Pontifícia Universidade Católica
do Rio de Janeiro

**Igor Rigolon La-Cava Veiga**

# Optimal Bus Subsidies: Evidence from Rio de Janeiro

**Dissertação de Mestrado**

Masters dissertation presented to the Programa de Pós–graduação em Economia, do Departamento de Economia da PUC-Rio in partial fulfillment of the requirements for the degree of Mestre em Economia.

Advisor: Prof. Leonardo Rezende

Rio de Janeiro
March 2025

Pontifícia Universidade Católica
do Rio de Janeiro

**Igor Rigolon La-Cava Veiga**

# Optimal Bus Subsidies: Evidence from Rio de Janeiro

Masters dissertation presented to the Programa de Pós–graduação em Economia da PUC-Rio in partial fulfillment of the requirements for the degree of Mestre em Economia. Approved by the Examination Committee:

**Prof. Leonardo Rezende**
Advisor
Departamento de Economia – PUC-Rio

**Prof. Lucas Lima**
Departamento de Economia – PUC-Rio

**Dra. Maína Celidonio de Campos**
Prefeitura da Cidade do Rio de Janeiro

Rio de Janeiro, March 27th, 2025

**Igor Rigolon La-Cava Veiga**

B.A. in Economics, Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), 2022.

Bibliographic data

Veiga, Igor Rigolon La-Cava

Optimal Bus Subsidies: Evidence from Rio de Janeiro / Igor Rigolon La-Cava Veiga; advisor: Leonardo Rezende. – 2025.

54 f: il. color. ; 30 cm

Dissertação (mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Economia, 2025.

Inclui bibliografia

1. Economia – Teses. 2. Transporte público. 3. Subsídio. 4. Regulação. 5. Tarifa Ótima. 6. Brasil. I. Rezende, Leonardo. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Economia. III. Título.

CDD: 004

# Acknowledgments

Thank you to my family for all the support, even when I tried to act my most nonchalant. Thank you to my father, Márvio, for keeping me grounded and stopping me from getting lost in the work. To my mother, Isa, thank you for being there and getting me to talk about things in life other than this dissertation.

I would like to thank my advisor, Leo, for encouraging me to pursue my ideas, as rough around the edges as they were. Our troubleshooting sessions were immensely helpful in fitting the puzzle pieces of this dissertation together. Thank you to Lucas and Maína for being in the committee and for all the constructive criticism. To Lucas, thank you for sifting through this text, catching my mistakes, and suggesting interesting additional exercises. To Maína, thank you for bringing a policy outlook and informing me of how this regulation works in practice.

I would like to thank Maína again, and all the workers at SMTR, for implementing data-driven policies to improve the lives of commuters in Rio de Janeiro. And thank you for the pleasant byproduct of collecting and publishing this data, which allowed this dissertation to exist.

I wouldn't have come this far without my master's friends pushing each other forward. I learned a lot from our endless debates over coursework. And maybe not so much from our pizza nights and poker games, but they were fun.

This study was financed in part by FAPERJ and by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

## Abstract

Veiga, Igor Rigolon La-Cava; Rezende, Leonardo (Advisor). **Optimal Bus Subsidies: Evidence from Rio de Janeiro**. Rio de Janeiro, 2025. 54p. Dissertação de Mestrado – Departamento de Economia, Pontifícia Universidade Católica do Rio de Janeiro.

In 2022, the city government of Rio de Janeiro began paying subsidies to the private companies that operate bus services. These subsidies are conditional on achieving daily goals and are paid as a flat fee per kilometer, regardless of the number of passengers. I use data from GPS devices installed on all buses and estimate that subsidies led to a considerable decrease in passenger wait times. To simulate counterfactuals with alternative policies, I build a structural model of how bus operators react to the subsidy rules set by the government. I compare the welfare implications of different regulatory and pricing schemes in the model. Results suggest that conditioning the payment of subsidies on a daily kilometer goal produces stronger incentives for bus companies to increase service levels than traditional subsidies.

## Keywords

Public Transit; Subsidies; Regulation; Optimal Fare; Brazil.

# Resumo

Veiga, Igor Rigolon La-Cava; Rezende, Leonardo. **Subsídio Ótimo aos Ônibus: Evidência do Rio de Janeiro**. Rio de Janeiro, 2025. 54p. Dissertação de Mestrado – Departamento de Economia, Pontifícia Universidade Católica do Rio de Janeiro.

Em 2022, a prefeitura do Rio de Janeiro passou a subsidiar as empresas que operam as linhas de ônibus na cidade. Esses subsídios são condicionais ao cumprimento de metas diárias, e são pagos em uma taxa fixa por quilômetro, independentemente do número de passageiros. Eu uso dados de dispositivos de GPS instalados nos ônibus e estimo que a introdução do subsídio reduziu, consideravelmente, o tempo de espera de passageiros por ônibus. Para simular contrafactuais com políticas alternativas, eu construo um modelo estrutural de como as operadoras de ônibus reagem às regras de subsídio estabelecidas pela prefeitura. Eu comparo as implicações de bem-estar de diferentes estruturas de regulação e preço. Os resultados sugerem que condicionar o pagamento de subsídios ao cumprimento de uma quilometragem diária produz incentivos mais fortes para que as empresas elevem a circulação de ônibus do que com subsídios tradicionais.



# Palavras-chave

Transporte público; Subsídio; Regulação; Tarifa Ótima; Brasil.

# Table of contents

## List of figures

## List of tables

## List of Abbreviations

SMTR – Secretaria Municipal de Transportes do Rio de Janeiro

GTFS – General Transit Feed Specification

OLS – Ordinary Least Squares

IV – Instrumental Variables

NLLS – Nonlinear Least Squares

GMM – Generalized Method of Moments

FOC – First-Order Condition

# 1
# Introduction

Bus fares are heavily subsidized in major cities around the world, and increasingly so in Latin America (RIVAS; BRICHETTI; SEREBRISKY, 2020). In Rio de Janeiro, public buses are the largest mode of public transit, transporting close to 2 million passengers each day. In response to the decline in bus frequency during the COVID-19 pandemic, the city government started to subsidize bus operators in 2022. They introduced a novel system of subsidies aimed at incentivizing bus companies to provide higher levels of service. It has two key features: (i) bus lines must reach a daily kilometer goal in order to be eligible for subsidies, and (ii) subsidies are not paid per passenger, but rather per kilometer traveled.

In this paper, I simulate the welfare gains created by this policy and compare them to counterfactuals with alternative regulation schemes. I model a profit-maximizing bus operator whose revenue is impacted by the subsidy rules, which, in turn, changes the number of buses it chooses to send into circulation. The revenue collected from passengers is a function of the demand for each bus line, which depends on the fare price – chosen by the city government – and on bus availability. If buses are scarce, wait times increase and fewer passengers will choose to take them.

I estimate how the demand for buses responds to price changes by leveraging a control group of non-paying passengers. Passengers such as senior citizens, toddlers, public school students, and people with disabilities are guaranteed by law to ride the bus for free. As a result, their demand should follow similar trends as paying passengers, while being unaffected by price changes. By comparing how the demand of the two groups changes around events of fare price adjustments, I estimate the price-elasticity of demand. I find that poorer passengers are less elastic to price changes than wealthier ones, indicating that they have fewer alternatives to substitute when the price increases.

I show that the introduction of the subsidies led to a decrease in the average wait times experienced by passengers, as it incentivized bus companies to increase bus circulation. Because of this, I treat this policy as an exogenous shifter of wait times during a window of time around its introduction, and exploit this variation to estimate how demand reacts to changes in wait time. This elasticity is essential to inform how the supply of buses is determined: as bus frequency increases, wait time falls and demand rises, until the profit-maximizing number of buses is reached.

I use this demand equation to build a model of how bus companies choose the supply of bus trips in response to the subsidy rules. I estimate cost parameters for this model using non-linear least squares (NLLS), and use it to simulate counterfactuals for alternate policy designs. I simulate what would have happened if, instead of implementing these kilometer-based subsidies, the city government had opted for (i) leaving the system unchanged, (ii) price hikes instead of subsidizing the companies, or (iii) traditional subsidization of each passenger's fare.

Results show that the subsidy system adopted by the city government has far superior welfare outcomes to increasing fare prices for passengers, and overall produces better outcomes for passengers than the alternatives. It is also less costly than subsidies per-passenger, and allows for a more even operation of bus lines, whereby companies are incentivized to not abandon bus lines with fewer passengers.

There is a growing literature in urban economics and in the economics of public transport studying optimal prices for public transit systems and discussing the need for government subsidization. Mohring (1972) lays out a theoretical groundwork that justifies public transit subsidies, on the basis of returns to scale. Bus demand drops when service is low, which creates a negative feedback loop leading the market provision of public transit to be below the social optimum.

Nelson et al. (2007) and Parry and Small (2009) incorporate the positive externalities generated by bus services when they substitute for cars, including reductions in traffic congestion, pollution, and accidents. Tsivanidis (2022) estimates the effect of the Bus Rapid Transit (BRT) system in Bogotá. He finds that a large portion of welfare gains was not due to time savings, but due to general equilibrium effects and interactions with the labor market.

In Rio de Janeiro, Campos (2019) measures the impact of the public transit expansion in preparation for the 2014 Football World Cup and the 2016 Olympics, which included BRT and rail. She found that effects were widely heterogeneous across worker's skill levels. Almagro et al. (2024) finds that optimal public transit prices for Chicago are much lower than the current ones.

I also draw from the literature in Industrial Organization and Regulation in building the model and conducting counterfactual exercises. Lewis and Bajari (2014) study how contractors respond to deadlines set by the government procuring their services. Gagnepain and Ivaldi (2002) evaluate different regulation schemes for public bus services in France.

I contribute to this literature by directly evaluating the impact of implementing bus subsidies, one of the most prevalent ways the government can support a privately-operated public transit system. I show that not just the money spent, but the design of the policy, can have large impacts on bus availability. By modeling how bus operators respond to subsidies and their incentives, I am able to simulate the outcomes caused by each policy and compare them.

The rest of the paper is structured as follows. In chapter 2 , I outline the institutional background behind the public bus system in Rio de Janeiro, and how it shaped the design of the subsidies policy. In chapter 3, I provide descriptive evidence that the introduction of the subsidies affected the decisions of bus companies, and that it contributed to increased bus frequency. The demand for bus services is estimated in chapter 4, which I then use to build the model describing the behavior of bus companies in chapter 5. Results of the model are showcased in chapter 6, which also compares outcomes across alternate policies. Finally, chapter 7 concludes.

# 2
# Institutional background and data

## 2.1
## Institutional background

Municipal public bus routes in Rio de Janeiro are operated by private companies, organized into four consortia. In 2010, these consortia took part in a bidding process, and each became responsible for a set of bus lines, shown in Figure 2.1. A major challenge in managing this system is to ensure all of these bus services are sufficiently supplied. As most of the revenue of bus companies originates from passengers directly paying fares, they are incentivized to allocate more buses towards their most passenger-dense bus lines, potentially neglecting lines with fewer passengers.

It may even be the case that some bus lines are not privately profitable for the bus companies, but serve a purpose in interconnecting the city and allowing residents of distant locations to commute to work. Because of this, some bus services may be abandoned by bus operators, which is not in the public interest. In 2022, the problem of insufficient bus circulation in Rio de

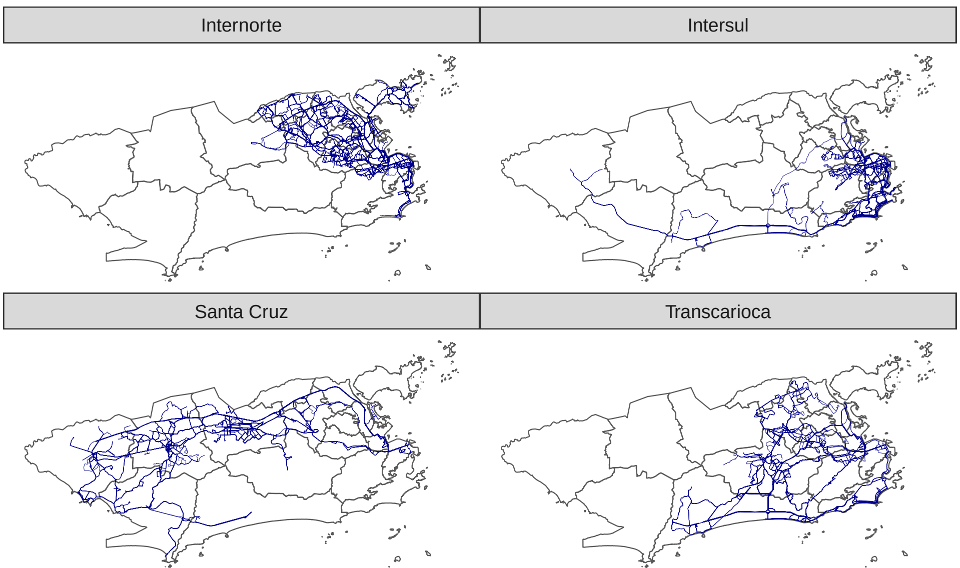


Figure 2.1: Bus services operated by each consortium

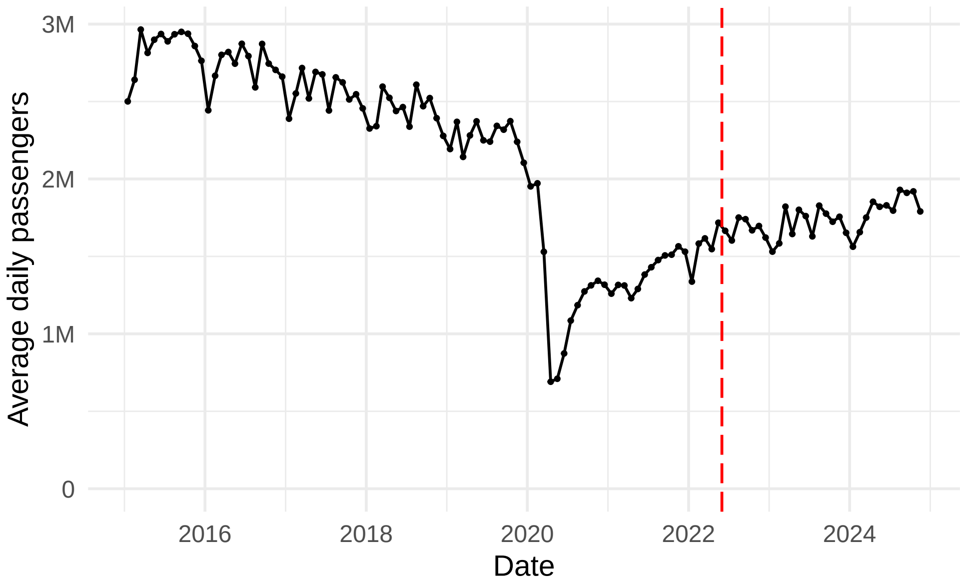


Figure 2.2: Average daily passengers. Red vertical line marks the introduction of the subsidies.

Janeiro reached a tipping point, leading to a renegotiation of the terms of the contract between city government and the bus consortia which culminated in the introduction of subsidies.

The first reason fewer buses began to operate in Rio de Janeiro was the decline in passengers. Figure 2.2 shows that the number of bus passengers declined sharply in early 2020 due to the COVID-19 pandemic, and has not, until 2024, returned to its previous levels. During the time period covered by the graph, there has also been an increase in the use of ride-hailing applications and the implementation of Bus Rapid Transit (BRT) in the city, which operates as a separate system to regular bus lines. Even before the pandemic, these alternatives drove a share of passengers away from regular buses. The decline in passengers cost the consortia a large share of their revenue, making the operation of many bus lines less sustainable.

The second reason was that the fare price lagged behind the costs faced by bus companies. The fare price, shown in Figure 2.3, is set by the city government and is the same across all municipal regular buses. Between early 2019 and 2022, bus fares were not readjusted, which compounded the pressure placed on the bus companies. According to a yearly report published by the

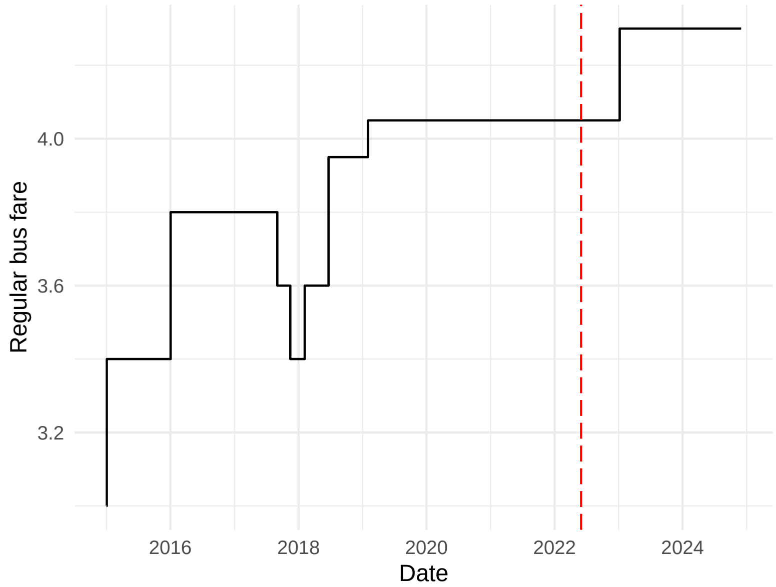


Figure 2.3: Regular bus fares, in R$ (Brazilian Reais). Red vertical line marks the introduction of the subsidies.

city's department of transportation, which takes into account cost components such as the price of fuel and auto parts, the bus fare should have increased 30% by the end of 2021, and another 10% by March of 2022 (Prefeitura da Cidade do Rio de Janeiro, Secretaria Municipal de Transportes, 2022b).

The outcome was that a large share of bus services ceased to operate at a high enough level, or ceased to operate entirely. The city's department of transportation, in May of 2022, deemed 58% of all bus lines to be inoperative, as they did not achieve minimum fleet requirements (Prefeitura da Cidade do Rio de Janeiro, Secretaria Municipal de Transportes, 2022a). Subsidies were a solution meant to increase the circulation of buses without increasing fare prices paid by passengers.

## 2.2 Subsidy rules

The subsidy system designed by the city's department of transportation was aimed at fixing the problem of inoperative bus lines, by complementing the revenue of bus companies enough to make their operation viable. The traditional approach would be to directly subsidize the fares of each passenger, whereby passengers pay a discounted price, and the government pays the difference directly to the bus companies. However, this would reinforce the incentives for bus companies to focus on bus lines in passenger-dense areas and disregard those with fewer passengers.

Instead, subsidies are tied to the number of kilometers traveled, and not to the number of passengers. The city government began to track all municipal buses by GPS, and calculated a flat rate to pay for each valid kilometer, detected to be within the bus' planned route. Because this revenue is independent of passengers, even the bus lines with less demand become more attractive to the bus companies and run a reduced risk of becoming inoperative.

To further ensure that each bus line provides enough supply, subsidies are only paid when a bus line achieves 80% of a set number of kilometers planned for it for the day by the city's department of transportation. Combining these two factors, the subsidies owed to a bus line in a given day are zero for bus lines that fall below the daily goal, and are a flat rate per kilometer for those that achieve the daily goal.

The rate paid per kilometer was readjusted over time, starting at R$2.13 (Brazilian Reais) and surpassing R$4 in 2024. This led subsidies to account for around a quarter of the total revenue of bus consortia. The subsidies owed to all bus lines operated by a consortium are added up and paid jointly, on a biweekly basis.

## 2.3
## Data sources

I access data published by Rio de Janeiro's department of transportation (SMTR) in the city's datalake (Prefeitura do Rio de Janeiro, 2025). The main dataset I employ is the Daily Operations Report (RDO), which compiles the number of passengers transported, revenue, number of trips, and kilometers traveled daily per bus line. This dataset ranges between 2015 and 2024. I only use data for regular bus services, and not Bus Rapid Transit (BRT).

I combine this with the data used by the SMTR to verify GPS data and calculate the amount of subsidies owed. I use it to gather information on the daily kilometer goals set for each bus line, and the validated percentage of the goal that was effectively reached. I also use a dataset containing all bus trips tracked by GPS, containing departure and arrival times, in order to measure the hours of operation of each bus line.

There are also auxiliary datasets in General Transit Feed Specification (GTFS) format, which is a standardized system used to store public transit schedules and routes. I use those to extract geographic data for the itineraries of each bus service. Since 2024, the SMTR has been publishing data on uses of their new bus smart card service (Jaé), which contains the time, location, and bus line of each card swipe. I combine this data with the 2010 Brazilian Census to obtain demographic information about passengers in different bus lines.

# 3
# Descriptive findings

Since subsidies were implemented, they have become a large portion of the revenue of the bus consortia. Table 3.1 contains descriptive statistics for bus lines since 2022, comparing the months before the subsidy, created in June, to the time period up until 2024. While the fare has been readjusted from R$4.05 to R$4.30, subsidies pay, on average, an additional R$1.1 per passenger. There has been a large increase in the number of planned bus lines, and a decrease in the proportion of inactive bus lines: those that are planned but do not actually operate.

Figure 3.1 shows the total number of bus trips operated each day, averaged out by month. After the sharp decline at the start of the COVID-19 pandemic, bus activity remained low and only increased again after the subsidies were put in place, marked by the red dashed vertical line.

To see how the subsidy system may have been responsible for this increase in bus circulation, Figure 3.2 shows the percentage of the daily kilometer goal

Table 3.1: Daily summary statistics for bus lines

| | Before subsidies | After subsidies |
|---|---|---|
| Number of bus lines | 253 | 336 |
| Share of inactive bus lines | 11% | 6% |
| Passengers | 6,931 | 5,623 |
| Number of bus trips | 60.6 | 58.9 |
| Kilometers traveled | 2,661 | 2,499 |
| Fleet | 11.6 | 10.3 |
| Median wait time | 6.2 | 6.1 |
| Fare price | 4.05 | 4.24 |
| Subsidies per passenger | 0 | 1.1 |
| Revenue from passengers | 18,051 | 14,491 |
| Revenue from subsidies | 0 | 6215 |

Data between 2022-2024. Monetary values in R$.

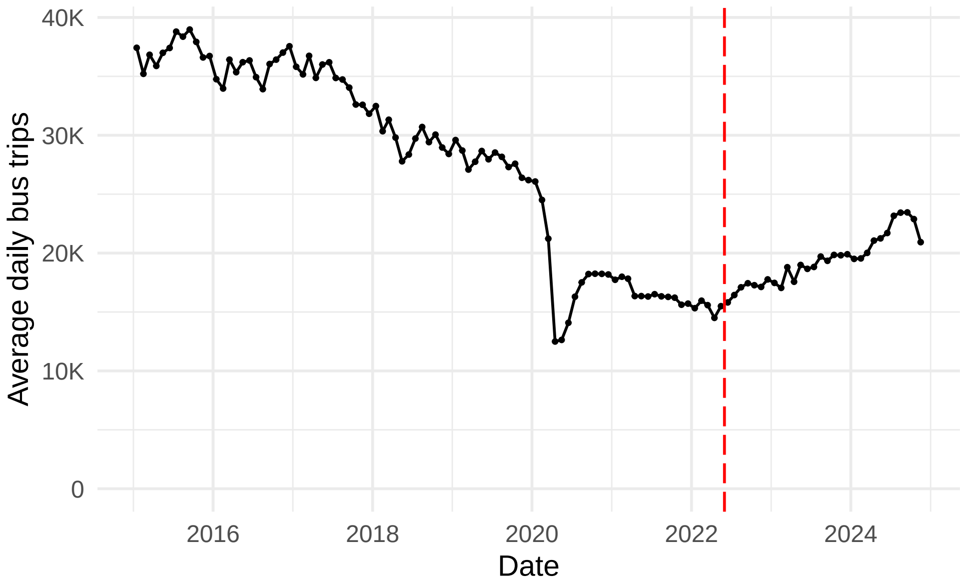


Figure 3.1: Daily bus trips

achieved by bus lines, before and after the subsidies. After the subsidies, each bus line must reach 80% of the daily goal, marked by the red dashed vertical line, to be eligible for subsidies. The histograms show a distinct pattern of bunching above the 80% mark since the system was implemented. A large mass of bus lines which sat consistently far below this threshold was nudged towards the right in order to earn subsidies. During the initial phases of the subsidy system there was a mass of inactive bus lines, traveling zero kilometers in some days, but this did not continue in the following months.

Another source of larger bus availability is at the extensive margin. While existing bus lines were incentivized to operate at a higher frequency by the subsidy system, there has also been a stark increase in the number of bus lines in circulation. Figure 3.3 plots the average number of active bus lines – those with at least one recorded trip – compared to the number of planned bus lines. Since the subsidies were put in place, many new bus services were created. Following an adjustment period of around six months after the subsidies were implemented, there was also a decline in the share of inactive bus lines.

Although I do not directly observe inactive bus lines in the pre-GPS data, I identify them by analyzing the documents that determined the bus

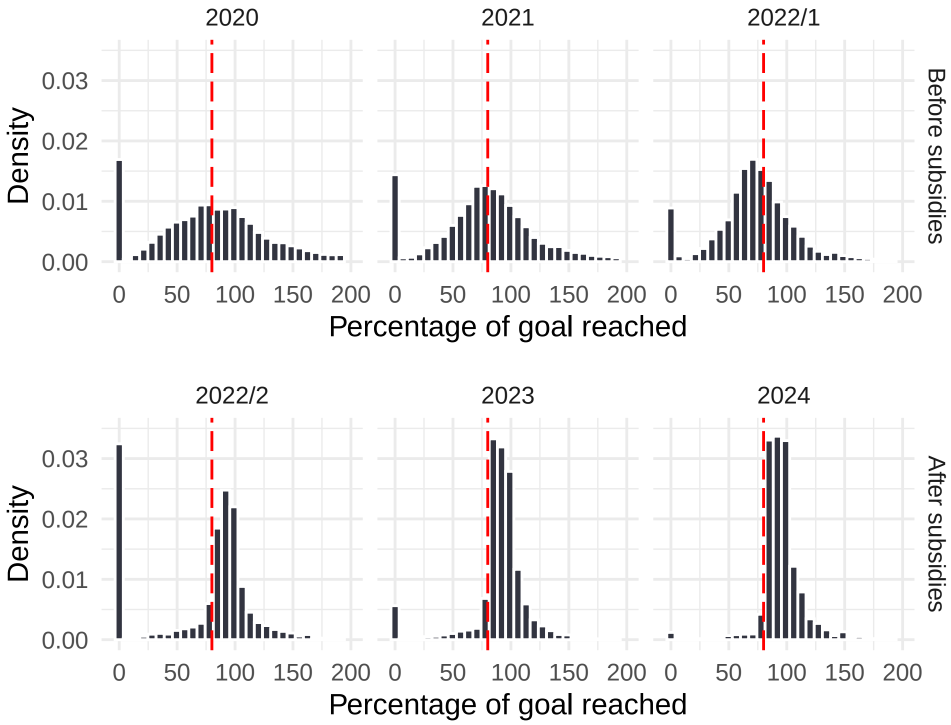


Figure 3.2: Distribution of goal achievement

The top panel is calculated according to the average goals of each bus line after the subsidies, as the goals did not exist beforehand.

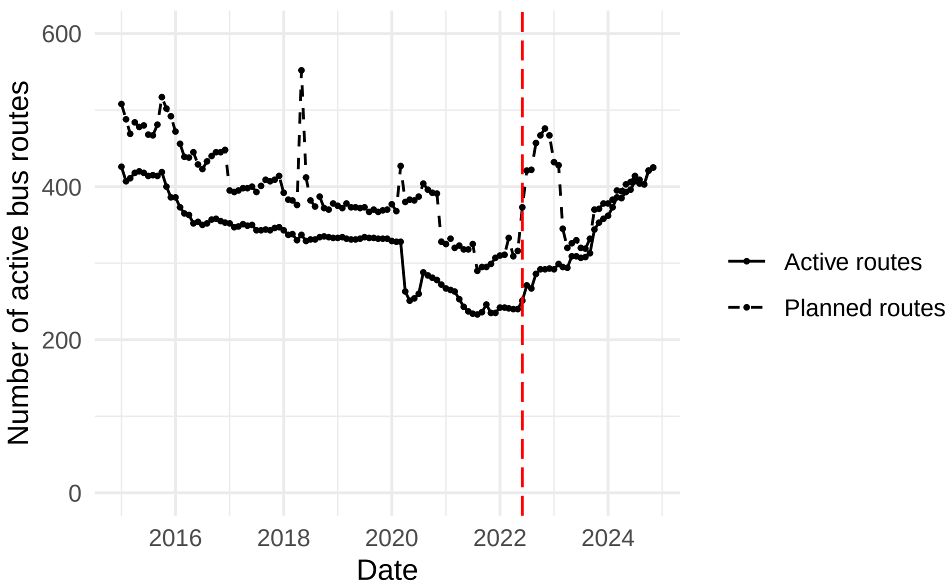


Figure 3.3: Active and total planned bus routes

fleet, called service orders ("ordens de serviço"), before the subsidies were created. From these documents, I know which bus lines were planned during each time period, and can impute a value of zero bus trips when a line does not operate. I account for bus lines which may not operate during weekends and holidays. With the new system in place since the subsidies, I directly observe bus lines with zero kilometers tracked by GPS, which should provide a much more accurate picture of the level of inactivity among bus lines. I expect the actual proportion of inactive bus lines before the subsidies to have been higher than I could infer from the data.

Figure 3.4 shows the average wait time experienced by passengers in each area of the city, before and after the subsidies, from 2022 onwards. I calculate the average wait time for each bus line daily[1] and use geolocated bus card swipe data to measure how many passengers take each bus service in each area. I then take a median of wait times across bus lines passing through each tile, weighted by the number of passengers who take each bus in that location during morning peak hours. Wait times have overall slightly decreased since the introduction of the subsidies, especially in the west and north regions of Rio, which had more tiles with very high wait times. As these are median wait times, they do not account for the benefits of the creation of new bus lines, or of reintroducing inoperant ones.

Around the time of the creation of the subsidies, many bus lines in the city were restructured, abolished, or created, which makes it hard to extract the causal effect of the subsidies from reduced-form estimates. This is why it is important to use a structural model that explicitly takes into account how the design of the subsidy system affects incentives faced by bus operators. Table 3.2 takes a naive approach and assumes all changes in wait times

[1] I follow Mohring (1972) in assuming wait times for each bus line are inversely proportional to the number of bus trips. The explicit formula for the average wait time for bus service $s$ on day $t$ is given by

$$\text{wait time}_{st} = \frac{\text{average \# of hours of operation}_s}{2 \times \text{\# of bus trips}_{st}}.$$

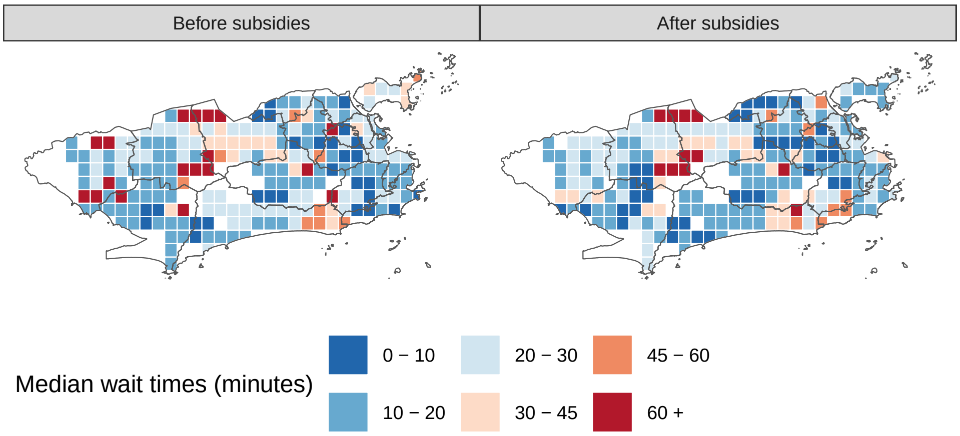


Figure 3.4: Wait times

around this window of time were due to the subsidies, using data from 2022 onwards. I restrict the sample to short time windows around the introduction the subsidies, in order to better approximate the short-term effect of the policy.

During the window of 1 month before and after its introduction, average wait times decreased by over one minute. The second and fourth columns also show that these reductions in wait time were mostly concentrated among the bus lines with the fewest average passengers per kilometer – the least profitable ones for bus operators. As seen in Figure 3.2, there was seemingly a learning period in 2022 while bus operators were still adapting to the new rules. This may lead these specifications to underestimate the true effect of the subsidies over time.

To gauge the effect of the policy in the long term, Table 3.3 regresses the wait time faced by passengers on the Subsidies/km rate, which was readjusted over time since the creation of the subsidies. Because this rate ranged between R$2.13 and R$4.04, the short-term effect of the subsidies should sit between a 0.7 to 1.2 minute reduction in wait time, and the long-term effect, considering the current rate, should sit between a 1.3 to 2.2 reduction in wait time, on average.

Table 3.2: Effect of subsidies on wait times

| | Wait time (minutes) | | | |
|---|---|---|---|---|
| | 3-month window | | 1-month window | |
| | (1) | (2) | (3) | (4) |
| Post subsidies | -0.462** | | -1.189*** | |
| | (0.166) | | (0.261) | |
| Post subsidies × 1st Pass/Km tercile | | -1.358** | | -2.112** |
| | | (0.414) | | (0.772) |
| Post subsidies × 2nd Pass/Km tercile | | -0.229 | | -0.957** |
| | | (0.202) | | (0.305) |
| Post subsidies × 3rd Pass/Km tercile | | 0.070 | | -0.642*** |
| | | (0.134) | | (0.171) |
| Median wait time | 11.1 | 11.1 | 10.7 | 10.7 |
| Weekday and holiday FEs | ✓ | ✓ | ✓ | ✓ |
| Bus Service FEs | ✓ | ✓ | ✓ | ✓ |
| Num.Obs. | 38616 | 38616 | 14085 | 14085 |
| R2 | 0.748 | 0.748 | 0.787 | 0.787 |

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.3: Effect of subsidy rate on wait times

| | Wait time (minutes) | |
|---|---|---|
| | (1) | (2) |
| Subsidies/km | -0.553*** | -0.336** |
| | (0.079) | (0.117) |
| Median wait time | 11.4 | 11.4 |
| Weekday and holiday FEs | ✓ | ✓ |
| Bus Service FEs | ✓ | ✓ |
| Year-Quarter FEs | × | ✓ |
| Num.Obs. | 70326 | 70326 |
| R2 | 0.768 | 0.768 |

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

# 4
# Demand estimation

Revenue from passenger fares is by far the most important component of the profit garnered by a bus line. Because of that, for a model of how decisions are made by bus operators to be accurate, I must first measure the behavior of demand and how it responds to changes in price and bus availability. In this chapter, I leverage exogenous variations in price and wait times to estimate the demand for buses, which I employ in the structural model of chapter 5.

## 4.1
## Price elasticity of demand

I exploit the price variations shown in Figure 2.3 to estimate how passengers respond to price changes. If the supply of buses did not also respond to these changes, a direct regression of passengers on fare prices could be used to estimate the price-elasticity of demand – directly assessing which percentage of passengers stop taking the bus following some increase in fares. However, similarly to the traditional simultaneity bias in supply and demand systems, this estimate would be biased. When fare prices are raised, bus companies begin to earn more for each passenger, and hence increase the frequency of buses. In turn, this lowers wait times and makes buses a more attractive mode of transportation. This induces a positive correlation between fare prices and passengers, creating a downward bias in the magnitude of OLS estimates of the price-elasticity of demand.

To identify causal effects, I use a group of non-paying passengers as a control. Senior citizens, toddlers, public school students, and people with disabilities are legally allowed to ride the bus for free. For them to serve as a control, the underlying assumption is that non-paying passengers have parallel trends to paying passengers, with the only difference being that they do not respond to price changes. Although senior citizens and students undergo

Table 4.1: Price elasticity of demand

| | Log passengers - Log non-paying passengers (Bandwidths around price changes) | | | |
|---|---|---|---|---|
| | Full Sample | | 4 weeks | |
| | (1) | (2) | (3) | (4) |
| Log Fare Price | -0.658*** | -0.343*** | -0.681*** | -0.550*** |
| | (0.006) | (0.008) | (0.013) | (0.019) |
| Log Fare Price × Log Income | | -0.230*** | | -0.273*** |
| | | (0.018) | | (0.040) |
| Weekday and holiday FEs | ✓ | ✓ | ✓ | ✓ |
| Bus Service FEs | ✓ | ✓ | ✓ | ✓ |
| Num.Obs. | 1118756 | 403447 | 161689 | 58899 |
| R2 | 0.545 | 0.463 | 0.578 | 0.532 |

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

different patterns of seasonality throughout the year when compared to paying passengers, those changes are not correlated to price changes, and will hence not bias the estimates.

I estimate the equation

$$\log \text{passengers}_{st} - \log \text{non-paying passengers}_{st} = \alpha_s + \alpha_t - \phi_p \log(\text{price}_t) + \varepsilon_{st}$$

for bus service $s$ and day $t$, including fixed effects for days of the week, holidays, and bus service. I remove variations in price that are too close together – those that are within 4 weeks of another price change. Results are shown in Table 4.1. Columns (1) and (2) include the entire sample, while columns (3) and (4) restrict the sample to only days within 4 weeks of a price change, to extract a clearer short-term effect. The price-elasticity of demand is around $\widehat{\phi_p} = 0.65$ in most specifications.

In columns (2) and (4), I also add an interaction term between the fare price and income. The Log Income variable is normalized to mean zero as to keep the price coefficient as the average elasticity. To approximate the income of passengers in each bus line, I use geolocated data on bus card swipes to pin down where passengers who take bus line $s$ during morning peak hours get on the bus, which I use as a proxy for where they live. I combine these locations with income data from the 2010 Census, for each census tract, to measure the average income of each bus line. This process was, however, only successful for less than half of the observations in the full dataset.

I find that low-income passengers are less price-elastic than high-income ones. Miller and Savage (2017) explain that this may be the case due to low-income passengers having fewer alternatives to public buses, whereas high-income passengers may choose to take a taxi or private car. But there are cases in which low-income passengers are more price-elastic than high-income ones, for not being able to afford higher prices. This result can have large welfare implications, as it means that price hikes are disproportionately more damaging to low-income passengers than to high-income ones, as they have fewer possibilities to substitute away from buses.

Table 4.2 shows the results of direct regressions of passengers on fare price, without using a control group. The price elasticity estimates are much larger in magnitude, which goes against the effect of an increase in supply leading to improvements in the service for passengers. This could be a reflection of overarching time trends of reductions in bus passengers since 2020, or could be linked to the particular timing of price changes. If price increases happen more often when bus operators are struggling, they will be correlated to a low service quality, inflating the negative correlation. These trends are, however, followed by both paying and non-paying customers, and therefore should not bias the main specifications.

Table 4.2: Price elasticity of demand

| | Log passengers | | | |
|---|---|---|---|---|
| | (Bandwidths around price changes) | | | |
| | Full Sample | | 4 weeks | |
| | (1) | (2) | (3) | (4) |
| Log Fare Price | -4.020*** | -4.020*** | -2.351*** | -2.456*** |
| | (0.009) | (0.015) | (0.022) | (0.036) |
| Log Fare Price × Log Income | | 0.178*** | | 0.335*** |
| | | (0.031) | | (0.075) |
| Weekday and holiday FEs | ✓ | ✓ | ✓ | ✓ |
| Bus Service FEs | ✓ | ✓ | ✓ | ✓ |
| Num.Obs. | 1129574 | 405372 | 163469 | 59337 |
| R2 | 0.803 | 0.792 | 0.852 | 0.838 |

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

### 4.2 Wait time elasticity of demand

I now estimate $\phi_w$, the coefficient measuring how much demand declines for each additional minute of wait time faced by the passengers. A direct OLS estimate for this coefficient suffers from simultaneity bias due to the supply increasing in response to demand shocks. If bus companies expect there to be more passengers, they will increase the supply of buses, decreasing wait times and inducing a negative correlation between wait times and passengers.

To identify causal effects, I use the exogenous decrease in wait time caused by the creation of the subsidies. By restricting the sample to narrow windows of time around the introduction of the policy, I expect most variation in wait times to be exogenous to demand, stemming from supply shifts. Results are shown in Table 4.3. The value of the coefficient indicates that, for each additional minute of wait time, 0.6% of passengers stop taking that bus line. Here, time fixed effects include fixed effects for each week, for each day of the week, and for holidays.

As a robustness check, Table 4.4 displays the same regressions run substituting non-paying passengers for total passengers as the dependent

Table 4.3: Wait-time elasticity of demand

| | Log passengers | | |
|---|---|---|---|
| | Full sample | 3-month window | 1-month window |
| Wait Time | -0.0075*** | -0.0061*** | -0.0052*** |
| | (0.0000) | (0.0004) | (0.0004) |
| Time FEs | ✓ | ✓ | ✓ |
| Bus Service FEs | ✓ | ✓ | ✓ |
| Num.Obs. | 1129574 | 43100 | 14085 |
| R2 | 0.870 | 0.947 | 0.958 |

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4.4: Wait-time elasticity of demand for non-paying passengers

| | Log non-paying passengers | | |
|---|---|---|---|
| | Full sample | 3-month window | 1-month window |
| Wait Time | -0.0082*** | -0.0062*** | -0.0056*** |
| | (0.0001) | (0.0005) | (0.0007) |
| Time FEs | ✓ | ✓ | ✓ |
| Bus Service FEs | ✓ | ✓ | ✓ |
| Num.Obs. | 1118756 | 43011 | 14044 |
| R2 | 0.837 | 0.923 | 0.941 |

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

variable. Under the parallel trends assumption of section 4.1, non-paying passengers should respond equally to changes in wait time as paying passengers do, despite not being affected by prices. This does seem to be the case, as the wait time coefficients remain very similar.

In terms of welfare implications, it is important to identify whether the passengers who stop taking a bus due to increases in wait time are able to choose another competing bus line, or must take another – potentially more expensive – mode of transportation. If passengers are able to substitute into another similar bus line, the welfare damage of higher wait times is mitigated. In Appendix A, I conduct two robustness exercises that suggest that this is not the case, and that the estimated value for $\phi_w$ is not driven by these substitution effects.

# 5
# Model

In this chapter, I build a model of how bus companies determine their supply in response to subsidy rules. The basis of the model is that bus companies know the behavior of demand – as estimated in the previous section – and freely choose the number of bus trips to send out each day in each bus line in order to maximize profits. I estimate parameters by matching the number of bus trips observed in the data to the number of bus trips simulated by the model. Once the model is estimated, I conduct a series of policy counterfactuals and evaluate their impact.

## 5.1
## Demand for buses

As in the previous section, I assume the demand for buses is of the form

$$\log \text{Pass}_{st} = \alpha_s + \alpha_t - \phi_p \ \log(\text{price})_t - \phi_w \text{wait time}_{st} + \varepsilon_{st}.$$

I plug in the estimated elasticity with respect to price of $\widehat{\phi}_p = 0.65$ and $\widehat{\phi}_w = 0.006$ to estimate the regression

$$\log \widehat{\text{Pass}}_{st} + \widehat{\phi}_p \ \log(p_t) + \widehat{\phi}_w \text{wait}_{st} = \underbrace{\widehat{\alpha}_s + \widehat{\alpha}_t}_{\text{bus line and time FEs}},$$

which I use to predict a counterfactual demand as a function of prices and wait times. Bus companies choose a number $n$ of bus trips, which determines how long wait times will be. The resulting demand, $\widetilde{\text{Pass}}_{st}(n)$, is calculated as

$$\log \widetilde{\text{Pass}}_{st}(n) = \widehat{\alpha}_s + \widehat{\alpha}_t - \widehat{\phi}_p \ \log(\widetilde{p}_t) - \widehat{\phi}_w \widetilde{\text{wait time}}_{st}(n) + \varepsilon_{st},$$

where wait times are inversely proportional to $n$. By rearranging and exponentiating, the demand function takes the form

$$\widetilde{\text{Pass}}_{st}(n) = \exp\left(\widehat{\alpha}_s + \widehat{\alpha}_t - \widehat{\phi}_p \log(\widetilde{p}_t) - \widehat{\phi}_w \widetilde{\text{wait time}}_{st}(n) + \varepsilon_{st}\right).$$

Under the assumption that bus companies do not observe demand shocks $\varepsilon_{st}$ but know them to follow a $\mathcal{N}(0, \sigma_\varepsilon^2)$ distribution, the number of daily passengers expected by bus service $s$ at day $t$ is

$$\mathbb{E}\left[\widetilde{\text{Pass}}_{st}(n)\right] = M_{st} \exp\left(-\widehat{\phi}_w \widetilde{\text{wait time}}_{st}(n)\right),$$

where $M_{st}$ represents the number of potential passengers who would take the bus if wait times were zero[1], and wait times are

$$\widetilde{\text{wait time}}_{st}(n) = \frac{\text{average \# of hours of operation}_s}{2 \times n}.$$

Figure 5.1 shows the shape of this function, with the number of bus trips in the $x$-axis. If $n \to 0$, the wait time $\to \infty$ and there are no passengers. Meanwhile, if $n \to \infty$, the wait time $\to 0$ and all potential passengers $M_{st}$ take the bus.

1

$$\mathbb{E}\left[\widetilde{\text{Pass}}_{st}(n)\right] = \exp\left(\widehat{\alpha}_s + \widehat{\alpha}_t - \widehat{\phi}_p \log(\widetilde{p}_t)\right) \exp\left(-\widehat{\phi}_w \widetilde{\text{wait time}}_{st}(n)\right) \mathbb{E}\left[e^{\varepsilon_{st}}\right].$$

From the moment-generating function of the normal distribution, $\mathbb{E}\left[e^{\varepsilon_{st}}\right] = e^{\sigma^2/2}$. I then define

$$M_{st} = \exp\left(\widehat{\alpha}_s + \widehat{\alpha}_t - \widehat{\phi}_p \log(\widetilde{p}_t) + \frac{\widehat{\sigma}^2}{2}\right).$$

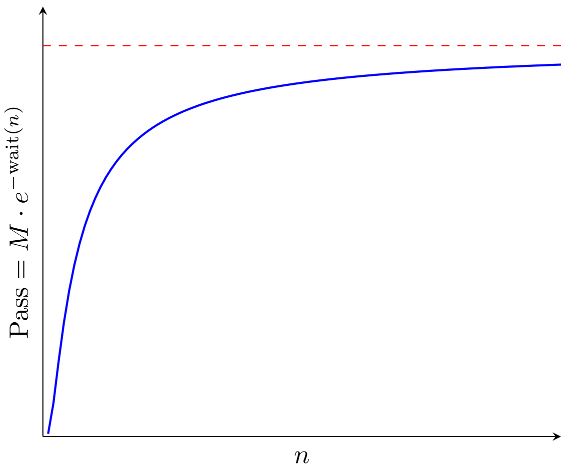


Figure 5.1: Demand function

## 5.2 Profit of a bus service

Each bus line $s$ at day $t$ chooses the number $n$ of bus trips to send out each day, for an expected profit of

$$\pi_{st}(n) = \lambda_s \, p_t \times \mathbb{E}\left[\widetilde{\text{Pass}}_{st}(n)\right] + \text{Subs}_{st}(n) - C_{st}(n),$$

where $p_t$ is the fare price, $\lambda_{st}$ is the average proportion of paying passengers[2], $\text{Subs}_{st}(n)$ follows the subsidies rule set by the city government, and $C_{st}(n)$ is the cost function.

The subsidies are given by

$$\text{Subs}_{st}(n) = \mathbb{I}[n \geq \overline{n}_{st}] \times L_s \times n \times \text{subskm}_t,$$

with $\overline{n}$ being the 80% of the daily goal threshold required by the government, $L_s$ being the number of kilometers per trip for route $s$, and subskm being how much the government pays in subsidies for each kilometer. To calculate the subsidies, I take into account the fact that only roughly 90% of the kilometers self-reported by bus companies in the Daily Operations Report are counted

[2] I take $\lambda_{st}$ to be the ratio between the revenue a bus line would obtain from all passengers (fare price $\times$ passengers) and its total reported revenue

for the subsidies, when compared to the GPS data used by city government to compute subsidies – which only spans from when the subsidies were introduced. This is because some parts of the trajectory buses take are discarded, such as when they move to and from their garage, or when they are within a few hundred meters of the start and end points.

There are also fines charged by the city government when a bus line operates below 60% of its daily goal (R$563.28), and a harsher fine when it operates below 40% (R$1126.55). These fines only took effect in January of 2023. Although they are small in magnitude, I also include these discontinuities in the model. In the following months, the city government also imposed a ceiling of 120% for most bus lines. Kilometers traveled above 120% of the goal would no longer qualify for additional subsidies. I take this ceiling as given for the entire period, as to avoid cases where infinite subsidies are paid.

While the daily kilometer goals are not fixed, and may be readjusted in negotiations between the city and bus companies, I treat them as exogenous to the behavior of bus companies. I assume bus companies do not employ dynamic strategies to manipulate the goals, such as choosing a below-optimum number of bus trips for a period of time in order to induce the city government to reduce the goal.

I assume the cost function to be linear with respect to diesel consumption and the number of trips, given by

$$C_{st}(n) = \delta_1 \times \text{diesel price}_t \times L_s \times n + \delta_2 \times n + q \times \overline{\text{fleet}}_{st} \times \mathbb{I}[n > 0],$$

along with a quasi-fixed cost of $q$ per bus, which can be interpreted as the cost of bringing each bus of the fleet from storage into circulation. It is only paid when the bus lines operates in a given day. I calibrate $q = \text{R\$}160$ for the model to match the overall proportion of inactive bus lines observed in the data. I do not take into account the fact that the number of trips is bound by the number of buses available in the fleet, as that is, in itself, another choice variable, which is implicitly tied to the choice of $n$. Although each consortium controls many bus lines, I assume that each choice of optimal bus trips takes place independently, rather than being a joint maximization of the sum of profits.

## 5.3 Estimation

Under this specification, the profit function $\pi_{st}(n)$ has the shape of Figure 5.2. Given parameters $\delta = (\delta_1, \delta_2)$, this allows me to compute the optimal number of trips chosen by the bus companies, $n^*(\delta)$. It is either a local maximum with $\pi'_{st}(n) = 0$, which can be above or below $\overline{n}_{st}$, or a corner solution with $n = 0$ or at the discontinuity, $n = \overline{n}_{st}$. The other candidates for optimization are the notches at 60% and 80% of the daily goal, where fines may be charged, and the kink at 120% of the daily goal, where marginal subsidies become zero.

By taking the first order condition of profit maximization and manipulating it, I am able to analytically solve for $n^*(\delta)$ using the Lambert W function, which I show in Appendix B. This means that given parameters $\delta$, it is computationally cheap to calculate all possible maxima and compare them. Hence, $n^*(\delta)$ is simply the one which yields the highest profits out of the candidates for maximum.

I then estimate the vector of parameters $\delta$ by NLLS, by numerically

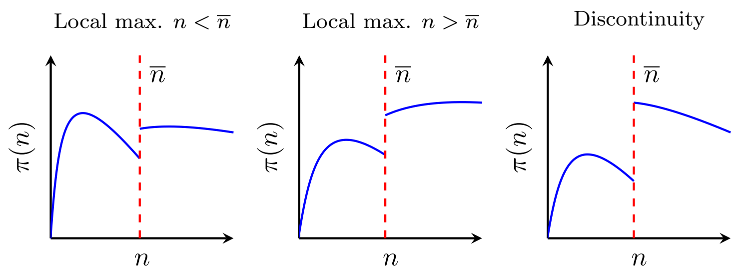


Figure 5.2: Profit function

solving the problem

$$\min_{\delta} \sum_{i} \Big( n_i - n_i^*(\delta) \Big)^2 .$$

This means that the estimated parameters $\widehat{\delta}$ minimize the distance between the vector of actual observed numbers of trips and the numbers of trips predicted by the model. Davidson, MacKinnon et al. (1993) show that, assuming the model is correctly specified as

$$n_i = n^*(\delta) + u_i,$$

where $u_i$ are independent with mean 0 and a diagonal covariance matrix $\Omega$, the NLLS estimator is, asymptotically,

$$\widehat{\delta} \sim \mathcal{N}\Big(\delta, (J'J)^{-1}(J'\Omega\, J)\,(J'J)^{-1}\Big),$$

where $\delta$ is the true parameter value, and $J$ is the jacobian matrix of the $n^* = (n_1^*, \dots, n_N^*)$ function with respect to $\delta$. This allows me to numerically compute heteroskedasticity-robust standard errors with an $\mathrm{HC}_0$ estimator, by choosing $\widehat{\Omega} = \mathrm{diag}(\widehat{u}^2)$. In the homoskedastic case where $\Omega = \sigma^2 I$, the variance of $\widehat{\delta}$ simplifies to $\sigma^2\,(J'J)^{-1}$.

# 6
# Results and counterfactuals

In this chapter, I display the estimates for the structural model and evaluate how closely it can reproduce the behavior of bus companies. I then use the model to conduct a series of counterfactual exercises with different policy designs. I compare bus availability, wait times, cost, and welfare under the system adopted by the city government, and under alternative subsidy systems. Then, I decompose the two main features of the subsidy system – the goal requirement and paying a flat rate per kilometer – in order to measure which one is most important for its effectiveness. Finally, I summarize the effects of the current subsidy system and discuss potential policy implications.

## 6.1
## Estimates and model fit

The NLLS estimates for the cost parameters $\delta$ are shown in Table 6.1, along with robust standard errors in parentheses. The value for $\delta_1$ can be interpreted as the number of liters of diesel consumed per kilometer, while $\delta_2$ is the excess cost per trip, regardless of distance. It can be interpreted as a fixed cost of sending another bus into circulation. I estimate the model for both the full time period (2015-2024) and for the time before the subsidies were created (before June 2022) to show that $\widehat{\delta}$ is stable. At an average diesel price of roughly R$5 per liter, the cost per kilometer is of around R$0.19, which is much smaller than would be expected considering the fuel expenditure of buses, and is also small compared to revenues. The additional cost per trip imposed by $\widehat{\delta}_2$ is very small in magnitude, as a single additional paying passenger per trip is enough to compensate for it.

In Figure 6.1, I add the total number of bus trips done each month, and compare it to the number of bus trips predicted by the model for each month. It shows that, despite having a lot of variation at the bus line level, the model

Table 6.1: NLLS Estimates

| | Full sample | Before subsidies |
|---|---|---|
| $\delta_1$ | | |
| Diesel L/km | 0.038 | 0.036 |
| | (0.0001) | (0.0002) |
| $\delta_2$ | | |
| R$/trip | 1.655 | 1.755 |
| | (0.0207) | (0.0223) |
| Obs. | 1,122,106 | 877,028 |
| R-squared | 0.52 | 0.48 |
| RMSE | 49 | 52 |
| MAE | 32 | 33 |

accurately matches the overall level of bus activity before the introduction of the subsidies, marked by the red dashed vertical line. On the other hand, the model overestimates the response of bus operators to the subsidies, leading the predictions after June 2022 to be consistently above the observed data.

Additionally, Figure 6.2 compares the distribution of the percentage of the daily goal reached in the data, in the first column, to the distribution predicted by the model, in the second column. The first row restricts the sample to 2022 before the subsidies, while the second row portrays the time period after the subsidies. It reveals that the model is able to reproduce the overall shape of the distribution, both before and after the subsidies. However, the bottom row shows that the model predicts that bus lines should bunch at the 120% of goal threshold much more often than at the 80% one.

Table 6.2 breaks down this distribution into its most relevant sections. I consider bunchers to be bus lines that operate between 80% and 85% of the daily goal, which is just enough to be eligible for subsidies. The model underestimates the number of bus lines choosing to operate below the goal, and predicts that very few bus lines would bunch with the subsidies in place – they should instead operate strictly above the 80% of goal threshold.

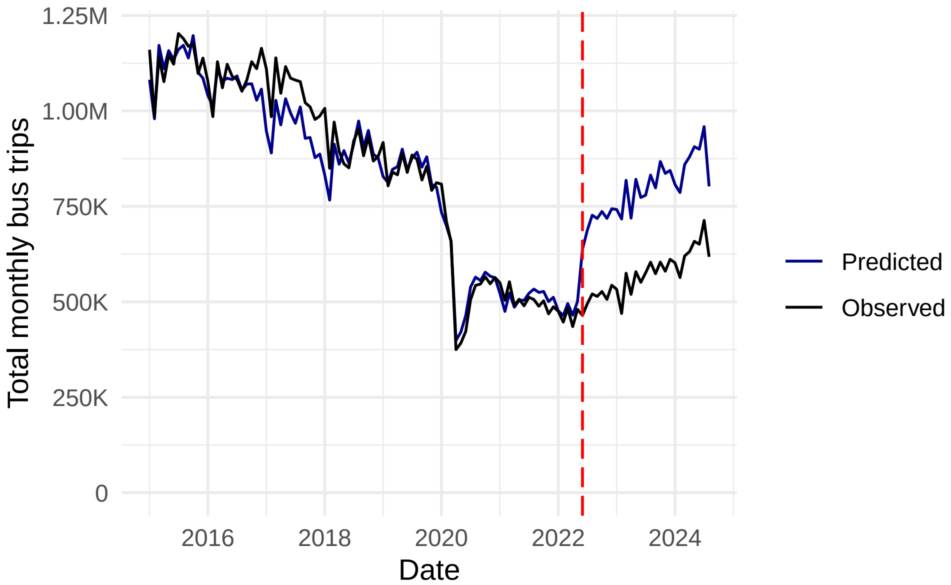


Figure 6.1: Number of bus trips: actual vs. predicted

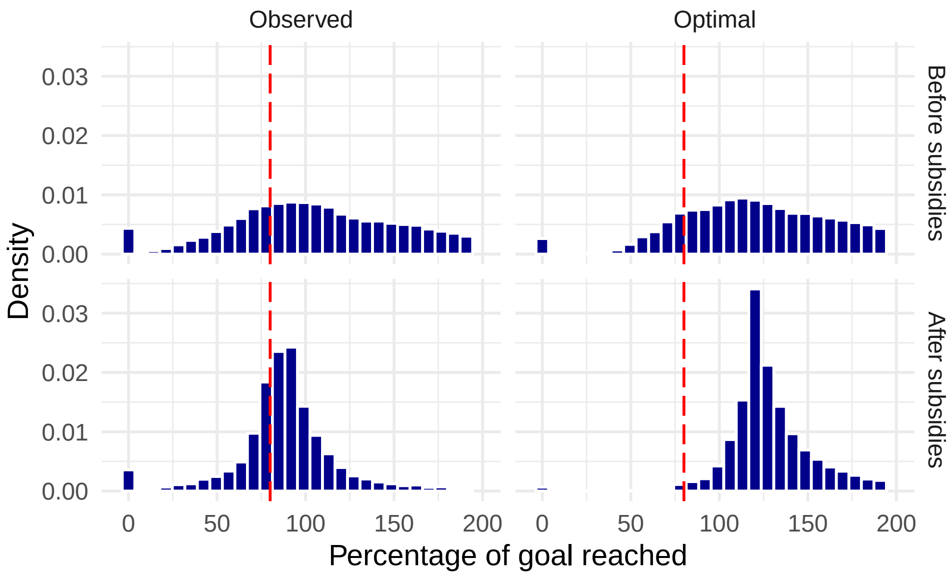


Figure 6.2: Goal achievement: actual vs. predicted

Table 6.2: Match across quantiles of the distribution

| | Before subsidies | | After subsidies | |
|---|---|---|---|---|
| | Observed | Predicted | Observed | Predicted |
| Inactive | 2.2% | 1.3% | 2.5% | 0.4% |
| Below 80% of goal | 21.4% | 11.7% | 30.4% | 1.6% |
| Bunching above 80% of goal | 3.1% | 2.6% | 11.2% | 0.6% |
| Above 80% of goal | 78.6% | 88.3% | 69.6% | 98.4% |

## 6.2 Counterfactuals

I now use the model to simulate counterfactual policy designs. I compare the current system with a scenario of making no policy changes – neither subsidizing nor raising prices. I also compare it with the alternatives of increasing prices instead of implementing subsidies and of subsidizing each passenger's fare, keeping prices constant for passengers, but readjusting them for the bus companies.

I run simulations for each of the 4 policies, and calculate bus availability and welfare outcomes in each, as well as the cost to the city government's budget. To obtain confidence intervals according to the NLLS properties discussed in section 5.3, I draw $K = 1000$ vectors of parameters $\delta$ from a $\mathcal{N}(\widehat{\delta}, \widehat{\mathbb{V}})$ distribution, where $\mathbb{V}$ is the asymptotic covariance matrix of $\widehat{\delta}$. I then calculate the number of bus trips, number of passengers, median wait times, goal achievement, and welfare for each, on a daily basis. From all the possible parameter draws, I calculate the mean, as well as the $2.5^{\text{th}}$ and $97.5^{\text{th}}$ percentiles, which serve as bounds for a 95% confidence interval.

Each year since 2021, Rio de Janeiro's department of transportation (SMTR) publishes a document calculating how much prices should increase following the changes in cost for bus companies, and the rate that companies should receive per passenger ("taxa de remuneração"). It represents the fare price that the SMTR predicts should be set were there no subsidies. This rate is used as a basis for calculating the rate of subsidies paid per kilometer, so I use it to calculate counterfactual price increases.

Full results of the simulations are displayed in Table 6.4. The number of average daily passengers is roughly consistent across the four scenarios, except for the third column, where fare prices are increased instead of creating subsidies. Due to the sizeable increases in price required to match the revenue granted by the subsidies, over 20% of passengers cease to take the bus in this scenario. In terms of bus availability, the first column – containing the policy

Table 6.3: Counterfactual policies

| | Actual | No change | Price hikes | Subs. per passenger |
|---|---|---|---|---|
| Passengers | 1.62M | 1.61M | 1.26M | 1.62M |
| | [1.62M, 1.62M] | [1.61M, 1.61M] | [1.26M, 1.26M] | [1.62M, 1.62M] |
| Wait time | 6.5 | 8.1 | 7.6 | 6.8 |
| | [6.45, 6.46] | [8.1, 8.14] | [7.62, 7.66] | [6.8, 6.83] |
| Bus trips | 26.18K | 18.97K | 20.16K | 22.52K |
| | [26.17K, 26.19K] | [18.93K, 19.02K] | [20.11K, 20.21K] | [22.47K, 22.58K] |
| Proportion inactive | 0.5% | 1.2% | 0.9% | 0.6% |
| | [0.005, 0.005] | [0.012, 0.012] | [0.009, 0.009] | [0.006, 0.006] |
| Proportion above 80% of goal | 99.5% | 83.9% | 88.3% | 93.7% |
| | [0.995, 0.995] | [0.836, 0.842] | [0.88, 0.886] | [0.935, 0.939] |
| Proportion with wait $\leq$ 15mins | 83.7% | 80.5% | 83.5% | 87.8% |
| | [0.837, 0.838] | [0.805, 0.806] | [0.834, 0.836] | [0.877, 0.879] |
| Profit | 6.26M | 3.51M | 4.10M | 5.44M |
| | [6.26M, 6.26M] | [3.51M, 3.51M] | [4.10M, 4.10M] | [5.44M, 5.44M] |
| Consumer surplus | 14.16M | 14.02M | 11.31M | 14.13M |
| | [14.16M, 14.16M] | [14.02M, 14.02M] | [11.31M, 11.31M] | [14.13M, 14.13M] |
| Government cost | 2.78M | 0.00 | 0.00 | 3.14M |
| | [2.78M, 2.78M] | [0.00, 0.00] | [0.00, 0.00] | [3.14M, 3.14M] |
| Welfare | 17.64M | 17.53M | 15.41M | 16.43M |
| | [17.64M, 17.65M] | [17.52M, 17.53M] | [15.41M, 15.42M] | [16.43M, 16.43M] |
| Pass $\times$ Km | 5.30B | 3.43B | 2.85B | 4.08B |
| | [5.30B, 5.30B] | [3.42B, 3.44B] | [2.85B, 2.86B] | [4.07B, 4.09B] |

actually introduced by city government – produces the highest number of daily buses and lowest wait times.

This policy also ensures that very few bus lines are inactive (with zero trips), and that the large majority of them operate above 80% of the imposed daily kilometer goal. This is because the alternative policies do not impose these daily goals. The system of directly subsidizing the fare of each passenger outperforms the actual subsidy system only in terms of achieving a higher proportion of bus lines with wait times below 15 minutes. This may be because this system incentivizes bus operators to concentrate on bus lines with more passengers, which already have low wait times, causing a larger amount of them to be able to dip below 15 minutes of average wait time.

On most metrics, however, the actual subsidy system implemented leads to better outcomes for passengers than the alternatives. Besides mostly outperforming the direct subsidies per passenger, its cost for the government is lower, as it can be thought of as a subsidy mostly targeted as bus lines with fewer passengers, whereas subsidizing on a per-passenger basis aims mostly at passenger-dense bus lines, which often already have high bus frequency. The scenarios of making no policy change and of increasing prices generate no cost for the government, as no subsidies are paid.

Now I show how to calculate the welfare of passengers in this model. From the demand specification, the number of passengers in each bus line is

$$\text{Pass}_{st} = \exp\left(\widehat{\alpha}_s + \widehat{\alpha}_t - \widehat{\phi}_p \log(\underline{p}) - \widehat{\phi}_w \text{wait}(n) + \varepsilon_{st}\right).$$

Consumer Surplus should hence be

$$\int_{\underline{p}}^{\infty} \text{Pass}_{st}\, \mathrm{d}p,$$

but this diverges for $\phi_p \in (0, 1)$. Instead, I set a ceiling of $\overline{p} = 20$, up to which I integrate. Essentially, I truncate the demand above this point, assuming that no passengers are willing to pay more that $\overline{p}$ for the bus. Welfare is then the sum of consumer surplus and private bus company profit, subtracted by the government cost. Towards the bottom of Table 6.4, I display the average daily welfare across bus lines – the sum of the welfare produced in each one.

According to the welfare calculations, the best policy for passengers among these four choices is the one adopted in reality by the city government. Due to the large cost to the city's government, the second best option is to neither subsidize nor increase prices, and accept a considerable decrease in the number of bus trips. Because direct subsidies on each passenger's fare are both more costly to the government and generate slightly lower consumer surpluses, its overall welfare is also lower. And finally, the alternative of increasing prices instead of subsidizing produces much lower consumer surpluses due to the reduced passenger demand, leading to the lowest welfare outcomes.

It is also important to consider the positive externalities generated by public transit for the environment, congestion, and for overall economic activity. By weighting these factors, total welfare for consumers should also contain an additional term. I assume that externalities are proportional to the total number of bus passenger-kilometers, which I calculate in the last row[1]. Total welfare should therefore be a linear combination of the welfare I calculated previously – which I will henceforth refer to as internal welfare – and the passenger-kilometers.

As the actual policy in the first column has both higher internal welfare and passenger-kilometers than the alternatives, it is strictly dominant within this framework. Meanwhile, not making any policy changes generates higher internal welfare than subsidies per passenger, but fewer passenger-kilometers. Therefore, the two cannot be directly compared, as the externalities may outweigh the direct benefits for passengers. However, both strictly dominate the scenario of raising prices.

As a final exercise, I isolate the two main features of the subsidy system adopted by the city government in order to assess what makes it successful. The system consists of paying subsidies per kilometer to bus lines which are able to reach 80% of a daily goal. Because of this, I compare this system to two other alternatives. The first is one paying subsidies per kilometer without requiring goals, at the same rate as the one actually paid by the city government. The second is one which only requires reaching the goal, but pays no additional subsidies per kilometer. In this scenario, when a bus line achieves 80% of its daily goal, it receives a lump-sum payment corresponding to the amount which would be paid for that number of kilometers in the other scenarios.

Results are displayed in Table 6.3. Even without imposing goals, subsidies per kilometer achieve virtually the same outcomes as the complete system. This is because, in the simulations, most bus lines are predicted to operate above the goal. The additional marginal revenue per kilometer given by the subsidies

[1] I calculate it as the sum, across bus lines and days, of Passengers $\times$ Length of the bus line, in kilometers, divided by two, assuming that the average passenger would travel half of the total length of the route.

Table 6.4: Counterfactual policies

| | Actual | Subs. per Km, no goal | Lump sum upon reaching goal |
|---|---|---|---|
| Passengers | 1.62M | 1.62M | 1.61M |
| | [1.62M, 1.62M] | [1.62M, 1.62M] | [1.61M, 1.61M] |
| Wait time | 6.5 | 6.5 | 7.7 |
| | [6.45, 6.46] | [6.45, 6.46] | [7.7, 7.73] |
| Bus trips | 26.18K | 26.18K | 20.58K |
| | [26.17K, 26.19K] | [26.17K, 26.19K] | [20.55K, 20.61K] |
| Proportion inactive | 0.5% | 0.5% | 0.6% |
| | [0.005, 0.005] | [0.005, 0.005] | [0.006, 0.006] |
| Proportion above 80% of goal | 99.5% | 99.5% | 99.4% |
| | [0.995, 0.995] | [0.995, 0.995] | [0.994, 0.994] |
| Proportion with wait $\leq$ 15mins | 83.7% | 83.7% | 81.1% |
| | [0.837, 0.838] | [0.837, 0.838] | [0.81, 0.812] |
| Profit | 6.26M | 6.26M | 5.36M |
| | [6.26M, 6.26M] | [6.26M, 6.26M] | [5.35M, 5.36M] |
| Consumer surplus | 14.16M | 14.16M | 14.06M |
| | [14.16M, 14.16M] | [14.16M, 14.16M] | [14.06M, 14.06M] |
| Government cost | 2.78M | 2.78M | 1.85M |
| | [2.78M, 2.78M] | [2.78M, 2.78M] | [1.85M, 1.85M] |
| Welfare | 17.64M | 17.64M | 17.56M |
| | [17.64M, 17.65M] | [17.64M, 17.65M] | [17.56M, 17.57M] |
| Pass $\times$ Km | 5.30B | 5.30B | 3.88B |
| | [5.30B, 5.30B] | [5.30B, 5.30B] | [3.88B, 3.89B] |

is enough to incentivize bus companies to reach the daily goals desired by the city government.

The system of paying a lump sum when the goal is reached is also able to incentivize a large majority of bus lines to achieve the goal, bus it leads to worse outcomes for passengers across the board. In this scenario, there are fewer bus trips, increased wait times, lower internal welfare, and fewer total passenger-kilometers. This is because this system pushes bus companies to achieve 80% of the goal, but creates no additional incentives for them to go above and beyond it. As a result, bus frequency is reduced. While the government cost is much lower, this is outweighed by reduced profits for the bus companies, leading to a slightly lower internal welfare. As there is a large drop in the number of bus trips and passenger-kilometers, it may also produce fewer positive externalities.

# 7
# Conclusions

Subsidies are a common form of government support to public transportation. They most often take the form of a fare discount for passengers, where the government pays the difference. However, the experience of Rio de Janeiro shows that a specially tailored regulation design can provide incentives for bus operators to increase bus frequency. In particular, paying subsidies per kilometer and conditioning payments on reaching a daily goal, monitored by GPS, successfully aligns public and private interests.

In this paper, I estimate the demand for buses, showing how passengers respond to price increases and changes in the wait time they face. I then build and estimate a model of bus supply and how it changes according to new rules and regulations. I use it to simulate several policy counterfactuals, which suggest that the most powerful tool at the government's disposal is to closely monitor bus activity and set goals tied to financial incentives.

In the simulations, paying subsidies per kilometer and conditional on achieving a daily kilometer goal leads to higher bus frequency and lower wait times than other policy alternatives. Paying subsidies per kilometer rather than per passenger incentivizes bus companies to provide a higher volume of bus trips across the board, instead of focusing solely on the most profitable passenger-dense bus lines. As a result, this subsidy system produces better outcomes for passengers and higher welfare, at a lower cost for the government's budget.

# 8
# Bibliography


ALMAGRO, M. et al. **Optimal Urban Transportation Policy: Evidence from Chicago**. [S.l.], 2024. (Working Paper Series, 32185). Disponível em: http://www.nber.org/papers/w32185.

CAMPOS, M. C. de. **Urban Mobility, Inequality and Welfare in Developing Countries: Evidence from 2016 Olympics in Rio de Janeiro**. Tese (Doutorado) — PUC-Rio, 2019.

DAVIDSON, R.; MACKINNON, J. G. et al. **Estimation and inference in econometrics**. [S.l.]: Oxford New York, 1993. v. 63.

GAGNEPAIN, P.; IVALDI, M. Incentive regulatory policies: The case of public transit systems in france. **RAND Journal of Economics**, v. 33, n. 4, p. 605–629, 2002. Disponível em: https://EconPapers.repec.org/RePEc:rje:randje:v:33:y:2002:i:winter:p:605-629.

LEWIS, G.; BAJARI, P. Moral hazard, incentive contracts, and risk: evidence from procurement. **Review of Economic Studies**, Oxford University Press, v. 81, n. 3, p. 1201–1228, 2014.

MILLER, C.; SAVAGE, I. Does the demand response to transit fare increases vary by income? **Transport Policy**, v. 55, p. 79–86, 2017. ISSN 0967-070X. Disponível em: https://www.sciencedirect.com/science/article/pii/S0967070X16302852.

MOHRING, H. Optimization and scale economies in urban bus transportation. **The American Economic Review**, JSTOR, v. 62, n. 4, p. 591–604, 1972.

NELSON, P. et al. Transit in washington, dc: Current benefits and optimal level of provision. **Journal of Urban Economics**, v. 62, n. 2, p. 231–251, 2007. ISSN 0094-1190. Essays in Honor of Kenneth A. Small. Disponível em: https://www.sciencedirect.com/science/article/pii/S0094119007000241.

PARRY, I. W. H.; SMALL, K. A. Should urban transit subsidies be reduced? **American Economic Review**, v. 99, n. 3, p. 700–724, 6 2009. Disponível em: https://www.aeaweb.org/articles?id=10.1257/aer.99.3.700.

Prefeitura da Cidade do Rio de Janeiro, Secretaria Municipal de Transportes. **Apresentação Novo Acordo**. [S.l.], 2022. Disponível em: https://transportes.prefeitura.rio/wp-content/uploads/sites/31/2022/06/ApresentacÌğaÌČo-Novo-Acordo.pdf.

Prefeitura da Cidade do Rio de Janeiro, Secretaria Municipal de Transportes. **Nota Técnica TR/SUBP Nº 01/2022 – Metodologia de Dimensionamento de Medida de Contingência de Apoio ao SPPO definida em Acordo Judicial**. [S.l.], 2022. Disponível em: https://transportes.prefeitura.rio/wp-content/uploads/sites/31/2023/06/Nota-Te%CC%81cnica-Metodologia-de-Dimensionamento-de-Medida-de-Continge%CC%82ncia-de-Apoio-ao-SPPO-para-Acordo-Judicial.pdf.

Prefeitura do Rio de Janeiro. **Datalake: Repositório Centralizado de Dados Institucionais**. 2025. Accessed: 2025-03-18. Disponível em: https://www.dados.rio/datalake.

RIVAS, M. E.; BRICHETTI, J.; SEREBRISKY, T. **Operating Subsidies in Urban Public Transit in Latin America: A Quick View**. [S.l.: s.n.], 2020.

TSIVANIDIS, N. Evaluating the impact of urban transit infrastructure: Evidence from bogota's transmilenio. **Unpublished manuscript**, v. 18, 2022.

# A
# Competition between bus lines

In section 4.2, I estimate the semi-elasticity of the demand for a bus line with respect to wait times, $\phi_w$. One possible concern is that this coefficient captures two distinct effects: the first is that, as wait times increase, passengers quit using the bus and find an outside alternative. The second is that passengers may substitute more frequent bus lines for scarcer ones, thus mitigating their welfare losses.

In this section, I compute the geographical overlap between each pair of bus lines, and use these values to conduct two robustness exercises showing that the effect in $\phi_w$ is not driven by substitution between bus lines. First, I calculate a Herfindahl-Hirschman Index (HHI) for bus lines, and show that demand reacts more strongly to changes in wait time in more monopolistic bus lines. If substitution between bus lines was prominent, passengers in bus lines subject to more competition should be more elastic with respect to wait times, as they can substitute more easily. In the second exercise, I calculate an index for the wait times of competing bus lines, and include it in the demand regression. This competition effect is small and leaves the estimate for $\phi_w$ mostly unaltered.

As a preparatory step, I use the shapefiles of the itineraries of each bus line provided by the SMTR to compute the degree of overlap between pairs of bus lines. I add a 500 meter buffer region around one of the lines and take their geometric intersection. Then, the overlap is calculated as

$$\text{overlap}_{i,j} = \frac{\mathrm{length}(i \cap j)}{\mathrm{length}(i)}.$$

I interpret it as being a proxy for the proportion of passengers from line $i$ who would also be willing to take line $j$, assuming passengers are uniformly distributed along the itineraries.

## A.1
## Wait-time elasticity for different market concentrations

To calculate an HHI, I define the relevant market for each bus line $i$ as the universe of passengers who would be willing to take it. The share of this market held by the bus line $j$ is given by

$$s_{j,i} = \frac{\text{overlap}_{j,i} \times \overline{\text{Pass}}_j}{\sum_k \text{overlap}_{k,i} \times \overline{\text{Pass}}_k},$$

where $\overline{\text{Pass}}_k$ is the average number of daily passengers who took bus line $k$ in a given month. The HHI measuring the level of competition faced by bus line $i$ is hence calculated as

$$\text{HHI}_i = \sum_j s_{j,i}^2.$$

This index is 1 if, and only if, bus line $i$ is a monopolist in its own market. Lower values indicate that market shares are dispersed across many bus lines, and therefore, bus line $i$ faces more competition. In Table A.1, I reproduce the demand regression of section 4.2 but add the HHI and an interaction term.

In the first column, the estimate for $\phi_w$ is similar to the earlier one. This does not continue in the next columns, but the interaction coefficient grows to be strongly negative. This means that in bus lines that are more monopolistic, passengers are more elastic with respect to wait times, even though they have fewer buses to substitute. This suggests that the effect captured by $\phi_w$ is mainly one of passengers substituting away from buses altogether, and not between competing bus lines.

Table A.1: Wait-time elasticity of demand with HHI

| | Log passengers | | |
|---|---|---|---|
| | Full sample | 3-month window | 1-month window |
| Wait Time | -0.008*** | -0.001* | -0.001+ |
| | (0.000) | (0.001) | (0.001) |
| HHI | 3.378*** | 7.908*** | 9.084*** |
| | (0.060) | (0.331) | (0.586) |
| Wait Time × HHI | -0.016*** | -0.234*** | -0.177*** |
| | (0.002) | (0.017) | (0.023) |
| Time FEs | ✓ | ✓ | ✓ |
| Bus Service FEs | ✓ | ✓ | ✓ |
| Num.Obs. | 951238 | 41505 | 13637 |
| R2 | 0.875 | 0.949 | 0.960 |

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

### A.2
### Cross wait-time elasticity between bus lines

For the second robustness exercise, I assume that when the wait time of bus line $j$ increases by 1, a proportion $\gamma \times \text{overlap}_{j,i}$ of its passengers will substitute line $i$ for line $j$. The cross elasticity of demand with respect to wait times is

$$\frac{\partial \log \text{Pass}_i}{\partial \text{wait time}_j} = \gamma \times \text{overlap}_{j,i} \times \frac{\overline{\text{Pass}_j}}{\overline{\text{Pass}_i}},$$

which is proportional to both the proportion of passengers of bus line $j$ who may also take bus line $i$, and the relative relevance of bus line $j$ compared to bus line $i$, in terms of the magnitude of passengers. From this framework, I combine all possible contributing bus lines $j$ to compute

$$\text{competing wait time}_{it} = \sum_j \text{wait time}_{jt} \times \text{overlap}_{j,i} \times \frac{\overline{\text{Pass}_j}}{\overline{\text{Pass}_i}},$$

and include it in the regression shown in Table A.2. Results show that the estimate for $\gamma$, although statistically significant, is of tiny magnitude. Most

Table A.2: Wait-time elasticity of demand adjusting for competition

| | Log passengers | | |
|---|---|---|---|
| | Full sample | 3-month window | 1-month window |
| Wait Time | -0.00863*** | -0.00388*** | -0.00415*** |
| | (0.00012) | (0.00037) | (0.00059) |
| Competing Wait Time | -0.00006*** | -0.00009*** | -0.00009*** |
| | (0.00000) | (0.00001) | (0.00002) |
| Time FEs | ✓ | ✓ | ✓ |
| Bus Service FEs | ✓ | ✓ | ✓ |
| Num.Obs. | 528657 | 29866 | 9325 |
| R2 | 0.909 | 0.961 | 0.966 |

$+$ p $<$ 0.1, * p $<$ 0.05, ** p $<$ 0.01, *** p $<$ 0.001

importantly, accounting for these confounding competition effects does not change the scale of the estimated $\phi_w$, measuring the drop in passengers driven by an increase of wait times. This once again suggests that the main estimates for $\phi_w$ are not driven by the substitution between bus lines.

# B
# Solving the model

In this section, I show how to obtain a closed-form solution for the model in chapter 5, using the non-elementary Lambert W function. This drastically speeds up computation compared to numerical optimization methods, allowing for the model to be estimated within a reasonable time frame, even with over a million observations. I first find the first order condition for optimizing the profit of each bus line and solve for the number of bus trips. I then show that the second order condition for profit maximization holds.

## B.1
## First order condition

The profit of each bus line can be written in the form

$$\pi(n) = \lambda\, p\, M \exp\Big( - \phi_w \mathsf{wait}(n)\Big) + \mathbb{I}[n \geq \overline{n}] \times L \times n \times \mathrm{subskm} - C(n)$$

where $M = \exp\Big(\widehat{\alpha}_s + \widehat{\alpha}_t - \widehat{\phi}_p \log(p) + \widehat{\sigma}^2/2\Big)$ and $\mathsf{wait}(n) = \overline{T}/(2n)$, $\overline{T}$ being the average number of minutes during which the bus line operates on a given day. The cost is given by

$$C(n) = \delta_1 \times \mathsf{diesel\ price}_t \times L \times n + \delta_2 \times n + q \times \overline{\mathsf{fleet}}_{st} \times \mathbb{I}[n > 0].$$

Differentiating with respect to $n$ and letting $C'(n) \equiv c$.

$$\begin{aligned}\pi'(n) = \lambda\, p\, M \exp(-\phi_w \mathsf{wait}(n)) \times (-\phi_w) \mathsf{wait}'(n) \\ + \mathbb{I}[n \geq \overline{n}] \times L \times \mathrm{subskm} - c.\end{aligned}$$

Substituting $\text{wait}(n) = \overline{T}/(2n)$ and $\text{wait}'(n) = -\overline{T}/(2n^2)$,

$$\pi'(n) = \lambda\, p\, M \exp\Big( - \phi_w \overline{T}/(2n)\Big) \times \phi_w \overline{T}/(2n^2)$$
$$+ \mathbb{I}[n \geq \overline{n}] \times L \times \text{subskm} - c.$$

Setting the derivative equal to zero and rearranging,

$$\exp\Big( - \phi_w \overline{T}/(2n)\Big) \times \phi_w \overline{T}/(2n^2) = \frac{c - \mathbb{I}[n \geq \overline{n}] \times L \times \text{subskm}}{\lambda\, p\, M}.$$

The Lambert W function is an inverse of $f(x) = x\exp(x)$, so we can solve for $n$ by writing the expression in the left-hand side in that form. Taking square roots on both sides,

$$\exp\Big( - \phi_w \overline{T}/(4n)\Big) \times \sqrt{\phi_w \overline{T}/2}\Big/ n = \sqrt{\frac{c - \mathbb{I}[n \geq \overline{n}] \times L \times \text{subskm}}{\lambda\, p\, M}}.$$

Multiplying both sides by $-\frac{1}{2}\sqrt{\phi_w \overline{T}/2}$ so that the constants match,

$$\exp\Big( - \phi_w \overline{T}/(4n)\Big)\Big( - \phi_w \overline{T}/(4n)\Big) = -\frac{1}{2}\sqrt{\frac{\Big(c - \mathbb{I}[n \geq \overline{n}] \times L \times \text{subskm}\Big)\phi_w \overline{T}}{2\,\lambda\, p\, M}}.$$

As the expression is of the form $x\exp(x) = k$, I use the principal branch of the Lambert W function to invert the left-hand side, leading to $x = W(k)$.

$$-\phi_w \overline{T}/(4n) = W\left( -\frac{1}{2}\sqrt{\frac{\Big(c - \mathbb{I}[n \geq \overline{n}] \times L \times \text{subskm}\Big)\phi_w \overline{T}}{2\,\lambda\, p\, M}}\right).$$

Solving for $n$,

$$n = -\frac{\phi_w \overline{T}}{4 \times W\left( -\frac{1}{2}\sqrt{\frac{\big(c - \mathbb{I}[n \geq \overline{n}] \times L \times \text{subskm}\big)\phi_w \overline{T}}{2\,\lambda\, p\, M}}\right)}.$$

Note that this has two possible solutions: one where the subsidy goal is achieved ($n \geq \overline{n}$) and another where it is not, changing the value of $\mathbb{I}[n \geq \overline{n}]$. To overcome this, we can simply calculate $n_1$ by setting $\mathbb{I}[n \geq \overline{n}] = 1$ and $n_0$ by setting $\mathbb{I}[n \geq \overline{n}] = 0$. Then consider $n_1$ only if $n_1 \geq \overline{n}$ and $n_0$ only if $n_0 < \overline{n}$. If both hold, we can take $n^*$ to be the one which produces the highest profits.

### B.2
### Second derivative test

The first derivative of the profit function is

$$
\begin{aligned}
\pi'(n) = \lambda\, p\, M \exp\Big( - \phi_w \overline{T}/(2n)\Big) \times \phi_w \overline{T}/(2n^2) \\
+ \mathbb{I}[n \geq \overline{n}] \times L \times \text{subskm} - c.
\end{aligned}
$$

By the product rule,

$$
\begin{aligned}
\pi''(n) = &-\lambda\, p\, M \exp\Big( - \phi_w \overline{T}/(2n)\Big) \times \phi_w \overline{T}/n^3 \\
&+ \lambda\, p\, M \exp\Big( - \phi_w \overline{T}/(2n)\Big) \times \Big(\phi_w \overline{T}/2n^2\Big) \times \phi_w \overline{T}/(2n^2) \\
= &-\lambda\, p\, M \exp\Big( - \phi_w \overline{T}/(2n)\Big) \times \phi_w \overline{T}/n^3 \times \left[1 - \frac{\phi_w \overline{T}}{4n}\right],
\end{aligned}
$$

which is negative whenever $\frac{\phi_w \overline{T}}{4n} < 1$. Since $\widehat{\phi}_w = 0.006$ and $\overline{T} < 1440$ hours in a day, this always holds for $n > 2.16$. Therefore, most local optima arising from the first order condition are local maxima. Otherwise, the solution will be at a discontinuity or at zero.