondents per barrio. Governance scales correspond to: 0 = Never, 0.33 = Occasionally, 0.66 = Frequently, 1 = Always. Legitimacy scales correspond to: 0 = Nothing, 0.33 = A little, 0.66 = Somewhat, 1 = Very.

Figure 3: Relative state governance by barrio, 2019



Notes: Blue indicates state governance > combo governance; red indicates combo governance > state governance. Each barrio's value is the average relative state governance (state–combo governance index) for all 17 items from Table 2. We did not survey high-income barrios.

Figure 4: Relationship between combo governance and combo legitimacy, 2019



Notes: Each dot is a barrio average, and the dotted line indicates fitted values. We did not survey high-income barrios.

Efficacy Combos also outperform state agents in some respects: 67% of survey respondents said the combo was easy to contact compared to 63% for the police and 32% for the Mayor's office. They also said the combo responded rapidly 58% of the time compared to 41% for the police and 27% for the Mayor's office.

This is not entirely surprising. With the exception of the police, the city's street-level bureaucrats are rarely available outside of business hours; offices are closed on Colombia's frequent holidays; and due to peculiarities in municipal budgeting and labor agreements, every December to January a large proportion of city staff on contracts are not working. The combo, by contrast, is always present.

Combos have other advantages. They have more local knowledge and deeper networks than state bureaucrats, and even local police. Community leaders have good information too, but combos have organized means of coercion to enforce rules and deals. Indeed, combos' freedom to use force can exceed that of the state: they can carry out swift and sometimes violent sanctions that some residents demand, such as expelling an abusive husband from the neighborhood. Also, whereas the state and community leaders are expected to be impartial and consistent, some combos openly resolve disputes and enforce contracts in favor of those who hire them or who are most closely connected. Residents have few mechanisms for accountability or voice in shaping and enforcing combo rule.

These differences help explain why many residents are conflicted about combo rule. Many

report they are happy to have access to both the combo and the state for protection. Just 46% of survey respondents agreed to the question that the neighborhood would be better off without the combo. Elaborating, some said they feared the vacuum of authority that might open up without this local actor. Others were simply satisfied with the work of the *muchachos* ("local boys"), a common term for combo members.

3.4 Why do combos govern?

Our interviews point to three important motives for gang rule: (i) direct revenues from protection as a business line; (ii) indirect benefits of governance on other business lines, especially drug retailing; and (iii) intrinsic rewards from ruling.

Motive 1: Protection as a business line First, for some combos, protection services are an important business line that yields significant revenue. For services such as debt collection or dispute resolution, combos commonly charge on a fee-for-service basis.[8]

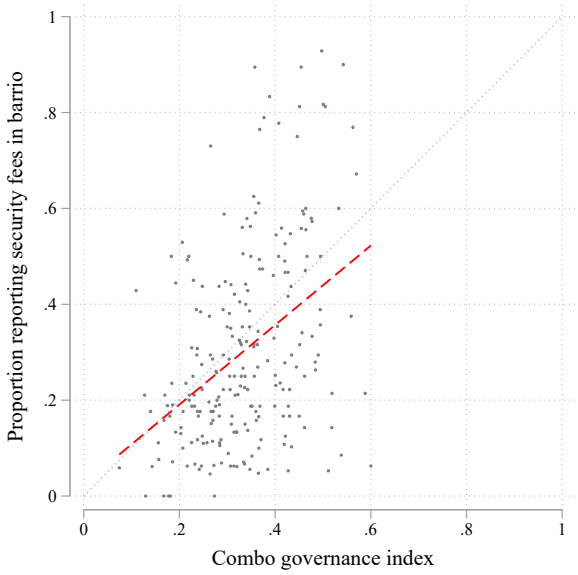
Revenues from other services, such as security and protection for homes and shops, are akin to semi-voluntary taxes or a subscription. Residents and businesses typically call this tax a *pago por la vigilancia* ("security" or "surveillance fee") or, more colloquially, a *vacuna*—literally, a vaccine. Among the combos where we have internal organization data, most organize vacuna collection and protection services as a specialized unit with dedicated staff. The manager of this business line generally reports directly to the coordinador, and is sometimes referred to as a *relacionista*, or relationship manager.[9]

Most combos tax only a fraction of local businesses and residents. In our survey, 85% reported that the combo charges vacunas in their neighborhood, typically weekly. But within these communities, only a quarter of businesses and a tenth of residents reported being charged this tax themselves. Figure 5 shows that the share of people who report that they or others in the neighborhood pay regular security fees is strongly positively correlated with

[8]As one community leader told us: "If a couple starts fighting, they [the gang] come to a kind of trial and fine them. It is the same with the problems between neighbors; they set fines of 100,000 [pesos]" —Community Leader 14/17, interview 1/1 [08/06/2020]. Another leader explained how "if you fight with someone, regardless of whether you provoked it or not, you must pay between 100,000 and 500,000 [pesos], depending on how serious the fight is. They decide what price to impose. There are also fines for theft. For example, something that happens a lot: a neighbor steals some plants from me, so she must buy or return those plants and also pay the fine to them. The price of the fine depends on what was stolen." —Community Leader 4/17, interview 1/2 [02/22/2020].

[9]As a typical example, one combo member told us how in his group, "Three people are in charge of business collection. Two of them collect in one zone and one of them in another. What they collect is delivered to the coordinator." He also told us that he used to tell businesses that the combo was going to charge 5,000 or 10,000 pesos weekly, and that he gave them a phone number for them to call if there were any disturbances [Criminal Group Leader 1/30, interview 1/6 [06/20/2019] and 2/6 [09/20/2019].

Figure 5: Relationship between gang governance and vacuna payments, 2019



Notes: Each dot is a barrio average, and the dotted line indicates fitted values. We did not survey high-income barrios.

combo governance levels—close to the 45 degree line. (Our survey experiment, discussed below, suggests these reporting rates are roughly accurate.)

Vacunas are also modest. Among those who pay, median weekly amounts were about US$1 for residents and US$2 for businesses—roughly 3% of business profits and 1% of sales at the median. While 89% of businesses said they disapprove of vacunas, just 27% of businesses said that vacunas were too high. By comparison, municipal taxes on these enterprises are about 6% of profits, and 54% said they were too high. Identifying a tax rate seemed to be relatively straightforward. Typically, a junior member of the relacionista staff observes a business for a few days, using the store characteristics and counting the number of customers per day to select an appropriate tax rate.

It is important not to exaggerate the voluntary nature of the vacuna. While in some places they are voluntarily paid, and in others they are an obligatory but generally-accepted tax, in some neighborhoods they resemble outright extortion. About half of respondents reported that refusing to pay the vacuna would result in threats or assaults. At the same time, that implies half did not expect coercion. In these cases, respondents said that the combo would simply stop providing them security if they failed to pay.

Motive 2: Intrinsic rewards In addition, power, authority, and the loyalty of subjects can be their own reward. Some combo leaders reported taking pride in ruling, or simply

enjoying the status and moral legitimacy it offers. As one said, "Personally, doing good work feels good. You can be the worst bandit, but you can also have a good heart of your own."[10]

Some also describe governance as a moral obligation or social duty to their community. Many combos emerged in the 1980s as local defense forces fighting left-wing militias affiliated with rural guerrilla movements. Today, most gangs in Medellín retain a socially conservative and anti-communist ideology. Many of the leaders we met saw themselves as responsible for upholding moral codes, protecting women and girls, and preserving conservative social mores. Others described themselves as critical "*anti-subversivos*"—bulwarks against socialism. They resent the lack of recognition of their contribution to Colombia's civil war.

Combos may derive utility from the status that their authority confers. In addition to pride and any appreciation of community respect, some of our interviews suggest respected combos may enjoy easier access to women.

Motive 3: Indirect impacts on other business lines The final and potentially most important motive for governing is that it helps protect the gang's physical security and illicit income from competitors and, perhaps more importantly, police.

First, providing order wins the loyalty of residents, leaders told us. Loyal residents rarely inform on combos to police, and may even actively protect them from police repression. Numerous combo leaders made this point: "The community shields you according to your behavior," explained one. "If you do not have the community in your hands and at your back, you have nothing. That is who takes care of you."[11] As another explained, "Caring for the neighbors gives a criminal more security. When the community feels comfortable and grateful, they open their houses, so that if you have to hide from the police, the community is going to welcome you. The community goes out to defend you."[12] In another case, asked about the benefits of governing, one member responded that they do it so that, "the neighbors love us, do not rat us out to the cops, watch us doing our stuff and do not interfere, and let us know when the police are coming."[13] Finally, as one public prosecutor told us, "They're very interested in winning over the community. That's why it's so hard to get witnesses against them."[14]

Moreover, providing local order may directly reduce police presence. As one combo leader put it, "There is a good relationship with the people," and therefore, "it is easier to bring order in the sector and so the police do not have to come around."[15] When police patrol

[10]Criminal Group Member 6/40, interview 2/3 [02/11/2020].
[11]Criminal Group Leader 13/30, interview 1/2 [05/02/2019].
[12]Criminal Group Member 6/40, interview 2/3 [02/11/2020].
[13]Criminal Group Member 8/40, interview 1/1 [12/30/2020].
[14]Official 12/17, interview 1/5 [10/16/2019].
[15]Criminal Group Member 6/40, interview 2/3 [02/11/2020].

or respond to service calls, it can scare off drug buyers, require a bribe to the officers, or increase the risk of a seizure. If crime and disorder are high, moreover, local police feel pressure from superiors to crack down. One active combo member offered a vivid example: "The police station is across from our headquarters and they never bother us. They know where our drug corners are and who works there. That's why it's important to keep the neighborhood calm: if nothing bad happens, the police don't squeeze us and let us work."[16]

We were also told that combos are especially worried about new or specialized police units in their territory. As one expert and former prison-gang leader explained, "there is always a police presence, but combos strike non-interference deals with the regular beat cops. When public order gets disrupted, the police must act and officers not part of the deal arrive. The area becomes visible and combos' activities become more vulnerable."[17] This suggests that even where bribery is part of how combos avoid police repression, governance can play a key supporting role, by avoiding the attention of less-corrupt police as well.

While our interviews suggest that combos mainly defend against the state, they also face rival criminal-market entrants, and the benefits of establishing loyalty extend to inter-gang competition as well. Even where combos enjoy firm local monopolies on crime and extortion, they must defend these against local coups, neighboring combos, and sometimes their razones (which could invade or sponsor a coup from combo or non-combo youth from the neighborhood). Residents' loyalty can mitigate these threats, while dissatisfaction with local combo rule could lead to collaboration with rivals. Coups are unusual, but this could be an equilibrium outcome. Combo leaders' wariness suggests that their rarity may well be due to vigilant deterrence.

These motives—fostering civilian loyalty and preventing police and rival entry—may explain some of the patterns we observe in the protection market: First, even though combos have the coercive power to extract higher sums, the tax itself is modest. It may also explain low rates of tax incidence in some neighborhoods, including the fact that combo rule and vacunas decrease as distance to the gang grows (see Appendix Table C.1).

Second, few of the combo leaders we spoke to saw protection as a highly profitable business line, and some said they provided the service at a discount because of the indirect benefits. One former gang member described fees and fines for dispute-resolution services as a way to limit demand and deter disputes, rather than as a money-making strategy.[18]

Third, combos avoid charging businesses whose ability to pay is low or whose loyalty is more fragile. For instance, when asked why some grocers were targeted and others were

[16]Criminal Group Member 5/40, interview 1/2 [10/09/2019].
[17]Criminal Group Leader 24/30, interview 5/5 [12/14/2020].
[18]Criminal Group Leader 24/30, interview 5/5 [12/14/2020].

not, one combo member explained that some were more likely to denounce the combo to the police if pressed to pay, and it is better to keep the population loyal.[19] Another explained that "There is no fixed fee; it is voluntary." He described how, when a new business opens, the combo talks to the owner and agrees on the weekly vacuna based on the size and type of business. If the business is doing badly, however, the combo does not demand payment. "We are here to help," he explained.[20]

For similar reasons, this indirect motive also helps explain the relatively extreme extortion of bus lines. Almost every combo with a passenger bus route in their territory charges drivers steep fees in return for little to no services. Bus lines are small, cash-heavy businesses with many small owners. Most importantly, both drivers and owners come from outside the combo's territory, and residents do not pay much attention to bus extortion.[21] In short, all aspects of the protection business, from extortion to semi-voluntary vacuna taxation to fee-for-service provision, are disciplined by a need for the loyalty of local residents.

Finally, the indirect motive is consistent with some of the correlations in combo governance we observe. Appendix Table C.1 reports simple correlations between block combo governance levels and neighborhood characteristics. We observe more combo governance near high drug value areas (proxied by the value of police drug seizures from 2014–19). We also see more governance close to their headquarters—both the combo's and their affiliated razón's. There is also more governance when there are more competing combos in the vicinity, in poorer neighborhoods, and in neighborhoods on the higher slopes of the mountainsides. Of course, none of these are causal relationships, and so we have to take them with caution. Rather, the focus of the remainder of this paper is how the presence of state protections services affects combo governance.

4 Conceptual framework

To structure our thinking, here we show how the conventional wisdom—that state and gang rule are strategic substitutes—is consistent with a gang's optimal response under duopolistic competition. Any model of imperfect competition should produce this result, and we illustrate with Cournot competition, where each side chooses a fixed quantity of protection services to provide and let prices clear the market.[22] (In an Appendix, we show how the

[19] Criminal Group Leader 23/30, interview 2/2 [12/28/2020].

[20] Criminal Group Leader 1/30, interview 1/6 [06/20/2019].

[21] As an aside, buses pass through many territories, and combos have solved the common pool resource problem by developing a collective norm whereby the combo at the bus' point of origin holds extortion rights.

[22] Note that Cournot fits some of our stylized facts well—especially that governing requires investments and advanced commitments, and that it is hard to adjust output capacity quickly. In modeling duopolistic competition, however, note that we abstract away from competition between combos. We do this in part

gang's response would be similar in other forms of imperfect competition, as well as a model of stationary bandits competing to provide public goods.)

Here we are mainly interested in the gang's best response to variations in state rule. In particular, what are the circumstances in which criminal and state rule will be strategic complements, and a gang has incentives to increase its governing in response to a more active state? To keep matters simple, we do not model the state's full objective function or its optimal policy choice. The approach taken here is intended to motivate the quasi-experimental analysis, in which highly localized variation in state rule around new administrative borders is essentially accidental.

4.1 Cournot competition in the market for protection

Consider a gang g and a state s offering distinct but substitutable services to residents in quantities q_g and q_s at a constant marginal cost of production c_i. We can write each organization i's utility function as:

$$V_i = p_i q_i - c_i q_i. \tag{1}$$

Price is determined by a linear inverse demand curve:

$$p_i = a_i - \beta q_i - \gamma q_j, \tag{2}$$

where $\gamma \in (0, 1]$ implies the two services are substitutes, and $\beta > 0$ implies downward-sloping demand.

We are interested in whether gang rule is crowded in or out when there is an exogenous increase in state governance: $\frac{\partial q_g^*}{\partial q_s}$. We derive each organization's best response function, their equilibrium values of q_g^* and q_s^*, and this comparative static in Appendix A, showing that:

$$\frac{\partial q_g^*}{\partial q_s} = -\frac{\gamma}{2\beta}. \tag{3}$$

So long as the two services are not complements, this comparative static implies that increases in one duopolist's supply of protection will reduce the other's—what we call "crowding out".

because gangs are insulated from territorial competition by the razones, who protect gang property rights.

4.2 Additional benefits to governing

Above, gang leaders described additional benefits to governing beyond the money it brings in as a business line. We summarize these diverse motives by adding a single, stylized term, $\rho(q_i, q_j)\pi_i$, to the players' objective functions:

$$V_i = p_i q_i - c_i q_i + \rho(q_i, q_j)\pi_i. \tag{4}$$

Set up this way, π_i is the return to full control of the neighborhood. For example, π_g includes the illicit rents from unimpeded retail drug sales, but it also includes non-material benefits, such as status, access to women, and other intrinsic rewards from loyalty and rule. The state has its own distinct π_s in the form of electoral rewards, achievement of policy aims, or preferences for dominance and citizen loyalty.

Meanwhile, $\rho(\cdot)$ scales each organization's ability to capture, retain, or enjoy these benefits. We can think of it as the share of π_i each player enjoys, one that is increasing in own governance and decreasing in the other's, such that: $\frac{\partial \rho(q_i,q_j)}{\partial q_i} > 0 > \frac{\partial \rho(q_i,q_j)}{\partial q_j}$. Importantly, however, we remain agnostic here about whether $\rho(\cdot)$ exhibits increasing or decreasing returns to own and other's governance provision.

The elasticity of gang governance to state governance now becomes:

$$\frac{\partial q_g^*}{\partial q_s} = \frac{\lambda \pi_g - \gamma}{2\beta - \delta \pi_g}. \tag{5}$$

where $\lambda = \frac{\partial^2 \rho(q_g,q_s)}{\partial q_g \partial q_s}$ represents the cross-partial derivative between gang and state governance, and $\delta = \frac{\partial^2 \rho(q_g,q_s)}{\partial q_g \partial q_g}$ reflects the rate of increasing or decreasing returns to governing.

Equation 5 helps us to identify two main channels by which more state governance could crowd in gang rule.

Case 1: Strategic response to state rule by the combo ($\lambda \pi_g > \gamma$) This case corresponds the closest to our gang leader interviews. When the state increases protection, they threaten the gangs' share of rents and non-material benefits, $\rho(q_g, q_s)\pi_g$. The cross-partial derivative, λ, will be positive when the gang has more-than-proportional returns to increasing its own rule in response to the state's efforts. This could arise, for example, from a gang leader who values dominance and sole allegiance, or who is averse to losing status. Alternatively, we could think of $\rho(\cdot)$ as a contest success function for control of illicit markets, with governing akin to arming. For many such functions and ranges of relative "armament" (loyalty-inducing q_g and q_s in our setting), the optimal response to an increase in opponent's

arming is to increase one's own arming, especially when one is strong to begin with (e.g., Hirshleifer, 1989; Skaperdas and Syropoulos, 1996; Konrad and Skaperdas, 1998, 2012).

Any $\lambda > 0$ will attenuate the state's attempts to crowd the gang out. The larger are λ and π_g, and the more differentiated are gang and state governance services (lowering γ), the more likely that we observe crowding in, where $\lambda\pi_g > \gamma$. Of course, this assumes that the denominator is negative, which is the case when gangs have diminishing returns to their own governance. The alternative brings us to the next case.

Case 2: Increasing returns to a gang's own level of governance ($\delta\pi_g > 2\beta$) Given downward-sloping demand ($\beta > 0$), the denominator in Equation 5 will be negative only if gangs enjoy large increasing returns to their own governance. This could arise, for instance, if residents reward protection with loyalty at increasing rates. While not impossible, increasing returns should not be assumed, and we saw no evidence of them. Generally, therefore, we expect diminishing returns to governance, making combo strategic response (Case 1) the more plausible of this pair of mechanisms.

Finally, outside of Equation 5, there is a third way the elasticity between state and gang governance could turn positive: growth and endogenous demand for protection.

Case 3: State rule generates growth and increases demand for governance An increase in state protection could raise the number and value of transactions in the local economy, thereby increasing demand for governance in general and gang governance in particular. This is a common feature of the political economy literature on stationary bandits, where a state monopolist has incentives to provide public goods to grow the very market they will later tax (Olson, 1993; McGuire and Olson, 1996; Grossman, 1996; Bates et al., 2002).

In our Cournot example, we could model such endogenous growth through the demand curve in Equation 2, writing a_i as an increasing function of q_s. This can produce crowding in ($\frac{\partial q_g^*}{\partial q_s} > 0$) even if we assume no additional benefits from dominance or loyalty (see Appendix A). That said, there are two major caveats. First, the demand effect would have to be quite large to overcome the crowding out that arises from normal duopolistic competition. Second, not all models make this prediction about endogenous demand. The prediction reverts to crowding out if we move away from a traditional model of duopoly to a model of stationary bandits competing to provide public goods for taxes (see Appendix A.2). Nonetheless, endogenous demand could contribute to a positive elasticity of gang rule to state rule, and we will look for evidence of this economic development in the empirical analysis below.

4.3 General formulation

Finally, as the foregoing suggests, our framework was designed to accommodate a variety of modeling approaches. In its general form, gang responsiveness to state rule is given by:

$$\frac{\partial q_g^*}{\partial q_s} = -\frac{\frac{\partial^2 D(q_g,q_s)}{\partial q_g \partial q_s} + \frac{\partial^2 F(q_g,q_s)}{\partial q_g \partial q_s}}{\frac{\partial^2 D(q_g,q_s)}{\partial q_g \partial q_g} + \frac{\partial^2 F(q_g,q_s)}{\partial q_g \partial q_g}}. \tag{6}$$

Here, $D(\cdot)$ represents the direct returns to governing, whether we model it through Cournot competition in private goods, Bertrand competition, or stationary bandits providing public goods. Meanwhile, $F(\cdot)$ represents the additional benefits to establishing rule, dominance, and loyalty—previously $\rho(\cdot)\pi_i$. Appendix A presents these alternative models. In general, whether we treat governance services as public or private goods, the cross-partial $\frac{\partial^2 D(q_i,q_j)}{\partial q_i \partial q_j} < 0$ is negative, making the overall elasticity negative in the absence of additional benefits. Introducing $F(\cdot)$ generally makes the sign of the elasticity indeterminate, for two reasons: first, the cross-partial $\frac{\partial^2 F(q_i,q_j)}{\partial q_i \partial q_j}$ itself is often indeterminate and sensitive to specific values of q_i and q_j;[23] and second, even if $\frac{\partial^2 D(q_i,q_j)}{\partial q_i \partial q_j} > 0$, it must be large enough to outweigh the crowding-out effect of $\frac{\partial^2 D(q_i,q_j)}{\partial q_i \partial q_j}$ in order to flip the sign of $\frac{\partial q_g^*}{\partial q_s}$.

5 Empirical strategy

Figure 6 plots the correlation between state and combo governance in our 2019 survey. We see a positive relationship, consistent with states crowding gangs in. Of course, this cross-sectional correlation could be confounded by any number of factors. Police could locate closer to drug-producing areas, or places with high levels of economic and social transactions could have higher demand for governance of all kinds. Initially, we shared the conventional wisdom, and regarded the positive correlation with suspicion. We looked for a natural experiment, and identified one that created discontinuous jumps in the distance to local state protection.[24]

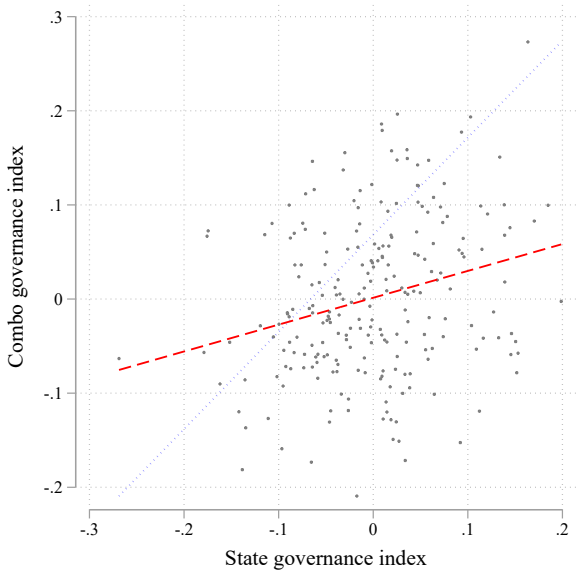
5.1 Medellín's new jurisdictional borders

In 1987, Medellín's elected council divided the city into 16 areas called comunas. Previously, the city was divided into 6 such areas. The new policy subdivided each into 2–3 smaller

[23] This is the case with virtually all contest success functions, for example.

[24] This emulates a strategy by Henn (Henn), who looks at the proximity of chiefdoms to the state.

Figure 6: Correlation between combo and state governance in 2019, adjusted for covariates



Notes: Each dot is a barrio average after partialing out block characteristics. The dotted line indicates fitted values. We did not survey high-income barrios.

units, producing 13 new internal borders.[25]

This comuna subdivision changed the jurisdiction of the state's security and justice apparatus. Residents can access some municipal services, like education and health, irrespective of their address (i.e., anyone can cross a comuna border to attend their nearest school or clinic). But the police and several municipal agencies—especially dispute resolution and family services officers—are mandated to patrol and address problems within a comuna, and so their jurisdiction and outreach ended at the new borders.

At the time the new borders were introduced, the city had 15 full-service police stations and 37 municipal security and justice agencies that provide dispute resolution and family services. They increased this to 39 shortly as a part of the border reform. These 54 historical headquarters represent the initial allocation of state services, and we focus on these for our main results. Figure 7 displays a map of Medellín with the original and new internal borders, as well as historical police stations and municipal headquarters.[26]

Starting around 2000, the city began expanding the number of headquarters. For police, they added no new full-service stations, but they did construct a large number of small

[25]These new comunas were created in the Bill 54 by the city council. The previous organization of the city dated back from more than 20 years before (city Bill 52 from 1963).

[26]We used phone books, satellite images, visits to the city and police archives, and visits to city and police infrastructure to document locations, openings, and closures. Municipal headquarters include *inspecciones*, *comisarías* and *Casas de Justicia*. All comunas had at least one police station or municipal agency in 1995, and so we can construct a Δd_{ij} measure based on this historical presence.

satellite offices, called *Comandos de Atención Inmediata*—CAIs, which mainly served as a stopping point for patrols, with few citizen services. They also constructed 8 more municipal security and justice headquarters. We exclude these from our main analysis, but we show in an appendix that the results are robust to including them, and that their inclusion does not affect causal identification.

5.2 Calculating block pairs and the treatment variable

Typically, for blocks close to the new border, the effect was to shift blocks on one side further from their nearest headquarters—a "shock" to state distance. Our shock, or treatment variable, is Δd_{ij}, the difference in distance to state security headquarters between two nearby blocks, i and j, that results from the introduction of a new border. (Note that we order the blocks in each pair so that i is the block furthest from its state headquarters. Thus Δd_{ij} rises as the border assigns i further from its original state protection services.)

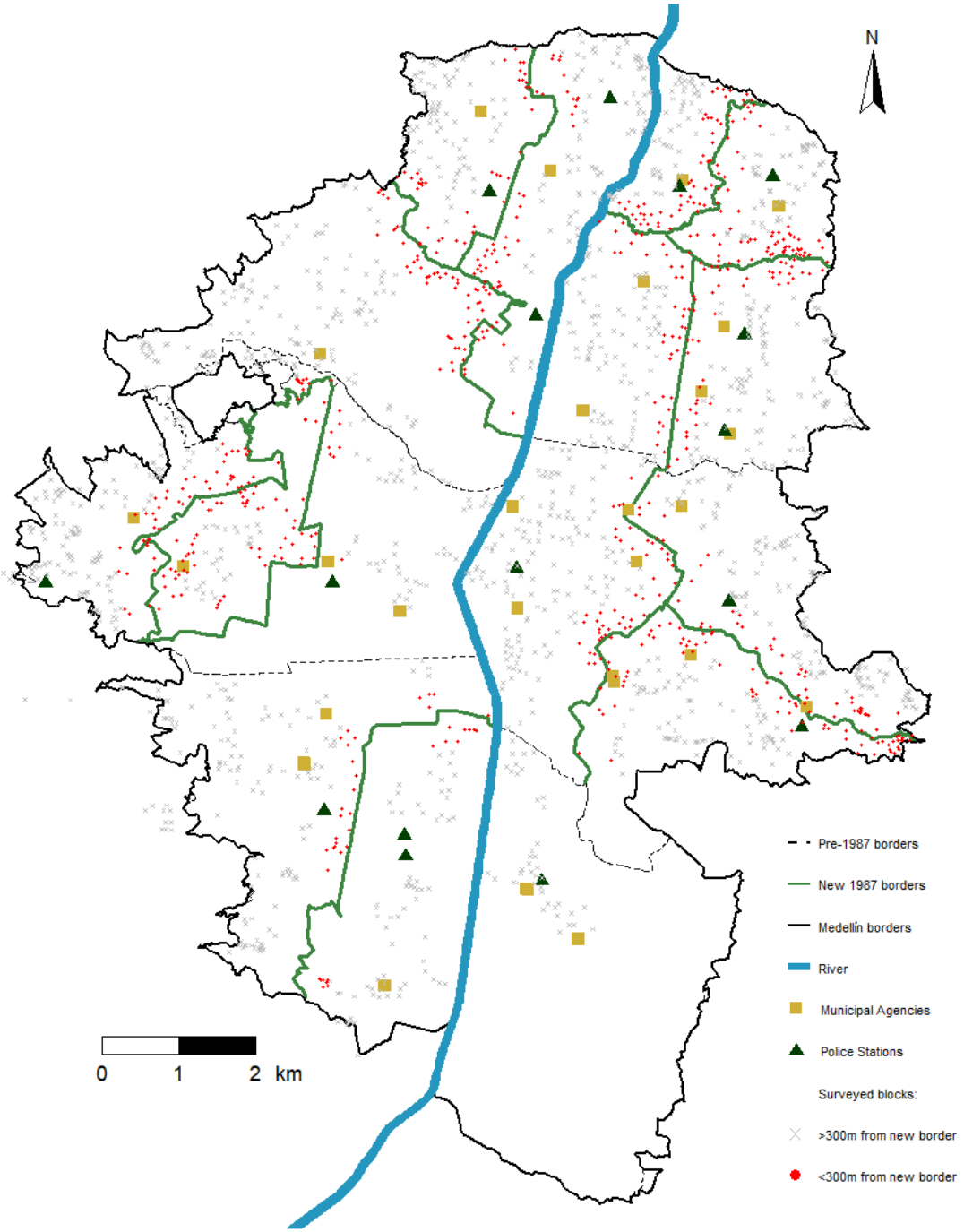
To create these nearest-neighbor block pairs, we use our representative sample of 2,066 city blocks from 2019. For each surveyed block, we calculated a matrix of distances to every other surveyed block. We then identified each block's nearest comuna border. To find a block to pair it with, we searched among the surveyed blocks in the neighboring comuna, limiting the search to ones that also have that border as their closest. From this set, we chose the closest cross-border block to make a pair. For each block, we then calculated the average distance to the comuna's historical police and municipal security headquarters.

Our analysis limits the sample to pairs where both blocks are within 300 meters of a new border, excluding borders that run along natural boundaries (such as impassable mountain ridges). We tested several pairing algorithms, as well as 200- and 100-meter bandwidths, and we will show that results are generally robust to different approaches.

Figure 7 highlights these eligible blocks in red, and Figure 8 displays the distribution of Δd_{ij} in the pairs within 300 meters of the new borders. The 10th percentile is 40 meters, the median is 402 meters, and the 90th percentile is 1,129 meters. The standard deviation is about 428 meters (similar to the median change).

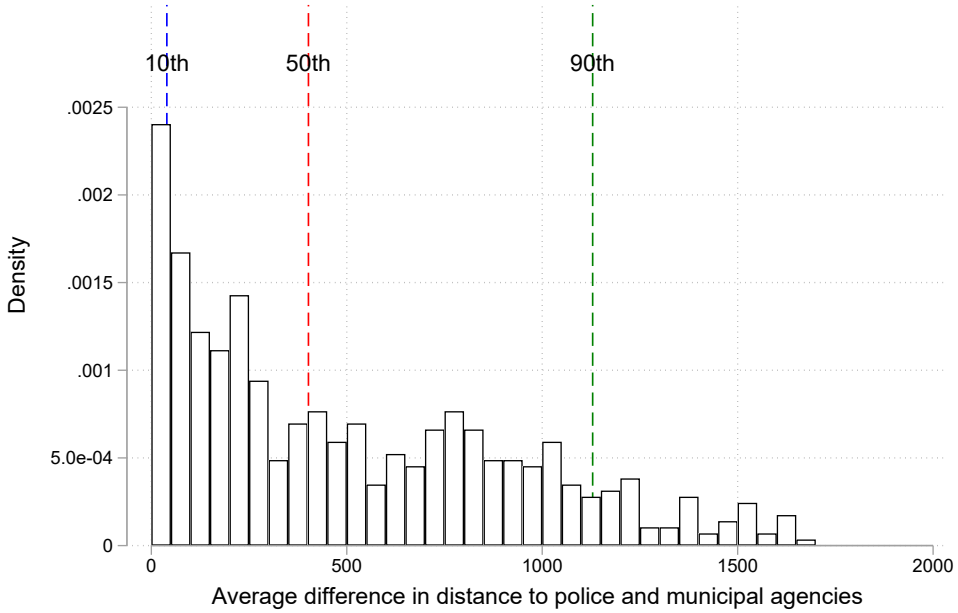
The reform produced variation in Δd_{ij} because of the irregular boundary shapes and the idiosyncratic position of state headquarters on either side. Consider another stylized illustration, in Figure 9. On the far left, the new border assigns block 1 to a more distant headquarters because its pair (block 2) is closer to its comuna headquarters. On the far right, the same border assigns a block on the opposite side (block 6) to a more distant headquarters. The distance shocks $\Delta d_{1,2}$ and $\Delta d_{6,5}$ are similar in magnitude, but the border

Figure 7: Surveyed blocks, historical state security headquarters, and pre- and post-reform comuna borders



Notes: Gold squares represent municipal dispute-resolution providers; green triangles are police stations. The dotted lines represent the old comuna borders, the solid green line represents the new borders, and the black line indicates Medellín's municipal boundaries. The blue center line shows Medellín's river.

Figure 8: Distribution of differences in the distance to municipal agencies between paired blocks within 300 meters of the new comuna borders



Notes: Vertical lines represent the 10th, 50th and 90th percentiles.

does not assign blocks on one side consistently to treatment or control.[27] The magnitudes of the shock also vary along the border. The distance shock experienced by the block pair in the middle is much smaller than the shocks to the block pairs on the far left and right.

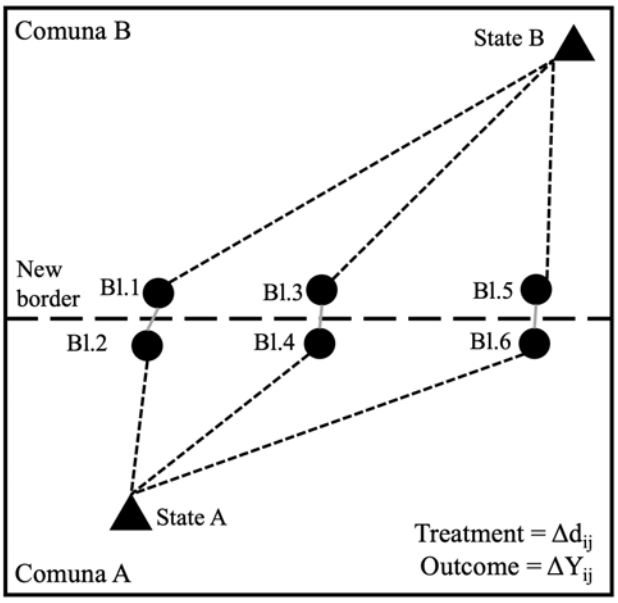
5.3 Estimation and identification

To identify the effect of Δd_{ij} on difference in long run block outcomes, the key assumption is that the only variable that changes discontinuously at the border is proximity to comuna-based state services, proxied by d (Keele and Titiunik, 2015). This would be violated if there were some other systematic difference between the paired blocks, ΔX_{ij}, that is correlated with both the difference in current combo governance as well as Δd_{ij}.

This is the advantage of the new administrative borders. Consider a potential confounder X, such as the distance to local business agglomerations. In the city-wide sample of blocks, we might expect businesses to locate themselves close to state protection services (or the state to locate close to them). Combos might also choose to offer combo governance near commercial centers. If so, this would confound a cross-sectional regression of combo gov-

[27] If it did, the border natural experiment would be akin to a clustered randomized trial with 13 clusters. As it stands, this is not a clustered experiment, although we show robustness to this simple clustering approach below.

Figure 9: Stylized illustration of variation in both treatment intensity and which side of the border is treated



Notes: Variation in treatment intensity across paired street blocks, as produced by the introduction of one new comuna border.

ernance on distance to the state. Once we look at cross-border differences, however, this confounding should dissipate. Paired blocks should be similarly close to business agglomerations, in part because the blocks are close to one another, but also because they can access and benefit from the economic activity regardless of the border, because the border only shapes access to state protection. That is, ΔX_{ij} should be small and generally uncorrelated with Δd_{ij}.

Table 3 tests these assumptions for observed covariates. We have data on the distance to other state services, such as schools and health centers, as well as distance to major business thoroughfares. We also have data on a range of block-level geographic and demographic characteristics from the 1993 census.[28]

Column (1) reports means and standard deviations for all blocks within 300 meters of a new border. Column (2) reports the main identification test—whether differences in paired blocks correlate with Δd_{ij}. Each estimate comes from a separate regression of the covariate on border fixed effects and each block's distance to a common fixed point on the border, as specified in Equation 7 below. We observe a handful of significant differences: blocks farther away from the state are located at higher altitudes, have a slightly lower probability of being present by 1948, have a higher probability of being present by 1985, and are closer to schools.

[28]Earlier rounds were not available, and though these data were collected slightly after treatment began, we nonetheless expect them to change negligibly at the border.

Table 3: Balance test: How block pair differences vary with Δd_{ij}

		Correlation between Δx_{ij} and Δd_{ij}	As % of sample mean
	Subsample Mean (SD)	Estimate (SE)	Estimate (SE)
	(1)	(2)	(3)
Block average elevation (meters)	1,608.11	1.536***	0.001
	(115.39)	(0.586)	
Block present in 1948	0.17	-0.009*	-0.054
	(0.37)	(0.005)	
Block present in 1970	0.66	0.000	0.000
	(0.47)	(0.011)	
Block present in 1985	0.84	0.036***	0.043
	(0.37)	(0.008)	
Block average slope	89.94	0.011	0.000
	(0.76)	(0.015)	
Log of total population (1993)	4.74	-0.013	-0.003
	(1.04)	(0.023)	
Median age (1993)	25.88	-0.095	-0.004
	(6.26)	(0.118)	
Share completed primary ed. (1993)	0.82	0.002	0.003
	(0.10)	(0.002)	
Share completed secondary ed. (1993)	0.46	0.001	0.003
	(0.20)	(0.003)	
Share completed higher ed. (1993)	0.10	-0.003	-0.026
	(0.11)	(0.002)	
Meters to schools	169.13	3.985**	0.024
	(93.54)	(2.013)	
Meters to health centers	484.14	-5.568	-0.012
	(322.49)	(3.489)	
Meters to business centers	326.91	1.075	0.003
	(55.18)	(1.277)	

Notes: Column (1) reports summary statistics. Column (2) tests whether differences in paired blocks are correlated with state proximity using Equation 7. Column (3) reports the magnitude of the estimate as a percentage of the sample mean. Standard errors after bootstrapping are nearly identical to our main specification.

Importantly, however, all these differences (however precise) are small in magnitude, as we report in Column (3). Typically they are less than 1–4 percent of the sample mean. This suggests the degree of confounding is limited. Nonetheless, we will want to control for these covariates as a precaution.

Estimation To estimate local treatment effects at the border, we run the following ordinary least squares regression:

$$\Delta Y_{ijb} = \alpha_b + \beta \Delta d_{ij} + \theta \Delta X_{ij} + \lambda B_{ij} + \varepsilon_{ijb} \tag{7}$$

Here, α_b is a vector of border fixed effects, one for each of the new borders. ΔX_{ij} is a vector of the control variables listed in Table 3. Finally, B_{ij} is a vector of the distance from each block to a common fixed point on the border, akin to a distance running variable, to account for differences in distances to the border within a pair.

This approach follows Keele and Titiunik (2015) on geographic regression discontinuities. Within a narrow bandwidth of the new border, it treats the distance shock as a random variable conditional on covariates and the distance running variables. Unlike many such quasi-experiments, however, treatment is not consistently on only one side of the border, but rather changes depending on the distance of each block to its corresponding comuna headquarters, relative to its matched pair.

Outcomes When the outcome variable, ΔY_{ij}, is the difference in current state service provision and legitimacy, we expect that $\beta_{state} < 0$. That is, as Δd_{ij} grows large, the block assigned to a more distant state headquarters reports lower state responsiveness. As all blocks receive some degree of state services, our treatment effects estimate the intensive margin of state presence.

We are principally interested in the effects on combo governance, however, where the sign of β_{combo} is ambiguous. As we discuss in our descriptive analysis, most blocks did not have significant levels of gang rule in the 1980s or 1990s, and so the treatment could have affected gang rule on both the intensive and extensive margins.

Within-combo estimates Note that for the most part this strategy will estimate *within-combo* effects on gang rule. Combo borders generally do not coincide with comuna borders, and oftentimes the paired blocks will be under the same combo. Since we do not have precise borders for most combos, we cannot say how often this is the case. Nonetheless, the estimates should generally reflect how combos respond within their territories to different levels of state penetration.

Economic development and sorting In Section 4, we noted how state governance could produce combo governance not only through a strategic response, but because state services produce economic and demographic changes that increase the demand for combo governance. For example, over three decades, proximity to police, dispute resolution, and other local services could increase prosperity and transactions, and with it demand for combo governance. In addition, households and businesses who demand governance may have moved to better-governed neighborhoods.

This is not an identification concern for β when we look at treatment effects on gang governance, but rather a question of mechanism and interpretation—to what degree is β attributable to the state's effects on growth versus the combo's strategic response. In all likelihood, the answer is "both," and our research design does not permit us to judge which is more influential. We can look at treatment effects on growth and demographics directly, however, and subjectively assess the plausibility of this channel.

Other identification concerns We address other concerns, including measurement error and placebo tests, after examining results.

6 Results

6.1 Local average treatment effects

Table 4 reports how increasing a block's distance from security headquarters affects residents' reports of state and combo governance and legitimacy in 2019. We scale the treatment variable so that the estimates reflect the effects of being 100 meters more distant from historical state headquarters. Column (1) reports mean state and combo governance in blocks <300 meters from a new border, and Column (2) presents our main specification.

For every 100 meters more distant, survey reports of state governance falls by 0.012, significant at the 5% level. Recall that the median difference in distance is roughly 400 meters. Compared to the control mean of 0.41, this implies the median change is associated with a 11% decline in citizen reports of state governance services. State legitimacy also declines somewhat as blocks get exogenously farther from the state, by 0.005, not statistically significant. (Recall this is a composite index of trust in the state, perceived fairness, and satisfaction.)

This is an important finding in itself, because it shows how proximity is important to projecting state power. It implies that the effects of police and municipal agencies are highly localized, even if they have a mandate to serve a wider area.

Table 4: Impacts on governance and legitimacy of being assigned to be 100 meters more distant from the state, using historical headquarters

Effect on ΔY	Subsample Mean (SD)	Correlation with Δd_{ij} Estimate (SE)	Median change as % of sample mean Estimate (SE)
	(1)	(2)	(3)
State			
Δ State Governance Index (0-1)	0.41	-0.012**	-11.294
	(0.19)	(0.005)	
Δ State Legitimacy Index (0-1)	0.58	-0.005	-3.227
	(0.14)	(0.005)	
Combo			
Δ Combo Governance Index (0-1)	0.32	-0.014**	-18.035
	(0.22)	(0.006)	
Δ Combo Legitimacy Index (0-1)	0.42	-0.010	-9.972
	(0.21)	(0.007)	
N for Governance outcomes		571	
N for Legitimacy outcomes		426	
Median $\Delta DistState$ (in 100m)		4.019	

Notes: Each estimate comes from a separate regression. Only residents (not business respondents) were asked about legitimacy, and some blocks have only residents, hence the lower sample size.

Combo governance co-moves with reports of state governance. The results in Column (2) suggest that, for every 100 meters further from current security headquarters, combo governance falls by 0.014, significant at the 5% level. At the median change in distance, this represents an 18% decline in gang governance services. Combo legitimacy also falls somewhat as the state grows more distant—by 0.010, a roughly 10% decrease for to the median change in distance—though the estimate is not statistically significant.

In Appendix Table C.2 we estimate the effects of police and municipal agencies separately. Broadly, we see negative point estimates for municipal offices and police stations for both state and combo governance. Combo governance seems to be slightly more responsive to the proximity of police.

Robustness to alternative estimation strategies Table 5 illustrates that results are highly robust to alternative estimation approaches, including: (i) using current headquarters instead of historical headquarters (which still produces causally-identified results, as described in Appendix B); (ii) dropping the two municipal headquarters constructed as part of the border reform; (iii) using 200 meter and 100 meter bandwidths around the borders; (iv) clustering standard errors at each of the 13 new borders; (v) using a machine-learning based algorithm to choose control variables; (vi) dropping all control variables; (vii) using latitude and longitude as an alternative to our running variable (the distance to a common fixed point on the border); and (viii) to (x) to different matching algorithms.[29]

In all cases the point estimates are generally similar in terms of both magnitude and precision. Using a 100 meter bandwidth generally increases the magnitude of results, though the smaller sample size produces less precise estimates. Using current headquarters produces somewhat more precise estimates for the legitimacy outcomes, in part because these incorporate the fact that some blocks have been exposed to new police and municipal headquarters since the early 2000s.

Alternative outcomes Finally, these results are consistent with related survey outcomes, reported in Table 6. When the discontinuity causes the state to be more distant, respondents reported that they found the state and combo were both 16–20% slower to respond, and about 9–11% more difficult to contact (not statistically significant). As state distance grows they were also 43% more likely to report that the combo's security fees are too high. This is consistent with combos being more extortionate in areas far from the state. They provide

[29]For instance: (i) an elevation-adjusted matching including elevation when computing distances to the comuna borders; (ii) a "relaxed" matching algorithm where we allow blocks to match with blocks for which the comuna with the closest border might not be the comuna of the original block; and (iii) an "unrestricted" matching algorithm where we search for the closest block in any other comuna.

Table 5: Robustness of impacts on state and combo rule of being 100 meters more distant from the local state

	Median Δd_{ij}	Δ State Governance	Δ State Legitimacy	Δ Combo Governance	Δ Combo Legitimacy
	Estimate (SE)	Estimate (SE)	Estimate (SE)	Estimate (SE)	Estimate (SE)
	(1)	(2)	(3)	(4)	(5)
Main specification	4.019	-0.012**	-0.005	-0.014**	-0.010
		(0.005)	(0.005)	(0.006)	(0.007)
Calculate Δd using current HQ	3.553	-0.010**	-0.008*	-0.013**	-0.013*
		(0.005)	(0.005)	(0.006)	(0.007)
Calculate Δd dropping 2 HQ constructed w. reform	3.858	-0.009*	-0.004	-0.010*	-0.007
		(0.005)	(0.004)	(0.006)	(0.007)
200m bandwith to border	3.918	-0.014**	-0.005	-0.011	-0.010
		(0.006)	(0.006)	(0.007)	(0.008)
100m bandwith to border	4.129	-0.011	0.002	-0.020*	-0.020
		(0.010)	(0.009)	(0.011)	(0.012)
Use all comuna borders	4.039	-0.011**	0.001	-0.014***	-0.011*
		(0.005)	(0.005)	(0.005)	(0.006)
Cluster std. err. on the 13 new borders	4.019	-0.012	-0.005	-0.014**	-0.010
		(0.008)	(0.005)	(0.005)	(0.009)
Choose controls with machine learning	4.019	-0.008	-0.003	-0.015**	-0.006
		(0.005)	(0.005)	(0.006)	(0.007)
Drop all control variables	4.019	-0.017***	-0.003	-0.019***	-0.010
		(0.005)	(0.005)	(0.006)	(0.006)
Use coordinates for running variable	4.019	-0.007	-0.007	-0.011*	-0.018**
		(0.005)	(0.005)	(0.006)	(0.007)
Elevation adjusted matching	4.039	-0.012**	-0.005	-0.015**	-0.011
		(0.005)	(0.005)	(0.006)	(0.007)
Relaxed matching	4.062	-0.012**	-0.006	-0.015***	-0.010
		(0.005)	(0.005)	(0.006)	(0.007)
Unconditional matching	4.068	-0.015***	-0.001	-0.013**	-0.008
		(0.005)	(0.004)	(0.005)	(0.006)

Notes: This table compares alternative specifications, changing one feature of the model at a time. Each row is a different estimation of treatment effects.

fewer services but continue to extract similar fees for the service, because they do not worry about citizen loyalty as much.

6.2 Heterogeneity by illicit rents

Our interviews and model suggest that this crowding-in effect in response to state rule should be strongest where gangs' rents from criminal activities other than extortion (like drug retailing) are greatest. Suppose we have a proxy for potential rents from such activities, π_{ij}. In that case, we expect state and combo governance to co-move (more) in high-π areas. We estimate heterogeneity by interacting treatment with an indicator for low-rent areas, π_{ij}^{low}:

$$\Delta Y_{ijb} = \alpha_b + \beta \Delta d_{ij} + \gamma \pi_{ij}^{low} + \delta \Delta d_{ij} \times \pi_{ij}^{low} + \theta \Delta X_{ij} + \lambda B_{ij} + \varepsilon_{ijb} \tag{8}$$

Data on combos' illicit rents is understandably scarce. As a proxy for Π, we obtained police data on the estimated value of drug seizures from 2014–19, geolocated to the seizure location. We take a 400-meter radius around each paired block, calculate the total drug seizure value for each block, and take the pair averages.[30] Since blocks share the same value, the heterogeneity analysis tells us how combos respond to a border difference in state proximity depending on the broader value of the territory.

For simplicity, we coarsen this seizure measure into a high/low indicator, based on whether the pair of blocks are located in an area above or below the median level of drug seizures. This aids interpretation. The coefficient on Δd_{ij} estimates the treatment effect in high drug value neighborhoods, and the coefficient on the interaction will give us the difference in the treatment effect between high- and low-value areas.

Naturally, one limitation of the drug-seizure data is that they are post-treatment, and so may be endogenous to state enforcement and combo governance. This is a limitation of the available data. To the degree that state governance is not a first order determinant of illicit rents, however, our coarsened measure could still accurately order the exogenous rent potential of the area. This is because the majority of seizures by local police patrols are small, and should not affect the classification into high/low seizures. Rather, high values are driven by large seizures of hard drugs, typically by national security organizations or special police units different than neighborhood patrols. Thus, they should not be affected by the location of local headquarters or relative distance to them. This should limit the endogeneity of our indicator. Nonetheless, these results should be taken with some caution.[31]

[30] 2014 is the first date that seizures are geolocated.

[31] Moreover, even if current rent potential accurately reflects baseline rent potential, drug seizure data suffers from measurement error. Most likely, combos in high-value markets may be more powerful, and would have higher incentives to bribe the police to reduce seizures and reports. This would lead to higher

Table 6: Impacts on other governance outcomes of being assigned to be 100 meters more distant from the state, using historical headquarters

Block-pair difference	Subsample Mean (SD)	Correlation with Δd_{ij} Estimate (SE)	Median change as % of sample mean Estimate (SE)
	(1)	(2)	(3)
Panel A: Efficacy outcomes			
How easy is it to contact the state	0.56	-0.009	-11.294
	(0.26)	(0.006)	
How fast is the state	0.52	-0.016**	-16.340
	(0.29)	(0.006)	
How easy is it to contact the combo	0.46	-0.013	-9.101
	(0.18)	(0.008)	
How fast is the combo	0.38	-0.026***	-20.391
	(0.21)	(0.010)	
Panel B: Payments to combos and state			
Payment rate of security fee	0.13	-0.001	-3.996
	(0.28)	(0.008)	
Says neighbors pay security fees	0.33	-0.001	-0.845
	(0.36)	(0.009)	
Approves of combo security fee	0.06	0.001	6.013
	(0.16)	(0.004)	
Says security fees are too high	0.28	-0.030**	-43.030
	(0.38)	(0.013)	
Percentage of bussines paying taxes	0.51	-0.003	-2.639
	(0.49)	(0.017)	
Approves of city's local taxes	0.60	0.004	2.639
	(0.33)	(0.009)	
Says local taxes are too high	0.62	0.006	3.580
	(0.34)	(0.010)	
Number of observations		574	
Minimum N		332	

Notes: Each estimate comes from a separate regression.

Table 7: Heterogeneous impacts on combo governance of being 100 meters more distant from the historical state, by local drug seizure value

	$\Delta Combo$ Governance
	Estimate (SE)
	(1)
Correlation with Δd_{ij}	-0.024***
	(0.007)
Correlation with Δd_{ij} X Low drug seizures	0.012*
	(0.006)
Low drug seizures	0.004
	(0.045)
Traditional and fixed point running var	✓
Border FE	✓
N	564
p-val for ATE in Low-drug area	0.062
Avg. combo gov in low-drug area	0.278
Avg. combo gov in high-drug area	0.356

Notes: Drug seizures are geographically matched to each block using a 400 meter buffers.

Table 7 estimates Equation 8 for combo governance.[32] The results suggest that treatment effects are indeed concentrated in high drug-seizure areas. The local treatment effect of the border is about twice as great in this half of the sample, and the difference between the two sub-samples is statistically significant at conventional levels. These above-median neighborhoods plausibly correspond to places where combos are most concerned with protecting drug profits from police incursions, while in low-seizure areas, combo governance does not respond as much to variation in distance to the state. Of course, combos still sell drugs and collect other illicit rents in below-median neighborhoods, and so they still experience statistically significant increases in combo governance in response to state presence.

under-reporting of drug seizures in high-value areas. Since large drug market areas are publicly known, and our data coincides with this conventional wisdom, we believe measurement error may affect the margin around the median but not the broad ordering of block pairs. Furthermore, measurement error would likely induce a change over a small part of the distribution (switching high- for low-value areas), that would tend to attenuate our estimates for high-value areas and exaggerate the estimates for low-value areas. Nonetheless, this is another reason for caution when interpreting our results.

[32]See Appendix Table C.5 for other outcomes.

6.3 Economic development and sorting as a channel of impact

As noted above, another possible channel for the crowding-in effects we detect is if state presence encourages economic development and migration to more state-intensive blocks, which raises overall demand for governance, some of which combos fill. Table 8 estimates the effects of state proximity to the state on a range of economic and demographic measures from administrative data, the 2018 census, and our 2019 survey. There is almost no evidence that economic activity or migration consistently falls or rises with proximity to the state. Most measures of poverty and business activity do not vary systematically with distance to the security services. Nor do measures of recent in-migration or population stock. There is some indication of lower human capital an lower soil value as the state grows more distant, which could indicate a minor sorting effect. These are relatively small, and so (combined with the absence of impacts on business activity) probably do not account for the large combo governance treatment effects we observe.

This is consistent with our qualitative interviews, which suggests that most people have deep social ties to their neighborhood, and choose location for a variety of reasons, and so the proximity of police and dispute resolution services is a second-order concern in their location decisions.

6.4 Sensitivity to potential identification threats

Evidence of endogenous state protection? One potential concern with border discontinuities, is that police could provide fewer patrolling services nearby the border of a jurisdiction because of the expectation of positive externalities from other police stations. That is, police agents would under-patrol peripheries if they believe that patrols at the other side of the border deter crime. We examine whether this is the case by estimating a street segment level regression of police patrolling time on distance to the border, distance to the state and the interaction.[33] The results are reported in Appendix Table C.4. Broadly, we do not find evidence of endogenous state protection nearby the borders.

Potential for other unobserved confounders Are there other unobserved block characteristics that are associated both with differences in proximity to historical state presence and to motives for combo governance? Our border discontinuity should reduce the likelihood of these confounders. In addition, they would need to have a stronger relationship with both combo and state governance than our observed confounders (such as the availability of

[33]Data on patrolling time per street segment were collected in 2015 for a hot spots policing experiment in the city Collazos et al. (2021).

Table 8: Impacts on neighborhood prosperity and demographics of being 100 meters more distant from the local state, comuna border discontinuity. current borders

ΔY	Subsample Mean (SD)	Correlation with Δd_{ij} Estimate (SE)	Median change as % of sample mean Estimate (SE)
	(1)	(2)	(3)
Income strata (2018)	2.56	0.007	1.069
	(0.97)	(0.012)	
Multidimentional Poverty Index (2018)	13.68	0.019	0.554
	(14.30)	(0.323)	
Log of mean profits (2019)	13.34	-0.007	-0.223
	(1.17)	(0.049)	
Log of mean sales (2019)	14.80	0.070	1.894
	(1.18)	(0.047)	
Number of employees (2019)	2.32	0.106	18.410
	(2.14)	(0.091)	
Unemployment rate (2018)	0.11	-0.002	-8.046
	(0.07)	(0.002)	
Schooling rate (2018)	0.91	-0.008***	-3.716
	(0.12)	(0.003)	
Log of total population (2018)	5.60	-0.004	-0.275
	(1.35)	(0.038)	
Absence of firms	0.23	0.001	1.305
	(0.42)	(0.010)	
Logarithm of total number of firms	1.52	-0.002	-0.494
	(1.18)	(0.023)	
Logarithm of average geoeconomic value of soil (2014)	13.34	-0.017**	-0.527
	(0.63)	(0.007)	
Logarithm of average housing value	11.84	-0.008	-0.275
	(0.63)	(0.014)	
Percent of women (2018)	52.54	-0.019	-0.145
	(4.47)	(0.126)	
Share with no ed. completed (2018)	2.00	0.013	2.615
	(1.70)	(0.045)	
Share completed primary ed. (2018)	79.98	-0.549***	-2.760
	(7.47)	(0.146)	
Share completed secondary ed. (2018)	63.70	-0.876***	-5.526
	(12.32)	(0.199)	
Share completed higher ed. (2018)	25.05	-0.154	-2.473
	(15.60)	(0.227)	
Percent of population aged 0 to 14 (2018)	16.04	0.084	2.105
	(6.01)	(0.116)	
Percent of population who recently migrated (2018)	4.44	-0.116	-10.548
	(3.65)	(0.085)	
Number of observations		574	
Minimum N		246	

Notes: Each estimate comes from a separate regression.

other state services, or the distance to business agglomerations). This is possible. For example, some borders might not have been arbitrarily drawn—although our anecdotal evidence on the process suggests otherwise. We address this by conducting a placebo exercise. We randomly matched 1,500 times our baseline sample of blocks located within 300 meters of comuna borders with other blocks inside the same comuna, ensuring that the matched blocks are at most 600 meters away from each other—so that we resemble our baseline specification. To build our treatment variable, we assign one of the block pairs to state services within the comuna and the other to services in the neighboring comuna (depending on the distance to average services for each). The distribution of treatment effects for combo governance and legitimacy is reported in Figure C.1. Our observed treatment effects lie at the edge of the distribution, suggesting they are unlikely to be explained by other confounders.

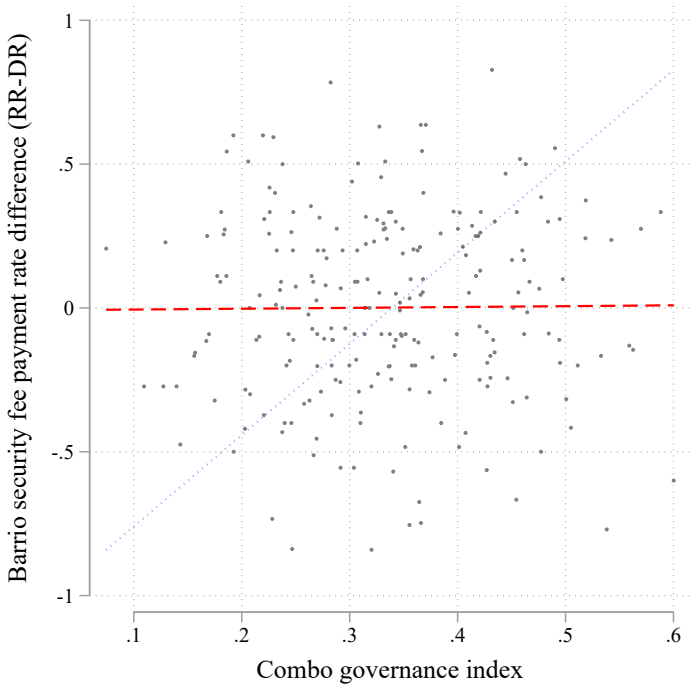
Measurement error correlated with treatment Finally, we consider different kinds of measurement error correlated with treatment, and judge that these are unlikely to account for the large crowding in effects we observe. First, it's important to ask what specific forms of measurement error would bias the estimate on combo governance upwards, if any. If people tend to understate gang rule in general, we will tend to underestimate crowding in. We show this in Appendix D with some simple models. In only one specific scenario will we overestimate crowding out: if citizens under-report gang governance when gang governance is high and the state is far away, but there is no misreporting when the government is close.

First note that the correlates of combo governance reported in Appendix Table C.1 suggest that people are comfortable talking about the combo. In particular, we see a positive correlation between the density of combos and combo governance, and negative correlations between distance from gang and razón central locations and combo governance. This accords with our qualitative experience: when interviewed in private, people speak freely about combo rule.

Second, we tested this using a survey experiment. We took one of the variables we thought could be the most sensitive—whether people paid "taxes" to the combo in the form of security fee and extortion payments. In our city-wide survey of thousands of residents, we randomly assigned respondents to either a direct question on whether they paid the combo, versus a randomized-response technique, where they privately flipped a coin and responded to the question honestly or not depending on the flip. In other contexts, this method has detected under-reporting of sensitive behaviors.[34] With the whole sample, randomized

[34]Others were asked the same question using a List Experiment, where half are asked to give the number of four nonsensitive actions they engaged in, and half see a list of five actions including paying extortion. In general, these list experiments are extremely noisy. Yet, the results are consistent with what we see in these randomized response and direct response questions. There is no evidence of systematic measurement error.

Figure 10: Survey experiment results: Difference between randomized response (RR) and direct response (DR) to security fee payment rate



Notes: Each dot represents a barrio average. The dotted line indicates fitted values. We did not survey high-income barrios.

response (RR) elicited an extortion rate of 22.6% from businesses and 6% from households, compared to 19.4% and 7.8% with directly responses (DR). The differences RR–DR run in opposite directions for households and businesses, and are not statistically significant.

Figure 10 plots this block averages for this RR–DR difference on block combo governance in the full city sample, and finds little systematic correlation. Table 9 reports summary statistics within our sample of blocks close to new borders, and examines correlates between differential reporting and our treatment variable. On average, the direct responses on vacunas are 6 percentage points lower than the randomized response (RR), as seen in Column (1). So there is evidence of slight under-reporting in this subsample, even if it is not statistically significant. There is no statistically significant correlation, however, between the RR–DR pair-block difference and difference in distance to state headquarters. As the state gets further away, direct reporting falls somewhat relative to randomized response. But this is small, imprecise, and unable to account for the large and highly significant crowding in estimated above.

We also investigate whether patterns of non-response are correlated with treatment. For

Table 9: Test of systematic measurement error: Coefficient of closeness to state services for blocks along the inner comuna borders of Medellín on measurement error proxies

		Correlation (ΔY) with $\Delta DistState$
	Subsample Mean (SD)	Estimate (SE)
	(1)	(2)
Extortion payment rate difference (RR-DR)	-0.08	-0.002
	(0.82)	(0.038)
Proportion of questions answered for state	0.90	0.001
	(0.13)	(0.003)
Proportion of questions answered for combo	0.85	0.011**
	(0.18)	(0.005)
Number of observations		574
Minimum N		251

Notes: This table examines the correlation between proxies for measurement error and being 100 meters more distant from the state, using the same estimation for our main treatment effects. The extortion rate difference computes the difference between randomized response and direct response to the question of whether the household pays extortion. The other measures capture non-response to sensitive items (the proportion of questions answered). We look at the proportion of questions answered for each index, and whether this is different for the state versus the combo. More questions answered for the state could indicate a reluctance to talk about or disclose combo activities.

instance, people might decide to skip combo governance questions if they are uncomfortable. As we show in Table 9, respondents answered about 86% of combo governance and legitimacy questions, compared to 90% for the state, again consistent with slight under-reporting of combo governance (which, as we noted, would act to understate crowding in). People are somewhat *more* likely to answer these questions the more distant they are from the state, however—the opposite of the direction we are worried about. If the proportion of questions indeed proxy for under-reporting, this pattern implies our findings are more likely a lower bound of the actual treatment effect.

7 Conclusions

Hundreds of major city governments operate in an uneasy duopoly of coercion, rule, and taxation with urban gangs. Many such governments, like Medellín's, are strong, in the sense that they levy taxes and provide extensive public and private goods. Many have even deliberately attempted to expand state capacity in slums and low-income neighborhoods. Yet organized crime and gang rule over civilians have proven enduring.

Our work suggests a few important insights. First, at least in Medellín, the market for protection was not the main reason gangs decided to rule. Rather, this was overshadowed by indirect motives to govern, ones that arose from other business lines, especially retail drug sales. The presence or absence of this indirect motive may help explain patterns of gang rule in other cities. Rio de Janeiro, for instance, has extensive retail drug markets and also expansive, militarized gang rule with little or no taxation of residents. San Salvador, on the other hand, has limited local drug markets and much more extractive gangs. Our theory and empirical analysis predicts that, *ceteris paribus*, efforts to expand state capacity would be more likely to lead to crowding in in Rio and crowding out in San Salvador.

Second, both our results and our theoretical discussion suggest that common policy interventions could backfire in the presence of these indirect motives to rule. For instance, popular responses to organized crime and extortion include police crackdowns, ease of anonymous denunciation, or facilitating collective action among merchants. Crackdowns and denunciations could actually increase incentives for the gang to govern and foster legitimacy, especially in the most valuable neighborhoods. And efforts to reduce extortion overlook the fact that many gangs would have an incentive to rule even if they were unable to collect fees at all. And since extortion is a modest percentage of business sales, merchants may have weak incentives to undertake costly, risky collective action, especially if they are receiving real protection in return.

Third, the results suggest that gang rule can be best weakened by going after a gang's

illicit revenues. For instance, lower profitability of drug markets (e.g., because of marijuana legalization) could reduce optimal gang size as well as incentives to govern. In defeating the American mafia, prosecutors attribute their success not simply to more aggressive investigation and sentencing, but to the slow erosion of the mob's main sources of revenues. In the late 20th century, loansharking, numbers games, and labor racketeering declined in response to the rise of widespread access to consumer credit, state lotteries, and the decline of unionization (Kroger, 2008). So long as there is high consumer demand for illicit drugs and loans, crackdowns and crowding out will not undermine the gang's main motives for existence and ruling.

But our results also point to a terrible policy trade-off for city governments: weakening gangs could make them more violent and coercive. This comes from our observation that gang abuses, including purely extractive extortion, are disciplined by the gang's need to protect drug rents by fostering community loyalty, collaboration, and respect. If cities legalized drug sales, treated addicts, or otherwise reduced retail demand for drugs, gangs would have fewer incentives to treat residents well, and maintain peace in their neighborhoods. Similarly, suppose a government or non-profit tried to de-legitimize gangs among residents and incentivize denouncement to authorities through social norms marketing campaigns or collective action. Even if effective, by eliminating the gang's incentives to win residents' loyalty this could inadvertently lead to more abusive and extortionate patterns of rule. In short, curbing criminal governance, which after all brings order to a significant share of the population, can have major unintended consequences, ones which city governments do not appear to be aware of.

Finally, our work highlights the need for further descriptive and theoretical work on gangs, and more policy evaluations. Organized crime is arguably the largest threat to national security and development in the century ahead, especially in the Western hemisphere. When gangs fight one another or the government, they provoke armed violence exceeding most civil wars (Lessing, 2017). As with both oligopolistic markets in industrial relations and nations in the international system, it is hard to predict when gangs will go to war and when they will strike ceasefires and pacts. Even if they were in agreement that gang peace is preferable to gang war, mayors and police chiefs lack policy best-practices for fostering such peace.

What's more, the problem is set to worsen and widen in the coming decades. Many organized criminal groups in Latin America emerged during and after wars, from demobilizing paramilitaries and rebels. Likewise, in Sicily, the first mafias emerged from the ranks of unemployed private security forces from the former feudal estates. Fighters in Iraq, Syria, Afghanistan, Myanmar, and other states could follow the same path in the coming decades.

In light of this, we see our methodological approach—combining qualitative interviews,

descriptive analysis, and a quasi-experiment—as proof-of-concept exercises for further research. While certainly not easy, we show that it is possible to develop systematic qualitative and quantitative data on criminal organizations and their governance practices. At the same time, whereas many crime and policing evaluations focus on individual-level outcomes, we show that it is possible to evaluate interventions rigorously focusing on relevant, gang-related outcomes.

References

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

Acemoglu, D. and J. A. Robinson (2006). *Economic Origins of Dictatorship and Democracy*. Cambridge University Press.

Albertus, M. and O. Kaplan (2013). Land Reform as a Counterinsurgency Policy: Evidence From Colombia. *Journal of Conflict Resolution 57*(2), 198–231.

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

Arias, E. D. (2006). The Dynamics of Criminal Governance: Networks and Social Order in Rio de Janeiro. *Journal of Latin American Studies 38*(2), 293–325.

Arjona, A. (2016). *Rebelocracy*. Cambridge University Press.

Bates, R., A. Greif, and S. Singh (2002). Organizing Violence. *Journal of Conflict Resolution 46*(5), 599–628.

Beath, A., F. Christia, and R. Enikolopov (2012). Winning Hearts and Minds Through Development? Evidence From A Field Experiment in Afghanistan. *Working paper*.

Berman, E., J. H. Felter, J. N. Shapiro, and E. Troland (2013). Modest, Secure, and Informed: Successful Development in Conflict Zones. *American Economic Review 103*(3), 512–17.

Berman, E. and A. M. Matanock (2015). The Empiricists' Insurgency. *Annual Review of Political Science 18*, 443–464.

Berman, E., J. N. Shapiro, and J. H. Felter (2011). Can Hearts and Minds be Bought? The Economics of Counterinsurgency in Iraq. *Journal of Political Economy 119*(4), 766–819.

Blattman, C., G. Duncan, B. Lessing, and S. Tobon (2021a). Gangs of Medellín: How Organized Crime is Organized. *Working paper*.

Blattman, C., G. Duncan, B. Lessing, and S. Tobon (2021b). The failed promise of the state: Experimental evidence on statebuilding in a context of organized crime. *Working paper*.

Bliss, C. J. and N. Stern (1982). *Palanpur: The Economy of an Indian village*. Oxford University Press.

Brown, Z. Y., E. Montero, C. Schmidt-Padilla, and M. M. Sviatschi (2021). Market structure and extortion: Evidence from 50,000 extortion payments. *Working Paper*.

Bueno de Mesquita, E. (2020). Territorial Conflict Over Endogenous Rents. *The Journal of Politics 82*(1), 162–181.

Carvalho, L. S. and R. R. Soares (2016). Living on The Edge: Youth Entry, Career and Exit in Drug-Selling Gangs. *Journal of Economic Behavior and Organization 121*, 77–98.

Castillo, J. C. and D. Kronick (2020). The Logic of Violence in Drug War. *American Political Science Review 114*(3), 874–887.

Castillo, J. C., D. Mejía, and P. Restrepo (2020). Scarcity Without Leviathan: The Violent Effects of Cocaine Supply Shortages in the Mexican Drug War. *Review of Economics and Statistics 102*(2), 269–286.

Collazos, D., E. García, D. Mejía, D. Ortega, and S. Tobón (2021). Hot Spots Policing in a High-Crime Environment: An Experimental Evaluation in Medellin. *Journal of Experimental Criminology 17*, 473–506.

Crost, B., J. H. Felter, and P. B. Johnston (2016). Conditional Cash Transfers, Civil Conflict and Insurgent Influence: Experimental Evidence from the Philippines. *Journal of Development Economics 118*, 171–182.

Dal Bó, E., P. Dal Bó, and R. Di Tella (2006). Plata o Plomo?: Bribe and Punishment in a Theory of Political Influence. *American Political Science Review*, 41–53.

De Feo, G. and G. D. De Luca (2017). Mafia in the Ballot Box. *American Economic Journal: Economic Policy 9*(3), 134–67.

Dell, M. (2015). Trafficking Networks and the Mexican Drug War. *American Economic Review 105*(6), 1738–79.

Dube, O., O. García-Ponce, and K. Thom (2016). From maize to haze: Agricultural shocks and the growth of the mexican drug sector. *Journal of the European Economic Association 14*(5), 1181–1224.

Ensminger, J. (1996). *Making a Market: The Institutional Transformation of an African Society*. Cambridge University Press.

Fiorentini, G. and S. Peltzman (1997). *The Economics of Organised Crime*. Cambridge University Press.

Gambetta, D. (1996). *The Sicilian Mafia: the Business of Private Protection*. Harvard University Press.

Giraldo, J., A. J. Rendón, and G. Duncan (2014). *Nuevas Modalidades de Captación de Rentas Ilegales en Medellín*. Universidad EAFIT.

Grossman, H. I. (1996). *Rival Kleptocrats: the Mafia Versus the State*, pp. 143–155. Cambridge University Press.

Henn, S. J. Complements or Substitutes? How Institutional Arrangements Bind Chiefs and the State in Africa. *Working Paper*.

Hirshleifer, J. (1989). Conflict and rent-seeking success functions: Ratio vs. difference models of relative success. *Public choice 63*(2), 101–112.

Kasfir, N. (2015). *Rebel Governance–Constructing a Field of Inquiry: Definitions, Scope, Patterns, Order, Causes*, Volume 2. Cambridge University Press.

Keele, L. J. and R. Titiunik (2015). Geographic Boundaries as Regression Discontinuities. *Political Analysis 23*(1), 127–155.

Khanna, G., C. Medina, A. Nyshadham, and J. A. Tamayo (2019). Formal Employment and Organized Crime: Regression Discontinuity Evidence from Colombia. *Working Paper*.

Konrad, K. A. and S. Skaperdas (2012). The Market for Protection and the Origin of the State. *Economic Theory 50*(2), 417–443.

Konrad, K. I. and S. Skaperdas (1998). Extortion. *Economica 65*(260), 461–477.

Kroger, J. (2008). *Convictions: A Prosecutor's Battles Against Mafia Killers, Drug Kingpins, and Enron Thieves*. Macmillan.

Leeson, P. T. (2007). An-arrgh-chy: The Law and Economics of Pirate Organization. *Journal of Political Economy 115*(6), 1049–1094.

Lessing, B. (2017). Counterproductive Punishment: How Prison Gangs Undermine State Authority. *Rationality and Society 29*(3), 257–297.

Lessing, B. (2020). Conceptualizing Criminal Governance. *Perspectives on Politics*, 1–20.

Lessing, B., D. Block, and E. Stecher (2019). Criminal Governance in Latin America: an Empirical Approximation. *Working paper*.

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

Levitt, S. D. and S. A. Venkatesh (2000). An Economic Analysis of a Drug-Selling Gang's Finances. *The Quarterly Journal of Economics 115*(3), 755–789.

Limodio, N. (2018). Terrorism Financing, Recruitment and Attacks: Evidence From a Natural Experiment in Pakistan. *Working Paper*.

Magaloni, B., E. Franco-Vivanco, and V. Melo (2020). Killing in the Slums: Social Order, Criminal Governance, and Police Violence in Rio de Janeiro. *American Political Science Review 114*(2), 552–572.

Martin, G. (2012). *Medellín Tragedia y Resurrección: Mafia, Ciudad y Estado, 1975-2012*. Planeta.

McGuire, M. C. and M. Olson (1996). The Economics of Autocracy and Majority Rule: the

Invisible Hand and the Use of Force. *Journal of Economic Literature 34*(1), 72–96.

Melnikov, N., C. Schmidt-Padilla, and M. M. Sviatschi (2021). Gangs, labor mobility, and development. *Working Paper*.

Murphy, T. E. and M. A. Rossi (2020). Following the Poppy Trail: Origins and Consequences of Mexican Drug Cartels. *Journal of Development Economics 143*, 102433.

North, D. C., J. J. Wallis, and B. R. Weingast (2009). *Violence and Social Orders: A Conceptual Framework for Interpreting Recorded Human History*. Cambridge University Press.

Olson, M. (1993). Dictatorship, Democracy, and Development. *American Political Science Review 87*(3), 567–576.

Reuter, P. (1983). *Disorganized Crime: Illegal Markets and the Mafia*. MIT Press.

Sánchez De La Sierra, R. (2020). On the Origins of the State: Stationary Bandits and Taxation in Eastern Congo. *Journal of Political Economy 128*(1), 000–000.

Sánchez De la Sierra, R. and K. Titeca. Corruption in hierarchies. *Working Paper*.

Sanchez-Jankowski, M. (1991). *Islands in the Street: Gangs and American Urban Society*. University of California Press.

Schelling, T. C. (1971). What is the Business of Organized Crime? *The American Scholar*, 643–652.

Skaperdas, S. (2001). The Political Economy of Organized Crime: Providing Protection When the State Does Not. *Economics of Governance 2*(3), 173–202.

Skaperdas, S. and C. Syropoulos (1996). *Gangs as Primitive States*, pp. 61–81. Cambridge University Press.

Skaperdas, S. and C. Syropoulos (1997). The distribution of income in the presence of appropriative activities. *Economica 64*(253), 101–1117.

Skarbek, D. (2011). Governance and Prison Gangs. *American Political Science Review 105*(4), 702–716.

Skarbek, D. (2014). *The Social Order of the Underworld: How Prison Gangs Govern the American Penal System*. Oxford University Press.

Sobrino, F. (2019). Mexican Cartel Wars: Fighting for the US Opioid Market. *Working paper*.

Sviatschi, M. M. (2018). Making a Narco: Childhood Exposure to Illegal Labor Markets and Criminal Life Paths. *Working paper*.

Tilly, C. (1985). War Making and State Making as Organized Crime. *Violence: A reader*, 35–60.

Tobón, S. (2020). Do Better Prisons Reduce Recidivism? Evidence from a Prison Construction Program. *The Review of Economics and Statistics*, 1–47.

Venkatesh, S. A. (2006). *Off the Books: The Underground Economy of the Urban Poor*. Harvard University Press.

Appendix

A Formal presentation and extensions of model

This section elaborates details of the model and several claims in Section 4.

A.1 Cournot competition

Setup In each neighborhood, a gang g and a state s compete to sell protection in quantities q_g and q_s. Each organization chooses q_i to maximize their respective pay-off, and each has constant marginal cost c_i. Products are differentiated, and the price of each one is given by the linear inverse demand function $p_i = a_i - \beta q_i - \gamma q_j$. Here, $\gamma \in (0, 1]$ since the services offered by both organizations are substitutes, and $\beta > 0$ for downward-sloping demand. The pay-off for each organization is $V_i = p_i q_i - c_i q_i$. For simplicity, we assume an interior solution.

Nash Equilibria We begin by deriving the best response function for each organization:

$$\begin{aligned} \max_{q_i} V_i &= (a_i - \beta q_i - \gamma q_j) q_i - c_i q_i \\ \frac{\partial V_i}{\partial q_i} &= a_i - 2\beta q_i - \gamma q_j - c_i = 0 \\ q_i^* &= \frac{a_i - c_i}{2\beta} - \frac{\gamma}{2\beta} q_j \end{aligned}$$

We obtain an identical best response function for the other organization analogously, and replacing values we obtain:

$$q_i^* = \frac{2\beta(a_i - c_i) - \gamma(a_j - c_j)}{(4\beta^2 - \gamma^2)}$$

and,

$$q_j^* = \frac{2\beta(a_j - c_j) - \gamma(a_i - c_i)}{(4\beta^2 - \gamma^2)}.$$

Comparative statics We are mainly interested in whether gang rule is crowded in or out when there is an exogenous increase in state governance: $\frac{\partial q_i^*}{\partial q_j}$. To obtain this comparative static, we begin by defining:

$$G(q_i, q_j) \equiv \frac{\partial V_i}{\partial q_i} = a_i - 2\beta q_i - \gamma q_j - c_i$$

which is a continuously differentiable function from $\mathbb{R}^2 \to \mathbb{R}$. At the optimum, we know:

$$G(q_i^*, q_j^*) = a_i - 2\beta_i q_i^* - \gamma q_j^* - c_i = 0.$$

Since $-2\beta \neq 0$, we can use the implicit function theorem to obtain our main comparative static:

$$\frac{\partial q_i^*}{\partial q_j} = -\frac{\partial G(q_i, q_j)/\partial q_j}{\partial G(q_i, q_j)/\partial q_i} = -\frac{\gamma}{2\beta}$$

Since the two services are not complements, this comparative static implies that increases in one duopolist's supply of protection will reduce the other's.

Cournot competition with benefits to governing

We now introduce a non-standard feature: externalities stemming from gang rule.

Setup As above, but now the payoff for each organization is $V_i = (a_i - \beta_i q_i - \gamma q_j)q_i - c_i q_i + \rho(q_i, q_j)\pi_i)$, where $\rho(q_i, q_j)\pi_i$ captures the externalities described in Section 4. For simplicity, we assume an interior solution.

Nash Equilibria We begin by deriving the best response function for each organization:

$$\begin{aligned}
\max_{q_i} V_i &= (a_i - \beta q_i - \gamma q_j)q_i - c_i q_i + \rho(q_i, q_j)\pi_i \\
\frac{\partial V_i}{\partial q_i} &= a_i - 2\beta q_i - \gamma q_j - c_i + \frac{\partial \rho(q_i, q_j)}{\partial q_i}\pi_i = 0 \\
q_i^* &= \frac{a_i - c_i + \frac{\partial \rho(q_i, q_j)}{\partial q_i}\pi_i}{2\beta} - \frac{\gamma}{2\beta} q_j
\end{aligned}$$

We obtain an identical best response function for the other organization analogously, and replacing values we obtain:

$$q_i^* = \frac{2\beta(a_i - c_i) - \gamma(a_j - c_j) + \left(2\beta\frac{\partial \rho(q_i, q_j)}{\partial q_i}\pi_i - \gamma\frac{\partial \rho(q_i, q_j)}{\partial q_j}\pi_j\right)}{(4\beta^2 - \gamma^2)}$$

with an identical function for q_j^*.

Comparative statics Again we are interested in whether gang rule is crowded in or out when there is an exogenous increase in state governance: $\frac{\partial q_i^*}{\partial q_j}$. To obtain this comparative

static, we begin by defining:

$$G(q_i, q_j) \equiv \frac{\partial V_i}{\partial q_i} = a_i - 2\beta q_i - \gamma q_j - c_i + \frac{\partial \rho(q_i, q_j)}{\partial q_i}\pi_i$$

which is a continuously differentiable function from $\mathbb{R}^2 \to \mathbb{R}$. At the optimum, we know:

$$G(q_i^*, q_j^*) = a_i - 2\beta q_i^* - \gamma q_j^* - c_i + \frac{\partial \rho(q_i^*, q_j^*)}{\partial q_i}\pi_i = 0$$

Additionally, we assume that $2\beta \neq \frac{\partial^2 \rho(q_i^*, q_j^*)}{\partial q_i \partial q_i}\pi_i$, thus:

$$\frac{\partial G(q_i^*, q_j^*)}{\partial q_i} = -2\beta + \frac{\partial^2 \rho(q_i^*, q_j^*)}{\partial q_i \partial q_i}\pi_i \neq 0$$

We can use the implicit function theorem to obtain our main comparative static:

$$\frac{\partial q_i^*}{\partial q_j} = -\frac{\partial G(q_i, q_j)/\partial q_j}{\partial G(q_i, q_j)/\partial q_i} = \frac{\lambda \pi_i - \gamma}{2\beta - \delta \pi_i}$$

where $\lambda = \frac{\partial^2 \rho(q_i, q_j)}{\partial q_i \partial q_j}$ represents the cross-partial derivative between gang and state governance, and $\delta = \frac{\partial^2 \rho(q_i, q_j)}{\partial q_i \partial q_i}$ reflects the rate of increasing or decreasing returns to governing. We discuss conditions for this comparative static to be positive in section 4.

A.2 Cournot competition with endogenous demand

We now consider the possibility that providing governance can produce economic growth, which in turn may produce greater demand for governance. This section incorporates this idea into the Cournot framework by "endogenizing demand". The next section considers a political economy model in which "stationary bandits" provide public goods in order to grow the economic pie that they will tax.

Setup As before, except we now generalize the functional form of demand such that products are differentiated so the price of each one is determined by $p_i = a_i(q_i, q_j) - \beta q_i - \gamma q_j$, where $\gamma \in (0, 1]$ as services provided by both organizations are substitutes, and $a_i(q_i, q_j)$ is twice continuously differentiable. The payoff function is $V_i = (a_i(q_i, q_j) - \beta q_i - \gamma q_j)q_i - c_i q_i + \rho(q_i, q_j)\pi_i$, where $\rho(q_i, q_j)\pi_i$ captures the externalities described in Section 4. Again, we assume an interior solution.

Nash Equilibria As above, we begin by deriving the best response function for each organization:

$$\begin{aligned}
\max_{q_i} V_i &= (a_i(q_i,q_j) - \beta q_i - \gamma q_j)q_i - c_i q_i + \rho(q_i,q_j)\pi_i \\
\frac{\partial V_i}{\partial q_i} &= a_i(q_i,q_j) + \frac{\partial a_i(q_i,q_j)}{\partial q_i}q_i - 2\beta q_i - \gamma q_j - c_i + \frac{\partial \rho(q_i,q_j)}{\partial q_i}\pi_i = 0 \\
q_i^* &= \frac{a_i(q_i,q_j) - c_i + \frac{\partial \rho(q_i,q_j)}{\partial q_i}\pi_i}{2\beta - \frac{\partial a_i(q_i,q_j)}{\partial q_i}} - \frac{\gamma}{2\beta - \frac{\partial a_i(q_i,q_j)}{\partial q_i}}q_j
\end{aligned}$$

We obtain an identical best response function for the other organization analogously, and replacing values we obtain:

$$q_i^* = \frac{\left(a_i(q_i,q_j) - c_i + \frac{\partial \rho(q_i,q_j)}{\partial q_i}\pi_i)\right)\left(2\beta - \frac{\partial a_j(q_i,q_j)}{\partial q_j}\right) - \gamma\left(a_j(q_i,q_j) - c_j + \frac{\partial \rho(q_i,q_j)}{\partial q_j}\pi_j)\right)}{\left(2\beta - \frac{\partial a_i(q_i,q_j)}{\partial q_i}\right)\left(2\beta - \frac{\partial a_j(q_i,q_j)}{\partial q_j}\right) + \gamma^2}$$

and similarly for q_j^*.

Comparative Statics To obtain the key comparative static, we define

$$G(q_i,q_j) \equiv \frac{\partial V_i}{\partial q_i} = a_i(q_i,q_j) + \frac{\partial a_i(q_i,q_j)}{\partial q_i}q_i - 2\beta q_i - \gamma q_j - c_i + \frac{\partial \rho(q_i,q_j)}{\partial q_i}\pi_i$$

which is a continuously differentiable function from $\mathbb{R}^2 \to \mathbb{R}$. At the optimum, we know that $G(q_i^*, q_j^*) = 0$. We also assume that $\frac{\partial G(q_i^*,q_j^*)}{\partial q_i} \neq 0$.

Then, we can use the implicit function theorem to obtain our main comparative static in the formulation with endogenous demand. Note this implies that there is a neighborhood of (q_i^*, q_j^*) such that when q_i is close enough to q_i^*, we have a unique q_j such as $G(q_i,q_j) = 0$, which makes q_j a continuous function of q_i. The comparative static is:

$$\frac{\partial q_i}{\partial q_j} = -\frac{\partial G(q_i,q_j)/\partial q_j}{\partial G(q_i,q_j)/\partial q_i} = -\frac{\frac{\partial a_i(q_i,q_j)}{\partial q_j} + \frac{\partial^2 a_i(q_i,q_j)}{\partial q_i \partial q_j}q_i - \gamma + \frac{\partial^2 \rho(q_i,q_j)}{\partial q_i \partial q_j}\pi_j}{2 * \frac{\partial a_i(q_i,q_j)}{\partial q_i} + \frac{\partial^2 a_i(q_i,q_j)}{\partial q_i \partial q_i}q_i - 2\beta + \frac{\partial^2 \rho(q_i,q_j)}{\partial q_i \partial q_i}\pi_i}$$

Now, a sufficient condition for having a positive cross partial is that $a_i(\cdot)$ is downward sloping on the product $\left(\frac{\partial a_i(q_i,q_j)}{\partial q_i} < 0\right)$, that the decrease is at decreasing rates $\left(\frac{\partial^2 a_i(q_i,q_j)}{\partial q_i \partial q_i} < 0\right)$, that the loyalty function is concave $\left(\frac{\partial^2 \rho(q_i,q_j)}{\partial q_i \partial q_i} < 0\right)$, and that $\frac{\partial a_i(q_i,q_j)}{\partial q_j} + \frac{\partial^2 a_i(q_i,q_j)}{\partial q_i \partial q_j}q_i - \gamma + \frac{\partial^2 \rho(q_i,q_j)}{\partial q_i \partial q_j} > 0$. The interpretation of the last condition would depend on what captures $a_i(\cdot)$.

A.3 Bertrand competition with differentiated products

Now we model a state and a gang engaging in Bertrand competition with differentiated products and externalities from gang governance. Each organization has a certain valuation of the loyalty of the people.

Setup A gang g and a state s compete over prices p_g and p_s. Each organization chooses to maximize their respective pay-off based on parameters. Both organizations have a constant marginal cost c. Products are differentiated so the quantity demanded of each one is given by $q_i = a_i - b_i p_i - \gamma p_j$, where $\gamma < 0$ as goods produced by both organizations are assumed to be substitutes. The pay-off for each organization is $\nu_i = (a_i - b_i p_i - \gamma p_j)(p_i - c) + F(p_i, p_j)$, where $F(p_i, p_j)$ captures externalities. For simplicity, we assume an interior solution.

Nash Equilibria We begin by deriving the best response function for each organization:

$$\begin{aligned}
\max_{p_i} \nu_i &= (a_i - b_i p_i - \gamma p_j)(p_i - c) + F(p_i, p_j) \\
\frac{\partial \nu_i}{\partial p_i} &= (a_i - 2b_i p_i - \gamma p_j - b_i c) + \frac{\partial F(p_i, p_j)}{\partial p_i} = 0 \\
p_i^* &= \frac{\left(a_i - b_i c + \frac{\partial F(p_i, p_j)}{\partial p_i}\right)}{2b_i} - \frac{\gamma p_j}{2b_i}
\end{aligned}$$

Analogously, we obtain an identical best response function for the other organization. Replacing values we obtain:

$$p_i^* = \frac{2b_j a_i - \gamma a_j - (2b_j b_i - \gamma b_j)c + 2b_j \frac{\partial F(p_i^*, p_j^*)}{\partial p_i} - \gamma \frac{\partial F(p_j^*, p_i^*)}{\partial p_j}}{(4b_i b_j - \gamma^2)}$$

Comparative statics To obtain our key comparative static, we define:

$$M(p_i, p_j) \equiv \frac{\partial \nu_i}{\partial p_i} = (a_i - 2b_i p_i - \gamma p_j - b_i c) + \frac{\partial F(p_i, p_j)}{\partial p_i}$$

which is a continuously differentiable function. In the optimum we know that:

$$M(p_i^*, p_j^*) = \frac{\partial \nu_i}{\partial p_i} = (a_i - 2b_i p_i^* - \gamma p_j^* - b_i c) + \frac{\partial F(p_i^*, p_j^*)}{\partial p_i} = 0$$

Additionally, we assume that $2b_i \neq \frac{\partial^2 F(p_i^*, p_j^*)}{\partial p_i \partial p_i}$, so:

$$\frac{\partial M(p_i^*, p_j^*)}{\partial p_i} = -2b_i + \frac{\partial^2 F(p_i^*, p_j^*)}{\partial p_i \partial p_i} \neq 0$$

Then we can use the implicit function theorem, obtaining the following result:

$$\frac{\partial p_i}{\partial p_j} = -\frac{\partial M/\partial p_j}{\partial M/\partial p_i} = -\frac{\gamma - \frac{\partial^2 F(p_i^*, p_j^*)}{\partial p_i \partial p_j}}{2b_i - \frac{\partial^2 F(p_i^*, p_j^*)}{\partial p_i \partial p_i}} \tag{9}$$

Thus, to have a crowding in effect $\left(\frac{\partial p_i}{\partial p_j} > 0\right)$ we require that $\gamma < \frac{\partial^2 F(p_i^*, p_j^*)}{\partial p_i \partial p_j}$. Since $\gamma < 0$, a sufficient condition for the state to crowd in the gang is that both services are complements in loyalty. Generally, it is enough that the complementarity in loyalty is higher than the degree of substitution of these services.

A.4 Public goods and encompassing interest

Section A.2 showed how state-provided protection and governance may "grow the pie", and how this can be incorporated into standard models of duopolistic competition. This idea lies at the very heart of standard political economy of governance and public-goods provision. This section adapts the classic Olson & McGuire (1996) (henceforth OM) model, in which stationary bandits face incentives to curtail their own coercive taxation and provide public goods at their own expense, precisely because doing so grows the pie that the bandit later taxes. We abstract from OM's comparison of autocracy and democracy, instead comparing the baseline OM model of a monopolistic, autocratic stationary bandit to a modified version in which two stationary bandits tax and provide public goods to the same subject population. A simple two-bandit model predicts crowding out. We first illustrate the single-bandit baseline model and then add the second stationary bandit.

A.4.1 Baseline: One Autocratic Stationary Bandit

- One player: The state (S) makes two independent choices, setting a level of public goods provision (G_s) and a uniform tax rate $t_s \in [0, 1]$.
- The output of the economy Y is increasing convexly in the total amount of public goods provided G (which here equals G_S since there is only one stationary bandit), and no production is possible without *some* amount of public goods. That is, for $Y(G)$ we assume $Y(0) = 0$, $\frac{\partial Y(G)}{\partial G} > 0$, and $\frac{\partial^2 Y(G)}{\partial G^2} < 0$.
- We assume that taxation distorts economic activity. Write $\tau(t_s) \in [0, 1]$ represent the loss factor due to taxation, so that final GDP is equal to $\tau(t_s) * Y(G_s)$. We assume $\frac{\partial \tau(t_s)}{\partial t_s} < 0$.

The state maximization problem is given by:

$$V_s = \tau(t_s) \cdot t_s \cdot Y(G) - cG_s \tag{10}$$

For simplicity, we normalize c to 1. By construction, S sets taxes independently of the desired level of public good. At the optimal t_s^*, S's gains from taxation and the increases of potential output losses due to further distortion into the economy are equal: $\tau(t_s) * t_s = \frac{\partial \tau(t_s)}{\partial t_s}$. This can be seen in the FOC for Equation 10:

$$\frac{\partial V_s}{\partial t_s} = \left(\tau(t_s) + t_s \frac{\partial \tau(t_s)}{\partial t_s}\right) \cdot Y(G_s) = 0$$

$$\tau(t_s) + t_s \frac{\partial \tau(t_s)}{\partial t_s} = 0$$

$$t_s^* = -\frac{\tau(t_s)}{\frac{\partial \tau(t_s)}{\partial t_s}} \tag{11}$$

Finally, the state selects the level of public good in the point were the marginal revenue is equal to the marginal cost of the public good c multiplied by the reciprocal of state's share of the national potential income.

$$\frac{\partial V_s}{\partial G_s} = \tau(t_s) \cdot t_s \cdot Y'(G_s) - 1 = 0$$

$$Y'(G_S^*) = \frac{1}{t_s^* \tau(t_s^*)} \tag{12}$$

A.4.2 Dual stationary bandits

The setup is similar but with two players, a state (s) and gang (g).

- Players $i \in \{s, g\}$ simultaneously choose levels of public-goods provision (G_i) and a uniform tax rate (t_i) which, as before, is independent of public-goods provision.
- Economic output depends on the total of the two actors' public good provision: $Y(G)$ where $G \equiv G_s + G_g$. As before, $Y(0) = 0$, $\frac{\partial Y(G)}{\partial G} > 0$ and $\frac{\partial^2 Y(G)}{\partial G \partial G} < 0$
- Distortion $\tau(t)$ depends on the total amount of taxes levied: $t \equiv t_s + t_g$. As before: $\tau(0) = 1$ and $\frac{\partial \tau}{\partial t} < 0$.
- To ensure that neither player sets $t_i > .5$, we assume that $\tau(.5) = 0$.

Players' utility functions and maximization problems are symmetrical:

$$V_i = t_i \cdot \tau(t_i + t_j) \cdot Y(G_i + G_j) - c_i G_i \text{ for } i, j \in \{s, g\} \tag{13}$$

In this simplest, symmetric-players iteration, we will assume that $c_i = c_j = 1$. Asymmetric costs raise important questions of sequencing, and will be considered in future iterations. As before, we solve the two maximization problems separately, starting with taxation.

Optimal taxation with dual stationary bandits

Lemma A.1. *In equilibrium, players' optimal tax rates are identical: $t_i^* = t_j^*$.*

Proof. From 13, the FOC for player i is

$$\begin{aligned} \frac{\partial V_i}{\partial t_i} &= \tau(t_i + t_j)Y(G) + t_i \frac{\partial \tau(t_i+t_j)}{\partial t_i} Y(G) = 0 \quad (14) \\ &= \tau(t_i^* + t_j^*) + t_i \frac{\partial \tau(t_i+t_j)}{\partial t_i} = 0 \quad (15) \end{aligned}$$

We can rewrite $\frac{\partial \tau(t_i+t_j)}{\partial t_i}$ as $\frac{\partial \tau(t_i+t_j)}{\partial t} \frac{\partial t}{\partial t_i} = \frac{\partial \tau(t_i+t_j)}{\partial t} = \tau'(t)$. This yields:

$$t_i^*(t_j) = -\frac{\tau(t_i^* + t_j)}{\tau'(t_i^* + t_j)}$$

By a similar derivation, $t_j^*(t_i) = -\frac{\tau(t_i+t_j^*)}{\tau'(t_i+t_j^*)}$

So in Nash Equilibrium:

$$t_i^* = -\frac{\tau(t_i^* + t_j^*)}{\tau'(t_i^* + t_j^*)} = t_j^*$$

□

In words, when i increases t_i he gets a larger share of a smaller pie. These two effects must be of equal size at the optimum t_i^*. But the negative effect on the size of the pie is the same whether i or j is raising their rate. Therefore, the increase in $i's$ share at t_i^* must be the same as the change in $j's$ share if she were to raise *her* rate. But these "shares" are just each player's tax rate. So these must be equal.

Optimal public-goods provision with dual stationary bandits In this simultaneous setup, we identify Nash equilibria in which player i's choice of G_i is a best response to player j's choice of G_j and vice versa. Solving FOC for Equation 13 for G_i and G_j yields

$$\begin{aligned} \max_{0 \leq G_i} V_i &= t_i \cdot \tau(t_i + t_j) \cdot Y(G_i + G_j) - c_i G_i \\ \frac{\partial V_i}{\partial G_i} &= t_i \tau(t_i + t_j) Y'(G_i + G_j) - c_i \leq 0 \end{aligned}$$

where the last condition hold with equality if $G_i > 0$. We can write player i's best response function implicitly (i.e. i wants to set G_i^* such that):

$$G_i^*(G_j) : Y'(G_i^* + G_j) \leq \frac{c_i}{t_i^* \tau(t_i^* + t_j^*)} \quad (16)$$

and player j wants to set G_j^* such that

$$G_j^*(G_i) : Y'(G_i + G_j^*) \leq \frac{c_j}{t_j^* \tau(t_j^* + t_j^*)} \quad (17)$$

If we assume $c_i = c_j = c$ and with no loss of generality that $c = 1$ then (because $t_i^* = t_j^*$)

there is a unique total G that is optimal for both players, call it G^*_{2B}:

$$Y'(G^*_{2B}) = \frac{1}{t^*_i \tau (t^*_j + t^*_j)} \tag{18}$$

And there is a continuum of Nash equilibria characterized by $G^*_i = G^*_{2B} - G^*_j$. Obviously, the cross-partial of this relationship, $\frac{\partial G^*_i}{\partial G^*_j}$ is negative.

In words, there is an optimal total amount of public-goods to be provided, and either player is happy to provide goods until total goods supplied reached that optimum. Obviously, each prefers that the other do it. But for any division of the optimal total amount between the two players, neither has an incentive to deviate. Whats more, if some force outside the model pushed the result from one equilibrium to another, in which one player's public goods provision increased, it is obvious that the other player's optimal response would decrease.

In this way, the basic stationary bandit model, which explicitly accounts for economic growth produced by governance provision, can be said to predict crowding out.

A.5 General formulation and alternative models

Here we abstract away from the examples of Cournot and Bertrand competition, or stationary bandits above. Instead of modeling competition with one model or the other, we could use a general form $D(q_i, q_j)$ that encompasses all of these models (including Bertrand. Likewise, instead of modeling the externality as $\rho(q_i, q_j)\pi_i$ we use a general form $F(q_i, q_j)$.

Comparative statics We now define a value function where we are agnostic about how duopolistic competition takes place:

$$V_i = D(q_i, q_j) + F(q_i, q_j)$$

Then we can define the first partial in q_i as:

$$G(q_i, q_j) \equiv \frac{\partial V_i}{\partial q_i} \quad = \quad \frac{\partial D(q_i, q_j)}{\partial q_i} + \frac{\partial F(q_i, q_j)}{\partial q_i}$$

which is a continuously differentiable function from $\mathbb{R}^2 \to \mathbb{R}$. As a technical note, we assume that there exists a point such that $\frac{\partial V_i}{\partial q_i} = 0$, and the functions $D(\cdot)$ and $F(\cdot)$ are concave so that the sum of both functions is also concave. This implies there is a unique solution. At the optimum, we know that $G(q^*_i, q^*_j) = 0$. We also assume that $\frac{\partial G(q^*_i, q^*_j)}{\partial q_i} \neq 0$.

Finally, we can use the implicit function theorem to obtain our main comparative static in the general formulation. Note this implies that there is a neighborhood of (q^*_i, q^*_j) such that when q_i is close enough to q^*_i, we have a unique q_j such that $G(q_i, q_j) = 0$. This makes q_j a continuous function of q_i. The comparative static is:

$$\frac{\partial q_i}{\partial q_j} = -\frac{\partial G(q_i, q_j)/\partial q_j}{\partial G(q_i, q_j)/\partial q_i} = -\frac{\frac{\partial^2 D(q_i, q_j)}{\partial q_i \partial q_j} + \frac{\partial^2 F(q_i, q_j)}{\partial q_i \partial q_j}}{\frac{\partial^2 D(q_i, q_j)}{\partial q_i \partial q_i} + \frac{\partial^2 F(q_i, q_j)}{\partial q_i \partial q_i}}$$

where $\frac{\partial^2 D(q_i,q_j)}{\partial q_i \partial q_i} \leq 0$, $\frac{\partial^2 F(q_i,q_j)}{\partial q_i \partial q_i} \leq 0$ would mean decreasing returns of production in loyalty and profit. With this assumption, a positive numerator is sufficient for a positive cross partial.

B Using current versus historical state headquarters

Our estimating equation uses historical headquarters to estimate Δd_{ij}. Conceptually, however, we could use current headquarters and also estimate a causally-identified treatment effect. The two estimates represent slightly different treatments, and we are interested in both. For example, to the extent that we want to evaluate the effects of the border introduction jointly with the state's subsequent investments in local protection services (a key part of the broader and longer-term reform), we are interested in the treatment effects that use the location of new headquarters built after 2000.

It is natural to worry that the use of current headquarters would undermine the credibility of the causal estimates. After all, we would expect the government to choose the location of new headquarters according to levels of combo activity and rule. This intuition is somewhat misleading, however, and here we show that the causal interpretation holds even if protection services are endogenously placed closer to high-crime or criminally-governed areas.

In practice, we have data on both current and historical headquarters and, as it happens, we see that the estimates are nearly identical. Ex ante, however, there was no reason to expect the treatment effects to be the same. There are reasons to expect current headquarter estimates to be more precise, however.

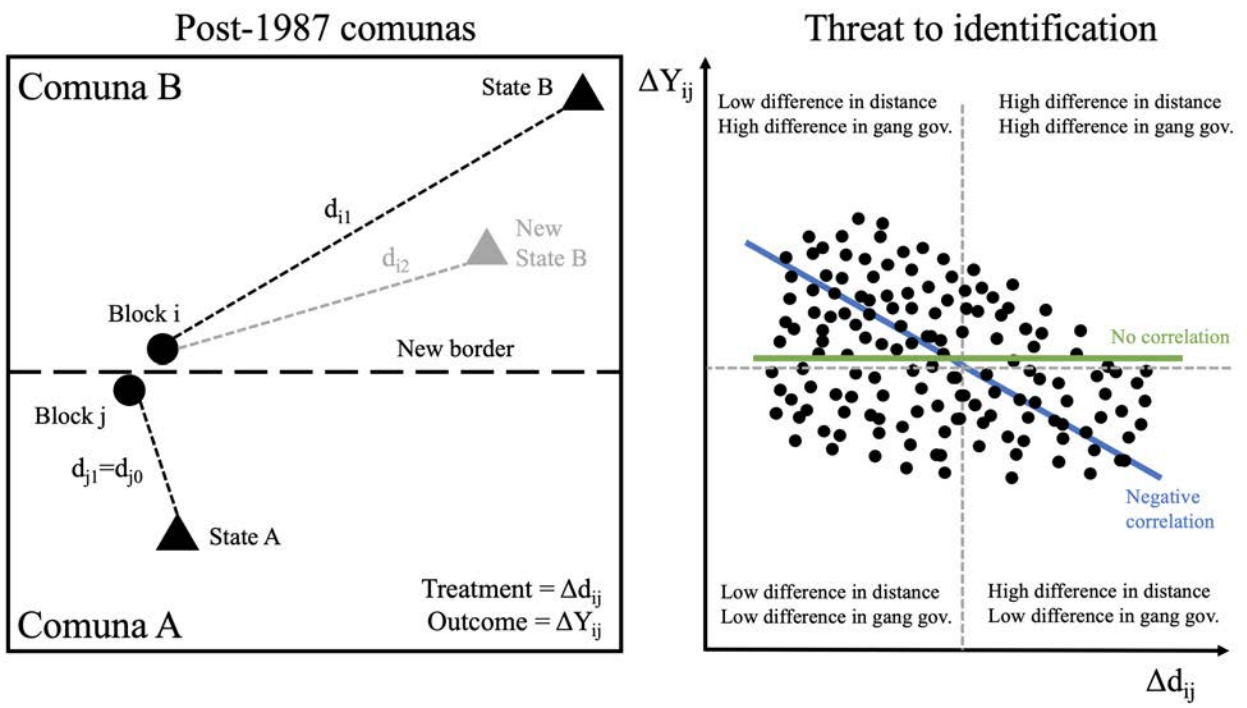
Figure B.1 helps to illustrate why endogenous headquarter construction does not bias estimates, but rather changes the interpretation. On the left is the same stylized comuna division as before. Now, however, we introduce a new state headquarters in Comuna B. First notice that, generally-speaking, this will reduce the absolute value of the shock Δd_{ij}. Suppose that the state systematically chooses to place new headquarters near stronger combos. This could induce a negative correlation between combo rule Y_i and d_i. This correlation would not be due to a combo strategic response, but rather as a mechanical function of government constriction choices. (Indeed, this is why the correlation in Figure 6 was confounded.)

Of course, we are not estimating the effect of d_i on Y_i, but rather the effect of Δd_{ij} on ΔY_{ij}. In general, blocks i and j are occupied by the same combo, and have similar baseline levels of governance. Thus, locating endogenously closer to a better-governed block means locating closer to better-governed pair. This should not affect our estimates of β. Rather, endogenously locating state headquarters will bias our estimated β only if *the government builds stations close to blocks not where gang governance is high or low, but rather where the cross-border difference in gang governance is high.* We illustrate this point on the right side of Figure B.1, that depicts a scatterplot of block-pairs. Suppose there were no correlation between the treatment variable, Δd_{ij}, and our outcome variable, ΔY_{ij}. This scenario is represented by the horizontal line in green. What construction strategy would create a false negative association between the two variables (the downward-sloping blue line)? Governments that build stations mainly in areas with high ΔY_{ij}. That is, governments that move block pairs from the upper right to the upper left quadrant.

Besides the absence of a compelling logic for such a decision rule, such endogeneity is unlikely for several reasons: the small number of new headquarters; the state's difficulty assessing differences in gang rule; and the logistical constraints on where land for new government buildings can be found.

There are reasons to expect estimates with current headquarters to be more precise, however. On occasion, new headquarters are constructed close enough to flip the direction

Figure B.1: Stylized illustration of threats to identification using current state headquarters



Notes: On the left is a sketch of the post-1987 comunas, including the development of new headquarters. Because of new infrastructure, block i can access the state at location New State B. On the right is a sketch of the relationship between the treatment and outcome variables, to illustrate under which scenario is endogenous location of headquarters problematic.

of treatment, changing the order of blocks i and j. Since most new construction happened more than a decade ago, state and gang rule on these blocks have evolved accordingly. The use of historical borders will generate noisier estimates of the treatment effects.

C Supplemental tables

Table C.1: Correlates of combo governance

		Combo Governance								
	Sample Mean (SD)	Estimate (SE) [p-value]	Estimate (SE) [p-value]	Estimate (SE) [p-value]	Estimate (SE) [p-value]	Estimate (SE) [p-value]	Estimate (SE) [p-value]	Estimate (SE) [p-value]	Estimate (SE) [p-value]	Estimate (SE) [p-value]
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Log of drug seizure value	18.124 (1.436)	0.056** (0.023) [0.013]								
Distance from drug corner	3.812 (2.550)		-0.022 (0.023) [0.337]							
Distance from gang HQ	2.510 (2.054)			-0.147*** (0.023) [0.000]						
Count of combo groups	0.358 (0.564)				0.086*** (0.023) [0.000]					
Distance to razón HQ	15.033 (9.699)					-0.128*** (0.025) [0.000]				
Meters to clinics and schools	369.508 (173.506)						-0.073*** (0.023) [0.001]			
Meters to business center	328.389 (61.341)							-0.029 (0.023) [0.213]		
Block average elevation	1,615.679 (133.649)								0.117*** (0.022) [0.000]	
Poverty Index (2018)	14.525 (15.752)									0.139*** (0.022) [0.000]
N		1,922	1,922	1,904	1,922	1,436	1,922	1,903	1,922	1,922

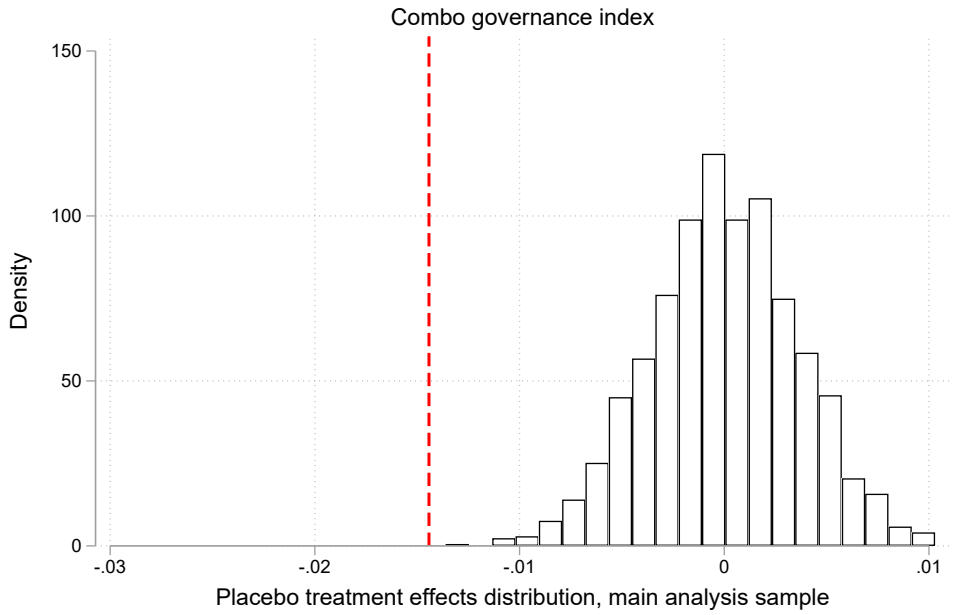
Notes: We run ordinary least squares (OLS) regressions of each measure on a range of available block- and neighborhood characteristics. Regressions are estimated at the block level. 143 missing values were imputed for the poverty index.

Table C.2: Coefficient of closeness to state services for police stations and municipal headquarters, separately

		Correlation with Δd_{ij}		
		Police & municipal agencies	Only municipal agencies	Only police stations
	Subsample Mean (SD)	Estimate (SE) [p-value]	Estimate (SE) [p-value]	Estimate (SE) [p-value]
	(1)	(2)	(3)	(4)
State				
Δ State Governance Index (0-1)	0.41	-0.012**	-0.009**	-0.002
	(0.19)	(0.005)	(0.004)	(0.003)
		[0.023]	[0.018]	[0.445]
Δ State Legitimacy Index (0-1)	0.58	-0.005	-0.003	-0.002
	(0.15)	(0.005)	(0.003)	(0.003)
		[0.322]	[0.383]	[0.428]
Combo				
Δ Combo Governance Index (0-1)	0.32	-0.014**	-0.005	-0.007*
	(0.21)	(0.006)	(0.004)	(0.004)
		[0.012]	[0.197]	[0.059]
Δ Combo Legitimacy Index (0-1)	0.42	-0.010	0.005	-0.002
	(0.21)	(0.007)	(0.005)	(0.004)
		[0.134]	[0.270]	[0.639]
N for Governance outcomes		571	571	571
N for Legitimacy outcomes		426	426	426

Notes: This table calculates the effect of being 100 meters further from either the nearest comuna police station, or the nearest municipal headquarter in the comuna. Note that other state agencies and service may be correlated with these two agencies, and so we cannot interpret these are causal effects of the individual headquarters with any certainty.

Figure C.1: Placebo treatment effects, inner border, main analysis sample, 1,500 simulations.



Notes: The figure depicts the distribution of average treatment effects of the difference in distance to the state on the difference in combo governance. Rather than matching pairs of blocks across the border, as in our main specification, we randomly matched our baseline sample of blocks with other blocks within the same comuna, using 1,500 simulations. We take blocks within 300 meters of the border and match them with blocks within a distance of at most 600 meters, resembling our baseline specification. We demean the distribution of treatment effects because of a mechanical bias resulting from the approach to the placebo exercise: typically, blocks closer to the border are assigned to state services farther away from the state, in a different comuna. The vertical line indicates our observed treatment effect.

Table C.3: Coefficient of closeness to state services for blocks along the inner comuna borders of Medellín on governance index components

		State	Combo
	Subsample Mean (SD)	Estimate (SE)	Estimate (SE)
	(1)	(2)	(3)
Relative Governance Index	0.09	-0.012**	-0.014**
	(0.24)	(0.005)	(0.006)
HH: Someone refuses to pay a big debt	-0.15	-0.010	-0.016*
	(0.33)	(0.009)	(0.010)
HH: There is domestic violence	0.16	0.004	-0.011
	(0.33)	(0.010)	(0.010)
HH: Two drunks fight on the street	0.14	0.007	-0.010
	(0.35)	(0.010)	(0.009)
HH: Kids fight on the street	-0.02	-0.015*	0.000
	(0.31)	(0.009)	(0.008)
HH: Home improvements affect neighbors	0.17	0.006	-0.003
	(0.33)	(0.011)	(0.008)
HH: Someone is making noise	0.25	0.003	-0.007
	(0.31)	(0.009)	(0.007)
HH: People smoking marijuana near children	0.06	-0.029***	0.010
	(0.31)	(0.009)	(0.008)
HH: Someone is mugged on the street	-0.02	-0.003	-0.015
	(0.34)	(0.009)	(0.010)
HH: A car or motorbike is stolen	0.06	0.016*	-0.013
	(0.36)	(0.009)	(0.009)
HH: Someone is threatening someone else	0.01	-0.017*	-0.004
	(0.34)	(0.010)	(0.009)
HH: It is necessary to prevent a theft	-0.01	-0.032***	-0.005
	(0.35)	(0.008)	(0.009)
HH: You have to react to a robbery	0.02	-0.003	-0.010
	(0.37)	(0.009)	(0.009)
Biz: Someone does not want to pay a debt	-0.04	-0.015	-0.048***
	(0.36)	(0.015)	(0.016)
Biz: Businesses in this sector are robbed	0.07	-0.023	-0.017
	(0.46)	(0.016)	(0.019)
Biz: Someone disturbs a business	0.14	-0.037***	-0.037**
	(0.45)	(0.013)	(0.016)
Biz: It is necessary to prevent a theft	0.10	-0.040***	-0.032*
	(0.48)	(0.013)	(0.017)
Biz: You have to react to a robbery	0.13	-0.036**	-0.016
	(0.48)	(0.014)	(0.017)
Number of observations		571	564

Notes: This table calculates the effect of being 100 meters further from the state on the 17 components of our governance indexes. HH indicates questions asked to households, and Biz represents questions asked to businesses.

Table C.4: Is police presence different at the border

Variable	Patrolling time	
	Estimate	Estimate
	[SE]	[SE]
	(1)	(1)
Distance to the border	0.007	0.001
	0.005	0.006
	0.205	0.920
Distance to the police station	-0.009	-0.012
	0.007	0.006
	0.242	0.96
Distance to the border x Distance to the police station		7.61e-06
		5.74e-06
		0.212
District fixed effects	Yes	Yes
Observations	36,946	36,946

Table C.5: Heterogeneous impacts on governance and legitimacy of being 100 meters more distant from the local state, by local drug seizure value

	State		Combo	
	Governance (Δ)	Legitimacy (Δ)	Governance (Δ)	Legitimacy (Δ)
	Estimate (SE)	Estimate (SE)	Estimate (SE)	Estimate (SE)
	(1)	(2)	(3)	(4)
Correlation with Δd_{ij}	-0.011*	-0.001	-0.024***	-0.008
	(0.006)	(0.006)	(0.007)	(0.009)
Correlation with Δd_{ij} X Low drug seizures	-0.002	-0.004	0.012*	-0.005
	(0.006)	(0.006)	(0.006)	(0.008)
Low drug seizures	0.023	-0.001	0.004	0.034
	(0.040)	(0.036)	(0.045)	(0.053)
Traditional and fixed point running var	✓	✓	✓	✓
Border FE	✓	✓	✓	✓
N	571	432	564	426

Notes: Each column comes from a separate regression. Drug seizures are geographically matched to each block using a 400 meter buffers.

D Measurement error

In this appendix, we study how measurement error can affect the interpretation of our results. We focus on three types of measurement error: reporting endogenous to gang rule, reporting endogenous to relative state governance, and reporting endogenous to both gang and state governance separately. In each case, we study how the reporting error changes the coefficient we estimate in the main results of the paper, and discuss the direction of the bias it induces.

Reporting endogenous to gang rule Suppose the true relationship between combo and state governance is given by:

$$g_c^* = \alpha + \beta g_s^* + \epsilon \tag{19}$$

However, suppose g_c^* is systematically under-reported in the survey at a rate proportional to combo governance, as follows:

$$g_c = \ g_c^* + \mu \tag{20}$$

Where $0 < \delta < 1$ is the reporting rate of gang governance. Then:

$$\frac{g_c - \mu}{\delta} = \alpha + \beta g_s^* + \epsilon \tag{21}$$

and hence we would estimate:

$$g_c = \delta\alpha + \delta\beta g_s^* + \nu \tag{22}$$

Where $\nu = \delta\epsilon + \mu$. Usingh observed data we will estimate $\hat{\delta\beta} < \beta$, which means that we underestimate the crowding out/in coefficient of gang governance.

Reporting endogenous to relative state governance Now let's continue to the same true relationship between g_c^* and g_c^*, but now under-reporting depends on relative state/combo governance:

$$g_c = \lambda(g_c^* - g_s^*) + \mu \tag{23}$$

Where $0 < \lambda < 1$ is the reporting rate of relative state governance. Then:

$$g_c^* = \frac{1}{\lambda} g_c + g_s^* - \frac{\mu}{\lambda} \tag{24}$$

and hence we would estimate:

$$g_c = \lambda\alpha + \lambda(\beta - 1) g_s^* + \eta \tag{25}$$

Where $\eta = \lambda\epsilon + \mu$. Using oberved data we will estimate $\lambda(\hat{\beta} - 1) < \beta$, which means that we, again, underestimate the crowding out/in coefficient of gang governance.

Reporting endogenous to gang and state governance, separately Now suppose we let reported gang governance depends on actual gang and state governance, where reporting rates are different can be different in each case:

$$g_c = \delta g_c^* + \lambda g_s^* + \mu \tag{26}$$

Where both: $0 < \lambda < 1$ and $0 < \delta < 1$ are the reporting rates of state and combo governance, respectively. Then:

$$g_c^* = \frac{1}{\delta}(g_c + \lambda g_s^* - \mu) \tag{27}$$

and we will estimate:

$$g_c = \delta\alpha + (\delta\beta + \lambda)g_s^* + \mu + \delta\epsilon \tag{28}$$

We generally would have $\delta\hat{\beta} + \lambda < \beta$, except in a specific situation: when gang governance is high and the state is far away, but there is no misreporting when the government is close.