

Le Roux, Leonard

Working Paper

Informality and violence: Evidence from South Africa

WIDER Working Paper, No. 45/25

Provided in Cooperation with:

United Nations University (UNU), World Institute for Development Economics Research (WIDER)

Suggested Citation: Le Roux, Leonard (2025) : Informality and violence: Evidence from South Africa, WIDER Working Paper, No. 45/25, ISBN 978-92-9256-604-3, The United Nations University World Institute for Development Economics Research (UNU-WIDER), Helsinki, <https://doi.org/10.35188/UNU-WIDER/2025/604-9>

This Version is available at:

<https://hdl.handle.net/10419/322229>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Informality and violence

Evidence from South Africa

Leonard le Roux*

June 2025

Abstract: This paper studies the emergence of violence in legal markets, with a particular focus on the informal public transport sector in South Africa, where minibuses are integral to urban mobility. Despite being a legal sector, the taxi industry is plagued by significant and persistent violence, imposing substantial social costs. Using novel administrative data from the government's operating licence system, route-level pricing information, and a unique dataset of taxi-related violence compiled from print and online media, I examine the interplay between competition, firm growth, and violence. I leverage variation from a sudden decline in rail service quality to estimate the effect of local demand shocks on taxi-related violence in a difference-in-differences framework. The taxi associations most positively affected by this demand shock experience substantial increases in violence. Excess entry appears to be the main mechanism driving this result. These findings suggest firms do not internalize the costs of violence in ways that limit its occurrence when faced with economic shocks, highlighting the limits of private organizing in the absence of well-functioning state institutions.

Key words: informality, public transport, violence, South Africa

JEL classification: O18, O17, K42, R40

Acknowledgements: I would like to thank Benjamin Marx, Pierre-Philippe Combes, Roberto Galbiati, Golvine de Rochambeau, Marleen Marra, Florian Oswald, Clara Santamaria, Marcos Sanso-Navarro, Lucas Conwell, Nonhlanhla Nyathikazi Nkosi, Sihle Thusi, Mothusi Motsosi, Bafana Mkhwebane, Sharon Matsila, Francois Lafont, Juan Sebastián Ivars, Ségal Le Guern Herry, and Kim Thomas for their help and advice. Thanks to seminar participants at Sciences Po, the SA-TIED work-in-progress workshop, the Warwick and HEC PhD conferences, AMSE, the Dauphine TrEnCE Working Group, the IAST-OxPo conference, the CSAE, and the ICDE. Work on this paper benefited from a research grant from the UNU-WIDER SA-TIED initiative (UNU Contract ID 605UU-3635). I gratefully acknowledge funding towards my doctoral studies by the Oppenheimer Memorial Trust, the Sciences Po Department of Economics, and the Aix-Marseille School of Economics. All errors remain my own.

* Department of Economics, Sciences Po, Paris, France; corresponding author: leonardleroux32@gmail.com

This study has been prepared within the UNU-WIDER project **Southern Africa—Towards Inclusive Economic Development (SA-TIED)**.

Copyright © UNU-WIDER 2025

UNU-WIDER employs a fair use policy for reasonable reproduction of UNU-WIDER copyrighted content—such as the reproduction of a table or a figure, and/or text not exceeding 400 words—with due acknowledgement of the original source, without requiring explicit permission from the copyright holder.

Information and requests: publications@wider.unu.edu

ISSN 1798-7237 ISBN 978-92-9256-604-9

<https://doi.org/10.35188/UNU-WIDER/2025/604-9>

United Nations University World Institute for Development Economics Research provides economic analysis and policy advice with the aim of promoting sustainable and equitable development. The Institute began operations in 1985 in Helsinki, Finland, as the first research and training centre of the United Nations University. Today it is a unique blend of think tank, research institute, and UN agency—providing a range of services from policy advice to governments as well as freely available original research.

The Institute is funded through income from an endowment fund with additional contributions to its work programme from Finland and Sweden, as well as earmarked contributions for specific projects from a variety of donors.

Katajanokanlaituri 6 B, 00160 Helsinki, Finland

The views expressed in this paper are those of the author(s), and do not necessarily reflect the views of the Institute or the United Nations University, nor the programme/project donors.

1 Introduction

A central question in economics is how markets function in the absence of the state. This question is especially important in developing countries, where firms often cannot rely on well-functioning state institutions to protect their property rights or enforce contracts. In such settings, firms depend on frameworks of private rules and constraints: systems of ‘self-governance’ (Dixit 2003). A longstanding literature in economics has documented how such private institutions can work well (Bernstein 2013; Ostrom 1990; Williamson 1979),¹ and recent work suggests they play a consequential role in economic behaviour (Banerjee et al. 2024; Breza et al. 2022). On the other hand, private institutions can break down and generate costs. It is important to understand the limits of self-governance and the nature of these costs, especially in developing countries, where we know the least about the consequences of how firms organize and interact.

In this paper, I study a legal sector, where, in the absence of well-functioning state enforcement, registered firms have come to rely on the threat of violence to enforce property rights. How do such markets respond to economic shocks? I study this question in the context of the informal commuter transport sector in South Africa, where privately owned minibus taxis are crucial to urban mobility.² Firms hold state-mandated monopolies on routes but—absent effective state enforcement—disputes often emerge when rival firms intrude on each others’ routes. These disputes regularly culminate in severe violence that has come to constitute a significant social cost. Criminologists suggest that 50% of all assassinations in the country are attributable to this industry (Thomas 2021). The issue is so commonplace that media houses have tags for ‘Taxi violence’ on their online media archives.³ What are the economic drivers of such violent outbreaks?

To answer this question, I use administrative data from the government’s operating licence system to construct a geolocated panel of firms—taxi associations—that operate on routes. I join this administrative data to a unique dataset on outbreaks of taxi-related violence I construct from an archive of print and online news media, and a novel dataset of collected route-level prices. I use these data to study the interactions between competition, firm growth, commuter prices, and violence in a legal market. Geographically, the project is focused on Gauteng Province (centred around Johannesburg and Pretoria) and the provinces immediately sur-

¹ Indeed, informal firms, which account for 70% of employment in emerging and developing countries, generally function in this way (Ohnsorge and Yu 2022).

² Sixty per cent of all commuters and 80% of public transport users commute by minibus taxi (Statistics South Africa 2021b). ‘Matatus’ in Kenya, ‘Chapas’ in Mozambique, and ‘Molue’/‘Danfo’ in Nigeria are all similar semi-informal urban transport service providers.

³ See, for example, <https://www.sabcnews.com/sabcnews/tag/taxi-violence/> and <https://www.news24.com/tags/topics/taxiviolence>.

rounding it. This area contains approximately one-quarter of South Africa's population (16 million inhabitants) and is one of Africa's largest metropolitan areas.

The relationship between violence and economic factors in the taxi sector is likely to be endogenous. To identify the causal effects of local demand shocks on violence when state institutions function poorly, I exploit quasi-experimental variation in the demand faced by some taxi associations but not by others. This variation in demand occurs in the aftermath of a negative supply shock to commuter rail, an important alternative commuting mode, that was unevenly distributed across space. Following relatively stable performance between 2010 and 2015, the state-owned Passenger Rail Agency of South Africa (PRASA) experienced dramatic declines in service quality from 2015 onwards, after revelations of widespread internal maladministration. The response to the scandal led to managerial instability, the termination of irregularly granted (but operationally crucial) contracts, and widespread organizational paralysis. The number of commuter trains running on time declined by 17% between 2015 and 2016 and by 65% by early 2020.⁴ Train ticket sales in Gauteng declined from 3.323 million in January 2015 to 532,575 in January 2020 (a decline of 84%). Commuters substituted away from passenger rail to, *inter alia*, minibus taxis. The decline in train operations began virtually as soon as the rail agency CEO implicated in the scandal resigned in March 2015.

I adopt a difference-in-differences framework pre- and post-2015 for associations with route networks closer to and more distant from the most-affected rail lines. Geocoding the text descriptions of association routes allows me to construct a measure of the distance of association route networks to the rail network. Using monthly data on the number of trains operating on each rail line, I construct a distance-weighted measure of the size of the demand shock affecting each taxi association. In my main specification, I estimate the impacts of this positive demand shock on levels of violence experienced by the taxi associations most exposed to it. To do so, I construct a new dataset on violence in this sector from an archive of print and online news media. I identify the association names mentioned in news reports of shootings and assassinations of taxi drivers or owners that are clearly related to taxi-industry violence—that is, where both the primary target and the perpetrator of the violence belong to the taxi industry.

The central results are that associations located near train stations that experienced the largest declines in train volumes experienced significant increases in taxi-related shooting incidents in the aftermath of the rail shock. The effect size is large and equates to a doubling of the probability of involvement in a shooting in a given year.

I test the robustness of this result in several ways. First, I find similar results when, instead of running an association-year-level regression, I re-estimate the baseline results using the

⁴ In January 2015, 23,034 trains ran on time, compared to 8,040 in January 2020. These figures are specific to Gauteng, but other commuter networks experienced similar declines.

geocoded locations of taxi-related shootings to run an area–year-level regression. In the aftermath of the decline in commuter rail, controlling for fixed differences between locations and years, there is a significant increase in the incidence and frequency of shootings in areas containing train stations. Second, though the outcome variable in my main specification is restricted to clear cases of taxi-related violence (and should thus not be construed as homicides in general), I also test whether the decline in rail quality might have had positive spillovers on local crime rates that might in turn have had positive spillovers on taxi-related violence. Using police precinct-level crime statistics, I find that precincts containing train stations experienced significantly lower total crime rates following the train shock. I also control for local crime rates directly in my main specification, without this affecting the main results, ruling out this potential channel. Third, the significance of the positive pooled average effect I estimate for the 2015–19 period is robust to dropping observations from any single post-treatment year, or any association from the sample. Finally, I run a placebo regression in which treatment assignment is based on association distances to non-commuter train stations, as opposed to the commuter stations used in the main analysis. In this specification, I find a null result. Moreover, I find similar results for other large demand shocks, by using the dates and locations of the openings of roughly 450 large shopping malls across the country. The opening of a mall is associated with a significant increase in taxi violence in the area it occurs, in the year after its opening.

I explore two policy-relevant axes of heterogeneity in this main result. The first is to understand how the results vary by baseline market structure. A clear question that emerges in this setting is whether the number, size, or locations of rival firms in the market makes violence more likely to occur in certain areas compared to others when shocks occur. For each association, I create a measure of competition in the market it operates on, based on the number and size of rival associations with similar route networks. I find that associations operating in more competitive markets at baseline (pre-2015) experience larger increases in violence in the aftermath of the train shock. These results suggest that the way in which routes are allocated, and the ways in which entry is regulated, can have material effects on the level of violence in the industry. Second, I study the role of local state capacity. To do so, I use a police precinct-level panel of local crime counts to construct two measures of how well each precinct is performing in reducing crime rates in the pre-train shock period. I link these precinct-level measures to taxi associations via their route networks and compare the outcomes of associations with route networks in well-performing police precincts to those in poorly performing precincts. I find that local state capacity, as measured by local police performance before the train shock, significantly mitigates the effect of the demand shock on the emergence of violence. This result suggests that well-functioning police can play an important role in decreasing violence in the taxi sector.

The main mechanism I propose to explain why violence increased in the aftermath of a positive demand shock is excessive market entry. While few new associations entered the market,

existing associations allowed many more new drivers to operate on their routes. This increase in supply was excessive and led to a decrease in the number of potential customers per taxi driver, without any clear price response. To document these responses, I use the administrative data and show that associations located near the rail network register significantly more new members in the aftermath of the decline in rail service quality and are less active in opening new taxi routes. I estimate the size of the change in demand at the association level by using two household travel surveys conducted before and after the decline in the rail network. I estimate a decline in the number of commuters per driver in Gauteng Province in general, and especially for associations most affected by the train shock. I then use these same survey data to approximate the change in the price for commuting by taxi in commuting zones most affected by the shock. I find that commuters in baseline rail commuting areas experience significant increases in their real all-mode monthly commuting costs. However, when I measure only costs paid in taxi fares, and control for time spent commuting by minibus taxi, I do not find any significant change in per-minute costs of commuting by taxi.

Taken together, these results suggest that the increase in violence I observe was driven by an increase in scarcity in the market that paradoxically emerged when participants in the industry did not respond efficiently to a large increase in demand. This is a puzzling result, because given the high costs of violence, one would expect that participants internalize the risk of violence in their economic decisions. While I cannot directly test why this does not occur, there are various potential explanations. One of these is that when associations take on the de-facto role of transport planners they have imperfect information on what all other actors in the market are doing, and so do not make socially optimal decisions. A second potential reason is that the organizational structure of associations means that association executives, whose incentives are to increase the size of the association, do not fully internalize the costs of doing so on their members (taxi drivers and owners) who are most exposed to violence.

The main contribution of this paper is to document that when state institutions do not function well, violence can emerge from economic competition in a legal sector, and that when such markets experience unexpected shocks, firms do not necessarily internalize the costs of violence in ways that would limit its occurrence. This leads to economic responses that increase the incentives for violence, highlighting the limits of private organizing in the absence of well-functioning state institutions. It is one of the first attempts to study the emergence of violence at the firm level in a legal sector. In doing so, it emphasizes that violence can emerge in certain legal markets when the state is absent, just as it does among criminal groups that are targeted by the state.

There is significant anecdotal evidence that violence is also used in similar markets throughout Sub-Saharan Africa and elsewhere. The Kenyan *matatu* industry, for example, shows strik-

ing similarities,⁵ as do the operations of transport unions in Lagos, Nigeria.⁶ In Colombia, the period between the 1960s and 1990s came to be known as the *guerra del centavo* or ‘penny war’, when an oversupply of poorly regulated private buses led to dangerous and sometimes violent competition among bus drivers (Rodriguez et al. 2016; World Bank 2014).

Many legal markets find ways to sustain mutually beneficial cooperation, relying on social ties for information and social sanctions as punishments, even when state institutions function poorly. I posit that the nature of economic activity can make this type of self-governance especially difficult to sustain in certain types of markets, and it is in such settings that violence emerges when the state is absent. The theoretical literature has long recognized that for private institutions to function well, participants need to be able to observe each other’s behaviour, and be able to immediately detect and effectively punish opportunistic behaviour (Dixit 2003, 2004). The transport market studied here makes this type of self-governance difficult to sustain in three interlinked ways. The first is the difficulty of defining property rights when routes span multiple kilometres, often intersect, and commuters can be picked up at any point. The second is the high returns from opportunistic behaviour. The pool of commuters is more or less constant and first-movers who can serve these commuters can enjoy high profits, while their rivals are left empty-handed. Finally, transport markets have inherently high information frictions. There is a large number of actors operating on markets that span large geographical distances. This means that even with the threat of extreme punishment, there is always the prospect of enjoying high profits by intruding onto the market of a rival without being caught.

I contribute to three main strands of the academic literature. The first of these is the literature studying functioning of firms in weakly institutionalized settings (La Porta and Shleifer 2014; Ulyssea 2018). Recent work suggests that social norms and community expectations play an important role in economic behaviour in developing countries. Banerjee et al. (2024), for example, suggest such norms limit profit-maximizing behaviour and keep firms small. Breza et al. (2022) suggest they are important in explaining downward wage rigidities. This literature motivates that markets in developing countries are less competitive than commonly thought (Bergquist and Dinerstein 2020). Other recent work has highlighted the importance of relational contracts (Macchiavello 2022; Macchiavello and Morjaria 2021, 2022), and is related to earlier literature on economic governance in the absence of the state (Bernstein 2013; Dixit 2004; Greif 1993; Ostrom 1990). In contrast to previous work, my paper highlights a previously under-studied private and social cost of contract enforcement when state institutions do not function well: violence. While the recent literature has documented the importance of pri-

⁵ See Katumanga (2005: 514) or Anderson (2002: 358). More recent incidents of violence in the matatu industry are covered by Kenyan news agencies (e.g. Citizen TV Kenya 2018).

⁶ See Fourchard (2023, 2024) and Agbiboa (2022).

vate institutions guiding firm behaviour, my paper additionally highlights their limits when large, complex markets are exposed to demand shocks, generating severe costs.

Second, my paper contributes to the literature on the economics of crime and violence. Traders in illegal goods cannot turn to the police or the courts, and often struggle to trust each other enough to cooperate (Gambetta 2006, 2009). A large body of recent work has studied violence among competing drug cartels in Latin America (Brown et al. 2024; Castillo et al. 2020; Dell 2015; Dell et al. 2019), and highlighted the economic motives of criminal networks (Arellano-Bover et al. 2024; Blattman et al. 2021; Dube et al. 2016; Herrera and Martinez-Alvarez 2022; Le Moglie and Sorrenti 2022; Magaloni et al. 2020; Sviatschi 2022).⁷ Important contributions have documented increased violence in the aftermath of revenue windfalls in criminal markets (Castillo et al. 2020; Kronick 2020; Mejia and Restrepo 2013), and provided theoretical rationalizations (Castillo and Kronick 2020; Garfinkel and Skaperdas 2007).⁸ The returns to illegal activity have been found to be an important driver of criminal behaviour (Dell et al. 2019; Draca et al. 2019). Finally, a related literature has studied the impacts of economic shocks on conflicts (Bazzi and Blattman 2014; Berman and Couttenier 2015; Berman et al. 2017; McGuirk and Burke 2020).⁹ One common factor in much of this literature is a focus on markets for illegal goods: drugs, arms, extortion. In contrast, the present article highlights how violence can result from institutional failures more broadly in fully legal sectors. Moreover, I conduct a study at the firm level, in contrast to much of this existing literature. Finally, I also contribute to a literature that has historically centred on Latin America and Italy, and where little work has been done on the direct interactions between economic competition and violence in legal markets on the African continent.

Finally, this paper contributes to the economic literature on urban mobility in lower- and middle-income countries. I build on existing work on the South African minibus taxi industry, spearheaded by Andrew Kerr (Antrobus and Kerr 2019; Kerr 2017, 2018). More recently, Conwell (2024) builds and estimates a model of privatized shared transit on data from Cape Town, finding significant welfare gains from reducing wait times through subsidization. Other contemporary work on urban public transport in lower- and middle-income countries includes that of Gaduh et al. (2022), Tsivanidis (2019), Lane et al. (2024), and Zenou (2011).¹⁰ In contrast to previous work, my paper highlights the central role of associations, presenting a large body of

⁷ A large political science literature also contributes to our understanding of gang violence (Dube et al. 2013; Durán-Martínez 2015; Kronick 2020; Skarbek 2011).

⁸ This result is not a-priori evident, as seminal theories in the economics of crime predict that windfalls would also increase the opportunity cost of violence (Becker 1968; Dal Bó and Dal Bó 2011).

⁹ McGuirk and Burke (2020) make the distinction between factor conflicts—where the goal of the conflict is to permanently displace contenders—and output conflict—where conflict arises over the appropriation of surplus.

¹⁰ Franklin (2018) documents the binding constraint that high transport costs place on the employment outcomes of young people in African cities.

new panel data on their evolution and the structure of this market. I also illustrate the responsiveness of informal transport providers to shocks to alternative public transport modes and document some of the effects of a large decline in urban commuter rail.

The remainder of this paper is organized as follows: Section 2 provides an overview of the functioning of the minibus taxi industry. Section 3 presents the collected and administrative data and the descriptive statistics. Section 5 presents the empirical strategy and Section 6 provides the results. The final two sections present a discussion and conclusion.

2 Background: the minibus taxi industry

This section provides a brief background of the functioning of the minibus taxi industry before moving on to the empirical analysis (a detailed overview is provided in Kerr (2018)).

Taxi associations and owners. Taxi associations are the unit of analysis of this paper, and can be seen as transport unions that control sets of taxi routes. Taxi owners buy 14-seater minivans and pay a joining fee and monthly membership fees to associations in order to have the right to operate on the associations' routes. This joining process is opaque, often reliant on informal connections, and associations do not automatically accept prospective owners.¹¹ Associations decide on the internal allocation of routes among their members and set prices (CCSA 2021). They play a crucial role in reducing competition by restricting market entry and coming to cooperative agreements with rivals.¹²

Routes and operating licences. Contrary to the conventional sense of the term 'taxi', minibus taxis operate on specific routes. By law, owners are required to obtain licences for the routes they operate on. In practice, owners generally apply for a licence to operate on all of their association's routes. This application requires a letter of support from an association. The provincial transport regulatory entity sends the application to the municipality in which the route is located, which makes a recommendation on whether the application should be accepted.¹³ The application is then published in a government gazette for public comment. If no public

¹¹ Associations often limit the on-boarding of new members because this can directly affect the profits of their existing members. Table D8 in the Appendix presents descriptive statistics from data collected in December 2023 for a sample of 92 taxi associations. Average joining fees are ZAR63,513 (US\$3,500). Average rank fees are ZAR1,015 per taxi per month. The data were collected by contacting the taxi associations via telephone or their Facebook pages.

¹² Bank (1990) provides an account of the emergence of associations and the ensuing emergence of organized violence on long-distance routes between the Qwa-Qwa 'Bantustan' area and Gauteng in the 1970s and 1980s. In that instance, associations emerged through a mixture of coercion by strongmen in the industry who aimed to reduce oversupply on popular routes, as well as voluntarily, as individual taxi drivers joined associations in order to access legal representation following a clampdown by the apartheid state.

¹³ Anecdotal discussions between the author, senior transport officials, and taxi drivers suggest this process is subject to widespread corruption in practice.

objections are received, the licence is granted. Operating licences are inexpensive (ZAR300 (US\$16)) compared to the fees involved in joining an association, and are valid for a period of up to seven years. Associations often have a long-running presence on routes and apply for new routes in a similar process.¹⁴ While most transport departments try not to issue operating licences for the same route to rival associations, associations often operate in very close proximity to one another. The state has no role in the *within-association* decision of how many drivers to allocate to each route. This allows associations to restrict supply on certain routes and to allocate lucrative routes to favoured or powerful owners.

Informality vs. formality. The taxi industry pays little to no corporate or personal income tax.¹⁵ Workers are generally employed without written contracts (87%), without union representation (93%), and without employer medical or unemployment insurance contributions (95%), as shown in Table D7. These elements of informality coexist with the fact that owners and associations register with the Transport Department and pay nominal fees for operating licences. Ulyssea et al. (2023) make the distinction between informality on the extensive margin—whether firms register and pay entry fees to become formal—and informality on the intensive margin—whether workers are hired without formal contracts. Strictly, the taxi industry falls into the latter category. However, this paper highlights how informality coexists with poorly implemented efforts at formalization.

Prices and collusion. Prices are set by taxi associations in an opaque process. The Department of Transport does not maintain a database of prices commuters pay, despite the central place of the taxi industry in urban mobility.¹⁶ There is widespread anecdotal evidence of collusion among taxi associations. For example, associations often agree not to pick up passengers on parts of routes they share with others.¹⁷ They also often engage in ‘joint ventures’, which are explicit agreements to share specific routes, agreeing on how many vehicles from each association should drive on the route and what the price should be.¹⁸ There is also widespread joint price setting, as illustrated in Figure C4, which the Competition Commission appears to

¹⁴ CCSA (2021) provides more information on the operating licence application process.

¹⁵ In 2021 the taxi industry paid only ZAR5 million (US\$275,000) in corporate income tax (MOF 2021), while being estimated to generate ZAR90 billion in yearly revenue (Mabogo 2023). Overall, 95% of taxi drivers report not being registered for personal income tax, as shown in Table D7.

¹⁶ Several independent private and public research initiatives have aimed to collect price data, such as WhereIsMyTransport, the CSIR, and TaxiMap.

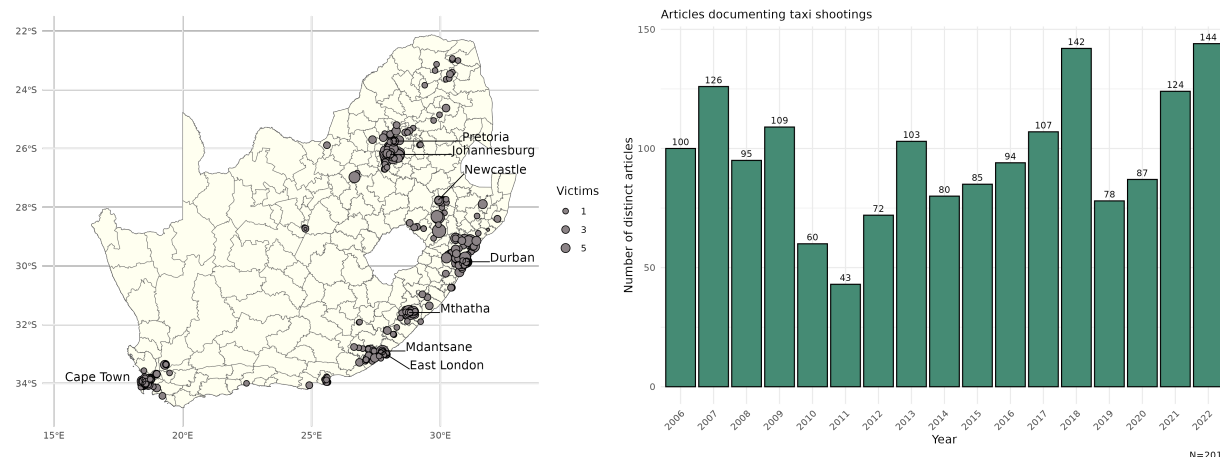
¹⁷ As a real-world example: two associations (NANDUWE and WATA) share a very similar route between Soweto (a major suburb) and Johannesburg city centre. There is an explicit agreement that each association does not pick up passengers on certain sections of the route. Each association stations armed members and ‘squad cars’ (vehicles with guards employed by associations to monitor compliance) on multiple street corners along the route to enforce the agreement.

¹⁸ An association executive told the author that joint ventures are one of the main mechanisms for resolving disputes.

accept.¹⁹ The overall regulatory structure is outlined in the National Land Transport Transition Act, 22 of 2000 (NLTTA) and National Transport Act 99 of 2009.

Violence. Taxi violence is a recurrent theme in South African current affairs. Criminologists have shown that targeted taxi-related assassinations have remained the largest single category of assassinations in the country over the last 15 years (Thomas 2021). Violence often manifests in the form of targeted assassinations of taxi owners and taxi drivers, often by hired hitmen.²⁰ At other times, it takes the form of shootings at taxi ranks or on taxi drivers while driving. Commuters are often injured or killed. Taxi association executive committee members are also sometimes assassinated. The violence is largely characterized as between taxi associations, although a minority of incidents take place within taxi associations as a result of internal disputes. Figure 1 presents an overview of the locations and evolution of taxi-related shooting events. The data sources used to construct these figures are presented in the next section.

Figure 1: Taxi industry shootings



Note: these figures are restricted to shooting incidents where taxi drivers, owners, or association executives are shot by their rivals—that is, taxi industry actors are both the primary target and the source of the violence. The data are constructed from local and national news clippings contained in the SABINET news archive over the period 2006–22.

Source: author's compilation based on data sources outlined in Section 3 and Appendix C3.

There is some consensus that pervasive violence in the taxi industry emerged following a period of deregulation in the late 1980s. The apartheid state initially repressed the industry and limited entry by black South Africans via various bureaucratic measures (Bank 1990; Khosa 1992). However, in the process of widespread deregulation that defined the final years of apartheid, the state released a white paper on transport in 1987 which promoted the industry and legalized 16-seater kombi taxis. A dramatic influx of new taxi operators ensued in this tumultuous period of the country's history. Taxi permits issued throughout South Africa increased from

¹⁹ See CCSA (2021: 163). Despite widespread anti-competitive behaviour, and the importance of the industry for urban mobility, the industry has never been prosecuted by the Commission.

²⁰ In South Africa, such hired assassins are known as 'Inkabis'.

1,453 in 1985 to 38,317 in 1990 (Khosa 1992). Violence associated with conflict over routes preceded this period but became heightened in the late 1980s and 1990s, when the term ‘taxi wars’ was first coined. This period also saw a rise in the importance and prominence of associations that took on the role of regulators in the absence of the state (Bank 1990; Dugard 2001; Khosa 1992; Sekhonyane and Dugard 2004).

3 Data and descriptive statistics

3.1 Data sources

There are two key barriers to quantitatively studying the relationship between competition and violence in the taxi industry. The first is the lack of publicly available systematic data on the economic behaviour of firms. I overcome this difficulty through access to a large administrative database from the Transport Department that has, to my knowledge not been used in previous work in economics, and collecting novel route-level data on prices set by associations. The second difficulty is the absence of firm-level and geocoded data on violence. I overcome this challenge by constructing such a database from an archive of print and online news media. I use these and other sources to create an association–year panel. This section presents a brief overview of the data sources used.

National Land Transport Information System (NLTIS). The primary source of data I use to study the taxi industry is a large administrative database maintained by the South African Department of Transport. The data obtained from the Department of Transport in May 2022 contain information on 179,944 registered individuals, the number of vehicles for 55,878 operators with valid operating licences; text descriptions of 17,578 taxi routes linked to the associations that operate on them; the date the taxi association was registered on the system and the date a route was registered on the system; and the names of 1,220 taxi associations. I construct a panel of taxi associations linked to a geocoded panel of routes for the 2010–20 period. More detailed information is provided in Appendix C1 and summary statistics are presented in Appendix A.

Taxi-related violence. I create a new database of violence in the taxi industry, sourced from local news media reports. To do so, I carry out a systematic search of print media and online press articles from 2006 to 2022. South Africa has a robust and active media landscape and taxi-related violence is regularly covered in online and print media. I rely on the SABINET archive of South African newspaper clippings and search for articles containing the terms ‘taxi’ and ‘shot’, ‘taxi’ and ‘shooting’, ‘taxi’ and ‘violence’, and ‘taxi’ and ‘geweld’ (the Afrikaans word

for violence) by year (SABINET 2022).²¹ This process yields 2,339 articles between 2006 and 2022. I extract the text from these articles and use it to identify a set of 1,300 articles that document distinct shooting events that are clearly related to taxi violence, where the primary target and clear origin of the violence are both part of the taxi industry.²² I extract the names of any taxi associations directly implicated in the shooting event (either by being affiliated with the victim or part of ongoing conflict the shooting is described to have stemmed from), the number of people killed and injured, the exact date of the shooting, and the exact location. I cross-reference the SABINET data with online news articles from the PressReader database of online news media to validate information and to ensure no events are missed (PressReader Inc. 2022). A more detailed discussion of the data construction is provided in Appendix C3.²³

Prices. I collect photographs and scans of 120 fare lists containing route-level prices of 92 taxi associations and over 2,200 specific taxi routes in Gauteng. This is done in three ways. First, through correspondence with the South African National Taxi Council (SANTACO)²⁴ I obtain current prices for all Tshwane (Pretoria) associations and some Johannesburg-based associations in March 2022. Second, I conduct an online search on Twitter and Facebook for the terms ‘taxi fare’ accompanied by the names of the associations in the NLTIS data.²⁵ Finally, a fieldworker visited the main taxi ranks in the Johannesburg region in December 2022 and took photos of any route-level fares advertised at ranks, or available from rank marshals. An example of one of these fare lists is presented in Figure C3.

Commuter rail. Through a series of written requests to the PRASA, I obtain monthly data on the number of trains running on each line of the commuter rail network in Gauteng for the 2009–22 period, as well as the number of tickets sold on each line for the 2013–22 period. I obtain the locations and names of train stations from OpenStreetMaps and cross-reference these with the names of stations on the PRASA commuter network. More information is provided in Appendix C2.

²¹ SABINET describes itself as: ‘A news clipping services collection comprising over 4.5 million articles. With an average of 2500 new articles added weekly, SA Media (News Clippings) is a comprehensive research tool that allows you to search local mainstream publications.’

²² For example, I exclude articles where taxi drivers are killed in robberies or in disputes that are clearly not related to the taxi industry. I also exclude cases where taxi drivers exert violence on external actors such as Bus Rapid-Transit or conventional bus drivers, which is not the subject of this paper.

²³ The Global Initiative Against Transnational Organized Crime (GI-TOC) reporting on assassinations in South Africa relies on a similar dataset, constructed using similar methodology, but which does not systematically capture detailed information on the associations involved, the nature of the conflict, or the exact locations of shootings (Thomas 2021). For these reasons I construct this dataset myself, additionally incorporating online news media articles.

²⁴ This is largest organizing body responsible for representing the taxi industry as a whole.

²⁵ Sometimes commuters post images of fare lists on social media to inform others.

Other data sources. In order to relate my study of transport providers to the responses of commuters, I rely on the National Household Travel Survey (NHTS) data from 2013 and 2020 (Statistics South Africa 2014, 2021b).²⁶ As a measure of local state capacity I also rely on the locations of police stations and a precinct-level panel of crime statistics for the 2010–20 period, published by the South African Police Service (SAPS 2020).

3.2 Descriptive statistics

Figure 2 presents the result of the geocoding of text descriptions of minibus taxi routes from the NLTIS data, following careful cleaning and verification. Figure D5 presents driving routes and the centroids of Gauteng associations and illustrates that many urban associations operate in very close proximity to one another.

Table A10 presents descriptive statistics on taxi owners and taxi associations from the NLTIS data. The median registered taxi owner has one vehicle, while the median taxi association has 103 registered members and operates on 13 distinct routes. There also exists significant variation in the concentration of the ownership of taxi operating licences within associations. The average concentration of vehicle ownership within associations is 0.16, while 16.6% of associations have a Herfindahl–Hirschman Index (HHI) score higher than 0.25.²⁷ Using information on when members register on the system and when routes are registered on the system, and averaging over all years from 2010 onwards, the median association takes on five new members per year and registers 0.09 new routes.²⁸ Additional summary statistics are presented in Appendix A.

Table D2 presents some of the provincial variation. On average, taxi associations in the Gauteng and the Eastern Cape have a larger membership than other provinces. The Northern Cape is anomalous in that it has a small number of relatively small taxi associations that have been very active in registering a large number of routes with transport authorities. Gauteng taxi associations also have more concentrated vehicle ownership within the association than in other provinces.

²⁶ The NHTS is a large cross-sectional survey carried out by Statistics South Africa. In 2013, 157,273 people were interviewed, and in 2020, 145,385 people were interviewed. The 2020 wave was completed in March, just before COVID-19 arrived in South Africa.

²⁷ This may be viewed as a measure of decision-making centralization at the association level. Where l_{ia} is the number of registered licences of operator i in association a with N members and L_a is the total number of licences under association a , $HHI_a = \sum_{i=1}^N (\frac{l_i}{L_a})^2$.

²⁸ The information system was introduced in 2003, illustrated by the large initial spike in operators and associations, followed by the roll-out of the NLTIS system. In analysis used later in this paper, only the period after 2009 is considered in order to avoid capturing variation in membership driven by the roll-out of the system.

Figure D4 presents some of the dynamics over time, using the dates that routes and members are added to the NLTIS system. In particular, this figure illustrates the very high membership growth among Gauteng taxi associations compared to other provinces.²⁹ Of the 219 registered associations in Gauteng, there are 113 (51% of the Gauteng sample) with over 100 members, and 35 associations (16%) have over 500 members. Descriptive statistics of prices are presented in Appendix D6.

3.3 Measuring market structure

Geocoding the routes in the administrative data allows me to measure the distance between different taxi association route networks. Using this distance measure, described below, I develop two association-specific measures of market structure: the number of competitor firms and the size of those competitor firms. Summary statistics of these measures are presented in Appendix D2.

Pairwise distances between association route networks. For each pair of taxi associations (a, b) I calculate the distances between all of their respective route pairs $dist(i_a, j_b)$. To reduce the computational requirements of this task, each route is defined to have two nodes, an origin and a destination. The distance of route i of association a to route j of association b is defined as:

$$dist(i_a, j_b) = \left(\min_1 [dist(o_{i,a}, o_{j,b}), dist(d_{i,a}, d_{j,b}), dist(o_{i,a}, d_{j,b}), dist(d_{i,a}, o_{j,b})] + \min_2 [dist(o_{i,a}, o_{j,b}), dist(d_{i,a}, d_{j,b}), dist(o_{i,a}, d_{j,b}), dist(d_{i,a}, o_{j,b})] \right) / 2 \quad (1)$$

where \min_1 and \min_2 are the first and second minima, and $dist(o_{i,a}, o_{j,b})$ is the straight-line distance between the origin of route i of association a and the origin of route j of association b (conversely, d represents the destinations). I treat origins and destinations symmetrically, because associations nearly invariably load passengers both on the forward and return journey. At the association-pair level $dist(a, b)$ can be defined as the median (or mean, or minimum) across all route-pair distances.

Holding demand constant, this would allow me to measure for each association a :

1. The average distance to potential rivals: $dist(a) = \frac{1}{n} \sum_{A-\{a\}} dist(a, a')$, where n is the number of rival associations in the market.³⁰

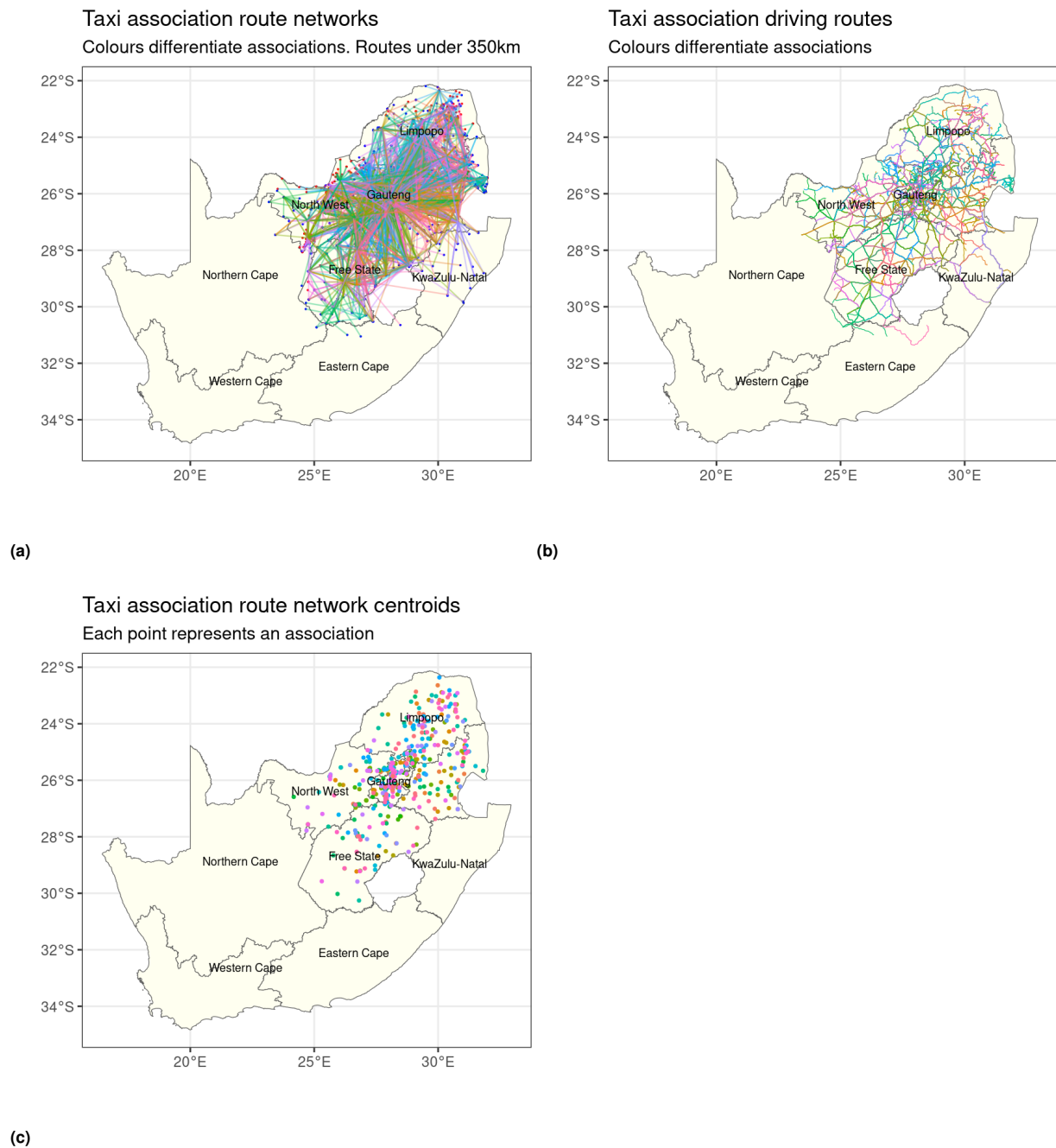
²⁹ This is in line with a key complaint raised by participants at the commission of inquiry into taxi violence (GDRT 2021).

³⁰ In the empirical setting, I define a market as a province.

2. The extensive-margin level of potential competition: $MS_a = \sum_{A-\{a\}} \frac{1}{dist(a,a')}$. This is a distance-weighted association-level count of the number of rival firms (other associations).
3. The intensive-margin level of potential competition: $MS_a = \sum_{A-\{a\}} \frac{N_{a'}}{dist(a,a')}$, where N is the number of drivers of association a' . This is a distance-weighted measure of the size of rival firms (the number of drivers of other associations).

I use these measures to study how levels of potential competition might affect violence in the taxi industry.

Figure 2: Geocoding of taxi association route networks



Note: the figures are restricted to taxi associations based in the provinces of Gauteng, North West, Limpopo, Free State, and Mpumalanga, which form part of the study area. The route centroids presented in the part (b) are the median coordinates of the taxi association origin and destination route nodes. The RAS data does not include the routes of Western Cape and KwaZulu-Natal taxi associations. Routes are registered with the transport authority in the province of the route's origin.

Source: author's compilation based on data sources discussed in the text.

4 Conceptual framework

Before turning to the empirical specification, I provide a brief discussion of the conceptual mechanisms underlying violence between taxi associations. Associations often coexist in peace for extended periods of time, but violence periodically flares up and subsides. In this paper, I argue that unexpected demand shocks can trigger economic responses that increase the incentives for violence.

Taxi associations frequently operate in close geographic proximity, competing over routes where property rights (e.g. passenger pickup locations) are ill-defined or weakly enforced by the state. Due to imperfect enforcement and significant monitoring costs, associations have substantial incentives to capture rivals' market shares opportunistically.

One way to illustrate some of the trade-offs in this setting is to consider two firms, A and B, that each operate on one route. Each firm can choose either to comply (respect current market segmentation) or to deviate (allow additional drivers onto routes contested by rivals, effectively encroaching on their market share). The strategic interaction resembles a prisoners dilemma:

Association A	Association B	
	Comply	Deviate
Comply	(C_A, C_A)	(L_A, W_B)
Deviate	(W_A, L_A)	(D_A, D_B)

In a symmetric setting, payoffs are ranked $W > C > D > L$. When both firms deviate, they compete for market share using violence v_i . Market share S_A for association A is determined by the relative intensity of violence, such that $S_A = \frac{v_A}{v_A + v_B}$. This is a standard formulation in the economics of conflict literature (Garfinkel and Skaperdas 2007). Violence also has a cost $\varphi(v)$ that is increasing in v .³¹ In this case their payoffs are $D_i = \frac{v_A}{v_A + v_B} - \varphi(v)$. If the game is played once or finitely, the dominant strategy is always to deviate. If the game is played indefinitely, and if actors adopt a grim trigger, the cooperative equilibrium can be sustained. As in Dixit (2003) or Macchiavello and Morjaria (2022), if a grim trigger strategy is followed by each firm, and each firm has a discount factor of δ , the necessary condition for cooperation is:

$$C + \frac{\delta}{1 - \delta}C \geq W + \frac{\delta}{1 - \delta}D_{\text{violence}} \quad (2)$$

which can be rearranged to give:

$$\underbrace{\frac{\delta}{1 - \delta}(C - D_{\text{violence}})}_{\text{Value of the relationship}} \geq \underbrace{W - C}_{\text{Temptation to deviate}} \quad (3)$$

³¹ Violence attracts police attention, diverts resources from productive activities (transportation), and the state sometimes closes routes as a result of violence.

In this framework, an unexpected positive demand shock on association A's route temporarily raises the payoffs (W) associated with deviation for association B. This increase alters the above inequality, reducing the relative value of cooperation compared to short-term gains from deviation. Given the repeated nature of interactions, associations would be expected to internalize the high costs of violence and cooperate to share the surplus. However, several frictions inherent to this market may inhibit such cooperation, including the difficulty of monitoring behaviour (implying actors can behave opportunistically for some time without being caught) and the presence of multiple competing associations, which makes coordination difficult. Conceptually, the above specification also suggests deviation, and violence would be higher in contexts where the costs of violence are lower.

5 Empirical strategy

The relationship between firms' economic behaviour and violence in the taxi industry is likely subject to endogeneity, particularly due to simultaneity bias. Economic factors, such as market entry, may increase competition and, as a result, the incentives for violence. At the same time, firms might strategically use violence to deter entry by rivals, causing reverse causality. To address this challenge, I exploit a natural experiment arising from a significant and unexpected positive demand shock driven by a substantial decline in commuter rail operations. This shock disproportionately affects certain taxi associations due to their geographic proximity to affected commuter routes. By leveraging this variation—unrelated to firms' strategic choices—I isolate the causal effect of changes in economic incentives on the incidence of violence.

5.1 Difference in differences

Shock to commuter rail. Commuter rail is a historically important mode of transport in urban areas. The rail network is operated by a large state-owned monopoly, the PRASA and its subsidiary, Metrorail.³² The commuter rail network in Gauteng operates on 27 service corridors.

While historically relatively stable, 2015 marked the beginning of a dramatic and unforeseen decline in the operational performance of the commuter rail network. This was sparked by a large corruption scandal at PRASA. The scandal started in 2015, when both the South African Auditor General and Public Protector released separate reports finding widespread maladministration at the rail agency. The CEO who was directly and personally implicated resigned in March 2015, following which the company saw six different CEOs in the next five years.

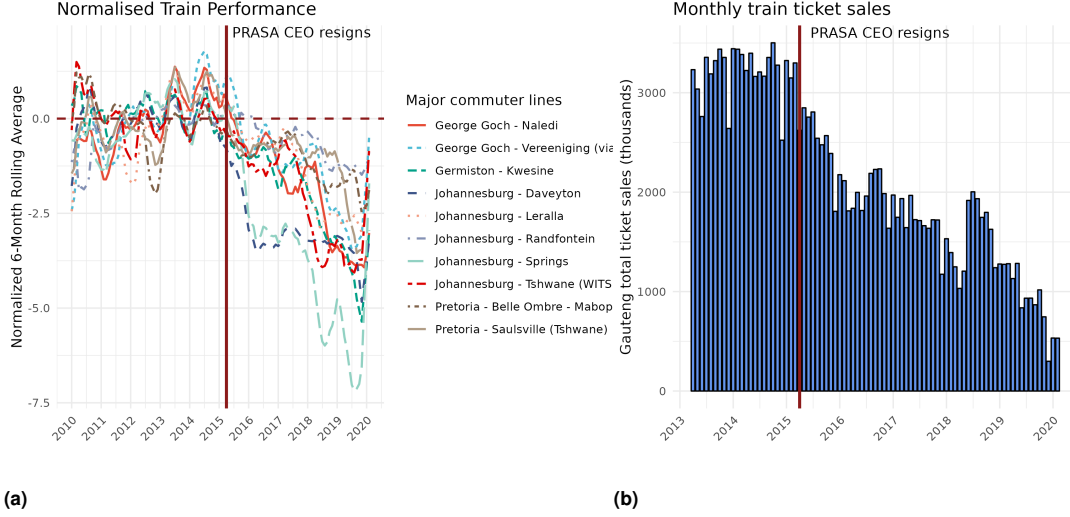
³² The commuter trains referred to here should not be confused with the Gautrain—a modern rail line built at the time of the 2010 Soccer World Cup. Low-income commuters (who rely on minibus taxis and commuter trains) generally cannot afford to take the Gautrain.

As shown in Figure 3, the decline in train operability and commuter ticket sales began almost immediately after this event. Along with the instability in senior management, the scandal led to the termination of multiple irregularly granted (but operationally crucial) contracts. This included contracts for train and train-line maintenance, security, and spare train parts. It also resulted in the initiation of forensic investigations of all large contracts, and direct oversight of procurement by the National Treasury. My conversations with a senior official at the rail agency suggest these well-intentioned measures resulted in severe organizational paralysis. Managers were scared of taking decisions for fear of being implicated in the scandal.³³ The board of control was also reconstituted on multiple occasions, with yearly changes. Appendix B1 provides more detailed context, including from conversations with senior PRASA employees. Figure 4 illustrates some of the geographic variation used in the empirical setting.

Commuters in historically rail commuting areas substituted away from the use of commuter rail, *inter-alia* towards the use of minibus taxis, as shown in Figure A1.

³³ One anecdote recounted by a PRASA official that is useful in understanding this is that members of the Special Investigation Unit set up temporary offices inside the PRASA building in Johannesburg.

Figure 3: Train performance and ticket sales



Note: part (a) presents a train-line-specific Z-score of the number of monthly trains running on time, normalized to the 2010–15 average. The vertical red line at March 2015 indicates the date the CEO of PRASA resigned in the wake of widespread corruption allegations and forensic investigations by the Public Protector. The horizontal line at 0 represents parity with the 2010–15 average number of trains per month for each line. The figure shows the ten largest commuter train lines. Part (b) presents monthly train ticket purchases in Gauteng. The ticketing data are available from 2013 onwards.

Source: author's compilation based on data from PRASA.

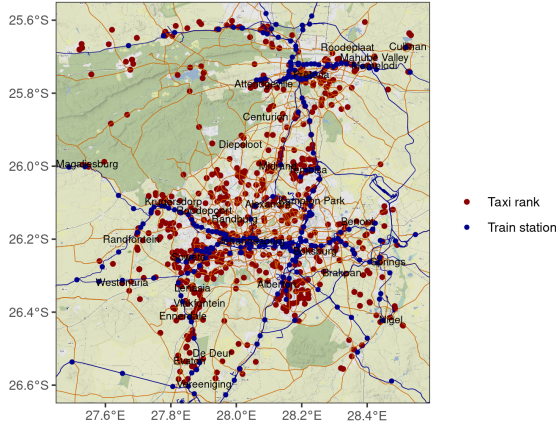
5.2 Measure of treatment intensity

I create a taxi-association-level measure of the size of the positive demand shock by using two sources of variation. The first is based on the distance between association route networks and the rail network. The second is based on the number of trains running on each rail line at each point in time. I combine these two sources into one measure of treatment intensity by weighting the size of the decline in train numbers on each rail line (pre- and post-2015) by the distance of that rail line to an association's route network.

Let the total number of trains running on time on line l at time t be denoted as $T_{l,t}$. I assign to each train station s the sum of the trains running on all lines that pass through the station $T_{s,t} = \sum_{l=1}^L T_{l,t}, \forall l$ intersecting s .³⁴ With this station-level measure of train volume (T), I construct Harris market potential measure to measure how exposed taxi associations are to changes in train frequency (Harris 1954). The intuition is that the importance of a shock at station s on association a depends both on the size of the shock ΔT_s and the distance between s and a . Changes in transport demand are important if they are large or close. In my DiD specification, I measure ΔT_s as the difference between the 2010–14 average number of trains per month and the 2015–19 average number of trains per month.

³⁴ The median station is used by one train line. The first and third quartiles of the distribution are one and three lines per station, respectively, and the maximum is nine.

Figure 4: Illustration of geographic variation in exposure to the rail network used in the empirical strategy



Note: this image illustrates the cross-sectional variation in exposure to the rail network used in the DiD estimation. Some taxi route origins or destinations are located near train stations. Others are located far from train stations.
Source: author's compilation based on data sources discussed in the text.

To create a measure of the proximity of taxi associations to the rail network, I calculate the distances between each association's taxi route nodes (origins and destinations) and commuter train stations. For each taxi route r , the distance of a taxi route node $n_r = \{o_r, d_r\}$ to a train station s is denoted $d(n_r, s)$. The distance of a route to a specific train station s is then $d(r, s) = \sum_n d(n_r, s)$. At the taxi route level, a measure of exposure to the rail network at a point in time t can then be defined as:

$$E_{r,t} = \sum_{s=1}^S \frac{(T_{s,t})}{d(s, r)} \quad (4)$$

I then aggregate this route-level variable to the association level $E_a = \sum_{R_a} E_{r_a,t}$. Using this measure of exposure to the rail network I measure the size of the change in train volume on each line pre- and post-2015, scaled by the distance of that line from an association's routes. At the route level:

$$\Delta E_r = \sum_{s=1}^S \frac{(\bar{T}_{s,pre2015} - \bar{T}_{s,post2015})}{d(s, r)} \quad (5)$$

where $\bar{T}_{s,pre2015(post2015)}$ is the average monthly number of trains running through station s between 2010 and 2014 (or between 2015 and 2019). At the association level $\Delta E_a = \sum_{R_a} \Delta E_r$. I use this measure as my main measure of treatment intensity. My main estimating equation is:

$$v_{a,t} = \beta_1 \Delta E_a \times post\ train\ shock_t + \gamma_a + \gamma_t + \delta X_{a,t} + \varepsilon_{a,t} \quad (6)$$

For ease of interpretation, I report results using various binary variables created by imposing treatment cut-offs based on the continuous treatment intensity variable (I present results with cut-offs at P70, P80, and P90). In the specification above, $v_{a,t}$ is measured principally as an

association-level shooting rate, the number of shootings in year t , divided by the number of members in year t . I also present results where v is alternately either a binary variable equal to 1 if association a is involved in a taxi-related shooting incident in period t , or a count of the number of shootings involving association a in period t , without weighting by the association size. *post train shock* is equal to 1 in all years after 2014 (the PRASA CEO tendered his resignation in March 2015). γ_a and γ_t are association and year fixed effects. Association fixed effects control for all time-invariant association-level factors that might be correlated with violence, such as the general location of the association, culture within the association, and the size of the association. Year fixed effects control for any temporal shocks that might affect all associations, such as changes in the petrol price. $X_{a,t}$ is a vector of time-variant association-level controls. I control for the number of routes, as well as the local crime rate on association route networks. In the main estimation table, I present both the main effects of these variables and their interactions with the *post train shock* variable.³⁵

The identifying assumptions of this type of DiD design with continuous treatment intensity are discussed in Callaway et al. (2024). Notably, there are two parameters of interest in such settings. The first is the causal effect of a change in the train exposure measure among those associations that are affected by the change, which can be identified under a standard parallel trends assumption. The second parameter of interest is the average effect of a marginal change in the exposure measure among units that are affected by the train shock. The authors suggest estimating this directly. I do so using a long-differences estimation.

Long differences

Using the exposure measure above, association-level changes in violence are related to the association-level size of the demand shock in a long-difference specification:

$$\Delta v_a = \alpha + \beta \Delta E_a + \delta_1 \Delta X_a + \delta_2 Y_a + \varepsilon_a \quad (7)$$

where Δ indicates the difference of each variable between its 2010–14 average and 2015–19 average. X is a vector of time-variant controls (the number of routes) and Y is a vector of association-level baseline controls (the number of routes and members in the 2010–14 period). ΔE is proposed as an association-level measure of the size of the demand shock that took place over the 2015–19 period.

³⁵ In order to provide visual evidence of whether the parallel trends assumption holds, a dynamic version of this specification, in which the treatment term is interacted with an indicator for each year around the treatment period, is also estimated: $v_{a,t} = \sum_{l=-4}^5 \beta_l \Delta_a \times \mathbf{1}[t = l] + \gamma_a + \gamma_t + \delta X_{a,t} + \varepsilon_{a,t}$.

6 Results

6.1 Descriptive results

I present two sets of descriptive results showing a negative correlation between prices and levels of competition at the route level, and a positive correlation between violence and levels of competition at the firm level. First, in Table 1, I use the cross-section of 1,465 collected route-level prices (described in Section 3 and Appendix C4) to present hedonic price regressions. I control directly for the driving distance between the route origin and destination, the population size at origin and destination (by means of origin and destination fixed effects), the overall level of demand and supply in the market that the association setting the price operates in (by means of district fixed effects at the association level), and various association-level controls, such as the number of drivers per route, the internal ownership structure, and the proximity to rival associations' routes.³⁶ Controlling for these factors, Table 1 documents a negative correlation between the route-level price paid by commuters and the total number of drivers with operating licences for the route (column 2) and the total number of associations with operating licences for the route (column 3) (The construction of these two measures is described in Section 3.3). The distance–price elasticity is 0.2, that is, on average a doubling of the distance of a route is correlated with a 20% increase in price.

Second, Table 2 presents association-level correlates with violence. The main insight is that there is a statistically significant positive correlation between the rate of involvement in violence and the level of competition at the association level. In the final columns, I control for district fixed effects (columns 7–9), which capture the most salient variation in demand. In addition, I control for other factors that may be correlated with violence, such as the size, ownership structure, and member-to-route ratio of each association. Within districts, associations that operate in markets with more rival drivers, or closer to the routes of rival associations, on average show higher rates of involvement in taxi-violence-related shooting events.

³⁶ A correlation table for the association-level variables is presented in Table D1.

Table 1: Descriptive hedonic price regression: trip-level price paid by commuters

	log(route level price)		
	(1)	(2)	(3)
Route-level observables			
Log(shortest drivable distance (km))	0.2138*** (0.0173)	0.2067*** (0.0154)	0.1994*** (0.0198)
Log(drivers on route)		-0.0722** (0.0233)	
Log(associations on route)			-0.1700** (0.0532)
Association-level controls			
Log(number of drivers per route)	0.0348 (0.0562)	0.0498 (0.0538)	0.0362 (0.0565)
Log(distance to rival associations)	-0.0791 (0.1425)	-0.0573 (0.1193)	-0.0670 (0.1455)
Log(license HHI within association)	0.0923 (0.0718)	0.0664 (0.0803)	0.0885 (0.0704)
Observations	1,465	1,465	1,465
R ²	0.90736	0.90811	0.90801
Within R ²	0.15812	0.16493	0.16402
Year fixed effects	✓	✓	✓
District fixed effects	✓	✓	✓
Origin fixed effects	✓	✓	✓
Destination fixed effects	✓	✓	✓

Note: The unit of observation is the price of an individual taxi route, collected from taxi fare lists (see Appendix C4 for an example and Section 3 for elaboration on the data capture). Origin and destination fixed effects are defined at the Uber-H3-6 grid-cell level, a 3.7 km edge-length hexagon. The number of potential drivers and associations on the route are two measures of potential competition calculated from the operating licences associations hold from the transport department. The origins and destinations of operating licences are geo-coded and linked to the origins and destinations of the taxi fare lists. The distance to rivals measure is calculated at the association level (see Section 3.3). District fixed effects are based on the district municipality where the median route of an association starts. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Source: author's calculations.

Table 2: OLS regression: association-level shooting rates and local market structure

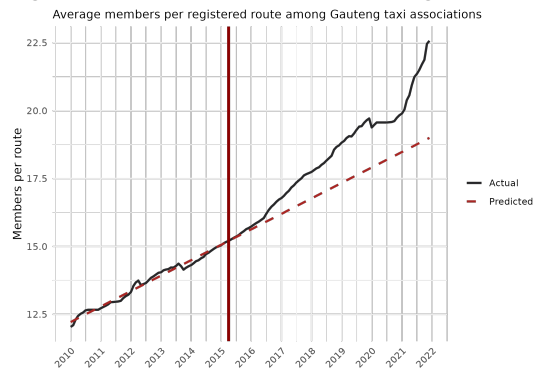
	Association-level shooting rate: shootings/drivers								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Market structure measures									
Measure 1: log(distance of competitor routes)	−0.0215*** (0.0044)			−0.0128** (0.0055)			−0.0147* (0.0073)		
Measure 2: log(number of rival associations)		0.0095*** (0.0018)			0.0051** (0.0026)			0.0034 (0.0023)	
Measure 3: log(number of rival drivers)			0.0090*** (0.0017)			0.0060** (0.0025)			0.0059** (0.0027)
Association controls									
Log(number of drivers)				−0.0120** (0.0051)	−0.0120** (0.0051)	−0.0120** (0.0050)	−0.0117* (0.0066)	−0.0113* (0.0064)	−0.0110* (0.0064)
Log(members per route)				0.0130** (0.0051)	0.0127** (0.0051)	0.0124** (0.0050)	0.0128* (0.0073)	0.0121* (0.0070)	0.0119 (0.0070)
Within-association vehicle licence HHI				−0.0338** (0.0154)	−0.0341** (0.0153)	−0.0347** (0.0153)	−0.0380* (0.0201)	−0.0382* (0.0195)	−0.0366* (0.0192)
Log(number of routes)				0.0157*** (0.0053)	0.0152*** (0.0053)	0.0152*** (0.0053)	0.0154 (0.0094)	0.0143 (0.0090)	0.0143 (0.0091)
Constant	0.1096*** (0.0219)	0.0049*** (0.0008)	−0.0348*** (0.0070)						
Province fixed effects				✓	✓	✓			
District fixed effects							✓	✓	✓
Standard errors				Heteroscedasticity-robust				District	
Observations	465	461	461	441	438	438	441	438	438
R ²	0.08579	0.08596	0.09681	0.16424	0.16221	0.16544	0.19818	0.19751	0.20093
Within R ²				0.06506	0.05904	0.06266	0.06000	0.05134	0.05538

Note: this table presents an OLS regression at the taxi association level. The outcome variable is an association-specific shooting rate: the total number of shootings the association is involved in, divided by the number of drivers registered under the association. This outcome is regressed on three measures of the supply side of the markets associations operate in (these measures are defined in Section 3.3). Measure 1 is the median distance to rival association route networks. Measure 2 is a distance-weighted count of rival associations. Measure 3 is a distance-weighted count of rival drivers. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Source: author's calculations.

Figure 5 plots the evolution of the number of minibus taxi licences over time, showing the clear deviation from the 2010–15 trend that commences following the start of the decline in train operability in 2015.

Figure 5: Minibus taxi association registered members per registered route



Note: this figure plots the monthly average cumulative number of licensed members per registered route among Gauteng-based taxi associations. The dotted line is a linear projection based on the 2010–15 period data. The vertical reference line at March 2015 indicates the date the CEO of the rail agency PRASA resigned, and widespread instability within the rail agency started.

Source: author's compilation based on data sources discussed in the text.

6.2 Difference-in-differences results

Table 3 presents the results of the baseline estimation presented in Equation 6 using a binary treatment cut-off of the treatment intensity variable at the 80th percentile. The outcome variable is the association-level shooting rate in a given year (the number of taxi-violence-related shootings divided by the number of drivers), which has a mean of 0.0008. The estimated coefficient on the DiD interaction term is 0.002, suggesting that the rate of involvement in taxi-related shootings more than doubled for associations most affected by the train shock in the period after the decline in the rail network. I provide visual evidence that the parallel trends assumption holds in Figure 6, where I additionally present two other binary treatment cut-offs at the 70th and 90th percentiles. This figure shows that the estimated coefficient size increases as the treatment cut-off increases; effects are larger for associations most affected by the rail shock. Figure A2 presents the dynamic specification of the fully continuous treatment intensity measure, showing a similar overall pattern to the results using binary treatment cut-offs. While the positive and significant pooled average treatment effect presented in Table 3 includes estimates from 2019, the coefficient estimate for 2019 itself is close to zero in Figure 6. This is due to widespread state intervention by the Gauteng provincial government in 2019, initiated in response to the severe taxi violence in the preceding years. I provide evidence to this effect in Figure A3 and Appendix B2. Specifically, the provincial government organized a taxi summit, where the first point of the transport head's speech was a call for unity and a cessation of the taxi war (see Appendix B2). This summit culminated in the creation of a taxi-monitoring committee, as well as a provincial commission of inquiry into the causes of taxi violence. These

mechanisms were mirrored by an increase in police enforcement (measured as the number of police-detected crime) in areas most affected by taxi violence, which I document using data on precinct-level crime statistics in Figure A3.

Table 3: DiD results: change in frequency of violence for associations most exposed to rail shock

	Shooting rate: # shootings/# members			
	(1)	(2)	(3)	(4)
Post train shock \times train routes	0.0023** (0.0009)	0.0023** (0.0010)	0.0020** (0.0009)	0.0020** (0.0010)
Standard errors				
Observations	3,110	3,107	3,107	3,107
Outcome mean	0.0008	0.0008	0.0008	0.0008
R ²	0.16843	0.16877	0.16978	0.17009
Within R ²	0.00767	0.00808	0.00929	0.00965
Association fixed effects	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓
Local crime rate control		✓		✓
Association controls		✓		✓
Local crime rate control \times post			✓	✓
Association controls \times post			✓	✓

Note: 'Train routes' is equal to 1 if an association is above the 80th percentile of the continuous train shock exposure measure defined in Section 5.1 and 0 if it is below the median. Specifically: the line-specific 2015–20 average rail operationality is subtracted from the 2010–15 average to create a measure of the size of the demand shock, which is then scaled by the distance to the association's route networks. The local crime rate control is the log of the total count of all crimes recorded in police precincts linked to association route networks. Association controls are the number of members and number of routes. Standard errors are clustered at the Association level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Source: author's compilation based on data sources discussed in the text.

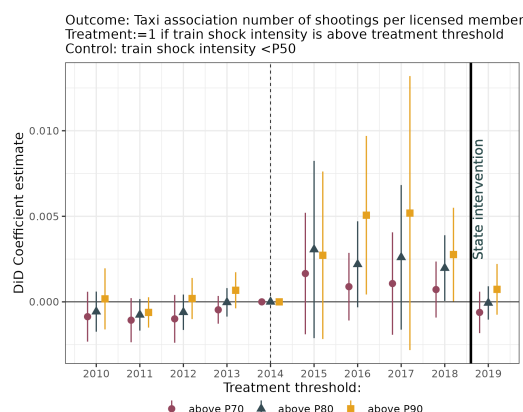
Robustness

I test the robustness of this result in several ways. The first is to test whether the results are driven by changes in local crime rates that have spillovers onto taxi violence. It should be stressed that the outcome variable in the main specification is constituted only of clear cases of taxi-industry-related violence, and should not be construed as incidents of violence in general.³⁷ However, using police-precinct-level crime statistics, Figure A4 presents the results of a Poisson regression where the unit of observation is the total count of crime in a police precinct in a given year and the treatment variable is equal to 1 if the precinct contains a train station.

These plots suggest that local crime rates declined in areas affected by the train shock. I also include a control for local crime rates directly in the main association-level regression by linking association route nodes to police precincts and taking the average crime rate across route nodes. Columns (2) and (4) of Table 3 include the log of the total crime rate at the association level as a direct control. Columns (3) and (4) include the interaction of local crime rates with

³⁷ For example, incidents where taxi drivers are killed in robberies do not form part of the shooting counts that constitute the outcome variable.

Figure 6: DiD: estimated change in shooting rates for associations most exposed to the positive demand shock



Note: the outcome variable is the number of shootings divided by the number of registered members of each association. The treatment is based on cut-offs of the continuous measure described in Equation 4. Specifically: the line-specific 2015–20 average rail operationality is subtracted from the 2010–15 average to create a measure of the size of the demand shock, which is then scaled by the distance to the association’s route networks. Association and year fixed effects are included.

Source: author’s compilation based on data sources discussed in the text.

the post train shock variable, to account for the potential that the nature of the effect of crime on taxi violence might have changed after the train shock. These controls do not affect the significance or size of the coefficient of interest. Second, in Table A7, I present the results of a robustness test where instead of estimating the effects of the train shock on violence at the association level, I estimate the effect on violence at the location level. These results echo the main results, in that there is a significant increase in taxi-related violence in areas containing train stations.

Third, in table A1 I re-estimate the main specification, sequentially dropping observations from any post-treatment year. In all cases the result remains statistically significant. Fourth, Table A8 shows that the estimated DiD coefficient remains significant irrespective of the treatment threshold and that the size of the coefficient estimate grows as the threshold is set closer to the maximum of the train exposure measure. Table A2 presents the estimation of the main result while changing the measurement of violence in the outcome variable. Finally, in Table A5 I run a placebo regression using the distance of association route networks to non-commuter train stations that were also affected by the decline of the rail agency (increased vandalism and declines in train services), but are not used by daily commuters. In this case I find a null effect. The results from the openings of shopping malls, presented in Table A11, also suggest that violence increases following positive demand shocks. Using the De Chaisemartin and d’Haultfoeuille (2024) estimator, the results in column (2) show that areas where new shopping malls open experience an increase in taxi-related violence. The dynamic results presented in Figure A5 suggest that violence increases in the year after a mall opens.

6.3 Heterogeneity by baseline characteristics

I present two sets of results documenting variation in the effect of the decline in commuter rail on violence in the taxi sector, by pre-shock differences across the most-affected taxi associations. The first of these concerns local state capacity, while the second concerns the structure of markets the associations were operating in.

Local police capacity

In order to better understand the role of the state in mitigating violence due to economic shocks in the taxi sector, I use crime statistics to create an association-level measure of police performance in the period leading up to the train shock. While crime statistics have long-established limitations (Langworthy 1999), in this case it is one of few sources of data that are available at a small area level that would allow for a heterogeneity analysis within cities. The South African Police Service publishes yearly disaggregated crime statistics at the police precinct level (SAPS 2020).³⁸ I create two precinct-level measures of baseline policing performance for the 2010–14 period. The first is the percentage change in total crime within the precinct over the period.³⁹ This complements changes in crime rates with changes in police arrests. Specifically, it is a binary variable equal to 1 if crime rates fell and the police-detected share of all crime increased. Following this, I link association route nodes to police precincts and create association-level averages of these measures across all their route nodes. I estimate a triple-differences specification, where I interact a baseline measure of police capacity with exposure to the demand shock:

$$\begin{aligned}
 v_{a,t} = & \beta_1(train_routes_a \times Post_train_shock_t \times baseline_police_capacity_a) \\
 & + \beta_2(train_routes_a \times Post_train_shock_t) \\
 & + \beta_3(train_routes_a \times baseline_police_capacity_a) \\
 & + \beta_4(Post_train_shock_t \times baseline_police_capacity_a) \\
 & + \gamma_a + \gamma_t + \varepsilon_{a,t}
 \end{aligned} \tag{8}$$

where *train_routes* is a binary treatment variable created by imposing a cut-off on ΔE . In this case, the coefficient of interest is β_1 , which is the additional effect of having routes in better-performing police precincts on violence for associations exposed to the rail shock after the decline in rail.

³⁸ An overview of the data is presented in Appendix C6

³⁹ Specifically: $Baseline\ police\ performance_p = \left(\frac{Total\ crime_{2010,2012} - Total\ crime_{p,2013,2014}}{Total\ crime_{p,2010,2012}} \right) \times 100$, where the *Total crime* terms are an average over the multiple years, in order to reduce the influence of outliers. I then standardize this measure.

Table 4 presents the results of this estimation. I find that associations exposed to the shock with route networks in areas with better-performing police experience lower increases in violence in the aftermath of the train shock. This suggests that local state capacity plays an important role in whether taxi-related violence emerges following demand shocks. Well-performing police departments serve as a deterrent to taxi-related violence.

Table 4: DiD results: heterogeneity by baseline police performance

	Shooting rate		
	(1)	(2)	(3)
Post train shock \times train routes	0.0023** (0.0009)	0.0015** (0.0007)	0.0049** (0.0019)
Post train shock \times baseline police performance A		0.0005** (0.0002)	
Post train shock \times train routes \times baseline police performance A		-0.0073*** (0.0024)	
Post train shock \times baseline police performance B			0.0004 (0.0003)
Post train shock \times train routes \times baseline police performance B			-0.0079** (0.0033)
Observations	3,110	3,110	3,110
R ²	0.16843	0.18411	0.17442
Within R ²	0.00767	0.02637	0.01481
Association fixed effects	✓	✓	✓
Year fixed effects	✓	✓	✓

Note: 'Train routes' is equal to 1 if an association is above the 80th percentile of the continuous train shock exposure measure defined in Section 5.1, and 0 if it is below the median. 'Baseline police performance (A and B)' are two measures of police performance on association route networks, based on changes in crime rates and police arrests in the pre-train shock period, described in Section 6.3. The outcome variable is the association-level shooting rate in a given year. Standard errors are clustered at the association level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Source: author's compilation based on data sources discussed in the text.

Market structure in the minibus taxi industry

Using the three market structure measures developed in Section 3.3, I present results on how the results vary by baseline market structure. Specifically, I measure the average distance to rival route networks, the distance-weighted number of other associations, and the distance-weighted number of other drivers in 2014. I then estimate a similar triple-differences specification, where I interact these measures of baseline market structure with the DiD interaction term to study the effect of the train shock on violence across affected associations with different baseline market structure.

Table 5 presents these results and suggests that the results in the main estimation equation are largely driven by associations in more competitive markets. The coefficient on the triple interaction term with distance to rival route networks is negative and significant at the 5% level, while the coefficient on the number of rival associations at baseline is positive and significant at the 5% level. This result suggests that the design of the market can accentuate or mitigate

the emergence of violence due to unexpected shocks, and highlights the importance of the transport planning process.

Table 5: DiD results: heterogeneity by baseline market structure

	Shooting rate			
	(1)	(2)	(3)	(4)
Post train shock × train routes	0.0023** (0.0009)	0.0353** (0.0170)	−0.0036* (0.0020)	−0.0170* (0.0090)
Post train shock × distance to rival associations		−0.0005 (0.0004)		
Post train shock × train routes × distance to rival associations		−0.0063** (0.0031)		
Post train shock × rival associations/distance			0.0004 (0.0003)	
Post train shock × train routes × rival associations/distance			0.0038** (0.0019)	
Post train shock × rival drivers/distance				0.0004 (0.0003)
Post train shock × train routes × rival drivers/distance				0.0034* (0.0017)
Observations	3,110	3,110	3,080	3,080
R ²	0.16843	0.17129	0.17373	0.17319
Within R ²	0.00767	0.01107	0.01403	0.01338
Association fixed effects	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓

Note: ‘Train routes’ is equal to 1 if an association is above the 80th percentile of the continuous train shock exposure measure defined in Section 5.1, and 0 if it is below the median. Three measures of baseline market structure are presented. The first is the distance to rival route networks. The second is the distance-weighted count of other associations. The third is the distance-weighted count of other associations’ drivers. The construction of these variables is presented in Section 3.3. All three measures are in logs. The outcome variable is an association-level shooting rate: the count of shootings in a given year, divided by the association size. Standard errors are clustered at the association level. *p<0.1, **p<0.05, ***p<0.01.

Source: author’s compilation based on data sources discussed in the text.

6.4 Mechanism

The main mechanism I propose for why violence occurred in the aftermath of a positive demand shock is excessive market entry, which increased scarcity. First, I document that associations most affected by the shock significantly increased the numbers of vehicles operating on their routes. These results are presented in Table 6, with the accompanying event plots in Figure 7. These results show that the number of registered vehicles per registered route increased significantly for associations most affected by the rail shock. Associations above the 80th percentile of the continuous train shock exposure measure increased the number of vehicles on their registered routes by roughly 11% in the aftermath of the train shock (6 column (2)).

Table 6: DiD: estimated change in vehicles per route for associations most exposed to decline in train operationality

Dependent variable:	log(vehicles per route)		
	(1)	(2)	(3)
Above P70 of train intensity measure \times post train shock	0.0978** (0.0387)		
Above P80 \times post train shock		0.1089** (0.0451)	
Above P90 \times post train shock			0.1216** (0.0556)
Association fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Observations	3,527	3,102	2,670
R ²	0.29773	0.30002	0.29395
Within R ²	0.00786	0.00825	0.00704

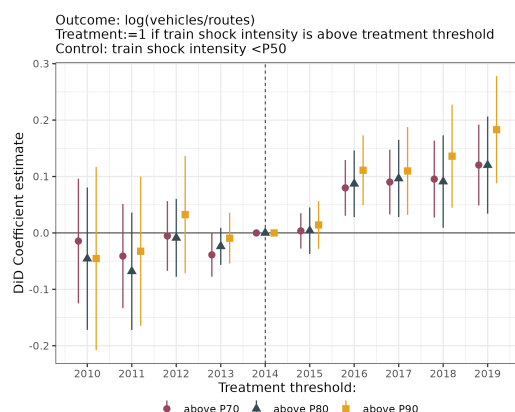
Note: this table presents DiD coefficient estimates from an OLS regression, for each year around the start of the decline in train operationality in 2015. The outcome variable is the log of the number of registered vehicles divided by the number of registered routes at the association–year level. Both variables are centred on their 2010–14 means at the association level. Treatment is based on binary cut-offs of the continuous measure of the size of the demand shock described in Equation 4. The line-specific 2015–20 average rail operationality is subtracted from the 2010–15 average to create a measure of the size of the demand shock, which is then scaled by the distance to the association's route networks. Association and year fixed effects included. Clustered standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Source: author's compilation based on data sources discussed in the text.

The fact that associations increase the number of vehicles operating on their routes after being exposed to a positive demand shock is to be expected. In order to understand the magnitude of this increase in supply in relation to the change in demand, I create a measure of the potential number of commuters per vehicle for each taxi association. I make use of two household travel surveys, conducted in 2013 and 2020 and described in Appendix C5. With the data from these surveys, which are designed to be representative of the population at the provincial level, I first document a decrease in the total number of commuters per licensed taxi owner in Gauteng Province as a whole, shown in Table A6.

Using these survey data, I create a measure of the potential number of commuters served by each association, based on the number of commuters in the transport analysis zones (TAZ) their routes start and end in. I assume taxi associations partition commuters in a certain TAZ in proportion to their number of vehicles. For example, if one association has 20 members and another 10, in a market where there are 60 commuters, I assume the former association serves 40 commuters and the latter 20. Using this measure, I calculate for each taxi association the number of potential commuters per registered vehicle in 2013 and 2020. I then re-estimate the baseline estimating equation with the number of commuters per taxi owner at the association level as the outcome variable. The results of this exercise are shown in Table 7, and suggest that associations most affected by the decline in commuter rail experienced a decline in the number of customers per registered driver.

Figure 7: DiD: estimated change in registered vehicles per registered route for associations most exposed to the positive demand shock



Note: this figure presents DiD coefficient estimates from an OLS regression, for each year around the start of the decline in train operability in 2015. The outcome variable is the log of the number of registered vehicles divided by the number of registered routes at the association–year level. Both variables are centred on their 2010–14 means at the association level. Treatment is based on cut-offs of the continuous measure described in Equation 4. The line-specific 2015–20 average rail operability is subtracted from the 2010–15 average to create a measure of the size of the demand shock, which is then scaled by the distance to the association’s route networks. Association and year fixed effects included.

Source: author’s compilation based on data sources discussed in the text.

Table 7: DiD: estimation of change in number of commuters per driver for associations most affected by rail shock

	Commuters per driver (1)
Post train shock × train routes	–26.51*** (6.052)
Observations	884
R ²	0.96148
Within R ²	0.06264
Time-variant association control	✓
Association fixed effects	✓
Year fixed effects	✓

Note: this table presents an estimation of the change in the number of commuters per registered driver for taxi associations most affected by the decline in commuter rail. The number of commuters is estimated from the 2013 and 2020 household travel surveys (Statistics South Africa 2014, 2021a).

Source: author’s compilation based on data sources discussed in the text.

A decline in the number of commuters per driver does not necessarily indicate an increase in scarcity if associations could cooperate to exploit the decrease in elasticity of demand for taxis to raise prices. While I do not have access to an association-level panel of prices, I estimate whether there is a price response using the same household survey data by running a regression at the individual level and using variation across TAZ. The central objective of this exercise is to understand whether taxi associations in baseline (2013) rail commuting areas were able to raise prices more rapidly than associations in non-rail commuting areas. I create an arbitrary cut-off for rail commuting areas as those TAZs where the number of commuters by train repre-

sent at least one-quarter (0.25) of commuters by taxi in 2013. I control for commuting time by minibus taxi (in minutes per month), and in total, as well as the number of days spent commuting. I estimate the increase in commuting cost by minibus taxi, taking into account changes in commuting times that may have resulted from the shock.⁴⁰ In a first specification I estimate the following equation:

$$\begin{aligned} \text{taxi cost}_{i,z,t} = & \beta_1 \text{rail area}_{i,z} + \beta_2 \text{post railshock}_t + \beta_3 \text{rail area}_{i,z} \times \beta_4 \text{post_railshock}_t \\ & + \beta_5 \text{taxi time}_{i,z,t} + \gamma_t + \delta_z + X_i + \varepsilon_{i,z,t} \end{aligned}$$

where $\text{taxi cost}_{i,z,t}$ is the real monthly cost commuter i in transport analysis zone z in year t pays in taxi fares, $\text{rail area}_{i,z}$ is a marker for whether the area had a high rail commuting share in 2013, $\text{taxi time}_{i,z,t}$ is the number of minutes individual i spent commuting by minibus taxi, γ_t is a year fixed effect, and δ_z is a TAZ fixed effect. X_i is a vector of individual controls (sex, age, age squared, population group).

I document in Table 8 (column (2)) that real all-mode monthly commuting costs increased by roughly 9% in areas most affected by the decline in commuter rail. The decline in commuter rail imposed substantial costs on commuters who had historically been dependent on the rail network. However, when restricting the sample only to commuters by minibus taxi, and restricting the outcome variable only to taxi fares, and controlling for changes in commuting times (column (5)), I do not find any significant changes in the per-minute price of commuting by minibus taxi in baseline rail commuting areas compared to non-rail commuting areas following the decline in commuter rail.

Taken together, the decline in commuters per driver, in the absence of any marked increase in prices, suggests that the real profits of drivers in areas most affected by the decline in commuter rail decreased over the study period, providing a potential explanation for the increase in violence. The results presented in this paper are thus also in line with results from illegal markets. Castillo et al. (2020), for example, find that exogenous increases in drug revenue in Mexico lead to increases in violence among cartels involved in the drug trade, while Kronick (2020) finds similar results, showing that increased profits increased the incidence of violence along a major drug trafficking route in Venezuela.

⁴⁰ The total time spent commuting is recorded consistently in both the 2013 and 2020 data. Unfortunately, commuting destinations at the sub-place level (with which origin–destination distances can be calculated) are not recorded in the 2020 data. For this reason, I use within-mode commuting time as a measure of distance, or quantities of transport.

Table 8: OLS regression: changes in real commuting costs in rail-commuting areas following the decline in commuter rail

	log(real all mode commuting cost)			log(real taxi commuting cost)		
	(1)	(2)	(3)	(4)	(5)	(6)
Baseline train commuting area \times post train shock	0.1057*** (0.0346)	0.0859** (0.0335)	0.0591* (0.0335)	0.1363*** (0.0391)	0.0554 (0.0381)	0.0430 (0.0392)
log(total monthly commuting time)		0.2151*** (0.0109)	0.1664*** (0.0136)			
Post train shock \times log(total monthly commuting time)			0.0949*** (0.0210)			
log(monthly commuting time by taxi)					0.2856*** (0.0149)	0.2689*** (0.0164)
Post train shock \times log(monthly commuting time by taxi)						0.0286 (0.0274)
Observations	17,936	17,931	17,931	10,815	10,157	10,157
R ²	0.23103	0.28568	0.28895	0.22057	0.33728	0.33798
Within R ²	0.14567	0.20643	0.21006	0.12322	0.24767	0.24847
Individual controls	✓	✓	✓	✓	✓	✓
TAZ fixed effects	✓	✓	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓	✓	✓

Note: this table presents changes in commuting costs of individuals located in TAZ with a high baseline (2013) share of commuting by train, following the decline in commuter rail. Train commuting districts are defined as districts where commuters by rail represent at least 25% of commuters by minibuss taxi. The sample is restricted to work and educational commuters for whom times and costs are consistently recorded in both NHTS years. The lowest level of representativeness in the NHTS data is the district municipality ($N = 52$); however, the data are also reported at a lower TAZ level ($N = 384$). Here, the sample is restricted to TAZ with more than 250 observations in each survey year, and located in the five provinces that form part of the analysis ($N = 98$ TAZ). Ten of these TAZ are classified as baseline train commuting areas. Survey weights are included. Individual controls include population group, gender, age, age squared, and days spent commuting to educational and work locations. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Source: author's compilation based on the 2013 and 2020 National Household Travel survey (Statistics South Africa 2014, 2021a).

Finally, Table A9 presents the results of the long-difference specification. District fixed effects are included throughout to account for some of the cross-sectional variation in demand. In column (1), the change in the association-level shooting rate is regressed on the change in the absolute treatment intensity. Associations that experienced larger declines in rail operationality close to their networks experienced larger increases in their rate of involvement in shooting incidents. In column (4), the positive coefficient on the interaction of the change in train operationality and the change in market structure suggests shooting rates increased for associations where large changes in train operationality coincided with increases in the number of competitor drivers.

7 Discussion

Countries often struggle with the regulation of private commuter transport operators. In Colombia, for example, the period between the 1960s and 1990s came to be known as the *guerra del centavo* or ‘penny war’, when an oversupply of poorly regulated private buses led to dangerous and sometimes violent competition among bus drivers (Rodriguez et al. 2016; World Bank 2014). Similar problems exist in Peru (Bielich 2007). In Nairobi, Katumanga (2005) documents the domination of the *matatu* minibus industry by violent vigilante groups that compete for control over lucrative routes.⁴¹ Agbibo (2022) documents similar behaviour in the control over *danfo* minibus routes in Lagos. A common feature across most of these cases is the dysfunction of state institutions to effectively regulate entry and behaviour in these markets, which leads to the emergence of private groups that take on the role of regulator. While such groups might be thought to internalize the costs of excessive entry and dangerous competition, this paper presents evidence to the contrary, with negative social consequences.

The following paragraphs offer a brief policy discussion. The clearest and most pressing goal should be for existing institutions to function more effectively. Better policing and better transport planning are central to reducing violence. The discussion below presents different ways of thinking about how regulation interacts with violence in the taxi sector.

Excessive entry. The purpose of the operating licence system is to regulate market entry. The transport planning process that should limit excessive entry does not appear to function well. Conversations with transport officials suggest there is significant corruption in the process of granting of operating licences, effectively rendering the transport planning process ineffective. More and more drivers competing for the same pool of commuters is a key cause of the vio-

⁴¹ ‘The control of matatu routes became a battleground as Mungiki and Kamjeshi began fighting for physical occupation of the routes and termini: “between July and September 2001 about 15 people were hacked to death consequent to engagements between Mungiki and Kamjeshi gangs over control of Dandora route” (Daily Nation, 20 February 2002)’ Katumanga (2005).

lence in the minibus taxi industry. Historically, the severe violence in the late 1980s, when the industry was initially deregulated (with similarities in Nairobi and Bogota), emerged under similar conditions. The democratic state attempted to re-regulate the industry, with the introduction of the NLTIS and formalization of the operating licence system. However, the 2010s saw effective deregulation when the state was unable to effectively manage and enforce regulation. Second, transport planners have no knowledge of the intensive margin allocations of members (taxi operators) to each route—a decision that is entirely up to the association, and out of the frame of information possessed by the Department of Transport. This severely limits the transport planning process, as it is the effective supply of vehicles to routes. More effective use of existing data that the Department of Transport collects can help make transport planning more effective.

Property rights and dispute resolution. The state has little capacity to immediately detect illegal incursions of taxi drivers onto the routes of others, making the enforcement of property rights very challenging. If the state does not exercise this function, taxi associations take on the role. This paper suggests that they do not act in a socially optimal manner. Developing effective means by which deviations (from licensed routes) can be detected and rapidly sanctioned, while taking more severe action on violence, raises the cost of violence and raises the benefit of other mechanisms for dispute resolution. Regularizing taxi stops may render property rights easier to define and enforce. While it may take time to develop more effective means of monitoring whether associations stick to licensed routes, creating more effective mechanisms for dispute resolution can be a good second-best measure in the short term.

Information frictions. Markets that govern themselves can function well under certain conditions. When participants interact repeatedly, are patient about the future, and are able to obtain information about their rival's behaviour, the threat of severe punishment can be enough for mutually beneficial cooperation to be sustained (Dixit 2003). However, these conditions become difficult to satisfy in markets where there is constant new entry, and where it is possible to deviate (operate on a rival's route) without being detected. It is possible that deriving ways in which information can flow more freely and by which deviation can be detected immediately will reduce the incentive for firms to deviate and reduce violent competition. Many taxi associations in Johannesburg now employ 'squad cars' and individuals are posted on street corners along their routes with the express role of detecting and retaliating against incursions onto their routes. If such private institutions can be adopted in a less costly manner, it may be an effective way for associations to internalize the costs of deviation and promote more peaceful interactions.

Pricing. Price setting in the taxi industry lacks meaningful government oversight despite the industry serving as the de-facto mode of public transport (80% of public transport users rely on minibus taxis (Statistics South Africa 2021b)). At present, transport departments do not main-

tain a database of the current prices of minibus taxi routes. More transparent price information may be a good starting point to improve the transport planning process. In other contexts, centralization of the price-setting process—for example through state-mandated or subsidized per-kilometre prices which leave a reasonable margin for profits—have been developed as ways to mitigate against price gouging by private actors, to improve the information problem by providing commuters and firms with one credible information source, and to reduce the costs of the inefficiency of the two-sided queuing problem documented by Conwell (2024). Given the importance of commuting costs for economic outcomes, and given the high commuting costs observed in South Africa (Kerr 2017), such measures should be considered.

The results presented here have several caveats that should be made clear. First, the NLTIS data presented here provide an important starting point from which the dynamics of this industry can be explored. However, much of the activity in the industry is unobserved by the state, which maintains the data. Second, the administrative data presented here exclude two important provinces, KwaZulu-Natal and the Western Cape, both of which are important drivers of overall violence in the taxi industry and are large population centres. Third, while I have done my best to capture all articles relating to taxi violence in a systematic and replicable manner, it is likely that more violence occurs than is documented by the news media. Finally, violence in the minibus taxi industry is a complex social phenomenon. Economic motivations can only explain a subset of violence, and it is not the intention of this paper to imply otherwise. However, the results in this paper do suggest that there exist important economic drivers of violence, and understanding these economic drivers may be useful in reducing its incidence.

8 Summary and conclusion

This paper studies the emergence of violence among registered private transport firms operating in a legal market in the absence of well-functioning state institutions. I present a large set of new data on the minibus taxi industry and empirically study its structure and responses to a large positive demand shock sparked by a decline in the commuter rail network. I find that this shock led to an increase in violence experienced by the most-affected firms and document the economic mechanisms that may be driving these results, namely excessive market entry. Taken together, these results stress that markets functioning in the absence of well-functioning state institutions can do so in ways that increase incentives for violence, despite its costs.

While a large literature has studied the emergence of violence between criminal networks competing in markets for illegal goods, this paper is one of a small set of studies in economics that shows how violence can emerge as a result of economic competition in a legal sector that operates in the absence of well-functioning state institutions. This violence has come to constitute a large social cost and understanding and addressing economic and regulatory drivers of

this pervasive problem is an important step in developing ways to counteract it. Finally, this article contributes to our understanding of the firms operating on the spectrum between formality and informality. In many rapidly urbanizing developing country contexts, states have the capacity to regulate, but limited capacity to enforce regulation, leading to the emergence of private institutions that govern their relations. This paper documents severe private and social costs associated with the reliance on such modes of self-governance in certain markets.

References

- Agbiboa, D.E. (2022). 'Nigeria's Transport Mafia'. In *They Eat Our Sweat: Transport Labor, Corruption, and Everyday Survival in Urban Nigeria*. Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780198861546.001.0001>
- Anderson, D.M. (2002). 'Vigilantes, Violence and the Politics of Public Order in Kenya'. *African Affairs*, 101(405): 531–55. <https://doi.org/10.1093/afraf/101.405.531>
- Antrobus, L., and A. Kerr (2019). 'The Labour Market for Minibus Taxi Drivers in South Africa'. Technical Report. Cape Town: SALDRU.
- Arellano-Bover, J., L. Guiso, R. Macchiavello, D.J. Marchetti, and M. Prem (2024). 'Mafias and Firms'. Discussion Paper 16893. Bonn: IZA.
- Banerjee, A., G. Fischer, D. Karlan, M. Lowe, and B.N. Roth (2024). 'Do Collusive Norms Maximize Profits? Evidence from a Vegetable Market Experiment in India'. Working Paper 30360. Cambridge, MA: NBER.
- Bank, L.J. (1990). 'The Making of the Qwaqwa 'Mafia'? Patronage and Protection in the Migrant Taxi Business'. *African Studies*, 49(1): 71–93. <https://doi.org/10.1080/00020189008707718>
- Bazzi, S., and C. Blattman (2014). 'Economic Shocks and Conflict: Evidence from Commodity Prices'. *American Economic Journal: Macroeconomics*, 6(4): 1–38. <https://doi.org/10.1257/mac.6.4.1>
- Becker, G.S. (1968). 'Crime and Punishment: An Economic Approach'. *Journal of Political Economy*, 76(2): 169–217. <https://doi.org/10.1086/259394>
- Bergquist, L.F., and M. Dinerstein (2020). 'Competition and Entry in Agricultural Markets: Experimental Evidence from Kenya'. *American Economic Review*, 110(12): 3705–47. <https://doi.org/10.1257/aer.20171397>
- Berman, N., and M. Couttenier (2015). 'External Shocks, Internal Shots: The Geography of Civil Conflicts'. *Review of Economics and Statistics*, 97(4): 758–76. https://doi.org/10.1162/REST_a_00521
- Berman, N., M. Couttenier, D. Rohner, and M. Thoenig (2017). 'This Mine Is Mine! How Minerals Fuel Conflicts in Africa'. *American Economic Review*, 107(6): 1564–610. <https://doi.org/10.1257/aer.20150774>

- Bernstein, L. (2013). 'Opting Out of the Legal System: Extralegal Contractual Relations in the Diamond Industry'. In R. Epstein (ed.), *Contract-Freedom and Restraint*. London: Routledge.
- Bielich, C. (2007). *La Guerra Del Centavo. Una Mirada Actual Al Transporte Público en Lima Metropolitana*. Lima: Instituto de Estudios Peruanos.
- Blattman, C., G. Duncan, B. Lessing, and S. Tobón (2021). 'Gang Rule: Understanding and Countering Criminal Governance'. Working Paper 28458. Cambridge, MA: NBER. <https://doi.org/10.3386/w28458>
- Breza, E., S. Kaur, and N. Krishnaswamy (2022). 'Coordination without Organization: Collective Labor Supply in Decentralized Spot Markets'. Working Paper 25880. Cambridge, MA: NBER.
- Brown, Z.Y., E. Montero, C. Schmidt-Padilla, and M.M. Sviatschi (2024). 'Market Structure and Extortion: Evidence from 50,000 Extortion Payments'. *Review of Economic Studies*. <https://doi.org/10.1093/restud/rdae057>
- Callaway, B., A. Goodman-Bacon, and P.H. Sant'Anna (2024). 'Difference-in-Differences with a Continuous Treatment'. Working Paper 32117. Cambridge, MA: NBER. <https://doi.org/10.3386/w32117>
- Castillo, J.C., and D. Kronick (2020). 'The Logic of Violence in Drug War'. *American Political Science Review*, 114(3): 874–87. <https://doi.org/10.1017/S0003055420000246>
- 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): 269–86. https://doi.org/10.1162/rest_a_00801
- CCSA (Competition Commission of South Africa) (2021). 'Market Inquiry into the Land Based Public Passenger Transport Sector'. Report. Available at: <http://www.compcom.co.za/wp-content/uploads/2021/04/PTMI-Non-Confidential.pdf> (accessed 15 April 2025).
- Citizen TV Kenya (2018). 'GANGSTER SACCO: Unearthing Rot in Matatu Industry'. Citizen TV Kenya. Available at: <https://www.youtube.com/watch?v=vSYzwrP2g9U> (accessed 15 April 2025).
- Conwell, L. (2024). 'Privatized Provision of Public Transit'. SSRN. Available at: <https://ssrn.com/abstract=4489658> (accessed 15 April 2025).
- Dal Bó, E., and P. Dal Bó (2011). 'Workers, Warriors, and Criminals: Social Conflict in General Equilibrium'. *Journal of the European Economic Association*, 9(4): 646–77. <https://doi.org/10.1111/j.1542-4774.2011.01025.x>
- De Chaisemartin, C., and X. d'Haultfoeulle (2024). 'Difference-in-Differences Estimators of Intertemporal Treatment Effects'. *Review of Economics and Statistics*. https://doi.org/10.1162/rest_a_01414
- Dell, M. (2015). 'Trafficking Networks and the Mexican Drug War'. *American Economic Review*, 105(6): 1738–79. <https://doi.org/10.1257/aer.20121637>

- Dell, M., B. Feigenberg, and K. Teshima (2019). 'The Violent Consequences of Trade-Induced Worker Displacement in Mexico'. *American Economic Review: Insights*, 1(1): 43–58. <https://doi.org/10.1257/aeri.20180063>
- Dixit, A. (2003). 'On Modes of Economic Governance'. *Econometrica*, 71(2): 449–81. <https://doi.org/10.1111/1468-0262.00415>
- Dixit, A.K. (2004). *Lawlessness and Economics: Alternative Modes of Governance*, volume 1. Princeton, NJ: Princeton University Press.
- Draca, M., T. Koutmeridis, and S. Machin (2019). 'The Changing Returns to Crime: Do Criminals Respond to Prices?' *The Review of Economic Studies*, 86(3): 1228–57. <https://doi.org/10.1093/restud/rdy004>
- Dube, A., O. Dube, and O. García-Ponce (2013). 'Cross-Border Spillover: US Gun Laws and Violence in Mexico'. *American Political Science Review*, 107(3): 397–417. <https://doi.org/10.1017/S0003055413000178>
- 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): 1181–224. <https://doi.org/10.1111/jeea.12172>
- Dugard, J. (2001). *From Low-Intensity War to Mafia War: Taxi Violence in South Africa, 1987–2000*, volume 4. Cape Town: Centre for the Study of Violence and Reconciliation.
- Durán-Martínez, A. (2015). 'To Kill and Tell? State Power, Criminal Competition, and Drug Violence'. *Journal of Conflict Resolution*, 59(8): 1377–402. <https://doi.org/10.1177/0022002715587047>
- Fourchard, L. (2023). 'Expanding Profit and Power: The National Union of Road Transport Workers in Nigeria'. *Canadian Journal of Development Studies/Revue canadienne d'études du développement*, 44(1): 97–112. <https://doi.org/10.1080/02255189.2022.2132924>
- Fourchard, L. (2024). 'In the Shadow of the State: The Making of Garage Laws in Lagos'. *Swiss Journal of Sociocultural Anthropology*, 29(1): 85–101. <https://doi.org/10.36950/sjsca.2023.29.8905>
- Franklin, S. (2018). 'Location, Search Costs and Youth Unemployment: Experimental Evidence from Transport Subsidies'. *The Economic Journal*, 128(614): 2353–79. <https://doi.org/10.1111/eoj.12509>
- Gaduh, A., T. Gračner, and A.D. Rothenberg (2022). 'Life in the Slow Lane: Unintended Consequences of Public Transit in Jakarta'. *Journal of Urban Economics*, 128: 103411. <https://doi.org/10.1016/j.jue.2021.103411>
- Gambetta, D. (2006). 'Trust's Odd Ways'. In A. Hylland, J. Elster, O. Gjelsvik, and K. O. Moene (eds.), *Understanding Choice, Explaining Behaviour. Essays in Honour of Ole-Jorgen Skog*. Oslo: Academic Press Oslo.

- Gambetta, D. (2009). *Codes of the Underworld: How Criminals Communicate*. Princeton, NJ: Princeton University Press. <https://doi.org/10.1515/9781400833610>
- Garfinkel, M.R., and S. Skaperdas (2007). 'Economics of Conflict: An Overview'. *Handbook of Defense Economics*, 2: 649–709. [https://doi.org/10.1016/S1574-0013\(06\)02022-9](https://doi.org/10.1016/S1574-0013(06)02022-9)
- GDRT (Gauteng Department of Roads and Transport) (2021). 'Final Report of Commission of Inquiry: Minibus Taxi-Type Service Violence'. Report. Available at: <https://tinyurl.com/nryvty62> (accessed 15 April 2025).
- Greif, A. (1993). 'Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders' Coalition'. *American Economic Review*, 83(3): 525–48.
- Harris, C.D. (1954). 'The Market as a Factor in the Localization of Industry in the United States'. *Annals of the Association of American Geographers*, 44(4): 315–48.
- Herrera, J.S., and C.B. Martinez-Alvarez (2022). 'Diversifying Violence: Mining, Export-Agriculture, and Criminal Governance in Mexico'. *World Development*, 151: 105769. <https://doi.org/10.1016/j.worlddev.2021.105769>
- Katumanga, M. (2005). 'A City Under Siege: Banditry & Modes of Accumulation in Nairobi, 1991–2004'. *Review of African Political Economy*, 32(106): 505–20. <https://doi.org/10.1080/03056240500466981>
- Kerr, A. (2017). 'Tax(i)ing the Poor? Commuting Costs in South African Cities'. *South African Journal of Economics*, 85(3): 321–40. <https://doi.org/10.1111/saje.12161>
- Kerr, A. (2018). *Minibus Taxis, Public Transport, and the Poor*. Washington, DC: World Bank. <https://doi.org/10.1596/30018>
- Khosa, M.M. (1992). 'Routes, Ranks and Rebels: Feuding in the Taxi Revolution'. *Journal of Southern African Studies*, 18(1): 232–51. <https://doi.org/10.1080/03057079208708312>
- Kronick, D. (2020). 'Profits and Violence in Illegal Markets: Evidence from Venezuela'. *Journal of Conflict Resolution*, 64(7–8): 1499–523. <https://doi.org/10.1177/0022002719898881>
- La Porta, R., and A. Shleifer (2014). 'Informality and Development'. *Journal of Economic Perspectives*, 28(3): 109–26. <https://doi.org/10.1257/jep.28.3.109>
- Lane, G., D. Schönholzer, and E. Kelley (2024). 'Information and Strategy in Lemon Markets: Improving Safety in Informal Transit'. Working Paper 192137. Washington, DC: World Bank.
- Langworthy, R.H. (1999). 'Measuring What Matters: A Policing Research Institute'. In *Measuring What Matters: Proceedings from the Policing Research Institute Meetings*. Washington, DC: National Institute of Justice.
- Le Moglie, M., and G. Sorrenti (2022). 'Revealing "Mafia Inc."? Financial Crisis, Organized Crime, and the Birth of New Enterprises'. *Review of Economics and Statistics*, 104(1): 142–56. https://doi.org/10.1162/rest_a_00942

- Mabogo, M. (2023). 'The SA Taxi Industry Is Worth Billions: Why Is It Paying so Little Tax?' *NEWS24 Business*. Opinion Article. Available at: <https://www.news24.com/fin24/opinion/opinion-the-sa-taxi-industry-is-worth-billions-why-is-it-paying-so-little-tax-20230920> (accessed 15 April 2025).
- Macchiavello, R. (2022). 'Relational Contracts and Development'. *Annual Review of Economics*, 14: 337–62. <https://doi.org/10.1146/annurev-economics-051420-110722>
- Macchiavello, R., and A. Morjaria (2021). 'Competition and Relational Contracts in the Rwanda Coffee Chain'. *Quarterly Journal of Economics*, 136(2): 1089–143. <https://doi.org/10.1093/qje/qjaa048>
- Macchiavello, R., and A. Morjaria (2022). 'Empirical Approaches to Trust and Relational Contracts'. *International Journal of Industrial Organization*, 84: 102849. <https://doi.org/10.1016/j.ijindorg.2022.102849>
- 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): 552–72. <https://doi.org/10.1017/S0003055419000856>
- McGuirk, E., and M. Burke (2020). 'The Economic Origins of Conflict in Africa'. *Journal of Political Economy*, 128(10): 3940–97. <https://doi.org/10.1086/709993>
- Mejia, D., and P. Restrepo (2013). 'Bushes and Bullets: Illegal Cocaine Markets and Violence in Colombia'. Working Paper 2013-53. Bogotá: Universidad de los Andes. <https://doi.org/10.2139/ssrn.2353979>
- MOF (Ministry of Finance) (2021). 'Minister of Finance Reply to National Assembly Question for Written Reply by Mr G. G. Hill-Lewis'. Question Number: 947.
- Ohnsorge, F., and S. Yu (eds.) (2022). *The Long Shadow of Informality: Challenges and Policies*. Washington, DC: World Bank. <https://doi.org/10.1596/35782>
- Ostrom, E. (1990). *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511807763>
- PressReader Inc. (2022). 'Pressreader: An endless Stream of Top News Stories'. Available at: <https://www.pressreader.com> (accessed 15 April 2025).
- Rodriguez, C., T. Peralta-Quiros, L.A. Guzman, and S.A.C. Reyes (2016). 'Bogota's Bus Reform Process: Accessibility & Affordability Effects, Lessons Learnt & Alternatives to Tackle Informal Services'. Working Paper. Washington, DC: World Bank.
- SABINET (2022). All SA Media (News Clippings). Available at: <https://discover.sabinet.co.za/search?Search=&ProductType=allmedianewsclippings> (accessed 15 April 2025).
- SACSC (South African Council of Shopping Centres) (2022). *2022 Shopping Centre Directory*. Sandton: SACSC.

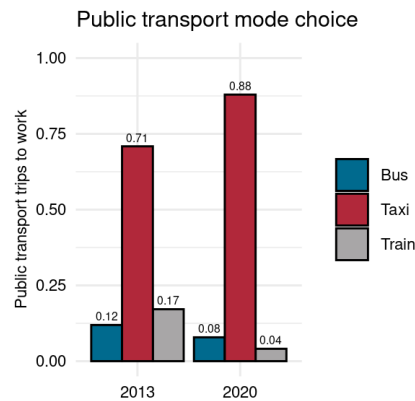
- SAPS (South African Police Service) (2020). 'Crime Statistics 2019/2020'. Available at: <https://www.saps.gov.za/services/crimestats.php> (accessed 15 April 2025).
- Sekhonyane, M., and J. Dugard (2004). 'A Violent Legacy: The Taxi Industry and Government at Loggerheads'. *South African Crime Quarterly*, 10.
- Skarbek, D. (2011). 'Governance and Prison Gangs'. *American Political Science Review*, 105(4): 702–16. <https://doi.org/10.1017/S0003055411000335>
- Statistics South Africa (2014). 'National Household Travel Survey 2013'. Revised version. Pretoria: SSA.
- Statistics South Africa (2017). 'Labour Market Dynamics Survey'. Version 1. Pretoria: SSA.
- Statistics South Africa (2018). 'Labour Market Dynamics Survey'. Version 1. Pretoria: SSA.
- Statistics South Africa (2019). 'Labour Market Dynamics Survey'. Version 1. Pretoria: SSA.
- Statistics South Africa (2021a). 'National Household Travel Survey 2020'. Pretoria: SSA.
- Statistics South Africa (2021b). 'National Household Travel Survey 2020, Statistical Release P0320'. Report. Available at: <http://www.statssa.gov.za/publications/P0320/P03202020.pdf> (accessed 15 April 2025).
- Sviatschi, M.M. (2022). 'Spreading Gangs: Exporting US Criminal Capital to El Salvador'. *American Economic Review*, 112(6): 1985–2024. <https://doi.org/10.1257/aer.20201540>
- Thomas, K. (2021). 'Murder by Contract: Targeted Killings in Eastern and Southern Africa'. Report. Geneva: Global Initiative Against Transnational Organized Crime (GITOC).
- Tsivanidis, N. (2019). 'Evaluating the Impact of Urban Transit Infrastructure: Evidence from Bogota's Transmilenio'. Mimeo.
- Ulyssea, G. (2018). 'Firms, Informality, and Development: Theory and Evidence from Brazil'. *American Economic Review*, 108(8): 2015–47. <https://doi.org/10.1257/aer.20141745>
- Ulyssea, G., M. Bobba, and L. Gadenne (2023). 'Informality'. *VoxDevLit*, 6(1).
- Williamson, O.E. (1979). 'Transaction-Cost Economics: The Governance of Contractual Relations'. *Journal of Law and Economics*, 22(2): 233–61. <https://doi.org/10.1086/466942>
- World Bank (2014). 'Colombia: Support to the National Urban Transport Program Project'. Report E2686. Washington, DC: World Bank.
- Zenou, Y. (2011). 'Search, Migration, and Urban Land Use: The Case of Transportation Policies'. *Journal of Development Economics*, 96(2): 174–87. <https://doi.org/10.1016/j.jdeveco.2010.11.001>

Appendices

Note that the appendices were not copy-edited.

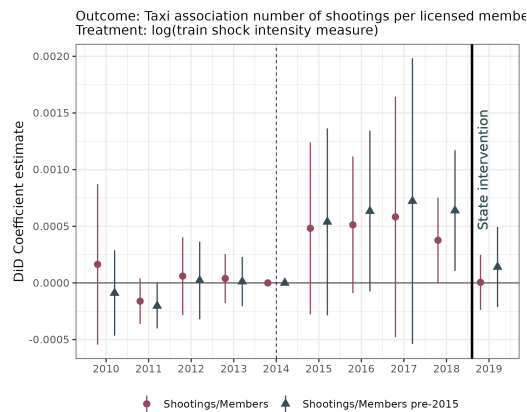
A Additional results

Figure A1: Commuter transport mode choice before and after the decline in commuter rail



Note and source: This figure presents survey responses from the National Household Travel Survey for residents in Gauteng, with roughly 2000 respondents in each year (Statistics South Africa 2014, 2021b).

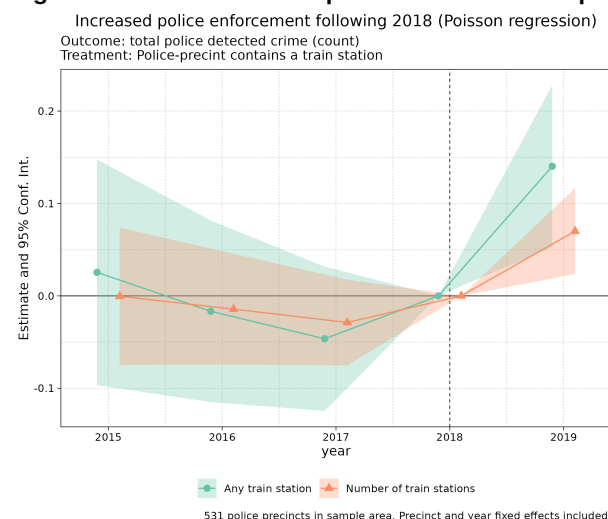
Figure A2: DiD: Estimated change in shooting rates and continuous treatment intensity



Note: The outcome variable is the number of shootings divided by the number of registered members. The treatment is the continuous measure of the size of exposure to the positive demand shock described in Section 4. Specifically: the line-specific 2015–2020 average rail operationality is subtracted from the 2010–2015 average to create a measure of the size of the demand shock, which is then scaled by the distance to association's route networks. Association and year fixed effects are included throughout.

Source: author's compilation based on data sources discussed in the text.

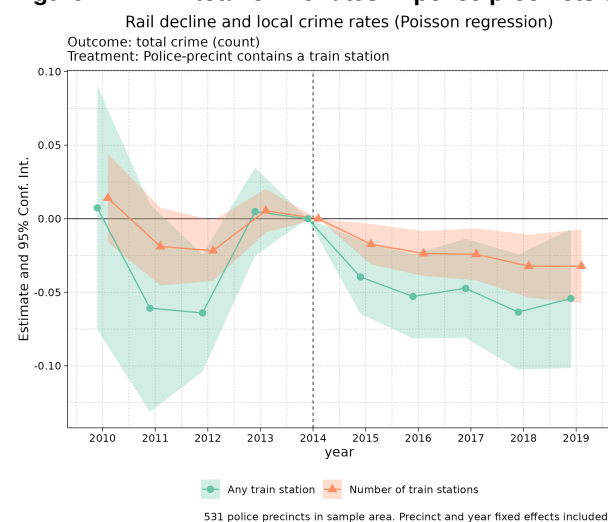
Figure A3: DiD: Increase in police enforcement in precincts containing train stations following 2018



Note: This figure documents a change in police action in areas affected by taxi violence in Gauteng following the high levels of violence in the 2015-2018 period. Appendix B2 provides further context. Notably, a Gauteng Taxi Summit was organized, where the first point in the MEC for transport's speech was a vow to act rapidly to quell the spiralling taxi violence in the province. Secondly, a Provincial Taxi Monitoring Committee was established. Finally, the premier of the province announced the formation of a Commission of Inquiry into Taxi Violence in 2019.

Source: author's compilation based on data sources discussed in the text.

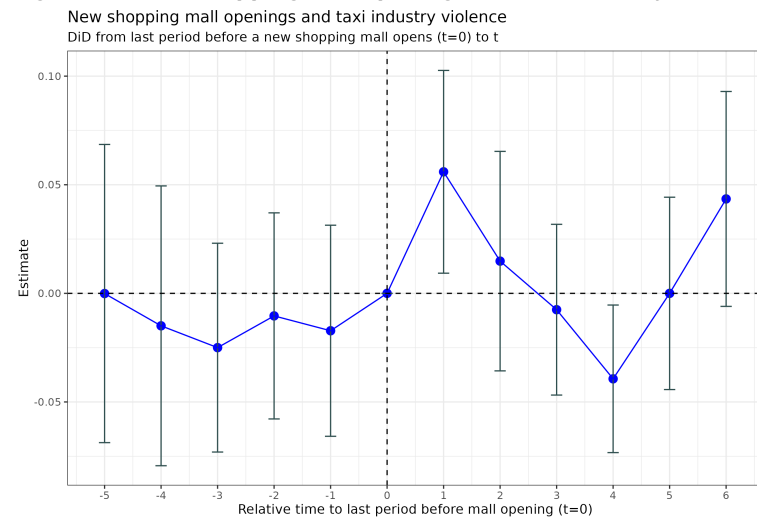
Figure A4: DiD: total crime rates in police precincts containing train stations



Note: In this police-precinct-level regression, treatment is equal to 1 if there is at least one train station located within the precinct bounds. The intensive-margin treatment measure is an inverse hyperbolic sine transformation of the number of train stations.

Source: author's compilation based on data sources discussed in the text.

Figure A5: New shopping mall openings and taxi industry violence



Source: author's compilation based on data sources discussed in the text.

Table A1: DiD results: Dropping observations from any one post-train-shock year

	Shooting rate: Number of shootings/Number of members				
	(1)	(2)	(3)	(4)	(5)
	Drop 2015	Drop 2016	Drop 2017	Drop 2018	Drop 2019
Post train shock × Train routes	0.0020** (0.0009)	0.0022** (0.0010)	0.0021** (0.0009)	0.0023** (0.0011)	0.0028** (0.0011)
Observations	2,796	2,797	2,796	2,795	2,796
Outcome mean	0.0008	0.0008	0.0008	0.0008	0.0008
R ²	0.20692	0.15124	0.17125	0.16577	0.17497
Within R ²	0.01002	0.00703	0.00812	0.00743	0.01044
Association fixed effects	✓	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓	✓
Local crime rate control	✓	✓	✓	✓	✓

Note: 'Train routes' is equal to 1 if an association is above the 80th percentile of the continuous train shock exposure measure defined in Section 5.1 and 0 if it is below the median. Specifically: the line-specific 2015–2020 average rail operationality is subtracted from the 2010–2015 average to create a measure of the size of the demand shock, which is then scaled by the distance to association's route networks. Standard errors are clustered at the Association level. *p<0.1; **p<0.05; ***p<0.01.

Table A2: DiD results: Varying the measurement of violence in the outcome variable

	Number of shootings/ Number of members	Number of shootings/ Baseline number of members	Any shooting	Number of shootings
	(1)	(2)	(3)	(4)
Post train shock × Train routes	0.0023** (0.0010)	0.0030*** (0.0011)	0.0430** (0.0193)	0.0850** (0.0336)
log(Members)	-0.0006 (0.0005)	-0.0003 (0.0005)	0.0072 (0.0105)	-0.0002 (0.0132)
log(Routes)	-0.0002 (0.0003)	-0.0004 (0.0003)	-0.0184* (0.0108)	-0.0250** (0.0119)
Local crime rate	0.0006 (0.0010)	0.0010 (0.0011)	-0.0223 (0.0206)	0.0006 (0.0327)
Observations	3,107	3,099	3,107	3,107
Outcome mean	0.00085	0.00085	0.0245	0.0305
R ²	0.16877	0.18570	0.26307	0.28363
Within R ²	0.00808	0.01052	0.00658	0.01047
Association fixed effects	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓

Note: This table presents the main results of the estimation of the effect of the shock to commuter rail on taxi-related violence, while varying the measurement of violence in the outcome variable. "Train routes" is equal to 1 if an association is above the 80th percentile of the continuous train shock exposure measure defined in Section 5.1 and 0 if it is below the median. The outcome variable in column (1) is used in the main analysis in the paper. Column (2) constructs a shooting rate based on the number of members in 2014. Column (3) is a binary association-level variable equal to one if the association was involved in a shooting in a given year. Column (4) is an inverse hyperbolic sine transformation of the number of shootings at the association level. Standard errors are clustered at the Association level. *p<0.1; **p<0.05; ***p<0.01.

Table A3: DiD results: Change in incidence of violence for associations most exposed to rail shock

	Shooting rate: Any shooting/Number of members				
	(1)	(2)	(3)	(4)	(5)
Post train shock	-0.0001 (0.0001)				
Train routes	0.0003 (0.0002)				
Post train shock × Train routes	0.0007*** (0.0002)	0.0007* (0.0004)	0.0007* (0.0004)	0.0009** (0.0004)	0.0009** (0.0004)
Standard-Errors	IID		Association		
Observations	3,110	3,110	3,107	3,107	3,107
Outcome mean	0.0005	0.0005	0.0005	0.0005	0.0005
R ²	0.01166	0.14752	0.14778	0.14892	0.14911
Within R ²		0.00318	0.00349	0.00483	0.00505
Association fixed effects		✓	✓	✓	✓
Year fixed effects		✓	✓	✓	✓
Local crime rate control			✓		✓
Association controls			✓		✓
Local crime rate control × Post				✓	✓
Association controls × Post				✓	✓

Note: 'Train routes' is equal to 1 if an association is above the 80th percentile of the continuous train shock exposure measure defined in Section 5.1 and 0 if it is below the median. Specifically: the line-specific 2015–2020 average rail operationality is subtracted from the 2010–2015 average to create a measure of the size of the demand shock, which is then scaled by the distance to association's route networks. The local crime rate control is the log of the total count of all crimes recorded in police precincts linked to association route networks. Association controls are the number of members and number of routes. Standard errors are clustered at the Association level. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A4: Violence panel regressions

	Involved in a shooting: year t		Number of shootings: year t	
	(1)	(2)	(3)	(4)
Involved in a shooting: year $(t - 1, t - 2)$	-0.0733*** (0.0190)		-0.0626** (0.0258)	
Involved in a shooting: year $t - 1$		-0.0541* (0.0281)		-0.0233 (0.0418)
Involved in a shooting: year $t - 2$		-0.0874*** (0.0249)		-0.0998*** (0.0359)
Observations	7,967	7,951	7,967	7,951
R ²	0.17925	0.18093	0.18286	0.18707
Within R ²	0.00822	0.01003	0.00425	0.00720
Association fixed effects	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓

Note: The number of shootings variable is an inverse hyperbolic sine transformation of the number of shootings: $\text{iht}(n) = \ln(n + \sqrt{n^2 + 1})$.

Table A5: Treatment assignment based on exposure to non-commuter train stations.

	Involved in a shooting	
	(1)	(2)
Non-commuter train route \times post train shock	-0.0049 (0.0080)	
Share of non-commuter train routes \times post train shock		-0.0049 (0.0080)
Observations	4,439	4,439
Association fixed effects	✓	✓
Year fixed effects	✓	✓

Note: In this specification, association-level treatment assignment is based on their distance to non-commuter train stations. A route is defined as a non-commuter train route if it has an origin and destination located within 3km of a non-commuter train station.

Table A6: Demand and supply changes in commuting by minibus taxi

Metric	2013	2020	% change
Membership in the taxi industry			
Number of members	38,666.00	57,032.00	47.50
Number of licensed members	8,390.00	14,119.00	68.28
Number of licensed vehicles	20,819.00	31,800.00	52.75
Number of registered routes	2,710.00	2,876.00	6.13
Commuting by taxi (per day)			
Taxi commuters to educational institutions	681,265.00	768,441.00	12.80
Taxi commuters to work	1,419,348.00	1,650,887.00	16.31
Total commuters by taxi	2,100,613.00	2,419,328.00	15.17
Rail agency statistics (per month)			
Train ticket sales	3,232.02	531.69	-83.55
Trains on time	22,764.00	7,234.00	-68.22
Commuting by train and bus (per day)			
Commuters by train	399,782.00	93,064.00	-76.72
Commuters by bus	434,145.00	267,954.00	-38.28
Taxi industry demand vs supply			
Commuters per member	54.33	42.42	-21.92
Commuters per licensed member	250.37	171.35	-31.56
Commuters per licensed vehicle	100.90	76.08	-24.60
Commuting times (minutes per month)			
Time spent commuting by taxi	2,557.64	2,778.69	8.64
Time spent commuting	2,147.74	2,127.17	-0.96
Real commuter expenditure on taxi fares (per month)			
Monthly taxi fare expenditure: educational commuters	657.35	841.34	27.99
Monthly taxi fare expenditure: work commuters	956.61	1,595.09	66.74
Monthly taxi fare expenditure	861.78	1,358.74	57.67
Monthly taxi fare expenditure per minute travelled	0.45	0.59	31.26
Taxi fare cost per minute travelled: educational commuters	0.28	0.48	71.75
Taxi fare cost per minute travelled: work commuters	0.34	0.64	90.10
Taxi industry revenue (work and educational commuters only)			
Taxi industry revenue: Commuters (1000s) x Monthly cost	1,805,587.57	3,279,829.76	81.65
Revenue per member	46.70	57.51	23.15
Revenue per licensed member	215.21	232.30	7.94
Revenue per vehicle	86.73	103.14	18.92

Note: This table presents weighted estimates from the 2013 and 2020 National Household Travel Survey data for Gauteng province. This data is combined with the administrative data on the minibus taxi sector to estimate changes in minibus taxi demand and supply, as well as the average cost to commuters. Only employment and educational commuters are considered here because monthly travel costs are captured consistently. All monetary estimates are reported in real terms and based on commuters who only travel by taxi.

Table A7: DiD results: change in minibus taxi-related shootings in areas with commuter rail stations

	Any shooting	
	(1)	(2)
Any train station \times post train shock = 1	0.0790*** (0.0246)	
Number of train stations \times post train shock = 1		0.0315*** (0.0111)
Observations	4,910	4,910
Outcome mean	0.047	0.047
R ²	0.23366	0.23277
Within R ²	0.00630	0.00514
Transport analysis zones fixed effects	✓	✓
Year fixed effects	✓	✓

Note: Estimation for the entire country. Minibus taxi-related shooting locations are geocoded and matched to transport analysis zones ($N = 381$). The treatment variable is based on the locations of commuter rail train stations. A total of 17% of all transport analysis zones contain train stations. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A8: DiD results: frequency of violence for associations most exposed to rail shock—varying treatment thresholds

	Shooting rate: # shootings/# members		
	(1)	(2)	(3)
Post train shock \times above P70	0.0014** (0.0007)		
Post train shock \times above P80		0.0023** (0.0009)	
Post train shock \times above P90			0.0032** (0.0015)
Observations	3,555	3,110	2,675
Outcome mean	0.00085	0.00085	0.00085
R ²	0.17240	0.16850	0.19005
Within R ²	0.00327	0.00774	0.01354
Association fixed effects	✓	✓	✓
Year fixed effects	✓	✓	✓

Note: The unit of observation is an association–year. ‘Above P70/P80/P90’ is equal to 1 if an association is above the 70th/80th/90th percentile of the continuous train shock exposure measure defined in Section 5.1 and 0 if it is below the median. The line-specific 2015–20 average rail operability is subtracted from the 2010–15 average to create a measure of the size of the demand shock, which is then scaled by the distance to the association’s route networks. All regressions include a control for local crime rates on association route networks. Standard errors are clustered at the association level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A9: Long-difference estimation of effects of change in rail operability on change in association-level taxi violence

	Δ Association-level shooting rate			
	(1)	(2)	(3)	(4)
Δ Train operability (E)	0.0044** (0.0022)	0.0053** (0.0023)	-0.0115** (0.0055)	-0.0121** (0.0055)
Δ Market structure		-0.0014 (0.0016)	-0.0234** (0.0111)	-0.0262** (0.0124)
Δ Train operability (E) \times Δ market structure			0.0049** (0.0023)	0.0054** (0.0025)
log(Δ licensed members)				-0.0033 (0.0024)
log(Δ routes)				0.0010 (0.0015)
Observations	454	441	441	441
R ²	0.02070	0.02611	0.04600	0.05122
Within R ²	0.01105	0.01587	0.03597	0.04124
District fixed effects	✓	✓	✓	✓

Note: Larger values of Δ Train operability (E) are associated with larger declines in train operability close to association route networks, and thus a larger positive demand shock. The outcome variable is the difference between the 2015–19 association average level of involvement in violence and the 2010–14 average. The market structure variable is measured as each rival association's membership size divided by its distance from the own-association route network, summed over all other associations. Standard errors clustered at the Association level. *p<0.1, **p<0.05, ***p<0.01.

Table A10: Registration Administration System summary statistics

Variable	Mean	Median	SD	Min	Max	N
Owners with registered vehicles:						
Number of vehicles	2.1	1	6.6	0	1064	55878
Number of licenses	2.1	1	6.9	0	1149	55878
More than one vehicle	0.45	0	0.5	0	1	55878
More than five vehicles	0.033	0	0.18	0	1	55878
Taxi associations:						
Total number of members	222	103	374	1	3909	811
Number of members w/vehicles	74	37	105	1	957	758
Number of vehicles	153	66	239	1	2013	758
Number of operating licenses	154	66	241	1	2021	758
Vehicles per member	1.9	1.7	1.5	1	36	758
HHI-licenses w/in association	0.16	0.043	0.28	0.002	1	758
Share of licenses held by largest owner	0.22	0.1	0.28	0.01	1	758
Number of routes	25	13	67	1	989	684
Members per route	7.1	4	18	0.039	367	656
Vehicles per route	13	7.5	32	0.078	625	656
New routes per year	0.75	0.091	2.8	0	40	684
New members per year	11	5	18	0	232	684

Note: This table presents summary statistics from the RAS/NLTIS system data for all seven provinces that use the information system. The first four rows present summary statistics based on the 55,878 individuals with registered vehicles and operating licenses. Not all registered individuals on the system (N=179,944) appear in the vehicle license file. The HHI measure is a Herfindahl-Hirschman Index of the concentration of the number of vehicles registered to individual owners within associations. The summary statistics are for the year 2022. The final two rows use information on when members join taxi associations, and when taxi associations register new routes to create per-year summary statistics over the period 2011–2022, when most provinces had adopted the use of the NLTIS system.

A1 Shopping centres

In order to test whether the results are generalisable to other demand shocks in this context, I use information on the timing and locations of new shopping mall openings. I geo-code the precise locations of all shopping centres opened in the country after the year 2000 (See appendix C7 for more detail.) I group the locations of these centre openings and the geo-coded locations of taxi-violence related shootings to Uber's H3 hexagonal polygons. Finally, using the administrative data and the market structure measures described in Section 3.3, I create area-level measures of the minibus taxi industry market structure in each location. I then estimate a difference-in-differences specification of the effect of the opening of shopping malls on taxi-related violence.

$$v_{ix,t} = \beta_1 Mall_{ix} \times Post\ mall\ opening_{ix,t} + \gamma_{ix} + \gamma_t + \varepsilon_{ix,t} \quad (9)$$

Where ix indexes polygon i at spatial scale x .

Finally, I also estimate a triple-difference specification where I interact the DiD interaction term with a measure of the number of taxi associations with a pre-existing presence in the area the mall is opened.

$$v_{ix,t} = \beta_1 Mall_{ix} \times Post\ mall\ opening_{ix,t} \times Baseline\ competition_{ix} + \gamma_{ix} + \gamma_t + \varepsilon_{ix,t} \quad (10)$$

Table A11: Violence following the opening of shopping malls

	Taxi related shooting in Hexagonal unit level 6 in year t				
	(1)	(2)	(3)	(4)	(5)
Post mall opening \times Mall opened in sample period = 1	0.0487*** (0.0104)	0.0563*** (0.0095)	0.0062 (0.0050)	0.0169*** (0.0047)	
Mall opened in sample period	0.0496*** (0.0068)				
Multiple associations in area			0.0376*** (0.0107)	0.0479*** (0.0105)	
Multiple associations in area \times Post mall opening \times Mall opened in sample period = 1			0.1009*** (0.0293)	0.1018*** (0.0305)	0.0379 (0.0323)
Observations	181,958	181,958	181,568	181,568	181,958
R ²	0.04012	0.64644	0.09439	0.06999	0.64347
Within R ²	0.03969	0.00884	0.02580	0.04127	0.00051
Year fixed effects	✓	✓	✓	✓	✓
Hexagonal unit level 6 fixed effects		✓			✓
Hexagonal unit level 5 fixed effects			✓		
Hexagonal unit level 4 fixed effects				✓	

Note: Multiple associations indicates the presence of more than one taxi association in the area, based on their route networks. Level 6 hexagons - the spatial scale at which the regression is run, have an average edge length of 3.7km. Level 5 and 4 hexagons have average edge lengths of 9.8km and 26.1km, respectively. Significance codes: ***=0.01, **=0.05, *=0.10

B Institutional background

B1 Collapse of commuter rail

This section provides more context for why the decline in service quality in commuter rail occurred and presents the passenger rail agency data used in the empirical strategy. The following extract from the 2016/2017 PRASA annual report provides a concise summary of the causes of the onset of declines in train operability.

Extract from the 2016/17 PRASA annual report:

Irregularities in contract awards identified by the AGSA in the 2014/15 audit and the report of the Public Protector titled “Derailed” and issued in August 2015, resulted in the review of procurement and contract processes. The contract irregularities identified negatively impacted the ability of the organisation to maintain rolling stock assets as procurement of spares and components were terminated. The theft and vandalism of cables and components aggravated the challenges of Maintenance resulting in increased numbers of rolling stock unavailability and unreliability. This negatively affected delivery of safe, clean and secure train service.

Conversations with PRASA officials corroborated the opinion expressed in the annual report. One senior official summarised the collapse in commuter rail that started in 2015 as being due to a culmination of factors, sparked by the events of 2015. Firstly, the instability of senior management led to a lack of consistency and a lack of decision making capability within the agency.

Secondly, important service delivery agreements that had been irregularly procured were set aside. This meant that these essential services were no longer available to PRASA. Thirdly, the investigations by the Public Protector, the Auditor General and subsequent remedial action (which including further forensic investigations of large contracts and increased oversight by the National Treasury) led to managers “treading carefully” and not making key decisions for fear of being involved in the maladministration scandal. This led to widespread organisational paralysis. To contextualise this, I was told by a senior PRASA employee that teams from both the Special Investigations Unit and the Hawks (an elite crime unit in the Police) took up offices inside the PRASA building.

Finally, in 2017/18 irregularly granted security contracts were also set aside. According to the official I spoke to, the affected service providers retaliated by vandalising PRASA lines. There are widespread rumours that the taxi industry also seized the opportunity to further sabotage PRASA lines and there are instances of such vandalism that have been reported in Cape Town. The PRASA official I spoke to mentioned PRASA had no clear proof of this in Gauteng.

Table B1: Managerial instability: Passenger Rail Agency of South Africa CEO turnover before and after 2015

Term started	Term ended	Surname	Name
2010	2015	Montana	Tshepo Lucky
2015	2016	Khena	Nathi
2016	2017	Letsoalo	Collins
2017	2018	Zide	Lindikaya
2018	2019	Sithole	Sibusiso
2019	2019	Sishi	Nkosinathi
2019	2020	Mpondo	Bongisizwe
2020	2021	Mabija	Thandeka
2021	2021	Matthews	Zolani Kgosie
2021	2022	Mphelo	David
2022	Current	Hishaam	Emeran

Note: The table illustrates the executive level instability that ensued at the Passenger Rail Agency after 2015—a key contributing factor in declining service quality.

B2 State intervention in Gauteng in 2019

Figure A3 suggests police responded to increases in taxi violence in 2019. Here I provide more context to the state intervention that reduced violence in 2019.

The first factor was the organization of the Gauteng Taxi Summit in July 2019. The previous summit had taken place in 2013.

The following is an extract from the Gauteng Department of Roads and Transport MEC Jacob Mamabolo's speech at the Provincial taxi summit held in Sedibeng in July 2019:

...Programme Director, allow me to share with you some of the profound and critical elements of the Modernisation of the Minibus Taxi Industry:

1) **Immediately bring to an end 'Taxi Wars' that are ravaging the Taxi industry.** This we will do by **getting the support of the key players, especially the Taxi Associations** to embed a two pronged approach. In the first instance, we need **law enforcement agencies and the judiciary to act decisively** on crimes committed in the Taxi Industry...

Following the Taxi Summit, a Gauteng Provincial Taxi Monitoring Committee (GPTMC) was established in conjunction with the Provincial Executive Committees of both the Gauteng National Taxi Alliance (GNTA) and the South African National Taxi Council (SANTACO). The taxi summit agreed the following:

...that provincial government working with the taxi industry should commit to modernising the industry through technology and **ensuring peace and stability** in the industry so that the sector contributes more towards growing of the Gauteng economy. It also agreed that

unity among the industry players would make it easier for the minibus taxi industry to participate in a meaningful way in the whole transport value chain. - Gauteng Province Department of Roads and Transport. Media Statement, 19 August 2019.

In addition, a commission of inquiry into Taxi violence was established by the Premier of the province in 2019.

...Premier Makhura, had established the Commission referred to in terms of Proclamation 101 of 2019 issued on 13 September 2019 and its brief was to investigate the underlying causes of taxi violence, make findings and recommendations. – Gauteng Department of Roads and Transport Media Statement 01 Dec 2020

C Data construction

C1 National Land Transport Information System (NLTIS) data

Access to the administrative data in the National Land Transport Information System (NLTIS) was obtained through written requests to the Gauteng Department of Roads and Transport and to the National Department of Transport. The information system joins two sub-systems, the Registration Administration System (RAS) and Operating License Administration System (OLAS). The main analysis presented here is based on the version of the data obtained from the National Department of Transport in May 2022. Provincial Transport Regulatory Authorities (PRE) input data into the NLTIS system and its sub-systems while processing road carrier permits and licenses for new routes. There is significant provincial and intra-provincial variation in the quality of route descriptions in the data and required extensive data cleaning.

The data sent to the author consists of a series of separate files in which rows correspond either to taxi associations, taxi and bus routes, or taxi operators.⁴²

Taxi operators have a unique RAS number from which the province they are registered in can be extracted. The taxi association under whom an individual member falls is also always listed in the data. Six different data files were sent to the author from the Department of Transport and the variables contained within them in each are listed here:

- Association routes (17,578 observations)
 - AssociationName: the name of the taxi association operating on the route
 - AssociationStatusDate: the date on which the Association is registered on the system
 - RouteDescription: A narrative text description of the route the license applies to, mentioning the origin, destination and specific directions. This data is entered by employees of the various PREs and as such, there exists substantial variation in the quality of the exact directions, but origins and destinations are always mentioned.
 - RouteDate: the date on which the route is registered on the system
- Association Names (1,220 observations)

⁴² Individual taxi operators can be linked to the associations they belong to in the data. Routes are allocated to taxi associations and not directly to operators on the system. This means that the within-association route-allocation is opaque.

- AssociationName: the name of the taxi association. This includes associations based in KwaZulu-Natal and the Western Cape, which have since migrated to different operating license systems.
- Member list (189,542 observations)
 - MemberRegistration: System unique identifier of each member e.g. "RASECA33444M116438". The 4th and 5th characters contain the province the member is registered under.
- Member vehicles (55,878 observations)
 - NumberVehicle: number of registered vehicles belonging to the operator
 - MemberRegistration: System unique identifier of each member
 - Names: Name and surname or Initial and surname of each member
 - AssociationName: Taxi association the member belongs to
 - NumberLicense: Number of operating licenses the operator has
- Members (179,994 observations)
 - AssociationName: Taxi association the member belongs to
 - MemberRegistrationNumber: System unique identifier of each member
 - MemberStatusDate: The date when the taxi operator is registered on the system
- Bus routes (22,457 observations)
 - RouteDescription: Description of bus route
 - date: Date the bus route was registered on the system

Figure C1: Example of RAS data structure

1) AOECCHLB11366 2) S. NKONJANE 6705235603089 3) P.O. BOX 4, ENGCOCO, ENGCOCO, 5050 United
4) not available New OL Engcoco Taxi Association
5) 1 x 15 (Seating)
6) Gazette Number: CHRIS HANI (20 Oct 2015)

7.1.) Proposed List of Routes
7.1.1.) Minibus taxi-type service
7.1.1.1. Engcoco (Origin) Durban Taxi Rank (Destination) - ecbrcchd2440(31005D20018410200557)From Engcoco Taxi Rang
towards Mqanduli via Idutywa, Port St Johns, Lusikisiki, Flagstaff, Bizana, Port Edward or via Bulwa, Underberg,
Bethlehem, Villiers, Standerton or towards Umtata via Tsolo Qumbu, Mount Frere, Mount Ayliff, Kokstad, Umzimkulu,
Ixopo, Pietermaritzburg or via Portshepston to Durban and return either way.

7.1.1.2. Engcoco (Origin), Welkom (Via Point 1), Rustenburg (Via Point 2), Klerksdorp (Destination) -
ecbrcchd244(31005D61009S10047016)From Ngcoco taxi rank join R30 to Welkom, Rustenburg, Klerksdorp and return
on the same route

Note: This is an example of a print-out of the RAS data from a Government Gazette. Operating license applications are published in government gazettes to allow for public objections. This image makes clear the identity of the individual applicant, their ID number, the association under which they are applying and the specific set of routes for which they are applying for an operating license. Generally individuals apply for the entire set of routes the taxi association they belong to has operating licenses for.

Source: South African Government Gazette No. 39389, 13 November, 2015. page 337. Available here: <https://search.opengazettes.org.za/text/1145>.

C2 Passenger Rail Agency of South Africa

The following section presents data from commuter rail lines using PRASA administrative data on the number of trains and ticket sales on each rail corridor. The variables in the data shared with the author are presented in Table C1. The daily data was only shared for the last day of the month—an idiosyncrasy of the data export process.

Table C1: Variables in Prasa line-level data

Variable	Example
line	Pretoria - Belle Ombre - Mabopane
actual_day_scheduled	139
actual_day_cancelled	2
actual_day_late	6.00
actual_day_minutes	98
actual_day_time	131
actual_day_perctime	95.62
revised_day_cancelled	0
revised_day_late	0.00
revised_day_minutes	0
revised_day_time	131
revised_day_perctime	95.62
actual_month_scheduled	3111
actual_month_cancelled	93
actual_month_late	760.47
actual_month_minutes	25835
actual_month_time	2257.53
actual_month_perctime	74.80
revised_month_cancelled	1
revised_month_late	0.00
revised_month_minutes	0.00
revised_month_time	2257
revised_month_perctime	74.79
area	north
date	30
day	thursday
month	april
year	2009

C3 Violence in the minibus taxi industry

This section describes the construction of the novel dataset on taxi related violence in more detail. The process involved the following steps:

1. Using the search terms "taxi" AND "shot"; "taxi" AND "violence"; "taxi" AND "shooting"; "taxi" AND "geweld" ("geweld" is the Afrikaans translation of violence), I search through the SABINET media archive for all years from 2006 to 2022. I then download the raw .pdfs of the news clippings yielded by this search. This process yields 2339 news articles.
2. Extracting the text from the news clippings using Optical Character Recognition (OCR) Software and storing the resulting text in a set of .csv files. I use the Amazon Textract layout function. I then combine the .csv files into a data set where each row is a news article. After dropping duplicate articles, this leaves 1563 news articles.

3. Extracting usable information from the text. First, I extract the the names of all taxi associations mentioned in each article. This is done in two ways. The first is searching for the term "association" or other common words that occur in taxi association names and extracting the four words preceding and two words after the term. Based on this - I match the specific strings to the names of taxi associations in the administrative data. For example, I match "the Alexandra, Randburg, Midrand, Sandton Taxi Association chairperson" in the news article to "ALEXANDRA RANDBURG MIDRAND SANDTON LOCAL AND LONG DISTANCE TAXI ASSOCIATION" - the name recorded in the administrative dataset. In a second method used ensure no association names are missed, I search through the news article for any direct matches between the lowercase value of administrative data name and the name in the news article. I then first manually code news articles.
4. Dropping duplicate articles and articles that do not describe specific cases of violence. This leaves 1311 news articles.
5. Identifying and keeping only articles describing specific types of violence related to the taxi industry. The present research article studies violence within the minibus taxi sector, where both the perpetrators and targets of violence are actors in the minibus taxi sector. I exclude all articles and violent events where violence is directed at non-taxi competitors (such as the Bus Rapid Transit System (Rea-Vaya, MiCiti) or Conventional bus operators (Golden Arrow, Intercape)). I also exclude articles involving actors in the minibus taxi sector that are clearly not related to the taxi industry (such as accounts of domestic abuse by taxi owners). This leaves 1274 articles in the final dataset.
6. Reshaping the article-level database to an association-year level database.

Examples of media reports that form the basis of this paper are presented in Figure C2, with relevant passages highlighted.

I am indebted to the Global Initiative Against Transnational Organised Crime (GI-TOC), and in particular to Kim Thomas who collected and shared their database on violence in the taxi industry (Thomas 2021). However, the GI-TOC database does not consistently collect data on the names of taxi associations mentioned in news reports, nor specific locations of taxi related shootings. For these reasons, I recollect and reconstruct the data entirely, using a similar methodology.

Figure C2: Example of media reports that form the basis of this paper

BONGANI NKOSI

WHILE the Ivory Park Taxi Association (IPTA) was still reeling from shock last night after the massacre of 11 of its members in a bloody ambush, Parliament conceded that violence in the industry had reached crisis levels.

Police yesterday confirmed that 11 people were killed in an attack on the R74 near Ladysmith in KwaZulu-Natal on Saturday night.

The assailants opened fire on a Toyota Quantum minibus taxi carrying 17 people. Two people escaped unharmed and four critically wounded were in hospital.

The taxi was shot at more than 200 times with high-powered rifles, including an AK47.

The operators and mechanics linked to a Joburg-based taxi association were ambushed on their way to Joburg from a funeral in Greytown.

Her daughter said she thought they were under attack. "The shooting went on for what seemed like eternity, but I think it was between five and 10 minutes.

"In the midst of the shooting I was receiving calls from some of our neighbours witnessing the 'fire' and asking me what was going on at my home, as it looked like the shooting was directed at us," she said.

General Khehla Sitole, the police's national commissioner, announced a 72-hour action plan to track down the killers. The crack unit probing it includes officers from crime intelligence, the Hawks and the Special Task Force.

Sitole described the attack as a "heinous crime".

"We will wait for the investigation to advance before speculating on a motive," he said.

IPTA was also still in the dark.

Taxi violence has claimed a number of lives in Gauteng recently. Just last week, two taxi bosses from different associations were killed.

Sanele Maseko, the Nancefield Dube West Taxi Association deputy chairperson, was ambushed and killed.

Themba Sam Budza, chairperson of the Alexandra, Randburg, Midrand and Sandton Taxi Association, died along with association member Fani Ramatfwane in a hail of bullets.

In May, four taxi drivers were chased in Brakpan and fatally shot by guards from a security firm called Ingonyama VIP Security.

The guards were hired to protect an official of the Greater Brakpan Taxi Association, which was hit by disputes. There had been attempts on his life before the Brakpan killing.

Source: The Star, 23 July 2018.

C4 Collected price data

Photographs of 120 fare lists are collected in total, OCR'd and converted to excel tables and combined into one excel sheet. The origin and destination points are then carefully cleaned and geocoded and the shortest driving distance between the points is calculated as an approximate indicator of the route distance.

Figure C3: Example of collected price data

Ivory Park Taxi Association
LOCAL AND LONG DISTANCE TRANSPORT

Richard 1555
ONE RICHARDS DRIVE & GALLAGHER AVENUE
GALLAGHER CORNER, BLOCK D, UNIT 2
MIDRAND, 1685

26/06/2020

TAXI FARE INCREASE AS FROM THE 1st JULY 2020

IVORY PARK EXT. 2/4/5 & KAALFONTEIN	TO WATERFALL	R18.00
IVORY PARK EXT 2/4/5 & KAALFONTEIN	TO KYALAMI	R19.00
IVORY PARK EXT 2/4/5 & KAALFONTEIN	TO MIDRAND, VORNAVELLY, VODAWORLD, NOORDWYK	R18.00
IVORY PARK 2/4/5 & KAALFONTEIN & EXT. 45	TO ALEXANDRA, SANDTON, SUNNINGHILL & RIVONIA	R21.00
IVORY PARK 2/4/5 & KAALFONTEIN & EXT. 45	TO ROSEBANK	R22.00
IVORY PARK EXT 2/4/5 & KAALFONTEIN & EXT. 45	TO JOHANNESBURG	R22.00
IVORY PARK EXT 2/4/5 & KAALFONTEIN & MAYIBUYE	TO CENTURION	R22.00
IVORY PARK EXT 2/4/5 & KAALFONTEIN & MAYIBUYE	TO SHOPRITE	R19.00
IVORY PARK EXT 2/4/5, KAALFONTEIN, MAYIBUYE	TO CORPORATE PARK	R19.00
IVORY PARK EXT 2	TO FOREST HILL MALL	R22.00
IVORY PARK EXT 2	TO ATTRIDGEVILLE	R26.00
IVORY PARK EXT 2/4/5, KAALFONTEIN, MAYIBUYE	TO LIMBRO PARK	R20.00
IVORY PARK EXT 3	TO LIMBRO PARK	R21.00
IVORY PARK EXT 2/3/4/5, KAALFONTEIN, MAYIBUYE	TO GERMISTON	R24.00
IVORY PARK EXT 2/3/4/5, KAALFONTEIN, MAYIBUYE	TO EDENVALE	R22.00
MIDRAND	TO DIEPSLOOT	R23.00
MIDRAND	TO DAVEYTON	R30.00
IVORY PARK 2/4/5	TO DIEPSLOOT & OLIVENHOUT	R23.00
IVORY PARK 2/4/5	TO ALBERTON	R27.00

Note: This image makes clear the taxi association, the date of the price, the route origin and destination and price paid by commuters. This project collected roughly 120 images similar to this from various sources described in Section 3, comprising a total of 2200 taxi routes and 90 taxi associations. Following collection, the data was transferred to an excel sheet. Origin and destination points were carefully cleaned to ensure accuracy and then batch Geo-located using the Google Maps API.

C5 Commuting costs

I rely on the National Household Travel Survey (NHTS) data for 2013 and 2020 (Statistics South Africa 2014, 2021b). I focus on work and educational commuters, for whom commuting modes, costs and times are recorded consistently, and in detail, across the two years. For each individual, I calculate their all-mode total monthly commuting cost, the cost only paid in taxi fares, and the amount of time spent commuting (in total and by minibuss taxi). Secondly, I identify commuting areas, Travel Analysis Zones (TAZ) in the data (N=384) that relied heavily on commuting by rail in 2013 (roughly 2 years before the decline in rail operability began). I create an arbitrary cut-off for these "Rail commuting zones" as those areas where the number of commuters by train represent at least a quarter (0.25) of commuters by taxi in 2013. There are 22 such "treated" TAZs, containing total of 23612 survey respondents. In order to ensure observed changes are not driven by very small sample sizes, or by very different areas, I restrict the analysis to TAZ's with at least 250 individual observations in each year (N=201). Using these measures, I compare changes in commuting costs and times across TAZs with high baseline rail commuting shares, with those with low baseline rail commuting shares, before and after the rail shock.

Figure C4: Example of joint price setting

NOTICE!!!! NOTICE!!!!NOTICE!!!! NOTICE!!!!

PRICE INCREAMENT PRS & RULLDTA

LOCATIONS	PRICES
PRETORIA TO DIEPSLOOT	R25
PRETORIA TO COSMO CITY / HONEYDEW	R35
OLIEVENHOUTBOSCH TO DIEPSLOOT	R18
WIERDAPARK TO DIEPSLOOT	

NOTICE!!!! NOTICE!!!!NOTICE!!!! NOTICE!!!!

PRICE INCREAMENT PRS, ZMTA, CTA & LAUDIUM

LOCATIONS	PRICES
LAUDIUM WIERDA PARK	R16
LAUDIUM CENTURION	R18

THANKS FOR YOUR CONTINUED SUPPORT!!!

Note: This image is of fare lists showing the Pretoria Randburg Soshanguve (PRS), Rustenburg United Local and Long Distance Taxi Association (RULLDTA), as well as the PRS, Zwanani Marabastad Taxi Association (ZMTA), Centurion Taxi Association (CTA) and Laudium Taxi associations setting prices together. Associations also sometimes operate what are colloquially called 'joint-ventures', by sharing routes or taxi ranks and splitting the profits.

C6 Police performance measures using crime statistics

Yearly, police-precinct level data published by the South African Police Service (SAPS). The data is publicly available on the SAPS website SAPS (2020). Table C2 presents an overview of the variables, as reported by SAPS.

Table C2: Summary of variables in crime statistics data

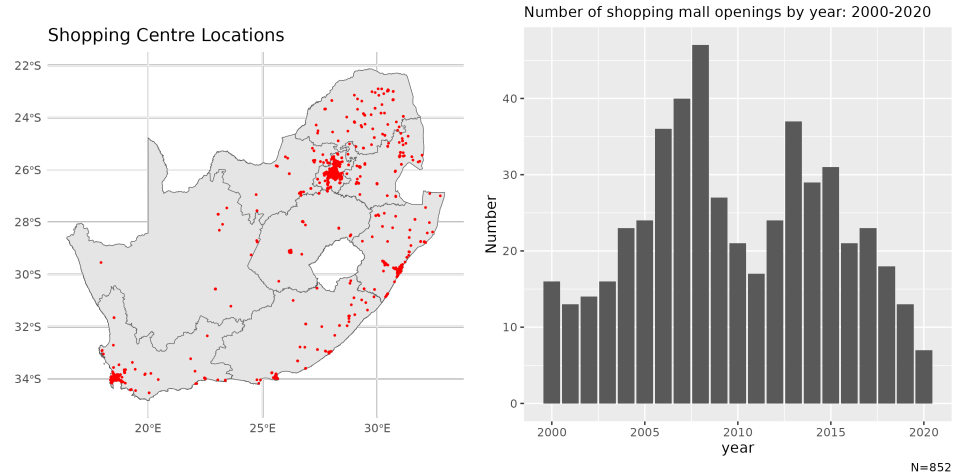
CRIME CATEGORY	2010/2011	2011/2012	2012/2013	2013/2014	2014/2015	2015/2016	2016/2017	2017/2018	2018/2019	2019/2020
CONTACT CRIMES (CRIMES AGAINST THE PERSON)										
Murder	15,893	15,554	16,213	17,023	17,805	18,673	19,016	20,336	21,022	21,325
Sexual Offences	64,921	60,539	60,888	56,680	53,617	51,895	49,660	50,108	52,420	53,293
Attempted murder	15,360	14,730	16,236	16,989	17,537	18,127	18,205	18,233	18,980	18,635
Assault with the intent to inflict grievous bodily harm	197,470	191,612	185,050	182,333	182,556	182,933	170,616	167,352	170,979	166,720
Common assault	184,103	180,165	171,653	166,081	161,486	164,958	156,450	156,243	162,012	165,494
Common robbery	54,442	52,566	53,196	53,505	54,927	54,110	53,418	50,730	51,765	51,825
Robbery with aggravating circumstances	101,039	100,769	105,488	118,963	129,045	132,527	140,956	138,364	140,032	143,990
Total Contact Crimes (Crimes Against The Person)	633,228	615,935	608,724	611,574	616,973	623,223	608,321	601,366	617,210	621,282
Total Sexual Offences										
Rape	48,158	47,069	48,408	45,349	43,195	41,503	39,828	40,035	41,583	42,289
Sexual Assault	7,006	7,194	6,967	6,597	6,087	6,212	6,271	6,786	7,437	7,749
Attempted Sexual Offences	3,599	3,535	3,293	2,913	2,641	2,573	2,073	2,066	2,146	2,076
Contact Sexual Offences	6,158	2,741	2,220	1,821	1,694	1,607	1,488	1,221	1,254	1,179
Total Sexual Offences	64,921	60,539	60,888	56,680	53,617	51,895	49,660	50,108	52,420	53,293
SOME SUBCATEGORIES OF AGGRAVATED ROBBERY										
Carjacking	10,541	9,417	9,931	11,180	12,773	14,602	16,717	16,325	16,026	18,162
Robbery at residential premises	16,889	16,766	17,950	19,284	20,281	20,820	22,343	22,261	22,431	21,130
Robbery at non-residential premises	14,637	15,912	16,343	18,573	19,170	19,698	20,680	20,047	19,991	20,651
Robbery of cash in transit	290	182	145	145	119	137	152	238	183	164
Bank robbery	39	35	7	21	17	6	3	13	4	0
Truck hijacking	999	821	943	991	1,279	1,184	1,183	1,202	1,182	1,202
CONTACT-RELATED CRIMES										
Arson	6,157	5,996	5,665	5,458	5,127	4,903	4,321	3,869	4,083	3,783
Malicious damage to property	122,814	119,907	119,026	117,983	120,662	119,901	116,409	111,492	113,089	108,461
Total Contact-Related Crimes	128,971	125,903	124,691	123,441	125,789	124,804	120,730	115,361	117,172	112,244
PROPERTY-RELATED CRIMES										
Burglary at non-residential premises	68,907	69,902	73,492	73,464	74,358	75,008	75,618	71,195	71,224	69,713
Burglary at residential premises	246,612	244,667	261,319	259,784	253,716	250,606	246,654	228,094	220,865	205,959
Theft of motor vehicle and motorcycle	64,162	58,800	58,102	56,645	55,090	53,809	53,307	50,663	48,324	46,921
Theft out of or from motor vehicle	122,334	129,644	138,956	143,801	145,358	139,386	138,172	129,174	125,076	118,213
Stock-theft	26,942	27,611	26,465	24,534	24,965	24,715	26,902	28,849	29,672	28,418
Total Property-Related Crimes	528,957	530,624	558,334	558,228	553,487	543,524	540,653	507,975	495,161	469,224
OTHER SERIOUS CRIMES										
All theft not mentioned elsewhere	361,222	370,916	356,847	363,517	360,541	340,372	328,272	302,656	300,457	282,817
Commercial crime	85,646	85,570	89,138	76,744	67,830	69,917	73,550	73,277	83,823	83,869
Shoplifting	78,326	71,810	71,267	70,487	71,327	68,786	67,454	62,180	60,167	59,883
Total Other Serious Crimes	525,194	528,296	517,252	510,748	499,698	479,075	469,276	438,113	444,447	426,569
Total 17 Community Reported Serious Crimes	1,816,350	1,800,758	1,809,001	1,803,991	1,795,947	1,770,626	1,738,980	1,662,815	1,673,990	1,629,319
CRIME DETECTED AS A RESULT OF POLICE ACTION										
Illegal possession of firearms and ammunition	14,385	14,372	14,813	15,362	15,116	14,772	16,134	17,558	15,736	15,779
Drug-related crime	150,561	176,218	206,721	260,596	266,902	259,165	292,689	323,547	232,657	170,510
Driving under the influence of alcohol or drugs	66,645	69,410	71,025	69,725	68,561	76,159	75,034	86,160	82,912	94,273
Sexual Offences detected as a result of police action	0	2,726	4,175	4,720	6,340	5,830	6,164	6,701	7,976	9,614
Total Crime Detected As A Result Of Police Action	231,591	262,726	296,734	350,403	356,919	355,926	390,021	433,966	339,281	290,176

Note: The statistics presented here are national total crime counts by year, taken directly from (SAPS 2020). The precinct-level data used in the main analysis contains the same variables for each police precinct (N=1156), for each year.

C7 Shopping centre openings

To locate the openings of shopping centres in South Africa, I use the South African Council of Shopping Centres 2022 Shopping Centre Directory (SACSC 2022). The directory contains the names, addresses, locations, opening dates, renovation dates and floor area of 1328 shopping centres opened in the country. Many South African shopping centres are very large in scale (the average Retail GLA in the post 2000 sample is $23366 m^2$, and the average number of stores is 57), employ thousands of workers, and attract tens of thousands of customers a day.

Figure C5: Locations and timing of shopping centre openings



D Additional summary statistics

D1 National Land Transport Information System (NLTIS) summary statistics

Table D1: Correlation matrix: NLTS data at the association level (2019)

	Cumulative Routes	Cumulative Licensed Members	Cumulative Members	Cumulative Vehicles	Members per Route	Licensed Members per Route	Vehicles per Route	HHI of Licenses
Cumulative Routes	1.00	0.15	0.12	0.15	-0.10	-0.08	-0.08	-0.08
Cumulative Licensed Members	0.15	1.00	0.90	0.94	0.27	0.24	0.27	-0.26
Cumulative Members	0.12	0.90	1.00	0.83	0.28	0.16	0.17	-0.20
Cumulative Vehicles	0.15	0.94	0.83	1.00	0.21	0.19	0.26	-0.17
Members per Route	-0.10	0.27	0.28	0.21	1.00	0.94	0.91	-0.08
Licensed Members per Route	-0.08	0.24	0.16	0.19	0.94	1.00	0.97	-0.11
Vehicles per Route	-0.08	0.27	0.17	0.26	0.91	0.97	1.00	-0.07
HHI of Licenses	-0.08	-0.26	-0.20	-0.17	-0.08	-0.11	-0.07	1.00

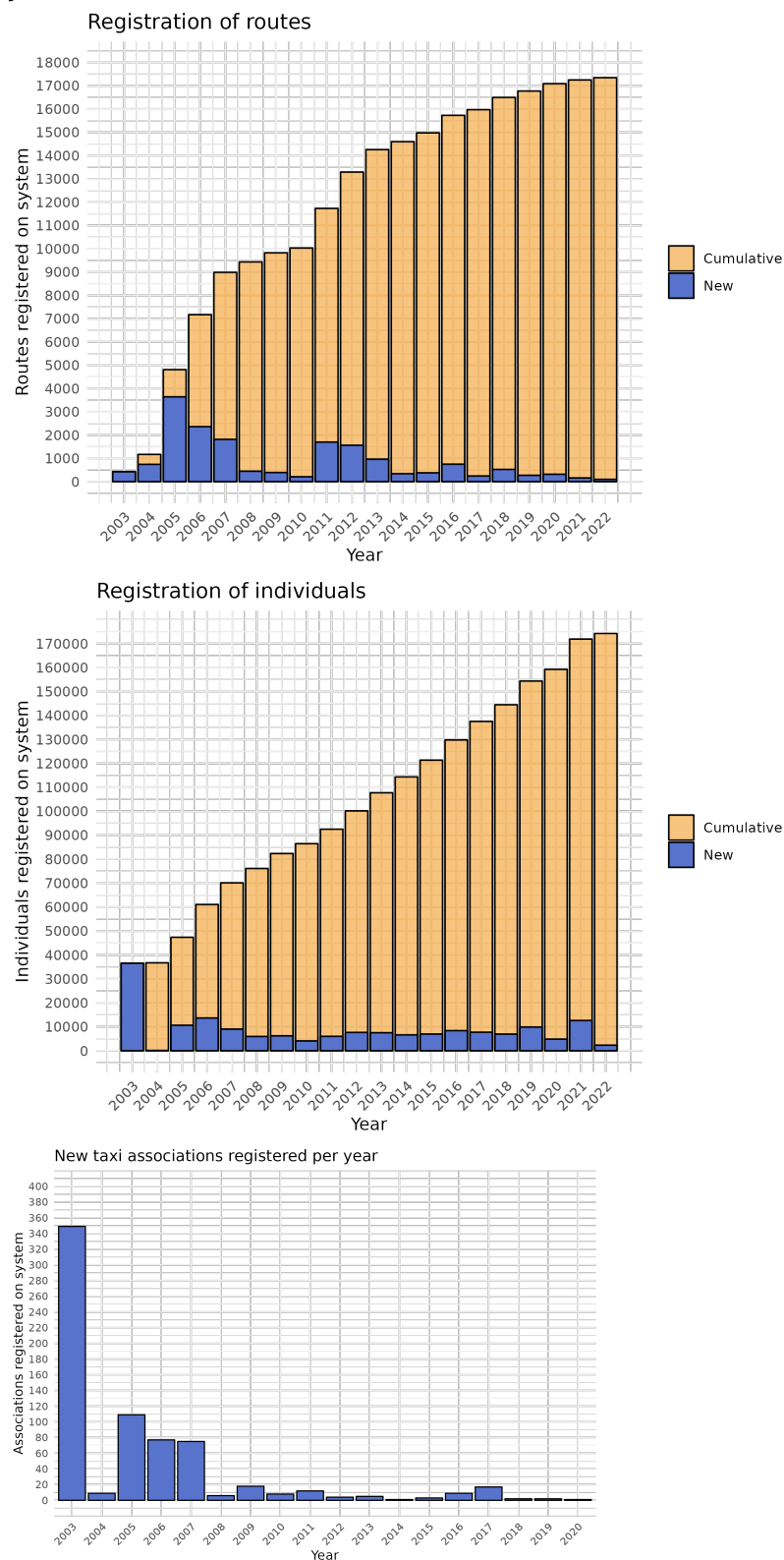
Note: This table presents pairwise correlations of the main variables created from the NLTIS data. The variables are all defined at the association level. The HHI of Licenses variable is a measure of vehicle license ownership concentration within the association. Where l_{ia} is the number of registered licenses of operator i in association a with N members and L_a is the total number of licenses under association a , $HHI_a = \sum_{i=1}^N (\frac{l_i}{L_a})^2$.

Table D2: Provincial summary statistics of the registration administration system data

Variable	Eastern Cape	Free State	Gauteng	Limpopo	Mpumalanga	Northern Cape	North West
Number of taxi associations	93	68	219	119	75	21	89
Number of routes	2198	1508	3226	1212	2090	4733	2380
Number of members	35224	16763	69220	16883	13933	2840	19371
Number of members with licensed vehicles	8852	5175	17347	7386	7159	1364	8125
Average members per association	379	247	316	142	186	135	218
Average members with licensed vehicles per association	95	76	79	62	95	65	91
Average routes per association	24	22	15	10	28	225	27
Average members per association	17.9	28.6	25.8	21.6	7.7	1.4	20.1
Average number of registered vehicles	192	138	169	117	234	125	187
Average HHI of licenses within associations	0.065	0.098	0.127	0.061	0.059	0.056	0.089
Average maximum share of licenses held by one owner within associations	0.13	0.18	0.19	0.12	0.11	0.13	0.14

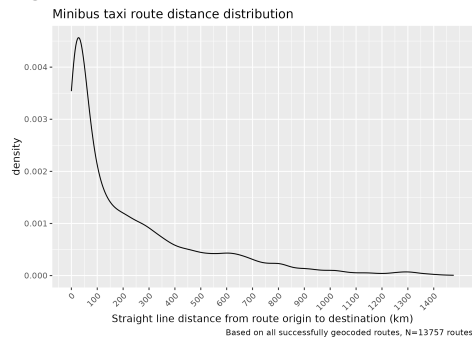
Note: This table presents provincial summary statistics of the RAS/NLTIS data. The first four rows are provincial totals. The final seven rows are provincial averages of association-level variables. The Western Cape and KwaZulu-Natal migrated to the use of different information systems for taxi licenses. As such, summary statistics are not available for these two provinces.

Figure D1: New routes, new members and new taxi associations registered on RAS/OLAS system per year



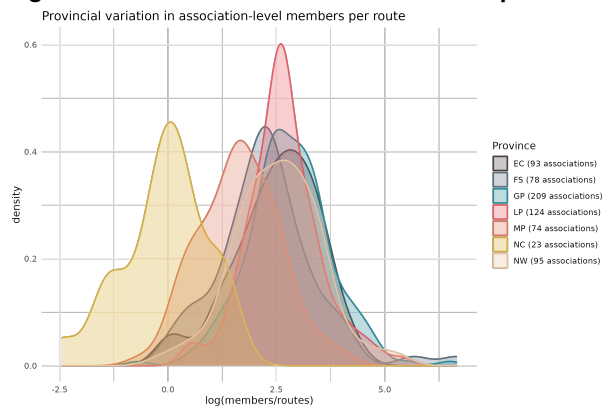
Note: The rollout of the system started in 2003, which is the reason for the large numbers of registrations in the early and mid 2000s. The main analysis of this article is restricted to the period after 2010 when the rollout had been completed.

Figure D2: Route distance



Note: This figure presents the straight line distance of geocoded routes in the RAS data for all provinces.

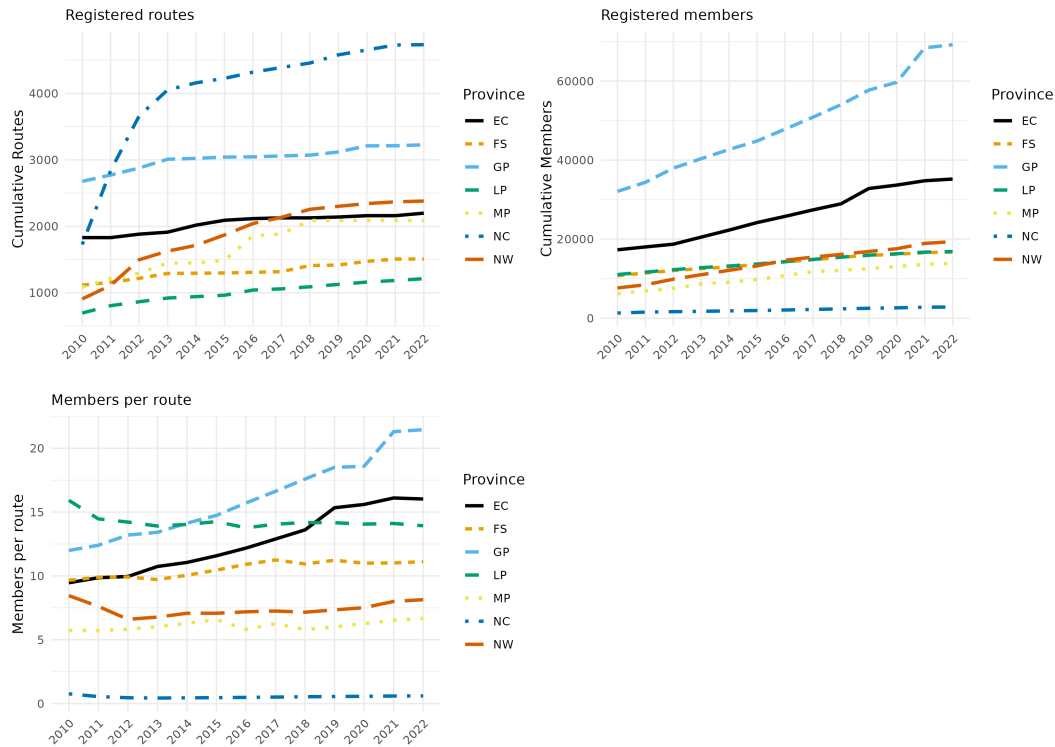
Figure D3: Provincial variation in members per route



Note: For each association (N=696) the number of members registered under the association is divided by the number of routes registered under the association. This value is logged and the provincial distribution is plotted for the seven provinces in the RAS data.

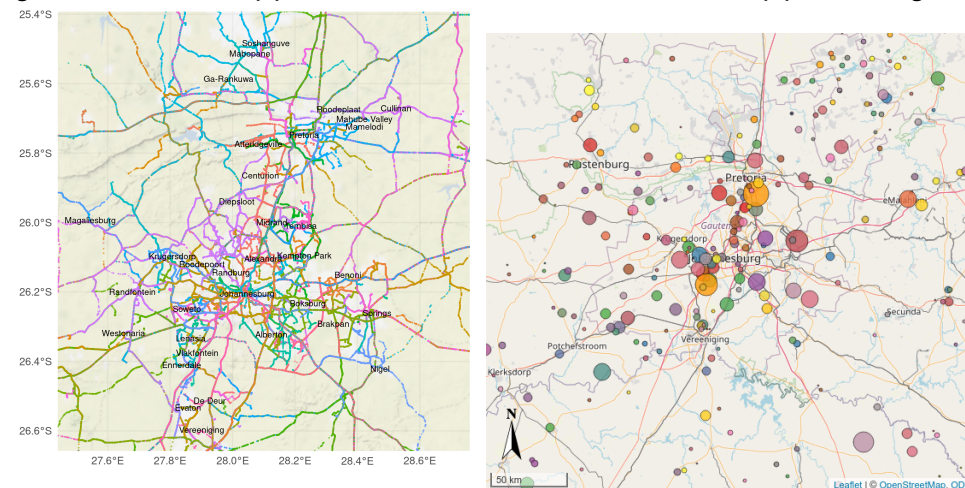
Source: This figure created using the 2022 version of the NLTIS data.

Figure D4: Total routes and members per year by province



Note: Each line represents a total across all taxi associations in the province. Each province has its own transport department. Taxi associations register routes and members in the provinces where the route originates. Nearly invariably, associations only appear in one province. In this figure, the information on when members or routes were first registered on the NLTIS is used to create cumulative provincial totals in each year.

Figure D5: Taxi routes (L) and Association route-network centroids (R) in Gauteng



Note: Different colours (circles) represent different taxi associations. The **left hand** figure presents the result of a preliminary geo-coding exercise of the NLTIS data for Gauteng. Route origins and destinations are carefully cleaned and then passed through the google-directions API to generate a shortest driving route. Different colours indicate different taxi associations, although it is often the case that multiple associations have operating licenses for the same route segment. The **right hand** figure plots the centroids of the network of origin and destination nodes of each taxi association and scales these by the number of routes of each association. There are 219 taxi associations operating in Gauteng and its immediate surrounding municipalities which have been included in the sample.

D2 Market structure summary statistics

Table D3: Competition based on route origins and destinations (N=474 associations)

Measure of competition	Mean
Other association(s) on at least one route	0.58
Number of routes with with presence of other associations	3.10
Share of routes with precence of other associations	0.22
Total number of other associations present on all routes	2.65

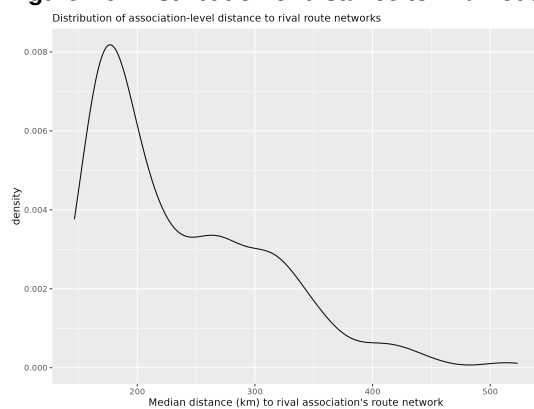
Note: Associations are matched based on their routes. An association matches with another association if at least one of their routes have the same origin and destination. Origins and destination latitudes and longitudes are rounded to 2 decimal places (1.11km).

Table D4: Competition based on route centroids (N=474 associations)

Distance from route network centroid	Proportion of associations with other associations within distance
1 km	0.19
2 km	0.25
5 km	0.45
10 km	0.60
25 km	0.87

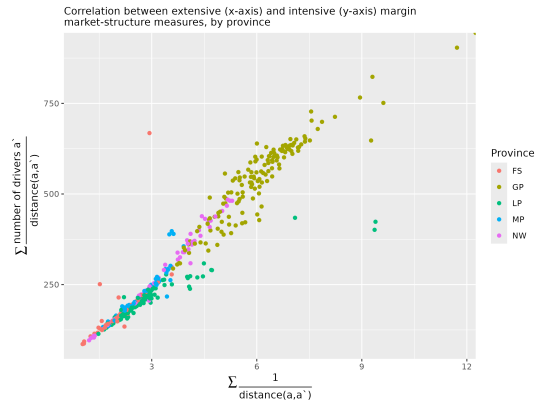
Note: Associations are matched based on their route network centroids. An association matches with another association if its route network centroid falls within the specified distance of another association.

Figure D6: Distribution of distance to rival route networks



Note: This distance measure is introduced in section 3.3. It is developed by geo-coding all associations' route origins and destinations, calculating the pairwise distances between all route-pairs of rival associations, then calculating the association-pair median across all route-pairs, and then creating an association-level mean of this measure across all rival associations. The data is presented for Gauteng and the four provinces surrounding it.

Figure D7: Correlation of market structure measures



Note: This figure presents the correlation between the two market-structure measures introduced in section 3.3. The extensive margin measure is the distance-weighted sum of all other associations, providing a measure of how many other associations are operating in an associations market. The intensive margin measure is the distance-weighted sum of all rival association drivers, providing a measure of how many rival drivers operate in an associations market. The data is presented for Gauteng and the four provinces surrounding it.

D3 Passenger Rail Agency of South Africa

Table D5: Overview of train performance on Gauteng Commuter Rail lines: Total on-time trains per line

line	area	2014	2019
De Wildt - Belle Ombre	north	1730	0
Hercules - Koedoespoorts	north	3037	6356
Hercules - Pretoria	north	2394	0
Koedoespoort - Pienaarspoort	north	1712	0
Koedoespoort - Pretoria	north	858	50
Mabopane - Belle Ombre	north	2699	1238
Mabopane - Pienaarspoort via CilPk	north	7003	3488
Pretoria - Belle Ombre - Mabopane	north	20694	11843
Pretoria - De Wildt	north	9654	6976
Pretoria - Pienaarspoort	north	16250	19109
Pretoria - Risik - Pretoria	north	0	0
Pretoria - Saulsville (Tshwane)	north	16245	8958
Rissik - Pienaarspoort	north	0	0
Saulsville - Pienaarspoort via Capital Pk	north	1403	1002
George Goch - Naledi	south	44049	16459
George Goch - Vereeniging (via MDY)	south	18388	4291
Germiston - Booyens - New Canada	south	5680	1112
Germiston - Kwesine	south	16658	8509
Germiston - Vereeniging (via MTN)	south	9986	4084
Johannesburg - Daveyton	south	26311	12889
Johannesburg - Leralla	south	27550	11379
Johannesburg - Oberholzer	south	4429	2601
Johannesburg - Randfontein	south	15525	11780
Johannesburg - Springs	south	15331	4197
Johannesburg - Tshwane (WITS)	south	15064	5328
Springs - Nigel	south	4977	2857
Springs - Oosrand	south	0	0
Total	All	287627	144506
Total	north	83679	59020
Total	south	203948	85486

Note: This table presents the total number of the trains-on-time in the PRASA administrative data, by rail corridor. The figures reported are the revised figures as captured by PRASA, summed over all months in the year.

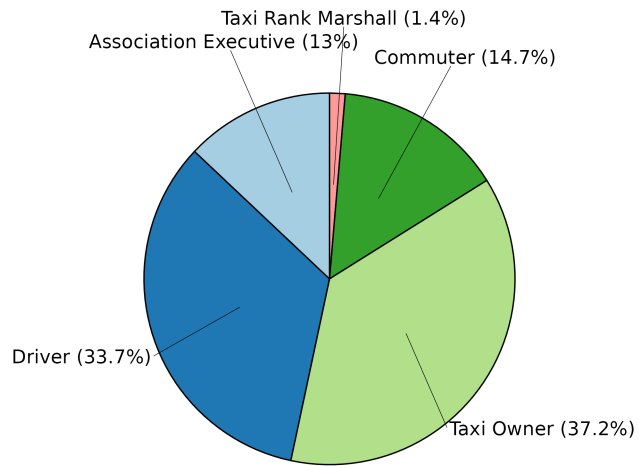
Table D6: Total ticket sales (thousands) per line – 2014 and 2019 only

line	2014	2019
Dewild - Rosslynn (De Wildt - Pta)	791	91
Doornfontein - President	2165	413
Fonteine - Oakmoor (Pretoria Exluded)	2308	566
Germiston - Leralla	8091	2791
Kutalo - Vereeniging (Via Meyerton)	1652	429
Mabopane - Belle-Ombre - Schuttestraat	3795	549
Mlamlankuzi - Duncanville (Via Midway Exl Vereeniging)	4910	1106
Naledi - Johannesburg	8272	2851
Pienaarspoort - Mears Street	3798	1882
Saulsville - Pretoria	3301	1202
Total	39082	11880

Note: This table summarizes the yearly total ticket sales in thousands for each line. The ticketing data and train operationality data are not reported in the same manner on the PRASA information system. As such, the line names presented here differ from those presented in the train operationality summary statistics. 2014 is the first full year for which data is available.

D4 Violence in the minibus taxi industry

Figure D8: Victims of taxi shootings



Note: This figure is based on mentions of victims in the set of news articles. One shooting can have multiple types of victims.

D5 Working conditions of taxi drivers

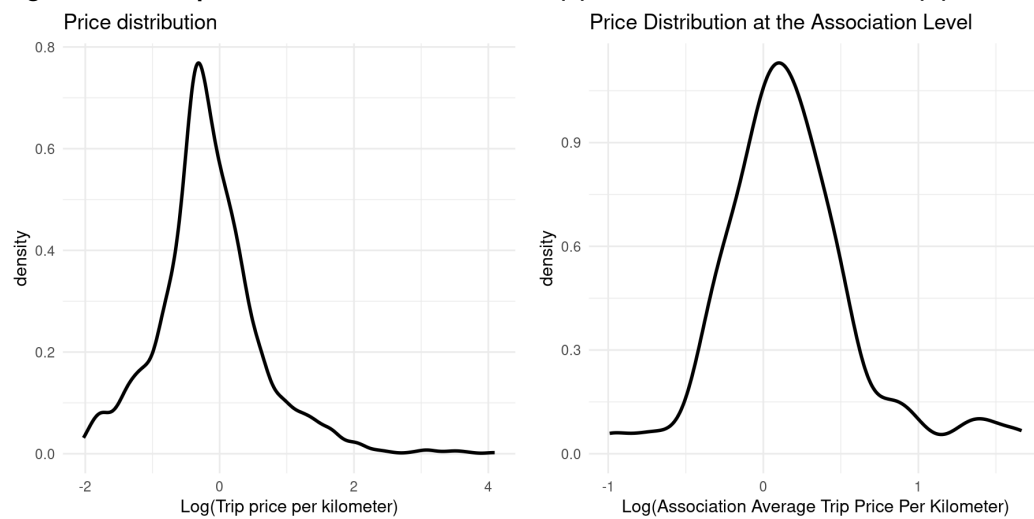
Table D7: Summary statistics: Working conditions of minibus taxi drivers

Variable	Non-taxi drivers			Taxi drivers		
	Mean	SD	N	Mean	SD	N
All employed respondents:						
Monthly earnings	7545	16938	50112	4784	6831	640
Registered for income tax	0.53	0.5	60825	0.05	0.22	775
Has a written contract	0.8	0.4	60825	0.13	0.33	775
Unemployment insurance contributions by employer	0.59	0.49	60825	0.077	0.27	775
Medical aid contributions by employer	0.29	0.45	60825	0.021	0.14	775
Member of a trade union	0.29	0.46	60825	0.068	0.25	775
Hours worked per week	43	13	71439	64	19	904
Years of potential experience	6.8	7.8	71440	4.7	5.8	904
Has a high school certificate	0.52	0.5	71440	0.37	0.48	904
Male	0.53	0.5	71440	0.97	0.17	904
Resident in an urban area	0.75	0.43	71440	0.73	0.44	904
Group = African	0.77	0.42	71440	0.96	0.19	904
Group = Coloured	0.11	0.31	71440	0.029	0.17	904
Low skilled employees only:						
Monthly earnings	3925	8504	25000	4821	7212	412
Registered for income tax	0.31	0.46	28549	0.041	0.2	512
Has a written contract	0.67	0.47	28549	0.11	0.31	512
Unemployment insurance contributions by employer	0.53	0.5	28549	0.066	0.25	512
Medical aid contributions by employer	0.1	0.3	28549	0.018	0.13	512
Member of a trade union	0.18	0.38	28549	0.062	0.24	512
Hours worked per week	43	15	34509	65	18	574
Years of potential experience	6.1	7.6	34510	4.9	6.1	574
Has a high school certificate	0	0	34510	0	0	574
Male	0.55	0.5	34510	0.99	0.12	574
Resident in an urban area	0.68	0.47	34510	0.75	0.43	574
Group = African	0.86	0.35	34510	0.96	0.2	574
Group = Coloured	0.12	0.32	34510	0.031	0.17	574

Note: Low skilled workers are defined as those without a secondary school leaving certificate. Taxi drivers are identified using occupation code 8320. "Taxi drivers, informal". The sample is restricted to respondents between 18 and 65 years of age. Observation counts vary across variables due to variation in missing responses across questions in the labour force surveys. Source: Pooled 2017-2019 Labour Market Dynamics Surveys (Statistics South Africa 2017, 2018, 2019).

D6 Collected price data

Figure D9: Price per kilometre at the Route level (L) and at the Association level (R)



Note: The data is for 2200 taxi routes run by 55 taxi associations for which price data was collected in Gauteng and surrounding provinces. Trip origins and destinations are geolocated using the Google API after careful cleaning and the shortest driving distance between them is calculated.

Figure D10: Price per kilometre and trip distance

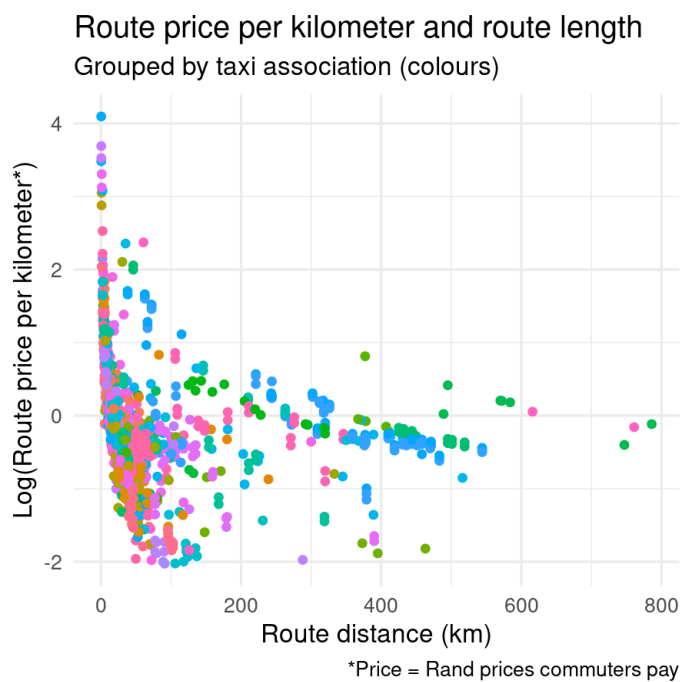


Table D8: Association joining fee survey

Variable	Mean	SD	Min	Median	Max	N
Rank fee per vehicle per month	1016	568	500	800	2580	63
Association Joining Fee	63514	20702	30000	60000	150000	74
Accept new members	0.83	0.38	0	1	1	92

Note: This table presents the results of a telephonic survey of 92 taxi associations for whom price data was collected. The taxi associations were contacted via the telephone numbers on their price lists, or on their facebook pages and asked what their monthly rank-fees, and their joining fees were. In addition, associations were asked whether they accept new members. Differing numbers of observations indicate the level to which associations did not respond to the survey questions.