Market Segmentation and Coordination Costs: Evidence from Johannesburg's Minibus Networks*

Oluchi Mbonu† (Job Market Paper)
F. Christopher Eaglin‡

October 28th, 2024

Click here for the latest version

Abstract

How does spatial market segmentation affect firms' ability to meet demand across space? We study the market for public transportation in Johannesburg, South Africa, where private associations of minibus owners segment the city into distinct territories. In contrast, the demand for urban mobility is inherently interconnected, with a quarter of commuter trips originating in one association's territory and ending in another's. Using GPS traces for over 40 million minibus trips and 9 million commuter trips, we present two complementary empirical results that quantify the frictions associations face on "across-territory" routes. First, we use an expected, cyclic mobility demand shock – the sharp increase in recreational mobility following monthly pay dates – to trace out the supply curve. The supply elasticity is close to 1 on routes contained within an association's territory but is significantly lower on across-territory routes (0.4). We estimate that if across-territory supply were as elastic as "within-territory" supply, then aggregate wait time for commuters would decrease by one million minutes per day, or approximately 4 minutes per trip. In our second exercise, we use exogenous fleet reductions due to bus breakdowns and repossessions to show that associations prioritize maintaining service on across-territory routes over within-territory routes, indicating that across-territory routes are more profitable at the margin. We next develop a model of minibus allocation to formalize the intuition from our empirical results. Our observed empirical patterns correspond to more convex costs on across-territory routes, reflecting the need for associations to coordinate with each other on these routes.

*We are grateful to Gabriel Kreindler, Emily Breza, and Ed Glaeser for their useful feedback and support throughout this project. We also thank Anhua Chen, Cassandra Cole, Lucas Do, Konstantin Poensgen, Diego Santa Maria, Yulu Tang, and participants at the Harvard Development and Labor/Public student workshops for useful feedback and suggestions.

†Harvard University (ombonu@g.harvard.edu)

‡Duke University (chris.eaglin@duke.edu)

1 Introduction

In many urban areas in developing countries, public transportation is dominated by decentralized networks of privately-owned minibuses. In cities across Africa, Asia, and Latin America, these networks account for between 30% and 80% of all mass transit in the region, filling in the gaps left by inadequate government-provided public transportation infrastructure (Vermeiren et al., 2015; Kumar, Zimmerman and Arroyo-Arroyo, 2021). In Johannesburg, South Africa, 38% of households use informal minibuses as their main mode of transportation; a higher share than any other travel mode, including private vehicles.

A common feature of these contexts is the organization of these minibuses into associations that segment the city into distinct territories. These regional associations separately control their routes and prices, restricting members' operations to the routes under their control.

Demand for urban transport, on the other hand, is not similarly fragmented. Cities are inherently interconnected, with commuters regularly traveling between these divided territories.[1] In Johannesburg, for example, 25% of all trips originate in one association's territory and end in another's.

In this paper, we study how spatial segmentation affects the supply of public transit, and analyze the resulting impacts on commuter mobility.[2] We study these issues in the context of Johannesburg, South Africa, where the minibus industry is divided among over twenty associations. We focus on the frictions that arise on "across-territory" routes which connect different association territories. To operate on across-territory routes, associations enter informal agreements about how to split operations; for example, each association agrees to a maximum number of minibuses they can allocate to the route. We examine whether coordination between associations creates inefficiencies in service provision by comparing supply on across-territory routes to supply on routes that require no inter-association coordination.

Our empirical analysis uses two rich datasets. On the supply side, we use novel data on minibus operations from a large minibus taxi financier in South Africa. The company provides insurance coverage for over 95% of the minibuses they finance, and they equip each minibus with a GPS tracking device. This enables us to observe over 40 million trips and more than 3,000 insurance claims for $19,214$ minibuses operating in Johannesburg between October 2022 and July 2023. By comparing our sample to published operating license

[1] Throughout this paper the term "commuter" is used broadly to refer to any individual traveling within the Johannesburg Metropolitan Municipality, and is not strictly limited to those traveling for work.

[2] Many papers have studied the efficiency of different types of decentralized transportation networks. For example, Chen (2024) examines the efficiency gains that occur from railroad company mergers due to reduced interchange costs. Brancaccio et al. (2023) look at the efficiency of decentralized transportation in oceanic shipping, and Rosaia (2020) studies network economies in ride-hailing services.

applications, we estimate that our data covers 48% of all minibuses operating in Johannesburg (see Appendix A.4). This granular data allows us to observe the daily operations of minibus networks at a scale and level of detail previously unavailable in the literature; we track minibus route patterns, service frequency, and temporal changes in service delivery. On the demand side, we construct commuter trips within Johannesburg using anonymized smartphone GPS data from over one million unique devices over the same time period.

Our analysis proceeds in three steps. First, we present our two main empirical results on supply response, which provide evidence of the presence of coordination frictions: i) associations prioritize maintaining supply on across-territory routes, suggesting that these routes are more profitable at the margin, and ii) short-term supply elasticity is lower on across-territory routes. Second, we develop a theoretical model that rationalizes these patterns through more convex operational costs on across-territory routes. Finally, we quantify the impact of market segmentation on commuters by estimating the additional wait times they face on across-territory routes.

For our first step, we examine how associations respond to exogenous and unexpected decreases in fleet size, focusing on how these shocks impact allocation decisions on within- and across-territory routes.[3] We use minibus breakdowns, accidents, and repossessions (observed in the insurance claims data) as shocks that reduce the number of minibuses available to the association.

Using a differences-in-differences design, we first quantify the impact of these incidents on fleet size. We find that if all minibuses in an association's fleet have an incident (as measured by the insurance claims data), the association's fleet size decreases by 32% on average. This reduction in fleet size leads to differential decreases in minibus allocation across route types. The same incident shock causes an average decrease of 11% in the number of minibuses on across-territory routes. In contrast, on within-territory routes there is a much larger decrease of 32%; associations allocate the shortfall more to within-territory routes. These results reveal a clear pattern: when faced with resource constraints, associations prioritize maintaining service on across-territory routes. This suggests that these routes are more profitable at the margin, and are under-serviced in equilibrium. We provide evidence for this interpretation by documenting higher per-kilometer fares on across-territory routes and ruling out alternative explanations such as strategic market share protection or binding agreements that prevent reductions in service.

Our second empirical exercise traces out the short-term supply elasticity using cyclical, anticipated demand shocks. We estimate the impact of an increase in demand on the supply of minibuses. To identify the supply response, we leverage variation in the demand for

[3] Associations allocate minibuses to specific routes, usually on a weekly schedule.

mobility induced by monthly pay cycles. In South Africa, most formal workers are paid on the 25th of the month; we show that residents take 30%-50% more recreational or leisure trips at the end of the month when their liquidity increases after getting paid. We use this spatial and temporal variation in mobility to construct an instrument for aggregate mobility by interacting an indicator for whether or not the date falls within the "end of the month" period, and whether a given route is "recreational".

Estimating this flexibility presents some econometric challenges. First, both our dependent and independent variables – the number of aggregate commuter trips and the number of minibuses – are counts containing numerous zeros and with long right tails. A linear model is thus inappropriate for both the first and second stage of our instrumental variable estimation. Using a logarithmic transformation is also not feasible given the presence of zeros in our data. Second, because we are using an instrumental variables approach, this non-linearity in the first stage precludes the use of two-stage least squares for estimating the causal relationship. To address these challenges, we employ two empirical methodologies to estimate the non-linear relationship with instrumental variables. The first methodology involves a non-parametric estimation approach outlined by Chen, Christensen and Kankanala (2024). The second uses a generalized control function approach proposed by Wooldridge (2014). Both methods produce similar estimates and point to a consistent pattern of results.

We find that across all route types, a 1% increase in demand increases the number of minibus trips on a route by between 0.85% and 0.96%, depending on the specific methodology used. However, this elasticity varies significantly by route type; across-territory routes have a markedly lower elasticity, with estimates ranging between 0.41 and 0.56. In contrast, on within-territory routes, the elasticity remains close to 1, ranging from 0.93 to 0.98.

We present descriptive evidence showing that this reduction in flexibility is likely because associations are less able to increase the number of minibuses on across-territory routes compared to within-territory routes. At the end of the month, the number of minibuses on non-recreational routes decreases, and the number of minibuses on recreational routes *within-territory* increases. However, on recreational routes *across-territory*, the number of minibuses does not increase significantly. This pattern suggests that associations can readily reallocate minibuses to meet changing demand patterns within their controlled territories, but they face restrictions in adding minibuses to across-territory routes – even when the demand increase is anticipated.

In sum, coordination frictions result in inefficiently low and rigid supply on across-territory routes. Associations prioritize these routes when there is a reduction in fleet size because they are more profitable at the margin; however, revenue-sharing agreements prevent associations from expanding service on across-territory routes even when demand increases.

We next develop a simple model of minibus ridership and allocation to formalize the implications of the two empirical patterns. In this model, associations allocate minibuses across the routes they control in order to maximize profits. Associations face allocation/operational costs that may vary by route type. Commuters traveling on a route decide whether to travel using a minibus taxi or some other outside option, trading off wait time – which is inversely proportional to the number of minibuses allocated to the route – and the minibus fare. This framework allows us to generate comparative statics related to our two empirical observations: i) the effect of overall fleet size on route allocation, and ii) the effect of aggregate demand on the number of minibuses allocated. In the model, if the operational costs on across-territory routes are more convex than on within-territory routes, we generate the same patterns observed empirically: a larger decrease in allocation on within-territory routes after a shock to fleet size, and lower flexibility on across-territory routes. More convex costs means that market segmentation gives rise to inefficiencies in supply on across-territory routes.

Finally, we quantify the costs to commuters in Johannesburg resulting from the segmentation of the informal public transportation network, focusing on the impact on wait times. We estimate by how much wait times would reduce if the supply elasticity on across-territory routes matched that on within-territory routes. To approximate wait times, we use the time between minibus departures, i.e., the minibus headway. Assuming uniform arrival of passengers over a given time period, the average wait time for each passenger is half of the minibus headway.

We estimate headway elasticities using our cyclical mobility instrument, with average minibus headway as the dependent variable. When aggregate mobility increases by 1%, headway on within-territory routes decreases by approximately 1.2%; on across-territory routes, headway decreases by only 0.3%.

By combining the estimated headways with approximations for the total number of minibus passengers on across-territory routes, we estimate the wait time savings that would accrue if the headway elasticity on across-territory routes (-0.3 in the status quo) mirrored the elasticity of -1.2 observed on within-territory routes. The results are economically significant: on recreational across-territory routes alone, passengers collectively lose a total of $988,142$ minutes each day due to the reduced flexibility. This corresponds to approximately four minutes of extra wait time on each trip.[4] Assuming a value of time between 50% and 100% of the average hourly wage for a minibus user, this translates to between R$757,576$ and R$2,852,437$ (\$$46,030$ and \$$173,313$) lost each day.[5]

[4] The average trip duration is 45 minutes

[5] This value of time approximation is based on estimates from the literature. Recent papers suggest that

While association-based organization may provide benefits in terms of local coordination, route management, and pricing, our findings show that the segmentation of these minibus networks creates significant inefficiencies in service provision on routes that connect association territories. These frictions impose significant costs on commuters, increasing wait times by an average of four minutes per trip on these across-territory routes. As in many settings with decentralized transportation networks, there is an inherent trade-off between market power from centralization, and the inefficiencies that arise from decentralized supply.

Related Literature

This paper contributes to four broad strands of literature: i) the (in)efficiency of decentralized transportation, ii) the impacts of public transportation infrastructure, iii) optimal transportation infrastructure, and iv) the political economy of transportation design.

First, there is an extensive literature on the efficiency of decentralized transportation infrastructure. Brancaccio et al. (2023) study the inefficiencies that arise in decentralized transportation networks, concentrating on oceanic shipping, Rosaia (2020) considers whether there are network efficiency gains in ride-hailing apps, and Buchholz (2021) and Fréchette, Lizzeri and Salz (2019) study matching frictions in the taxi industry. Chen (2024) studies the efficiency gains from mergers of U.S. railroad freight companies. He finds that most of the gains arise from the elimination of interchange costs between railroad companies – another type of coordination cost arising from segmented markets. There is also a growing literature that considers the impacts of political decentralization on transportation infrastructure. For example, Bordeu (2023) studies how municipal fragmentation leads to over-investment in transportation infrastructure close to municipal boundaries, and under-investment in the core of the municipalities. A report by the OECD also highlights how decentralization to authorities at the sub-national level can lead to efficiency losses and asymmetric outcomes across local governments (OECD, 2019). We contribute to this literature by highlighting a private transportation market where decentralization leads to inefficiencies in supply – the minibus taxi network. We also empirically quantify these costs and their impacts on consumers.

Second, many studies have examined the impacts of public transportation infrastructure on various economic outcomes (Gibbons and Machin, 2005; Glaeser, Kahn and Rappaport, 2008; Donaldson and Hornbeck, 2016; Baum-Snow et al., 2017; Gonzalez-Navarro and Turner, 2018; Donaldson, 2018; Billings, 2011; Heblich, Redding and Sturm, 2020)). Despite the dominance of informal transportation in many developing countries, most existing research on mass transit in these contexts focuses on formal public transportation systems. Studies have

these valuations may be even larger. See section 9 for more details.

examined the impacts of subways and trains (Gu et al., 2021; Zárate, 2022) and bus rapid transit (BRT) lines (Majid, Malik and Vyborny, 2018; Gaduh, Gračner and Rothenberg, 2022; Tsivanidis, 2019) on outcomes such as congestion, commute times, pollution levels, and overall welfare.

Recently, two studies in economics have explored the relationship between informal minibus networks and commuter welfare. Conwell (2024) examines how policymakers can improve upon informal public transportation provision, given that minibuses and passengers do not internalize their spillovers on overall wait times. By fitting an urban model of transportation demand, he finds that governments should subsidize the operation and use of minibus taxis on specific routes to correct for these externalities. The second paper, Björkegren et al. (2024), studies how informal transportation in Lagos, Nigeria, responds to the expansion of state-planned formal buses. They find that minibus departures fall following the entry of more formal buses, and minibus fares also fall on routes with cheaper formal bus options. They then estimate the welfare effects of formal bus expansion, accounting for the informal transportation response. We contribute to the nascent literature on minibus taxi operations by examining how the organizational structure of the minibus taxi industry affects its supply and operations, and the subsequent costs to commuters.

Third, there is extensive research on the optimal design of transportation infrastructure networks. Studies have primarily focused on transportation infrastructure in developed economies, looking at optimal road (e.g., Fajgelbaum and Schaal (2020), Allen and Arkolakis (2022)), and multimodal transit (Almagro et al., 2024; Wong and Fuchs, 2022) networks. More recently, Kreindler et al. (2023) simulate optimal bus transportation networks in Jakarta, Indonesia, using estimated commuter preference parameters. By documenting the costs related to the prevalent design of informal minibus networks, we contribute to the literature on how governments should organize their public transportation infrastructure.

Fourth, our study is also relevant to the literature examining the impact of political economy on transportation investment and design. In Johannesburg, minibuses are required to organize into associations by government mandate. Association territories are also closely related to historical apartheid era townships (see Appendix A.3). This aligns with studies such as Fajgelbaum et al. (2023), who explore how political preferences shaped the placement of train stations on the California high-speed rail, and Glaeser and Ponzetto (2018) who show that voter perceptions of infrastructure costs can lead to over- or under-investment in transportation projects depending on which costs are salient.

The rest of the paper is organized as follows. Section 2 describes the minibus taxi industry in Johannesburg. Section 3 describes a simple conceptual framework. Section 4 describes our data and data processing. Section 5 describes the mobility patterns of minibuses and commuters. Sections 6 and 7 describe our empirical methodology and results. Section 8 presents our theoretical framework and model implications. Section 9 presents our findings on the impacts of segmentation on commuter wait times, and section 10 concludes.

2 Informal Transportation in Johannesburg: The Minibus Taxi

In South Africa, the most prevalent form of informal mass transit is the minibus taxi. These are privately operated 16 to 18-seater buses that serve as the primary mode of transportation for households in Johannesburg, particularly for low-income commuters. The 2020 National Household Travel Survey (NHTS) shows that in Gauteng province, where Johannesburg is located, minibus taxis accounted for 87% of all public transportation trips.[6] Moreover, 38% of households stated that the minibus taxi is their main mode of transportation (see Figure A.1). This represents the largest share among all transportation modes in the survey, including the use of private vehicles. Figure A.2 illustrates that low-income households rely heavily on minibus taxis for their travel needs; as income increases, households substitute away from the minibus taxi and walking, towards private vehicles.

Minibuses operate an unscheduled service on mostly fixed routes. They begin their route at a minibus stop, known as a taxi rank, wait until the minibus is at capacity, and then ply their designated route, picking up, and dropping off passengers along the route. Minibuses queue at taxi ranks by route, and can only start to fill up with passengers after the minibuses ahead of them have departed.[7,8] Minibuses are largely uniform in size and capacity. Almost all are 16-18-seater buses, and South African legislation prohibits minibuses from having more than 16 seated persons while in operation. Notably, it is not the case that larger minibuses service busier routes. This uniformity in size applies across all routes, regardless of their popularity or passenger demand.

Minibuses are organized into associations that control entry, pricing, and routing. In Johannesburg, associations are regional and control the routes within their geographic territory. Figure 1 displays a map of inferred association territories within the Johannesburg metropolitan municipality, constructed using our GPS data on minibus taxi operations.[9]

[6] Surveys were carried out from January 2020 to March 2020.

[7] Different routes and drivers assign queue positions differently, some queue on a first come first serve basis, while others have a rotation for queue position each day.

[8] The vast majority of minibuses operate using this model. However, a minority of minibuses "tout" for passengers instead. When touting, minibuses drive around and look for passengers to transport locally, as opposed to queuing at a minibus rank.

[9] Details about this data are outlined in Section 4.

Each association allocates minibuses to the routes they control based on a centralized schedule; in most cases, associations rotate minibuses weekly so that all owners have access to the most profitable routes. Route prices are set and reviewed annually by the association, and drivers cannot adjust them at will.[10] In order to operate a minibus in South Africa, you must be a member of an association. This is written into the law in South Africa, but is enforced by the associations themselves, usually through threats of violence.[11]

The organization of minibuses into associations is not unique to South Africa. In many contexts where privatized minibuses make up an important share of mass-transit, these buses organize into more centralized cooperatives. For example, in Kampala, Uganda, minibus taxi associations control different routes, and minibus owners must register with an association in order to operate a route (Ndibatya and Booysen, 2020). In Nairobi, Kenya, minibuses must be a member of a Savings and Credit Cooperative (SACCO) in order to operate. Each SACCO controls operations on a subset of routes (Kerzhner, 2022; Kelley, Lane and Schoenholzer, 2018). In Accra, Ghana, minibus owners are members of unions which set route prices centrally and manage the minibus terminals (Saddier and Johnson, 2018). Thus, the dynamics outlined in this paper are not unique to our setting; the informal minibus industry is often comprised of cooperatives that segment the market.

The geographic segmentation of Johannesburg by minibus associations means that associations must coordinate with each other in order to service a route from one association's territory to another (across-territory routes); the associations must enter what they call a "joint venture". They agree together on the price of the route and how they will share revenue. For example, some joint ventures stipulate a cap on the number of minibuses each association can have servicing the route. Another common contract is to allow each association to pick up passengers in their own territory only. That is, they can only transport passengers one-way and must drive back empty.

3 Conceptual Framework: Coordination Costs and Supply Efficiency

To contextualize our analyses, consider two potential states of the world for "across-territory" routes that require associations to coordinate:

[10] For more details on the minibus taxi industry, see Appendix A.3

[11] The minibus taxi industry in Johannesburg is known to be very violent. Associations engage in "turf wars" for control of routes, and often use violence to maintain market share. Operating a minibus taxi without the express permission/protection of an association is ill-advised.

1. Efficient Coordination Associations achieve efficient supply through effective coordination, despite the decentralized market structure. Associations are profit maximizing firms that think carefully about pricing, entry, and allocation. In this world, they are also able to coordinate without frictions and reach optimal supply on across-territory routes.

2. Coordination Frictions Coordination frictions lead to inefficiencies in supply, resulting in suboptimally rigid or low supply on across-territory routes. Here, frictions may arise from difficulties in implementing or enforcing efficient terms, the presence of hidden information, or the presence of transaction costs.

In this paper, our empirical exercises will help us to distinguish between these two scenarios and quantify the magnitude of coordination costs, if present. We will use within-territory routes as a benchmark; on these routes, inter-association cooperation is not required. We will present a model that formalizes the intuition of our empirical exercises. Through the lens of the model, we can interpret our empirical results as a function of the cost structures on within- and across-territory routes.

4 Data

We use three main datasets in this paper: a census of minibus taxi routes, data on minibus taxi operations from a large minibus financing company in South Africa, and data on aggregate commuter flows, constructed using anonymized smartphone location data. For most of our analyses, we use data observed between October 2022, and July 2023.

We divide Johannesburg into 1km-by-1km grids and aggregate all geographic data to this level. A route is defined as an origin-destination pair where each origin and destination is a distinct grid cell.

Minibus Taxi Routes Our sample of interest consists of all potential minibus taxi routes. We define these as the permutations of origins and destinations where taxi ranks exist, regardless of whether these routes are in operation. We use taxi ranks because they serve as the primary infrastrucure for minibus taxi operations, effectively determining where routes can start and end.[12] This approach allows us to capture all operationally feasible routes given the existing infrastructure. To construct this sample, we use a census of all minibus taxi routes within Johannesburg in August 2022, collected by WhereIsMyTransport (WIMT), a

[12]By focusing on the existing minibus infrastructure, we do not consider any longer-term dynamics that could involve infrastructure investments and changes to the route network.

mobility technology company based in South Africa. WIMT field workers collected this data by traveling on the minibus routes multiple times, collecting both the GPS path and the fare charged. In total, they documented 3,861 minibus routes in the Johannesburg area. Figure A.3 plots a map of these routes. Using this data, we identified the universe of grids containing minibus taxi ranks – these are the grids at the origin and destination of each route. 319 grids in Johannesburg contain a minibus taxi rank, representing 17% of all grids. Figure A.4 displays the location of these grids. This gives us 101,442 unique minibus taxi routes in our sample.

In our setting, some routes are completely within one association's territory, while others connect two associations' territories. We refer to these as "within-territory routes" and "across-territory routes" respectively. To classify our minibus routes into one of the two categories, we start by first assigning each grid in the Johannesburg Municipality to an association. For each grid, we calculate the total number of minibuses per association that started a trip in that grid during our observation period (see the section below for details on minibus trips). A grid is categorized as belonging to an association's territory if the plurality of minibuses that started a trip in that grid belong to that association. The left panel of Figure 1 plots the association territories created through this process. We further restrict the definition of an association's territory to create a more conservative measure of an association's domain of control. We only consider a grid to be under the "control" of an association if i) there were at least 15 unique minibuses starting a trip in the grid during our time period, and ii) the plurality share of minibuses is greater than 22%.[13] As such, we only consider a grid to be controlled by an association if we have enough observations to make this determination, and if the association has a large enough share of operators in the grid. The right panel of Figure 1 plots the restricted association territories.

With each grid assigned to an association's territory, we classify a *route* as being "within-territory" or "across-territory" depending on the association of the origin and destination grids. We use our more conservative definition of territories. A route is classified as being "across-territory" if the origin grid and destination grid are assigned to different territories. All other routes are classified as being "within-territory". This includes: i) routes where the origin and destination grids belong to the same territory, ii) routes where either the origin or destination grids do not belong to any territory, and iii) routes where both the origin and destination grids do not belong to any territory. Using this definition, 42,546 routes, or 42% of routes, are classified as being across-territory routes. Columns (3) and (4) of Table A.1 provide some statistics of within-territory routes and across-territory routes. Within-territory routes are on average 4km shorter than across-territory routes, but there is

[13] These cut-off points represent the 25th percentile of both distributions.

significant overlap in the distribution of route distances (see Figure A.5). All our empirical exercises control for route length.

Minibus Taxi Operations Our data on minibus operation comes from a large minibus financing company. The company provides asset-backed loans for the purchase and operation of minibus taxis in South Africa. As part of their risk management, they outfit each minibus they finance with a GPS tracker. The company also provides insurance to 95% of the vehicles they finance. We have access to pre-processed trip data and insurance claims data for $19,214$ vehicles in operation in Johannesburg between October 2022 and July 2023. We also have data on the association to which the minibus belongs.

Our data partner first processes the raw GPS data into "trips" before sharing this data with us. Each trip has a start and stop latitude and longitude, time stamps, and total distance traveled. They provide data on $40,124,891$ trips within Johannesburg.[14] The endpoints of the pre-processed data corresponds roughly to locations where i) the minibus has regularly been observed to stop for more than five minutes over the full dataset of GPS readings, and ii) the minibus's calculated average speed was less than 5km/h as they passed through the location.[15] There are several start and stop locations in this pre-processed data that do not correspond to a minibus taxi rank grid. This can happen for several reasons: for example, at gas stations or busy intersections. We thus process this data to generate trips on our routes of interest by aggregating up to our minibus taxi route level. We define "stop points" for each minibus taxi which correspond to the start or stop location of a minibus as classified by our data partner. We classify a minibus taxi as taking a trip on a minibus route o-d if:

1. A stop point for the minibus intersects with a rank grid, o, and
2. Rank grid d is the first minibus rank grid with which a subsequent stop point intersects, within a 6-hour time span.

Note that we do not allow $o = d$. This algorithm throws out any intermediary "stops" the minibus taxi made, as defined by our data partner, if it does not intersect with a minibus rank grid in our data. After this procedure, we have $7,124,646$ trips on $31,433$ minibus routes. Column 2 of Table 1 and column 1 of Table A.2 provide statistics on the minibus route trips. Our final dataset includes zeros for all route-day pairs with no minibus trips. As such, we have 30,838,368 observations corresponding to 101,442 minibus routes over 304 days of observation (October 2022 to July 2023).

We use insurance claims data to identify incidents that may affect a minibus's ability to

[14] We define a trip as being within Johannesburg if both the origin and destination grids are within the border of the Johannesburg Metropolitan Municipality

[15] For more information about the specific company algorithm, see appendix A.5

operate. There are $3,475$ incidents that take place during our study period, impacting 14% of the minibuses in our sample. The data includes the date of the incident and the claim cost for each incident. 66% of incidents have accompanying claim descriptions. We use these descriptions to classify incidents into different types: 33% of all incidents are described as repossessions, and 23% are classified as accidents.[16] We further define a separate category – "severe accidents" – which corresponds to accidents with claim costs exceeding the median claim cost for all accidents. Severe accidents make up 12% of our claims data.

Our minibus data is broadly representative of the minibus taxi industry as a whole. In South Africa, all applications for minibus taxi operating licenses are gazetted. We digitize these published applications and find that our data set contains 95.2% of all associations that applied for an operating license between March 2017 and March 2020.[17] Figure A.17 shows that our data is also broadly representative of association shares in the industry. See Appendix A.4 for more details.

Aggregate Commuter Flows To study overall mobility patterns in Johannesburg, we purchased anonymized smartphone GPS data from a mobility technology company. This data contains a selection of smartphone GPS pings from devices in operation within South Africa between October 2022 and July 2023. In total, we have $373,715,157$ smartphone pings from $1,087,889$ unique devices in Johannesburg during this time period. See column (1) of Table 2 for other statistics from this raw data.

We translated these smartphone pings into trips using a modification of the algorithm developed by Kreindler (2023). The algorithm combines sequential GPS pings within a distance and duration threshold into "stays" or "trips". After running the algorithm, we have approximately $9,049,991$ trips within Johannesburg from $516,558$ unique devices (see column (2) of Table 2). We also restrict our data to subset of trips that are on minibus taxi routes – these are trips whose origin and destination intersect with a minibus rank grid with $o \neq d$. There are $510,918$ trips on $33,198$ minibus taxi routes. Column (3) of Table 2 and column (2) of Table A.2 provide statistics on aggregate commuter trips on minibus routes.

In general, higher income quartiles of the population are over-represented in the smartphone data (see Figure A.19). As a robustness check, we re-weigh our trip data using the 2010 South Africa census as outlined in appendix A.6 to create a more representative sample of aggregate trip flows within Johannesburg. We then re-run our main analyses using this representative sample. The pattern of results is similar but the estimates are more noisy.

[16] Accident descriptions include: "multiple vehicles", "single collision", collision with animal", "collision with building", "collision with object", and "collision with pedestrians". The remaining 10% of incidents are not classified as either repossessions or accidents, for example, "Fire" and "Hail damage"

[17] Gazettes were no longer published during the COVID pandemic.

5 Mobility Patterns and Minibus Services in Johannesburg

5.1 Minibus Taxi Prices

Associations set the prices of routes centrally and do not change them in the short-term. Instead, prices are reviewed annually and the association decides whether to increase fares or leave them as is. Fares are more expensive on across-territory routes than on within-territory routes after controlling for route length (Table A.3). This suggests that associations pass through some of the costs of operating on these routes to commuters.

Minibus taxi routes are also more expensive on average than government-provided BRT buslines. Figure A.6 plots the average price of both binned by total route distance. Government-provided buses are subsidized but service far fewer locations than minibus taxis.[18]

The average minibus taxi fare is between 8% and 16% of the average user's hourly wage. Transportation costs in South Africa are high relative to household incomes. Direct transportation costs are approximately 17% of average wages, and this share increases to 37% for households in the lowest income quintile (Shah and Sturzenegger, 2022).

5.2 Where do Minibuses Operate?

Minibus operations are concentrated within their association's regional territory and between these territories and the Central Business District (CBD).[19] Figure 2 illustrates this pattern of operation for a selection of associations within Johannesburg. Each dot on the map represents the starting point of a minibus taxi trip in March 2023, color-coded by association.[20] Panels (B), (C), and (D) highlight the operations of three different associations individually. The pattern is stark – minibuses mostly operate within a confined region (their association territory), and in the CBD.

Regression analysis at the day-route level confirms this pattern. We classify routes as either within-territory or across-territory and use Pseudo-Poisson Maximum Likelihood (PPML) to estimate:

[18]Government provided buses and trains are the main mode of transportation for only 2.31% and 2.35% of households, respectively. The modal reason households gave for not using these transportation modes more often was their lack of availability (see Figure A.7).

[19]The CBD region is not controlled by any one minibus association. Most associations have at least one route from their territory into the CBD (and back).

[20]This map is generated using data described in Section 4.

$$\text{minibus\_trips}_{rt} = \exp\left(\nu_o + \eta_d + \gamma_t + \text{route\_length}_r + \text{within\_terr}_r\right) \tag{1}$$

Where $\text{minibus\_trips}_{rt}$ is the number of minibus trips on route r and day t. ν_o are origin fixed effects, η_d are destination fixed effects, and γ_t are day fixed effects. route_length_r is the length of route r (i.e., the distance between o and d).

Column (2) of Table A.4 displays the results. A 1km increase in distance reduces trips by 15%. This decay over distance is the same order of magnitude when we compare it to aggregate commuter trips (column (3) of Table A.4); a 1km increase in distance reduces commuter trip by 22%. However, there are 1.22 log-points more trips on within-territory routes than on across-territory trips – this is approximately 238% more trips within- than across-territory. The difference is an order of magnitude higher than for aggregate commuter trips (all modes of travel); commuter trips are 18% higher within-territory than across-territory.

Though minibus trips are concentrated within their association's territory, each minibus still operates on several different routes over the course of month, exemplifying the inherent flexibility of the sector. The median minibus in our data operates on an average of 4.62 unique routes each month, and has operated on 13 unique routes over our 10-month study time period. The same minibus can operate on both within- and across-territory routes. For the average minibus, 15% of all trips are on across-territory routes.

5.3 Short-term Shocks to Supply and Demand

Short-term shocks are common in the minibus industry, occurring on both the supply and demand sides. We take advantage of both kinds of shocks to study associations' response.

Shocks to Minibus Fleets The number of minibuses available to operate fluctuates daily. While associations allocate minibuses to routes, the decision to operate and the frequency of operation are up to the minibus owner and driver. Various factors affect association fleet size: drivers may be sick or on vacation, or choose not to work; minibuses may breakdown, get in an accident, be repossessed, or be undergoing maintenance. Our data shows that on a typical workday, an average of 2,429 minibuses are operating in Johannesburg, and the variance is 410 – 18% of the mean.

We use exogenous shocks in fleet size from bus breakdowns, accidents, and repossessions to study how associations allocate these fleet reductions in section 6.

Cyclical Mobility Mobility in South Africa is highly cyclical. Residents take more trips at the end of the month after they have been paid – this is especially true for recreational/leisure trips. Most formal jobs in South Africa pay their employees on the 25th of the month. This liquidity infusion leads to higher aggregate demand for trips to (and from) retail, recreation and grocery locations. We can see this pattern distinctly using Google Mobility data for Johannesburg. Google published the percent change in mobility in 2022 compared to a baseline day in 2020, broken down by category of visit.[21] Figure 3 plots this percent change by day of the month after taking out day of the week fixed effects. That is, the coefficient on the 25th, for example, is the average percentage point difference in mobility between the 25th of the month and the 15th (the reference day), holding the day of the week constant. Retail, recreation, grocery, and pharmacy trips spike sharply on the 25th of each month, and then decrease slowly as residents spend down their salary. On average, the number of retail and recreational trips increases by 32% from the 24th to the 25th of the month, and by 51% for grocery and pharmacy trips. As we might expect, there is no such trend for trips to the workplace; commuters travel to work every day regardless of their liquidity.

We use this cyclical trend to construct an instrument to identify the minibus response to fluctuations in demand. The instrument and the assumptions needed for identification are described in detail in Section 7.1.1.

6 The Impact of a Shock to Fleet Size on Route Allocation

Our first set of results examines how associations respond to an exogenous decrease in fleet size. We use minibus breakdowns, accidents and repossessions, identified in the insurance claims data, as quasi-exogenous shocks that reduce the number of minibuses an association can allocate to routes. We analyze how these shocks impact the association's choices in allocating their remaining buses between across-territory and within-territory routes.

Our preferred specification uses all categories of incidents as a shock to vehicle operation, as this provides the most statistical power. We also present results using repossessions, accidents, and severe accidents as separate shock categories.

[21] The baseline day's mobility is the median value from the 5-week period between Jan 3 and Feb 6, 2020 for the same day of the week. See https://www.google.com/covid19/mobility/ for more details

6.1 "Zeroth Stage" Impacts of Incidents on Minibus Operation

We first study whether an insurance claim incident causes the probability of minibus operation to go down. To study this, we estimate the following equation:

$$\text{operating}_{vt} = \phi_v + \gamma_t + \beta \text{post}_{vt} + L_{>10,vt} + L_{<-10,vt} \quad (2)$$

Where operating_{vt} is an indicator for whether a minibus v is operating in week t, ϕ_v are minibus fixed effects, and γ_t are week fixed effects. Our variable of interest, post_{vt}, is an indicator for whether vehicle v had an incident in the previous 10 weeks before t. We also control for $L_{>10,vt}$ and $L_{<-10,vt}$ which are indicators for whether vehicle v had an incident 10 weeks after week t and ten weeks before week t respectively. β captures the average effect of having an insurance incident in the 10 weeks after the incident relative to 10 weeks before the incident.

The impacts of insurance claims on minibus operation are shown in panel A of Table 3 for the different type of insurance claims. Having any type of insurance claim decreases the probability of operation in the next 10 weeks by 12 percentage points. Repossession leads to a decrease by 26 percentage points, and severe accidents decrease the probability by 18 percentage points.

Figure A.8 plots results from the event-study version of equation (2) for different types of incidents. For incidents in general, we observe no pre-trends in the probability of operation, and there is sharp drop the week after an incident, persisting up to 10 weeks later. Accidents and severe accidents show a similar pattern, except that the operation probability steadily increases over time, as minibuses undergo repairs. Repossessions, however, show some evidence of pre-trends; minibuses that are repossessed are more likely to be operating approximately two months before the repossession. This pattern suggests that minibus operators are likely aware of impending repossessions and may increase their operation in an attempt to avoid losing their vehicle.

Minibuses that file an insurance claim are similar to those that do not on observable characteristics. In particular, they are not more or less likely to operate on within-territory routes (see Table A.5). Minibuses that are repossessed operate more on across-territory routes, belong to larger associations, and operate on fewer routes in total. Adding controls for these characteristics does not materially change our results (see Table A.6).

Minibuses do not operate on the same routes from week to week – they rotate between multiple routes based on a central schedule made by the association. To demonstrate this rotation in the data, we plot the proportion of a minibus's trips that are within territory against the proportion of trips that are within-territory in the subsequent week, residualized by the association size. Figure A.15 shows the results. The two are largely uncorrelated. Consequently, when a minibus becomes inoperable, its impact is felt at the association level as the entire fleet must be reallocated, rather than affecting a single fixed route. Changes in the supply of minibuses on any given route therefore reflect deliberate allocation decisions by associations rather than mechanical effects from specific minibuses going offline

6.2 "First Stage" Impacts of Incidents on Association Fleet Size

We next turn to examining whether the decrease in vehicle operation after an incident translates to a reduction in their association's fleet size. This may not happen if, for example, associations have a backlog of wait-listed operators who are ready to begin operating if a minibus goes offline.

To look at this first-stage impact, we estimate a similar equation to (2) above:

$$\text{No. Minibuses}_{at} = \exp\left(\phi_a + \gamma_t + \beta \text{post}_{at} + L_{>10,at} + L_{<-10,at}\right) \tag{3}$$

Our data is at the association week level. No. Minibuses$_{at}$ is the number of operating minibuses in association a's fleet in week t. ϕ_a and γ_t are association and week fixed effects respectively. post$_{at}$ is the total number of incidents that occurred to association a's minibuses in the 10 weeks before week t. We weight this count by association a's fleet size over our full study period. $L_{>10,at}$ and $L_{<-10,at}$ are the weighted number of incidents that occurred more than 10 weeks before and less than 10 weeks after week t respectively.

We estimate (3) using Pseudo-Poisson Maximum Likelihood (PPML). Using PPML together with the weighted treatment variables allows us to account for non-linearities that occur due to the varying association fleet sizes in our data.[22] Thus, $\beta * 100$ is the average percent decrease in the number of an association's minibuses in operation in the 10 weeks following an incident – relative to 10 weeks before – if all the minibuses in the association's fleet have an incident.

Panel B of Table 3 show the results. An association-wide incident leads to a decrease in fleet size of 32% on average over the next ten weeks. Repossessions lead to a decrease in fleet size of 70%,[23] and severe accidents lead to a 51% decrease. Figure 4 plots the event-study version of the regression.[24]

[22]We have 697 associations that have at least one trip in Johannesburg during our time period. The average association size is 20.7 minibuses and the standard deviation is 4.47

[23]Note here that repossessions do not have a similar pre-trend problem as in the zeroth-stage regression as they are not endogenous to association fleet size in the same way as minibus operation.

[24]Effects are less than 100% because an incident does not lead to a minibus going offline 100% of the time (see Section 6.1)

6.3 Impacts of Fleet Size on Route Allocation

Finally, we look at how the number of minibuses an association has in it's fleet affects its allocation to within- and across-territory routes. To do this, we instrument for fleet size using insurance incidents.

The reduced form effects of incidents on the number of minibuses on within- and across-territory routes are estimated using the below equation:

$$\text{No. Minibuses}_{atr} = \exp(\phi_a + \gamma_t + \beta_w \text{post}_{at} \times \text{within\_terr}_r + \beta_a \text{post}_{at} \times \text{across\_terr}_r + L_{>10,at} + L_{<-10,at}) \quad (4)$$

Where our data is at the association/week/route type (within vs. across) level. No. Minibuses_{atr} is the number of minibuses in operation for association a, in week t, on route type r, within_terr_r is an indicator for route type r being within territory, and across_terr_r is an indicator for route type r being across territory. All other variables are defined the same as in equation (3). Coefficients are estimated using PPML.

Panel C of Table 3 show the results. On average, an incident leads to a decrease in the number of minibuses operating on within-territory routes of 32%. The corresponding decrease on across-territory routes is 11% and the difference between within and across-territory allocation is statistically significant. The pattern carries through for all incident types – associations decrease the number of minibuses within territory more than they do across territory. Figure 5 displays the event study versions of these regressions.

We now turn to the instrumental variable specification. Our structural equation is given by:

$$\text{No. Minibuses}_{atr} = \exp(\phi_a + \gamma_t + \alpha \log(\text{fleet\_size})_{at} + \beta_w \text{within\_terr}_r \times \log(\text{fleet\_size})_{at} + L_{>10,at} + L_{<-10,at}) \quad (5)$$

Where we instrument for fleet size using the number of incidents for association a that occurred in the last 10 weeks before week t, weighted by the total size of the association. We add the controls for $L_{<-10}$ and $L_{>10}$ to continue to control for incidents that occurred more than 10 weeks before and after week t.[25] We estimate the equation using PPML-IV

[25] This is needed in the first-stage regression in order to retain the interpretation as the effect in the 10 weeks after the incident relative to the value 10 weeks before the incident.

that allows for the inclusion of fixed effects, following Lin and Wooldridge (2019).[26]

Table 4 presents the results of this estimation, using different incident types as instruments. Using all incident types as an instrument, when fleet size decreases by 1%, the number of minibuses allocated to across-territory routes decreases by 1.4% and the number of minibuses allocated to within-territory routes decreases by 2.1%. This difference (0.7%) is statistically significant.

6.4 Why do Associations Maintain Supply Across-Territory?

When an association faces a shock to its fleet size, it prioritizes maintaining supply on across-territory routes while decreasing supply on within-territory routes. We consider three potential explanations for this behavior: these routes could be more profitable at the margin, associations might want to maintain market share, or joint-venture agreements could be binding. While all three scenarios are consistent with coordination frictions impacting supply, we present evidence showing that higher marginal profitability is the most likely explanation in this context.

First, descriptive evidence supports the profitability mechanism: across-territory routes have higher fare prices per kilometer than within-territory routes, as shown in Table A.3.

Second, we examine whether associations' behavior reflects concerns about maintaining market share on across-territory routes. If associations reduced service out of fear that their partner would seize market share, we would expect to see lower variation in the total number of minibuses on a route compared to the variation in each association's individual contribution. Our data shows the opposite pattern: the distribution of standard deviations has a lower mean for individual associations than for the combined total (see Figure A.14). Additionally, when we regress one association's minibus allocation on its partner's allocation on across-territory routes, we find a coefficient of -0.4 (Table A.7). This is far from the coefficient of -1 we would expect if associations were perfectly cannibalizing each other's market share, suggesting that while some competitive response exists, it is not the primary driver of behavior.

Third, we investigate whether joint-venture agreements constrain associations' ability to reduce supply across-territory. Evidence from Section 7.3 shows that associations can and do remove minibuses from across-territory routes when demand is low; specifically they do so at the end of the month when demand on non-recreational routes decreases. This flexibility in reducing service suggests that joint-venture agreements do not impose binding downward constraints on supply. This ability to reduce service is logical - associations are likely more

[26]The procedure involves estimating first-stage (with FEs) using OLS, obtaining residuals, and then inserting the residuals into the second stage and estimating via PPML.

concerned about their partners expanding service and capturing additional revenue than about temporary service reductions.

Given these results, we conclude that the most likely mechanism is that across-territory routes are more profitable at the margin. This implies that they are also under-serviced in equilibrium.

7 The Impact of Aggregate Mobility on Minibus Supply

We estimate the supply response to short-term shifts in aggregate mobility and document how this elasticity differs by route type. We present results for two distinct econometric strategies and summarize the results of both methods in section 7.2. We also present descriptive patterns detailing the likely mechanism for the difference in flexibility on within- and across-territory routes.

7.1 Estimation Framework

To measure the average minibus supply elasticity, our ideal specification is:

$$\log(\text{supply}_{rt}) = \phi_r + \gamma_t + \beta \log(\text{aggregate\_trips}_{rt}) \tag{6}$$

where $\text{aggregate\_trips}_{rt}$ is the number of aggregate commuter trips on route r on day t. ϕ_r are route fixed effects, and γ_t are day fixed effects. The fixed effects absorb any differences across routes and any time-varying shocks across all routes. We consider two related measures of minibus supply, supply_{rt}. The first is the number of minibus taxis on the route – the extensive margin response. The second is the number of minibus taxi trips taken on the route – the intensive margin response. β is our measure of the elasticity. There are two reasons why we cannot estimate the model in equation (6) as-is in our setting. First, β is an equilibrium parameter and thus, in this model, it does not measure the causal impact of increasing aggregate demand on minibus supply. Second, both aggregate_trips_r and supply_r are counts which include 0. This means we cannot take logs of the variables in order to easily estimate an elasticity. Further, a linear model is likely to fit the data poorly given that our counts have a long right-tail (see Figure A.9 for the distribution of aggregate_trips_r and supply_r).

To deal with the first issue of endogeneity, we construct an instrumental variable which shifts the aggregate number of trips on a route. Our instrument uses the fact that mobility increases at the end of the month, when most residents of South Africa get paid. The use of an instrumental variable for identification further complicates our ability to carefully model the non-linear relationship between the two counts. Ideally, the first-stage of our estimation – the impact of the instrument on the aggregate number of trips – is non-linear. However, with a non-linear first-stage, two-stage least squares is no longer appropriate. We use two methods to estimate this non-linear relationship with instrumental variables. Both methods rely on slightly different assumptions, but result in similar estimates. The first estimates the relationship non-parametrically.[27] The second uses a generalized control function approach, sometimes called two-stage regression inclusion, proposed by Wooldridge (2014).

Section 7.1.1 details the construction of our instrument and the assumptions needed for identification. Section 7.1.2 presents our non-parametric estimation in detail, as well as the results from this strategy, and section 7.1.3 details the generalized control function approach, and presents the results. Section 7.2 summarizes the results from the two methodologies.

7.1.1 Instrumental Variable: Cyclical Mobility

We use variation in demand for mobility induced by monthly pay cycles to construct an instrument to estimate minibus flexibility. As described in section 5.3, mobility is highly cyclical in Johannesburg. Residents take more discretionary trips at the end of the month after they have been paid.

Based on these patterns, we posit that on recreational routes, at the end of the month, the aggregate number of trips increases exogenously and we can use this as an instrument to estimate the supply elasticity. We construct our instrument by first classifying routes in our sample as recreational and then defining high "demand days" at the end of the month. We classify a route as recreational if there is a large mall at either the origin or destination grid of the route. We use the Google Place of Interest (POI) database to identify all the malls within Johannesburg, and to extract the number of Google reviews each mall has. The number of reviews serves as a proxy for the size and popularity of the mall. We classify a mall as large if it is in the 75th percentile of the number of reviews. In our sample, this translates to having $2,959$ or more Google reviews.[28] Our aim is to isolate large malls that draw large numbers of visitors such that a trip originating or ending within the 1km grid has a high probability of being recreational. With this definition – having a large mall in either the origin or destination – we classify $26,640$ routes (26% of all routes) as recreational. Table A.1 displays the summary statistics of these routes as compared to other categories of routes in the data. In general, recreational routes do not look different on average from all routes. Figure A.11 displays the spatial distribution of grids with large malls within Johannesburg.

[27]Non-parametric instrumental variable models have been widely studied, and we closely follow the approach outlined in Chen, Christensen and Kankanala (2024).

[28]Figure A.10 plots the distribution of the total number of mall reviews for each grid.

Large malls are not concentrated in any one part of Johannesburg, though there are few large malls in minibus taxi grids in the outskirts of the municipality, especially in the South.

We first examine whether our aggregate mobility data replicates the cyclical mobility patterns observed in the Google mobility data. To do so we estimate the below model:

$$\text{aggregate\_trips}_{rt} = \exp\left(\phi_r + \sum_{d=1,d\neq 24}^{31} \beta_d \mathbb{1}\{\text{day}_t = d\} + \mathbb{X}_t\right) \tag{7}$$

where aggregate_trips$_{rt}$ is the total number of trips on route r and day t, and ϕ_r are route fixed effects. $\mathbb{X}_t$ include day of week, month, and holiday fixed effects. The top panel of Figure 6 plots the estimates of β_d for each day of the month (estimated using Pseudo-Poisson Maximum Likelihood). Mobility is indeed cyclical – the number of trips on a route are lower in the middle of the month, and there is a discrete jump on the 25th, when most workers in South Africa get paid. These results are averaged over all types of routes in our dataset. Next, we examine the added effect on recreational routes by estimating the below model:

$$\text{demand\_trips}_{rt} = \exp\left(\phi_r + \gamma_t + \sum_{d=1,d\neq 24}^{31} \alpha_d \mathbb{1}\{\text{day}_t = d\} \times \text{has\_mall}_r\right) \tag{8}$$

where γ_t are day fixed effects, and has_mall$_r$ is an indicator for whether the route has a large mall at the origin or destination, i.e., whether the route is recreational. The bottom panel of Figure 6 plots the estimates of α_d for each day of the month. There is an added mobility effect at the end of the month on recreational routes from the 25th of the month, until the 10th of the subsequent month.

Given these patterns, our main instrument is the interaction of whether or not a route is recreational with whether or not it is a high demand day :- demand_day$_{rt}$ $\times$ has_mall$_r$. We define a high demand day to be from the 25th of the month to the 10th of the next month. Table 5 shows the first stage impact of the instrument on aggregate mobility for both the linear specification using OLS, and the non-linear version which uses PPML. On recreational routes on demand days, the number of aggregate trips increases by 0.001 trips, or by 4.1%. Our assumption for identification is that this instrument does not affect the supply of minibus taxis except through its impact on the aggregate number of trips. In this setting, most minibus taxi drivers have daily or weekly rental contracts, and thus do not also experience a discrete jump in liquidity at the end of the month. It could be the case, however, that drivers' total household income increases at the end of the month, if, for example, their spouse gets paid on the 25th. Such a scenario would lead to a downward bias in our estimate of elasticity – drivers would want to work less at the end of the month. In our setting, almost all minibus taxi drivers are male, and in South Africa, men tend to be the main household breadwinners. As such, we believe that the assumption of exogeneity is plausible and we can use this instrument to identify the supply elasticity of minibus taxis.

Before proceeding to our formal econometric analysis, we present reduced-form estimates of our instrument's effects using PPML. While we cannot directly take the ratio of these coefficients to estimate elasticities due to the non-linear model, they provide transparent and intuitive preliminary evidence. The results in Table 6 show that on within-territory routes, the instrument increases supply by 3.3%, comparable to the 3% increase of commuter trips on these routes. In contrast, on across-territory routes, there is smaller supply response (2%) relative to the increase in commuter trips (7%). Taken together, these patterns suggest that the supply elasticity is much lower across-territory. These patterns persist when we control for route length, as shown in Table A.8. We now turn to more rigorous estimation approaches to precisely quantify these supply elasticities.

7.1.2 Method 1: Non-Parametric Instrumental Variables (NPIV)

To estimate the supply elasticity non-parametrically, we use a B-spline nonparametric instrumental variable estimator (see Chen and Qiu (2016); Chen, Christensen and Kankanala (2024)). In this setting, the identifying assumption is given by:

$$\mathbb{E}\left[Y - h_0(X)|Z\right] = 0 \tag{9}$$

where h_0 is some function of X, our endogenous variable. We can approximate h_0 using a linear combination of J cubic B-spline basis functions which have a vector of coefficients c_J. Our instrument can similarly be approximated using K cubic B-spline basis functions of Z, where $K > J$. Then we can estimate c_J using two-stage least squares or Generalized Method of Moments (GMM).

As proposed by Chen, Christensen and Kankanala (2024) (henceforth CCK), we use a data-driven approach to choose J, the number of B-spline basis functions used to estimate $h_0(X)$ (and consequently K, the number of basis functions for our instrument). This approach ensures that estimates of $h_0(X)$ and its derivatives converge at the best possible rate.

Our structural equation is given by:

$$\text{supply}_{rt} = h_0(\text{aggregate\_trips}_{rt}) + \phi_r + \gamma_t + \epsilon_{rt} \tag{10}$$

The key assumption for identification is that our instrument, route, and day fixed effects are exogenous. For this non-parametric estimation, we use a continuous version of our instrument to ensure that we are able to non-parametrically fit K B-spline basis functions of

Z. This instrument is an extension of our binary instrument – $\text{demand\_day}_{rt} \times \text{has\_mall}_r$. Instead of the binary indicator has_mall_r, which is 1 if there is a large mall at either the origin or destination of route r, we use mall_ratings which is the maximum number of mall ratings at the origin and destination of the route. mall_ratings is zero if there is no large mall at either the origin or destination.

The CCK approach to estimating data-driven number of splines does not extend to estimating linear fixed effects. Thus, as outlined in the paper, we also make the additional assumption that $\text{aggregate\_trips}_{rt}$ is conditional mean independent of route and day fixed effects given our instrument. That is: $\mathbb{E}\left[h_0(\text{aggregate\_trips}_{rt})|z_{rt}, \phi_r, \gamma_t\right] = \mathbb{E}\left[h_0(\text{aggregate\_trips}_{rt})|z_{rt}\right]$. With this assumption, our reduced form can be written as:

$$\text{supply}_{rt} = g(z_{rt}) + \phi_r + \gamma_t + e_{rt} \tag{11}$$

where $g(z_{rt}) = \mathbb{E}\left[h_0(\text{aggregate\_trips}_{rt})|z_{rt}\right]$ and $\mathbb{E}\left[e_{rt}|z_{rt}, \phi_r, \gamma_r\right] = 0$. We can then eliminate the fixed effects in a first stage by regressing supply_{rt} on route and day fixed effects and B-splines of z_{rt}. We then apply the data-driven approach using $Y_{rt} = \text{supply}_{rt} - \hat{\phi}_r - \hat{\gamma}_t$ as the dependent variable.[29]

Heterogeneity by Route Type We use the same procedure to estimate $h_0(X)$ separately for within-territory routes and across-territory routes. For this heterogeneity analysis, we also control for the interaction between the aggregate total number of trips, and the length of the route. This ensures that differences in elasticities between the two route types are not driven by differences in travel behavior based on the route length.

Here, our structural equation is given by:

$$\text{supply}_{rt} = h_0(\text{aggregate\_trips}_{rt}) + l_0(\text{aggregate\_trips}_{rt} \times \text{route\_length}_r) + \phi_r + \gamma_t + \epsilon_{rt} \tag{12}$$

Which we run separately for within-territory routes and across-territory routes. Our instruments are: $\text{demand\_day} \times \text{mall\_ratings}_{rt}$ and $\text{demand\_day} \times \text{mall\_ratings}_{rt} \times \text{route\_length}_r$. We again eliminate the fixed effects in a first stage and then use the data-driven approach with the residualized dependent variable.

Results

Using this procedure, our data-driven choice of B-spline segments is 5. Figure A.12 plots our estimate of $h_0(\text{aggregate\_trips}_{rt})$ over the full range of x values in our data. We focus

[29] See Chen, Christensen and Kankanala (2024) for implementation details.

our inference over the range $x \leq 4$ as this is where the mass of our data lies.[30] Figure 7 plots $h_0(\text{aggregate\_trips}_{rt})$ and the implied elasticity for $\text{aggregate\_trips}_{rt} \leq 4$. We calculate the elasticity using the non-parametric estimates for $h_0(x)$ and $h_0'(x)$.[31] The median elasticity of minibus trips is 0.85, and the median elasticity of the number of minibus taxis is 0.90. That is, on average, when the number of aggregate trips on a route increase by 1%, the number of minibus trips on that route increases by 0.85%, and the number of minibus taxis increases by 0.90%. We calculate standard errors via bootstrapping and cluster at the origin level.

Within- vs Across-Territory Route Elasticities The left panel of Figure 8 plots our estimated $h_0(x)$ for within-territory routes and across-territory routes evaluated at the average route length in our data (21.58km). The right panel plots the associated elasticities. The median elasticity of minibus trips is 0.94 for within-territory routes, and 0.55 for across-territory routes. The median elasticity of the number of minibus taxis is 0.98 for within-territory routes, and 0.55 for across-territory routes.

7.1.3 Method 2: Generalized Control Function (GCF)

Our second approach to estimating the supply elasticity follows from Terza (2009) and Wooldridge (2014). They propose a control function procedure with two steps. First, estimate a parametric model for $\mathbb{E}[x|z,w]$ with parameters θ, and define a control function $\hat{e}(\theta)$ which includes the standardized residual of the model. Second, estimate a parametric model for $\mathbb{E}[y|x,w,e]$, inserting $\hat{e}$ for e. The key assumption in this setting is that controlling for the residual of the first stage solves the endogeneity problem. In the case of a linear model, this assumption follows directly from the assumption of exogeneity of the instrument. However, if the model in the first step is non-linear, as it is in our case, the assumption that the control function takes care of all endogeneity does not necessarily hold.[32]

For our setting, we model $\mathbb{E}[x|w,z]$ as following a Poisson distribution:

$$\mathbb{E}[\text{aggregate\_trips}_{rt}|z_{rt},\phi_r,\gamma_t] = \exp(\pi z_{rt} + \phi_r + \gamma_t) \tag{13}$$

Where z_{rt} is our binary instrument:- $\text{demand\_day}_t \times \text{has\_mall}_r$. Define $\hat{g}_{rt} \equiv \exp(\hat{\pi} z_{rt} + \hat{\phi}_r + \hat{\gamma}_t)$, then our generalized error is $\text{aggregate\_trips}_{rt} - \hat{g}_{rt}$. We standardize the error using $\sqrt{Var(\text{aggregate\_trips}_{rt}|z_{rt},\phi_r,\gamma_t)}$. Because we are assuming our model follows a Poisson distribution, $\text{Var}(\text{aggregate\_trips}_{rt}|z_{rt},\phi_r,\gamma_t) = \mathbb{E}[\text{aggregate\_trips}_{rt}|z_{rt},\phi_r,\gamma_t]$. Thus our

[30] See figure A.13 for the distribution of aggregate demand counts.
[31] The elasticity is given by $h_0'(x) \cdot (x/h_0(x))$
[32] See Wooldridge (2014) for details.

standardized residual is:

$$\hat{e}_{rt} = \frac{\text{aggregate\_trips}_{rt} - \hat{g}_{rt}}{\sqrt{\hat{g}_{rt}}}$$

To get our causal estimate of the elasticity under this procedure, we estimate $\mathbb{E}\left[\text{supply}_{rt}|\text{aggregate\_trips}_{rt}, \phi_r, \gamma_t, e_{rt}\right]$ using a Poisson model and inserting our standardized residual from the first stage for e_{rt}.

The presence of fixed effects again presents some challenges in the implementation of this method. We estimate the first stage and second stage using Pseudo-Poisson Maximum Likelihood (PPML). PPML drops observations for fixed effects with no variation, e.g., routes that have only zero trips for our full observation period. As such, we can only generate residuals for routes that are not dropped during the first stage.[33] To avoid dropping a large number of observations, we instead estimate our model using origin and destination fixed effects in lieu of route fixed effects. Route fixed effects add little information over and above origin and destination fixed effects. Table 5 provides evidence of this. Estimates of our first-stage effects are identical with (column 3) and without (column 4) route fixed effects. Tables A.9 shows the same comparison for the reduced form regressions. Given that the estimates are identical for both specifications, we are confident that the omission of route fixed effects does not bias our results. Our new first-stage model is:

$$\mathbb{E}\left[\text{aggregate\_trips}_{rt}|z_{rt}, \nu_o, \eta_d, \gamma_t, w\right] = \exp(\pi z_{rt} + \alpha \text{has\_mall}_r + \nu_o + \eta_d + \gamma_t) \tag{14}$$

Recall that $z_{rt} = \text{demand\_day}_t \times \text{has\_mall}_r$ and so we must add back has_mall_r to the model if we do not include route fixed effects. With this specification, our second stage model is:

$$\mathbb{E}\left[\text{supply}_{rt}|\text{aggregate\_trips}_{rt}, \nu_o, \eta_d, \gamma_t, e_{rt}\right] = \exp(\beta_1 \text{aggregate\_trips}_{rt} + \beta_2 \text{has\_mall}_r + \nu_o + \eta_d + \gamma_t + \hat{e}_{rt}) \tag{15}$$

Where $\hat{e}_{rt}$ is the standardized residual from the first stage, and β_1 is our estimate of the elasticity. We calculate the standard errors via bootstrapping, standard errors are clustered at the route level.

Heterogeneity by Route Type We use the above procedure separately for within-territory routes and across-territory routes.

For the heterogeneity results, we also run a version in which we control for route length. We instrument for both $\text{aggregate\_trips}_{rt}$ and $\text{aggregate\_trips}_{rt} \times \text{route\_length}_r$. We de-mean route length so that estimates are interpreted at the average route length. Our first

[33]This is an issue only for route fixed effects as each day has variation in the number aggregate trips.

stages are modeled as:

$$\begin{aligned}\mathbb{E}\left[\text{aggregate\_trips}_{rt}|z_{1rt}, z_{2rt}, \nu_o, \eta_d, \gamma_t, w\right] = \exp(&\pi_1 z_{1rt} + \pi_2 z_{2rt} + \alpha \text{has\_mall}_r + \\ &\text{route\_length} + \\ &\text{treat\_day} \times \text{route\_length} + \\ &\nu_o + \eta_d + \gamma_t)\end{aligned} \tag{16}$$

$$\begin{aligned}\mathbb{E}\left[\text{aggregate\_trips}_{rt} \times \text{route\_length}|z_{1rt}, z_{2rt}, \nu_o, \eta_d, \gamma_t, w\right] = &\pi_1 z_{1rt} + \pi_2 z_{2rt} + \alpha \text{has\_mall}_r + \\ &\text{route\_length} + \\ &\text{treat\_day} \times \text{route\_length} + \\ &\nu_o + \eta_d + \gamma_t\end{aligned} \tag{17}$$

Where $z_{1rt} = \text{treat\_day}_t \times \text{has\_mall}_r$, and $z_{2rt} = \text{treat\_day}_t \times \text{has\_mall}_r \times \text{route\_length}_r$ are the instruments for $\text{aggregate\_trips}_{rt}$ and $\text{aggregate\_trips}_{rt} \times \text{route\_length}_r$ respectively. The standardized residuals are then added to our second stage model:

$$\begin{aligned}\mathbb{E}\left[\text{supply}_{rt}|x_{1rt}, x_{2rt}, \nu_o, \eta_d, \gamma_t, e_{rt}\right] = \exp(&\pi_1 x_{1rt} + \pi_2 x_{2rt} + \alpha \text{has\_mall}_r + \\ &\text{route\_length} + \\ &\text{treat\_day} \times \text{route\_length} + \\ &\nu_o + \eta_d + \gamma_t + \hat{e}_{1rt} + \hat{e}_{2rt})\end{aligned} \tag{18}$$

Where $x_{1rt} = \text{aggregate\_trips}_{rt}$ and $x_{2rt} = \text{aggregate\_trips}_{rt} \times \text{route\_length}_r$. We run this specification separately for within-territory routes and across-territory routes.

Results

Column (2) of Table 5 shows our estimates for the first-stage PPML regressions. On high-demand days on recreational routes, the aggregate number of trips increases by 4.1%. Panel A of Table 7 shows the results from our second stage regression using the generalized control function described above. An increase in the aggregate number of trips by 1%, increases the number of minibus taxi trips by 0.96%, and the number of minibuses on the route by 0.99%.

Within- vs Across-Territory Route Elasticities Panels B and C of Table 7 shows the results when we run the specification separately for within-territory routes and across-territory routes. Supply is more elastic on within-territory routes for both the number of minibus trips and the number of minibus taxis. On within-territory routes, an increase in the aggregate number of trips by 1% increases the number of minibus taxi trips by 0.93%. This elasticity is only 0.41% on across-territory routes. Similarly, an increase in the aggregate number of trips by 1% increases the number of minibus taxis by 0.93% compared to 0.41% on across-territory routes.

Table A.10 shows the results when we include controls for route length. Here, we lose significance in our elasticity estimate for within-territory routes, and this estimate is negative for across-territory routes. One caveat here is that these estimates are unreliable because the specification is too demanding; Table A.11 shows the first stage when we include controls for route length. Our instruments are no longer strong shifters for aggregate_trips and $\text{aggregate\_trips}_{rt} \times \text{route\_length}$. In general, there is a lot of overlap in route lengths for within-territory routes and across-territory routes; figure A.5 shows the distribution of route lengths for both types of routes. It is unlikely that differences in our elasticity estimates are driven by differences in travel behavior based on the length of the route.

7.2 Summary: Minibus Taxi Short-term Elasticity

Table 8 displays all our elasticity estimates from this section. We also include estimates from a non-parametric estimation *without* instrumental variables (columns (1) and (4)). For the non-parametric estimations, we display the median elasticity over our data range.

In general, both methodologies produce similar estimates and point to a similar pattern of results. On average, across all routes, the minibus taxi industry has an elasticity close to 1. When the number of aggregate trips increases by 1%, the number of minibus trips on the route increases by between 0.85%, and 0.96%. This elasticity remains high on within-territory routes, ranging from 0.93 to 0.94. However, on across-territory routes, the responsiveness is more muted. On these routes, elasticity ranges between 0.55 and 0.41. In general, our GCF estimates are more precise, and the NPIV estimation is more demanding on our data. The difference in elasticity is statistically significant when using the generalized control function methodology.

Our elasticity estimates rely on the assumption that minibus taxi availability on a route does not affect commuters' decisions to travel on that route. We assume that all commuters who want to travel on a route can find an alternative option, rather than forgoing the trip entirely. Under this assumption, our observed aggregate mobility data is equal to latent demand. If this assumption is violated, our observed increase in commuter trips would be biased down, and we would thus be overestimating the supply elasticity (and it would be biased towards 1). We do not think this is a big concern in this setting for two reasons. First, commuters have varied mobility options outside of minibus taxis, with a large proportion of households walking or using private vehicles as their primary mode of transportation (see Figure A.1). Second, a violation of this assumption would bias us away from observing a difference in elasticity between within-territory and across-territory routes, as both elasticities would be biased towards one, diminishing any difference.

In the next section, we provide evidence showing that associations cannot reallocate minibuses to across-territory routes as effectively as they can on within-territory routes. This analysis does not rely on aggregate mobility measures or the assumption outlined above. We arrive at consistent results demonstrating frictions in supply, further supporting our confidence in our elasticity estimates.

7.3 Why is the Impact Higher on Within-territory Routes?

Thus far we have seen that the elasticity of minibus taxis is lower on across-territory routes than on within-territory routes, and that marginal profits are higher on across-territory routes. In this section, we outline some descriptive evidence showing that associations face frictions in adding minibuses to across-territory routes. On these routes, associations must agree on how they are going to share the revenue generated from the route. One association cannot unilaterally add a new minibus as this could potentially eat into the profit of the other association with whom they are operating the route. These frictions limit flexibility.

At the end of the month, associations would like to reallocate their minibuses from non-recreational routes to recreational routes, where aggregate mobility is higher. They are able to freely allocate minibuses to recreational routes within-territory, but face additional frictions when trying to allocate these minibuses across-territory. We use our rich minibus operation data to test whether we observe this pattern of results. For this exercise, we use data on the 22 associations which have at least one within-territory route and one across-territory route. On within-territory routes, we consider only trips made by the association which controls the route; on across-territory routes, we consider only trips made by the associations who control the origin and destination grids of the route. This data refinement allows us to focus on the relevant associations' reallocation patterns.

We estimate the below model:

$$\begin{aligned}\text{supply}_{art} = \exp(\phi_r + \xi_a &+ \alpha \text{demand\_day}_t \times \text{within\_terr}_r \\ &+ \nu \text{demand\_day}_t \times \text{across\_terr}_r \\ &+ \beta \text{demand\_day}_t \times \text{has\_mall}_r \times \text{within\_terr}_r \\ &+ \rho \text{demand\_day}_t \times \text{has\_mall}_r \times \text{across\_terr}_r + \theta X_{rt})\end{aligned} \tag{19}$$

where ξ_a are association fixed effects. We include additional controls, X_{rt}, which include day of the week, month, and holiday fixed effects, as well as a control for the interaction of route length and demand_day. Based on our hypothesis, we would predict that:

1. $\alpha < 0$: Associations will remove minibuses from non-recreational routes at the end of the month.
2. $\beta > 0$: Associations will reallocate minibuses to recreational routes within their territory.
3. $0 \leq \rho < \beta$: Associations want to reallocate minibuses to recreational routes across-territory but face negotiation frictions which prevent them from doing so as efficiently as on within-territory routes.

There are no direct predictions on ν. $\nu = 0$ would suggest that associations in "joint ventures" face constraints in both directions – in the addition and removal of minibuses on across-territory routes. $\nu < 0$ would suggest that associations can reduce the number of minibuses across-territory, and the main constraint is in adding more minibuses.

Table 9 displays the estimated coefficients from equation (19). Columns (1) and (2) display the results for the total number of minibus taxis operating on a route, and columns (3) and (4) display the results for the total number of minibus trips. For both measures of supply, $\alpha < 0$. On within-territory routes that are not recreational, the number of minibus taxis decreases by 3.4% at the end of the month, and the number of minibus trips decreases by 3%. We also see that $\nu < 0$; associations are able to decrease the number of minibus taxis on across-territory routes. $\beta > 0$ for both measures of supply – on recreational within-territory routes, the number of minibus taxis increases by 2.4%, and the number of trips increases by 3.3%. On recreational across-territory routes, the response is more muted and not statistically significant – the number of minibus taxis increases by 1.0%, and the number of trips increases by 1.5%.

We cannot reject that $\beta = \rho$; however, it is worth noting that these results do not take into account the first-stage effect on demand. Column (5) of Table 9 reports the estimates using the number of aggregate demand trips as the outcome. On across-territory trips, aggregate demand increases by 7.1% on recreational routes, compared to an increase of 3.3% within-territory. As such, these differences in minibus response are even more stark – associations are severely under-allocating vehicles onto across-territory routes.

8 Theoretical Framework: Minibus Allocation

We present a simple model of how associations allocate their minibus taxis across different types of routes. The framework allows us to generate comparative statics for our two empirical exercises; i) the effect of a decrease in fleet size on route allocation, and ii) the effect of an increase in aggregate mobility on supply. We can then interpret our observed empirical patterns in terms of the model's parameters, formalizing our intuition on the impact of coordination costs on supply.

We consider the decisions of one association operating two types of routes – one within-territory, and one across-territory. Over some short-term time horizon (each day, or each week), the association decides how many minibuses to allocate to each route in order to maximize total profit. In this setting, prices are fixed in the short-term, and are thus treated as exogenous during these short-term dynamics.[34]

8.1 Minibus Ridership

Each route has λ_r agents traveling from the origin of the route to its destination. They decide whether to take a minibus taxi or some outside option (private vehicle, Uber, bus, etc.). In this set-up, λ_r is exogenous and is not affected by the availability of minibus taxis on the route.

Agents traveling on the route decide whether to use a minibus taxi or their outside option based on the price p_r of the minibus, and the wait time at the taxi rank . Wait time is decreasing in the number of minibuses b_r on the route. The proportion of agents on r who take a minibus is given by $\pi(p_r, b_r)$, where $\frac{\partial \pi}{\partial p} < 0$ and $\frac{\partial \pi}{\partial b} > 0$. That is, a higher proportion of λ_r will take a minibus taxi if the price of the minibus is lower, and there are more minibuses operating on the route (i.e, the wait time is shorter). Total minibus ridership on a route r is thus given by $\lambda_r \pi\left(p_r, b_r\right)$.

8.2 Minibus Supply

The association chooses the number of buses to allocate to their across-territory route and to their within-territory route – b_a and b_w respectively. They pay a common cost $c_0(b_a + b_w + b_{out})$ to allocate a minibus to a route. Here, b_{out} represents the number of

[34] Associations set prices for all routes annually and drivers cannot change them from day to day.

minibuses that are unavailable to operate. An increase in b_{out} increases the marginal cost of allocation such that the association must allocate fewer minibuses. Associations also pay additional costs (which may be zero) to allocate minibuses to within-territory routes $c_w(b_w)$ and across-territory routes, $c_a(b_a)$.

The association solves:

$$\begin{aligned} \underset{b_w,b_a}{max} \quad & p_w{\cdot}\lambda_w\pi(p_w,b_w) + p_a{\cdot}\lambda_a\pi(p_a,b_a) - c_0(b_w+b_a+b_{out}) - c_w(b_w) - c_a(b_a) && (20)\\ s.t \quad & b_w \geq 0, && (21)\\ & b_a \geq 0 && (22) \end{aligned}$$

Where, as discussed in the section above, $\lambda_w\pi\left(p_w,b_w\right)$ and $\lambda_a\pi\left(p_a,b_a\right)$ is the total minibus ridership on within-territory routes and across-territory routes respectively.

The first order conditions for the association's optimization problem are given by:

$$[b_w]: \quad p_w\lambda_w\pi'(b_w) = c_0'(b_w+b_a+b_{out}) + c_w'(b_w) \tag{23}$$

$$[b_a]: \quad p_a\lambda_a\pi'(b_a) = c_0'(b_w+b_a+b_{out}) + c_a'(b_a) \tag{24}$$

8.3 Model Implications

To simplify our results, we make the below assumption

Assumption 1. $\pi_r(b_r)'' \approx 0$ *for* $r \in w,a$ That is, demand is locally linear for small increases in the number of minibuses on the route.

Proposition 1. $\eta_w \equiv \dfrac{\mathrm{d}b_w}{\mathrm{d}b_{out}}, \eta_a \equiv \dfrac{\mathrm{d}b_a}{\mathrm{d}b_{out}}$ *, then under assumption 1.,* $\dfrac{\eta_w}{\eta_a} \approx \dfrac{c_a''}{c_w''}$

When there is a decrease in the overall fleet size, associations allocate the shortfall to routes with less convex costs.

Proposition 2. $\epsilon_w \equiv \dfrac{\mathrm{dlog}b_w}{\mathrm{dlog}\lambda_w}, \epsilon_a \equiv \dfrac{\mathrm{dlog}b_a}{\mathrm{dlog}\lambda_a}$ *, then under assumption 1.,* $\dfrac{\epsilon_w}{\epsilon_a} \approx \dfrac{\dfrac{C_a''\,b_a}{C_a'}}{\dfrac{C_w''\,b_w}{C_w'}}$,

where $C_r \equiv c_0 + c_r$ for $r \in \{w,a\}$

The ratio of short-term supply elasticity depends on the relative convexity of costs within- and across-territory. Elasticity will be higher within territory when costs across territory are relatively more convex than costs within territory.[35]

[35]Relative convexity measures not just how quickly marginal costs are changing (second derivatives c_a'' and

8.4 Summary: Model Implications

Our simple model provides a clear, unified story about allocation costs in this market. Our two empirical patterns are consistent with allocation costs being more convex on across-territory routes. This is in line with qualitative evidence on how associations operate these routes – by negotiating informal operating agreements with each other. Allocation to across-territory routes thus face more frictions which translates to higher costs that are also more convex.

9 The Costs of Segmentation in Informal Transportation

To quantify the costs incurred by Johannesburg commuters as a result of the segmentation of the informal public transportation network, we consider the additional wait time these commuters face on across-territory routes during periods of high demand. Our analyses above show that when aggregate mobility increases, the supply of minibuses on across-territory routes increases less proportionally than on within-territory routes. We estimate how much wait times would be reduced if the supply elasticity on across-territory routes matched the elasticity on within-territory routes. We then assign a monetary value to the total time lost using value of time travel estimates from the literature.

We use the average minibus headway – time between minibus departures – to approximate wait times. Assuming passengers arrive uniformly over a given time interval, the average wait time for each passenger will be half of the minibus headway.[36] Thus, we can estimate the total wait time lost if we know i) the total minibus ridership, ii) the average minibus headway, and iii) how these values vary within-territory and across-territory. Our counterfactual scenario does not consider any long-term readjustments to service or ridership if there are no coordination costs.[37]

Total Minibus Ridership We first quantify the total number of commuters who use minibus taxis on both within-territory routes and across-territory routes each day in Johannesburg. To do this, we combine data from three sources: our smartphone pings dataset, the 2011 South Africa census, and the 2020 South Africa National Household Transport Survey

c''_w), but also how these changes relate to current marginal costs (first derivatives c'_a and c'_w) and the scale of operations (b_a and b_w) (Palmer, 2003)

[36] This assumes that the bus capacity is large relative to the passenger arrival rate. That is, every passenger who arrives within a headway period is able to board the next bus that arrives. Based on our aggregate minibus ridership counts and estimated headway, this assumption is satisfied.

[37] For example, we do not take into account any changes in ridership as a result of increased flexibility across-territory.

(NHTS). We assign each device in our smartphone data to an income category based on the device's inferred home location and the average income level for that location, measured by the census (see Appendix A.6 for more information). Using the NHTS survey, we calculate the proportion of all trips that are minibus taxi trips by income category.[38] We then assign each device a probability of using a minibus taxi for a trip based on its proxied income group and the corresponding probability of using a minibus for that income group. We also weight each device based on the 2011 census so that our smartphone dataset is representative of the total Johannesburg population. Using these data adjustments to our smartphone trips, we calculate the total number of minibus trips on all within and across-territory routes for different time periods throughout the day.[39] Using this methodology, we estimate that on the average day, there are $2,047,438$ total commuter minibus trips within Johannesburg. The 2020 NHTS survey reports $1,979,758$ daily minibus trips within Johannesburg. The similarity between these two estimates increases the confidence in our aggregation method.[40]

Average Headway We calculate the average headway on different routes using our data on minibus trips. We define headway as the average amount of time between minibus departures. On route-days with less than two minibus departures, we impute a headway of 24 hours. We estimate headway elasticity using our estimation strategy in section 7, with average headway as our outcome. On within-territory routes, a 1% increase in demand decreases headway by approximately -1.2%. On across-territory routes, this elasticity is -0.3 (see table A.13). To approximate the level shift in headway, we use the fact that our data provider estimated that they finance 20% of all minibus taxis in operation in South Africa. We thus scale our minibus count and associated headway to account for all minibuses in operation. We assume that headway reduces proportionally with the number of minibus taxis on the route, and that the 20% market share is constant across all routes.[41] We thus scale down out observed headway estimates by 5 uniformly across all routes.

Total Time Lost Our estimate of the total time lost is how much total wait time would reduce if the headway elasticity on across-territory routes changed from -0.3 to -1.2 (to match the elasticity within-territory). We consider the total time lost on recreational across-

[38]NHTS asks participants about their trips on a randomly assigned "travel day". It asks how many trips they took, and how many trips were taken using a minibus taxi (and other modes).

[39]Table A.12 provides an example of our estimated minibus ridership on the average non-demand day

[40]Note that we do not use the NHTS survey for the absolute number of trips; we only use this survey for the proportion of trips taken by minibus taxi for each income group.

[41]Note that in Gauteng province, where Johannesburg is located, and where our data provider is headquatered, we estimate that the market share of our data provider is closer to 48% (See appendix A.4). In this section, we use the more conservative market share of 20% which decreases our headway (and thus wait time) by more than if we used the 48% estimate.

territory routes on high-demand days. Table 10 displays this calculation. We first document the baseline statistics on across-territory recreational routes on non-demand days. Column (1) of table 10 displays the average baseline headway (in minutes) for different time periods, and column (2) displays the total number of commuter minibus trips. On average there are over $200,000$ minibus trips every non-demand day on recreational, across-territory routes.

Column (3) documents the percent increase in aggregate commuter trips moving from non-demand to high-demand days. Using this percentage change, we calculate the corresponding decrease in headway under two scenarios: the status quo where the headway elasticity is -0.3, and the counterfactual scenario where the elasticity is -1.2. Column (4) presents the total wait time at baseline, calculated as the product of the total number of minibus riders and half of the newly calculated status-quo headway. Column (5) shows this result given the counterfactual headway reduction. The difference between these total wait times gives us our estimate of the total time lost due to this decreased flexibility. In total, we estimate that commuters in Johannesburg would save $988,142$ minutes on every high-demand day on recreational routes if the elasticity on across-territory routes was the same as that on within-territory routes. This represents a time saving of approximately 4 minutes per trip on every trip made on these types of routes.[42]

Value of Time Lost To quantify this lost time monetarily, we assume that the value of travel time (VTT) is between 50% and 100% of the average wage. This approximation is based on studies in developed countries which use both stated preferences (See Small (2012) for a review) and more recently, quasi-experimental methods (Goldszmidt et al., 2020; Buchholz et al., 2020). Research in developing contexts suggests that the VTT may be significantly higher in these contexts. Kreindler (2023) estimates that the VTT in Bangalore, India is 370% of the average hourly wage. Kreindler et al. (2023) also finds that commuters in Jakarta, Indonesia, value wait time 2.4 times more than actual travel time on the bus. Given these findings, our approach using 50 - 100% of the average wage likely underestimates the true monetary value of the additional wait time in this context.

In Johannesburg, South Africa, the modal weekly income category for commuters who use minibus taxis is R3696 - R6928 (calculation based on the 2020 National Household Transport Survey). Assuming a 40-hour work week, this translates to an hourly wage range of R92 to R173.2 (\$5.59 to \$10.40 using 2020 exchange rates). Applying our VTT estimate of between 50% and 100% of wages, the frictions to flexibility on across-territory routes costs commuters in Johannesburg between R$757,576$ and R$2,852,437$ (\$$46,030$ and \$$173,313$) per day on recreational routes.

[42]The average trip duration is approximately 45 minutes.

10 Conclusion

Informal transportation plays a crucial role in developing countries, yet we know little about the best way to organize the sector. While governments often mandate that minibuses organize into associations or cooperatives, our research reveals that this can impose significant costs for commuters.

This paper demonstrates that the segmented organizational structure of informal transportation creates supply frictions. The need for inter-association negotiations to allocate operations introduces frictions that increase operational costs on routes connecting different association territories. This reduces the system's overall responsiveness and leads to under-provision on these across-territory routes. This reduced flexibility has tangible costs for consumers, resulting in approximately 1 million minutes of additional wait time daily on across-territory routes.

While we focus on the costs of decentralization, it is important to acknowledge potential benefits of this organizational structure. Associations may facilitate the creation of new routes in response to long-term demand shifts, offering a different type of flexibility (Kerzhner, 2022). These benefits may depend on associations' incentives to grow and compete, suggesting that multiple associations may be necessary to realize these advantages. Moreover, competition among associations likely leads to lower prices for commuters, as operators vie for market share near territory boundaries.

Our research contributes to the understanding of optimal design for informal public transportation systems. By quantifying the inefficiencies introduced by fragmentation, we provide valuable insights for policymakers considering how to organize and regulate their informal public transportation industry.

References

Allen, Treb, and Costas Arkolakis. 2022. "The welfare effects of transportation infrastructure improvements." *The Review of Economic Studies*, 89(6): 2911–2957.

Almagro, Milena, Felipe Barbieri, Juan Camilo Castillo, Nathaniel G Hickok, and Tobias Salz. 2024. "Optimal Urban Transportation Policy: Evidence from Chicago." National Bureau of Economic Research Working Paper 32185, March.

Baum-Snow, Nathaniel, Loren Brandt, J. Vernon Henderson, Matthew A. Turner, and Qinghua Zhang. 2017. "Roads, Railroads, and Decentralization of Chinese Cities." *The Review of Economics and Statistics*, 99(3): 435–448.

Billings, Stephen. 2011. "Estimating the value of a new transit option." *Regional Science and Urban Economics*, 41(6): 525–536.

Björkegren, Dan, Alice Duhaut, Geetika Nagpal, and Nick Tsivanidis. 2024. "Public and Private Transit: Evidence from Lagos." *Working Paper.*

Bordeu, Olivia. 2023. "Commuting infrastructure in fragmented cities." *Job Market Paper, University of Chicago Booth School of Business*, 6.

Brancaccio, Giulia, Myrto Kalouptsidi, Theodore Papageorgiou, and Nicola Rosaia. 2023. "Search frictions and efficiency in decentralized transport markets." *The Quarterly Journal of Economics*, 138(4): 2451–2503.

Buchholz, Nicholas. 2021. "Spatial Equilibrium, Search Frictions, and Dynamic Efficiency in the Taxi Industry." *The Review of Economic Studies*, 89(2): 556–591.

Buchholz, Nicholas, Laura Doval, Jakub Kastl, Filip Matějka, and Tobias Salz. 2020. "The value of time: Evidence from auctioned cab rides." National Bureau of Economic Research.

Chen, Xiaohong, and Yin Jia Jeff Qiu. 2016. "Methods for Nonparametric and Semiparametric Regressions with Endogeneity: A Gentle Guide." *Annual Review of Economics*, 8(Volume 8, 2016): 259–290.

Chen, Xiaohong, Timothy Christensen, and Sid Kankanala. 2024. "Adaptive Estimation and Uniform Confidence Bands for Nonparametric Structural Functions and Elasticities."

Chen, Yanyou. 2024. “Network Structure and the Efficiency Gains from Mergers: Evidence from US Freight Railroads.” *Available at SSRN 4742990.*

Conwell, Lucas. 2024. “Subways or Minibuses? Privatized Provision of Public Transit.” *Working Paper.*

Donaldson, Dave. 2018. “Railroads of the Raj: Estimating the Impact of Transportation Infrastructure.” *American Economic Review*, 108(4–5): 899–934.

Donaldson, Dave, and Richard Hornbeck. 2016. “Railroads and American Economic Growth: A “Market Access” Approach.” *The Quarterly Journal of Economics*, 131(2): 799–858.

Fajgelbaum, Pablo D, and Edouard Schaal. 2020. “Optimal transport networks in spatial equilibrium.” *Econometrica*, 88(4): 1411–1452.

Fajgelbaum, Pablo D, Cecile Gaubert, Nicole Gorton, Eduardo Morales, and Edouard Schaal. 2023. “Political Preferences and the Spatial Distribution of Infrastructure: Evidence from California’s High-Speed Rail.” National Bureau of Economic Research.

Fréchette, Guillaume R., Alessandro Lizzeri, and Tobias Salz. 2019. “Frictions in a Competitive, Regulated Market: Evidence from Taxis.” *American Economic Review*, 109(8): 2954–92.

Gaduh, Arya, Tadeja Gračner, and Alexander D Rothenberg. 2022. “Life in the slow lane: Unintended consequences of public transit in Jakarta.” *Journal of urban economics*, 128: 103411.

Gibbons, Stephen, and Stephen Machin. 2005. “Valuing rail access using transport innovations.” *Journal of Urban Economics*, 57(1): 148–169.

Glaeser, Edward L, and Giacomo AM Ponzetto. 2018. “The political economy of transportation investment.” *Economics of Transportation*, 13: 4–26.

Glaeser, Edward, Matthew Kahn, and Jordan Rappaport. 2008. “Why do the poor live in cities? The role of public transportation.” *Journal of Urban Economics*, 63(1): 1–24.

Goldszmidt, Ariel, John A List, Robert D Metcalfe, Ian Muir, V Kerry Smith, and Jenny Wang. 2020. “The value of time in the United States: Estimates from nationwide natural field experiments.” National Bureau of Economic Research.

Gonzalez-Navarro, Marco, and Matthew A. Turner. 2018. “Subways and urban growth: Evidence from earth.” *Journal of Urban Economics*, 108: 85–106.

Gu, Yizhen, Chang Jiang, Junfu Zhang, and Ben Zou. 2021. “Subways and road congestion.” *American Economic Journal: Applied Economics*, 13(2): 83–115.

Heblich, Stephan, Stephen J Redding, and Daniel M Sturm. 2020. “The Making of the Modern Metropolis: Evidence from London.” *The Quarterly Journal of Economics*, 2059–2133.

Kelley, Erin, Gregory Lane, and David Schoenholzer. 2018. “The Impact of Monitoring Technologies on Contracts and Employee Behavior: Experimental Evidence from Kenya’s Matatu Industry with Gregory Lane and David Schoenholzer.”

Kerzhner, Tamara. 2022. “Is informal transport flexible?” *Journal of Transport and Land Use*, 15(1): 671–689.

Kreindler, Gabriel. 2023. “Peak-Hour Road Congestion Pricing: Experimental Evidence and Equilibrium Implications.” National Bureau of Economic Research Working Paper 30903, January.

Kreindler, Gabriel, Arya Gaduh, Tilman Graff, Rema Hanna, and Benjamin A Olken. 2023. “Optimal Public Transportation Networks: Evidence from the World’s Largest Bus Rapid Transit System in Jakarta.” National Bureau of Economic Research.

Kumar, Ajay, Sam Zimmerman, and Fatima Arroyo-Arroyo. 2021. “Myths and realities of “informal” public transport in developing countries: Approaches for improving the sector.” *Washington, DC: SSATP.*

Lin, Wei, and Jeffrey M. Wooldridge. 2019. “Chapter 2 - Testing and Correcting for Endogeneity in Nonlinear Unobserved Effects Models.” In *Panel Data Econometrics.* eds. Mike Tsionas, 21–43. Academic Press.

Majid, Hadia, Ammar Malik, and Kate Vyborny. 2018. “Infrastructure Investments, Public Transport Use and Sustainability: Evidence from Lahore, Pakistan.” Working Paper.

McKay, Tracey, and Jarred Bell. 2011. “The Rise of ’Class Apartheid’ in Accessing Secondary Schools in Sandton, Gauteng.” *Journal of Southern African Studies*, 17: 27–48.

Ndibatya, Innocent, and M.J. Booysen. 2020. "Minibus taxis in Kampala's paratransit system: Operations, economics and efficiency." *Journal of Transport Geography*, 88: 102853.

OECD. 2019. *Making Decentralisation Work: A Handbook for Policy-Makers. OECD Multi-level Governance Studies*, Paris: OECD Publishing.

Palmer, James A. 2003. "Relative convexity." *ECE Dept., UCSD, Tech. Rep.*

Rosaia, Nicola. 2020. "Competing platforms and transport equilibrium: Evidence from New York City." *Working Paper URL https://economics. sas. upenn. edu/system/files/2021-09/Rosaia. pdf.*

Saddier, Simon, and Alex Johnson. 2018. "Understanding the Operational Characteristics of Paratransit Services in Accra, Ghana: A Case Study." 37th Annual Southern African Transport Conference (SATC 2018).

Shah, Kishan, and Federico Sturzenegger. 2022. "Search, Transport Costs, and Labor Markets in South Africa." Red Nacional de Investigadores en Economía (RedNIE) Working Paper 158.

Small, Kenneth A. 2012. "Valuation of travel time." *Economics of transportation*, 1(1-2): 2–14.

Terza, Joseph V. 2009. "Parametric Nonlinear Regression with Endogenous Switching." *Econometric Reviews*, 28(6): 555–580.

Tsivanidis, Nick. 2019. "Evaluating the impact of urban transit infrastructure: Evidence from bogota's transmilenio."

Vermeiren, Karolien, Els Verachtert, Peter Kasaija, Maarten Loopmans, Jean Poesen, and Anton Van Rompaey. 2015. "Who could benefit from a bus rapid transit system in cities from developing countries? A case study from Kampala, Uganda." *Journal of Transport Geography*, 47: 13–22.

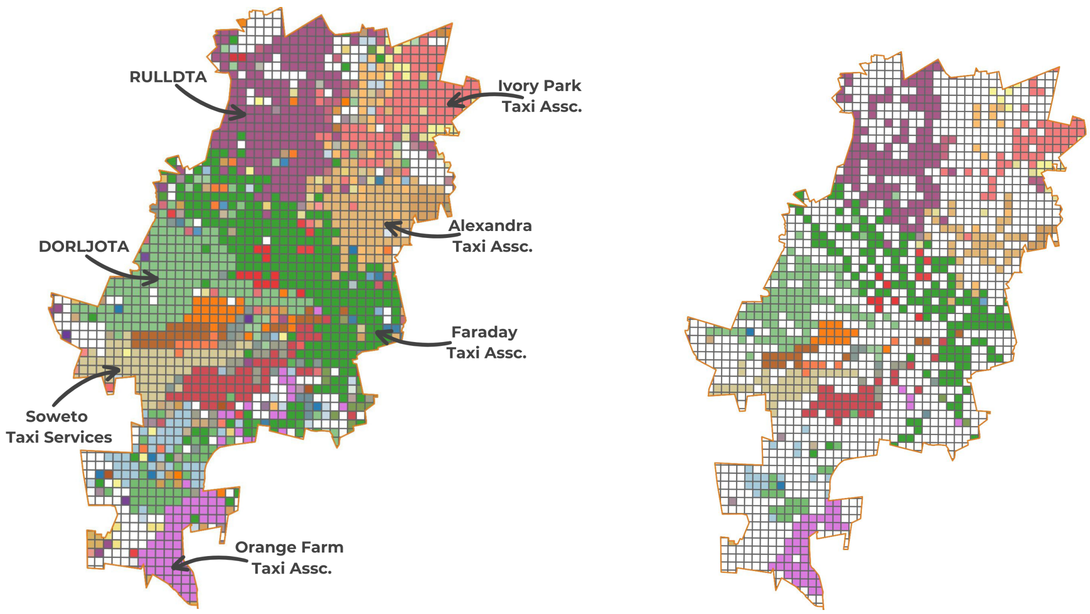
Wong, Woan Foong, and Simon Fuchs. 2022. "Multimodal Transport Networks." FRB Atlanta Working Paper 2022-13, October. Available at SSRN: https://ssrn.com/abstract=4277194 or http://dx.doi.org/10.29338/wp2022-13.

Wooldridge, Jeffrey M. 2014. "Quasi-maximum likelihood estimation and testing for nonlinear models with endogenous explanatory variables." *Journal of Econometrics*, 182(1): 226–234.

Zárate, Román D. 2022. "Spatial misallocation, informality, and transit improvements: Evidence from mexico city."

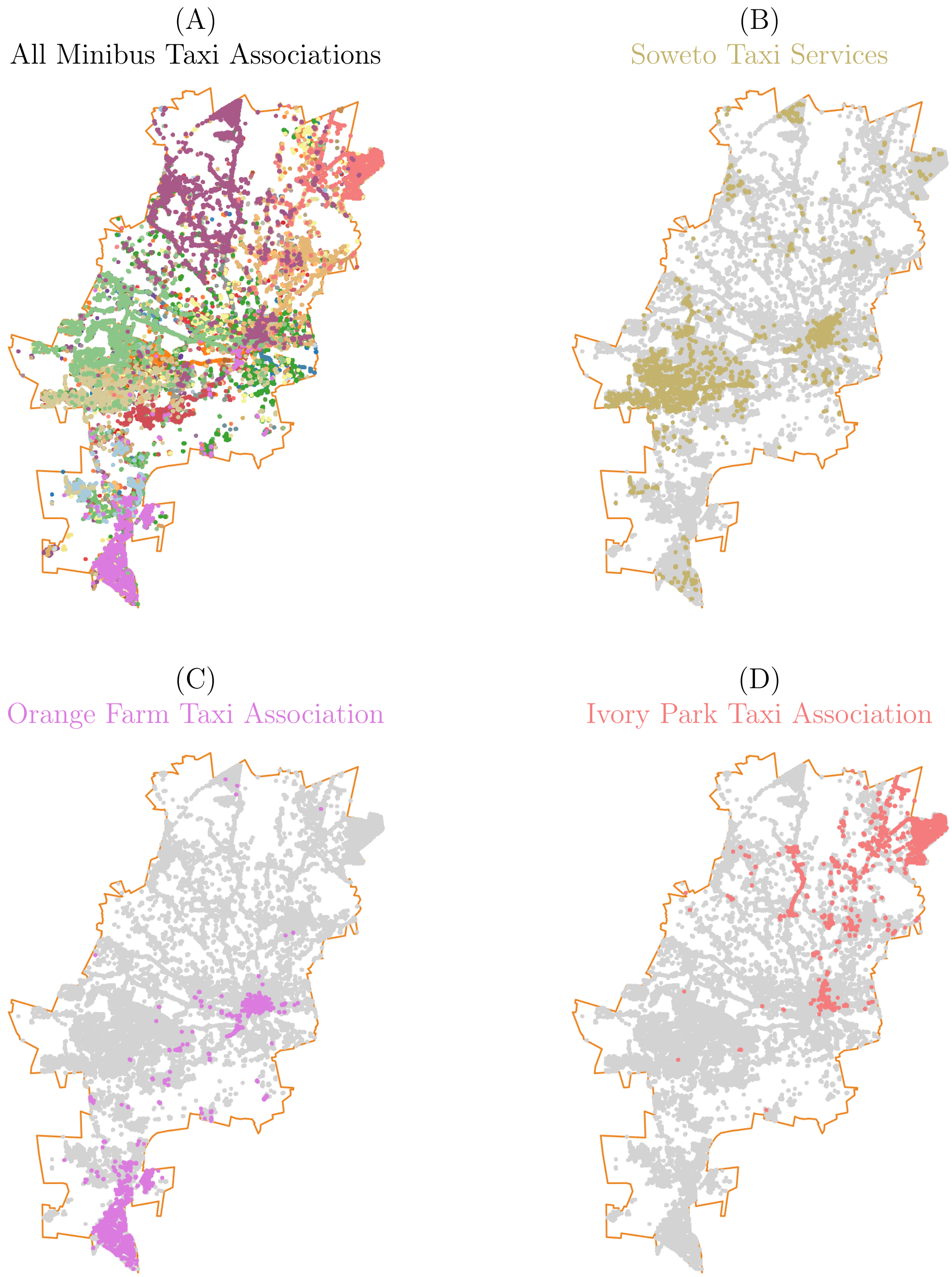
Figures

Figure 1: Minibus Association Territories



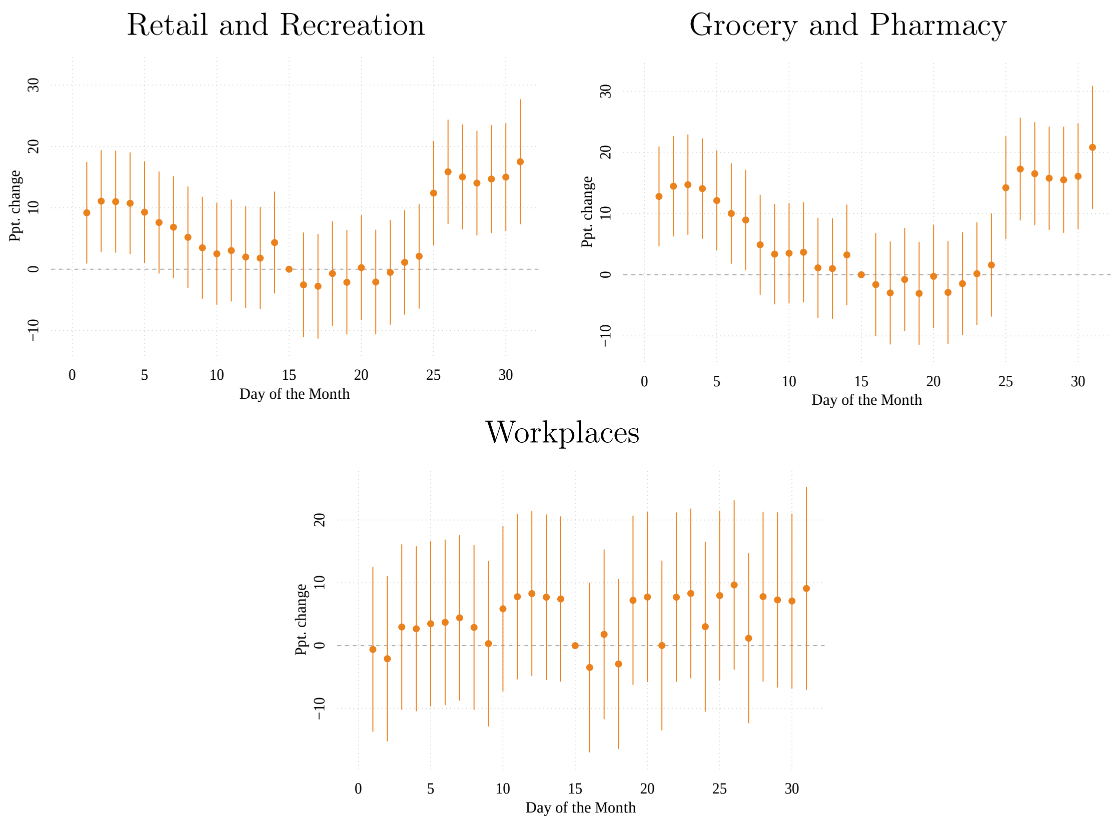
Note: The left panel plots the map of grid associations in Johannesburg, based only on the plurality of minibus associations operating in that grid during our study period. The right panel displays the same information filtered to show only grid cells where at least 15 unique minibuses operated and where the plurality association held more than 22% of the vehicle share. RULLDTA represents the Randburg United Local and Long Distance Taxi Association, and DORLJOTA represents Dobsonville, Roodepoort, Leratong, Johannesburg Taxi Association.

Figure 2: Minibus Mobility Patterns by Association



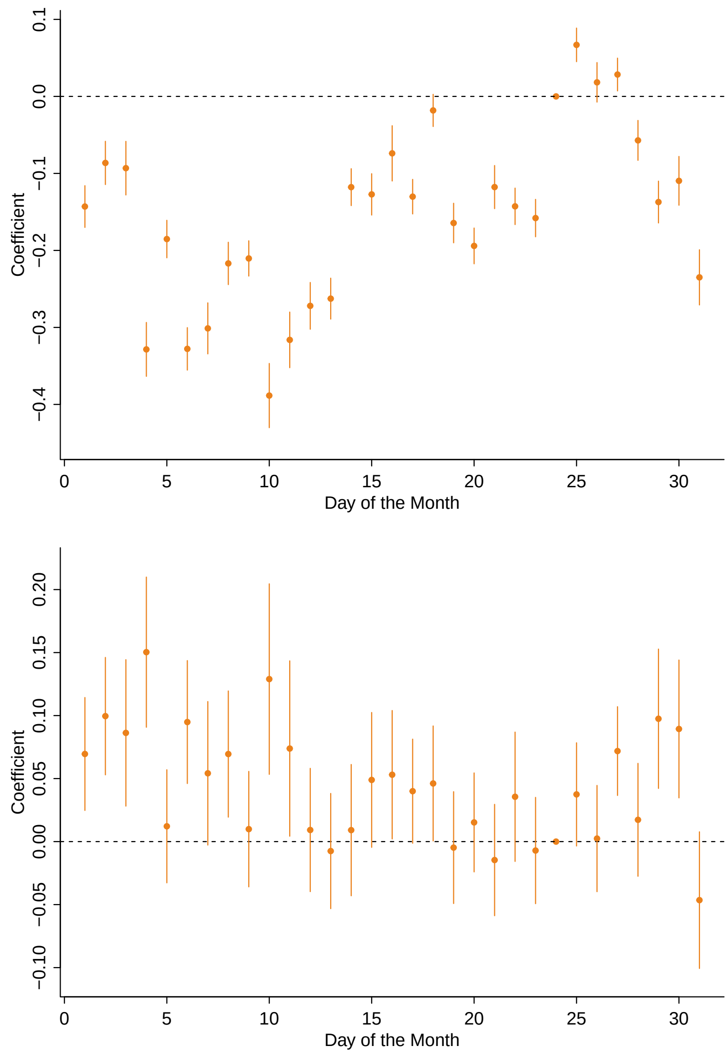
Note: This figure displays the operation of associations in Johannesburg in March 2023. Each dot represents a starting location of a recorded trip, with colors indicating the taxi association. Panel (A) shows the complete network of all minibus movements. Panels (B), (C), and (D) isolate the movement patterns of three specific associations: Soweto Taxi Services, Orange Farm Taxi Association, and Ivory Park Taxi Association respectively, with other associations shown in gray for context.

Figure 3: Day of Month Mobility by Destination Type (Google Mobility Data)



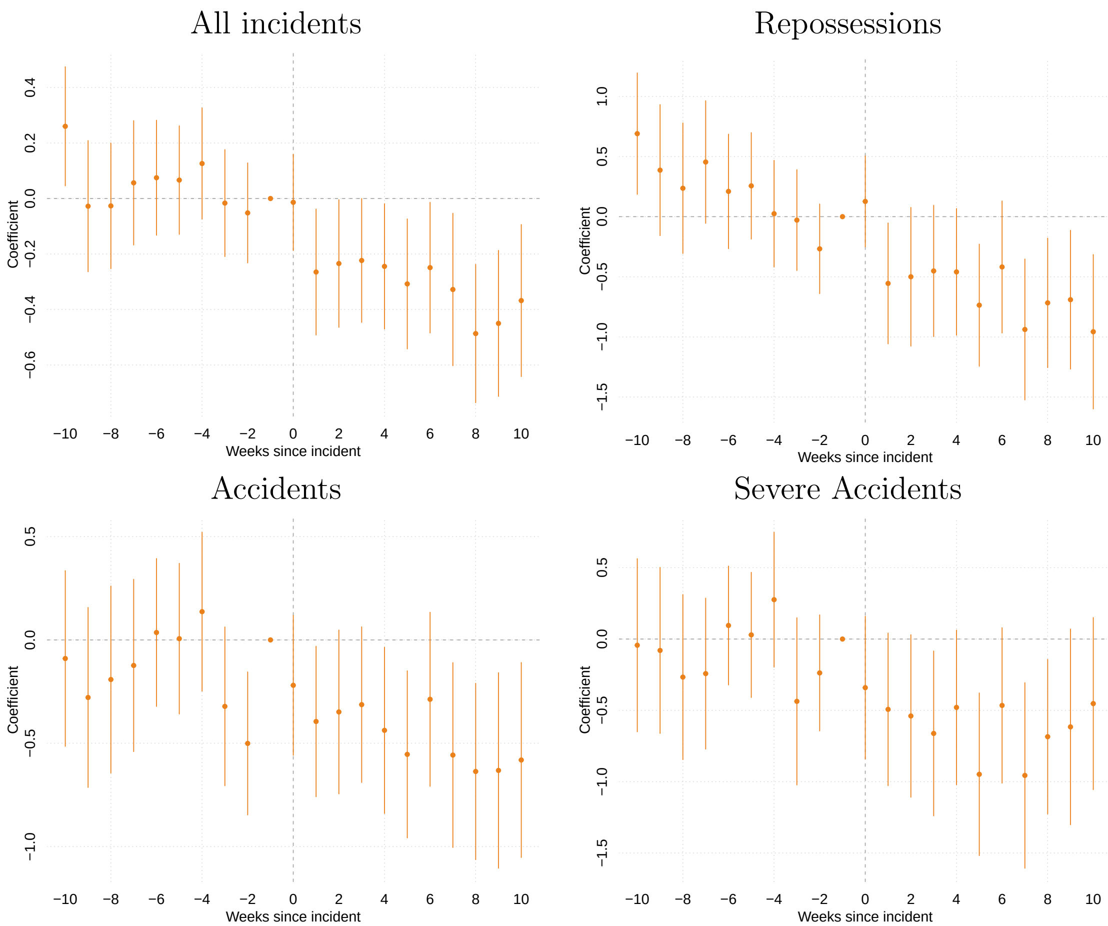
Note: Residualized change in mobility by trip category. These plots display the change in mobility in 2022 compared to the same day of the week in 2020, residualized by day of the week. Each data point is the estimated average difference in mobility on that day, compared to the 15th of the month holding day of the week constant. The top left panel displays the mobility pattern for trips to retail and recreational locations, the top right panel displays the mobility pattern for trips to grocery stores and pharmacies, and the bottom panel displays the mobility pattern for trips to workplaces.

Figure 6: Aggregate Commuter Trips Cyclicality



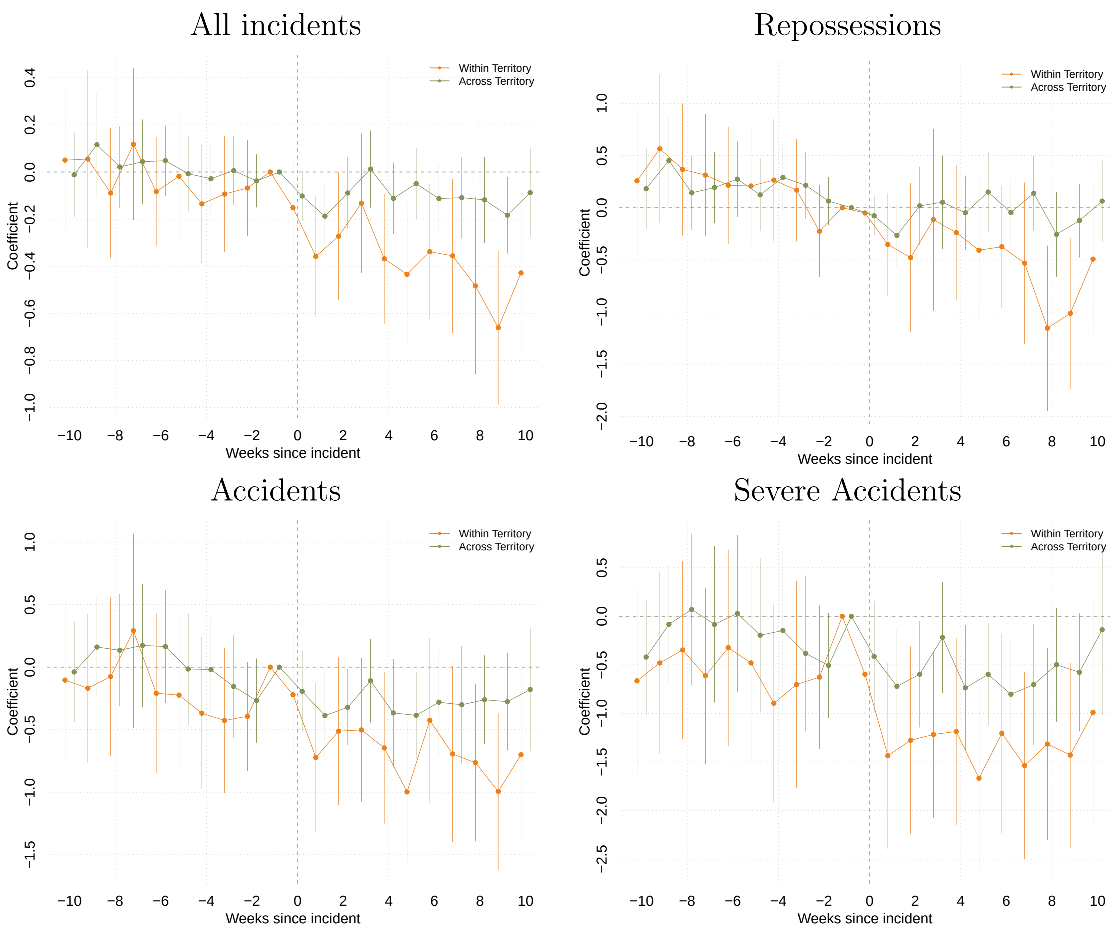
Note: These plots show how aggregate mobility varies by day of the month using the 24th as the reference day. The top panel plots coefficients for each day of the month (eq. 7). The bottom panel plots coefficients for the interaction of the day of the month and an indicator for if the route is recreational (eq. 8). Standard errors are two-way clustered at the origin-destination level.

Figure 4: Impact of Insurance Incidents on Asociation Fleet Size



Note: These plots show the weekly changes in association fleet size before and after different types of insurance incidents. The horizontal axis represents weeks relative to the incident (week 0), while the vertical axis shows the estimated percentage change in fleet size. The panels display effects for different incident types: all incidents (top left), repossessions (top right), accidents (bottom left), and severe accidents (bottom right). Events are weighted by the total association fleet size over the full study period. Standard errors are clustered at the association level and coefficients are estimated using PPML.

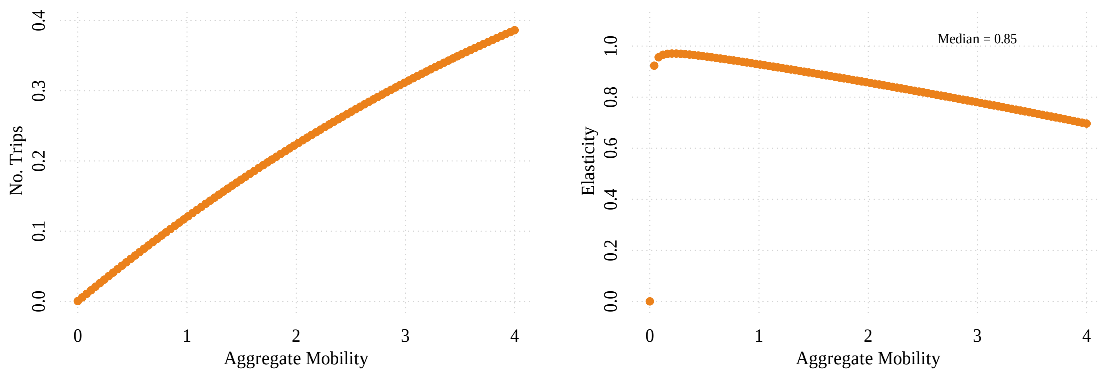
Figure 5: Effect of Insurance Incidents on Route Allocation



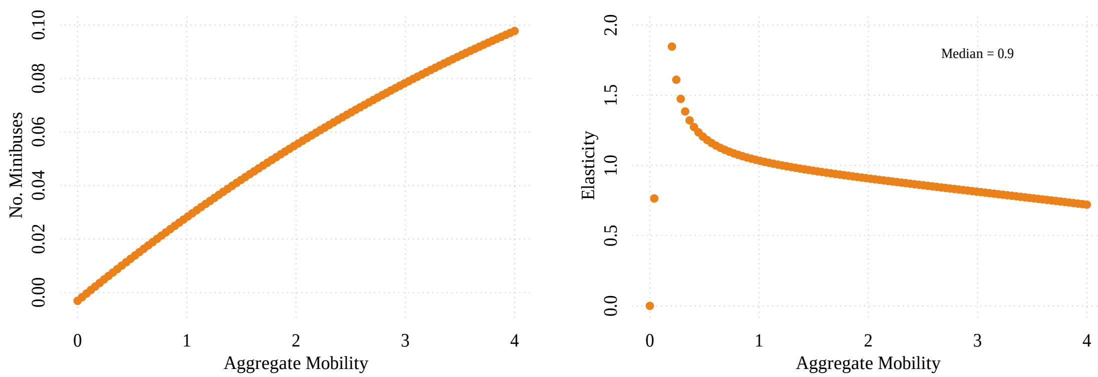
Note: These plots show how different types of insurance claims affect minibus operations on within- and across-territory routes over time. The horizontal axis shows weeks relative to the incident (week 0), and the vertical axis represents the estimated percentage change in number of minibuses assigned to the route. Orange lines with confidence intervals show effects on within-territory routes, while blue lines show across-territory routes. The panels display effects for: all incidents (top left), repossessions (top right), accidents (bottom left), and severe accidents (bottom right). Events are weighted by the total association fleet size over the full study period. Coefficients are estimated using PPML. Standard errors are clustered at the association level.

Figure 7: Non-parametric Estimation of Minibus Supply as a Function of Aggregate Demand

(a) Minibus Trips



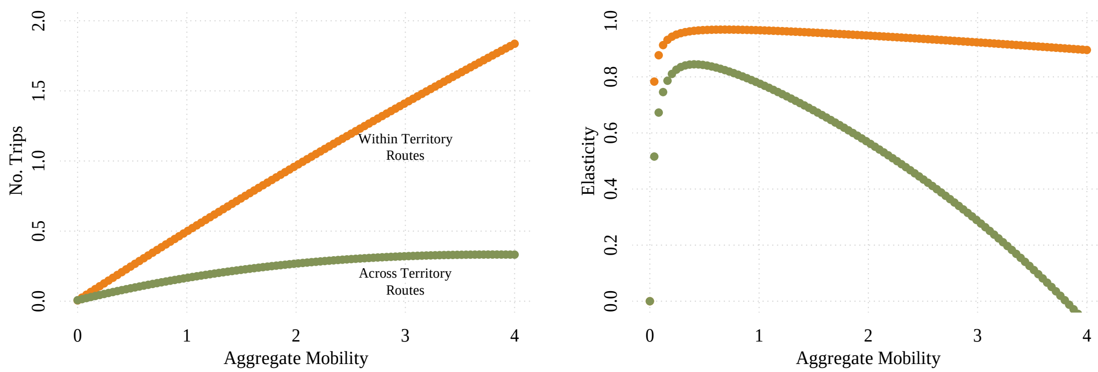
(b) No. Minibuses



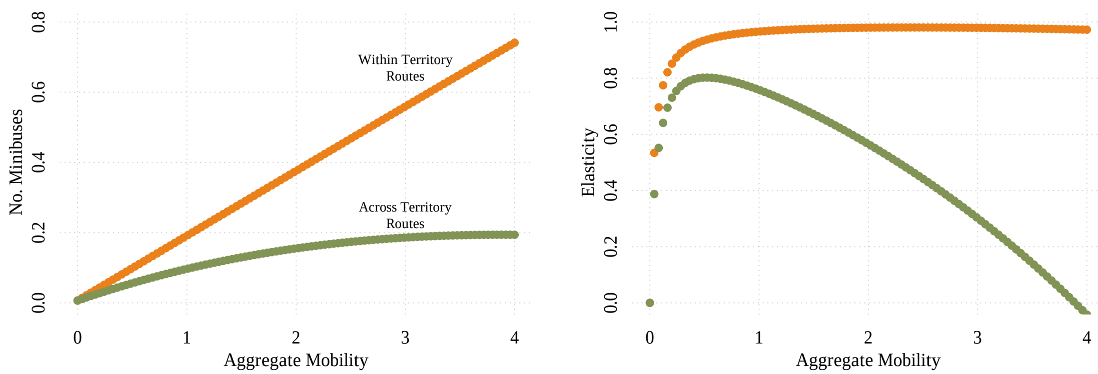
Note: These plots show the non-parametric relationship between aggregate mobility demand and minibus taxi supply. The top row estimates the relationship between the number of minibus trips and aggregate demand, and the bottom row estimates the relationship between the number of minibuses and aggregate demand. The left column plots the estimated function, and the right column plots the implied elasticities.

Figure 8: Non-parametric Estimation of Minibus Supply Elasticity by Route Type

(a) Minibus Trips



(b) No. Minibus



Note: These plots show the non-parametric relationship between aggregate mobility demand and minibus taxi supply separated by route type. The top row estimates the relationship between the number of minibus trips and aggregate demand, and the bottom row estimates the relationship between the number of minibuses and aggregate demand. The left column plots the estimated function, and the right column plots the implied elasticities. The orange points correspond to the relationship on within-territory routes, and the green points correspond to the relationship on across-territory routes.

Tables

Table 1: Summary Stats: Minibus Taxi Trips

	Raw Trips (1)	Minibus Route Trips (2)
N.	40,124,891	7,142,646
N. Unique Vehicles	19,214	8,947
N. Unique Associations	720	527
Median No. Vehicles per Assoc.	8.00	5.00
Mean No. Vehicles per Assoc.	23.87	15.44
Median No. Daily Trips per Veh.	12.00	5.00
Mean No. Daily Trips per Veh.	14.65	6.74

Note: This table presents summary statistics comparing all raw taxi trips as defined by the minibus financier (Column 1) with trips restricted to established minibus routes (Column 2).

Table 2: Summary Stats: Smartphone Data

	Raw Pings (1)	All Algorithm Trips (2)	Minibus Route Trips (3)
N.	373,715,157	9,049,991	510,918
N. Unique Devices	1,087,889	516,558	139,889
Median No. Daily Obs per Device.	19.00	2.00	1.00
Mean No. Daily Obs per Device.	34.56	2.11	1.46

Note: This table summarizes smartphone location data collected from October 2022 to July 2023. Column (1) shows statistics for raw smartphone pings. Column (2) presents data for all trips identified by our algorithm within Johannesburg. Column (3) shows the subset of trips that match established minibus taxi routes.

Table 3: Effects of Insurance Incidents on Minibus Operation, Association Fleet, and Allocation

	All Incidents	Repossessions	Accidents	Severe Accidents
	(1)	(2)	(3)	(4)
Panel A: Minibus Operation	-0.121^{***}	-0.264^{***}	-0.111^{***}	-0.182^{***}
	(0.007)	(0.015)	(0.012)	(0.018)
N.	718,155	718,155	718,155	718,155
Panel B: Fleet Size	-0.323^{***}	-0.701^{***}	-0.315^{***}	-0.511^{***}
	(0.068)	(0.167)	(0.118)	(0.139)
N.	29,925	29,925	29,925	29,925
Panel C: No. Minibuses by Route Type				
Within Terr.	-0.319^{***}	-0.658^{***}	-0.471^{***}	-0.745^{***}
	(0.111)	(0.242)	(0.173)	(0.246)
Across Terr.	-0.114^{**}	-0.218^{*}	-0.286^{**}	-0.374^{*}
	(0.054)	(0.124)	(0.136)	(0.216)
Difference	-0.205^{***}	-0.44^{**}	-0.185	-0.371^{**}
	(0.075)	(0.172)	(0.114)	(0.157)
N.	47,070	47,070	47,070	47,070
Minibus/Assoc. FEs	Yes	Yes	Yes	Yes
Week FEs	Yes	Yes	Yes	Yes

Note: The average effect of incidents on the probability of minibus operation, association fleet size, and the number of minibuses allocated to within- and across-territory routes. Coefficients represent the average effect in the first 10 weeks after the incident. Panel A is at the minibus-week level. Each event is an indicator for whether a minibus had an incident in the previous 10 weeks. Standard errors are clustered at the vehicle level and coefficients are estimated using OLS. Panels B and C are at the association-week level. Each event is the total number of incidents in the previous 10 weeks weighted by the total association fleet size over the full study period. Standard errors are clustered at the association level and coefficients are estimated using PPML. Column (1) shows the effect for all incidents in the insurance claims data, column (2) shows the effects for repossessions, column (3) shows the effects for accidents and column (4) shows the effect for severe accidents (accidents with a claim value larger than 75,000 rand). *p<0.1; **p<0.05; ***p<0.01.

Table 4: The Impact of Fleet Size on Route Allocation

	No. Minibuses Allocated				
	PPML	PPML-IV			
Instrument:		All Incidents	Repossessions	Accidents	Severe Accidents
	(1)	(2)	(3)	(4)	(5)
Fleet Size	0.742***	1.41***	0.507**	3.01***	2.14***
	(0.026)	(0.300)	(0.258)	(0.746)	(0.679)
Fleet Size × Within Terr.	0.181***	0.718***	0.731***	0.805***	0.630***
	(0.017)	(0.069)	(0.106)	(0.172)	(0.201)
N.	31,130	31,130	31,130	31,130	31,130
Assoc FEs	Yes	Yes	Yes	Yes	Yes
Week FEs	Yes	Yes	Yes	Yes	Yes
F-test (1st stage)					
Cragg-Donald F-stat		27.984	21.502	17.341	21.996

Note: This table displays the relationship between the total association fleet size and the number of vehicles allocated to within- and across-territory routes. Regressions are at the association-week-routetype level. Model (1) estimates the relationship using PPML and models (2) - (5) instrument for the association fleet size in a given week using the number of incidents (repossessions/accidents/severe accidents) that occurred 10 weeks prior weighted by the total association fleet size over the full study period. Coefficients (2) - (5) are estimated using PPML-IV. The last panel of the table reports Crag-Donald F-statistics from the first-stages of the PPML-IV models. Standard errors are bootstrapped and clustered at the association level. *p<0.1; **p<0.05; ***p<0.01.

Table 5: "First-Stage": Impact of Cyclicality Instrument on Commuter Trips

	Commuter Trips			
	OLS		PPML	
	(1)	(2)	(3)	(4)
demand_day	-0.0003^{***}		-0.0273^{***}	
	(0.0001)		(0.0054)	
has_mall	-0.042^{***}		0.0265	
	(0.012)		(0.0532)	
demand_day × has_mall	0.001^{***}	0.001^{***}	0.0409^{***}	0.0409^{***}
	(0.0002)	(0.0002)	(0.0074)	(0.0074)
N.	30,838,368	30,838,368	30,838,368	10,092,192
Day FEs		Yes		Yes
Route FEs		Yes		Yes
Weekday FEs	Yes		Yes	
Holiday FEs	Yes		Yes	
Month FEs	Yes		Yes	
Origin FEs	Yes		Yes	
Destination FEs	Yes		Yes	

Note: This table analyzes how the cyclicality instrument (interaction of end of the month with recreational route) impacts the aggregate number of commuter trips. Coefficients in columns (1) and (2) are estimated using Ordinary Least Squares, and coefficients in columns (3) and (4) are estimated using Pseudo-Poisson Maximum Likelihood. demand_day is an indicator for whether the day is at the "end of the month" (after the 25th and before the 10th). has_mall is an indicator for whether a route has a large mall at either its origin or destination. The dependent variable is the number of aggregate commuter trips on a given route on a given day. Regressions without day FEs have weekday, holiday and month FEs. Standard errors are two-way clustered by origin and destination. *p<0.1; **p<0.05; ***p<0.01.

Table 6: Cyclicality "First Stage" and "Reduced Form" by Route Type

	First-Stage	Reduced Form	
	Commuter Trips	Minibus Trips	No. Minibuses
	(1)	(2)	(3)
Panel A: All Routes			
demand_day × has_mall	0.041***	0.030***	0.023***
	(0.007)	(0.007)	(0.006)
N.	10,092,192	9,555,936	9,555,936
Panel B: Within Terr. Routes			
demand_day × has_mall	0.033***	0.030***	0.023***
	(0.009)	(0.007)	(0.006)
N.	6,637,536	6,270,608	6,270,608
Panel C: Across Terr. Routes			
demand_day × has_mall	0.071***	0.021**	0.017**
	(0.015)	(0.009)	(0.008)
N.	3,454,656	3,285,328	3,285,328
Day FEs	Yes	Yes	Yes
Route FEs	Yes	Yes	Yes

Note: This table displays the effect of the cyclicality instrument on commuter and minibus trips by route type. Coefficients are estimated using Poisson Pseudo-Maximum Likelihood (PPML). Column (1) shows the first-stage effect on commuter trips, while columns (2) and (3) show reduced-form effects on minibus supply, measured by the number of minibus trips and unique minibuses respectively. Panel A shows results for all routes, Panel B for routes within taxi association territories, and Panel C for routes that cross territory boundaries. All specifications include route and day fixed effects. Standard errors (in parentheses) are two-way clustered at the origin-destination level. *p<0.1; **p<0.05; ***p<0.01.

Table 7: Supply Elasticity Estimates Using Generalized Control Function (GCF) Approach

	Minibus Trips	No. Minibuses
	(1)	(2)
Panel A: All Routes		
Commuter Trips	0.960***	0.989***
	(0.035)	(0.034)
N	30,645,024	30,645,024
Panel B: Within Terr. Routes		
Commuter Trips	0.929***	0.928***
	(0.017)	(0.016)
N	17,673,040	17,673,040
Panel C: Across Terr. Routes		
Commuter Trips	0.406***	0.401***
	(0.047)	(0.047)
N	12,933,984	12,933,984
Difference		
Within Terr. - Across. Terr	0.523***	0.527***
	(0.049)	(0.049)
Day FEs	Yes	Yes
Origin FEs	Yes	Yes
Destination FEs	Yes	Yes

Note: This table presents estimates of how minibus taxi supply responds to changes in commuter demand across different route types. Coefficients are estimated using Pseudo-Poisson Maximum Likelihood (PPML) with a generalized instrumental variable control function. Commuter Trip is the number of aggregate commuter trips on route r and day t. Column (1) shows the effect of aggregate demand on the number of minibus taxi trips, and column (2) shows the effect on the number of minibus taxis on the route. Panel A shows the average elasticity for all routes, panels B and C show the elasticity for within- and across-territory routes respectively. The last panel shows the difference between the estimates for within-territory routes and across-territory routes. Standard errors are obtained via Bayesian bootstrap and are clustered at the origin level. *p<0.1; **p<0.05; ***p<0.01.

Table 8: Summary of Estimated Elasticities

	Minibus Trips			No. Minibuses		
	NP	NPIV	GCF	NP	NPIV	GCF
	(1)	(2)	(3)	(4)	(5)	(6)
Overall Elasticity	1.352***	0.854***	0.960***	1.441***	0.900	0.989***
	(0.209)	(0.119)	(0.035)	(0.184)	(0.694)	(0.034)
Within Terr. Elasticity	1.281***	0.943***	0.929***	1.403***	0.976*	0.928***
	(0.250)	(0.139)	(0.017)	(0.208)	(0.539)	(0.016)
Across Terr. Elasticity	0.005	0.546	0.406***	1.875	0.546	0.401***
	(1.949)	(0.722)	(0.047)	(1.441)	(0.709)	(0.047)
Within Terr. − Across Terr.	1.276	0.397	0.523***	−0.472	0.430	0.527***
	(2.045)	(0.756)	(0.049)	(1.481)	(0.952)	(0.049)
Day FEs	Yes	Yes	Yes	Yes	Yes	Yes
Route FEs	Yes	Yes		Yes	Yes	
Origin FEs			Yes			Yes
Destination FEs			Yes			Yes

Note: This table summarizes our estimates of elasticities across methodologies. Columns (1) and (2) estimate the elasticity using non-parametric estimation (NP) without an instrumental variable. Columns (3) and (4) use non-parametric instrumental variable (NPIV) estimation, and columns (5) and (6) use a generalized control function approach (GCF). Columns (1), (3), and (4) display elasticities for the number of minibus trips, while (2), (4), and (6) display results for the number of minibus vehicles. All standard errors are obtained via bootstrapping and are clustered at the origin level.

*p<0.1; **p<0.05; ***p<0.01.

Table 9: Reallocation by Route Type

	No. Minibuses		Minibus Trips		Commuter Trips
	(1)	(2)	(3)	(4)	(5)
demand_day× within_terr (α)	-0.034^{***} (0.005)		-0.029^{***} (0.006)		-0.021^{**} (0.007)
demand_day× across_terr (η)	-0.029^{***} (0.007)		-0.024^{***} (0.007)		-0.050^{***} (0.010)
demand_day×has_mall× within_terr (β)	0.024^{**} (0.010)	0.024^{**} (0.010)	0.033^{***} (0.011)	0.033^{***} (0.011)	0.033^{***} (0.009)
demand_day×has_mall× across_terr (ρ)	0.010 (0.011)	0.010 (0.011)	0.015 (0.012)	0.015 (0.012)	0.071^{***} (0.015)
within_terr - across_terr		0.014		0.017	-0.038^{**}
[p-value]		[0.373]		[0.291]	[0.050]
N.	15,692,466	15,692,466	15,692,466	15,692,466	10,092,192
Day FEs		Yes		Yes	
Route FEs	Yes	Yes	Yes	Yes	Yes
Assoc. FEs	Yes	Yes	Yes	Yes	Yes
Weekday FEs	Yes		Yes		Yes
Holiday FEs	Yes		Yes		Yes
Month FEs	Yes		Yes		Yes

Note: This table examines how end-of-month timing affects associations' allocation of minibuses across different route types. Coefficients are estimated using Pseudo-Poisson Maximum Likelihood. Columns (1) and (2) use the number of minibus taxis as the dependent variable, columns (3) and (4) use the number of minibus trips as the dependent variable, and column (1) uses the number of aggregate trips as the dependent variable. Greek letters in parentheses correspond to the coefficients in eq.(19). within_terr - across_terr reports the difference between demand_day×has_mall× within_terr and demand_day×has_mall× across_terr and reports the p-value of the difference in brackets. Standard errors are two-way clustered by origin and destination. *p<0.1; **p<0.05; ***p<0.01.

Table 10: Total Lost Time on Recreational Routes

	Baseline Headway (mins) (1)	Baseline Minibus Ridership (2)	Mobility Increase (%) (3)	Total Wait Time w/ elasticity of -0.33 (4)	Total Wait Time w/ elasticity of -1.2 (5)	Difference (6)
12:00am - 4:59am	55.55	12,779	−6.63	336,118	355,078	−18,960
				(240)	(501)	(555)
5:00am - 9:59am	51.30	57,478	11.47	1,565,857	1,403,451	162,406
				(2,053)	(4,293)	(4,758)
10:00am - 2:59pm	48.89	74,545	28.01	2,138,026	1,564,041	573,984
				(7,257)	(15,174)	(16,816)
3:00pm - 7:59pm	57.93	59,853	13.78	1,830,529	1,600,662	229,867
				(2,906)	(6,077)	(6,734)
8:00pm - 11:59pm	56.05	9,295	16.01	277,652	236,806	40,846
				(516)	(1,080)	(1,197)
Total		213,951		6,148,180	5,160,038	988,142
				(12,494)	(26,123)	(28,950)

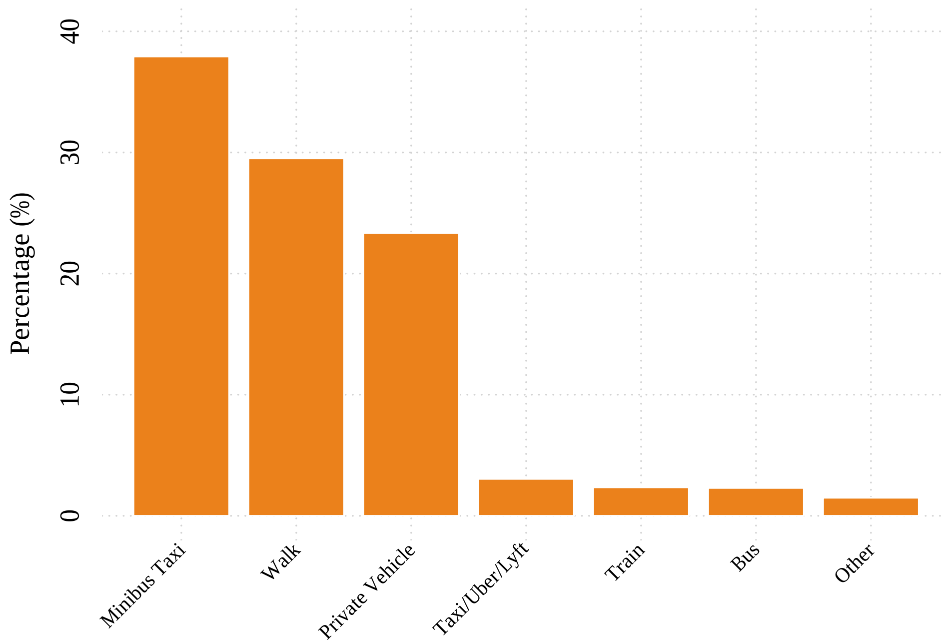
Note: This table calculates the total time lost on across-territory recreational routes at the end of the month. Columns (1) and (2) display the baseline headway and number of minibus rides respectively for each time period. Column (3) displays the percent increase in demand moving from non-demand days to demand days. Column (4) shows the estimated total wait time using the baseline headway elasticity of -0.33, and column (5) shows what the wait times would be with an elasticity of -1.2. Column (6) displays the difference. Standard errors are displayed in parentheses and calculated using the headway elasticity standard errors (Table A.13)

A Appendix

Contents

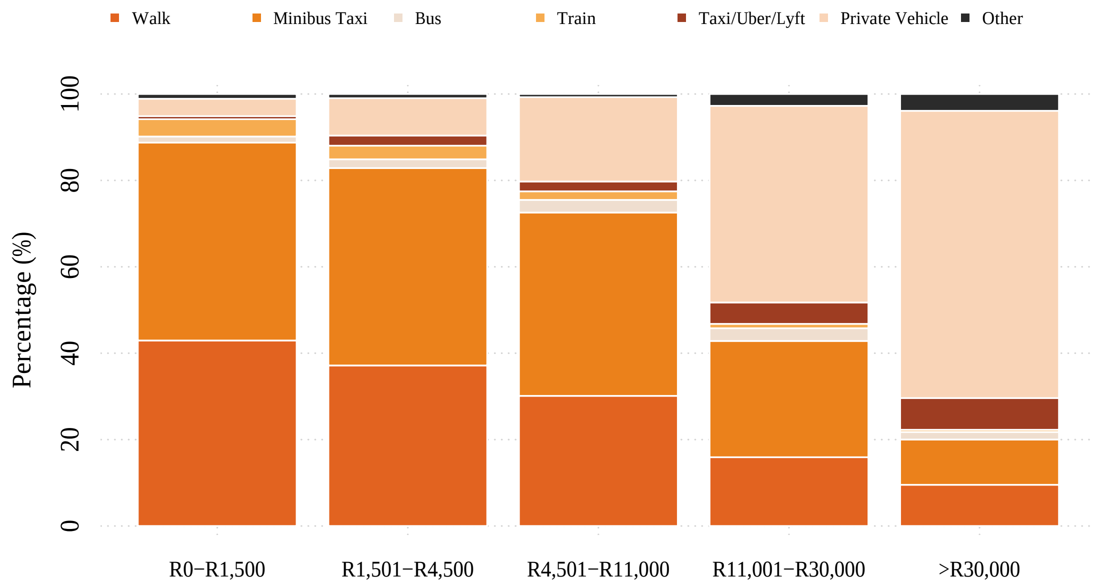
A.1 Appendix Figures

Figure A.1: Main Mode of Transportation



Note: Proportion of households that mentioned each mode as their main mode of transportation in Gauteng Province. Source: South Africa National Household Travel Survey, 2020

Figure A.2: Main Mode of Transportation by Income

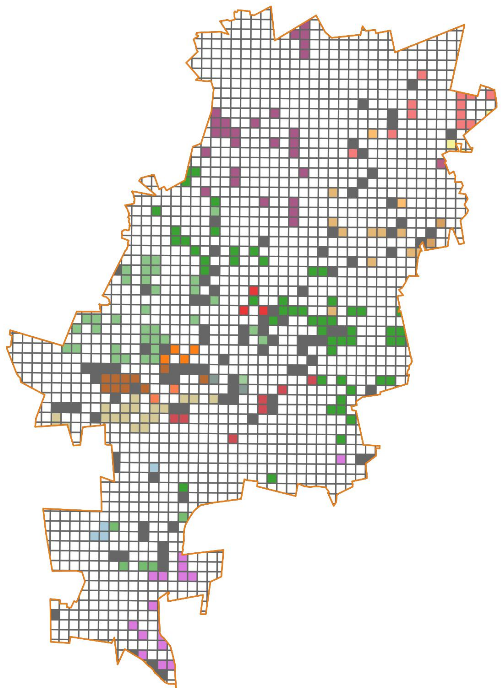


Note: Proportion of households that mentioned each mode as their main mode of transportation by income group in Gautend Province. Source: South Africa National Household Travel Survey, 2020

Figure A.3: Minibus Taxi Route Census – 2022

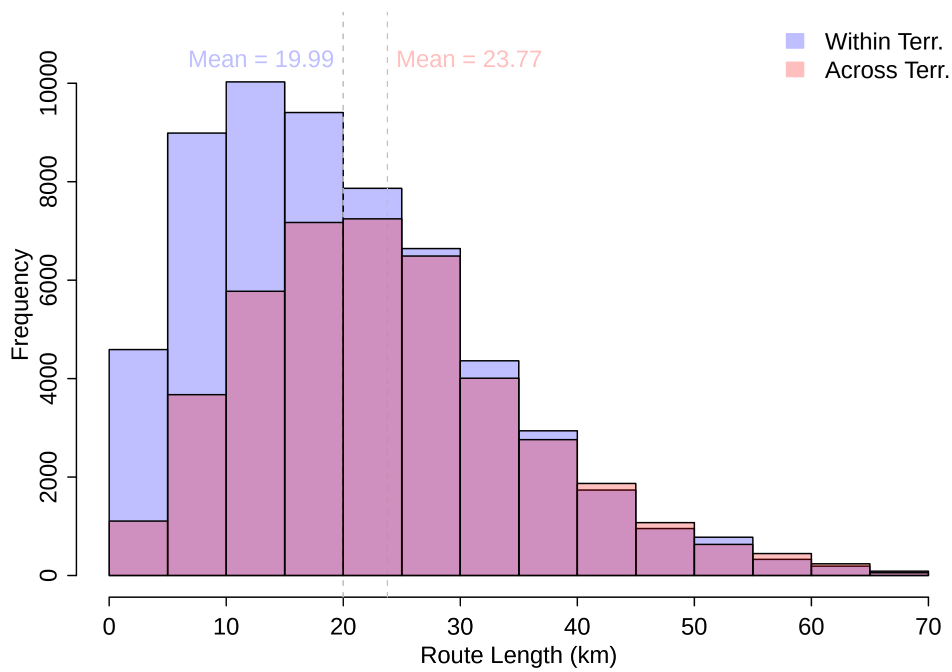
Note: This map plots all routes collected during a census of minibus taxi routes in the Johannesburg area in 2022 by WhereIsMyTransport (WIMT).

Figure A.4: Grids with Minibus Ranks



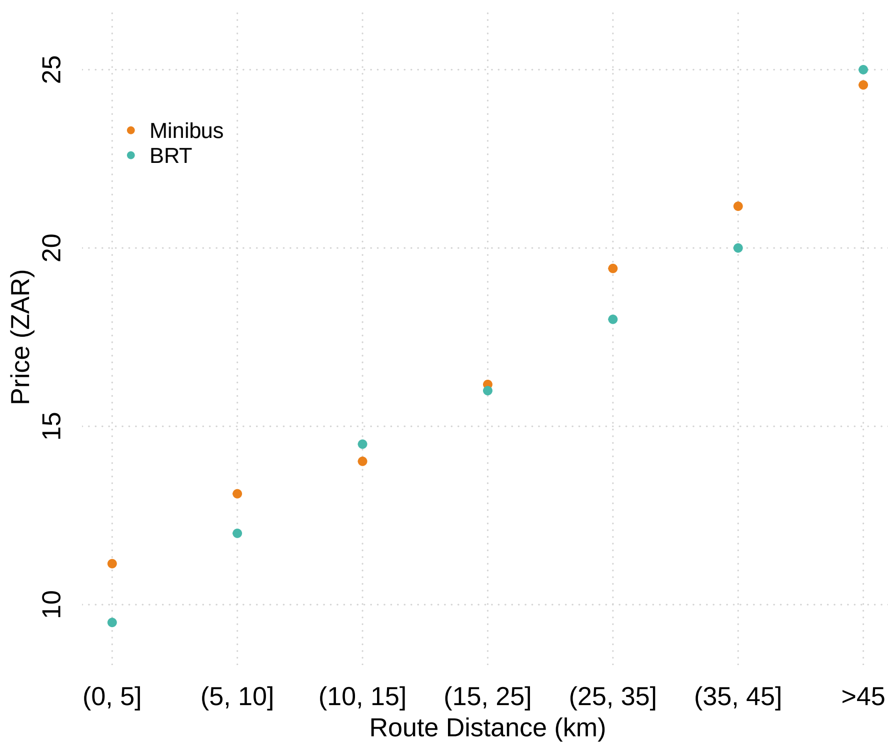
Note: Grids with a minibus taxi rank (stop) as observed using the census of minibus taxi routes in Johannesburg. Grids are colored based on the association they are assigned to. Dark grey grids have a minibus rank but are not assigned to any association. Light grey grids do not have a minibus rank.

Figure A.5: Distribution of Route Lengths



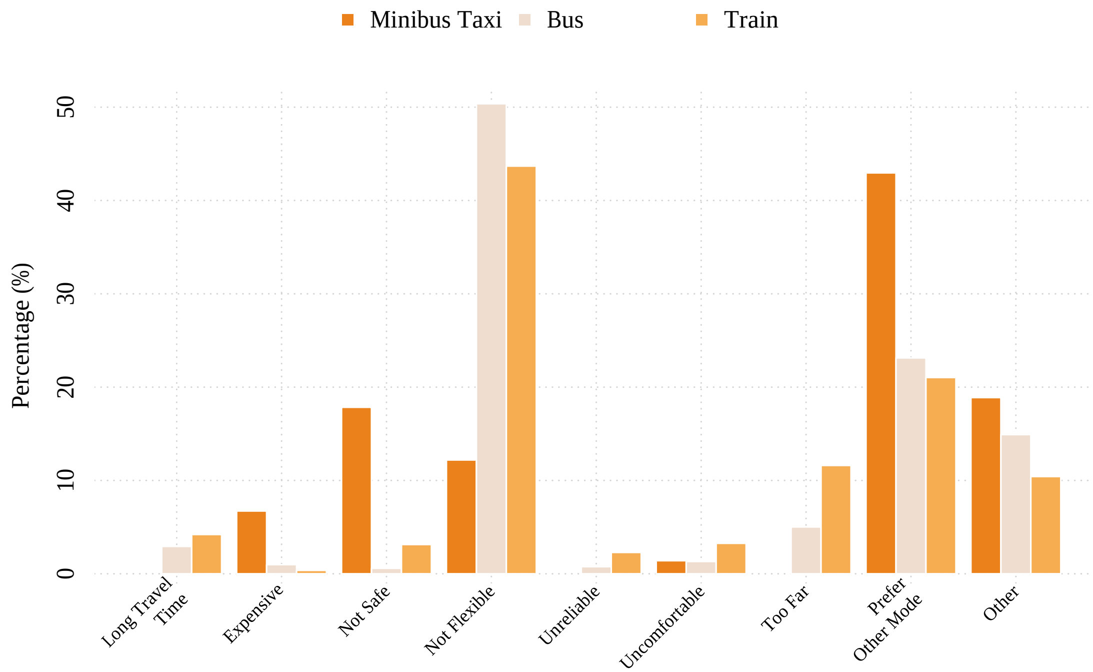
Note: Distribution of the route lengths by type of route. Route length is defined as the Euclidean distance between the centroids of the origin and destination grids. The blue bars plot the distribution of route lengths for within-territory routes, and the red bars plot this distribution for across-territory routes.

Figure A.6: Route Prices: Minibus Taxis and BRT Lines



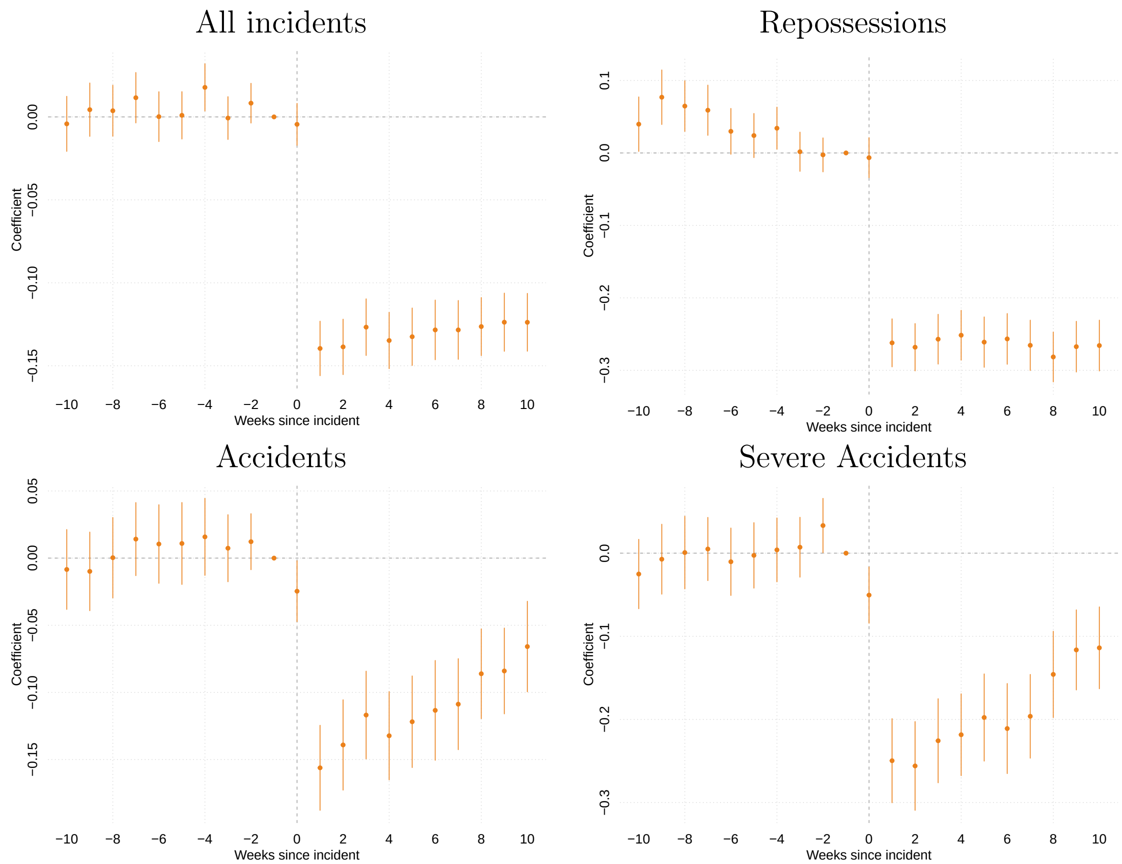
Note: Route fares for minibus taxis and government provided bus rapid transit (BRT). BRT lines are priced based on total km travelled. Route distance bins are based on the BRT pricing schedule. Plotted minibus taxi prices are the average fare for all minibus routes in each distance bin

Figure A.7: Reasons for not Using Modes of Transportation



Note: Reasons households gave for not using each mass transit mode. The following options are classified as "Long Travel Time": "Travel time is too long/slow" and "Have to change transport (transfer)". "Expensive" has the following options: "Mode too expensive". "Not Safe" has the following options: "Too much crime (too dangerous)" and "Too many accidents". For minibus taxis only, the following options were also given and categorized under "Not Safe": "Too much violence/wars", "Taxis not roadworthy", "Drivers drive recklessly". "Not Flexible" has the following options: "No «mode» available at all", "«mode» not available often enough", "«mode» not available at the right times", "«mode» don't go where needed". "Unreliable" has the following options: "«mode» always late". "Uncomfortable" has the following options: "«mode» too crowded". "Too Far" has the following options: "Station too far from home", "Station too far from destination". Source: South Africa National Household Travel Survey, 2020

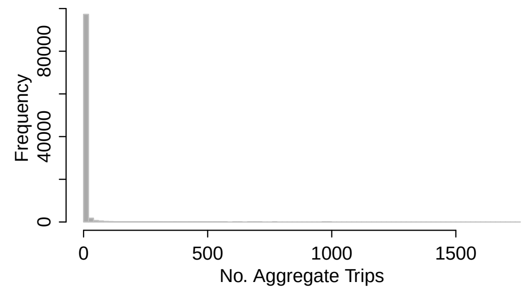
Figure A.8: Effect of Insurance Incidents on Vehicle Operation



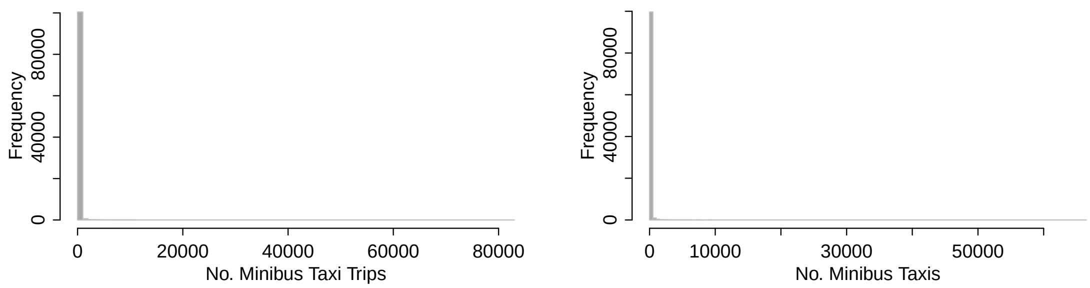
Note: Weekly coefficients for the effect of different insurance claim types on the probability of minibus operation. Standard errors are clustered at the minibus level.

Figure A.9: Distribution of Commuter Trips and Minibus Supply

(a) Aggregate No. Trips

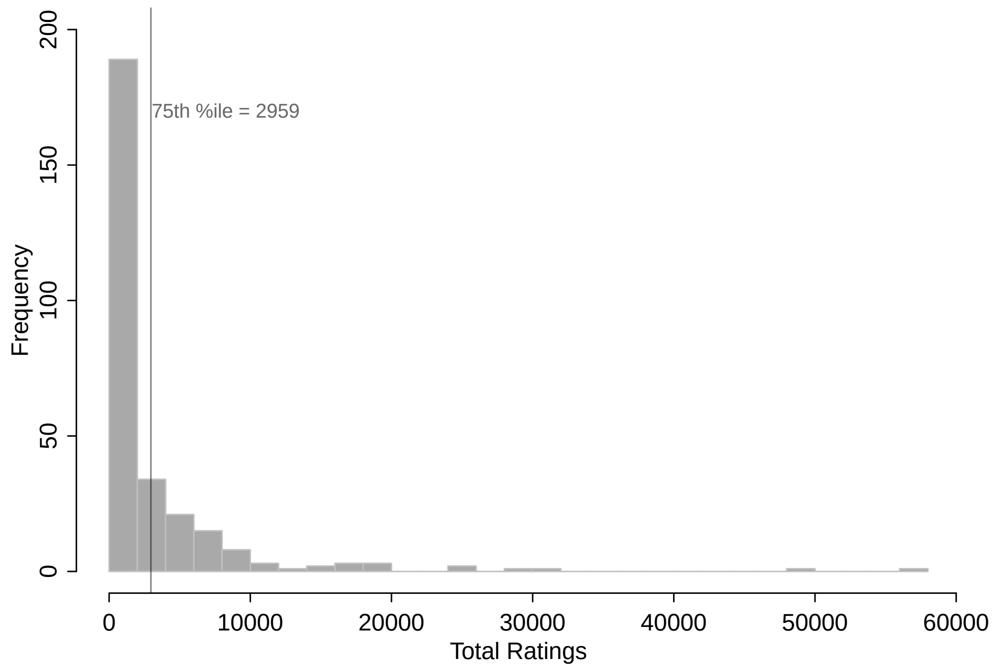


(a) Minibus Supply



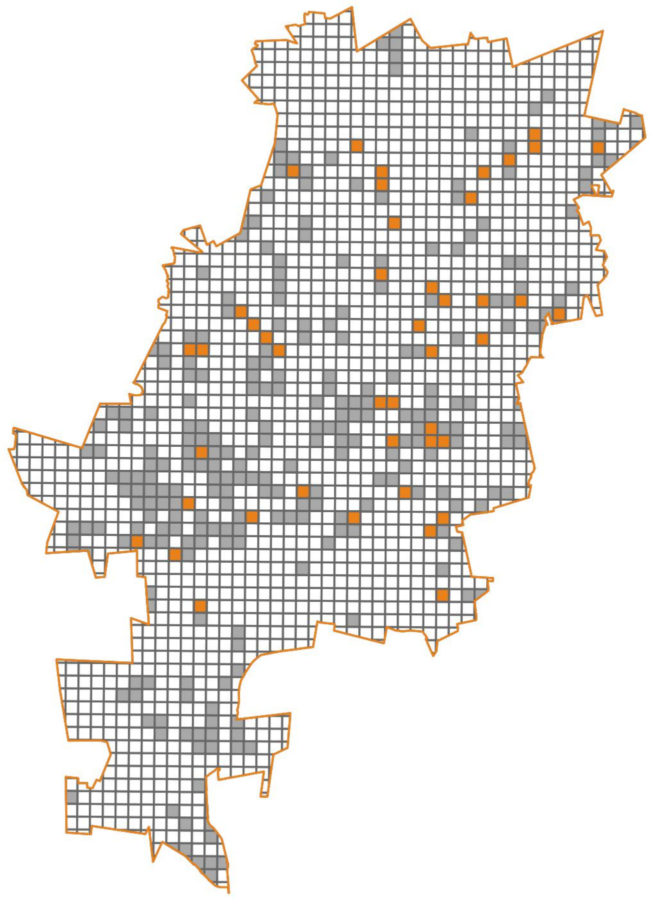
Note: Distribution of aggregate no of trips and minibus supply at the route level. This figure plots the histogram of the number of trips on each route across the sample's 10-month period. Panel (a) plots the count of commuter trips and panel (b) plots the distribution for our two measures of minibus supply – the number of minibus trips on the left, and the number of minibus taxis on the right.

Figure A.10: Distribution of the number of Google Reviews



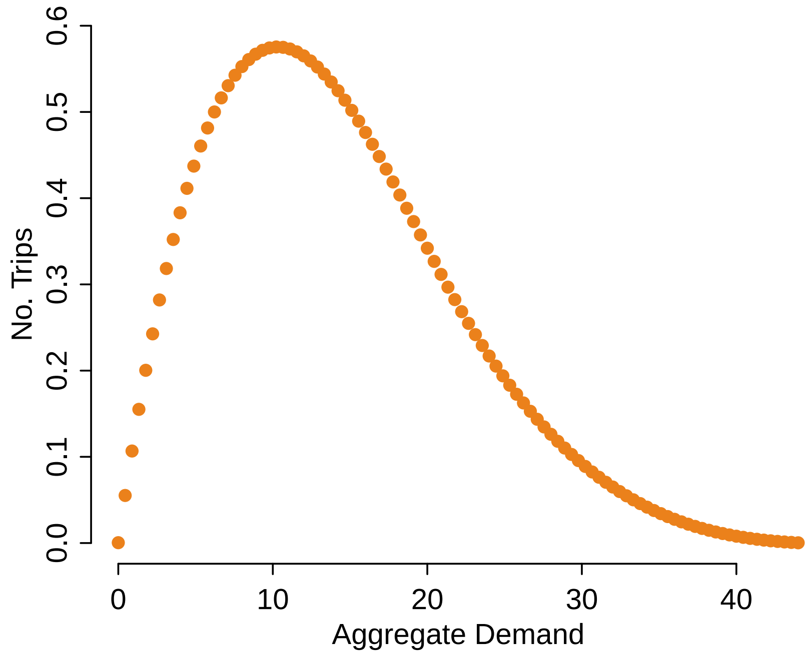
Note: Distribution of the number of mall ratings on Google for grids with with at least one mall.

Figure A.11: Minibus Grids with a Large Mall



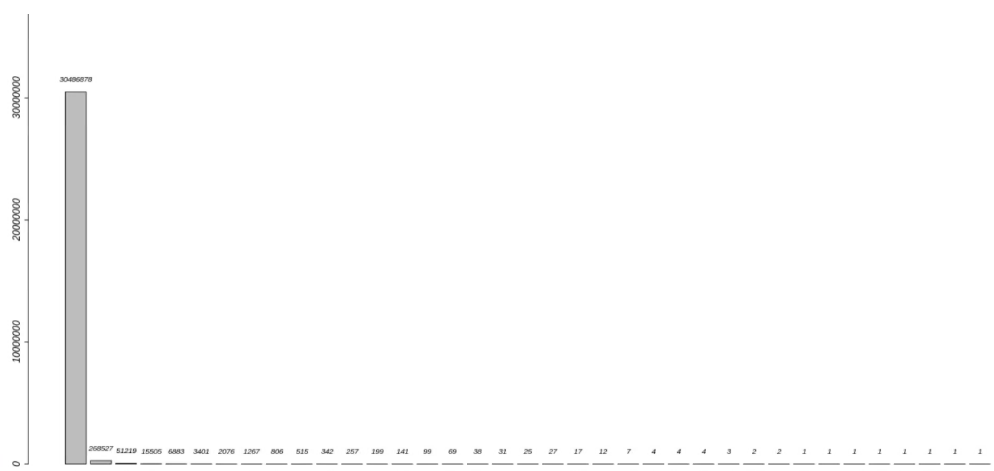
Note: Grids with a large mall in the Johannesburg Metropolitan Area. Grey grids represent our full sample of grids which contain a minibus taxi rank. Grids highlighted in orange contain a large mall.

Figure A.12: Non-parametric Estimation of Minibus Trips as a function of Aggregate Demand



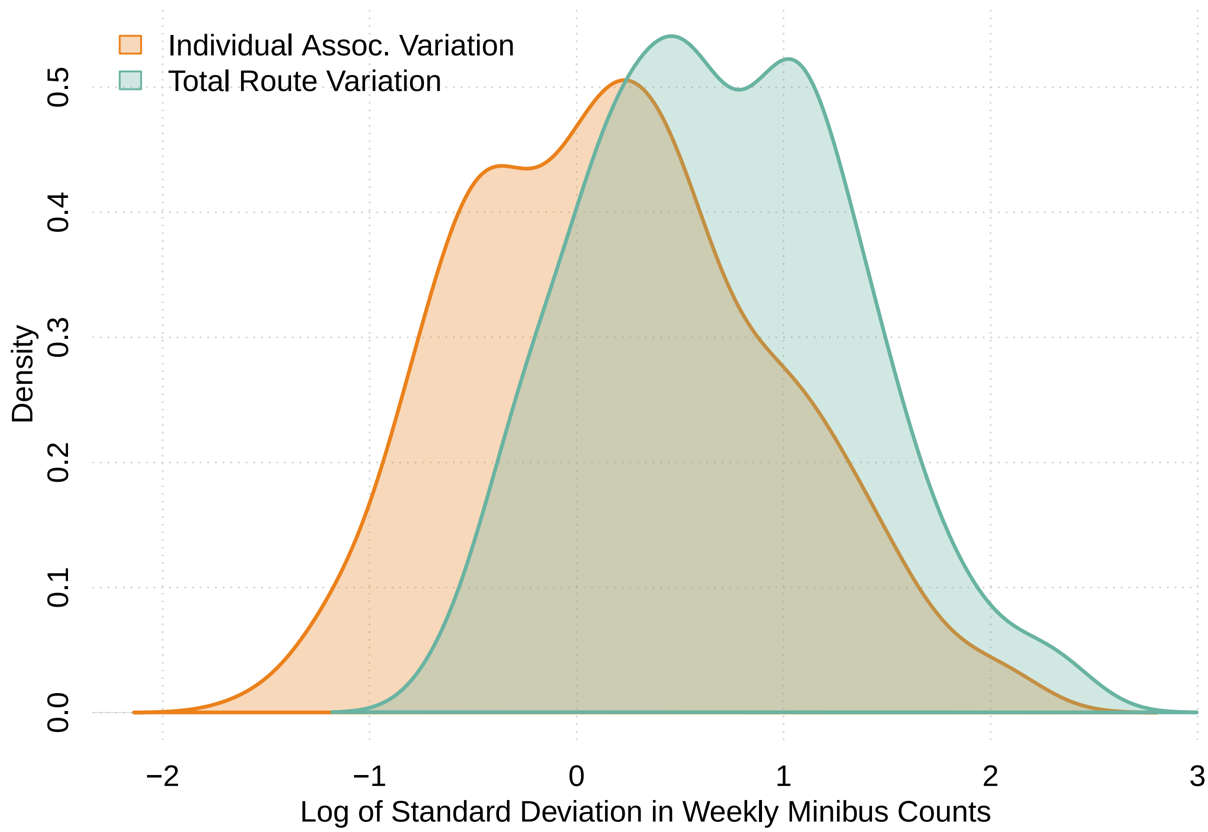
Note: This figures plot the estimates from the non-parametric estimation of the number of minibus trips as a function of aggregate demand, evaluated over the full range of aggregate demand values.

Figure A.13: Distribution of Aggregate Demand Counts



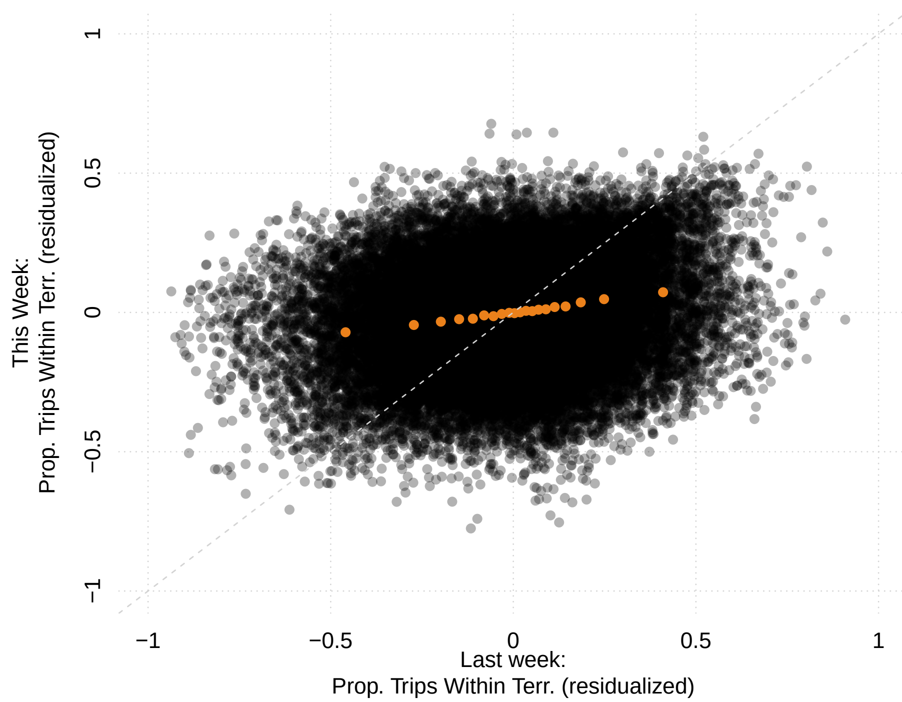
Note: This figure presents the bar chart of the commuter trip counts at the route-day level.

Figure A.14: Variation in Minibus Supply on Across-Territory Routes



Note: This figure compares variation in minibus supply across routes. For each across-territory route, we calculate the standard deviation of weekly minibus counts over our study period, separately for total route supply (combined service from both associations) and for each individual association's contribution. We then take the log of these standard deviations to account for right-skewed distributions. The figure plots the distribution of these log standard deviations across routes.

Figure A.15: Proportion of Trips Within-Territory By Week



Note: Figure shows scatter plot at the minibus-week level. Y-axis shows proportion of within-territory trips in current week, X-axis shows proportion in previous week, both residualized by association size. Orange dots represent binned averages. Grey dotted line indicates 45-degree reference line.

A.2 Appendix Tables

Table A.1: Summary Stats – Minibus Taxi Routes

	All Minibus Routes	Recreational Routes	Within Terr. Routes	Across Terr. Routes
Prop. Within Terr.	0.58	0.52	1.00	0.00
Prop. CBD Route	0.07	0.01	0.08	0.07
Avg. Route Length	21.58	21.08	19.99	23.77
N.	101,442	26,640	58,896	42,546

Note: Summary Statistics for different categories of routes. The first column provides statistics for all routes in our data. The second column provides statistics for routes categorized as recreational, the third column for routes categorized as being within-territory, and the fourth column for routes categorized as being across-territory.

Table A.2: Summary Stats – Minibus Route Trips

	Minibus Trips	Commuter Trips
N. Routes w/ $\geq$ 1 Trip	31,433	33,198
Median Trip Duration (min)	31.32	46.22
Mean Trip Duration (min)	76.54	101.74
Median No. Trips per day	0.00	0.00
Mean No. Trips per day	0.51	0.05
Median No. Vehicles/Devices per day	0.00	0.00
Mean No. Vehicles/Devices per day	0.44	0.05

Note: Statistics on trips on minibus routes. The first column displays statistics for the minibus taxi trips and the second column displays statistics for aggregate trips constructed from smartphone pings. "N. Routes w/ $\geq$ 1 Trip" is the number of routes that have at least one minibus or aggregate trip during our observation period.

Table A.3: Price by Route Type

	Price		
	(1)	(2)	(3)
Constant	15.4***	15.6***	
	(0.283)	(0.343)	
Within Terr.	−0.689**	−0.746**	−0.069
	(0.315)	(0.337)	(0.280)
Route Length	0.284***	0.278***	0.237***
	(0.008)	(0.009)	(0.013)
Dist. to CBD		−0.016	−0.217***
		(0.014)	(0.042)
N.	1,167	1,167	1,167
Origin FEs			Yes
Destination FEs			Yes

Note: Route prices by route type. Pricing data is from the WIMT census of minibus taxi routes collected in 2022. Within terr is an indicator for whether the route is within territory, Route length is the demeaned total route distance, dist. to cbd is the minimum distance to the Johannesburg central business district. Standard errors are two-way clustered by origin and destination. *p<0.1; **p<0.05; ***p<0.01.

Table A.4: Minibus and Commuter Trips by Route Type

	Minibus Trips		Commuter Trips	
	(1)	(2)	(3)	(4)
Route Length	−0.17***	−0.15***	−0.23***	−0.22***
	(0.02)	(0.01)	(0.01)	(0.01)
Within Terr.		1.22***		0.18***
		(0.15)		(0.05)
N.	30,645,024	30,645,024	30,838,368	30,838,368
Origin FEs	Yes	Yes	Yes	Yes
Destination FEs	Yes	Yes	Yes	Yes

Note: Coefficients are estimated using pseudo-Posson Maximum Likelihood. Standard errors are two-way clustered at the origin-destination level. Route length is the Euclidean distance in Km between the origin grid centroid and destination grid centroid.

Table A.5: Covariate Balance for Minibuses with and without Insurance Claims

	Prop. Trips Within Terr.	Assoc Fleet Size	No. Unique Routes
Had an Incident	0.003	-13.4^{**}	-2.44
	(0.008)	(5.37)	(1.58)
Was Repossessed	-0.036^{***}	17.6^{*}	-7.93^{***}
	(0.014)	(9.51)	(2.81)
Had an Accident	0.016	-14.9	7.53^{***}
	(0.014)	(9.67)	(2.86)
Had a Severe Accident	0.015	-13.7	0.652
	(0.019)	(13.2)	(3.92)
Control Mean	0.847	147.1	36.5
N.	8,841	8,042	8,841

Note: Difference in characteristics between minibuses with an incident and those without for minibuses that operate on a minibus route in Johannesburg. Each row is a separate regression. Prop. Within-terr trips is the proportion of all trips taken by that vehicle that are on within territory routes. Assoc fleet size is the size of the association (total number of minibuses) that the vehicle belongs to, and No. Unique Routes is the number of unique routes the vehicle operated on during the time period. The control mean is the average value of each independent variable for the minibuses with no incident reported during our time period. *p<0.1; **p<0.05; ***p<0.01.

Table A.6: The Impact of Insurance Incidents on Minibus Operation, with Controls

	Minibus Operation			
	All Incidents	Repossessions	Accidents	Severe Accidents
	(1)	(2)	(3)	(4)
10-week Average	-0.092^{***}	-0.198^{***}	-0.090^{***}	-0.129^{***}
	(0.007)	(0.013)	(0.013)	(0.019)
N.	647,235	647,235	647,235	718,155
Minibus. FEs	Yes	Yes	Yes	Yes
Week FEs	Yes	Yes	Yes	Yes

Note: Same regression as in Panel A of Table 3 with controls for the interaction of post-incident indicator and minibus association fleet size, total number of routes minibus operated on, and number of trips taken.

Table A.7: Supply Response to Partner Association on Across-Territory Routes

	Number of Minibuses	
	Association A	Association A
	(1)	(2)
Contant	1.44^{***}	
	(0.075)	
Association B	-0.404^{***}	-0.372^{***}
	(0.098)	(0.054)
N.	33,414	33,414
Day FEs		Yes
Route FEs		Yes

Notes: *p<0.1; **p<0.05; ***p<0.01.

Note: This table examines how associations respond to changes in their partner's supply on across-territory routes. The dependent variable is the number of minibuses allocated by one association, and the key independent variable is the number of minibuses allocated by their partner association on the same route-day. The unit of observation is a route-day. The sample includes all across-territory routes served by two associations between October 2022 and July 2023. *p<0.1; **p<0.05; ***p<0.01.

Table A.8: Cyclicality "First Stage" and "Reduced Form" by Route Type

	First-Stage	Reduced Form	
	Commuter Trips	Minibus Trips	No. Minibuses
(1)	(2)	(3)	
Panel A: All Routes			
demand_day × has_mall	0.045***	0.032***	0.026***
	(0.008)	(0.005)	(0.005)
demand_day × has_mall × dist_quintile_2	−0.020	−0.018***	−0.013***
	(0.013)	(0.006)	(0.004)
demand_day × has_mall × dist_quintile_3	−0.025	−0.006	−0.006
	(0.021)	(0.005)	(0.005)
demand_day × has_mall × dist_quintile_4	0.018	0.005	−0.006
	(0.029)	(0.008)	(0.007)
demand_day × has_mall × dist_quintile_5	−0.034	−0.022***	−0.013
	(0.063)	(0.007)	(0.009)
N.	10,092,192	9,555,936	9,555,936
Panel B: Within Terr. Routes			
demand_day × has_mall	0.038***	0.031***	0.025***
	(0.010)	(0.006)	(0.006)
demand_day × has_mall × dist_quintile_2	−0.051***	−0.014**	−0.009***
	(0.018)	(0.006)	(0.003)
demand_day × has_mall × dist_quintile_3	−0.011	0.002	0.004
	(0.031)	(0.006)	(0.004)
demand_day × has_mall × dist_quintile_4	0.025	0.006	−0.007
	(0.043)	(0.009)	(0.007)
demand_day × has_mall × dist_quintile_5	−0.091	−0.030***	−0.024**
	(0.072)	(0.009)	(0.010)
N.	6,637,536	6,270,608	6,270,608
Panel C: Across Terr. Routes			
demand_day × has_mall	0.071***	0.030*	0.026**
	(0.018)	(0.017)	(0.011)
demand_day × has_mall × dist_quintile_2	0.020	−0.024	−0.022**
	(0.026)	(0.017)	(0.011)
demand_day × has_mall × dist_quintile_3	−0.044	−0.013	−0.018
	(0.030)	(0.018)	(0.013)
demand_day × has_mall × dist_quintile_4	0.007	−0.003	0.006
	(0.043)	(0.036)	(0.032)
demand_day × has_mall × dist_quintile_5	0.061	0.058	0.085**
	(0.110)	(0.038)	(0.039)
N.	3,454,656	3,285,328	3,285,328
Day FEs	Yes	Yes	Yes
Route FEs	Yes	Yes	Yes

Note: This table replicates the estimates from Table 6 but includes controls for route distance quintiles. *p<0.1; **p<0.05; ***p<0.01.

Table A.9: Reduced Form – Cyclical Instrument on Minibus Supply

	Minibus Trips			No. Minibuses		
	(1)	(2)	(3)	(4)	(5)	(6)
demand_day	0.015***	0.0003		0.003	−0.008***	
	(0.005)	(0.004)		(0.005)	(0.003)	
has_mall		0.027			0.028	
		(0.080)			(0.082)	
demand_day × has_mall		0.030***	0.030***		0.023***	0.023***
		(0.007)	(0.007)		(0.006)	(0.006)
N.	9,555,936	30,645,024	9,555,936	9,555,936	30,645,024	9,555,936
Day FEs			Yes			Yes
Route FEs	Yes		Yes	Yes		Yes
Weekday FEs	Yes	Yes		Yes	Yes	
Holiday FEs	Yes	Yes		Yes	Yes	
Month FEs	Yes	Yes		Yes	Yes	
Origin FEs		Yes			Yes	
Destination FEs		Yes			Yes	

Note: Coefficients are estimated using Pseudo-Poisson Maximum Likelihood. demand_day is an indicator for whether the day is at the "end of the month" (after the 25th and before the 10th). has_mall is an indicator for whether a route has a large mall at either its origin or destination. For (1), (2), and (3), the dependent variable is the number of minibus taxi trips on a given route on a given day. For (4), (5), and (6), the dependent variable is the number of minibus taxis on a given route on a given day. Standard errors are two-way clustered by origin and destination. *p<0.1; **p<0.05; ***p<0.01.

Table A.10: The effect of aggregate demand on minibus taxi supply by route type, controlling for route length

	Minibus Trips		No. Minibuses	
	Within Terr.	Across Terr.	Within Terr.	Across Terr.
	(1)	(2)	(3)	(4)
Demand Trips	0.150	-0.907^{**}	0.140	-0.881^{**}
	(0.130)	(0.406)	(0.113)	(0.387)
Within Terr. - Across Terr.	1.057		1.021	
Day FEs	Yes	Yes	Yes	Yes
Origin FEs	Yes	Yes	Yes	Yes
Dest FEs	Yes	Yes	Yes	Yes
N.	17,673,040	12,933,984	17,673,040	12,933,984

Notes: *p<0.1; **p<0.05; ***p<0.01.

Note: Coefficients are estimated using Pseudo-Poisson Maximum Likelihood with a generalized instrumental variable control function and include controls for route length. aggregate_trips is the number of aggregate demand trips on route r and day t. Column (1) and (3) show the effect of aggregate demand on the number of minibus taxi trips, and columns (2) and (4) show the effect on the number of minibus taxi vehicles on the route. Standard errors are two-way clustered by origin and destination. *p<0.1; **p<0.05; ***p<0.01.

Table A.11: First Stages by Route Type and Distance

	Within Terr.		Across Terr.	
Dependent Variables:	commuter_trips	commuter_trips × route_length	commuter_trips	commuter_trips × route_length
Model:	(1) Poisson	(2) OLS	(3) Poisson	(4) OLS
has_mall	0.047	1.21***	-0.052	0.063
	(0.064)	(0.407)	(0.096)	(0.043)
route_length	-0.237***	0.049***	-0.179***	0.014***
	(0.007)	(0.006)	(0.008)	(0.003)
demand_day × route_length	0.011**	-0.017***	-0.009	-0.003***
	(0.005)	(0.003)	(0.006)	(0.0009)
demand_day × has_mall	-0.356**	0.146**	0.265*	-0.029***
	(0.149)	(0.060)	(0.144)	(0.008)
demand_day × has_mall × route_length	-0.025***	0.078***	0.017	0.012***
	(0.009)	(0.012)	(0.012)	(0.004)
N.	17,866,080	17,904,384	12,933,984	12,933,984
Day FEs	Yes	Yes	Yes	Yes
Origin FEs	Yes	Yes	Yes	Yes
Destination FEs	Yes	Yes	Yes	Yes

Note: First stage regressions for the instruments of demand_trips and demand_trips × route_length. route_length is demeaned so that other variables can be interpreted at the average route length. *p<0.1; **p<0.05; ***p<0.01.

Table A.12: Total Minibus Commuter Trips – Recreational Routes on Non-demand days

	Within Terr. Routes	Across Terr. Routes
12:00am - 4:59am	27,455	12,779
5:00am - 9:59am	153,924	57,478
10:00am - 2:59pm	258,603	74,545
3:00pm - 7:59pm	157,010	59,853
8:00pm - 11:59pm	22,612	9,295
Total	619,604	213,950

Note: This table provides the estimated count of the number of minibus taxi trips taken by commuters in Johannesburg on recreational routes on non-demand days using a combination of smartphone data, census data, and household travel surveys.

Table A.13: Estimated Headway Elasticity

	Minibus Headway		
	NP	NPIV	GCF
	(1)	(2)	(3)
Overall Elasticity	1.352***	−0.820	−0.981***
	(0.348)	(0.588)	(0.022)
Within Terr. Elasticity	−0.797	−1.007	−1.190***
	(0.862)	(1.231)	(0.023)
Across Terr. Elasticity	2.108**	−0.599*	−0.331***
	(1.023)	(0.353)	(0.011)
Within Terr. − Across Terr.	−2.905**	−0.408	−0.859***
	(1.346)	(1.295)	(0.027)
Day FEs	Yes	Yes	Yes
Route FEs	Yes	Yes	
Origin FEs			Yes
Destination FEs			Yes

Note: This table summarizes our estimates of the elasticity of average headway. Columns (1) and (2) estimate the elasticity using non-parametrics (NP) without an instrumental variable. Columns (3) and (4) use non-parametric instrumental variable (NPIV) estimation, and columns (5) and (6) use a generalized control function approach (GCF). Columns (1), (3), and (4) display elasticities for the number of minibus trips, while (2), (4), and (6) display results for the number of minibus vehicles. All standard errors are obtained via bootstrapping and are clustered at the origin level.

*p<0.1; **p<0.05; ***p<0.01.

A.3 More on Minibus Taxi Operations

In South Africa, the minibus taxi industry is organized around associations. All minibus owners and operators must be a member of an association. These associations control most aspects of the industry's operations, forming the backbone of its structure. Associations tightly control membership, typically only allowing new owners who are currently drivers within the association, thus maintaining a close network of operators. Members are required to pay monthly fees to retain status. To prevent the concentration of power within associations and maintain equity, associations often limit the number of minibuses each individual can own.

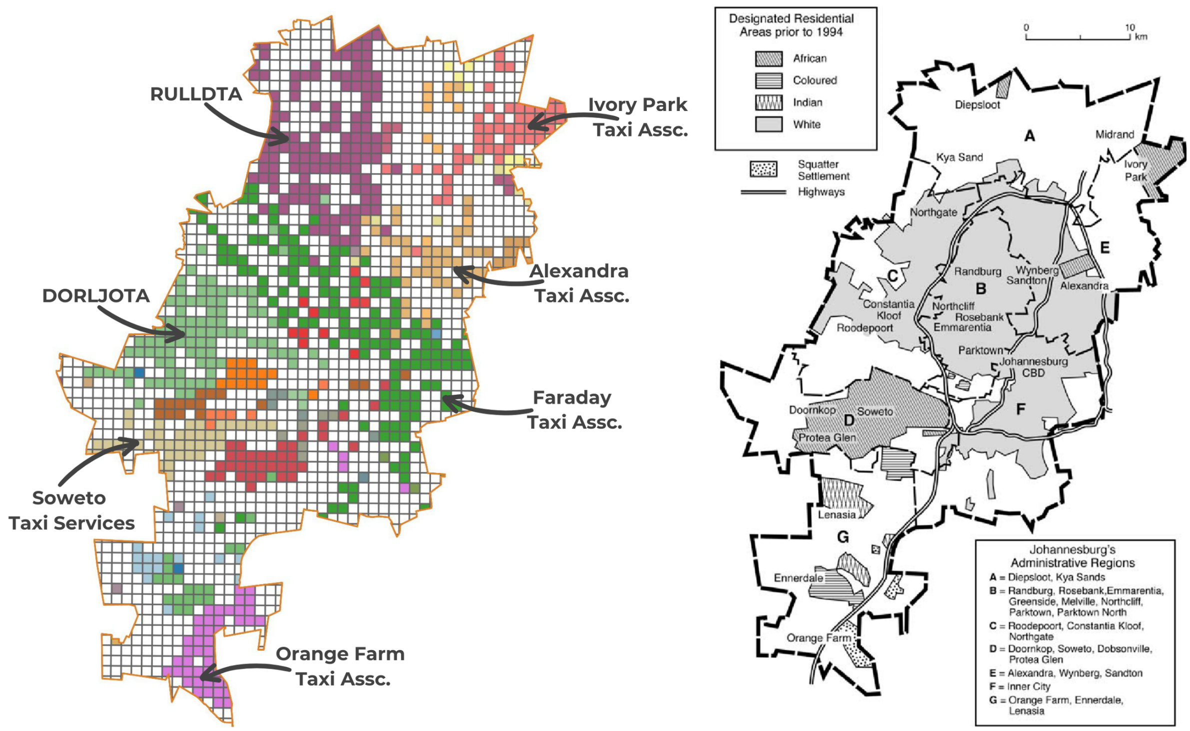
South African associations' territories are closely tied to historical apartheid "townships". The non-white populace was forcibly relocated to these townships outside the Central Business District (CBD) in Johannesburg during the 1950s. The minibus taxi emerged as a way to transport residents within the townships and into the CBD. After the deregulation of the industry post-apartheid, the industry grew exponentially, and local minibuses aggregated into regional associations. As such, association territories are highly correlated with historical township boundaries; see Figure A.16.

Within this organizational framework, the relationship between drivers and vehicle owners is structured around rental agreements. Drivers typically operate on daily or weekly fee structures, paying a set amount to owners and retaining the remaining profits. Drivers are responsible for fuel costs, and owners cover all other expenses related to the vehicle – maintenance, licensing, etc.

The day-to-day operations of minibus taxis are centered around taxi ranks, which serve as hubs for passenger pickup and dropoff. These ranks come in two main types: large, formal ranks that resemble garages and serve multiple routes, and smaller roadside ranks that are essentially curb-side stops serving single routes. At these ranks, operations are often managed by conductors who organize the supply of minibuses. They maintain the queues of minibuses, and can call for additional vehicles when demand is high.

Despite the organized structure of associations and ranks, the minibus taxi industry faces significant challenges, particularly related to competition and violence. The industry is plagued by conflicts stemming from both inter-association and intra-association rivalries. Associations often engage in turf wars, fighting for control over ambiguous or overlapping routes. Within associations, owners and drivers compete for access to the most profitable routes. This fierce competition can escalate to property damage, such as broken windshields, and in more severe cases, has led to shoot-outs and even assassinations.

Figure A.16: Association Territories and Historical Townships



Note: The left panel plots the map of grid associations in Johannesburg, as measured by minibus operations in that grid. Each color corresponds to a different association. RULLDTA is the common acronym for the Randburg United Local and Long Distance Taxi Association, and DORLJOTA is the common acronym for the Dobsonville, Roodepoort, Leratong, Johannesburg Taxi Association. The right panel plots the historical map of designated residential areas during Apartheid. Source: McKay and Bell (2011).

A.4 Minibus Data Representativeness

We examine how representative our data on minibuses is of the full market. To do so, we rely on published gazettes of all applications for minibus operating licenses.

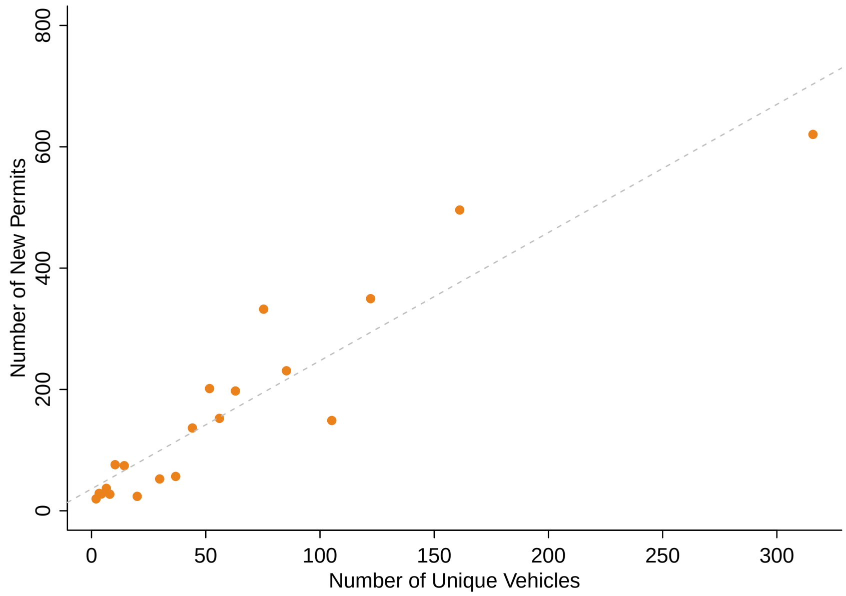
The Department of Transport for each province in South Africa is required to publish all operating license applications for minibus taxi operations each week. These gazettes contain information on the type of application (new operating license or transfer of operating license), the routes the owner is applying for, and the association the owner belongs to. We digitize these gazettes from March of 2017 until March of 2020. After 2020, the printing works department was backlogged due to COVID, and as far as we are aware, there have been no further gazettes published for Johannesburg since.

We compare the number of associations and the flow of vehicles in our minibus data to the associations and flow of applications for operating licenses in the gazette. Our minibus data contains observations from 95% of all associations in the gazette that applied for a new operating license for a route within Johannesburg between March 2017 and March 2020. The data share of these associations is also correlated with the number of applications of the association in the gazette. Figure A.17 displays this correlation. Associations that applied for more licenses also have more vehicles in our data on average.

Our data partner estimates that they finance 20% of all minibus taxis nationwide. Our analyses using the gazette data estimate that within Johannesburg, this proportion is closer to 48%. We calculate the average flow of new vehicles in our data set and compare this to the average flow of new operating licenses (note that we cannot calculate the stock number of minibuses in operation using the published licenses because we do not observe when a minibus stops operating). In our data, on average, 270.8 new minibuses are added each month. According to the gazette, there are on average 563.4 new operating license applications each month suggesting that our data partner finances approximately 48% of these new minibuses. This discrepancy with their nationwide estimate of 20% is likely because their market penetration is much lower in other parts of South Africa. The company headquarters are in Gauteng and this encompasses the lion's share of their operations.

The gazette is our best measure of the flow of minibuses operating in South Africa. Of course, it may be the case that some minibuses operate without applying for an operating license. It may also be the case that applications are denied and thus the number of applications is an overestimate of the total number of minibuses operating. Anecdotally, applying for an operating license is low-cost and the department of transportation eventually approves all the applications they receive (though they typically have a very large backlog).

Figure A.17: Association Representation



Note: Binned correlation between the number of unique vehicles per association in our data and the number of new operating licenses the association applied for between 2017 and 2020. Each observation is an association which applied for at least one operating licence for a rout in Johannesburg between March of 2017 and March of 2020.

A.5 Minibus Financier Pre-processed Data

Our minibus financier's tracker receives pings from each minibus taxi every 20 seconds. They first pre-process the raw pings into their own definition of trips before sharing this data with us. We use the pre-processed trips data to define minibus operations on our sample of routes.

Below are the events as defined by the financier;

Stop: A stop is any event in which the average speed of the minibus drops below 5km/h. Average speed is calculated using the time and distance between GPS pings.

Long-stop: A long-stop occurs when more than 5 minutes have elapsed, and less than 200m of odometer distance is traveled.

Base: A base a the clustering of long-stops for the minibus based on a DB scan clustering algorithm with minimum distance set to 150m, and minimum number of observations set to 3.

Trip: A trip (what we have access to) is any event stating at a base and ending at a base with more than 1km of distance traveled.

A.6 Smartphone Data Representativeness

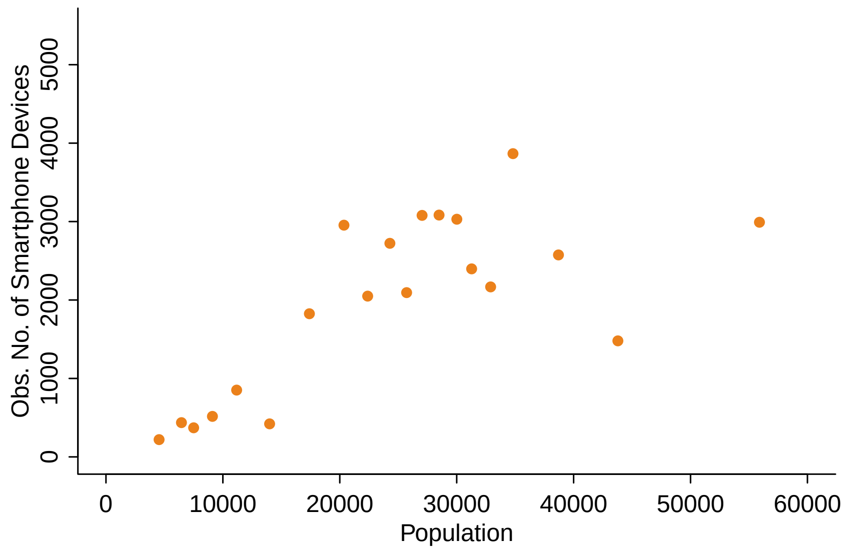
We examine how representative our smartphone data is of the total Johannesburg population and the income distribution. We compare the proportion of devices from a given home location to the population proportion and income proportion of that home location as measured by the South Africa 2011 census.

We first assign each smartphone device in our data to a census ward based on the modal ward where the device is located between 1am and 5am. The South Africa census divides Johannesburg into 506 wards.

Figure A.18 shows that the number of smartphone devices in our data is positively correlated with the population in Johannesburg. However, the data is biased towards higher income residents in Johannesburg. Figure A.19 plots the proportion of smartphone devices in our data on the percent of the population under the poverty line. We observe more devices from wards with a lower proportion of the population under the poverty line as defined by South Africa.

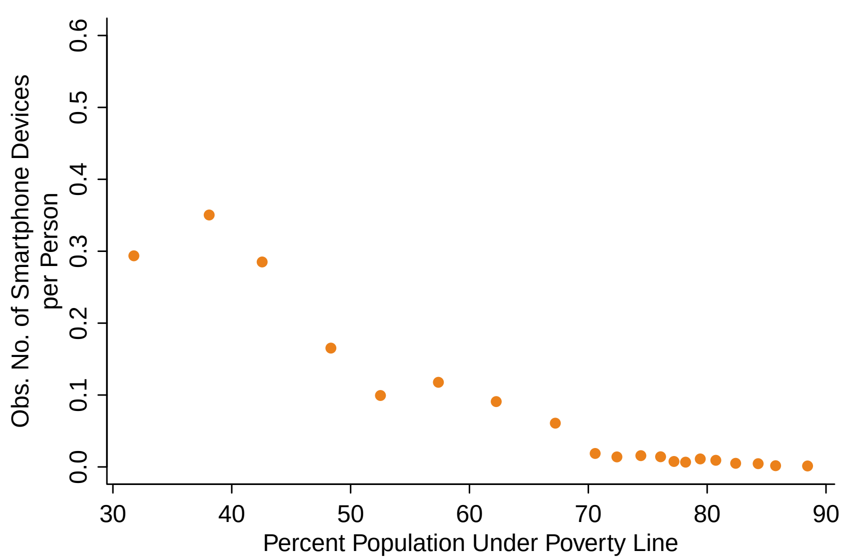
Given this income imbalance, we re-run our main results using data that is weighted in order to be more representative of all income levels. In each ward, we calculate the proportion of the population that lives in that ward and the proportion of all our smartphone devices that are assigned a home in that ward. The device weight is the ratio of these two values. Table A.14 shows the first stage of our instrumental variable using weights and Table A.15 displays the main results. The pattern of results is similar to our main specification, but estimates are more noisy.

Figure A.18: Smartphone Representativeness: By Population



Note: Binscatter plot of the correlation of observed smartphone devices on the total population. Each observation is a ward in South Africa (from the 2011 South Africa Census).

Figure A.19: Smartphone Representativeness: Share below Poverty



Note: Binscatter plot of the number of observed smartphone devices per person on the percent of the population under the poverty line. Each observation is a ward in South Africa (from the 2011 South Africa Census).

Table A.14: First Stage Using Reweighed Data

	OLS		PPML	
	(1)	(2)	(3)	(4)
demand_day	-0.0004		-0.039^{*}	
	(0.0003)		(0.023)	
has_mall	-0.018^{**}		-0.022	
	(0.008)		(0.068)	
demand_day × has_mall	0.001^{**}	0.001^{**}	0.070^{**}	0.070^{**}
	(0.0007)	(0.0007)	(0.031)	(0.031)
N.	30,838,368	30,838,368	30,838,368	10,092,192
Day FEs		Yes		Yes
Route FEs		Yes		Yes
Weekday FEs	Yes		Yes	
Holiday FEs	Yes		Yes	
Month FEs	Yes		Yes	
Origin FEs	Yes		Yes	
Destination FEs	Yes		Yes	

Note: This table replicates results from Table 5 but uses reweighed smartphone trips data. *p<0.1; **p<0.05; ***p<0.01.

Table A.15: Summary of Results: Estimated Elasticities Using Reweighed Data

	Minibus Trips		No. Minibuses	
	NPIV	GCF	NPIV	GCF
	(1)	(2)	(3)	(4)
Overall Elasticity	0.345	0.581**	0.345	0.577**
	(0.562)	(0.238)	(0.592)	(0.239)
Within Terr. Elasticity	0.831	0.731***	0.830	0.605***
	(0.732)	(0.225)	(0.915)	(0.228)
Across Terr. Elasticity	0.633	0.599***	0.628	0.348
	(0.711)	(0.266)	(0.510)	(0.287)
Within Terr. − Across Terr.	0.199	0.132	0.203	0.257
	(0.845)	(0.416)	(1.052)	(0.431)
Day FEs	Yes	Yes	Yes	Yes
Route FEs	Yes		Yes	
Origin FEs		Yes		Yes
Destination FEs		Yes		Yes

Note: This table replicates results from Table 8 but uses reweighed smartphone trips data. *p<0.1; **p<0.05; ***p<0.01.

A.7 Theoretical Proofs

Proposition 1.

Fully differentiate the first order conditions wrt b_{out}:

$$\eta_w \equiv \frac{\mathrm{d}b_w}{\mathrm{d}b_{out}} = \frac{-c_0'' \cdot \left(c_a'' - (c_0' + c_w')\frac{\pi''(b_w^*)}{\pi(b_w^*)'}\right)}{\left(c_a'' + c_0'' - (c_0' + c_a')\frac{\pi''(b_a^*)}{\pi'(b_a^*)}\right)\left(c_w'' + c_0'' - (c_0' + c_w')\frac{\pi''(b_w^*)}{\pi'(b_w^*)}\right) - c_0''^2} \tag{25}$$

$$\eta_a \equiv \frac{\mathrm{d}b_a}{\mathrm{d}b_{out}} = \frac{-c_0'' \cdot \left(c_w'' - (c_0' + c_a')\frac{\pi''(b_a^*)}{\pi'(b_a^*)}\right)}{\left(c_a'' + c_0'' - (c_0' + c_a')\frac{\pi''(b_a^*)}{\pi'(b_a^*)}\right)\left(c_w'' + c_0'' - (c_0' + c_w')\frac{\pi''(b_w^*)}{\pi'(b_w^*)}\right) - c_0''^2} \tag{26}$$

$$\tag{27}$$

Using **Assumption 1.**, we get:

$$\eta_w \equiv \frac{\mathrm{d}b_w}{\mathrm{d}b_{out}} \approx \frac{-c_0''c_a''}{(c_0'' + c_a'')(c_0'' + c_w'') - c_0''^2} \tag{28}$$

$$\eta_a \equiv \frac{\mathrm{d}b_a}{\mathrm{d}b_{out}} \approx \frac{-c_0''c_w''}{(c_0'' + c_a'')(c_0'' + c_w'') - c_0''^2} \tag{29}$$

Then $\frac{\eta_w}{\eta_a} \approx \frac{c_a''}{c_w''}$

Proposition 2.

Fully differentiate first order conditions wrt λ_w and λ_a

$$\epsilon_w \equiv \frac{\mathrm{d}b_w}{\mathrm{d}\lambda_w} \cdot \frac{\lambda_w}{b_w} = \frac{(c_0' + c_w')\left[c_a'' + c_0'' - (c_0' + c_a')\frac{\pi''(b_a^*)}{\pi'(b_a^*)}\right]}{\left[\left(c_a'' + c_0'' - (c_0' + c_a')\frac{\pi''(b_a^*)}{\pi'(b_a^*)}\right)\left(c_w'' + c_0'' - (c_0' + c_w')\frac{\pi''(b_w^*)}{\pi'(b_w^*)}\right) - c_0''^2\right] b_w} \tag{30}$$

$$\epsilon_a \equiv \frac{\mathrm{d}b_a}{\mathrm{d}\lambda_a} \cdot \frac{\lambda_a}{b_a} = \frac{(c_0' + c_a')\left[c_w'' + c_0'' - (c_0' + c_w')\frac{\pi''(b_w^*)}{\pi'(b_w^*)}\right]}{\left[\left(c_a'' + c_0'' - (c_0' + c_a')\frac{\pi''(b_a^*)}{\pi'(b_a^*)}\right)\left(c_w'' + c_0'' - (c_0' + c_w')\frac{\pi''(b_w^*)}{\pi'(b_w^*)}\right) - c_0''^2\right] b_a} \tag{31}$$

Using **Assumption 1.**, we get:

$$\epsilon_w \approx \frac{(c_0' + c_w')(c_a'' + c_0'')}{\left[(c_0'' + c_w'')(c_0'' + c_a'') - c_0''^2\right] b_w}$$

$$\approx \frac{(C_w')(C_a'')}{\left[(C_w'')(C_a'') - c_0''^2\right] b_w}$$

$$\epsilon_a \approx \frac{(c_0' + c_a')(c_w'' + c_0'')}{\left[(c_0'' + c_w'')(c_0'' + c_a'') - c_0''^2\right] b_a}$$

$$\approx \frac{(C_a')(C_w'')}{\left[(C_w'')(C_a'') - c_0''^2\right] b_a}$$

Then $\dfrac{\epsilon_w}{\epsilon_a} \approx \dfrac{\dfrac{C_a''\, b_a}{C_a'}}{\dfrac{C_w''\, b_w}{C_w'}}$