

Subways or Minibuses? Privatized Provision of Public Transit*

Lucas Conwell [†]

First Version: October 2022

This Version: June 2023

[Please [click here](#) for updated version]

Abstract

Workers in developing countries waste significant time commuting, and gaps in public transit constrain access to productive jobs. In many cities, privately-operated minibuses provide 50–100% of urban transit, at the cost of long wait times and poor personal safety for riders. Should developing-country cities follow the typical recommendation of bus rapid transit or subway investments or rather optimize this existing, home-grown network? I build a micro-founded model of privatized shared transit subject to externalities in matching between buses and passengers. I then estimate the model with newly-collected data on minibus and passenger queues in Cape Town and stated user preferences for exogenously-varied commute attributes. I find that Cape Town's existing bus rapid transit decreased welfare, net of costs, but socially-optimal minibus fares and commuter taxes correct matching externalities, particularly benefit low-skill workers, and reduce spatial misallocation. Government actions to improve security bring even more substantial welfare gains.

*I am deeply indebted to my advisors Costas Arkolakis, Michael Peters, Mushfiq Mobarak, and Orazio Attanasio for their invaluable guidance and generosity with their time and ideas. I additionally thank Lorenzo Caliendo, Richard Carson, Will Damron, Fabian Eckert, Cecilia Fieler, Claudia Gentile, Antonia Paredes-Haz, Sam Kortum, Ryungha Oh, Mark Rosenzweig, Nicholas Ryan, Matthew Schwartzman, Sam Slocum, Michael Sullivan, Kaushik Vasudevan, Trevor Williams as well as a host of seminar participants and visitors at Yale for their helpful comments and suggestions. I thank Philip Krause, Aslove Mateyisi, and Mokgadi Mehlape from GoAscendal/GoMetro for their herculean efforts to obtain permission for and organize the data collection. Finally, I acknowledge generous financial support from the Yale Economic Growth Center and Ryoichi Sasakawa Young Leaders Fellowship Fund that made this project possible. IRB Exemption Determination obtained through Yale University, ID 2000032302.

[†]University College London. lucas.conwell@yale.edu

I. INTRODUCTION

Workers in low- and middle-income countries waste significant time commuting (OECD (2016)), and barriers such as personal safety further constrain their access to the most productive jobs (Uteng (2011) and Moreno-Monroy (2016)). International organizations typically focus on one policy solution, namely expansion of government-provided “formal” subway or bus rapid transit lines beyond established high-income neighborhoods.¹ The economics literature, too, neglects an alternative: the privately-operated minibuses which provide 50 to 100% of urban transit in many developing-world cities (Tun and Hidalgo (2022)). These chaotic networks offer the urban poor broad connectivity at the cost of substantial wait times – up to one-third of the typical commute in my context of Cape Town – and poor personal safety. In this paper, I ask whether developing-country cities should extend bus rapid transit or train lines to the lower-income masses or, instead, optimize existing *privatized shared transit* systems. I then consider how this optimization can help connect high-amenity neighborhoods, ameliorate spatial mismatch between workers and jobs, and reduce carbon emissions.

Answers to these questions require a model of privatized shared transit. Mine dispenses with the standard fixed commute costs and instead highlights externalities in matching between minibuses and passengers as well as the social planner’s optimum. I collected new data in Cape Town on both sides of this matching market with which I directly estimate the matching function. Simultaneously, I extend recent advances in model-tailored stated preference surveys to a particularly plausible context and thus quantify user preferences for exogenously-varied commute attributes. The estimated model reveals that the high fixed costs of Cape Town’s existing MyCiti bus rapid transit line exceed its benefits. In contrast, optimization of the minibus system increases low-skill commuter welfare by almost one-half percent as wait times, spatial misallocation of workers, and carbon emissions all fall. However, the welfare gains from government-provided security guards at minibus stations exceed even those from optimization of the existing system.

I start from first principles to build a model of privatized shared transit with two key features. First, minibuses enter distinct origin-to-destination routes and match with passengers, subject to an increasing cost of entry which reflects minibus collectives’ potentially violent attempts to restrict entry. Second, commuters with heterogeneous incomes choose a mode of transport as well as home and work locations based on factors such as com-

¹Bus rapid transit refers to high-capacity vehicles with their own stations and lanes. For examples of international organization recommendations, see articles and guides by the [World Bank](#) and [Institute for Transportation & Development Policy](#).

mute times and safety. At the heart of the model, a frictional matching market between minibuses and passengers determines the wait times of each. Passengers first wait in long lines to board buses and subsequently wait on these buses, which depart only when full. Crucially, the number of buses affects these two wait time components in opposite ways. “Off-bus” wait times fall, and “on-bus” wait times rise, with minibus entry due to opposing thick-market and congestion externalities in matching. On higher-demand routes, status-quo minibus fares fail to compensate the increasingly large thick-market, or *boarding*, externality of minibus entry, relative to the congestion, or *filling*, externality. Thus, under-provision of minibuses ensues.

To quantify the model, I collected two forms of primary data in Cape Town. First, enumerators tracked passenger and bus queues on a random sample of 44 minibus routes, from which I measure bus loading rates and commuters’ wait times. Second, I employ under-utilized stated preference methods to generate exogenous variation in commute choices. Commute choice, given its tangibility and familiarity, provides a particularly suitable application. In my survey, 526 respondents chose hypothetical minibus commute options with exogenously-varied travel times, costs, and quality improvements. Preferences across modes, not only minibuses but also car and “formal” public transit, come from a separate stated preference survey conducted by the City of Cape Town.

With the help of these two datasets, I estimate the minibus matching function and the commuter demand system. Because I observe both sides of the matching market in the queue data, I can estimate the matching elasticities without initially imposing the constant returns to scale assumption common in the literature. The primary threat to identification comes from time-varying shocks correlated with routes’ matching efficiencies, such as weather. Thus, I instrument for year-2022 relative demand with year-2013 commute start times; the exclusion restriction holds provided hourly matching efficiency trends do not strongly persist over these nine years. From commuters’ stated preferences, I estimate a discrete choice model that yields commuters’ mode-specific utility costs as well as values of time and quality improvements, such as security, in dollar equivalents.

Finally, I employ the estimated model to analyze how policymakers can best improve the provision of urban transit. First, Cape Town’s internationally-admired MyCiti bus rapid transit line actually reduced welfare. In a city of dispersed commute flows, the benefits concentrate in a single neighborhood and thus fail to outweigh the high fixed costs. Second, the social planner’s optimal minibus network and commute choices increase low-skill welfare by almost one-half percent. The social planner corrects the under-provision of minibuses on high-wage and amenity routes through higher fares, and off-bus wait times

fall substantially. In turn, emissions as well as spatial misallocation decline as commuters reallocate towards higher-wage locations. Third, low-skill commuters gain substantially more from government-provided minibus station security guards than from optimization of the existing system. Thus, improved privatized transit could help solve the transport problems of a host of rapidly growing, resource-poor African cities similar to Cape Town.

Related Literature

Existing work studies the effects of transport infrastructure on the wider economy through careful causal empirics (Baum-Snow et al. (2005), Baum-Snow (2007), Glaeser et al. (2008), Baum-Snow et al. (2017), and Gonzalez-Navarro and Turner (2018)) and quantitative spatial models (Allen and Arkolakis (2014), Ahlfeldt et al. (2015), Donaldson and Hornbeck (2016), Donaldson (2018), Monte et al. (2018), Heblich et al. (2020), and Nagy (2023)). The most closely-related papers quantify how travel time reductions from bus rapid transit or subway lines translate into job matching (Tsivanidis (2022) and Severen (2023)), informality (Zarate (2023)), or gentrification (Balboni et al. (2020) and Warnes (2021)). I contribute the first model of the privatized transit sector and then study policies whose impact depends crucially on these operators' decisions.

Another strand of the literature explicitly considers feedback between transport costs and agents' decisions via road congestion (Duranton and Turner (2011), Kreindler (2022), Barwick et al. (2022), and Akbar et al. (2023)), network effects, and environmental externalities (Almagro et al. (2023)). Recent advances then characterize optimal road (Fajgelbaum and Schaal (2020) and Allen and Arkolakis (2022)), public transit (Kreindler et al. (2023)) and ride-sharing (Almagro et al. (2023)), commuting (Fajgelbaum et al. (2021)), or ocean shipping networks (Brancaccio et al. (2023)). My supply-side model instead highlights externalities key to privatized transit and extends the nascent study of socially-optimal transport networks to developing countries' most common commute "technology."

Methodologically, my paper draws on matching models of transport markets such as ride-hailing (Castillo (2022), Castillo et al. (2022), and Rosaia (2023)), taxis (Fréchette et al. (2019) and Buchholz (2022)), and ocean shipping (Brancaccio et al. (2020)). In particular, I start with the latter's one-to-one matching framework and then develop a model of minibus operations and commute mode choice. Unusually, my new data on both sides of the matching market allows me to estimate a many-to-one matching function with non-constant returns to scale.

Finally, stated preference surveys, originally employed in economics to study consumer

credit (Juster and Shay (1964) and Juster (1964)), have increasingly provided exogenous variation to estimate structural models (Caplin (2021), Bernheim et al. (2021), and Almås et al. (2023)) of long-term care (Ameriks et al. (2020)) or marriage decisions (Andrew and Adams-Prassl (2023)). I extend the stated preference methodology to a particularly concrete, familiar context where respondents can likely accurately predict their behavior.

I structure the remainder of the paper as follows. In Section II, I describe my data, both newly collected and from existing sources, and the context. In Section III, I discuss a series of facts to rationalize my modeling choices and counterfactuals. I then lay out my theory in Section IV, followed by the estimation procedure in Section V. After validating my model’s fit in Section VI, I discuss the welfare gains from alternative transport policy interventions in Section VII. Finally, I conclude in Section VIII.

II. MINIBUS DATA COLLECTION AND CONTEXT











In this section, I discuss the collection of my primary data, with additional details regarding all datasets in Online Appendix A. I then provide an overview of the minibus market in Cape Town.

Newly-Collected and Household Survey Data

Little micro-data on privatized transit exists, so I collaborated with the mobility advisory firm GoAscendal to design a custom two-part data collection effort in Cape Town. First, minibus “station counts” tracked the matching-like process by which minibuses on a given route load at designated lanes in stations. For each route, enumerators recorded bus arrival and departure times, the number of passengers on board each bus, and the length of the queue to board at five-minute intervals. I later employ these records to calculate passenger and bus wait times along with the *loading rate* at which buses fill up. The counts covered a two-stage cluster sample of $N = 44$ minibus routes in Cape Town, where the eight clusters sampled in the first stage corresponded to the routes that originate from a given station. I stratified the station and the second-stage route samples by a proxy for the number of entering buses to reduce the number of zero bus observations.

Second, I designed a stated preference survey to estimate commuters’ monetary values of quality improvements and time as well as cost sensitivity. I asked respondents to consider a hypothetical work commute trip and then choose a preferred option in a series of *choice sets* composed of two minibus alternatives, as in Figure 1. Each option varied exogenously

FIGURE 1. STATED PREFERENCE SURVEY CHOICE SET BETWEEN MINIBUS OPTIONS

Q1.1	Option 1.1.1	Option 1.1.2
Cost	R18.00 	R6.00 
Travel Time	50 Minutes 	50 Minutes 
Security	Security at taxi rank 	No security at taxi rank 
Driver Behaviour	Adheres to speed limit 	Exceeds speed limit 
Bus Loading	Enough seats for all passengers 	Overloaded: more passengers than seats 

Notes: This figure shows an example of a choice set from my stated preference survey, consisting of two hypothetical minibus commute alternatives, from which respondents indicated their preferred option. The rows list the attributes associated with each option, which vary exogenously across choice sets and respondents. Note that “taxi rank” is the South African term for a minibus station.

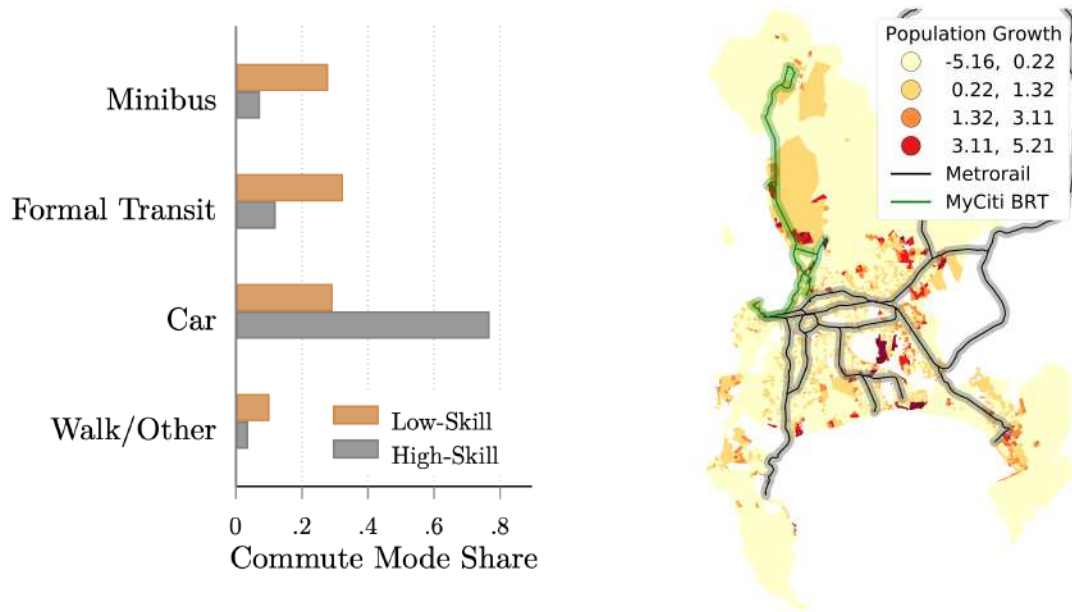
in cost and travel time as well as in the presence or absence of three quality improvements, chosen based on commuter concerns in past city surveys. The latter included security guards at the publicly-owned, shared minibus stations, driver adherence to speed limits, and whether the minibus loads more passengers than seats. Respondents completed one of two randomized “blocks” of five choice sets; I chose the levels of the attributes with a *d-efficiency* algorithm that maximizes parameter precision (Rose and Bliemer (2009)). After a pilot survey, I reduced the number of choice sets and alternatives per set to maintain respondent attention.

To conduct interviews, enumerators randomly approached respondents at one intermodal transport hub and two minibus stations on weekdays in June 2022. Resource constraints precluded interviews at a greater variety of locations that would have avoided the over-representation of minibus commuters in my sample (see Online Appendix A.2.6). Therefore, I employ this data only in the estimation of relative preferences for different *minibus* attributes and later attempt to quantify disparities in preferences between my sample and the population.

Finally, I make use of representative household survey data collected by the City of Cape Town. A city-run stated preference survey omitted quality improvements but varied the mode: car, various types of formal public transit, or minibus. Additionally, for a larger

FIGURE 2. DIFFERENT MODES OF TRANSPORTATION IN CAPE TOWN

(A) COMMUTE MODE SHARES BY SKILL GROUP (B) FORMAL PUBLIC TRANSIT NETWORK



Notes: Panel (A) displays the shares of low- (non-college) and high-skill commuters in Cape Town who use each mode, from the 2013 Cape Town Household Travel Survey. I exclude residents who work in their home transport analysis zone, and “minibus” includes any who use minibuses during a typical commute. Panel (B) displays the networks of formal transit modes with dedicated infrastructure in Cape Town, namely MyCiti bus rapid transit and Metrorail commuter trains. Shading indicates population growth from 1996–2011 at the small area layer level.

sample of residents, the same survey provides home and workplace locations as well as incomes.

Minibuses in Cape Town

Privatized shared transit in South Africa takes the form of ubiquitous 15-passenger minibuses. Figure 2a displays the shares of commuters in Cape Town on each mode of transport. Throughout the paper, I distinguish between *low*-skilled workers, or those with less than a college education, and the *high*-skilled. A full 28% of low-skill workers and 7% of high-skill workers commute via minibus; the overwhelming majority of high-skill commuters instead drive to work. Around one-third of low-skill commuters use limited publicly-provided “formal” transit alternatives. These include infrequent Golden Arrow buses that run in mixed traffic as well as higher-speed MyCiti bus rapid transit and Metrorail commuter rail lines. However, the latter two networks, overlaid on recent population growth in Figure 2b, miss many fast-growing suburbs.

The minibus market consists of a large number of private firms who own, on average, less than two buses each (Woolf and Joubert (2013)). Firms pay an entry fee to an owner *association* to operate on a specific origin-destination pair, or *route* (City of Cape Town (2014) and Kerr (2018)). City government grants these collectives exclusive rights to one or more routes, which they violently defend (Kerr (2018)). Associations could in theory restrict the quantity of buses on their routes, but because their “main income derives from owners’ membership fees. . . it is in [their] interest to have as many members as possible” (Schalekamp (2017) p. 59). Associations also set fares (Kerr (2018)), subject to government approval of the “cost to the user (portion of monthly income spent on public transport)” (City of Cape Town (2014) p.65).²

III. FACTS ABOUT THE MINIBUS MARKET

I now present six facts about the minibus sector in Cape Town. The first three facts each motivate a specific modeling choice. Routes begin at a so-called *taxi rank*: a large covered minibus station with clearly-marked loading lanes for each route. Henceforth, I deviate from local terminology and call these facilities *minibus stations*.

Fact 1. *Multiple minibuses load simultaneously for each route, and large numbers of passengers wait to board, at minibus stations.*

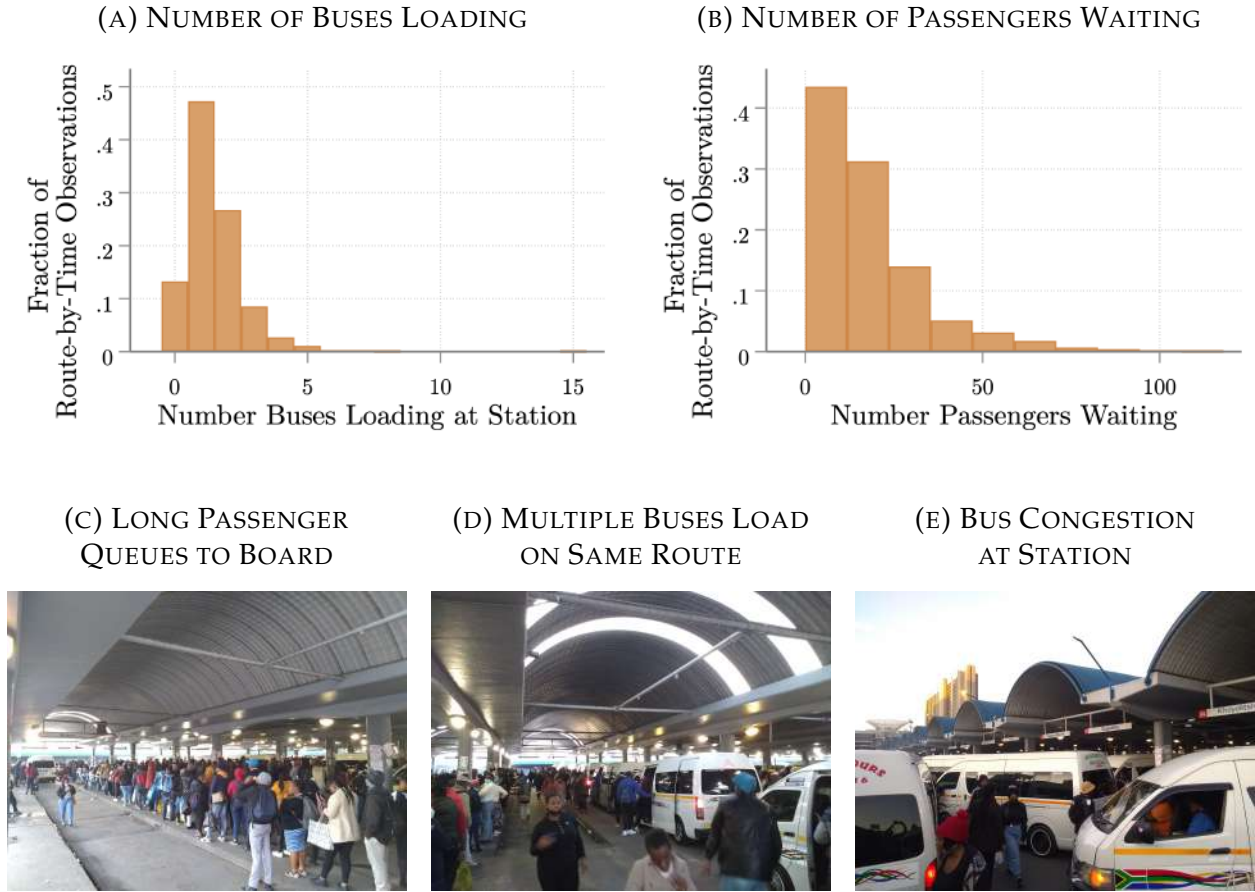
The histogram in Figure 3a displays the number of buses on the *same* route loading simultaneously in the station, across five-minute time blocks and routes in my station count data. Multiple buses load simultaneously in about half of the route-by-time observations, as pictured in Figure 3d. Furthermore, the histogram of the corresponding number of waiting passengers in Figure 3b demonstrates that passengers wait in queues that can exceed 50 people, echoing the visual in Figure 3c. In practice, passengers board not in queue order but in a disorderly scramble. To replicate the large numbers of buses and passengers who simultaneously wait and the lack of an orderly queue, I model the minibus loading process as random matching between buses and passengers.

Fact 2. *15-passenger vans account for 94% of licensed minibuses, and 96% of minibuses depart with a full load of at least 15 passengers.*

Though the law allows for several discrete bus sizes, the 15-passenger-plus-driver variant

²Virtually all firms join associations (Antrobus and Kerr (2019)). Additionally, the law requires individual firms to obtain an effectively pro-forma government permit and operate a vehicle with one of several approved seat capacities (Jobanputra (2018) p. 290). However, up to half of firms lack these permits (City of Cape Town (2014) p. 77).

FIGURE 3. MINIBUS ROUTE LOADING PROCESS AT ORIGIN STATION



Notes: Panel (A) displays the distribution of the number of minibuses loading at the origin station on a specific minibus route, over minibus routes and five-minute periods in my station count data from Cape Town. Panel (B) displays the distribution of the number of passengers waiting at the origin station for a specific route, also over routes and five-minute periods. Panels (C)-(D) display images of the minibus loading process for a single origin-destination route at its designated loading lane within the origin station, here the Cape Town CBD station. Panel (E) displays the exits from these loading lanes. Images by author, June 2022.

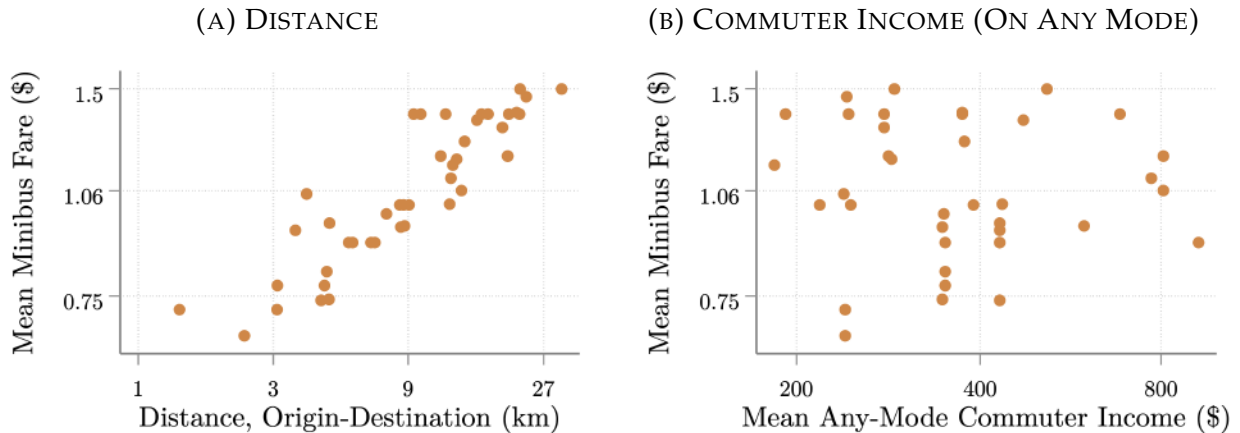
dominates in practice (Jobanputra (2018) p. 290), so I impose a single exogenous bus size in my model. Because buses in my station count data virtually never leave less than full, I further require buses to reach this exogenous capacity before they depart.³

Fact 3. *Minibus fares increase strongly with route distance but not with commuters' ability to pay.*

Figure 4a plots distance on the horizontal axis and the mean fare charged on a minibus

³The passenger experience once underway more closely resembles that of scheduled transit. Most routes follow a “line-haul” mode of operation during peak hours: they travel on highways to their destination and do not expect to pick up additional passengers.

FIGURE 4. MINIBUS FARES VERSUS DISTANCE AND COMMUTER INCOME

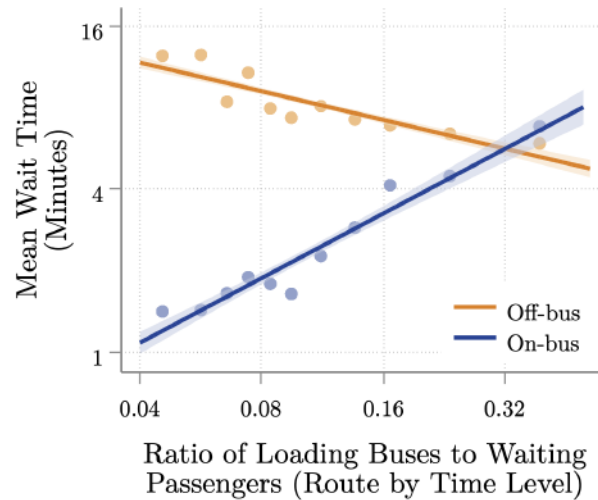


Notes: Scatterplots display, all on log scales, the mean fare on a minibus route versus straight-line distance from the route’s origin to its destination in Panel (A) and versus the mean income of commuters (using any mode) from the transport analysis zone (TAZ) where the route originates to the TAZ of its destination in Panel (B). Fares come from my on-board tracking data, and incomes come from the 2013 Cape Town Household Travel Survey.

route in my sample on the vertical axis; fares increase almost log-linearly with distance. Figure 4b instead displays the mean income of commuters, using any mode, between the route’s origin and destination neighborhoods on the horizontal axis. Surprisingly, the fares on the vertical axis bear no apparent relationship to income. Negotiations between the minibus associations and city government during the route approval process thus produce a primarily distance-based fare scheme. In consequence, in my baseline model, I calibrate fares as a function of route distance; my key social planner-optimal fare counterfactual does not require information on the determinants of status-quo fares.

The latter two facts rationalize my focus on specific counterfactual transport policies. I begin with the problem of passenger wait time, which takes two forms. First, in my route sample, the average minibus passenger waits 8.5 minutes *off-bus* to board; indeed, local journalists complain that “queues, especially during certain times of the day, are impossibl[y]” long (Theway (2018)). Second, the World Bank laments the “inefficien[cy]” of the behavior cited in Fact 2, namely “that minibus taxis generally only leave when they are full” (Kerr (2018)). Traffic jams at the exits of bus loading areas, as pictured in Figure 3e, further delay departure. As a result, the average passenger waits an additional 2.7 minutes *on-bus* for departure, out of an average 36-minute minibus commute. Both forms of wait time, in turn, depend on bus entry.

FIGURE 5. OFF- AND ON-BUS PASSENGER WAIT TIMES VERSUS MINIBUS ENTRY



Notes: Binned scatterplots and best fit lines display the log-scale relationship between the relative number of loading buses to passengers waiting at the station for a given minibus route, across routes and five-min. periods in the station count data, and mean off-bus (orange) and on-bus (blue) passenger wait times.

Fact 4. *Minibus passengers' off-bus wait time falls and on-bus wait time rises, the greater the number of loading minibuses relative to the number of waiting passengers on a route.*

On the horizontal axis, Figure 5 plots the number of loading buses relative to waiting passengers for a given route across routes and five-minute time periods in my station count data. The vertical axis displays minibus passengers' mean off- and on-bus wait times. Note first that higher numbers of loading buses go hand-in-hand with lower off-bus wait times since passengers more easily find seats on buses. Bus entry, however, increases passengers' on-bus wait times between the moment they board and the bus's departure out of the station. This congestion effect stems from greater passenger confusion, which slows the loading process, and more acute in-station traffic jams, as pictured in Figure 3e. As a result, bus entry has an ambiguous effect on *total* off- plus on-bus wait time. My second policy counterfactual, then, involves minibus fares that optimize this tradeoff and implement the social planner solution.

Fact 5. *Commuters in a past satisfaction survey rated security second-most negatively among all minibus attributes.*

In a 2013 survey, security led the list of minibus-related grievances that included road safety, crowdedness, cleanliness, timetable adherence, ease of use, and distance to the stop, as displayed in Appendix Figure A.1. The fundamental public-good nature of security in shared public spaces such as minibus stations suggests state intervention. Thus, as a

second policy counterfactual, I consider government-financed minibuses station security guards.

IV. A THEORY OF PRIVATIZED SHARED TRANSIT

In this section, I build a model of the privatized shared transit sector. My model has two key features: first, minibuses enter and match with passengers, and second, commuters with heterogeneous incomes optimally choose their home location, work location, and mode of transport. I lay out the environment, discuss the problems of each type of agent, define equilibrium, and then derive its efficiency properties.

Environment

I consider a city made up of a finite number of locations indexed by $i, j \in \{1, \dots, I\} \equiv \mathcal{L}$ and two types of agents: minibuses and commuters. Time in my dynamic model is continuous, and commuters discount the future at rate r . Minibuses freely enter each origin-destination pair, or what I term a route, ij and complete multiple trips, subject to matching frictions. Fixed masses $\{N^g\}$ of commuters of skill $g \in \{\text{low } (l), \text{high } (h)\}$ are born per unit time. Commuters choose a home location, indexed by i and with fixed amenity θ_i^g , a work location, indexed by j , and one mode $m \in \{\text{minibus } (M), \text{formal transit } (F), \text{car } (A)\}$ for a single commute to work. They collect a present-value wage ω_j^g upon arrival. I leave the parameters related to formal transit and cars as exogenous and solve the model exclusively in steady-state.

Minibuses

Minibuses enter each origin-destination route ij at a cost $F_{ij} \equiv \bar{\psi} b_{ij}^\phi$. The cost “intercept” $\bar{\psi}$ accounts for entry fees paid to minibus associations as well as the monetary costs of bus purchase. The elasticity ϕ summarizes the extent to which entry costs rise with the mass of loading buses, b_{ij} , on a route. This setup captures associations’ attempts to restrict entry: they might charge higher entry fees on increasingly lucrative routes or directly restrict quantities through threats of violence.

After entry, minibuses complete multiple trips on the same route, as follows. First, minibuses load passengers at the minibus station in a frictional matching process. Importantly, I assume that buses depart only upon reaching an exogenous passenger capacity $\bar{\eta}$. Second, they collect exogenous fares τ_{ijM} from passengers, which reflect the government-

mediated fare scheme in the status quo. Third, the minibus travels toward its destination, subject to a per-kilometer operating cost χ . Fourth, the minibus arrives at its destination j at a fixed Poisson rate d_{ij} and exits the market with a probability $x \equiv g(T_{ij})$ that increases with the expected total trip time T_{ij} , i.e., $g'(\cdot) > 0$. This structure incorporates the tradeoff between the number of trips and trip time inherent to a finite work shift and simultaneously preserves the stationarity of the model. Buses that do not exit costlessly return to the route origin i and begin another trip.

Matching

Each trip starts at a minibus station, where minibuses and passengers meet in a matching market segmented by route. The idea of incorporating matching frictions into a model of transportation comes from Brancaccio et al. (2020), who consider ocean shipping. I extend their one-to-one to a many-to-one matching framework approximating the minibus loading process, where large crowds jostle to board one of several buses. Given a mass p_{ij} of waiting passengers and a mass b_{ij} of loading buses, \mathcal{M}_{ij} matches between a minibus and a single passenger form per unit time, where

$$\mathcal{M}_{ij} \equiv \mu_{ij} p_{ij}^{\alpha} b_{ij}^{\beta}. \quad (1)$$

The exogenous matching efficiency μ_{ij} reflects station infrastructure and factors such as weather. In turn, the passenger and bus matching elasticities α and β capture the degree to which passengers rival each other and the extent of congestion among loading minibuses, respectively. The relative magnitudes of these two elasticities then determine how strongly additional minibus entry decreases passengers' off-bus wait times and increases their on-bus wait times.

As in any matching model, each additional minibus that enters a route imposes two externalities on passengers and fellow buses. On the one hand, through the usual thick-market, or what I term *boarding*, externality, bus entry increases the boarding rates at which passengers meet buses,

$$\lambda_{ij} \equiv \frac{\mathcal{M}_{ij}}{p_{ij}} = \mu_{ij} b_{ij}^{\beta} p_{ij}^{\alpha-1}. \quad (2)$$

This positive externality is more significant, the less potent bus congestion relative to passenger congestion, i.e., the higher the bus matching elasticity β compared to the passenger elasticity α . On the other hand, through a negative congestion, or *filling*, externality, bus

entry decreases the loading rate ι_{ij} at which buses meet passengers, given by

$$\iota_{ij} \equiv \frac{\mathcal{M}_{ij}}{b_{ij}} = \mu_{ij} p_{ij}^\alpha b_{ij}^{\beta-1}. \quad (3)$$

Lower levels of bus congestion go hand-in-hand with higher relative values of the bus matching elasticity β and shrink the magnitude of the filling externality. Thus, though the two externalities work in opposite directions, they do not necessarily cancel one another, even under constant returns to scale in matching.

As I now highlight, these two externalities affect passenger wait times in opposite ways. I rewrite the expected total passenger wait time on a given route as

$$\text{total wait}_{ij} \equiv \underbrace{\frac{1}{\lambda_{ij}}}_{\text{off-bus}} + \underbrace{\frac{1}{2} \frac{\bar{\eta}}{\iota_{ij}}}_{\text{on-bus}} = \underbrace{\left(\frac{b_{ij}}{p_{ij}}\right)^{-\beta}}_{\text{boarding externality}} + \underbrace{\frac{\bar{\eta}}{2} \left(\frac{b_{ij}}{p_{ij}}\right)^{1-\beta}}_{\text{filling externality}}, \quad (4)$$

where the second equality relies on $\mu_{ij} = 1$ and constant returns in matching. The first term in (4), or expected off-bus wait time, equals the inverse of the boarding rate. The second term, the expected on-bus wait time, equals half the time $\bar{\eta}/\iota_{ij}$ required for the bus to fill because passengers, in expectation, board half-full buses. Now, consider an increase in the relative mass of loading buses, b_{ij}/p_{ij} . The higher this ratio, the lower passengers' off-bus wait time, precisely the boarding externality visible in the data in Figure 5. Simultaneously, the lower loading rates associated with additional entry increase the expected on-bus wait time. The former positive externality grows, and the latter negative filling externality falls in magnitude, the higher the bus matching elasticity β relative to the passenger elasticity α . Thus, bus entry has an ambiguous effect on total passenger wait time. Potentially sizable heterogeneity in route profitability nevertheless precludes a direct mapping from elasticities to under- or over-provision, as I now describe.

Profits and Free Entry

Lower bus loading rates not only inconvenience passengers but also reduce minibus profits by increasing total trip time. More precisely, the expected total time T_{ij} required to complete one trip on a route ij equals the sum of the time a bus waits to load and the expected travel time from i to j ,

$$T_{ij} \equiv \frac{\bar{\eta}}{\iota_{ij}} + \frac{1}{d_{ij}}. \quad (5)$$

Loading time equals bus capacity $\bar{\eta}$ divided by the loading rate ι_{ij} , while average travel time equals the reciprocal of the Poisson destination arrival rate d_{ij} . The number of trips completed post-entry then follows a geometric distribution whose expectation, $1/g(\cdot)$, decreases in the exit rate function $g(\cdot)$ and thus in average trip time, as in a model with a fixed work shift length.⁴ Net revenue per trip, in turn, equals fare revenue $\bar{\eta}\tau_{ijM}$ minus the product of operating cost χ and the exogenous distance Δ_{ij} driven. Finally, expected minibus profits equal the product of the expected number of trips and net revenue per trip. Free entry then ensures a bus loading rate, an expected trip time T_{ij} , and thus an expected number of trips, such that expected profit equals entry costs:

$$\underbrace{\frac{1}{g(T_{ij})}}_{\text{expected number trips}} \underbrace{(\bar{\eta}\tau_{ijM} - \chi\Delta_{ij})}_{\text{net revenue per trip}} = \underbrace{\bar{\psi}b_{ij}^{\phi}}_{\text{entry cost}}. \quad (6)$$

Importantly, free entry generates considerable heterogeneity in passenger wait times across routes. In particular, a route's profitability decreases with trip time T_{ij} and, through loading rates ι_{ij} , with bus entry. Amid congestion in entry costs, i.e. $\phi > 0$, bus entry must thus rise less than proportionately with demand to increase loading rates and compensate higher entry costs. Any policy that alleviates the resulting long off-bus wait times on high-demand routes might then significantly increase commuter welfare.

Commuters

Commuters of skill g choose the combination of home location i , work location j , and mode m , either minibus, formal transit, or car, that offers the highest utility, $\theta_i^g + U_{ijm}^g + \omega_j^g + v\varepsilon_{ijm}$. Total utility comprises (i) the home location's exogenous amenity value θ_i^g , as in standard spatial models; (ii) the deterministic commute value U_{ijm}^g ; (iii) the work location's fixed wage ω_j^g ; and (iv) a Gumbel-distributed idiosyncratic preference ε_{ijm} with variance scaled by the parameter v .⁵

The commute value U_{ijm}^g from home i to work j on mode m depends on the utility, time, and monetary costs of each mode and linearly approximates a micro-founded commute

⁴Specifically, the number of trips follows the geometric distribution of the number of Bernoulli trials until, but including, the first success with probability of success given by $x \equiv g(\cdot)$.

⁵I do not allow workers to combine modes; in my stated preference survey, only 4.6% of minibus riders report also using formal buses or trains over the course of their commutes.

model, detailed in Online Appendices B.1-B.3. For the minibus mode,

$$U_{ijM}^g = -\underset{\substack{\uparrow \\ \text{utility cost}}}{\kappa_M^g} - r\omega_j^g \left(\underset{\substack{\uparrow \\ \text{off-bus wait}}}{\frac{1}{\lambda_{ij}}} + \underset{\substack{\uparrow \\ \text{on-bus wait}}}{\frac{1}{2} \frac{\bar{\eta}}{l_{ij}}} + \underset{\substack{\uparrow \\ \text{travel time}}}{\frac{1}{d_{ij}}} \right) - \underset{\substack{\uparrow \\ \text{fare}}}{\tau_{ijM}}. \quad (7)$$

The skill-specific utility cost κ_m^g represents commuters' non-pecuniary taste for each mode; improved station security and other quality improvements affect utility through κ_M^g . In turn, the *product* of the time preference rate r and the wage ω_j^g determines commuters' value of reductions in the total commute time in parentheses. The latter equals the sum of expected off- and on-bus wait as well as travel time. Finally, commuters consider the fare τ_{ijM} charged; their fare sensitivity varies inversely with the Gumbel shape parameter ν . In my stated preference survey, I exogenously vary quality improvements to κ_M^g , travel times, and fares. I then compare the quality improvement and time effects to commuters' fare sensitivity to estimate commuters' utility costs and value of time in dollar terms.

Commuters on other modes make similar tradeoffs. Formal transit commuters meet a vehicle and depart instantly, pay a utility cost κ_F^g as well as exogenous fares τ_{ijF} , and arrive at their destinations at an exogenous rate d_{ijF} . Car commuters similarly pay a utility cost κ_A^g as well as a fixed monetary cost τ_A and reach their destinations at the same Poisson rate d_{ij} as minibuses. The formal transit and car commute values mirror those for minibus commuters.⁶

Aggregate commuter demand then adheres to three choice-probability equations of the familiar Gumbel form. First, skill- g commuters choose to commute by minibus from home i to work j with a probability π_{ijM}^g that satisfies

$$\pi_{ijM}^g = \exp \left[\frac{\bar{W}^g}{\nu} \right]^{-1} \exp \left\{ \theta_i^g - \kappa_M^g - r\omega_j^g \left(\frac{1}{\lambda_{ij}} + \frac{1}{2} \frac{\bar{\eta}}{l_{ij}} + \frac{1}{d_{ij}} \right) - \tau_{ijM} + \omega_j^g \right\}^{1/\nu}. \quad (8)$$

Note that the choice probability of a given home by work location by mode tuple increases in the home location amenity θ_i^g and work location wage ω_j^g in addition to the aforementioned commute utility. The denominator is a function of commuters' ex-ante expected Gumbel utility over home and work locations as well as modes, $\bar{W}^g \equiv$

⁶Specifically, $U_{ijm}^g \equiv -\kappa_m^g - r\omega_j^g \left(\frac{1}{d_{ijm}} \right) - \tau_{ijm}$ for $m = F, A$, where $d_{ijA} \equiv d_{ij}$ and $\tau_{ijA} \equiv \tau_A$.

$v \log \left[\sum_{i,j,m} \exp \left[\theta_i^g + U_{ijm}^g + \omega_j^g \right]^{1/v} \right]$. Second and third, the home i , work j , formal transit or car choice probability π_{ijm}^g instead increases with the destination arrival rate d_{ijm} and decreases with the utility cost κ_m^g as well as monetary cost τ_{ijm} ,

$$\pi_{ijm}^g = \exp \left[\frac{\bar{W}^g}{v} \right]^{-1} \exp \left\{ \theta_i^g - \kappa_m^g - r\omega_j^g \left(\frac{1}{d_{ijm}} \right) - \tau_{ijm} + \omega_j^g \right\}^{1/v} \quad \text{for } m = F, A, \quad (9)$$

where $d_{ijA} \equiv d_{ij}$ and $\tau_{ijA} \equiv \tau_A$.

Equilibrium

First, I summarize the equilibrium conditions that determine demand and supply. The former adheres to the commuter choice-probability equations (8)-(9). The free entry condition (6) characterizes supply.

Second, I characterize the matching rates on each minibus route. In steady-state, the inflow to the stock, p_{ij}^g , of minibus passengers of skill g waiting off-bus equals the outflow, or $N^g \pi_{ijM}^g = \lambda_{ij} p_{ij}^g$. The boarding rate, however, depends on the sum of (off-bus) waiting passengers across skill groups, $p_{ij} \equiv \sum_g p_{ij}^g$. These insights, combined with (2)-(3), yield equations for the passenger boarding rate,

$$\lambda_{ij} = \mu_{ij}^{1/\alpha} \iota_{ij}^{-\beta/\alpha} \left[\sum_g N^g \pi_{ijM}^g \right]^{(\alpha+\beta-1)/\alpha}, \quad (10)$$

and bus loading rate,

$$\iota_{ij} = \frac{\sum_g N^g \pi_{ijM}^g}{b_{ij}}. \quad (11)$$

Definition. (Equilibrium) Given parameters $\{r, v, \kappa, \alpha, \beta, \mu, \tau_A, \bar{\psi}, \phi, \bar{\eta}, g(\cdot), \chi, N\}$ and model geography $\{\theta, \omega, d, d_F, \tau_M, \tau_F, \Delta\}$, an equilibrium is a vector $\{\mathbf{b}, \boldsymbol{\pi}, \boldsymbol{\lambda}, \boldsymbol{\iota}\}$ such that (i) free entry of minibuses (6) holds; (ii) commuter demand is consistent with (8)-(9); and (iii) passenger boarding rates in (10) as well as (iv) bus loading rates in (11) are consistent with matching technology.

Efficiency

To permit a characterization of efficiency, I first define welfare in my model.

Definition. (Welfare) Welfare, Ω , equals the aggregate ex-ante expected utility of newly-born commuters, net of aggregate rebates Π received upon birth and external emissions costs E ,

$$\Omega \equiv \sum_g N^g \bar{W}^g + \Pi - E. \quad (12)$$

The ex-ante expected utility \bar{W}^g includes the deterministic utility of each commute as well as the expectation of the idiosyncratic shocks, under the assumption that commuters make optimal choices of home, work, and mode. Commuters receive rebates, Π in the aggregate, of minibus profits and entry costs.⁷ Finally, E accounts for external emissions costs: each commuter emits, at the start of his or her trip, an amount of CO2-equivalent emissions proportional to the driving distance Δ_{ij} from origin to destination. Each mode differs in its per-kilometer and commuter emissions χ_m^e , and emissions decrease welfare by some social cost of carbon ζ , so that $E \equiv \zeta \sum_{i,j,m,g} N^g \pi_{ijm}^g \chi_m^e \Delta_{ij}$. Importantly, as I show in Appendix B.1, welfare does not depend on minibus fares.

Next, I define a social planning problem where the planner chooses allocations subject to matching frictions.

Definition. (Planning Problem) The social planner chooses searching buses \mathbf{b} for each route as well as skill-group-specific commute choice probabilities $\boldsymbol{\pi}$ for each home, work, and mode to maximize welfare, Ω , as in (12), subject to the matching technology (10)-(11) as well as the constraint $\sum_{i,j,m} \pi_{ijm}^g = 1$ for each skill group g .

I later confirm that the social planner can indeed implement these choice probabilities through budget-neutral taxes on commuters. However, due to the boarding and filling as well as environmental externalities,

Proposition 1. The competitive equilibrium minibus entry \mathbf{b} and commuter choices $\boldsymbol{\pi}$ are generically inefficient.

Proof. See Appendix B.2. □

Decentralization of the social planner's allocation requires two instruments: directly-set minibus fares τ_{ijM}^* and budget-neutral commuter taxes $t_{ijm}^{\mathcal{S}*}$. Consistent with my later estimates, I now assume that the matching function exhibits constant returns to scale.

⁷Per-newly-entering minibus profits equal the difference between the left- and right-hand sides of (6), so the aggregate rebate satisfies $\Pi \equiv \sum_{i,j} b_{ij}^E \left[\frac{1}{g(T_{ij})} (\bar{\eta} \tau_{ijM} - \chi \Delta_{ij}) - \bar{\psi} b_{ij}^\phi \right] + \sum_{i,j} b_{ij}^E \bar{\psi} b_{ij}^\phi$, where b_{ij}^E denotes the flow of newly-entering minibuses on route ij .

Proposition 2. *Under constant returns to scale in matching, $\alpha + \beta = 1$, the social planner implements the optimal allocation via*

i. optimal minibus fares

$$\tau_{ijM}^* \equiv \frac{\chi \Delta_{ij}}{\bar{\eta}} + \frac{\bar{\psi}}{\bar{\eta}} \mathfrak{g} \left[\frac{\bar{\eta}^\beta}{\mu_{ij}} \left(\frac{2\beta}{1-\beta} \right)^{1-\beta} + \frac{1}{d_{ij}} \right] \left[\left(\frac{2\beta}{\bar{\eta}(1-\beta)} \right)^{1-\beta} \frac{\sum_g N^g \pi_{ijM}^{g*}}{\mu_{ij}} \right]^\phi \quad (13)$$

ii. commuter transfers

$$t_{ijm}^{g*} = \varsigma \chi_m^e \Delta_{ij} - T^g - 1 \{m = M\} \left(\tau_{ijM}^* - \frac{\chi \Delta_{ij}}{\bar{\eta}} \right) \quad (14)$$

where π_{ijM}^{g*} denotes the optimal choice probabilities, $\mathfrak{g}[\cdot]$, an increasing function, and T^g , constants set to ensure within-skill-group budget neutrality, $\sum_{i,j,m} N^g \pi_{ijm}^{g*} t_{ijm}^{g*} = 0$.

Proof. See Appendix B.3. □

Intuitively, optimal minibus fares correct the boarding and filling externalities to induce optimal minibus entry, while taxes ensure that commuters internalize any operating costs or environmental externalities induced by their mode choices. Given the intuitive nature of the latter, I focus on the optimal fares. The first term in (13) accounts for distance-dependent operating costs, and the second term corrects the net matching, or boarding minus filling, externality. Recall that higher values of the bus matching elasticity β generate a larger (positive) boarding externality, relative to the negative filling externality. The planner then induces this additional beneficial entry through fares which increase with the bus matching elasticity β , given sufficiently large $\bar{\eta}$. Notice the last term in brackets: due to congestion in entry costs, the boarding externality grows in relative magnitude on routes with higher optimal commuter demand, $\sum_g N^g \pi_{ijM}^{g*}$. As a result, correction of the matching externalities requires disproportionately higher fares on these routes with superior amenities or wages.

V. ESTIMATION OF MINIBUS AND DEMAND PARAMETERS

I combine my newly-collected data with existing micro-data to estimate the model for a geography composed of the $I = 18$ *transport analysis zones* in Cape Town.

Structural Parameters

I estimate the model's structural parameters in five main steps. First, I devise an instrumental variables strategy that employs the station count data to identify the matching elasticities α and β . Second, from commuters' stated preferences, I estimate their mode-specific utility costs κ_m^g , rate of time preference r , and Gumbel shock shape ν . Third, I externally calibrate the geography and other secondary parameters. Fourth, I calibrate the minibus entry cost elasticity ϕ and fares τ_{ijM} . Finally, conditional on all other parameters, I internally calibrate the minibus capacity $\bar{\eta}$, entry cost intercept $\bar{\psi}$, and matching efficiency μ . As part of this final step, I also invert the model to obtain home location-specific amenities θ_i^g and work location-specific wages ω_j^g . Table 1 summarizes all calibrated parameters.

Minibus Matching

First, I estimate the passenger and (mini)bus matching elasticities, α and β , using variation in demand over time within a given minibus route in my station count data. In contrast to much of the transport matching literature, I observe both sides of the matching market, so I can transparently estimate the matching elasticities without additional assumptions. Across 44 routes, indexed by origin i and destination j , and 48 five-minute clock time intervals indexed by t , I estimate the empirical equivalent of equation (3) for the bus loading rate:

$$\log \iota_{ijt} = \alpha \log p_{ijt} + (\beta - 1) \log b_{ijt} + \bar{\mu}_{ij} + \bar{\mu}_t + \bar{\mu}_{it} + \epsilon_{ijt}.^8 \quad (15)$$

I calculate the bus loading rate ι_{ijt} as the number of boarding passengers on a route per bus and minute; recall that p_{ijt} and b_{ijt} denote waiting passengers and buses, respectively. Furthermore, I model matching efficiency, $\log \mu_{ijt}$, as a combination of route ($\bar{\mu}_{ij}$), time ($\bar{\mu}_t$), and origin-by-time ($\bar{\mu}_{it}$) fixed effects, as well as an idiosyncratic shock ϵ_{ijt} to account for out-of-steady-state dynamics in the data. In particular, the origin-time fixed effects control for any shocks that equally affect the matching efficiencies of all routes that originate from a given station. Most plausible threats to identification, such as localized rain or special events which might both slow the loading process and affect demand, likely fall into this category. Rain shocks would nevertheless bias the OLS estimates to the extent that they affect routes at the same station, perhaps ones without roofs over their loading lanes, differently.

⁸The theory suggests many equivalent matching function estimation methods. I choose equation (3) for ι_{ij} to permit estimation with a single instrumental variable as well as later validation against actual bus wait times.

TABLE 1. CALIBRATED PARAMETERS

Parameter	Description	Value	Parameter	Description	Value
<i>Externally Calibrated</i>			<i>Stated Preference</i>		
I	Number Locations	18	r	Commuter Rate of Time Pref.	0.001
N^g	Commuter Populations		ν	Gumbel Shape	4.76
d_{ij}	Road-Based Destination Arrival Rate		κ_M^l	Low-Skill Minibus Util. Cost	7.7
d_{ijF}	Formal Destination Arrival Rate		κ_M^h	High-Skill Minibus Util. Cost	15
τ_{ijF}	Formal Fare		κ_F^l	Low-Skill Formal Util. Cost	3.6
τ_A	Car Commute Cost	5.2	κ_F^h	High-Skill Formal Util. Cost	9.2
δ_0	Minibus Shift Length	240	<i>Internally Calibrated</i>		
δ_1	Minibus Inverse # Trips	0.01	$\bar{\psi}$	Minibus Entry Cost Intercept	3.1
χ	Per-km. Operating Cost	0.06	$\bar{\eta}$	Minibus Capacity	6.2
Δ_{ij}	Route Driving Distance		μ	Minibus Matching Efficiency	0.2
χ_M^e	Minibus CO2-equiv./km.	0.06	<i>Model Inversion</i>		
χ_F^e	Formal CO2-equiv./km.	0.04	θ_i^g	Amenities	
χ_A^e	Car CO2-equiv./km.	0.55	ω_j^g	Wages	
ζ	Social cost of carbon	0.0485			
<i>Minibus Supply</i>					
α	Passenger Elasticity	0.84			
β	Bus Elasticity	0.16			
ϕ	Entry Cost Elasticity	0.602			
Γ_0	Fare Intercept	2.23			
Γ_1	Fare Distance Slope	0.29			

Notes: This table displays the full set of estimated model parameters. The externally calibrated parameters and geography come primarily from the 2013 Cape Town Household Travel Survey as well as the Google Maps and Azure APIs, and emissions parameters come from Borck (2019) and the U.S. Department of Energy (see Online Appendix C.2). The minibus matching elasticities and entry cost elasticity are estimated using the station counts; the minibus fares are calibrated using on-board tracking data. The stated preference estimation uses my new survey and an existing module from the aforementioned 2013 survey. The internal calibration minimizes the distance to moments in the station counts, and the model inversion again employs the 2013 survey data.

To address this more remote possibility, I pursue an instrumental variables strategy. I assume constant returns, $\alpha + \beta = 1$, to rewrite the estimating equation as a function of the relative number of waiting passengers to buses, i.e., $\log \iota_{ijt} = \alpha \log (p_{ijt}/b_{ijt}) + \bar{\mu}_{ij} + \bar{\mu}_t + \epsilon_{ijt}$.⁹ Then, I instrument for $\log (p_{ijt}/b_{ijt})$ with the log number of commuters resident in i who reported, in a 2013 survey, that they leave around time t to go to work. Since I measured loading rates ι_{ijt} in the year 2022, minute-by-minute matching efficiency trends in ϵ_{ijt} would have to persist over nine years to violate the exclusion restriction. Even under

⁹Because I have only one instrument, I cannot estimate flexible returns to scale in the IV specifications.

TABLE 2. MATCHING FUNCTION ESTIMATES

Parameter	OLS			IV with $\alpha + \beta = 1$	
	(1) log bus loading rate	(2) log bus loading rate	(3) log bus loading rate	(4) log bus loading rate	(5) log bus loading rate
α	0.687	0.662	0.645	0.841	0.665
<i>Passenger Matching Elasticity</i>	(0.0130)	(0.0188)	(0.0264)	(0.106)	(1.060)
β	0.433	0.425	0.435		
<i>Bus Matching Elasticity</i>	(0.035)	(0.042)	(0.043)		
95% CI for $\alpha + \beta$	[1.03,1.21]	[0.98,1.19]	[0.97,1.19]		
Route FE	✓	✓	✓	✓	✓
Time FE		✓	✓		✓
Origin-Time FE			✓		
Observations	1,627	1,627	1,607	1,316	1,316
R-Squared	0.587	0.818	0.839	0.50	0.56
First-Stage F Statistic				14.80	0.24

Notes: Robust standard errors in parentheses, clustered at the origin level. Columns 1–3 present estimates of (15) over five-minute time blocks and 44 routes in my station count data, with fixed effects included, as noted, for route, time, or origin station by time. In Columns 4–5, I assume constant returns to scale, $\beta = 1 - \alpha$, so that I can regress the log bus loading rate on the log ratio of the stock of waiting passengers to the stock of loading buses. I instrument for this ratio using the log number of commuters living in the mesozone spatial unit where the route originates who report leaving their home during the 15-minute period including time t , calculated from the 2013 Cape Town Household Travel Survey.

such persistence, the distribution of commute start times remains exogenous provided commuters do not choose their residential locations or the start times of their working days based on within-route matching efficiency trends over the course of the morning commute.

Table 2 displays the matching elasticities that result from the estimation of (15), first by OLS. In the specification in Column 1, I include only route fixed effects; in Columns 2 and 3, I add time and origin-time fixed effects, respectively. In Column 3, I cannot reject constant returns to scale at the 5% level, so I can safely assume the latter in the instrumental variables specifications in Columns 4–5. The passenger elasticity $\hat{\alpha} = 0.841$, thus estimated in Column 4, is larger than in prior specifications, but the corresponding 95% confidence interval includes the OLS estimate. In Column 5, I include time fixed effects and obtain a noisy but broadly similar estimate. In my model calibration, I revert to the IV estimates from Column 4 due to their greater precision. The estimates change little when, in Online Appendix C.1.1, I weight by routes' inverse sampling probabilities or estimate an alternative specification for the total, rather than per-bus, number of loading

passengers.

Stated Preferences

Second, stated preference data informs four sets of demand parameters, as follows. Respondents faced choices, conditional on their existing home and work locations, among commute alternatives l in a choice set c . Alternatives vary exogenously in their wait time w_{cl} , travel time t_{cl} , fares τ_{cl} , and quality improvements that affect utility costs. As for the latter, let $\bar{\kappa}_{cl}^g$ denote the utility cost of alternative l in choice set c for skill group g , $m(c, l)$, the associated mode of transport, and $q_{cl}(z)$, an indicator for the presence of (binary) quality improvement z . I assume that the mode-specific utility cost depends linearly on quality improvements, with intercept $\kappa_{m(c,l)}^g$, according to $\bar{\kappa}_{cl}^g \equiv \kappa_{m(c,l)}^g + \sum_z \zeta_z^g q_{cl}(z)$. I then estimate the multinomial logit discrete choice model consistent with commuter demand, (8)-(9), via maximum likelihood. I use the responses of individuals between 25 and 65 years old employed outside the home in my own and the city-run survey.

Individual i from group g then chooses alternative l from choice set c with probability

$$\pi_{icl}^g = \frac{\exp \left[\zeta_{m(c,l)}^g + \sum_z \beta_z^g q_{cl}(z) + \beta_{\text{time}} \omega_i (w_{cl} + t_{cl}) + \beta_{\text{fare}} \tau_{cl} + \beta_{\text{resid}} w_{cl} \tau_{cl} \right]}{\sum_{l'} \exp (U_{icl'}^g / \nu)}. \quad (16)$$

The ratio of the group-mode fixed effect $\zeta_{m(c,l)}^g$ to the fare effect β_{fare} places no-quality-improvement utility costs κ_m^g in dollar terms, where I normalize the car utility cost $\kappa_A^g = 0$. Crucially, while I over-sampled minibus commuters, my survey included only minibus alternatives and so does not contribute to the identification of the relative utility costs across modes. The choices of respondents in my data do, however, yield the effect of each quality improvement z on the minibus utility cost, calculated as $\zeta_z^g = \beta_z^g / \beta_{\text{fare}}$. I obtain the rate of time preference r from the ratio $\beta_{\text{time}} / \beta_{\text{fare}}$, or the relative increase in the disutility of commute time $(w_{cl} + t_{cl})$ with personal income ω_i . Finally, commuters' inverse responsiveness, $|1 / \beta_{\text{fare}}|$, to fares τ_{cl} identifies the Gumbel shape ν .¹⁰

Table 3 displays the parameters estimated from the multinomial logit model in (16). I begin with the rate of time preference, $\hat{r} = 0.001$, which implies that commuters would sacrifice only 1% of their daily wage to save ten minutes of commute time to and from

¹⁰Note that I have used the relationship between arrival rates and average travel time, $t_{cl} = 1/d_{cl}$, and a similar relationship for wait time. Because respondents make their choices conditional on their existing home and work locations, the wage and amenity terms fall out of the choice probability. Finally, unlike in the main model, I do not suppress the final term—the product of fares and wait time—implied by the micro-founded model in Online Appendices B.1-B.3.

TABLE 3. STATED PREFERENCE SURVEY ESTIMATES

Parameter	Value		Parameter	Value	
r	.001		<i>Effects on Utility Costs</i>	<i>Low-Skill</i>	<i>High-Skill</i>
<i>Commuter Rate of Time Pref.</i>	(.0004)				
ν	4.76		ζ_{security}	-1.09	-2.75
<i>Gumbel Shape</i>	(1.26)		<i>Station Security</i>	(0.390)	(0.84)
<i>Utility Costs</i>	<i>Low-Skill</i>	<i>High-Skill</i>	$\zeta_{\text{no overloading}}$	-1.38	-1.39
			<i>Overloading Ban</i>	(0.437)	(0.543)
κ_F	3.63	9.17	$\zeta_{\text{follows speed limit}}$	-1.36	-0.825
<i>Formal Transit Utility Cost</i>	(0.51)	(1.89)	<i>Speed Limit Enforcement</i>	(0.445)	(0.465)
κ_M	7.68	15.03			
<i>Minibus Utility Cost</i>	(1.56)	(3.55)			

Notes: Robust standard errors in parentheses. Estimates reflect $N = 19,712$ individuals by choice sets by alternatives in either my newly collected minibus stated preference survey or a stated preference module of the 2013 Cape Town Household Travel Survey. The estimated parameters come from a multinomial logit model with choice probabilities given by (16). I normalize $\kappa_A^s = 0$ and restrict the sample to individuals employed outside the home between 25 and 65 years of age.

work. Next, I estimate a Gumbel shape parameter $\hat{\nu} = 4.76$ and, by implication, a low cost sensitivity: minibus relative choice probabilities would fall by only 24% in response to a 100% increase in fares. The low- and high-skill minibus utility costs, $\hat{\kappa}_M^l = 7.68$ and $\hat{\kappa}_M^h = 15.03$, demonstrate that both groups dislike minibuses relative to cars. If commuters experienced identical utility costs for minibuses and cars, their relative minibus choice probabilities would rise by a factor of 4.35 for low-skill workers and 23.5 for high-skill workers. The same is true for formal transit, albeit to a lesser degree.

In the right panel of Table 3, all three quality improvements significantly decrease minibuses' utility cost for the low-skill group; high-skill commuters place a premium on security but not on speed limit adherence. For example, the provision of station security would almost double the high-skill minibus mode share relative to any other mode. To quantify any heterogeneity in preferences for minibus attributes, I re-estimate the logit model (16) among only respondents interviewed outside of minibus stations. Separately, I weight my survey by realized commute mode choices; the results, in Online Appendix C.1.2, change little.

External Calibration

Third, I externally calibrate secondary parameters as well as the model geography. I parameterize the function g that determines the rate x at which a minibus exits upon

arrival at its destination as

$$x \equiv g(T_{ij}) \equiv 1 / \{1 + \exp[-\delta_1 (T_{ij} - \delta_0)]\}. \quad (17)$$

The workday “length” parameter δ_0 corresponds to the total time available to the bus to make trips. The parameter δ_1 , then, determines how quickly the expected total trip time T_{ij} decreases the feasible number of daily trips. I calibrate these two parameters to produce reasonable expected numbers of trips for various trip times. To quantify emissions costs, I use the social cost of carbon ζ from Borck (2019). Finally, I directly obtain the emissions factors χ_m^e , car commute cost τ_A , minibus operating cost χ , and the model geography from the data. The latter includes commute populations N^g , origin-destination driving distances Δ_{ij} , destination arrival rates $\{d_{ij}, d_{ijF}\}$, and formal transit fares τ_{ijF} . I provide additional details regarding all externally calibrated parameters in Online Appendix C.2.

Minibus Supply

Fourth, I quantify additional minibus supply parameters. The less strongly bus entry in the station count data responds to trip time, the greater the congestion elasticity, ϕ , of minibus entry costs and the more severe any minibus association entry restrictions. In particular, from free entry, I derive that the number of loading buses b_{ij} on route ij depends on the (log) expected number of daily trips – the second term below – which itself depends on bus loading and travel times,

$$\log b_{ij} = \zeta_0 + \frac{1}{\phi} \log \left\{ 1 + \exp \left[-\delta_1 \left(\frac{\bar{\eta}}{t_{ij}} + \frac{1}{d_{ij}} - \delta_0 \right) \right] \right\} + \mathbf{X}_{ij}\zeta + \varepsilon_{ij}. \quad (18)$$

The vector \mathbf{X}_{ij} includes the log of per-trip net revenue, $\bar{\eta}\tau_{ijM} - \chi\Delta_{ij}$, as implied by the model. Thus, the threats to identification in ε_{ijt} include determinants of profitability other than the number of trips, fare revenue, or gas costs: for example, variation in bus maintenance costs across neighborhoods. I estimate this modified free entry equation (18) by OLS across routes in my data and, conditional on the parameters δ that control the number of trips minibuses can accomplish per day, find an entry cost elasticity of $\hat{\phi} = 0.602$.¹¹ This congestion elasticity suggests that minibus associations, for example, impose moderate barriers to entry. Furthermore, I calibrate minibus fares τ_{ijM} as a log-

¹¹Standard error = 0.326, $N = 43$ routes, $R^2 = 0.163$. The station count data provides the average number of loading buses on a route and the corresponding mean bus loading time. The on-board tracking data provides average travel time as well as per-trip revenue for each route, and I employ the externally calibrated values of the remaining parameters.

TABLE 4. INTERNAL CALIBRATION

Moment			Parameter		
<i>Description</i>	<i>Data</i>	<i>Model</i>	<i>Description</i>	<i>Value</i>	
Median Loading Buses/Waiting Passengers	0.09	0.09	$\bar{\psi}$ Entry Cost Intercept	3.1	
Median Bus Loading Time	4	4	$\bar{\eta}$ Minibus Capacity	6.2	
Median Off-Bus Passenger Wait Time	7.18	7.18	μ Matching Efficiency	0.2	

Notes: This table displays the moments used in internal calibration: medians across routes and five-minute periods in the station count data and medians across routes in the model. I also list the model parameter heuristically corresponding to each moment, along with its internally-calibrated value. For calibration, I choose a close-to-optimal starting point, which I then feed into the simplex search method to numerically minimize the sum of squared (percentage) deviations from the three moments. In each iteration, I invert the model equations for employment by residence and workplace to obtain implied residential amenities and workplace wages.

linear function of route distance by way of an OLS specification across routes in the minibus tracking data.¹²

Internal Calibration and Inversion

Finally, I match aggregate moments in the station count data to obtain the remaining supply parameters, and, for each iteration of moments, invert the model to obtain location-specific amenities and wages. Specifically, I match (i) the median, over routes and time periods, of the relative number of loading buses, b_{ijt}/p_{ijt} ; (ii) the median bus loading time, $\bar{\eta}/\iota_{ijt}$; and (iii) the median off-bus passenger wait time, $1/\lambda_{ijt}$. As in standard spatial models, I then invert the model: I exactly match each location's observed employment, by residence to obtain home location-specific amenities θ_i^s and by workplace to obtain workplace-specific wages ω_j^s .¹³

The second and third columns of Table 4 list the values of each moment in the data and in the calibrated model across routes. Heuristically, the relative number of loading buses identifies entry costs $\bar{\psi}$, bus loading time identifies bus capacity $\bar{\eta}$, and the passenger off-bus wait time determines matching efficiency μ . The final column lists the calibrated parameter values; the median realized entry cost exceeds the fare revenue from the median trip by a factor of eight.

¹²Constant $\hat{\Gamma}_0 = 2.23$, slope $\hat{\Gamma}_1 = 0.29$ with standard error 0.023. $N = 43$ routes, $R^2 = 0.8$. Origin-destination straight-line distance and fares reflect averages over observed trips on a route.

¹³I measure employment by residence and workplace in the 2013 Cape Town Household Travel Survey and normalize, for each skill group, (i) the amenity of one location to zero and (ii) the average wage to the empirical skill-group average.

VI. MODEL-PREDICTED MINIBUS OPERATIONS AND AGGREGATE PATTERNS

I now demonstrate that my model matches non-targeted data: the minibus network, the boarding and filling externalities, and aggregate mode shares. First, the model-predicted bus entry on each origin-destination pair mirrors a rough proxy of entry in the data, namely the number of distinct minibus routes that link each pair of neighborhoods (see Appendix Figure A.2).¹⁴ Importantly, the model matches the concentration of minibuses in central neighborhoods and their ability to directly link outlying suburbs, in contrast to the usual service patterns of publicly-provided transit.

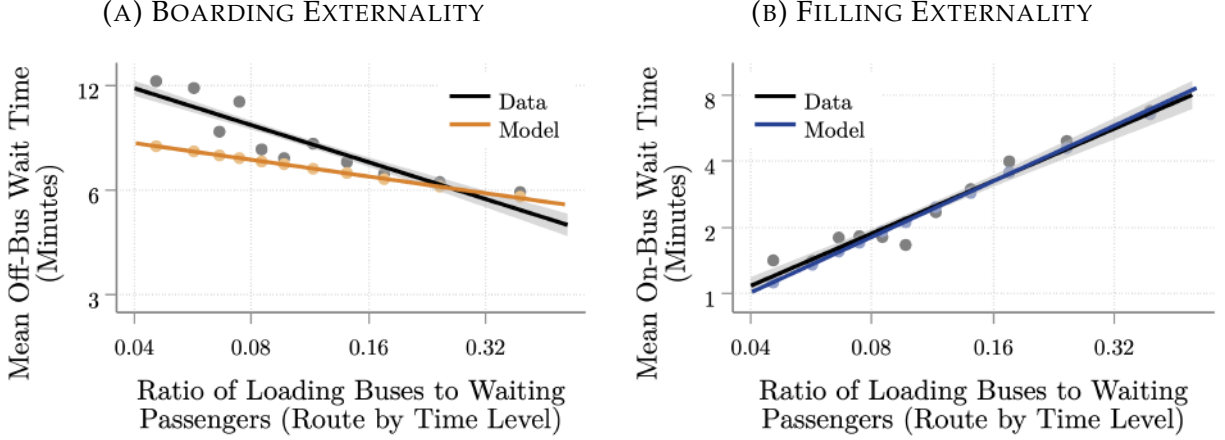
Second, the model replicates the boarding and filling externalities which play a central role in my theory. Figures 6a and 6b display the relative number of loading buses to passengers on the horizontal axis and off-bus or on-bus wait times on the vertical axis. I plot route-by-time observations in the station count data in black and routes in the model in color. In both figures, my calibration targets the *medians* of the horizontal and vertical axes but not the relationship between bus entry and wait times. Reassuringly, the model nevertheless generates off-bus wait times that fall with bus entry in Figure 6a, precisely as in the data. I match the filling-externality relationship in Figure 6b even more closely.

Third, the model accurately predicts commute choices. The bars in Figure 7 indicate the skill-group-level shares of Cape Town commuters in the household survey data (light colors) and model (dark colors) who use each mode. In both data and model, a sizable share of low-skill commuters use minibuses, while high-skill commuters choose minibuses at comparatively low rates and cars at high rates. Furthermore, I confirm in Online Appendix D.2 that model-predicted origin-destination-level mode shares correlate strongly with data; predicted minibus choice probabilities also decrease, and those for cars increase, with average income, albeit less strongly than their empirical counterparts. The close match between these non-targeted mode shares computed from 2013 survey data and those implied by the model suggests that the stated preference methodology accurately captures real-world preferences. Indeed, I demonstrate in Online Appendix D.1 that the skill differential in the model results almost entirely from differences in the utility costs – estimated from the stated preference surveys.

Because my later results depend crucially on the associated estimates, I demonstrate the

¹⁴Note that, in reality, unlike in my model, many neighborhood (transport analysis zone) pairs are linked by multiple distinct minibus routes.

FIGURE 6. MATCHING EXTERNALITY RELATIONSHIPS IN DATA VERSUS MODEL



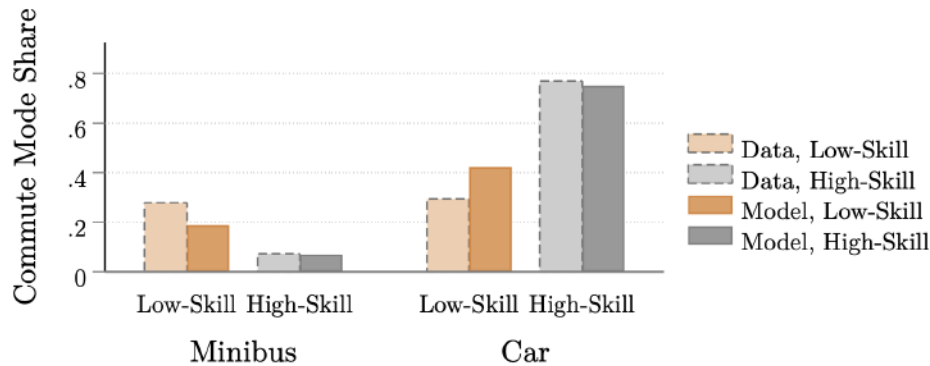
Notes: Panel (A) displays a binned scatterplot and best-fit line corresponding to the boarding externality of entry, i.e. the log-scale relationship between the relative number of loading buses to waiting passengers, $\log(b_{ij}/p_{ij})$, across routes and five-min. periods in the station count data and across routes as predicted by the model, and expected passenger off-bus wait time, $\log(1/\lambda_{ij})$. Panel (B) displays the filling externality, instead plotting expected passenger on-bus wait time, $\log[\bar{\eta}/(2t_{ij})]$, on the vertical axis.

plausibility of respondents' stated preferences in two additional ways, laid out in Online Appendix D.3. First, the model predicts not only citywide mode shares, as in Figure 7, but also the reported commute modes of the stated preference respondents. Second, demographic heterogeneity in commuters' values of time and quality improvements largely follows intuition. Women, for example, place a higher value on time saved, as Borghorst et al. (2021) similarly find.

VII. URBAN TRANSPORTATION POLICIES

Finally, I use the estimated model to compare the gains from the typical recommendation of formal bus rapid transit investments to alternative policy strategies which optimize the existing minibus provision. I investigate the welfare effects of (i) the existing MyCiti bus rapid transit line in Cape Town; (ii) the implementation of the social planner optimum; and (iii) the provision of security at publicly-owned minibus stations. In terms of model objects, these policies modify (i) formal transit destination arrival rates d_{ijF} , (ii) minibus fares τ_{ijM} and commuter taxes t_{ijm}^g , and (iii) utility costs κ_m^g . I present changes in the welfare measure Ω defined in (12) as equivalent variation: the proportionate change in a skill group's wages ω_j^g , at baseline values of $\{\lambda, \iota, d_F, \tau_M, \kappa\}$, that leaves the average commuter

FIGURE 7. COMMUTE MODE SHARES BY SKILL GROUP IN DATA VERSUS MODEL



Notes: This figure displays the shares of low- (non-college) and high-skill commuters in Cape Town who use each mode in the data and as predicted by the model. Data comes from the 2013 Cape Town Household Travel Survey; in the model, I sum choice probabilities π_{ijm}^g by skill and mode.

equally well off as in the counterfactual.¹⁵ Table 6 at the end of this section summarizes all counterfactuals.

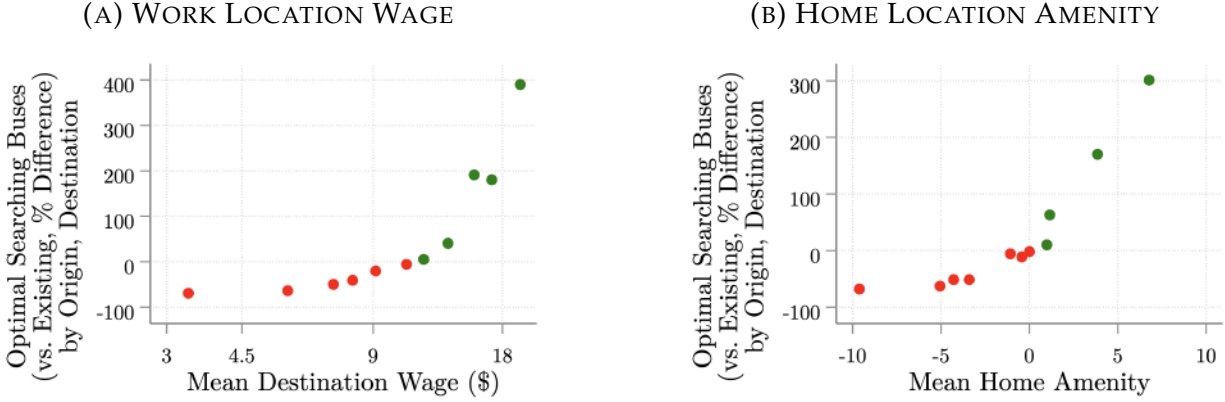
MyCiti Bus Rapid Transit

First, I consider the solution most often recommended to developing-country cities, namely the construction of formal bus rapid transit lines. Indeed, Cape Town has already built one such *MyCiti* line, with exclusive bus lanes and elaborate stations, to the northern suburbs. To investigate whether this new commute option justifies the construction cost of around \$250 million, I start with a pre-policy economy without the Northern Suburbs-CBD formal transit network link which corresponds to MyCiti. I then reimpose the status-quo formal transit destination arrival rates d_{ijF} —which include MyCiti—but commuters pay (annuitized flows of) construction plus operations costs via equal lump-sum taxes.¹⁶ Due to the limited spatial reach of MyCiti infrastructure, only a handful of home and work locations, displayed in Figure 10a, benefit from inflows of residents and workers. In

¹⁵To calculate welfare at the skill-group level in a manner unaffected by minibuses fares and transfers, I rebate minibuses profits and entry costs as follows: for each route, I multiply route-level profits and entry costs by a group's share among minibuses commuters on that origin-destination and then sum across routes (i.e. origin-destination pairs). I allocate other monetary costs of policies as well as emissions costs according to population.

¹⁶MyCiti costs derive from the City of Cape Town (2015). Construction costs include R4,157,851,000 in Phase 1 infrastructure, vehicle, and compensation costs (2006-2013) in Table 9-1 plus R1,062,320 in planning and transition costs (2006-2013) in Table 9-3. Operating costs per-commute are R390,447,000 in "deficit before funding," i.e. net of fare revenue, in 2015/16, from Table 9-4, divided by two times the number of working days per year.

FIGURE 8. SOCIAL PLANNER: OPTIMAL MINIBUS ENTRY, VERSUS...



Notes: Panels (A) and (B) display the the percent deviation of the optimal minibus entry b_{ij}^* from the status-quo entry on the vertical axis and the average across skill groups, weighted by aggregate populations, of the corresponding work-location wage and home-location amenity, respectively, on the horizontal axis, each as binned scatterplots.

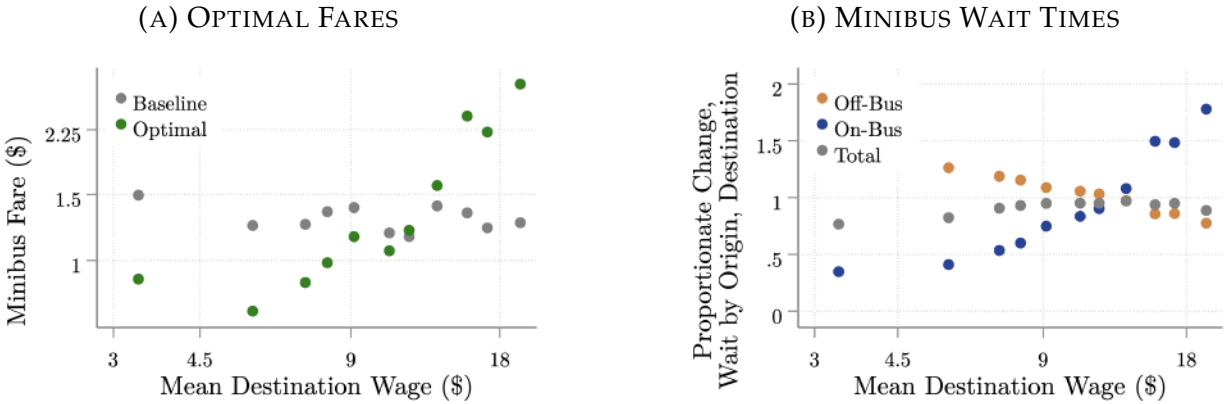
turn, the fixed costs of MyCiti significantly exceed these benefits, and welfare inequality increases. The high-skill lose approximately 2% in equivalent variation terms, and the low-skill, over twice as much, as detailed in Table 6. Indeed, in sprawling cities like Cape Town, widely dispersed home and work locations may not justify the economies of scale inherent to costly formal transit infrastructure.

Social Planner Optimum

Second, I implement the socially optimal minibus entry and commuter choices. Figures 8a and 8b display two determinants of commuter demand, namely the wages ω_j^g at the destination and amenities θ_i^g at the origin of a route, respectively, on the horizontal axis. On the vertical axis, I plot the optimal bus entry b_{ij}^* , relative to the status quo bus entry. On low-demand routes, i.e. those with low wages or amenities, the filling externality outweighs the boarding externality, and too many minibuses enter in equilibrium. In contrast, on higher-demand routes, the boarding externality fosters under-provision of minibuses.

The optimal minibus fares τ_{ijM}^* , when imposed on the minibus sector, and mode-specific commuter taxes t_{ijm}^{g*} from Proposition 2 correct these externalities. Figure 9a plots the existing fares, which reflect both government restrictions and minibus association market power, and the optimal minibus fares against destination wages. As suggested in Section IV, the social planner sets increasingly higher fares on higher-demand routes to induce

FIGURE 9. SOCIAL PLANNER: OPTIMAL FARES REDUCE WAIT TIMES



Notes: Panel (A) displays the status quo and optimal minibus fares, τ_{ijM} and τ_{ijM}^* , by origin and destination, versus the average across skill groups, weighted by aggregate populations, of the corresponding destination-location wage as a binned scatterplot. Panel (B) instead displays, on the vertical axis, the proportionate change in off- and on-bus minibus passenger wait times, from the status quo to the social optimum.

additional beneficial bus entry and correct the matching externalities. In Figure 9b, I again display the mean destination wage on the horizontal axis and, on the vertical axis, the proportionate change in off- and on-bus wait times from the status quo to the optimum. On lower-demand routes, the lower fares reduce congestion and on-bus waits; on higher-wage routes, increased fares alleviate the boarding externality, and off-bus waits fall significantly. Importantly, unlike in the case of the bus rapid transit line to a single neighborhood, total wait and thus commute times fall across the board.

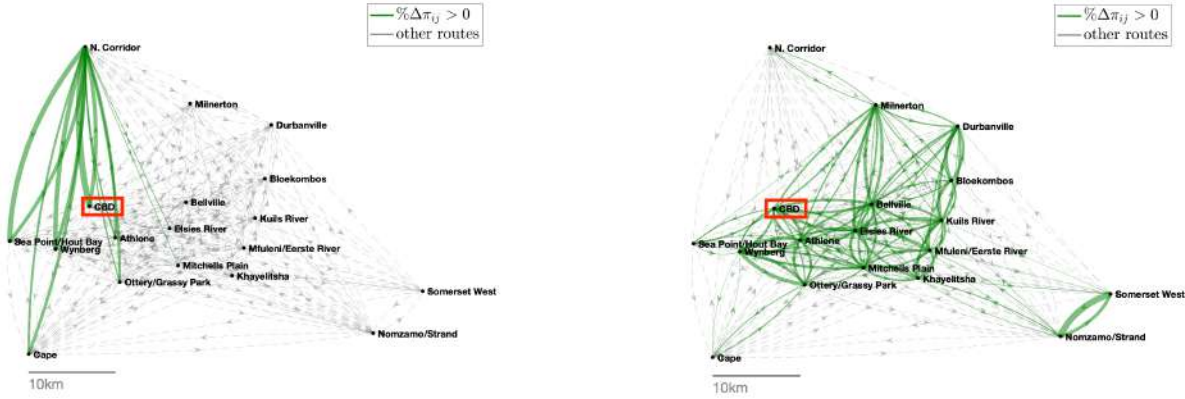
This social optimization of the minibus system has important effects – on the spatial structure of the city and on emissions – beyond direct commute time gains. Figure 10b highlights origin-destination commute flows which increase in size from the status quo to the social optimum, with the line width proportional to the increase. Not only radial connections to the central business district (CBD), but also a broad array of suburb-to-suburb commutes grow in popularity. This broad reallocation contrasts with MyCiti, which, recall, primarily reallocates commuters towards the Northern Suburbs it serves (Figure 10a). In consequence, workers, on average, reallocate towards higher-wage jobs, as enumerated in Table 6, and carbon emissions decrease by a full 10%. As a result, low-skill commuters gain 0.4% in equivalent variation from the social optimum, relative to the status quo, and the high-skill gain 0.1%. As Table 5 lays out, low-skill commuters’ changes of home and work location generate around 6% of their gains.¹⁷

¹⁷Note that high-skill workers’ location choice component is negative because correction of environmental externalities shifts them away from otherwise more-attractive home and work locations.

FIGURE 10. GROWING COMMUTE FLOWS, VERSUS STATUS QUO

(A) MYCITI FORMAL BUS RAPID TRANSIT

(B) SOCIAL PLANNER OPTIMUM



Notes: Panels (A) and (B) map the origin-destinations with larger numbers of commuters after the MyCiti formal bus rapid transit and social planner optimum counterfactuals, respectively, relative to the status quo, with the line width proportional to the percentage increase in commuters.

TABLE 5. SOCIAL PLANNER: DECOMPOSITION OF WELFARE GAINS

Skill	Share of Total Welfare Effect $\Delta\Omega$ from...		
	Commute Time	Mode Choice	Location Choice
Low	9.6	84.0	6.4
High	8.2	94.8	-3.0

Notes: This table displays the shares of the raw change in the skill-group-equivalent of welfare Ω from implementation of the social optimum obtained by changing, from the baseline to the optimal values: (i) expected *commute times* (via passenger boarding and bus loading rates λ_{ij} and ι_{ij}), holding choice probabilities constant; (ii) *mode choice* probabilities, holding home and work location choice constant; and (iii) *location choice* probabilities. Each step of the decomposition conditions on the socially-optimal values of the quantities adjusted in previous steps; I display the additional change in welfare associated with each step.

Minibus Security

Third, I evaluate the government provision of security guards at minibus stations. I adjust the minibus utility cost κ_M^g by the estimated binary security effect $\zeta_{security}^g$, and commuters pay their lump-sum share of guard wages.¹⁸ Because they place a larger premium on security, high-skill commuters shift even more strongly towards minibuses than their low-skill counterparts, as summarized in the left panel of Table 6. The right panel, in turn,

¹⁸The hourly guard wage, at only twice the median minibus fare, plays a minuscule role in welfare. I assume 2 guards per route; commuters pay a lump-sum tax to cover four hours of guard costs during the morning peak commute at a wage quoted by a local security firm.

TABLE 6. COUNTERFACTUAL URBAN TRANSPORTATION POLICIES

Policy	Skill:	Change in Mode Share				% Change in...				
		Minibus		Car		Mean Wage		Emissions	Welfare	
		Low	High	Low	High	Low	High		Low	High
Bus Rapid Transit		-0.0006	-0.0003	-0.001	-0.004	0.04	0.02	-0.29	-3.9	-1.7
Social Planner		0.05	0.02	-0.05	-0.04	0.16	0.14	-10.2	0.4	0.1
Minibus Security		0.04	0.05	-0.02	-0.04	0.03	-0.01	-3.76	2.3	1.2

Notes: This table displays the results of three counterfactuals: construction of the existing MyCiti formal bus rapid transit line, implementation of the social optimum via optimal minibus fares τ_{ijM}^* and mode-specific commuter taxes t_{ijm}^{g*} , and adding security to all minibus stations. The first four columns show the changes in the minibus and car mode shares by skill group. The final five show the percent change in group-level average wages ω_j^g , total emissions, and group-level welfare, the latter measured as equivalent variation, net of any associated costs.

demonstrates the extent to which poor safety constrains workers' choices. A policy as simple as posting station security guards can actually improve the allocation of workers to higher-wage jobs and reduce carbon emissions. Because the low-skill predominate among existing minibus riders, low-skill welfare increases, on net, by a striking 2.3%, compared to 1.2% for the high-skill.

Finally, I investigate additional minibus-targeted policies, namely mobile-app-motivated increases in matching efficiency, construction of exclusive minibus lanes, and reorganization of the minibus loading process so that buses wait in an orderly queue and load one at a time. In Online Appendix D.4, I detail the limited gains from these alternatives. As a robustness exercise, I incorporate road congestion effects on minibus and car travel times in Online Appendix D.5, but doing so scarcely changes the gains from my three main counterfactuals.

VIII. CONCLUSION

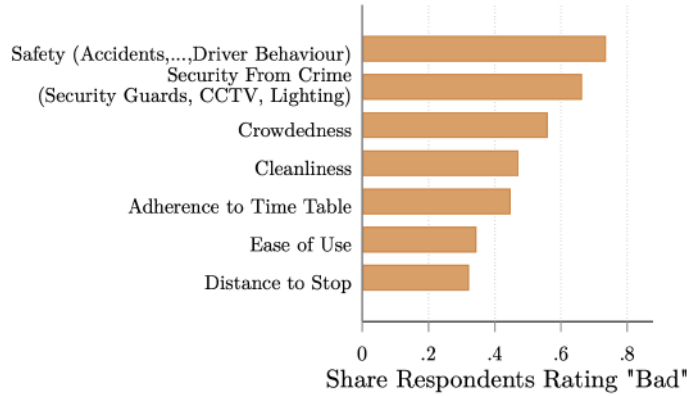
In this paper, I build a model of the privatized shared transit sector that dominates many developing-country cities. My theory unpacks the literature's typically exogenous transport costs and highlights opposing boarding and filling externalities in minibus matching. Unusually, newly-collected data on both the passenger and bus sides of this matching market in Cape Town permits direct estimation of the matching function. Furthermore, I

introduce the stated preference approach to the urban literature in order to identify the commuter demand system. Finally, I compare the model-implied gains from the typical formal bus rapid transit infrastructure to alternative policy strategies whose impact depends crucially on the response of the minibus sector.

The existing MyCiti formal bus rapid transit line's benefits fail to justify its high construction costs, in contrast to policies which directly target the minibus sector. In particular, the positive boarding outweighs the negative filling externality of minibus entry on higher-wage and amenity routes. The social planner thus optimally decreases minibus fares on the former routes, increases them on the latter, and additionally corrects environmental externalities with mode-specific taxes. On net, low-skill commuters gain almost one-half percent in welfare terms. However, personal safety turns out to distort home, work, and mode choices so severely that the gains from government-provided security guards at minibus stations dwarf those of other policies. Thus, even developing-country cities with the resources to build subways or bus rapid transit should consider the lower-cost alternatives.

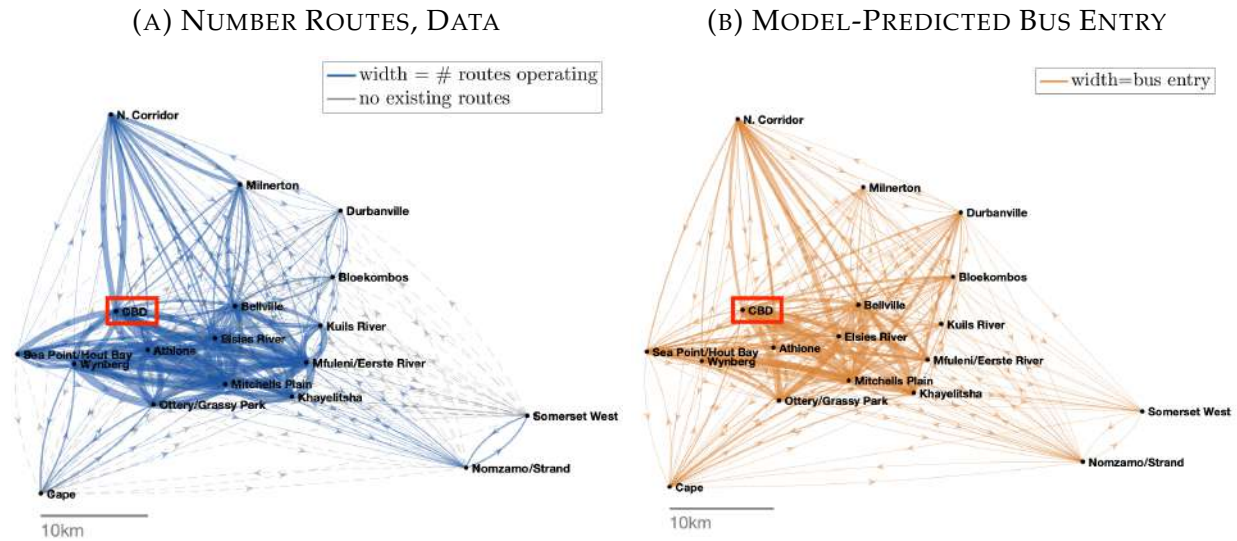
A. ADDITIONAL FIGURES

FIGURE A.1. RATINGS OF MINIBUS ATTRIBUTES IN PAST SATISFACTION SURVEY



Notes: Bar graph displays share of respondents from representative sample of commuters across all modes ($N = 1685$) in the 2013 Cape Town Household Travel Survey rating each minibus attribute as “bad,” as opposed to “acceptable” or “good.”

FIGURE A.2. MINIBUS NETWORK IN DATA VERSUS MODEL



Notes: Map in Panel (A) displays number of minibus routes linking each origin-destination pair of transport analysis zones according to a GIS shapefile created through a collaboration between GoMetro and the City of Cape Town. Note that, since these neighborhood units include multiple minibus stations in the real world, many pairs are linked by multiple “routes” in the data, in contrast to my model. The map in Panel (B) displays origin to destination lines whose thickness corresponds to the model-predicted minibus entry per unit time.

B. THEORY

B.1 Alternative Welfare Expression

Lemma B.1. *Welfare satisfies*

$$\begin{aligned}
 \Omega = & \sum_{i,j,g} N^g \pi_{ijM}^g \left[\theta_i^g - \kappa_M^g - r\omega_j^g \left(\frac{1}{\lambda_{ij}} + \frac{1}{2} \frac{\bar{\eta}}{t_{ij}} + \frac{1}{d_{ij}} \right) + \omega_j^g - \nu \log \pi_{ijM}^g - \frac{\chi \Delta_{ij}}{\bar{\eta}} - \varsigma \chi_M^e \Delta_{ij} \right] \\
 & + \sum_{i,j,g} N^g \pi_{ijF}^g \left[\theta_i^g - \kappa_F^g - r\omega_j^g \frac{1}{d_{ijF}} - \tau_{ijF} + \omega_j^g - \nu \log \pi_{ijF}^g - \varsigma \chi_F^e \Delta_{ij} \right] \\
 & + \sum_{i,j,g} N^g \pi_{ijA}^g \left[\theta_i^g - \kappa_A^g - r\omega_j^g \frac{1}{d_{ij}} - \tau_A + \omega_j^g - \nu \log \pi_{ijA}^g - \varsigma \chi_A^e \Delta_{ij} \right] \quad (\text{B.1})
 \end{aligned}$$

and thus does not depend on minibus fares τ_{ijM} .

Proof. In (12), consider first the ex-ante expected utility of group- g commuters, \bar{W}^g , given their optimal choices of home, work, and mode, subject to idiosyncratic Gumbel-distributed preference shocks.¹⁹ Denoting total deterministic utility of alternative ijm by $\bar{U}_{ijm}^g \equiv \theta_i^g + U_{ijm}^g + \omega_j^g$, I rewrite expected utility as

$$\begin{aligned}
 \bar{W}^g & \equiv E \left[\max_{i',j',m'} \left(\bar{U}_{i'j'm'}^g + \nu \varepsilon_{i'j'm'} \right) \right] \\
 & = \sum_{i,j,m} \pi_{ijm}^g \left[\bar{U}_{ijm}^g + \nu E \left(\varepsilon_{ijm} | ijm \in \operatorname{argmax}_{i',j',m'} \left(\bar{U}_{i'j'm'}^g + \nu \varepsilon_{i'j'm'} \right) \right) \right] \\
 & = \sum_{i,j,m} \pi_{ijm}^g \left[\theta_i^g + U_{ijm}^g + \omega_j^g - \nu \log \pi_{ijm}^g \right]. \quad (\text{B.2})
 \end{aligned}$$

The final equality uses a well-known result that the expected value of Gumbel preference shocks of agents who have optimally chosen a given alternative equals the negative of the corresponding log choice probability.²⁰ Next, the aggregate profits plus entry-cost rebate

¹⁹Note that the social planner will choose choice probabilities directly and then implement these choice probabilities with appropriately-set transfers.

²⁰To see this, note that

$$\begin{aligned}
 E \left[\varepsilon_{ijm} | ijm \in \operatorname{argmax}_{i',j',m'} \left(\bar{U}_{i'j'm'}^g + \nu \varepsilon_{i'j'm'} \right) \right] & = \frac{1}{\nu} \left[E \left[\max_{i',j',m'} \left(\bar{U}_{i'j'm'}^g + \nu \varepsilon_{i'j'm'} \right) \right] - \bar{U}_{ijm}^g \right] \\
 = \frac{1}{\nu} \left[\nu \log \left[\sum_{i',j',m'} \exp \left[\bar{U}_{i'j'm'}^g \right]^{1/\nu} \right] - \bar{U}_{ijm}^g \right] & = \log \left[\sum_{i',j',m'} \exp \left[\bar{U}_{i'j'm'}^g \right]^{1/\nu} / \exp \left(\bar{U}_{ijm}^g \right)^{1/\nu} \right] = -\log \pi_{ijm}^g
 \end{aligned}$$

Π depends on the flow of newly-entering buses b_{ij}^E and satisfies

$$\Pi \equiv \sum_{i,j} b_{ij}^E \left[\frac{1}{g(T_{ij})} (\bar{\eta} \tau_{ijM} - \chi \Delta_{ij}) - \bar{\psi} b_{ij}^\phi \right] + \sum_{i,j} b_{ij}^E \bar{\psi} b_{ij}^\phi = \sum_{i,j,g} N^g \pi_{ijM}^g \left[\tau_{ijM} - \frac{\chi \Delta_{ij}}{\bar{\eta}} \right]$$

where the second equality comes from a steady-state flow balance relationship, $b_{ij}^E = \sum_g \frac{g(T_{ij}) N^g \pi_{ijM}^g}{\bar{\eta}}$.²¹ Finally, recall that $E \equiv \zeta \sum_{i,j,m,g} N^g \pi_{ijm}^g \chi_m^e \Delta_{ij}$; substituting each element into (12), using the mode-specific definitions of commute utility U_{ijm}^g , and rearranging, I obtain the expression in (B.1). \square

B.2 Proof of Proposition 1 (Efficiency)

Proof. First, I derive the conditions characterizing the social planner optimum. From the definition of optimality in the main text, the planner solves

$$\max_{\{\pi\}_{i,j,m,g}, \{b\}_{i,j}} \Omega \text{ subject to (10), (11), and } \sum_{i,j,m} \pi_{ijm}^g = 1. \quad (\text{B.3})$$

By substituting (10) and (11) into the expression in Lemma B.1 to obtain welfare as $\Omega(\mathbf{b}, \boldsymbol{\pi})$, I can rewrite the planner's problem as $\max_{\mathbf{b}, \boldsymbol{\pi}} \Omega(\mathbf{b}, \boldsymbol{\pi})$ s.t. $\sum_{i,j,m} \pi_{ijm}^g = 1$.²² The planner's first-order conditions for the mass of searching buses b_{ij} on each route read

$$\frac{\partial \Omega}{\partial b_{ij}} = \frac{\beta}{\alpha} \mu_{ij}^{-1/\alpha} \left[\sum_g N^g \pi_{ijM}^{g*} \right]^{\frac{1-\alpha}{\alpha}} b_{ij}^{*\frac{-\beta-\alpha}{\alpha}} - \frac{1}{2} \frac{\bar{\eta}}{\sum_g N^g \pi_{ijM}^{g*}} = 0. \quad (\text{B.4})$$

The first-order conditions for optimal commuter choice probabilities π_{ijm}^{g*} , in turn, can be combined with the condition for a reference choice probability for each group g , π_{klA}^{g*} , to

where the first equality uses the well-known property of discrete choice models with Gumbel shocks whereby the conditional equals the unconditional expected utility, the second substitutes in for \bar{W}^g , and the final uses the choice probability equations (8)-(9).

²¹To maintain a constant mass of minibus commuters waiting or traveling, the mass of minibus commuters on route ij reaching j must equal the inflow, $\sum_g N^g \pi_{ijM}^g$. Given bus capacity, the mass of buses reaching j must be $1/\bar{\eta}$ the passenger mass, and a fraction $g(T_{ij})$ of these buses exit. For these bus outflows from service to equal inflows, we must have $b_{ij}^E = \sum_g \frac{g(T_{ij}) N^g \pi_{ijM}^g}{\bar{\eta}}$.

²²This welfare function includes the expectation of the idiosyncratic shocks, under the assumption that these choice probabilities can be implemented through appropriate transfers to commuters. I later confirm that the social planner can indeed do so.

derive an optimal relative choice probability:

$$\begin{aligned}
\log \left(\pi_{ijm}^{g^*} / \pi_{klA}^{g^*} \right) &= \exp \left(\theta_i^g - \kappa_m^g - r\omega_j^g \frac{1}{d_{ijm}} + \omega_j^g - \varsigma \chi_m^e \Delta_{ij} - \tilde{\gamma}^g / N^g \right. \\
&+ 1 \{m = M\} \left\{ -r\omega_j^g \left(\mu_{ij}^{-1/\alpha} b_{ij}^{*\beta/\alpha} \left[\sum_{g'} N^{g'} \pi_{ijm}^{g'^*} \right]^{\frac{1-\alpha}{\alpha}} + \frac{1}{2} \frac{\bar{\eta} b_{ij}^*}{\sum_g N^g \pi_{ijm}^{g^*}} \right) - \frac{\chi \Delta_{ij}}{\bar{\eta}} \right. \\
&\left. - \left(\sum_{g'} N^{g'} \pi_{ijm}^{g'^*} r\omega_j^{g'} \right) \left[\mu_{ij}^{-1/\alpha} \frac{1-\alpha}{\alpha} \left[\sum_g N^g \pi_{ijm}^{g^*} \right]^{\frac{1-2\alpha}{\alpha}} b_{ij}^{*\beta/\alpha} - \frac{1}{2} \frac{\bar{\eta} b_{ij}^*}{\left(\sum_{g''} N^{g''} \pi_{ijm}^{g''*} \right)^2} \right] \right\} \\
&\left. - 1 \{m \neq M\} \tau_{ijm} - \theta_k^g + \kappa_A^g + r\omega_l^g \frac{1}{d_{kl}} + \tau_A - \omega_l^g + \varsigma \chi_A^e \Delta_{kl} \right)^{1/\nu} = 0, \quad (\text{B.5})
\end{aligned}$$

where $d_{ijA} = d_{ijM} \equiv d_{ij}$ and $\tau_{ijA} \equiv \tau_A$. Equations (B.4), (B.5), and the adding-up constraints in (B.3) together define the *social planner's allocation* and can be solved for the I^2 optimal masses of searching buses b_{ij}^* as well as the $G \cdot I^2 \cdot 3$ optimal commuter choice probabilities $\pi_{ijm}^{g^*}$.

Second, I show that the social planner solution defined by (B.4)-(B.5) deviates from the competitive equilibrium defined in the main text. Substituting the bus loading rate (11) into free entry (6), I obtain $(\bar{\eta} \tau_{ijM} - \chi \Delta_{ij}) / g \left(\bar{\eta} b_{ij} / \sum_g N^g \pi_{ijM}^g + d_{ij}^{-1} \right) = \bar{\psi} b_{ij}^\phi$. This condition generically defines an equilibrium mass of searching buses b_{ij} distinct from the optimal entry b_{ij}^* pinned down by the planner's bus entry condition (B.4). From (8)-(9), it is immediate, after substituting in the matching technology (10)-(11), that relative choice probabilities, $\log \left(\pi_{ijm}^g / \pi_{klA}^g \right)$, will deviate from those chosen by the social planner in (B.5). \square

B.3 Proof of Proposition 2 (Optimal Fares and Taxes Under Constant Returns)

Proof. First, consider minibus fares. I substitute $\alpha + \beta = 1$ into (B.4) and solve for the planner's $b_{ij}^* = [2\beta / (\bar{\eta}(1-\beta))]^{1-\beta} \sum_g N^g \pi_{ijM}^{g^*} / \mu_{ij}$. Combining this equation with free entry (6) and the bus loading rate (11) demonstrates that the social planner can implement the optimal level of minibus entry b_{ij}^* by setting "optimal" minibus fares τ_{ijM}^* according to (13). Second, I derive *budget-neutral commuter taxes* $t_{ijm}^{g^*}$ differentiated by skill, home, work, and mode which induce the optimal choices $\pi_{ijm}^{g^*}$. Taxes $t_{ijm}^{g^*}$ are paid along with the fare upon choice of a commute. By comparing the social planner-optimal choice probabilities

(B.5) with the equilibrium (relative) choice probabilities implied by (8)-(9) and substituting in optimal minibus fares (13), we can see that the optimal choice probabilities can be implemented by taxes as in Proposition 2. \square

REFERENCES

- Ahlfeldt, Gabriel M., Stephen J. Redding, Daniel M. Sturm, and Nikolaus Wolf. 2015. "The Economics of Density: Evidence from the Berlin Wall." *Econometrica* 83 (6): 2127–2189.
- Akbar, Prottoy A., Victor Couture, Gilles Duranton, and Adam Storeygard. 2023. "Mobility and Congestion in Urban India." *American Economic Review* 113 (4): 1083–1111.
- Allen, Treb, and Costas Arkolakis. 2014. "Trade and the Topography of the Spatial Economy." *The Quarterly Journal of Economics* 129 (3): 1085–1140.
- . 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 Hickok, and Tobias Salz. 2023. "Optimal Urban Transportation Policy: Evidence from Chicago."
- Almås, Ingvild, Orazio Attanasio, and Pamela Jervis. 2023. "Economics and Measurement: New measures to model decision making." NBER Working Paper 30839.
- Ameriks, John, Joseph Briggs, Andrew Caplin, Matthew D. Shapiro, and Christopher Tonetti. 2020. "Long-Term-Care Utility and Late-in-Life Saving." *Journal of Political Economy* 128 (6): 2375–2451.
- Andrew, Alison, and Abi Adams-Prassl. 2023. "Revealed Beliefs and the Marriage Market Return to Education."
- Antrobus, Lauren, and Andrew Kerr. 2019. "The labour market for minibus taxi drivers in South Africa." SALDRU Working Paper No. 250.
- Balboni, Clare, Gharad Bryan, Melanie Morten, and Bilal Siddiqi. 2020. "Transportation, Gentrification, and Urban Mobility: The Inequality Effects of Place-Based Policies."
- Barwick, Panle Jia, Shanjun Li, Andrew R. Waxman, Jing Wu, and Tianli Xia. 2022. "Efficiency and Equity Impacts of Urban Transportation Policies with Equilibrium Sorting." *Revise & Resubmit, American Economic Review*.
- Baum-Snow, Nathaniel. 2007. "Did Highways Cause Suburbanization?" *The Quarterly Journal of Economics* 122 (2): 775–805.
- 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.

- Baum-Snow, Nathaniel, Matthew E. Kahn, and Richard Voith. 2005. "Effects of Urban Rail Transit Expansions: Evidence from Sixteen Cities, 1970-2000." *Brookings-Wharton Papers on Urban Affairs*, 147–206.
- Ben-Akiva, Moshe, Daniel McFadden, and Kenneth Train. 2019. "Foundations of Stated Preference Elicitation: Consumer Behavior and Choice-based Conjoint Analysis." *Foundations and Trends in Econometrics* 10 (1–2): 1–144.
- Bernheim, B. Douglas, Daniel Björkegren, Jeffrey Naecker, and Michael Pollmann. 2021. "Causal Inference from Hypothetical Evaluations." NBER Working Paper 29616.
- Borck, Rainald. 2019. "Public transport and urban pollution." *Regional Science and Urban Economics* 77:356–366.
- Borghorst, Malte, Ismir Mulalic, and Jos van Ommeren. 2021. "Commuting, children and the gender wage gap."
- Brancaccio, Giulia, Myrto Kalouptsidi, and Theodore Papageorgiou. 2020. "Geography, Transportation, and Endogenous Trade Costs." *Econometrica* 88 (2): 657–691.
- Brancaccio, Giulia, Myrto Kalouptsidi, Theodore Papageorgiou, and Nicola Rosaia. 2023. "Search Frictions and Efficiency in Decentralized Transport Markets." *Forthcoming, Quarterly Journal of Economics*.
- Buchholz, Nicholas. 2022. "Spatial Equilibrium, Search Frictions, and Dynamic Efficiency in the Taxi Industry." *The Review of Economic Studies* 89 (2): 556–591.
- Caplin, Andrew. 2021. "Economic Data Engineering." NBER Working Paper 29378.
- Castillo, Juan Camilo. 2022. "Who Benefits from Surge Pricing?" *Conditionally accepted, Econometrica*.
- Castillo, Juan Camilo, Daniel T. Knoepfle, and E. Glen Weyl. 2022. "Matching in Ride Hailing: Wild Goose Chases and How to Solve Them." *Revise & Resubmit, Management Science*.
- City of Cape Town. 2014. *Operating Licence Strategy 2013-2018*. Technical report. City of Cape Town Transport and Urban Development Authority, October. <https://tdacontenthub.store.blob.core.windows.net/resources/53226657-22e8-4795-b9f8-144f2b535636.pdf>.
- . 2015. *MyCiti Business Plan 2015 Update*. Technical report. Cape Town City Council, March. <https://www.myciti.org.za/docs/categories/1605/MyCiTi%5C%20Business%5C%20Plan%5C%202015%5C%20Update.pdf>.

- Coetzee, Justin, Christoff Krogscheepers, and John Spotten. 2018. "Mapping minibus-taxi operations at a metropolitan scale - methodologies for unprecedented data collection using a smartphone application and data management techniques."
- De Beer, Lourens, and Christo Venter. 2021. "Priority infrastructure for minibus-taxis: An analytical model of potential benefits and impacts." *Journal of the South African Institution of Engineering* 63 (4): 53–65.
- 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.
- Duranton, Gilles, and Matthew A. Turner. 2011. "The Fundamental Law of Road Congestion: Evidence from US Cities." *American Economic Review* 101 (6): 2616–2652.
- Fajgelbaum, Pablo D., Amit Khandelwal, Wookun Kim, Cristiano Mantovani, and Edouard Schaal. 2021. "Optimal Lockdown in a Commuting Network." *American Economic Review: Insights* 3 (4): 503–522.
- Fajgelbaum, Pablo D., and Edouard Schaal. 2020. "Optimal Transport Networks in Spatial Equilibrium." *Econometrica* 88 (4): 1411–1452.
- Fréchet, Guillaume R., Alessandro Lizzeri, and Tobias Salz. 2019. "Frictions in a Competitive, Regulated Market: Evidence from Taxis." *American Economic Review* 109 (8): 2954–2992.
- Glaeser, Edward L., Matthew E. Kahn, and Jordan Rappaport. 2008. "Why do the poor live in cities? The role of public transportation." *Journal of Urban Economics* 63:1–24.
- Gonzalez-Navarro, Marco, and Matthew A. Turner. 2018. "Subways and urban growth: Evidence from earth." *Journal of Urban Economics* 108:85–106.
- 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.
- Jobanputra, Rahul. 2018. *Comprehensive Integrated Transport Plan 2018 – 2023*. Technical report. City of Cape Town Transport and Urban Development Authority, January. <https://tdacontenthubstore.blob.core.windows.net/resources/fd3ddc0d-b459-4d26-bb01-7f689d7a36eb.pdf>.

- Johnston, Robert J., Kevin J. Boyle, Wiktor (Vic) Adamowicz, Jeff Bennett, Roy Brouwer, Trudy Ann Cameron, W. Michael Hanemann, et al. 2017. "Contemporary Guidance for Stated Preference Studies." *Journal of the Association of Environmental and Resource Economists* 4 (2): 319–405.
- Juster, F. Thomas. 1964. "Consumer Sensitivity to the Price of Credit." *The Journal of Finance* 19 (2): 222–233.
- Juster, F. Thomas, and Robert P. Shay. 1964. *Consumer Sensitivity to Finance Rates: An Empirical and Analytical Investigation*. NBER.
- Kerr, Andrew. 2018. *Background note: Minibus Taxis, Public Transport, and the Poor*. Technical report. World Bank. <https://openknowledge.worldbank.org/handle/10986/30018>.
- Kreindler, Gabriel E. 2022. "Peak-Hour Road Congestion Pricing: Experimental Evidence and Equilibrium Implications." *Conditionally accepted, Econometrica*.
- Kreindler, Gabriel E., 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."
- Mangham, Lindsay J., Kara Hanson, and Barbara McPake. 2009. "How to do (or not to do)...Designing a discrete choice experiment for application in a low-income country." *Health Policy and Planning* 24:151–158.
- Monte, Ferdinando, Stephen J. Redding, and Esteban Rossi-Hansberg. 2018. "Commuting, Migration, and Local Employment Elasticities." *American Economic Review* 108 (12): 3855–3890.
- Moreno-Monroy, Ana I. 2016. "Access to public transport and labor informality." *IZA World of Labor* 274. <https://wol.iza.org/articles/access-to-public-transport-and-labor-informality/long>.
- Nagy, Dávid Krisztián. 2023. "Hinterlands, city formation and growth: Evidence from the U.S. westward expansion." *Forthcoming, Review of Economic Studies*.
- OECD. 2016. *Time spent travelling to and from work*. Technical report LMF2.6. OECD Social Policy Division - Directorate of Employment, Labour and Social Affairs, December. https://www.oecd.org/els/family/LMF2_6_Time_spent_travelling_to_and_from_work.pdf.
- Rosaia, Nicola. 2023. "Competing Platforms and Transport Equilibrium."

- Rose, John M., and Michiel C.J. Bliemer. 2009. "Constructing efficient stated choice experimental designs." *Transport Reviews* 29 (5): 587–617.
- Schalekamp, Herrie. 2017. "Lessons from building paratransit operators' capacity to be partners in Cape Town's public transport reform process." *Transportation Research Part A* 104:58–66.
- Severen, Christopher. 2023. "Commuting, Labor, and Housing Market Effects of Mass Transportation: Welfare and Identification." *Forthcoming, Review of Economics and Statistics*.
- Theway, Chesway. 2018. *Pros & Cons of Minibus Taxis: The Transport System in South Africa*. <https://theway.medium.com/pros-cons-of-minibus-taxis-23afd16de783>.
- Tsivanidis, Nick. 2022. "Evaluating the Impact of Urban Transit Infrastructure: Evidence from Bogotá's TransMilenio." *Revise & Resubmit, American Economic Review*.
- Tun, Thet Hein, and Darío Hidalgo. 2022. *Learning Guide: Toward Efficient Informal Urban Transit*. Technical report. WRI Ross Center for Sustainable Cities and Transformative Urban Mobility Initiative (TUMI). <https://thecityfixlearn.org/en/learning-guide/toward-efficient-informal-urban-transit>.
- U.S. Department of Energy. 2023. *Alternative Fuels Data Center: Public Transportation*. Technical report. https://afdc.energy.gov/conserve/public_transportation.html.
- Uteng, Tanu Priya. 2011. *Gender and Mobility in the Developing World*. Technical report. World Development Report 2012: Gender Equality and Development, Background Paper. <https://openknowledge.worldbank.org/server/api/core/bitstreams/a9b17dbd-1764-569c-b8c5-f8fb8522d5d3/content>.
- Warnes, Pablo Ernesto. 2021. "Transport Infrastructure Improvements and Spatial Sorting: Evidence from Buenos Aires."
- Wooldridge, Jeffrey. 2019. "The Use of Survey Weights in Regression Analysis." Presented at The Use of Test Scores in Secondary Analysis PIAAC Methodological Seminar, Paris, 14th June 2019.
- Woolf, S.E., and J.W. Joubert. 2013. "A people-centred view on paratransit in South Africa." *Cities* 35:284–293.
- Zarate, Roman David. 2023. "Spatial Misallocation, Informality, and Transit Improvements: Evidence from Mexico City." *Revise & Resubmit, American Economic Review*.

ONLINE APPENDIX

A. DATA

A.1 Minibus Station Counts

Here, I provide more detail on the minibus station counts used to characterize the matching process. I designed the counts in cooperation with input from the South African firm GoAscendal, who also organized the logistics of data collection.

A.1.1 *Sample*

Resources allowed for the enumeration of 6 minibus routes per minibus station at 8 stations. For supervisory purposes, routes had to be enumerated in groups of 6, all operating from the same station.

Sampling Frame Complicating the sampling procedure is the fact that no fully comprehensive, accurate list of stations and routes exists. Thus, as a sampling frame, I employ a roster of routes by origin station derived from a 2018 collaboration between GoMetro and the City of Cape Town’s Transport and Urban Development Authority.²³ This listing is, according to stakeholders, as comprehensive and accurate as any available, and includes the number of minibus trips mapped by route in this previous data collection effort. Stakeholder discussions revealed that the number of trips mapped is an indicator of the number of buses on a route.

Population Since the team of 12 enumerators, who could cover 6 routes, had to be employed at the same station on a given day, my survey population consists of minibus routes originating from stations in Cape Town with at least 6 routes. The aforementioned sampling frame lists 519 routes operating from 107 stations. Of these stations, 31 have at least 6 routes, and these stations account for 328 of the 519 total routes, leaving a total population of $N = 328$ routes.

Clusters and Stratification I employed a two-stage stratified cluster sample, sampling 8 stations and then 6 routes originating at each of the 8 stations. I stratify within each stage

²³In order to update the city’s record of on-the-ground minibus route paths and operations, GoMetro sent enumerators armed with a smartphone app to ride on close to 30,000 minibus trips on the approximately 800 established minibus routes. The results showed the official, city-designated routes to be outdated. For example, 250 of the official city routes no longer operated (Coetzee et al. (2018)).

by a proxy for station and route-level bus entry, namely the aforementioned number of trips mapped in the previous data collection effort, to over-sample stations and routes with higher levels of bus entry and thus reduce the number of zero bus and passenger observations in the resulting data.

In the first stage, I took a stratified random sample of stations. First, I took a 100% sample of the 5 highest-bus-traffic feasible stations that operate in the morning peak, as measured by my proxy of bus entry, i.e. the total trips mapped originating at that station in the previous 2018 study.²⁴ Second, I sampled 3 stations, or a 16% sample, from the remaining 19 non-duplicate lower-bus-traffic stations with no permission issues, redrawing stations until obtaining 2 feasible ones. The final “busy” sample includes Du Noon, Bellville, Mfuleni, Khayelitsha Site C, and Mitchells Plain Station Eastern Side (North); the “less busy” sample includes Elsie’s River, Wesbank, and Nomzamo.²⁵

In the second stage, I took a stratified random sample of routes within each cluster, or station. Specifically, I sampled 4 routes (or the maximum possible, up to 4) per station from among those in the top ten percent of bus entry, as measured by trips mapped in the previous study, across all routes serving the 8 sampled stations. Then, I draw the remaining two or more routes, for a total of six, from those below the top ten percent of trips mapped. Thus, while feasibility constraints do not permit a constant sampling rate across stations, I obtain a random sample with variation in the traffic levels across routes. My final two-stage cluster sample of routes thus contains 48 routes clustered across six stations.

Practical Implementation In the field, 14 sampled routes unexpectedly did not operate at all in the morning peak. Field supervisors then randomly selected replacement routes at the same station on the spot, to the extent that there were additional routes operating at the station. One station, Elsie’s River, was discovered to have only 2 total routes in operation upon commencement of the day’s data collection, and logistical considerations meant that

²⁴From the full listing of 31 stations, some which would have otherwise counted among the 5 busiest had to be skipped due to minibus associations denying permission (Nyanga Central, Gugulethu Eyona) or not containing 6 routes that load off-road (Claremont Station). The 5 busiest feasible stations are in fact numbers 1-2, 4, 6, and 8.

²⁵Wynberg Station (Western Side) had to be excluded from the less busy sampling frame due to lack of permission, Cape Town CBD station due to lack of AM peak operations, and Mitchell’s Plain Station (North) and Promenade as well as Mitchells Plain Station Eastern Side (South) due to being adjacent to an already sampled station. After drawing the sample, further stations had to be excluded and resampled as follows. One station sampled, Khayelitsha (Vuyani), turned out to be part of an already sampled station (Khayelitsha (Nolungile Site C)); several others (Athlone, Vasco Station) do not operate as minibus stations with queues and loading off-road, and another set (Zevenwacht Mall, Mitchells Plain (Promenade), Tableview (Bayside)) do not operate in the AM peak.

FIGURE A.1. STATION COUNT DATA COLLECTION FORMS

(A) PASSENGER QUEUES

40036 Yale Cape Town Surveys

Rank Count Form | Passenger Waiting and Vehicle Arrival

Enumerator:		Rank:		Route:		Date:	
Time	People Waiting	Time	People Waiting	Time	People Waiting	Time	People Waiting
06:00 – 06:05		06:05 – 06:10		06:10 – 06:15		06:15 – 06:20	
06:20 – 06:25		06:25 – 06:30		06:30 – 06:35		06:35 – 06:40	
06:40 – 06:45		06:45 – 06:50		06:50 – 06:55		06:55 – 07:00	
07:00 – 07:05		07:05 – 07:10		07:10 – 07:15		07:15 – 07:20	
07:20 – 07:25		07:25 – 07:30		07:30 – 07:35		07:35 – 07:40	
07:40 – 07:45		07:45 – 07:50		07:50 – 07:55		07:55 – 08:00	
08:00 – 08:05		08:05 – 08:10		08:10 – 08:15		08:15 – 08:20	
08:20 – 08:25		08:25 – 08:30		08:30 – 08:35		08:35 – 08:40	
08:40 – 08:45		08:45 – 08:50		08:50 – 08:55		08:55 – 09:00	
09:00 – 09:05		09:05 – 09:10		09:10 – 09:15		09:15 – 09:20	
09:20 – 09:25		09:25 – 09:30		09:30 – 09:35		09:35 – 09:40	
09:40 – 09:45		09:45 – 09:50		09:50 – 09:55		09:55 – 10:00	
Vehicle ID	Arrival Time	Vehicle ID	Arrival Time	Vehicle ID	Arrival Time	Vehicle ID	Arrival Time

(B) BUS LOADING AND DEPARTURE

40036 Yale Cape Town Surveys

Rank Count Form | Departure Counts

Enumerator:		Rank:		Route:		Date:	
Vehicle ID	Time start loading	Time vehicle departs	Passenger onboard at departure	Vehicle ID	Time start loading	Time vehicle departs	Passenger onboard at departure

Notes: This figure displays the data collection forms used by enumerators to record station count data by hand for later digitization. Form (A) was used by the first of two enumerators to record the length of the passenger queue on an assigned route every 5 minutes from 6-10am as well as each minibuss that arrived on the station premises and its time of arrival. Form (B) was used by the second enumerator to record, for every minibuss that loaded passengers between 6-10am on a given route, the time it began actively loading passengers, the time of departure from the station, and the number of passengers on-board at departure.

no replacement routes at another station could be chosen. As a result, my final sample comprises 44 rather than 48 routes.

A.1.2 Data Collected

Station counts occurred on weekday mornings during one morning peak period (6-10am) per station, on weekdays from June 20-28 and 30, 2022. Two enumerators recorded data on each of 6 sampled routes over the course of the four-hour period. One enumerator stationed at the beginning of the lane and passenger queue corresponding to a route recorded the time a minibuss vehicle arrived and also the number of passengers waiting in the queue every 5 minutes on forms such as that in Online Appendix Figure A.1a. The second enumerator per route monitored bus loading and departures, recording the time a vehicle began loading passengers, the time of departure, and the number of passengers on board, as in Online Appendix Figure A.1b.

A.1.3 Cleaning

To calculate dwell times, I pair vehicle arrivals with the next possible departure; when multiple arrivals occur in a row without any departures, I pair only the last arrival with the next departure. For vehicle departures or arrivals that are not paired, I impute the corresponding missing arrival or departure.²⁶ I perform additional cleaning by adjusting vehicle ID formats to facilitate merging, removing duplicates in terms of vehicle, route, and arrival/departure time (separately), dropping records from the departure file where the loading start time is after the departure time, and dropping departures after 10:00. The resulting final dataset covers 44 routes.

A.1.4 Calculations

I discretize time into 5-minute periods, each beginning at some clock time t . I calculate the number of waiting buses on route l , b_{lt} , as the number of vehicles which arrive at the station before $t + 5$ and depart after t , i.e. the total number present during that 5-minute block, based on my imputations of arrivals and departures.

Calculating the number of passengers boarding buses, or “matches,” in a period is a bit more involved. Denote by loadingtime_{sl} the number of minutes between the time the minibus for trip s on route l begins loading and the time it departs, as described in Online Appendix Section A.1.2. I then assume that the passengers I observe departing from the origin station on trip s , deppax_{sl} , board the bus at a uniform rate. Then, I proportionately apportion these passengers who depart on a trip to the five-minute blocks during which the bus loads to calculate the total number of passengers boarding buses, P_{lt}^L , on route l from t to $t + 5$ as $P_{lt}^L \equiv \sum_s \frac{\text{loadingtime}_{sl} \cap [t, t+5]}{\text{loadingtime}_{sl}} \text{deppax}_{sl}$. Here, $\text{loadingtime}_{sl} \cap [t, t + 5)$ indicates the number of minutes of trip s 's loading time that overlap temporally with clock times t to $t + 5$. The number of departing passengers, P_{lt}^D , is the sum of all passengers on buses on route l which depart the station between t and $t + 5$, i.e. $P_{lt}^D = \sum_s 1 \{s \text{ departs in } [t, t + 5)\}$ deppax_{sl} .

Finally, I seek the number of unique passengers who wait during $[t, t + 5)$. For each route l

²⁶The process is simpler for vehicles that show up only once in the dataset. For these, I set the corresponding missing arrival (loading and departure) time to 6:00am (10:00am), under the assumption that such vehicles arrive before or load and depart after the hours during which the station counts were conducted. I perform a similar imputation if the non-paired departure is the first of the day or the non-paired arrival is the last of the survey period. For vehicles that show up multiple times and have a non-paired departure that is not the first of the day, I average the preceding departure time and the current loading start time to obtain an imputed arrival time. If a vehicle shows up multiple times and has a non-paired arrival that is not the last of the day, I impute the loading start time and the departure time as the average of the current and next arrival time.

and period t , I observe (a static snapshot of) the number of passengers waiting exactly at t , p_{lt} . Under the assumption that no passengers give up and stop waiting before boarding a bus, I can then calculate the total number of waiting passengers during period t , or between $[t, t + 5)$, as $p_{lt} \equiv P_{lt}^L + p_{lt}$. The average waiting time of these passengers to board buses, consistent with the model, is passenger wait time $_{lt} \equiv 5 \cdot \frac{p_{lt}}{P_{lt}^L}$ and is measured in minutes. The bus loading rate, denoted by ι in the model, is $\iota_{lt} \equiv \frac{P_{lt}^L}{5 \cdot b_{lt}}$.

A.2 Minibus Stated Preference Survey

Next, I detail my stated preference survey, also implemented by GoAscendal.

A.2.1 Questionnaire Design

I designed the questionnaire to maximize statistical power while retaining respondent attention. Since the 2013 Cape Town Household Travel Survey already contains discrete choice experiments containing different modes of transport, I focus exclusively on minibus commutes with different non-pecuniary attributes and cost.

Choice of Attributes and Levels In a discrete choice experiment, attributes should be chosen that are important and relevant to the decision at hand (Mangham et al. (2009) and Johnston et al. (2017)). Conveniently, the aforementioned Cape Town Household Travel Survey asks respondents to rate the importance of a variety of factors in their mode choice decisions; the three factors most frequently rated “most important” in mode choice were comfort, safety, and security. In separate questions asking respondents to rate various aspects of existing minibus service on a four-point scale, “Safety (accidents, maintenance, driver behavior),” “Security from crime,” and “Availability of a seat / crowdedness” are also those most frequently rated “bad.”²⁷

I thus choose three nonpecuniary attributes, or “quality improvements,” corresponding to mode users’ three main concerns: the presence or absence of security guards, driver adherence to speed limits, and whether the minibus loads more passengers than seats. Additionally, I stipulate a travel time and cost (fare) for each minibus alternative. In line with guidance in the literature (Johnston et al. (2017) and Mangham et al. (2009)), I choose attribute levels for the quantitative attributes that are plausible and within the range of typically experienced values in Cape Town yet allow for sufficient variation.²⁸

²⁷Other aspects included timetable adherence, cleanliness, distance to stop, and ease of use.

²⁸Fares can take values of R6, R10, R14, and R18, while travel time is either 20, 30, 40, or 50 min., corresponding to the lengths of typical minibus rides in the morning peak.

D-Efficiency Algorithm I use the Stata package `dcreate` to choose a questionnaire design, namely the combinations of attribute levels in each alternative of each choice set presented to respondents. This *d-efficiency* algorithm minimizes the determinant of the variance-covariance matrix of the estimated parameters of a discrete choice model under some priors (Ben-Akiva et al. (2019) and Rose and Bliemer (2009)). In addition, I specified, in the introductory script, a wait time of 10 minutes, which is constant across all choice sets and alternatives. As the discrete choice model whose statistical power is maximized, I use a version of my model where passengers pay a flow utility cost only while traveling on a minibus, rather than a one-time utility cost.

Coefficient Priors I now require priors for the demand model parameters. I obtain priors $\nu = 12.7$, $r = 0.002$, and $\kappa_M = 0.88$ from estimating a mode choice model on the Cape Town household travel survey stated preference module.²⁹ Then, I use the results of my 1-day stated preference pilot survey (with a very similar format, $N = 20$) to estimate priors for ξ_z , $z \in \{\text{security, no speeding, no overloading}\}$, obtaining $\xi_{\text{security}} = -0.18$, $\xi_{\text{no speeding}} = -0.16$, and $\xi_{\text{no overloading}} = -0.21$.³⁰ I set $\theta_z = -0.1$ for each z since the pilot estimates are noisy and use the median household income per working day from the Cape Town Household Travel Survey to set $\omega_i = 427$.

Questionnaire Dimensions I employ these priors in the *d-efficiency* algorithm to generate 2 “blocks,” or versions, of 5 choice sets with 2 alternatives each.³¹ The 6th choice set, common to both blocks, had a strictly dominant option and thus could provide a measure of comprehension. I do not include an outside option (Ben-Akiva et al. (2019)), as my survey is intended to test relative, rather than absolute, demand for different minibus options; my quantitative model will yield the overall demand for minibus commuting. Furthermore, all attributes have pictograms to aid comprehension in a lower-education context (Mangham et al. (2009)). In addition, I collected demographic information: education, gender, age, personal income, and car ownership. I also collected transport-related information such as current trip purpose, usual commute modes, and frequency of minibus

²⁹I restrict the sample to choice sets that do not contain car as a mode and to respondents aged 25-65 who work outside the home.

³⁰In estimating the multinomial logit on pilot data, I restrict the coefficients on travel time, cost, and the travel-time income interaction to be consistent with the aforementioned two priors and use the midpoints of household income bins from a separate income question.

³¹I create two blocks, which are randomized across respondents, because doing so increased power in Monte Carlo simulations without increasing respondent burden. As for the numbers of choice sets and alternatives, I reduced these from 8 to 5 and 3 to 2, respectively, after the pilot revealed respondent frustration and inattention towards the end of the survey – and a version with 2 rather than 3 alternatives proved less problematic in this regard.

use.

A.2.2 Pilot Survey Lessons

The enumerator team and I conducted a pilot survey at the Cape Town CBD minibus station on 15 June 2022 from approximately 11am to 1pm, where we contacted 36 respondents, 25 of whom qualified for and completed the pilot questionnaire, which had one version (block) with 9 choice sets of 3 alternatives each and a second block with 9 choice sets of 2 alternatives each. Anecdotally, the respondents I interviewed seemed to be taking the scenarios seriously and understanding the aim of the exercise, musing out loud, for example, “I can’t take this bus because it will make me late to work!” However, after the pilot survey, I reduced the number of choice sets and alternatives per choice set to maintain respondent attention.

A.2.3 Sample

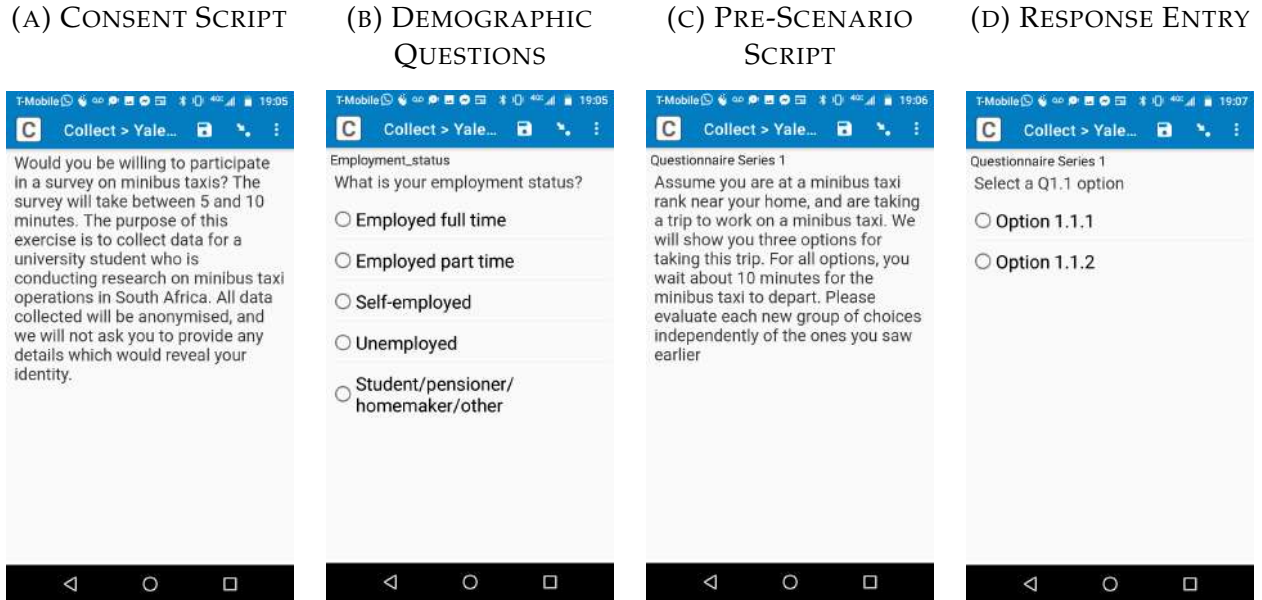
Stated preference surveys were conducted at one mall and transport interchange and at two minibus stations, for 5 weekday hours per location (11am-4pm) on 21, 27, and 30 June 2022. Security considerations did not permit a random sampling of minibus stations or other locations, as many were not deemed safe to approach strangers for this kind of survey. At the Middestad Mall/Bellville transport interchange, enumerators were instructed to conduct surveys inside the mall, at the Golden Arrow formal bus station, and on surrounding streets, but explicitly *not* within the minibus station. On the other hand, at the Khayelitsha Site C and Somerset West Shoprite minibus stations, interviews were conducted only within the station. The aim here was to obtain, at the mall and transport interchange, a representative sample of different mode users as well as, at the minibus stations, a sample of respondents intimately familiar with minibuses, for whom the hypothetical alternatives would be similar to their existing commutes. Only (full-, part-time, or self-) employed respondents were interviewed, so that the scenarios correspond to my quantitative model.³²

A.2.4 Administration and Script

Survey enumerators randomly approached respondents and asked their consent to participate in the survey, according to a script in Online Appendix Figure [A.2a](#). They were

³²Enumerators approached 586 people. Of these, 333 were full-time employed, 97 part-time employed, and 96 self-employed, for a total of 526 respondents who qualified for the survey. Of the remaining people not qualifying, 14 were students/pensioners/homemakers/other and 42 were unemployed.

FIGURE A.2. STATED PREFERENCE SURVEY APP SCREENSHOTS



Notes: These images show screenshots from the Survey CTO app used by enumerators to conduct and record stated preference responses, specifically (A) the consent script; (B) an example of a demographic question; (C) the script that introduces the stated preference choice sets; and (D) the screen used to enter stated preference responses.

offered a chocolate as an incentive. Enumerators then proceeded to read them questions shown in the Survey CTO Android app (see Online Appendix Figures A.2b-A.2d), which automatically progresses through the questionnaire, showing follow up questions or terminating the survey where appropriate. The stated preference scenarios themselves were shown on laminated paper, and enumerators also entered responses directly into the app.

A.2.5 Field Experience

All 526 employed respondents who began the survey also completed all questions. Note that there were no corner solutions: every alternative of every question was chosen by a nonzero number of respondents.

A.2.6 Sample Characteristics

In later estimation, I stack my own stated preference survey with the stated preference module of the 2013 Cape Town Household Travel Survey. Online Appendix Table A.1 compares the demographic characteristics of each sample to the aggregate city commuter population, as measured by the same 2013 survey. Along basic demographic dimensions,

TABLE A.1. STATED PREFERENCE SAMPLE CHARACTERISTICS

Variable	<i>Stated Pref. Samples</i>		<i>Data</i>
	Own	City-Run	Cape Town
Share Auto Owners	0.448	0.581	0.561
Share Female	0.458	0.494	0.458
Share College-Educated	0.295	0.228	0.190
Median Monthly Personal Income [bin]	\$182-\$364	\$182-\$364	\$182-\$364
Median Age	35	39	39
<i>Commute Mode Shares of...</i>			
Minibus	59.56	22.56	23.55
Formal Transit	19.61	27.69	22.81
Auto	12.11	40	39.40
Share Using Minibuses > 1x/Week	0.951	0.635	
<i>N</i>	413	407	

Notes: This table's first two columns display demographic characteristics of my newly-conducted stated preference survey sample as well as the 2013 Cape Town Household Travel Survey stated preference sample. The third lists the corresponding statistics in the aggregate Cape Town population, as inferred from a separate module of the latter survey. In each case, statistics reflect those samples used for estimation, namely respondents between the ages of 25 and 65 who work outside the home.

including gender, education, income, and age, both stated preference samples are representative of the aggregate population. However, respondents in my new sample are less likely to own cars, and, not surprisingly, given that many were recruited at minibus stations, more likely to report that they typically commute by minibus. I later pursue multiple strategies to quantify any bias resulting from this oversampling of minibus users.

A.3 City of Cape Town Household Travel Survey (2013)

The City of Cape Town conducted the 2013 Cape Town Household Travel Survey (CTH-HTS) on a representative sample of residents. In addition to demographics and car ownership, this survey records the addresses of residence and work, which I geocode with the Google Geocoding API ($N = 17,395$), along with the details of respondents' commute. For descriptive statistics and moments, I define the commute mode as follows: minibus includes any commuter who uses minibuses during his or her commute; formal transit includes commuters who use train, bus, or MyCiti bus but not minibuses; auto

includes car and motorcycle drivers and passengers who do not use minibuses or formal transit; and the non-motorized and other category, all others. A subset of respondents also completed a commute stated preference survey with a format similar to my own except that respondents chose among different *modes* of transport: car, formal MyCiti bus, (regular) formal bus, formal train, or minibus (taxi). However, this city-run survey did not include non-pecuniary quality improvements such as station security.

A.4 Minibus On-Board Tracking Data

I make use of GPS-tracked minibus trips, also newly collected for this paper by the South African firm GoMetro. This data, logged by enumerators via smartphone app, covers two trips from the beginning to the end of each route in my station count data and provides stop-level information within each trip. For each stop, I observe the number of passengers boarding and alighting, the arrival and departure time, and the fare paid by passengers boarding. In total, my sample includes $N = 582$ stops, made by 60 vehicles on 43 routes over 2 trips per route.

B. MICRO-FOUNDATIONS OF COMMUTE UTILITY

B.1 Minibus

I now lay out the micro-founded model of minibus commuting which underlies the linear approximations to commute utility U_{ijm}^g used in the main text. Minibus passengers immediately enjoy their home-location-specific amenity θ_i^g , pay a one-time mode-specific utility cost κ_M^g and then wait off-bus to match at the minibus station, so their total (deterministic) utility satisfies $\bar{U}_{ijM}^g = \theta_i^g - \kappa_M^g + u_{ijM}^g$. The value of waiting off-bus, u_{ijM}^g , then follows an HJB equation:

$$ru_{ijM}^g = \lambda_{ij} \left[E_n \left(\tilde{u}_{ij}^g(n) \right) - u_{ijM}^g \right]. \quad (\text{B.1})$$

Because passengers enjoy no flow value of waiting, the right-hand side equals the product of the rate λ_{ij} at which passengers board minibuses and the change in value from boarding a minibus. I assume that commuters cannot observe the mass of passengers already on board, so they never reject a boarding opportunity.

Passengers then wait on-bus for departure. Since passengers board any bus, the change in value from boarding in (B.1) depends on the expectation of the value $\tilde{u}_{ij}^g(n)$ of waiting on-bus, which is itself a function of the passenger mass n already on board because fuller buses depart sooner. The (annuitized) on-bus waiting value equals the product of the rate

ι_{ij} at which the bus fills and the change in value from a higher on-board passenger mass n , $r\tilde{u}_{ij}^g(n) = \iota_{ij}\tilde{u}_{ij}^{g'}(n)$.

Upon bus departure, passengers enter a traveling state until their bus receives the Poisson destination arrival shock at rate d_{ij} . The traveling value V_{ijM}^g thus increases in the destination arrival rate d_{ij} and the skill-specific wage ω_j^g received upon arrival according to $rV_{ijM}^g = d_{ij}(\omega_j^g - V_{ijM}^g)$. Since I assume that minibuses only depart when full and passengers pay fares upon departure, the passenger value at the bus capacity $\bar{\eta}$ satisfies $\tilde{u}_{ij}^g(\bar{\eta}) = V_{ijM}^g - \tau_{ijM}$.

Through repeated value function substitution, I can show that the total deterministic utility \bar{U}_{ijM}^g of skill- g minibus commuters from home i to work j satisfies

$$\bar{U}_{ijM}^g = \theta_i^g - \kappa_M^g + \frac{\lambda_{ij}}{r + \lambda_{ij}} \left[1 - \exp\left(\frac{-r\bar{\eta}}{\iota_{ij}}\right) \right] \frac{\iota_{ij}}{r\bar{\eta}} \left[\frac{d_{ij}\omega_j^g}{r + d_{ij}} - \tau_{ijM} \right]. \quad (\text{B.2})$$

Taking a first-order approximation, around $r = 0$, to this total utility yields

$$\bar{U}_{ijM}^g \approx \underbrace{\theta_i^g - \kappa_M^g - r\omega_j^g \left(\frac{1}{\lambda_{ij}} + \frac{1}{2} \frac{\bar{\eta}}{\iota_{ij}} + \frac{1}{d_{ij}} \right)}_{\equiv U_{ijM}^g} - \tau_{ijM} + \omega_j^g \quad (\text{B.3})$$

where I have suppressed the term $r\tau_{ijM} \left(\frac{1}{\lambda_{ij}} + \frac{1}{2} \frac{\bar{\eta}}{\iota_{ij}} \right)$, which will be close to zero due to the multiplication of a small time preference rate and a small fare. I then define commute utility U_{ijM}^g as the component corresponding to the costs of the actual commute.

B.2 Formal Transit

Workers living in i who choose formal transit to go to work location j directly board a formal transit vehicle which also departs immediately, mirroring the fact that formal transit runs on a schedule to which commuters can adjust to ensure zero wait time. Thus, the total deterministic utility of a formal transit commuter equals $\bar{U}_{ijF}^g = \theta_i^g - \kappa_F^g - \tau_{ijF} + V_{ijF}^g$. Group- g workers enjoy the home location amenity, suffer the formal utility cost κ_F^g , pay the exogenous fare τ_{ijF} , and immediately receive the value V_{ijF}^g of traveling by formal transit from i to j . The traveling value, analogously to the minibus case, reads $rV_{ijF}^g = d_{ijF}(\omega_j^g - V_{ijF}^g)$, where d_{ijF} is the formal-transit-specific rate of arriving into j from i . Combining these formal transit values yields the exact total deterministic utility and then

the corresponding first-order approximation around $r = 0$, as in the minibus case. Again, I define commute utility U_{ijF}^g as the sum of the commute costs:

$$\bar{U}_{ijF}^g = \theta_i^g - \kappa_F^g + \frac{d_{ijF}\omega_j^g}{r + d_{ijF}} - \tau_{ijF} \approx \theta_i^g - \kappa_F^g - \underbrace{r\omega_j^g \left(\frac{1}{d_{ijF}} \right)}_{\equiv U_{ijF}^g} - \tau_{ijF} + \omega_j^g. \quad (\text{B.4})$$

B.3 Car

Commuters choosing car immediately pay the utility and monetary costs of car commuting, κ_A^g and τ_A , and depart, so that the deterministic utility of driving to work is simply $\bar{U}_{ijA}^g = \theta_i^g - \kappa_A^g - \tau_A + V_{ijA}^g$. The corresponding traveling value V_{ijA}^g follows $rV_{ijA}^g = d_{ij} [\omega_j^g - V_{ijA}^g]$; car commuters arrive at their destinations at the same road-based arrival rate d_{ij} as traveling minibus commuters. Again combining value functions, I obtain the exact and linearly-approximated (around $r = 0$) deterministic total car utility, which includes the car commute utility U_{ijA}^g :

$$\bar{U}_{ijA}^g = \theta_i^g - \kappa_A^g + \frac{d_{ij}}{r + d_{ij}}\omega_j^g - \tau_A \approx \theta_i^g - \kappa_A^g - \underbrace{r\omega_j^g \left(\frac{1}{d_{ij}} \right)}_{\equiv U_{ijA}^g} - \tau_A + \omega_j^g. \quad (\text{B.5})$$

C. ESTIMATION

C.1 Robustness

C.1.1 Matching Function

Weighted Estimates I stratified each stage of my sample by the number of trips mapped on a minibus route in a previous 2018 city-commissioned on-board tracking study, which discussions with stakeholders revealed to be a measure of bus traffic on a route. In my model, then, this empirical measure should most closely proxy the number of loading buses b_{ij} that appears on the right-hand side of estimating equation (15). Under common assumptions on error term exogeneity, the unweighted estimates using a sample stratified on an independent variable will be consistent (Wooldridge (2019)). However, one might be concerned that this past-trips-mapped measure might also be correlated with my

outcome, namely the speed at which buses fill up, or loading rate. To account for any such correlation, I reestimate all specifications, in Online Appendix Table C.1, while weighting each observation by the inverse sampling probability of the route with which it is associated. Reassuringly, the estimates, in particular those in Column 3, change little.

Measurement Error Second, I address the concern that my baseline bus loading rate specification (15) might be susceptible to measurement error because the loading rate is in fact a function of an independent variable, namely the number of loading buses b_{ij} . Provided my year-2013 work start time instrument is uncorrelated with year-2022 measurement error in the number of loading buses, the IV specifications will provide consistent estimates. Nevertheless, the fact that the bus loading rate is a function of a right-hand side variable might magnify any bias from such a correlation. I thus re-estimate equivalent specifications for the number of total passengers (“matches”) boarding buses per minute on route ij at time t .³³ Online Appendix Table C.2 shows that the estimated matching elasticities do not greatly change.

C.1.2 Stated Preference

Sample Representativeness I demonstrated in Online Appendix A.2.6 that my new stated preference survey oversamples minibus users. I test for bias resulting from this non-representative sample in two ways. First, I re-estimate the model using the city-conducted survey plus only those respondents in my survey interviewed at the Middestad Mall/Bellville intermodal interchange, a recruitment location less prone to oversampling of minibus riders. Column 2 of Online Appendix Table C.3 shows that the estimated parameters, though somewhat noisier, mirror my full-sample estimates quite closely, in particular the rate of time preference r , Gumbel shape ν , and the high value the high-skill place on minibus station security. Even this “intermodal sample,” however, still oversamples minibus commuters. Thus, in Column 3, I weight my own sample, which, critically, does not contribute to identification of the relative utility costs across modes, by the ratio between the citywide mode share, from the 2013 Household Travel Survey, and the in-sample mode share of a respondent’s self-reported commute mode. Reassuringly, the key takeaways remain.

³³I have only one instrument, and the total matches depend on both passengers and buses—not only on the ratio of the two, even under constant returns. Thus, I cannot estimate IV specifications for $match_{ijt}$.

TABLE C.1. WEIGHTED MATCHING FUNCTION ESTIMATES

Parameter	OLS			IV with $\alpha + \beta = 1$	
	(1) log bus loading rate	(2) log bus loading rate	(3) log bus loading rate	(4) log bus loading rate	(5) log bus loading rate
α	0.703 (0.0148)	0.677 (0.0249)	0.651 (0.0309)	1.000 (0.281)	1.348 (39.88)
β	0.405 (0.049)	0.393 (0.053)	0.403 (0.045)		
95% CI for $\alpha + \beta$	[0.99,1.22]	[0.94,1.19]	[0.94,1.17]		
Route FE	✓	✓	✓	✓	✓
Time FE		✓	✓		✓
Origin-Time FE			✓		
Observations	1,627	1,627	1,607	1,316	1,316
R-Squared	0.600	0.846	0.861	0.45	0.0759
First-Stage F Statistic				6.20	0.00

Notes: Robust standard errors in parentheses, clustered at the origin level. Each specification is weighted by the route-level inverse sampling probabilities yielded by my two-stage stratified cluster design. The unit of analysis is a minibus route (defined by origin and destination, $n = 44$) by five-minute time block in my station count data over the course of the 6-10am morning commute. Columns 1-3 present estimates of (15), with fixed effects included, as noted, for route, time (5-minute clock time block), or origin station (of the route) by time. In Columns 4-5, I assume constant returns to scale, $\beta = 1 - \alpha$, so that I can regress the log bus loading rate on the log ratio of the stock of waiting passengers to the stock of loading buses. I instrument for this ratio of passengers to buses using the log number of commuters living in the mesozone where the route originates who report leaving their home during the 15-minute period including time t , calculated from the 2013 Cape Town Household Travel Survey.

C.2 Externally Calibrated Parameters

C.2.1 Geography

The model geography consists of the $I = 18$ transport analysis zones (TAZ) in Cape Town. Commuters can choose to live in any location. My model is not suited to short-distance commutes, so I do not allow commuters to choose to live and work in the same location.

C.2.2 Commuter Populations N^g and Average Wages

I calibrate commuter populations N^g using the 2013 Cape Town Household Travel Survey (see Online Appendix A.3) and the accompanying sample weights. Since commuters in my model cannot work in their home zone, I exclude those who work in the same transport analysis zone from these commuter populations. I define two skill groups g , high and low,

TABLE C.2. MATCHING FUNCTION ESTIMATES: MATCHES AS OUTCOME

Parameter	OLS		
	(1) log passengers boarding/min.	(2) log passengers boarding/min.	(3) log passengers boarding/min.
α	0.703 (0.0148)	0.677 (0.0249)	0.651 (0.0309)
β	0.405 (0.0490)	0.393 (0.0526)	0.403 (0.0452)
Route FE	✓	✓	✓
Time FE		✓	✓
Origin-Time FE			✓
Observations	1,627	1,627	1,607
R-Squared	0.724	0.902	0.915

Notes: Robust standard errors in parentheses, clustered at the origin level. Each specification is weighted by the route-level inverse sampling probabilities yielded by my two-stage stratified cluster design. The unit of analysis is a minibus route (defined by origin and destination, $n = 44$) by five-minute time block in my station count data over the course of the 6-10am morning commute. Columns 1-3 present estimates derived from a regression of the (log) number of passengers boarding buses on the (log) numbers of waiting passengers and loading buses, with fixed effects included, as noted.

where high-skill includes those with a tertiary degree. For the stated preference estimation and normalization of model wages, I also use the 2013 Cape Town survey to compute average wages by skill group, taking the weighted mean daily per-person household income of workers in a skill group employed outside the home.³⁴

C.2.3 Driving Distance Δ_{ij} and Destination Arrival Rates d_{ij}, d_{ijF}

The driving distance Δ_{ij} that determines minibuses' operating costs equals the driving distance under free-flow (Sunday, 11pm) conditions between centers of employment of transport analysis zones (TAZ) i and j predicted by the Google Maps Distance Matrix API.³⁵ I then calibrate the destination arrival rates, the road-based rate d_{ij} for minibuses

³⁴To calculate daily personal income ω_i in the Cape Town survey, I take the midpoint of the household's income bin, divided by 22.5 (the number of working days in a month) times the number of people in the household. Additionally, I multiply by $\frac{1}{2}$ since I only model a one-way commute. For my own survey, I make similar adjustments, except that I have personal income directly instead of needing to impute it from household income. Finally, I convert all monetary amounts, including also fares, to USD for scaling purposes.

³⁵I calculate the center of employment as the weighted average centroid latitude and longitude of *small area layer* (SAL) units in the TAZ, where the weights are employment by residence in the SAL times the proportion of SAL area within the TAZ. Note that I also use only the part of the SAL unit within the TAZ to calculate centroids. SAL is the smallest census geography.

TABLE C.3. STATED PREFERENCE: ROBUSTNESS TO SAMPLE

Parameter	Skill	(1)	(2)	(3)
		Baseline	Intermodal Sample Only	Commute Mode- Weighted
r		0.001 (0.0004)	0.0014 (0.0007)	0.0011 (.0005)
ν		4.76 (1.26)	6.83 (2.73)	5.84 (1.99)
κ_M	<i>Low</i>	7.68 (1.56)	10.61 (3.54)	9.25 (2.55)
	<i>High</i>	15.03 (3.55)	21.16 (7.82)	18.3 (5.67)
ζ_{security}	<i>Low</i>	-1.09 (0.39)	-2.13 (1.06)	-1.55 (0.69)
	<i>High</i>	-2.75 (0.84)	-4.91 (2.29)	-5.1 (1.86)
$\zeta_{\text{no overloading}}$	<i>Low</i>	-1.38 (0.437)	-2.02 (1.01)	-1.26 (0.596)
	<i>High</i>	-1.39 (0.543)	-1.25 (1.28)	-1.43 (0.83)
$\zeta_{\text{no speeding}}$	<i>Low</i>	-1.36 (0.44)	-3.03 (1.38)	-2.12 (0.85)
	<i>High</i>	-0.825 (0.465)	-1.86 (1.39)	-0.582 (0.73)
κ_F	<i>Low</i>	3.63 (0.51)	4.53 (1.08)	4.14 (0.80)
	<i>High</i>	9.17 (1.89)	12.5 (4.20)	10.96 (3.05)
N Respondents		820	546	820

Notes: Robust standard errors in parentheses. The unit of analysis is an alternative by choice set by individual respondent in either my newly collected minibus stated preference survey (in Cape Town, estimates reflect $N = 489$ unique individuals) or a stated preference module of the 2013 Cape Town Household Travel Survey ($N = 646$ unique individuals). The estimated parameters are derived from the coefficients in a multinomial logit model with choice probabilities given by (16). Column 1 displays the baseline estimates, as in Table 3; Column 2 estimates the model on only the 2013 city-run survey respondents plus the respondents in my survey interviewed at the Middestad Mall/Bellville transport interchange (i.e. excluding those sampled at minibus stations); Column 3 estimates the model on the full sample but weights the respondents in my survey by the aggregate citywide share of their reported commute mode divided by that mode’s share among respondents to my survey.

and cars and the formal transit rate d_{ijF} , using stylized transport networks specific to each mode. The nodes of each mode’s network consist of the 18 TAZ with links connecting each pair of TAZ which share a border; the Microsoft Azure API provides mode-specific

travel times along each link.³⁶ For a given mode m between a given origin-destination pair ij , I calculate the minimum travel time through the mode-specific network, and, consistent with the properties of Poisson arrival processes, the road-based and formal transit destination arrival rates d_{ij} and d_{ijF} equal the inverse of the respective travel time.

C.2.4 Formal Transit Fares τ_{ijF} and Car Commute Cost τ_A

Formal transit fares for a given route ij are calculated using the Cape Town MyCiti bus rapid transit distance-based fare **scheme**, where I again make the calculation using the straight-line distance between TAZ centroids. I calculate the car monetary commute cost τ_A using an “average [monthly] total mobility cost ” calculated by WesBank of South Africa.³⁷

C.2.5 Minibus Operations Parameters δ_0, δ_1, χ

I set the minibus shift length δ_0 and inverse “number of trips” δ_1 to generate reasonable numbers of expected trips in half of a typical work shift. I use half of a typical 8-hour work day as minibus drivers’ time budget because my model depicts only a one-way commute. $\delta_0 = 240$ ensures that minibus drivers whose expected total trip time $\bar{\eta}/\iota_{ij} + 1/d_{ij}$ equals 240, i.e. consumes their entire work shift, complete one trip and start a second, in expectation, where I assume that drivers can complete any trip they start. In turn, I assume that, even if total trip time equals zero, minibuses lose 20 minutes per trip for repositioning or driver breaks that are, of course, outside my model. As a result, under zero trip time, a minibus should be able to run 12 trips and start a 13th, on average, in the half-day shift that I model. Setting $\delta_1 = 0.0104$ ensures that the expected number of trips under $\bar{\eta}/\iota_{ij} + 1/d_{ij} = 0$ approximately equals 13. I calibrate the minibus per-kilometer operating cost χ with figures provided by the firm GoMetro.³⁸

³⁶I calibrate the driving time \bar{t}_{ikA} along each link ik by querying the Microsoft Azure API (similar to Google Maps) for a trip between the centers of employed population of the two TAZ on a Wednesday at 7am. From the Azure API, I also obtain formal transit travel time \bar{t}_{ikF} for each link in the transit network as the average trip time over 6 evenly-spaced trips on a Wednesday between 7:00 and 8:00am via formal transit, including Metrorail commuter trains, Golden Arrow private scheduled buses, and MyCiti bus rapid transit.

³⁷The monthly **average total mobility cost** equals ZAR 9,356.80; I divide this figure by the number of working days per month (22.5), multiply by the share of all car trips made *to* work, 0.21, in the 2013 Cape Town survey travel diary, and convert to USD.

³⁸I multiply the Toyota Quantum minibus’s litres of diesel used per kilometer, 0.099, by the June 2022 diesel per-litre price in South Africa, ZAR22.63, convert to USD as with other prices in my model, and adjust by the ratio between my calibrated $\bar{\eta}$ and the actual bus capacity, 15, to set operating costs in proportion to revenue from the calibrated fares.

C.2.6 Emissions Parameters χ_m^e and ζ

I obtain mode-specific carbon-equivalent emissions from calculations in Borck (2019) combined with U.S. Department of Energy estimates. Specifically, I take Borck (2019)'s estimate of $\chi_A^e = 0.554$ kg CO₂-equivalent emissions per kilometer from driving, a figure which includes actual CO₂ emissions as well as "local pollutant" emissions. To obtain emissions from other modes, I use relative passenger-mile-per-gallon (pmpg) estimates from my partner firm GoMetro and the Alternative Fuels Data Center (U.S. Department of Energy (2023)).³⁹ For the social cost of carbon, I follow Borck (2019)'s benchmark of $\zeta = \$0.0485$ per kilogram CO₂.

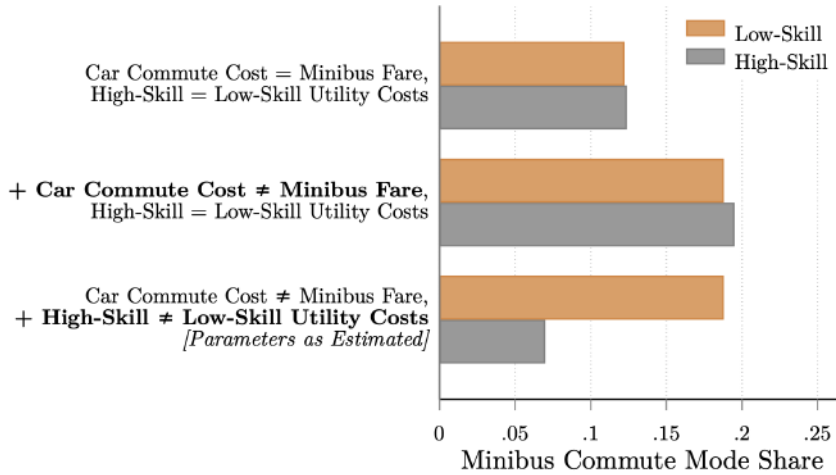
D. VALIDATION AND RESULTS

D.1 Decomposing Mode Choices

The low-skill minibus mode share in Cape Town exceeds that of the high-skill by 21 percentage points in the data. This stark skill differential could stem from the the high monetary cost of cars or the starkly higher minibus utility costs of the high-skill. Online Appendix Figure D.1 demonstrates that utility costs account for the difference in minibus mode shares between skill groups, as follows. I begin with a model with a car cost equal to the average minibus fare, $\tau_A = \bar{\tau}_M$, and equal utility costs across groups, $\kappa_m^h = \kappa_m^l$. The top line in Online Appendix Figure D.1 plots the resulting similar minibus mode shares across skill groups. When I increase the car cost to its calibrated value, both skill groups shift towards minibuses, as displayed in the second line. Only when commuters experience the estimated utility costs of each mode, in the third line, do the high-skill abandon minibuses. Policies that reduce these costs, such as providing security, might thus offer substantial welfare gains.

³⁹Fuel efficiency estimates provided by GoMetro suggest that the most common minibus vehicle, the gasoline-powered Toyota Quantum, requires 0.143 liters/km, equivalent to 248.05 passenger miles per gallon with a full load of 15 passengers. The AFDC estimates an average of 27.5pmpg for single-occupancy cars, so minibuses have 0.11 the energy use of cars per passenger-distance. Applying this factor to χ_A^e yields $\chi_M^e = 0.0615$. The AFDC estimates 137.2pmpg for high-ridership bus and 600pmpg for high-ridership train. Taking a weighted average of these using the 53% share of formal transit commuters in the 2013 Cape Town Household Travel Survey who use trains at some point during their commutes, I obtain a formal transit average of 382.48pmpg. Thus, formal transit has 0.07 the fuel use of cars, yielding $\chi_F^e = 0.0388$.

FIGURE D.1. SKILL DIFFERENTIAL IN MINIBUS USE: MONETARY COSTS OR UTILITY COSTS?



Notes: This figure displays model-predicted skill-group-level minibus mode shares, under different parameter assumptions that shut down alternative explanations for the skill differential in mode shares. In the first line, I set the car commute cost equal to the mean minibus fare, $\tau_A = \bar{\tau}_M$, and set high-skill equal to low-skill utility costs for each mode, $\kappa_m^h = \kappa_m^l$. In the second, I reintroduce the true car commute cost τ_A and, in the third, different utility costs across skills, $\kappa_m^h \neq \kappa_m^l$.

D.2 Mode Shares

First, in Online Appendix Table D.1, I show that, for both minibus and car, the model-predicted mode shares by origin-destination pair and skill are significantly positively correlated with their empirical counterparts. Second, in the cross-section across origin-destination tuples, the model replicates the negative (positive) signs of the relationships between average minibus (car) mode shares and income (Online Appendix Figure D.2).

D.3 Stated Preference Approach

Stated preference data is only as useful as respondents' ability to predict their future non-hypothetical choices. I have already demonstrated in Section VI and Online Appendix D.1 that the model matches aggregate mode shares and does so primarily due to the utility costs estimated from stated preferences; to further confirm respondents' ability to predict their future non-hypothetical choices, I pursue two strategies.

First, I employ the model to predict the mode choices of the respondents in the combined stated preference sample. In particular, respondents' reported personal income and education, as well as the median model-calibrated wait times, travel times, and fares for each mode, weighted by origin-destination commute flows, imply choice probabilities for each commute mode when inserted into (8)-(9). Online Appendix Figure D.3 demonstrates that

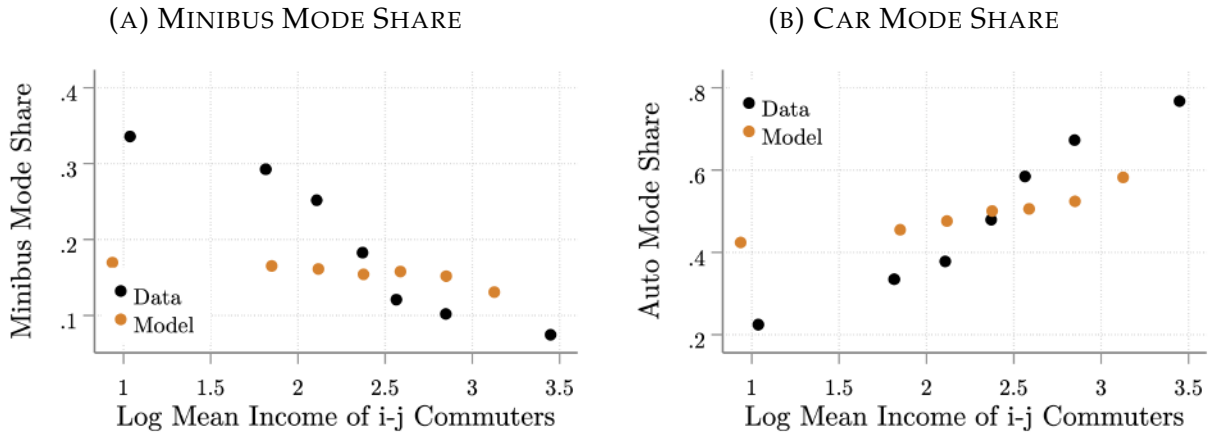
TABLE D.1. MODE CHOICE PROBABILITIES, DATA VS. MODEL

Variables	Minibus	Car
	Mode Share, Data	Mode Share, Data
Mode Share, Model	1.209 (0.153)	0.992 (0.0814)
Constant	-0.00558 (0.0208)	0.0335 (0.0493)
Observations	507	507
R-squared	0.106	0.230

Robust standard errors in parentheses

Notes: This table presents the results of OLS regressions of empirical mode shares on those predicted by the model at the skill group by origin by destination transport analysis zone level. In the data, I calculate skill-specific origin-destination transport analysis zone (TAZ) commute mode shares from the 2013 Cape Town Household Travel Survey, taking shares of respondents working outside the home who commute by each mode. In the model, I take the share of commuters with a given skill and chosen home and work location (TAZ) who use a given mode.

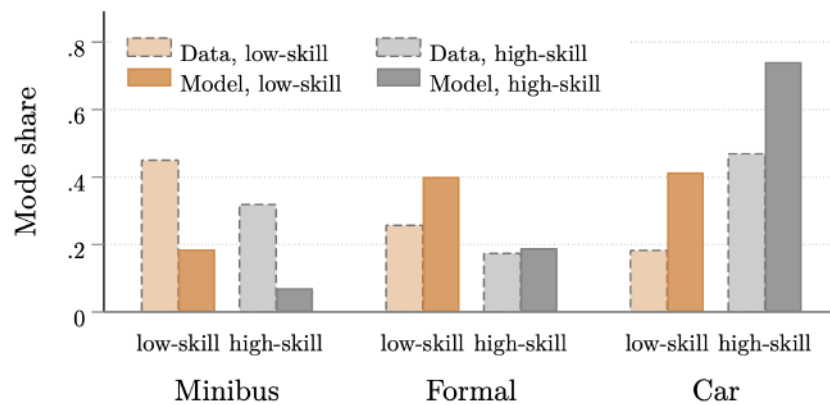
FIGURE D.2. ORIGIN-DESTINATION MODE SHARES VERSUS INCOME



Notes: This figure’s binned scatterplots display the relationship between mean income and mode shares, where the unit of analysis is an origin by destination transport analysis zone tuple in Cape Town. In the 2013 Cape Town Household Travel Survey data, for each origin-destination, I compute average household income per-person and (half) workday as well as the shares of respondents working outside the home who commute by each mode. In the model, I take a weighted average of the group-level mode shares and income over skill groups using commuter inflows.

my model, as estimated using respondents’ choices, replicates the broad patterns in their actual mode choices. However, the model cannot fully explain the over-representation of

FIGURE D.3. MODE SHARES, REPORTED VERSUS MODEL-PREDICTED IN COMBINED STATED PREFERENCE SAMPLE



Notes: This figure displays commute mode shares for the combined stated preference sample (from my own survey plus the 2013 Cape Town Household Travel Survey stated preference module), comparing respondent-reported (actual) commute modes with the mode shares predicted by the estimated model. In the data, minibus includes all who use minibuses at some point during their commutes. In the model, I take an average of the corresponding predicted choice probabilities π_{ijm}^s across respondents.

minibus commuters in the stated preference sample noted in Online Appendix [A.2.6](#).

Second, I allow for heterogeneous effects by a series of demographic characteristics and find plausible heterogeneity in preferences. I take equation (16) and interact a dummy variable for the binary demographic characteristics listed in Column 1 of Online Appendix Table [D.2](#) with the terms in the multinomial logit model that identify utility costs, their dependence on policy, and the value of time. Women and college workers have a higher value of time saved, even conditional on income; the former result echoes Borghorst et al. (2021), who find that women’s marginal cost of commuting increases after the birth of children. That college-educated workers value their time more highly, even conditional on income, might similarly reflect a higher value of home production. Surprisingly, women place a lower value on security; perhaps men are more likely to be involved in gang activities that would put them at risk. Older workers place a higher value on security and especially on driver adherence to speed limits, suggesting an intuitive greater risk aversion.

D.4 Additional Policy Counterfactuals

I now explore three additional minibus-targeted policy counterfactuals, detailed in Online Appendix Table [D.3](#). First, I simulate the effects of mobile minibus-booking apps or station

TABLE D.2. STATED PREFERENCE HETEROGENEITY: DIFFERENCE IN PARAMETER ESTIMATE, VERSUS BASE CATEGORY

Dimension	r	<i>Mode Utility Cost</i>		<i>Effects on Minibus Utility Cost</i>		
		κ_M	κ_F	$ \zeta_{\text{overload}} $	$ \zeta_{\text{security}} $	$ \zeta_{\text{speed}} $
Female	0.0013 (0.0006)	-3.61 (1.06)	-3.27 (0.924)	-0.222 (0.419)	-1.33 (0.535)	-0.49 (0.436)
College	0.0019 (0.0007)	6.66 (1.94)	4.62 (1.28)	0.052 (0.481)	1.71 (0.659)	-0.458 (0.499)
Age>45	0.0027 (0.001)	-1.03 (0.709)	-1.80 (0.671)	0.494 (0.640)	1.72 (0.770)	2.50 (0.906)

Notes: Robust standard errors in parentheses. Each cell gives the coefficient on the interaction of a dummy variable for the demographic characteristic listed in Column 1 with the parameter at the top of each column in a multinomial logit equivalent to (16). I estimate all interaction effects in each row in one specification across alternatives, choice sets, and individuals in my own and the 2013 Cape Town H.H. Travel Survey stated pref. modules.

expansions as a doubling of matching efficiency μ to the 90th percentile of separate route-level estimates. The benefits pale in comparison to the gains from the social optimum or security, even though I do not account for any costs of this improvement. Second, I explore a reorganization of the minibus loading process such that only one bus loads at a time. Buses enter and wait in an orderly queue; their total wait time from entry until departure remains as in my baseline model. I then make the optimistic assumption that the full inflow of commuters, $\sum_g N^g \pi_{ijM}^g$, can instantly board the loading bus, so expected total minibus passenger wait time is $\frac{1}{2} \frac{\bar{\eta}}{\sum_g N^g \pi_{ijM}^g}$. Importantly, I do not account for the potentially high costs of infrastructure improvements to facilitate such efficient loading. The before-cost benefits, however, are on the same order of magnitude as the gains from the social optimum. Third, I simulate the construction of exclusive minibus lanes which allow minibuses to enjoy free-flow travel times on the top 10% of road network links, by model-predicted minibus traffic. However, the construction costs estimated by De Beer and Venter (2021) outweigh the associated travel time gains.

D.5 Robustness: Road Congestion

Finally, I consider how road congestion might alter the welfare gains from my policies. The travel time over an individual link in the road network, $t_{ik} \equiv \bar{t}_{ik} v_{ik}^\gamma$, now depends on a calibrated intercept \bar{t}_{ik} and increases with the vehicle inflow v_{ik} , raised to a road congestion elasticity γ . For tractability, I require that minibuses and cars follow the sequence of

TABLE D.3. ADDITIONAL COUNTERFACTUAL URBAN TRANSPORTATION POLICIES

Policy	Skill:	Change in Mode Share				% Change in...				
		Minibus		Car		Mean Wage		Emissions	Welfare	
		Low	High	Low	High	Low	High		Low	High
Matching Efficiency		0.002	0.002	-0.001	-0.002	0.04	0.01	-0.2	0.1	0.1
Minibus Lanes		0.002	0.001	-0.001	-0.001	0.06	0.01	-0.2	-0.1	-0.1
Loading Reorganization		0.004	0.004	-0.002	-0.003	0.08	0.01	-0.4	0.3	0.1

Notes: This table displays the results of three counterfactuals: increasing matching efficiency to the 90th percentile of route-level matching efficiencies; building exclusive minibus lanes on the top 10% of road network segments by model-predicted minibus traffic; and reorganization of the loading process such that only one minibus loads at a time. The first four columns show the changes in the minibus and car mode shares by skill group. The final five show the percent change in group-level average wages ω_j^s , total emissions, and group-level welfare, the latter measured as equivalent variation. Only the lanes counterfactual accounts for the associated costs.

TABLE D.4. COUNTERFACTUAL URBAN TRANSPORTATION POLICIES WITH ROAD CONGESTION

Policy	Skill:	% Change in Welfare	
		Low	High
Bus Rapid Transit		-3.9	-1.7
Social Planner		0.4	0.2
Minibus Security		2.3	1.2

Notes: This table displays the percent change in welfare (equivalent variation) by skill group from three counterfactuals: construction of the existing MyCiti formal bus rapid transit line, implementation of the social optimum via optimal minibus fares τ_{ijM}^* and mode-specific commuter taxes t_{ijm}^{g*} and adding security to all minibus stations. Unlike in the main text, I account for the road congestion effects of minibus and car traffic on the road-based destination arrival rates of minibuses and cars.

links that minimizes free-flow travel time but calculate d_{ij} as the inverse of the sum of congestion-affected travel times t_{ik} over links traversed. In separate results available upon request, I employ road-segment-level data for Cape Town from TomTom’s Traffic Stats API to estimate $\hat{\gamma} = 0.0917$, closely tracking the baseline instrumental variables estimate of the equivalent log-linear effect of traffic volume on urban-segment travel time in Allen and Arkolakis (2022). Minibuses make up only a small share of overall traffic, so the gains from each policy remain almost unchanged (Online Appendix Table D.4). Nevertheless, the high-skill benefit slightly more from the social optimum due to associated reductions in (car) traffic.