University of São Paulo

School of Economics, Business Administration and Accounting at Ribeirão Preto

Department of Economics

Graduate Programs in Economics - Area: Applied Economics

The welfare costs of traffic congestion in São Paulo Metropolitan

Area

Ricardo Campante Cardoso Vale

Supervisor: Cláudio Ribeiro de Lucinda

Ribeirão Preto - SP

April 19th, 2018

Prof. Dr. Vahan Agopyan

Rector, University of São Paulo

Prof. Dr. Dante Pinheiro Martinelli

Dean, School of Economics, Business Administration and Accounting at Ribeirão

Preto

Prof. Dr. Renato Leite Marcondes

Head of Economics Department

Prof. Dr. Sérgio Naruhiko Sakurai

Chief of Graduate Programs in Economics - Area: Applied Economics

RICARDO CAMPANTE CARDOSO VALE

The welfare costs of traffic congestion in São Paulo

Metropolitan Area

Master's Thesis submitted to the Graduate Program in Economics - Area: Applied Economics in the School of Economics, Business Administration and Accounting at Ribeirão Preto / University of São Paulo, to obtain a Master's degree in Sciences. Corrected version. The original is available at FEA-RP / USP.

Supervisor: Cláudio Ribeiro de Lucinda

Ribeirão Preto - SP

April 19th, 2018

I authorize the disclosure and total or partial reproduction of this work, by any conventional or electronic mean, for study and research purposes, provided that the source is mentioned.

Name: Ricardo Campante Cardoso Vale

Title: The welfare costs of traffic congestion in São Paulo Metropolitan
Area

Master's Thesis submitted to the Graduate
Program in Economics - Area: Applied Eco-
nomics in the School of Economics, Business
Administration and Accounting at Ribeirão
Preto / University of São Paulo, to obtain
a Master's degree in Sciences. Corrected ver-
sion. The original is available at FEA-RP /
USP.

Approved in:

Examination Committee

_____

**Prof. Dr. Cláudio Ribeiro de
Lucinda (Supervisor)**
FEA-RP/USP

_____

Prof. Dr. Ciro Biderman
EESP/FGV

_____

Prof. Dr. Danilo Camargo Igliori
FEA/USP

_____

**Prof. Dr. Rodrigo Menon Simões
Moita**
Insper

Acknowledgements

Firstly, I would like to acknowledge Instituto Escolhas for providing a fellowship for this thesis and also for offering me courses and seminars that expanded my knowledge and horizons. My sincere thanks to all the Escolhas' staff operating the Economics and Environmental Chair.

I would like to thank to my thesis supervisor Prof. Dr. Cláudio Lucinda for the helpful guidance, that always steered me in the right direction to complete this thesis.

I also would like to thank all the professors I had during the graduation for sharing their knowledge. I would like to particularly mention Prof. Dr. Márcio Laurini for all the attention given to my doubts, Prof. Dr. Luiz Guilherme Scorzafave and Prof. Dr. Daniel Santos for the insights given during the intermediate examination committees.

I have to express my profound gratitude to Dr. Petterson Vale, he kindly dedicated himself as second reader of this thesis, giving me so many helpful comments and holding valuable discussions to improve it.

Besides that, I would like to thank the other members of the Gibbs Land Use and Environment Lab who are contributing to my formation as researcher. Special thanks to Prof. Dr. Holly Gibbs who provided me the opportunity to join her team.

I am grateful to my parents for providing me with unfailing support and continuous encouragement throughout all my years of study, to my sister for the companionship and the sincere talks. To my girlfriend for all the patience, affection and care.

To the friends met in Ribeirão Preto that made the long study days more pleasant, Bruno and Giulio for sharing home. Also to the old friends of Belo Horizonte: the 'Bancada' that always welcomed me back to my hometown and the friends of UFMG that are always close regardless of any physical distance.

Finally, I would like to acknowledge the Brazilian society as a funder of my studies through the public higher education system I benefited from. I hope to be worth the investment.

Abstract

VALE, R. C. C. **The welfare costs of traffic congestion in São Paulo Metropolitan Area**. Thesis (Master's Degree) – School of Economics, Business and Accounting at Ribeirão Preto, University of São Paulo, Ribeirão Preto, 2018.

This thesis presents new evidences of negative impacts of traffic congestion in São Paulo Metropolitan Area, indicating that the omission or retardation of policy to alleviate that charges society with a high cost. Besides that, we shed light on Waze application's impacts, evidencing GPS-based participatory navigation is a key element of the traffic in the contemporary cities. Induced by the economic theory, one can see that traffic slowness derived from the general demand for automotive vehicles corresponds to a social cost. With the objective of translating this cost to monetary units, we developed an approach that matches the 2012 Origin-Destination survey to Google Maps data. An econometric travel mode choice model is estimated and through that we infer the Value of Time for trips. From counterfactual analysis for travel duration, we estimate the delays of the trips in relation to a hypothetic free flow situation. Joining the calculations, we estimate that 89% of the trips motivated by work in São Paulo are retarded by traffic frictions, generating an annual welfare cost that amounts to R$7.338 billions. Additionally, we analyze widespread Waze malfunctioning at 10/23/2017. Combining congestion data to the Value of Time, we estimate that the Waze bug approximatedely tripled the welfare costs in regards to a typical day, showing the magnitude of mass effects that a social navigation application is able to trigger.

Keywords: urban mobility, congestion, costs of transportation, social network, GPS-navigation.

Resumo

VALE, R. C. C. **Os custos de bem-estar do congestionamento do trânsito na Região Metropolitan de São Paulo**. Dissertação (Mestrado) – Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2018.

Esta dissertação traz novas evidências dos efeitos negativos do congestionamento do trânsito na Região Metropolitana de São Paulo, mostrando que a morosidade na implantação de soluções custa muito caro para a sociedade. Além disso, evidências empíricas dos impactos do aplicativo Waze são elucidadas, mostrando que a navegação participativa por meio de aparelhos com GPS não pode ser esquecida ao pensarmos o trânsito nas cidades contemporâneas. À luz da teoria econômica, é possível ver que a lentidão do trânsito proveniente da demanda coletiva por automóveis corresponde a um custo social. Para quantificar este custo em unidades monetárias, é desenvolvida uma abordagem que combina dados da Pesquisa de Mobilidade de 2012 com dados do Google Maps. Através de um modelo econométrico de escolha de modal de transporte, são inferidos os valores do tempo em viagem e, através da análise contrafactual da duração dos deslocamentos, é estimado o tempo de atraso das viagens em relação a uma situação hipotética de fluxo livre. Juntando os cálculos, é estimado que 89% das viagens motivadas por trabalho em São Paulo são atrasadas por fricções do tráfego, gerando um custo de bem-estar da ordem de R$7,338 bilhões por ano. Ademais, são analisados empiricamente os impactos de uma falha generalizada do Waze ocorrida em 23/10/2017. Combinando dados de congestionamento com os valores de tempo já inferidos, é estimado que a falha aproximadamente triplicou os custos de bem-estar em relação a um dia típico, dando magnitude aos efeitos de massa que um aplicativo de navegação social é capaz de provocar.

Palavras-chaves: mobilidade urbana, congestionamento, custos de transporte, redes sociais, navegação por GPS.

List of Figures

List of Tables

Contents

1 Introduction

1.1 Motivation

Everyday people struggle to avoid traffic gridlocks while trying to reach their destinations in São Paulo. The largest South American Metropolis has a daily volume of 43,715mi trips(METRÔ-SP, 2013) and 1472 km of congested streets, on average[1]. How much does the daily congestion cost for people? To what extent do GPS-based applications affect traffic and commuting? In this thesis we address these questions by estimating the welfare effects of traffic congestion in a microeconomic framework. We give a monetary measure to the time people lose in traffic jams by estimating a mode choice demand model. It is a structural econometric model using combined data from the 2012 São Paulo Urban Mobility Survey (PMU 2012) and Google Maps. From that, we infer people's marginal valuation of time. Then, using Google Maps' information on travel duration conditional on traffic flow, we estimate how long the trips would have lasted in the absence of any congestion. Having estimates of time loss and of the value of time, we are able to gauge the congestion welfare cost in São Paulo Metropolitan Area (SPMA). Further, we quantify the welfare effects of a day when Waze, the most popular navigation app in Brazil, experienced malfunction in São Paulo, contributing to debates on the consequences of new technologies for traffic patterns.

The extreme levels of traffic congestion in São Paulo are widely known and have been reported by both the national and international media(CABRAL, 2012; Folhapress, 2014). The Traffic Engeneering Company (Portuguese acronym CET) reports that trips at rush hours have a 31% rate of delay in comparison to the same route without congestion (CET, 2013b; CET, 2017). This has direct adverse effects on the economy as freights become more costly and there is an opportunity cost of workers being stuck in traffic. There are effects on the environment as the automobile engines work less efficiently at low speeds and pollute more. And traffic congestion has effects on human well-being as people are locked inside a vehicle. The goal of this thesis is to empirically assess the latter welfare

---

[1]   Calculated from 2017 data provided by CET under request

effects.

The topic of urban mobility affects most of the citizens' lives directly and also involves real estate interests, so it is constantly under debate. The discussions are not exclusive to São Paulo or Brazil. In 2008, in New York City, the Mayor Michael Bloomberg tried to implement a congestion pricing system that would charge a fee for inbound car trips to the core of Manhattan between 6 a.m. and 6 p.m. on weekdays, but his plans were blocked by the state legislature (BLISS, 2018). New York Governor Andrew Cuomo raised the theme again in 2017, sending a revised report to the Legislature proposing a US$11.52 fee for personal vehicles entering the core of Manhattan (FIX, 2018). The project is once more being discussed by legislators and generating divergent opinions among the population. The urban toll model has already been implemented and kept in large urbanizations such as London, Stockholm and Singapore. In São Paulo, disputes surrounding traffic issues have also been occupying the public and electoral agendas in recent years . The issues under debate range from speed limits to the placement of metro stations (CAITANO, 2011; LOBEL, 2017). In light of budgetary constraints, the former mayor Fernando Haddad called for federal tax revenue over the commerce of fuels (CIDE, in Portuguese acronym) to be transferred to Municipal Governments for investment in public transportation. Following the global trend, zone pricing schemes have also been considered in São Paulo (MENA; VELOSO, 2013).

A key challenge to tackling the problem is the congestion feedback loop. Brazilians bother with traffic and recognize it as a main problem, but when experiencing bad public transportation they respond by increasing the use of individual motorized vehicles, which in turn intensifies congestion. In the electoral year of 2018, a survey at the national level pointed out that over 80% of respondents would vote for a candidate whose priorities include investment in public transportation and bicycle lanes, but only 34% support urban tolls and only 30% support a tax rise for cars using fossil fuels (IDEIA; SOCIEDADE; ESCOLHAS, 2018). If on the one hand people want investment in public transportation, on the other they immediately associate it to bad conditions and desire private modes for their own use ('awful' is the second top of mind word when people are queried about public

transport and 'car' is the top answer when people are asked about which is their ideal mode of transport (IDEIA; SOCIEDADE; ESCOLHAS, 2018)). In this context, usage rate of individual motorized vehicles rose from 31.9% to 45.7% in the 1967-2012 period and, despite a deceleration in the last few years, it kept increasing (METRÔ-SP, 2013). This means that travelers are considering individual motivations when choosing a mode of transport, but in return they are generating negative externalities that affect everyone.

Finding definitive solutions is difficult. For example, the continuous widening of roads to accommodate traffic has become increasingly challenging due to high urban population density and to land scarcity. Evidence shows that Brazilians are unwilling to pay more taxes or to be charged with zoning fees, but they might not be aware of the magnitude of the losses provoked by traffic immobility. We estimate a R$ 7.338 bi annual welfare cost, so policies able to untie the current traffic knots could make people better-off even if this is not the prior perception. Quantifying impacts is an important tool, especially in the public policy domain. Whatever the proposed policy, it imposes a cost on society either through trade-offs in public budget allocation or through new taxation. Hence for any policy under analysis, it would be recommendable to assess costs and benefits. This thesis contributes to this discussion by estimating how much an extra minute in traffic is valued by individuals in the São Paulo Metropolitan Area and by calculating the total cost of congestion in terms of deducted welfare.

What policy is more effective in solving the traffic jam problem is unknown and it is outside of the scope of this thesis to find out. However, one thing that is highly likely is that any efficient solution today and in the future will use new technologies to some extent. Rapid technological development in the past decade has already affected the economy and transportation patterns, for example by enabling the search and acquisition of goods or services over the internet or by making the 'home-office' a more common phenomenon. Technologies affect transportation indirectly, by shifting the demand for travel, but also directly by providing low-cost (when not free) easy-access applications of GPS navigation, carpooling and on-demand ride services. The main effect of carpooling and ride services is mode substitution because it attracts passengers from public and private transportation

(SHAHEEN; COHEN, 2018; IDEIA; SOCIEDADE; ESCOLHAS, 2018). For navigation apps, the main effect is route substitution, which has broader impacts on congestion – if vehicles could be perfectly sprawled across the streets, traffic jams would be considerably reduced.

One way in which a route substitution model can work out is by connecting drivers to an information central that is able to redistribute their choices of direction considering the choices of other drivers. To date, Waze is the most popular such application with 65mi monthly active users worldwide (Waze, 2016). Its tremendous success can be largely explained by an interactive system where a network of live information is created, and users can feel as community members by helping each other. Despite the power of its technology, Waze is not exempt from polemics. Residents of previously quiet neighborhoods started to complain about outsiders' cars crowding their streets because of Waze's suggestions, which in one case culminated with a borough (Leonia, New Jersey-USA) closing its streets to non-locals (FODERARO, 2017; FODERARO, 2018). Moreover, a recent simulation by the Institute of Transportation Studies at UC Berkeley shows that Navigation apps could aggravate the traffic problem compared to the situation where they are not used (UC Berkeley, 2017; Institute of Transportation Studies UC Berkeley, 2018). The simulation is restricted to a few local streets and does not investigate the wider system. Also, it depends on assumptions about driver behavior and about the suggestions given by the application that can be contested, as everybody is being redirected to the same side street in the model. Nevertheless, there is a lack of empirical studies assessing the effects of navigation applications that could support these discussions.

Taking advantage of a major event evolving Waze in Brazil, we run an empirical study of the welfare effects that arise over commuters due to the wide use of GPS Navigation. On October 23rd 2017, a widespread malfunctioning in Waze led to an increased traffic gridlock throughout the day as was largely reported by the media(CAPELAS; RIBEIRO; MENGUE, 2017b; CAPELAS; RIBEIRO; MENGUE, 2017a; KLEINA, 2017; SANTINO, 2017; Canaltech, 2017). The problem was acknowledged by the company, although they did

not provide details[2]. Looking at CET's data, we can confirm the high volume of congestion in São Paulo from 10am to 4pm. The number of kilometers of congestion are noticeably above the average for a day when no other jam trigger was going on - no adverse weather, no big accidents nor crowded games and concerts - except for many drivers reporting Waze was directing them to the congested areas that they usually avoid. By combining these data to the valuation of time that we produce in the congestion analysis, we are able to calculate the welfare effects of this particular event.

1.2   Related Work

In addition to addressing questions of potential policy interest, this dissertation contributes to the urban and transportation economics literature as an empirical study in one of the largest metropolises in the world. It has been some time since economic theory seeks to incorporate the subject of congestion(BECKMANN; MCGUIRE; WINSTEN, 1955), producing models and explanations for the phenomenon(VICKREY, 1963; WALTERS, 1961). The economic literature interprets problems arising from the use of automobiles within the theoretical framework of externalities(PARRY; WALLS; HARRINGTON, 2007). That is to say, situations in which the consumer's welfare or the production possibilities of a firm are directly affected by the actions of other agents, without market pricing mediation.[3] In this case, each agent that decides to take their car, taxi, or other mode of transportation, is increasing traffic and generating costs. Since a person only observes her own cost at the time of the choice, several individuals can decide to leave their houses at the same time going through the same routes, so that each one will impose a charge on the others, which was not initially taken into account in the individual decision. In an urban world where the number of automobiles keeps rising, the economics of congestion are an increasingly relevant topic. (PARRY, 2009; SCARINGELLA, 2001).

This thesis is not the first study to discuss congestion costs in São Paulo. Concerned with the growth of the fleet without a counterpart in infrastructure investment,

---

[2]   We have contacted Waze's Public Relations Office in Brazil, which has answered with an official email classifying the malfunctioning as a single isolated event that has already been solved
[3]   Definition based on Mas-Colell et al. (1995).

Cintra (2014) studies the costs of São Paulo's immobility in two ways: by estimating the business opportunity cost due to job hours lost by people in transit plus fuel and freight disbursements, and by estimating the environmental costs of increased pollutant emissions. The present thesis differs from Cintra (2014) in many ways. First, Cintra (2014) is focused on business costs while we study welfare costs associated with individuals. Second and most importantly, the methodology and data types used are different. Cintra (2014)'s computation relies upon aggregate data and secondary information. We estimate discrete choice econometric models using microdata from thousands of individuals and trips, which also permits the simulation of several possible situations.

Another study estimating the economic costs of congestion in São Paulo is Haddad & Vieira (2015). They use a general equilibrium approach to simulate the value of time and to estimate the amount of GDP and families' consumption lost due to traffic. Their study yields results applicable to a broader geographic scope, such as effects in the state and in whole country. They also make projections for different time ranges, including what they call "very short run", "short run" and "long run".

The disadvantages of this more aggregated approach as compared to a microeconomic focus lies in the traditional trade-offs between general equilibrium and partial equilibrium analysis. To increase coverage, they use aggregated data and estimate more inputs, having estimates that can be noisier and less efficient. For example, they obtain value of time estimates based on labor productivity which, in turn, is estimated from travel duration and cost estimates aggregated by zones. Their data on the extra travel time due to congestion is derived from an econometric regression of commuting travel duration across Brazilian municipalities while ours comes from Google's local microdata. And, while our welfare estimation is based on theoretical utility, which can embrace more aspects of human wellbeing, their estimation is based on consumption, which is less embracing but much more tangible. Nevertheless, the two approaches are complementary and it is useful to assess their convergence.

When looking at works that used similar methodology, there are others estimating discrete travel demand models for São Paulo Metropolitan Area, but they have different

objectives. Barcellos (2014) studies the proposed decentralization of CIDE and its implications on São Paulo's traffic. Pacheco & Chagas (2016) estimate the potential shift in the demand for trips caused by the establishment of a congestion toll in São Paulo's expanded center, while Lucinda et al. (2017) try to understand welfare effects of this policy and its distribution over the population compared to the current policy of license plate. Similarly, Feres (2015) analyses a hypothetic new rule to public transport fare that would vary accordingly to demand dynamics along the day. Luz (2017) discusses the introduction of Google Maps data combined to Origin-Destination Survey in order to estimate travel demand. Finally, Moita & Lopes (2016) also estimate a discrete choice model, but, differently of this thesis and the others cited above that model the transport choices at the individual level, they apply a model that ultimately aggregates demand at zone level to analyze effects of public policies over the transportation share of each mode.

All these empirical jobs used the Origin-Destination Survey of 2007 (O-D 2007), while we are using the PMU 2012, which is an update of O-D 2007. The earlier has a bigger sample and more disaggregated zone information than its update. Even though, we are dealing with a dynamic urban center and Mobility Surveys were created exactly to fill the gap between decennial Origin-Destination Surveys. If we are already late with a 2012 snapshot, a 2007 can be even worse. Another important difference from our model to most of the others is inputting Google Maps' data. Luz (2017) was the first to complement the São Paulo's O-D 2007 with Google's data as far as we know, however he dealt with some downloading constraints and accomplished it only for trips motivated by work occurring during the morning peak. We were able to download Google data to all PMU 2012's observations, although our main regression sample consists only of trips motivated by work as well. Data used and other details of the model will be more in depth discussed throughout the next sections of this thesis.

Regarding the specific Waze event studied here, we are unaware of published work with similar objectives. In the broader social sciences, there is more qualitative discussion accompanied by descriptive data showing evidence and discussing the consequences of the spread of Information and Communication Technologies (ICT) on traffic in the forthcoming

years(SHAHEEN; COHEN, 2018; WEE; GEURS; CHORUS, 2013; DUTZIK; INGLIS; BAXANDALL, 2014). This literature studies GPS navigation and its consequences among other ICT innovations that are changing travel behavior. From the other innovations, the internet platforms of commerce, services and telecommuting have been more largely explored and deeper empirical investigations were developed (HANDY; YANTIS, 1997; MOKHTARIAN; HANDY; SALOMON, 1995; MOKHTARIAN; ORY, 2005).

Focusing on GPS navigation, the engineering literature has been concerned with technical issues concerning the improvement of converting the data available into better traffic management (PATIRE et al., 2015; HERRERA et al., 2010). Despite the availability of engineering estimates of traffic flow and travel duration as in the simulation we cited above, or by field experiment (HERRERA et al., 2010), its objective is usually to better measure travel time from GPS (HAGHANI et al., 2010; WESTERMAN; LITJENS; LINNARTZ, 1996; WRIGHT; JOY, 2001), not to empirical evaluate time savings in a complete urban system, let alone welfare assessing. Narrowing to Waze application, Faria (2014) discusses crowdsourcing engagement under the social communication perspective based on Waze's case of success. And, under informatics lens, Silva et al. (2013) characterize the activity and information provided by this crowdsourced network.

In Economics, there is a significant number of papers whose objective is to evaluate consumer surplus effects from an industrial innovation or regulation change and some are closely related to ICT evolution (QUAN; WILLIAMS, 2017; CRAWFORD; YURUKOGLU, 2012; MORTIMER, 2007). The most similar work in objective and object of study from ours is Uber's consumer surplus paper by Cohen et al. (2016), however it is very different methodologically. Its objective is to estimate welfare effects for the use of a transportation application, Uber, the most popular ride services provider that has shaken markets worldwide.

In terms of methodology, Cohen et al. (2016) benefit of a rich Uber's dataset to estimate demand across different levels of price, extrapolating the traditional local elasticity estimate to multiple elasticity estimates upon a price range. Although their approach is very innovative, it would not fit our problem of estimating consumer surplus considering

multiple alternatives of transport. Furthermore, we are interested in welfare effects from congestion, derived from travel duration counterfactuals, not in direct consumer surplus from price variation. Knowing price elasticities is a mean to monetize the utility for us, not our end. Thus, despite some similarities in scopes, our analysis is completely different.

1.3   Thesis Structure

After this Introduction, the theoretical aspects of the microeconomics of congestion are discussed and, in sequence, the econometric model of travel demand is described in Section 2. There is a brief presentation of the study area and of the main local institutions linked to traffic in Section 3. The Section 4 is dedicated to describing the 2012 Mobility Survey and the Google Maps data to better depict the empirical work and what information it provides. The estimates of the econometric models are presented in Section 5. The subsequent Section 6 describes the simulation concerning congestion and its result. Section 7 relates the consequences of Waze's malfunction and presents the empirical study derived from that. Finally, Section 8 concludes the thesis.

2   Theoretical Framework and Empirical Implementation

2.1   The Microeconomics of Congestion

Two models dominate the economic literature on urban congestion: the speed-flow model, presented by Walters (1961), and the bottleneck model, presented by Vickrey (1969) and improved in Arnott, Palma & Lindsey (1993) . The speed-flow model is static and has some simplifications, such as uniform traffic in space and equal costs across individuals, which makes it difficult to accommodate hypercongestion or dynamic decision making. The dynamic bottleneck models try to overcome these shortcomings but they require a substantial amount of input data, such as the specification of road networks. Thus, the empirical implementation of a bottleneck model is severely constrained by data availability.

As a consequence, most empirical studies use the speed-flow model (PARRY, 2009, p. 8), which is simpler but satisfies most purposes. Our assumptions are thus based on this

model, especially the articles by Walters (1961)'s and Parry (2009). We assume that traffic conditions can be measured by flow of vehicles. The flow is the product of two variables: average speed, measured in kilometers per hour, and the density of vehicles on the roads, measured in vehicles per kilometer of lane:

$$\text{Average Speed (km/h)} \times \text{Density (V/km)} = \text{Flow (V/h)} \tag{1}$$

At a first stage, additional vehicles getting into traffic increase density that results in a growing flow. On the other hand, the relationship between density and mean speed is negative. The higher the density of vehicles, the slower one must drive to maintain a comfortable and safe distance from the vehicle in front. Thus, there is a moment when an extra vehicle leads to a drop in speed larger than the marginal increase in density, which yields a decreasing flow. At that point, it is said that the road capacity has been surpassed and that a congestion has taken place. The congestion often occurs at peak times, in which the demand for trips keeps pushing the traffic even if the road capacity level has been reached.
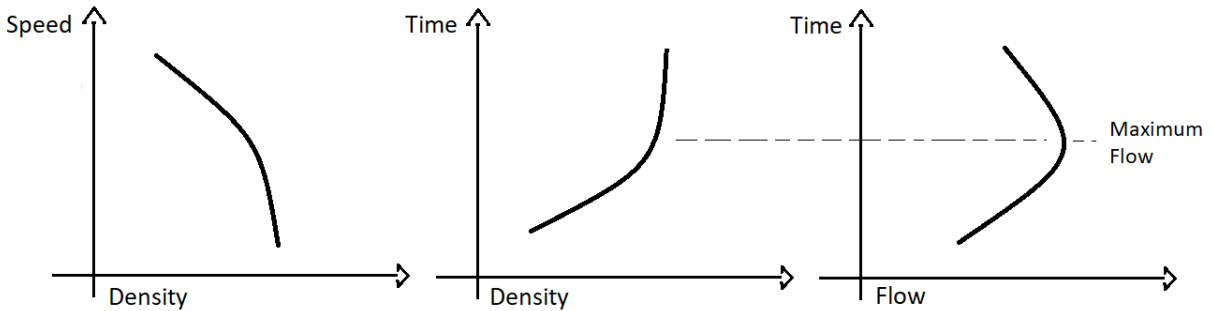
Figure 1 – Speed-Flow Model

The speed-flow model can be directly associated with an economic partial equilibrium model. There is a demand for travel that depends on prices, appointments and personal preferences. There is also a cost of these trips, which, in addition to direct expenses such as public transport tickets or fuel and vehicle depreciation, depends on the time spent on travel. Time, in turn, depends directly on the flow of vehicles, which becomes negatively related to demand when there is congestion.

Social arrangements make most users commute at similar times (peak hours) – work time, school time - that can be straightforwardly associated to demand shocks. On the opposite side, increased demand can lead to congestion, which may be linked to cost shocks as the cost of commuting under congestion is higher. First, because constantly starting and stopping engines causes more fuel consumption. Second, because the stressful and frustrating situation of standing in middle of other slow vehicles and spending more time for travel than could be expected leads to disutility. It is a matter of paying a higher price per distance travelled.

Based on that, we can draw a partial equilibrium model in which the travel demander loses well-being due to limited road capacity. Such loss of well-being is what we want to estimate according to the methodology detailed in the next session. The figure below illustrates the model. AA' is the off-peak demand and BB' is the peak demand, which is bigger than AA' irrespective of the cost. CC' is the cost curve for flowing traffic and DD' is the cost curve under congestion. The colored parallelogram is the loss of consumer surplus in Marshallian terms.
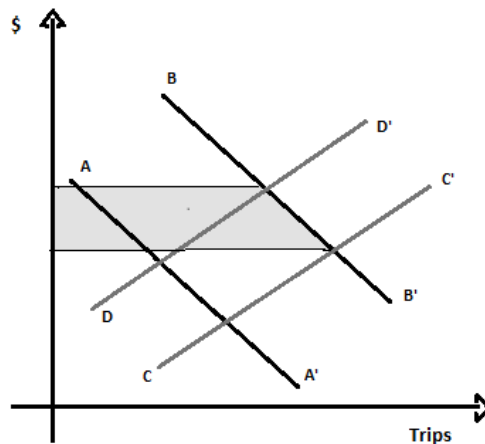
Figure 2 – Travel Partial Equilibrium

Consumer surplus is by definition the net utility derived from consuming a product, measured in monetary units. We assume that consumers make the best choice they can in response to the information available. Based on that, we model the commuter's choice of mode of transport and infer consumer utility, as we explain in detail in the next section. Keep in mind the theoretical model presented in this section is a simplified framework

to better understand the economic analysis regarding traffic congestion. The structural model estimated involves a set of variables that could influence the choices, which are more than a 'price - number of trips' pair.

Also, for clarity, we have to disclaim we are not estimating the externalities. To exactly estimate externalities, we would need to full estimate the cost curve and the demand curve. The working paper by Akbar & Duranton (2017), come to our knowledge after the thesis defense, is proposing a novel approach to deal with that. Here we are estimating either the expected and non expected costs of vehicles imposing slowness on the others, or observed increased costs plus externalities[4]. That is to say, our approach captures the whole welfare effect from not having enough space to the trips demanded, which is also driven by externalities but is not their net effects.

2.2   The Travel Demand Model

The empirical analysis applying a utilitarian framework resembles the seminal works of McFadden (1974) and Adler & Ben-Akiva (1979). These authors have developed transport demand models by incorporating microeconomic theory.

Historically, travel demand models made by engineers and other traffic planners follow two traditions (see Bhat & Koppelman (2003) for more details): the first and oldest is known as the Trip-Based Approach, which is statistically oriented, while the second, called Activity-Based Approach, tries to improve upon the first by using behavioral theory.

The traditional Trip-Based Approach takes every trip that the individuals realize as units of observation, that is, if a commuter is going to work and strategically stops at the grocery midway through, the larger trip is modelled as two independently demanded trips. The interaction between the two trips, which could even be seen as one, is not taken into account. For example, at the decision time in which a person is deciding which mean of transportation to use to go to the grocery, he was already envisioning that he would

---

[4]   The optimal demand for trips during the peak time (given by the point that marginal costs intersect the demand) is above the equilibrium point reached during free flow, but also above the equilibrium demand during peak time (where average costs intersect the demand). In practice, it means commuters know they are paying a higher price when they are not traveling with free flow, despite they do not know how much the extra costs that they are imposing and being imposed on are (the externalities).

go to work afterwards. And, after the stop at the grocery, the person does not choose a mode of transport again, as he had already planned before. But the Trip-Based Approach considers that there were two independent car driving decisions, a double counting that hinders the analysis of different traffic management policies.

The Activity-Based Approach, or Behavioral Oriented Approach, has emerged from the contribution of psychology and economics to transportation science (MCFADDEN, 1974). The demand for travel came to be seen as the aggregation of choices of individuals, induced by their needs to participate in different activities distributed in space. In this way, social interactions, personal characteristics and the influence of the environment are factors that can determine the behavior of the demand for travel. Using time intervals as units of analysis (which can be entire days or periods of the day, for example), the activity-based approach accommodate trips interactions and allows the identification of real trips' ends. In the words of Bowman & Ben-Akiva (2000):

"The most important elements of activity-based travel theory can be summarized in two basic ideas. First, the demand for travel is derived from the demand for activities. (...)Travel causes disutility and is only undertaken when the net utility of the activity and travel exceeds the utility available from activities involving no travel. Second, humans face temporal spatial constraints, functioning in different locations at different points in time by experiencing the time and cost of movement between the locations (HäGERSTRAND, 1970). They are also generally constrained to return to a home base for rest and personal maintenance."

From the behavioral approach, the demand for transport can be related to Consumer Theory, where rational individuals are assumed to choose the option that maximizes their utilities. The determinants of utility are then estimated from the choices made by groups of individuals. However, due to the lack of full information about the decision process, it is necessary to add a stochastic factor to the deterministic component. This term is a way to account for the uncertainty that may arise from unobserved attributes, unobserved taste

variations, measurement errors, or instrumental variables included that do not accurately represent the true determinant (BEN-AKIVA; LERMAN, 1985). As such, an individual i derives utility from a mean of transport j, as represented below:

$$U_{ij} = V_{ij} + \epsilon_{ij}\,, \qquad j = 1, 2, ..., m, \tag{2}$$

In which V_{ij} is the deterministic component of utility, depending on : $V_{ij} = \boldsymbol{x'_{ij}\beta_i}$ for observable characteristics that vary according to transport alternative and individual (e.g. ticket cost, fuel cost, travel time) or $V_{ij} = \boldsymbol{z'_i\beta_j}$ for observable characteristics that vary among individuals but have no dependency on the transport alternative (e.g. gender, income). ϵ_{ij} captures the random component and unobservable factors. Individual i chooses $U_{j=J}$ if, and only if, $U_J \geq U_j \, \forall j \neq J$. Thus, from a probabilistic perspective for an individual:

$$
\begin{aligned}
Pr[y = J] &= Pr[U_J \geq U_j, \quad \forall j \neq J] \\
&= Pr[U_j - U_J \leq 0, \quad \forall j \neq J] \\
&= Pr[(V_j + \epsilon_j) - (V_J + \epsilon_J) \leq 0, \quad \forall j \neq J] \\
&= Pr[\epsilon_j - \epsilon_J \leq V_J - V_j, \quad \forall j \neq J] \\
&= Pr[\epsilon'_{jJ} \leq -V'_{Jj}, \quad \forall j \neq J]
\end{aligned}
\tag{3}
$$

Therefore, our estimates depend on the distributions we assume for the components of the model, especially for the error term. Here we assume that ϵ has a random logistic distribution $L_{ij}(\beta_i)$. In addition, we assume that β_{time} and β_{cost} - parameters associated to travel time and travel expenditures respectively - vary over our population with normal density $f(\beta) \sim N(b, W)$. Thereby, we have:

$$
\begin{aligned}
Pr[y_i = j] = p_{ij} &= \int L_{ij}(\beta_i) f(\beta) d\beta, \qquad j = 1, ..., m. \\
p_{ij} &= \int \left(\frac{e^{V_{ij}(\beta)}}{\sum_{l=1}^{m} e^{V_{ij}(\beta)}} \right) \phi(\beta|b, W) d\beta, \qquad j = 1, ..., m.
\end{aligned}
\tag{4}
$$

Introduce m binary variables for each i such that:

$$
y_{ij} = \begin{cases} 1, & \text{if } y_i = j, \\ 0, & y_i \neq j. \end{cases}
\tag{5}
$$

Thus, we can fully characterize the likelihood function of the decision process as:

$$
L_N = \prod_{i=1}^{N} \prod_{j=1}^{m} p_{ij}^{y_{ij}}
\tag{6}
$$

This model was called Mixed Logit by Train (2009) and is sometimes cited as Random Parameters Logit (CAMERON; TRIVEDI, 2005). A good description of microeconomic foundations of random utility models (RUMs) is found on Ben-Akiva & Lerman (1985) and a proof of how mixed logit can approximate any RUM is found at McFadden & Train (2000). The standard logit is a special case of mixed logits when $f(\beta)$ is a degenerated distribution that equals 1 if $\beta = b$ and 0 if $\beta \neq b$. For comparison, we estimate a standard logit as well, but our preferred estimator is the Mixed Logit for many reasons. First, because in comparison to a standard multinomial logit it allows for intra-cohort preferences, which means that individuals can have the same income but valuate costs and time differently. In addition, it allows for unrestricted substitution patterns and correlation in unobserved factors over time, which better approximates reality.

The Nested Logit is a formulation that also relaxes the independence of irrelevant alternatives hypothesis (IIA, (MCFADDEN; TYE; TRAIN, 1977)), yet we do not think that there are clear nested structures to be modelled. Moreover, taste variation is a more important feature to be retained. Finally, as compared to a Probit model the Mixed Logit is not restricted to a normal distribution and the estimation is computationally simpler (TRAIN, 2009). We have estimated the mixed logit in Stata™ software, which implements Train (2009)'s methods. It integers probabilities p_{ij} by simulation to build-up a simulated log likelihood (SLL). The estimator is the vector θ that maximizes the SLL. For the standard logit under comparison, we have estimated an alternative-specific conditional logit proposed since McFadden (1974), also implemented in Stata™.

Independently of the chosen distribution class, working with a discrete choice model makes welfare analysis possible. (CAMERON; TRIVEDI, 2005, pp. 507) Welfare effects are estimated in terms of the consumer surplus, also called accessibility measure (BEN-AKIVA; LERMAN, 1985). As stated in the subsection 2.1 , the consumer surplus is the utility derived from a choice in monetary units. Therefore, we can define the surplus for consumer i as $CS_i = \frac{1}{\alpha_i} max_j(U_{ij} \quad \forall_j)$ (TRAIN, 2009, pp.59), where α is the marginal utility of income. Dividing by α monetizes the utility, since $\frac{1}{\alpha_i} = \frac{dY}{dU}$, the income value of an utility marginal unit. Here we are particularly interested in estimating the value a person attributes to one additional minute in traffic, which we denominate Value of Time (VOT):

$$VOT_i = \frac{dY}{dU}\frac{dU}{dt} \tag{7}$$

We can compute individual VOTs from our parameter estimates. Since the utility variation from a one-dollar costs reduction should be equivalent to a one-dollar increase in income (in both cases one ends up with an extra dollar), we have $\frac{dU}{dY} = -\frac{dU}{dc}$ and thus $\frac{dY}{dU} = -\frac{dc}{dU}$. We use this formulation to translate utility into money. It is not possible to utilize income directly for that, because it is constant across alternatives (BEN-AKIVA; LERMAN, 1985).

Moreover, we want to estimate the congestion effects on welfare. Ergo, we have to compute the aggregated consumption surplus loss:

$$\begin{aligned}\Delta CS &= \sum_{i=1}^{n} \frac{1}{\alpha_i}\frac{dU_i}{dt_i}\Delta t_i \\ &= \sum_{i=1}^{n} VOT_i(t\text{congestion}_i - t\text{nocongestion}_i)\end{aligned} \tag{8}$$

For statistical reasons and sampling limitations, we model the decisions aggregating the alternatives of transport into six groups. The aggregation was based on vehicle typology logic and mainly on expenditures logic, as we are modeling demand. Of course, there is some arbitrary at choosing any aggregation rule and some features are prioritized to the detriment of others. We made our imperfect choice, that we found better than other options.

Our alternatives are $j = 1$) Buses, 2) Railway, 3) Driving, 4) Motorcycle, 5) Taxi, 6) Shared Ride, Walking or Bike. In the section 4.1 we come back to the discussion of this point and show the distribution of these alternatives in the survey, now we only want to illustrate our utility model. For any of these alternatives, we assume individual derives a utility U_{ij} as shown above. And, as already commented, this utility has a deterministic component divided into variables that vary according to the alternative and the individual (e.g.: cost and time), called alternative-specific, and variables that vary according to the individual but not to the alternative, called case-specific. For a case-specific variable, we estimate one associated parameter for each transport alternative ($z_i\beta_j$). For an alternative-specific, we can assume some parameters have a distribution dependent on the representative individual - the mixed logit - or have a population parameter - standard logit ($x_{ij}\beta_i$ or $x_{ij}\beta$). We model the deterministic component of U_j using the following variables:

Alternative-specific ($\boldsymbol{x_{ij}}$):

- Financial costs (c)

- Travel duration (t)

- c / Household Income (I)

- t * I

  Case-specific ($\boldsymbol{z_i}$):

- Gender of traveler

- Household Income

- Age of traveler

- Age^2 of traveler

- Dummy for peak time (Departure Times: 6-8a.m, 12-2p.m, 5-7p.m)

- Dummy for origin or destination point being downtown

- Dummy for car out of home in the moment of the trip

In our mixed logit is assumed the financial costs and the travel duration parameters are normally distributed across the population, but the parameters for the interaction of both with income do not vary. Therefore, for a given mode of transport j, the individual's utility and VOT is:

$$U_{ij} = c_{ij}\beta_{ci} + t_{ij}\beta_{ti} + c_{ij}/I_i\beta_{c/I} + t_{ij}I_i\beta_{tI} + \boldsymbol{z_i'}\boldsymbol{\beta_j} + \epsilon_{ij}$$

$$VOT_i = -\frac{dc_{ij}}{dU_{ij}} \times \frac{dU_{ij}}{dt_{ij}}$$

$$= -\frac{1}{\beta_{ci} + \frac{\beta_{c/I}}{I_i}} \times (\beta_{ti} + I_i\beta_{t.I})$$

$$= -\frac{\beta_{ti} + I_i\beta_{t.I}}{\beta_{ci} + \frac{\beta_{c/I}}{I_i}}$$

$$(9)$$

From this derivation, we know how to use the estimated regression to assess the welfare effects. The VOT is combined to the trip to analyze the traffic impact, as we detail in Section 6. The data we have applied to estimate the model and the congestion effects is presented in the Section 4. Section 5 provides a discussion of the results. Before that, the area under analysis is presented in the Section 3.

3   Study Area

São Paulo is the most populous municipality in Brazil and also capital of the state of São Paulo. Founded in 1554, it has an estimated population of 12.1 million inhabitants (IBGE, 2017a), of which 99.1% live in the urban zone of the municipality – the city of São Paulo. The São Paulo Metropolitan Area (SPMA), also known as Greater São Paulo, is organized into 39 municipalities (Figure 3) and has 21.4 million inhabitants (IBGE, 2017a).

For simplicity, the toponym "São Paulo" will henceforth be used in reference to the city and its adjunct urban agglomeration that sprawls the metropolitan area. To emphasize

Figure 3 – São Paulo Metropolitan Area

the whole metropolitan area, we will often use SPMA. When the intention is to refer to the strict São Paulo municipality area, we add the word "municipality", as we do when mentioning the state. This can be confusing because the metropolitan dynamics revolves around the main center in the city of São Paulo, and some outlying municipalities are completely integrated with the São Paulo municipality.

Here we present some important characteristics of São Paulo, as well as its main transport administration institutions.

The gross domestic product (GDP) of the municipality, R$650.5 billion in 2015, is around 11% of the entire Brazilian GDP and 34% of all production of goods and services in the state of São Paulo. In that same year, the SPMA's GDP amounted to R$1056.9 billion GDP, corresponding to approximately 18% of the Brazilian total and to 54% of the State (EMPLASA, 2018; IBGE, 2017b).

Meeting the transportation demand of a city with the economic dynamism of São

Paulo is a great challenge. The city's public transport system is made up of municipal buses, intermunicipal buses, subway and trains.

At the municipality level, the urban road transportation system is supervised by the Municipal Department of Mobility and Transportation. Under the Department, São Paulo Transport, Inc. (SPTrans, in Portuguese acronym) is responsible for the management and inspection of the bus system. The actual operation of the services is outsourced to private companies working in consortiums that currently have more than 1,300 bus lines, with approximately 15 thousand vehicles(SPTRANS, 2018; Prefeitura da Cidade de São Paulo, 2018).

Under the command of the Municipal Department of Mobility and Transportation, there is also the Traffic Engineering Company (CET, in Portuguese acronym), whose objective is to plan and control the operation of the road system, in order to ensure greater safety and traffic flow within municipality limits.

The State of São Paulo Department of Metropolitan Transport is responsible for the São Paulo Metropolitan Company (Metrô, in Portuguese abbreviated name). The Metrô is in charge of São Paulo's subway. Currently, there are 5 lines operating with 155 trains over 71.5 kilometers of railroad across 64 stations (METRÔ-SP, 2018).

Complementing the subway, there is São Paulo Metropolitan Trains Company (CPTM, in Portuguese acronym) that was created from existing railroads in SPMA to manage over ground trains. The system is radial and each line connects the city of São Paulo to other city of the metropolitan area. Currently, there are 94 active stations in seven lines, which totalize 273 km of rail network (CPTM, 2017; CPTM, 2018).

The Urban Transportation Metropolitan Enterprise (EMTU, in Portuguese acronym) is also under the Department of Metropolitan Transport's oversight and is in charge of inspecting and regulating intermunicipal transport. EMTU supplies SPMA with bus lines connecting the city of São Paulo to its neighboring cities, some of those having dedicated lanes, called metropolitan corridors. The operation of the lines is delegated to private-finance initiative.

Finally, São Paulo Metropolitan Planning Enterprise, Inc. (EMPLASA, in Portuguese acronym) is a company directly under the governor's office, able to influence SPMA's traffic dynamics through its policies and projects of urban and regional development. It also carries out studies and provides cartographic products, geospatial information systems and technical knowledge about metropolitan planning to public and private managers and citizens.

In this thesis we study welfare effects within the Metropolitan Area by using the trips recorded in the Urban Mobility survey. The survey has a representative sample of trips for the whole area of the Figure 3. In the next section, the data of the survey is presented.

4 Data for the Travel Demand Model

4.1 The Mobility Survey

The main dataset we use in this work is the Urban Mobility Survey of São Paulo Metropolitan Area from 2012 (*Pesquisa de Mobilidade Urbana da Região Metropolitana de São Paulo 2012* - (PMU)). The PMU is an update of the Origin-Destination Survey (*Pesquisa Origem/Destino*- (O/D)) realized each 10 years since 1967, having the objective to rise information about population mobility patterns as an input to urban and transportation planning. In 2012, 8.1 thousand households were randomly selected and 32.4 thousand people were surveyed, covering for a geographic division of São Paulo metro area in 31 zones and population representative (METRÔ-SP, 2013). Among the information collected, there are the used transport modes, the traveling reasons, departure and arrival times, and also socioeconomic data that is very important to understand travelers background and then to model demand.

To this thesis, some PMU's info are particularly relevant. The first is the chosen mode of transport, because we are estimating demand through a discrete choice model. Except for walking trips, that are mostly done in short distances, we see in Figure 4 that car is the most used mean by population. Cars favor traffic jam, because they are not

dense occupied as buses or vans and are not thin as bicycles and motorcycles.
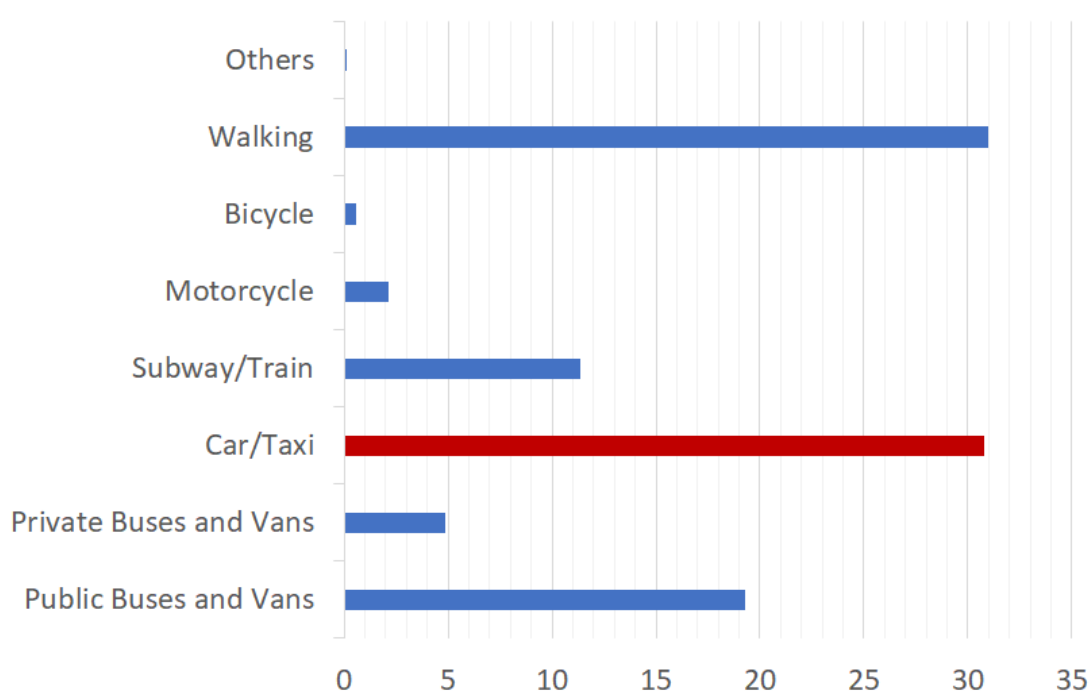
Figure 4 – Main modes of transport. Source: *Pesquisa de mobilidade da RMSP 2012*

To make the discrete choice computationally feasible, we aggregate the main transportation choice into six categories. As we are modeling demand, one of the main attributes one is likely to consider in his decision is the financial cost of the trip. Other features that influence decision are comfort, flexibility of departure time, speed and security. Thinking on that, the aggregation follows a vehicle typology logic and expenditures logic:

1) Buses

2) Railway (Subway and Train)

3) Driving

4) Motorcycle

5) Taxi

6) Shared Ride, Walking and Bike

All types of buses are aggregated into a single category, irrespective of whether

38

they travel only inside the city limits or not, whether they are micro-buses or larger buses, public or charters. Rail transportation is a different category. Car driving and taxi were separated because the second is a differentiated and much more expensive service. At last, shared rides, walking and bike compound a unique alternative because they are usually free. It can be seen in Figure 5 this category is the most demanded.
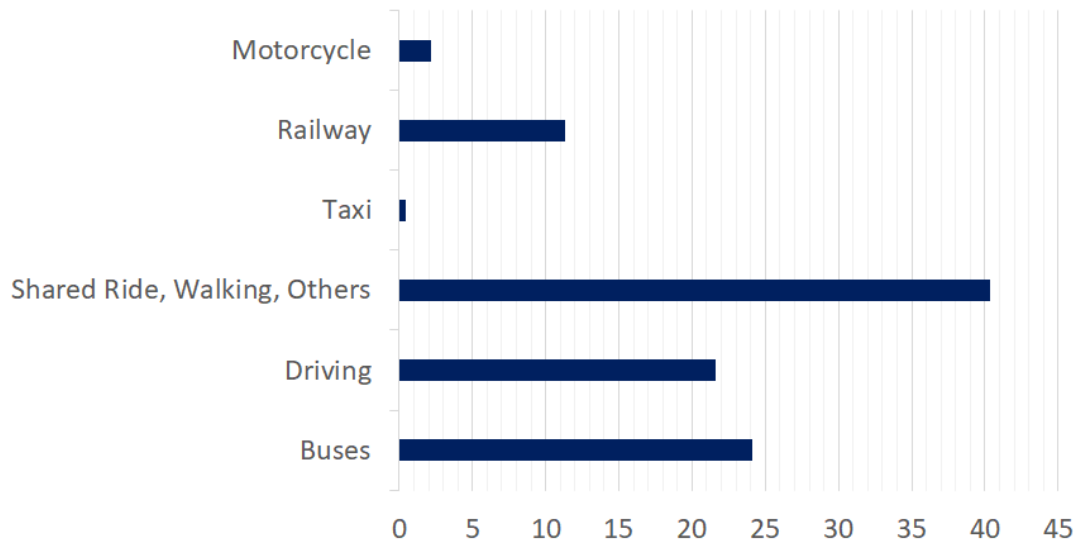
Figure 5 – Aggregated main modes. Source: *Pesquisa de mobilidade da RMSP 2012*

Another important characteristic of the PMU is that it records up to four changes of mode between the origin and destination points. In Figure 6 we see the transports most used as second mean in a single trip. The secondary modes are essential to understand modes integration and total cost of the trip. It can be seen that most of it are buses and the railway, so people changing mode along the trips are most likely to be public transport users.

Besides of the travel modes, we are concerned about congestion. Therefore, we have to check whether our data is likely to be representative of that, and, if so, when and where the gridlocks are likely to happen. In Figure 7, we see the departure time distribution, showing peak times during the early hours of day (6 and 7 a.m.), the lunch time (12 noon), and the end of the day (5 to 7 p.m.). This confirmed our expectations, gridlocks are likely to happen if lots of people travel at the same period. Also, they only happen if people travel through the same routes. By Figures 8 and 9, we see most of the trips in the Metropolitan Area have the capital as destination and, by far, the downtown zone is the
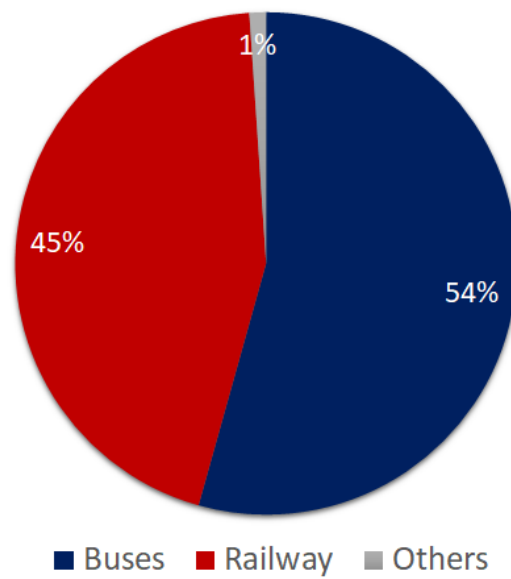
Figure 6 – Second modes taken when trips are not realized by a single mode. Source: *Pesquisa de mobilidade da RMSP 2012*

main destination.
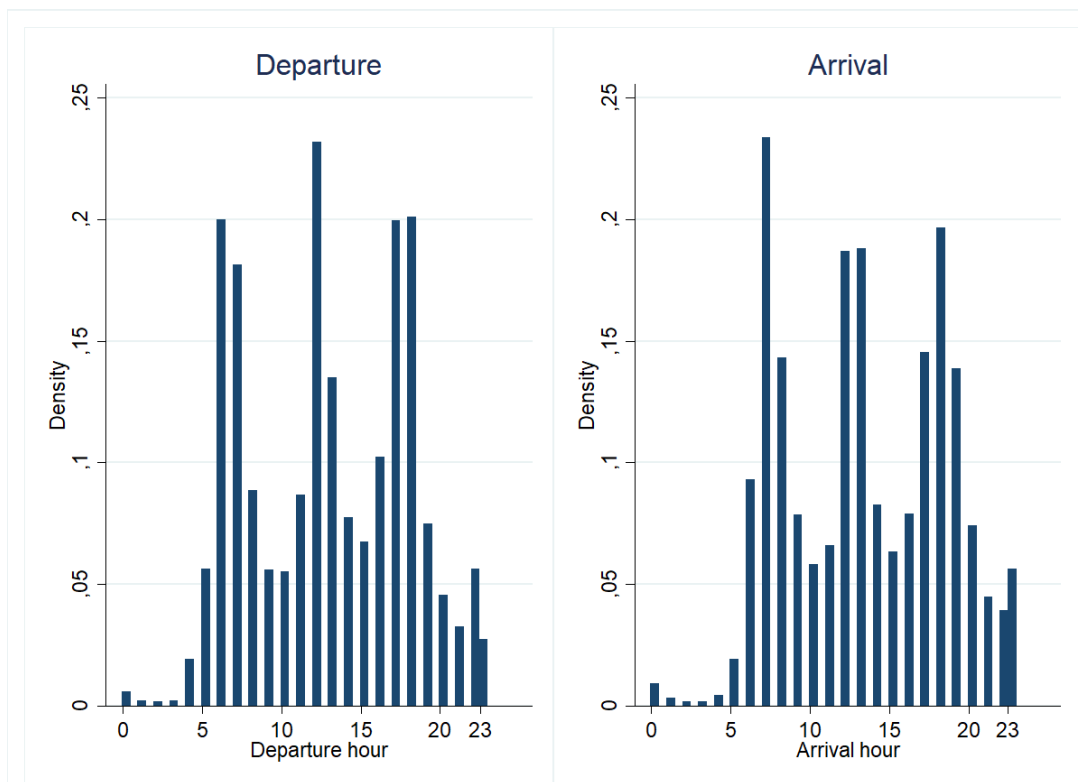
Figure 7 – Departure time distribution. Source: *Pesquisa de mobilidade da RMSP 2012*

The distributions of departure times and destination zones indicate that congestion is likely to occur. In Table 1, we have checked if the pattern predicted by the speed-flow
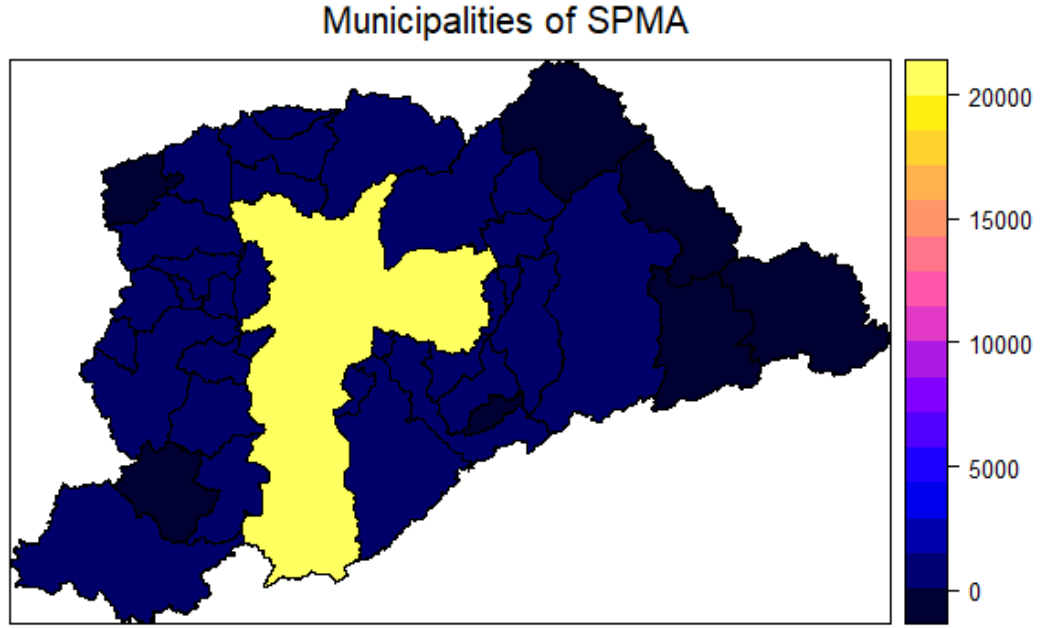
Figure 8 – Number of trips in the sample going to municipalities within SPMA - Source: *Pesquisa de mobilidade da RMSP 2012*

model is empirically evident. For comparison, we define an interval of distance equal to one standard deviation from the sample average distance ($]\overline{X}, \overline{X} + \hat{\sigma}]$), and we confirm that the average travel duration is shorter at non-peak hours than at rush hour. This corroborates the underlying idea that when many vehicles are getting into traffic, the route's capacity is overreached, the flow decreases, and travel times increase.

Table 1 – Average travel duration in average distance interval $]\overline{X}; \overline{X} + \hat{\sigma}]$

Departure Time	Average Duration (Min)
9 p.m. to 5 a.m.	56.5
Overall	60.7
Peak times (6,7,12,17,18)	63.6
Evening Peak (5 to 7 p.m)	69.3

Source: *Pesquisa de mobilidade da RMSP 2012*

Wages shall influence on the transport people are willing to pay for and on individual values of time . Considering a 8 hours work journey, twenty two working days a month, we show in Table 2 the descriptive statistics for gross individual income and gross familiar income - excluding missing and null values.

One of the features we have considered to estimate the model is the reason why
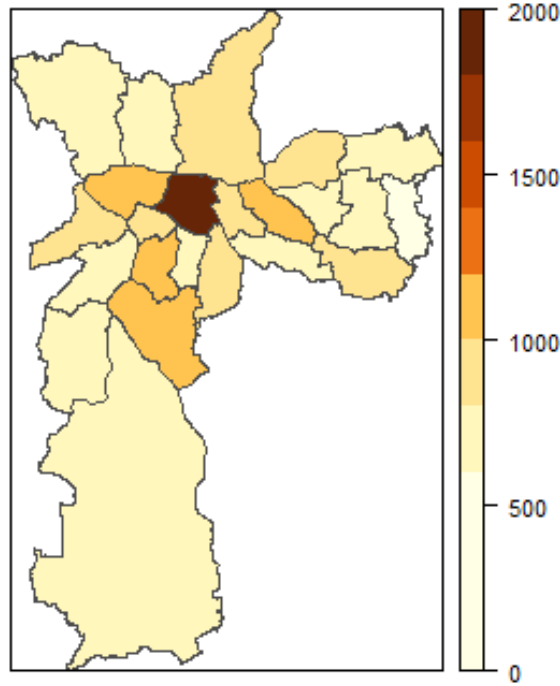
Figure 9 – Number of trips in the sample going to zones within São Paulo Municipality -
Source: *Pesquisa de mobilidade da RMSP 2012*

Table 2 – Gross Wage per hour (R$ / 2012)

	Individual Wage/ Hour	Familiar Wage[5]/ Hora
N	26105	26105
mean	11.34	24.58
sd	15.31	28.67
min	0.21	0.28
max	340.91	454.55

Source: *Pesquisa de mobilidade da RMSP 2012*

the trips are undertaken. Structural parameters can have different values according to the
underlying causes of the trips , for example, delays are less acceptable in work-related
trips than in pleasure-related trips, and this have to be taken into account in the utility
estimates. In Figure 10, we can see that most of the observations are commuting trips.

So far, we have presented a dataset that provides characteristics of the trips and
socioeconomic information, allowing for several types of analysis. Descriptive statistics for
all the variables used in the modeling are provided in the Appendix A and more can be
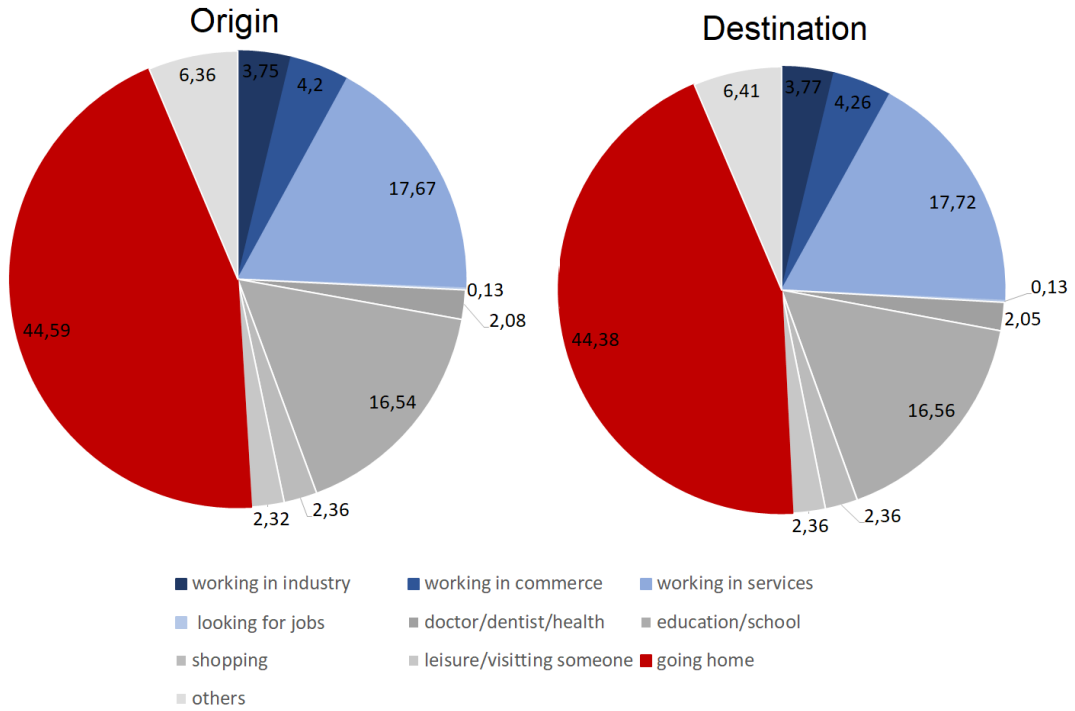
Figure 10 – Travelling reasons at origin and destination. Source: *Pesquisa de mobilidade da RMSP 2012*

found in the Survey Report (METRÔ-SP, 2013). Yet the PMU lacks of some important data that are indispensable to estimate the discrete choice model and to do the welfare analysis, such as direct travel costs, traveled distances, and potential travel times for unchosen alternatives. We fulfill these gaps by using Google Maps data and some secondary sources as described in the next two sub-sections.

4.2   Google Maps API

One common problem in using origin/destination surveys to estimate discrete choice models is that surveys have no attributes for unchosen modes. For example, when a surveyed person has chosen the bus as her mode of transport, the survey records the time she spent at the bus and does not report the potential duration of the trips she could choose (for example, the duration of the same trip in a car). Discrete choice models of demand require the specification of variables for the unchosen modes since the alternatives need to be compared (mathematically, equation 4's denominator depends on V_{ij} for all kinds of transport j that individual i has available).

There are four ways that the literature proposes to address this lack of information (LUZ, 2017; HENSHER; ROSE; GREENE, 2005). The first is to use sample averages and medians to fill in the blanks. The second is to use the distribution of the variable of interest across the chosen modes. This can be achieved with any matching technique. The third is to implement an extra survey presenting hypothetical choice options to people and inferring values to populate the dataset. The fourth is called data synthesis and consists of predicting data conditional on observed variables. The most known and used method to do that is regression.

The survey method is resource-consuming and requires a lot of planning ahead. The other methods can result in substantial errors, and the more variability in the data the larger the error from using averages or matching. In addition, there is likely to be selection bias on the estimates, even if matching or regression is used. For example, if walking trips are mostly short distance and consequently done at short travel times, the fitted values will underestimate the walking duration for long distance. Because of feasibility and cost-benefit, different authors have used econometric regressions to estimate the missing data in recent years (LUCINDA et al., 2017; BARCELLOS, 2014; FERES, 2015).

Yet a more interesting alternative has recently emerged: the use of a Google Maps API to obtain the missing data (JAVANMARDI et al., 2015; LUZ, 2017). Google Maps has made its Application Programming Interface publicly available, so it is possible to collect data from Google's data bank in a large scale. Google's data is generated by GPS and chronometer technologies as it come from users that navigate using the Google Maps application, so the data is highly accurate (it might be even superior to self-declared information). Compared to estimates generated by econometric OLS[6] estimators, we expect that the Google data will address the selection bias problem.

We request information to Google API as if we were to travel from the point where the PMU's trip begins to the point it ends. We include queries by car, by bus, by railway or by walk. Google returns the estimated travel duration and a set of fundamental variables that were unavailable in the original survey dataset, such as public transport costs, travel

---

[6]   Standard initials to refer to Ordinary Least Squares

distance and walking distance to bus stop or rail station. Besides time, financial cost is another essential variable that we need to specify in the travel demand model and we use Google's data to fill that. This is an improvement compared to previous models in the literature where expenditure estimates were produced based on straight line distances (LUCINDA et al., 2017; BARCELLOS, 2014; FERES, 2015). Additionally, distance can be used to estimate average speed. The distance to a bus stop or station was tested in the demand model specification, but it had no significance to explain speed. Our Google Maps queries cover for the whole georeferenced observations we had in dataset, 46,861 trips.

Google Maps only accept queries for present or future dates, not allowing access to historical data. Thus, we are susceptible to missing changes in street configuration, infrastructure alterations and traffic dynamics. In one hand, we would expect travel duration to be higher today, as the fleet of vehicles is growing. On the other hand, we find that the average traffic speed as measured by CET actually increased from 2012 to 2016, taking into account all the four period-direction combinations available (morning: downtown-suburb or suburb-downtown and afternoon: downtown-suburb or suburb-downtown)(CET, 2013b; CET, 2017), so it is less likely we are overestimating travel times for 2012's data. Moreover, there were few reported trips in modes for which Google Maps could not provide an answer. In Table 3 we show the number of queried trips that Google missed stratifying by our aggregated modes of transport.[7]

Table 3 – Queries by mode and trips for which Google Maps API could not provide an answer

Mode	Freq.	Miss.	%
Bus	11297	1619	14.33
Rail	5325	306	5.75
Driving	10121	9	0.09
Motorcycle	998	1	0.10
Taxi	220	0	0.00
Walking	14504	2	0.01

---

[7] When attributing a potential time to our free cost category (Shared Ride, Walking and Bike), we always indicated the walking information. The reasons are there is more uncertainty surrounding shared ride and bikes, for example, a shared ride can completely deviate from the expected way between origin-destination to pick up people in different points along the way. Bikers can use sidewalks or streets, whether Google is expecting they rather use bicycle lanes etc. And also going by foot is an option that is almost always available and does not rely on the availability of a car or a bicycle.

4.3   Travel Costs

There are no expenditures by trips recorded in the survey. This is one of the fundamental variables for a demand model and for our analysis, so we need to retrieve it using different strategies. As mentioned in 4.1, one PMU trip can include up to four intermediary steps in different modes of transport. We calculate the financial cost for each step independently and sum over all the costs to produce the full cost by trip.

Car driving

For the cost of traveling by car we use the formula suggested by Lucinda et al. (2017). The travel cost through the track l in a trip realized at the month m is:

$$C_{lm} = \left[\frac{d}{t} \middle/ \frac{d}{g}\right] \times p_{gm} \times t_l$$

where d is the distance traveled, t is the time spent, g is the gasoline consumption and p_{gm} is the gasoline price in the month of travel. The first term of the division inside the brackets is the speed and the second term represents the energy efficiency. Since d cancels with d, the resulting quotient in the brackets is the consumption of gasoline (measured in liters) per unit of time (hours). This is multiplied by the gasoline price (R$ / l) and by the travel duration (hours) to yield the disbursements with car trips. The PMU has no distance information, except for straight line distances computed between the initial and arrival points. Thus, we estimate speed using the distance indicated by Google Maps for its suggested route between the departure and destination points, divided by the travel duration as reported in the survey. The energy efficiency is calculated from INMETRO (National Institute of Metrology, Quality and Technology information of the National) as an average of the performance of different car models weighted by the number of models sold units in the period 2008-2012 (Inmetro, 2017). The data for car sales in the period is taken from Anfavea (2017) (National Association of Manufacturers of Automotive Vehicles). For gasoline prices, we use the monthly average retail prices in Sao Paulo as recorded by ANP (2017) (National Petroleum Agency). Finally, as in the speed estimates, the last term is the travel time as reported in the survey.

Motorcycles

For motorcycles, the formula is the same as for cars, with differences in the data sources used. The energy efficiency is gathered from specialized websites (MotoClube, 2017; Motonline, 2017; UOL Carros, 2017; Consumo Combustível, 2017) (to the best of our knowledge, INMETRO does not run performance tests for motorcycles). The sales data used in the weighted average come from Fenabrave (2017) (National Federation of the Distribution of Motor Vehicles).

Taxi

Taxi fare values are available on the website of the municipality of São Paulo (Prefeitura da Cidade de São Paulo, 2012). The service charge included a fixed rate of R$ 4.10 plus variable travel time factor of R$ 33.00 / h plus a distance-based factor of R$ 2.50 / km on Fare 1 or R$ 3.25 / km on Fare 2 (8 p.m. to 6 a.m. and weekends).

The Metro and Train

For routes traveled by subway or train, the price of R$ 3.00 was charged by the São Paulo Metropolitan Trains Company (*Companhia Paulista de Trens Metropolitanos*) in 2012. In addition, trips preceded by another train or subway transfer have zero cost and tickets preceded by buses from the São Paulo Municipality are priced at 20% discount or R$ 2.40.

Municipal Bus

Buses in the city of São Paulo were also charged according to the "*Bilhete Único*"(Single Ticket) policy, with R$ 3,00 being charged for the first trip and zero for bus trips made thereafter. In case of integration with railways, the discount 20% is applied.

Chartered Bus

For privately hired buses, the value of R$ 1,548 / km divided by 10 was adopted, considering the average capacity of passengers per vehicle. This price was provided by one of the most important companies in the sector. For private school buses, we added the condition that the minimum price is R$ 3, the price of a bus ticket.

Inter Municipal Bus

The most difficult part of inputting the cost data relates to the metropolitan and intercity bus lines. The 2012 Charge Report of the Municipal Transportation Company of São Paulo (EMTU, 2012), which is responsible for supervising this type of transportation, contains information on 1,231 bus lines. We developed an algorithm that matches inter-municipal trips and line fairs in a four-stage procedure. In the first stage, the fare was matched whether the origin and destination zones of the line were exactly the same as the ones reported in Mobility Survey. To assign price across zones where multiple values were available, we computed an average price[8]. In the second stage, the municipalities in the SPMA were aggregated according to their geographical position in relation to the capital (as in Figure 8). If the trip price had not been filled in the first step, it is matched with the average price on routes at this more aggregated scale. After that, for the prices still missing, the zones within the Municipality of São Paulo are aggregated into nine different regions (as in Figure 9) and we match using average prices of lines across these aggregated zones. Finally, for the remaining missing observations, R$ 3.96 was assumed as the average price for all lines.

4.4   Travel Durations and Financial Costs for Unchosen Alternatives

As anticipated when discussing the Google Maps data (Section 4.2), we need to input data for unchosen alternatives. Since our model has six aggregated alternatives (Bus, Railway, Driving, Motorcycle, Taxi, Walking/Bike/Shared ride), even if the commuter chooses only one we have to fully specify information for the alternative options in the

---

[8]   excluding the "Selective" and "Executive" classes, more expensive and differentiated services

model. The alternative-specific variables are time, costs and interactions between these and income. As income is individual-specific, we have to input only potential time and costs.

Travel Durations

We use the travel durations provided by Google as our proxies. Google provides bus duration, railway duration, walking duration and two kinds of car duration: a general average and one that depends on traffic conditions and period of the day. We match bus and railway with their respective values. For driving, motorcycle and taxi we assign the car traffic duration, which could better predict trip durations than the traffic-independent duration. For the free expenditure alternative, we used the walking time as a general indicator of a freely available option (to most people).

Financial costs

For disbursements to unchosen alternatives, we use the guidelines of Section 4.3, but inputting Google's information. For bus and railway alternatives, we use the total price provided by Google corrected by the ticket's inflation rate of 26.67% between 2012 (year of the survey) and 2017 (when we queried Google). For driving, motorcycle and taxi alternatives, we use the formulas in Section 4.3, replacing surveyed travel duration by estimated car travel duration under traffic.

Descriptive statistics of the variables when unchosen alternatives are included in the dataset are available in the Appendix A.

5   Travel Demand Model Estimated Results

5.1   Discrete Choice Model of Travel Demand

We estimate mixed logit models for travel demand of transport modes as presented in Section 2.2. As discussed in Section 4.1, we aggregated alternatives into 1) Bus, 2) Railway, 3) Driving, 4) Motorcycle, 5) Taxi, 6) Shared Ride, Walking and Bike. We have

alternative-specific and case-specific independent variables. Alternative-specific variables reflect the costs of each alternative in monetary terms as well as in time opportunity. Case-specific variables concern socio-economic individual characteristics and conditions associated with the places being travelled to or from. The list of variables used was presented on Section 2.2 and can be found at the tables of the present section.

The literature recommends the estimation of different structural models according to the travel motivation, since the valuation of utility can differ substantially (BEN-AKIVA; LERMAN, 1985). Think, for example, about being late when going to a job meeting and when going to the beach. The first seems more stressful and one minute less in traffic can be much more costly. Because of that, we have a main specification where the purpose for all observations is coded as work for the origin or destination. In this model we also allow time and cost parameters to vary normally.

Besides the main specification, which follows the literature and includes the most reasonable assumptions, we run two additional regressions for robustness check. One is the standard multinomial logit, where there is no varying-parameter, and the other is an extra mixed logit that uses all the trips in the survey, independent on the motivation. The latter specification controls for travel purposes with a dummy for educational motivated trips and a dummy for labor motivated trips.

Table 4 shows the coefficients estimated for the alternative-specific variables and Table 5 is for the case-specific variables. Almost all coefficients are significant under a 95% significance level and there is no important difference in magnitude or direction among the three models. Moreover, most of the coefficients have the expected sign – larger costs and trip duration result in larger disutility.

When cars are away of the houses, people are less likely to choose any other mode of transport. The female effect is that women are more likely to take a bus or the subway than get a free alternative, but less likely to go by car, motorcycle or taxi, which reflects still prevalent cultural and otherwise disadvantages for women in Brazilian households. At peak times, people are relatively more likely to choose the free alternative, which includes bike and walking, than the other modes, which probably is a consequence of congestion

side effects. The only variable that does not show a pattern according to our expectations is income, as the free cost alternative is more likely than some options for higher incomes. Possible explanations are a bias due to the large amount of walking trips in the dataset, which is possibly capturing small trips to go to lunch or to acquire small services by richer people, or there is a confounding effect due to the urban phenomenon that richer people live closer to work. Nevertheless, the effect we are more interested in is the Value of Time, which we analyze in the sequence.

Table 4 – Alternative-Specific Estimated Coefficients and Regression Stats

	Main Mixed Logit	Std. Logit	Alternative Mixed
		Mean	
fin. costs	-0.108***	-0.042*	-0.122***
	(0.010)	(0.017)	(0.007)
travel time	-0.017***	-0.016***	-0.015***
	(0.001)	(0.001)	(0.001)
cost/income	-0.043**	-0.088***	-0.037***
	(0.017)	(0.026)	(0.010)
time x income	-0.001***	-0.001***	-0.002***
	(0.000)	(0.000)	(0.000)
		SD	
fin. costs	0.061***		0.075***
	(0.004)		(0.003)
travel duration	0.000		0.000
	(0.000)		(0.000)
Working only	Yes	Yes	No
Dummy for Motivation	No	No	Yes
N	93147	93147	177343
chi2	6525.144	46742.843	9279.755

$*p < 0.05, \quad **p < 0.01, \quad ***p < 0.001$

Table 5 – Case-specific Estimated Coefficients

	Main Mixed Logit	Std. Logit	Alternative Mixed
Bus			
female	0.431***	0.416***	0.230***
	(0.053)	(0.052)	(0.035)
income	-0.108***	-0.116***	-0.031***
	(0.014)	(0.014)	(0.007)
age	0.058***	0.052***	0.026***
	(0.004)	(0.004)	(0.003)
age squared	-0.001***	-0.001***	-0.000***
	(0.000)	(0.000)	(0.000)
peak	-0.343***	-0.349***	-0.125***
	(0.053)	(0.052)	(0.036)
downtown	-1.183***	-1.174***	-1.020***
	(0.089)	(0.088)	(0.065)
car out	-3.017***	-2.937***	-2.692***
	(0.267)	(0.260)	(0.241)
Railway			
female	0.160**	0.133*	-0.206***
	(0.056)	(0.055)	(0.042)
income	-0.015	-0.022	-0.017*
	(0.012)	(0.012)	(0.008)
age	0.033***	0.026***	0.022***
	(0.004)	(0.005)	(0.003)
age squared	-0.001***	-0.001***	-0.000***
	(0.000)	(0.000)	(0.000)
peak	-0.192***	-0.202***	-0.325***
	(0.056)	(0.055)	(0.042)

downtown	-0.547***	-0.516***	-0.146*
	(0.094)	(0.093)	(0.073)
car out	-2.214***	-2.115***	-1.712***
	(0.270)	(0.267)	(0.246)

Driving			
female	-0.857***	-0.845***	-1.192***
	(0.057)	(0.057)	(0.041)
income	0.006	0.004	-0.032***
	(0.008)	(0.008)	(0.006)
age	-0.028***	-0.034***	-0.040***
	(0.004)	(0.004)	(0.003)
age squared	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)
peak	-0.430***	-0.418***	-0.901***
	(0.056)	(0.056)	(0.039)
downtown	-0.982***	-0.912***	-0.609***
	(0.083)	(0.083)	(0.070)
car out	2.618***	2.644***	3.781***
	(0.074)	(0.074)	(0.079)

Motorcycle			
female	-2.206***	-2.195***	-2.528***
	(0.150)	(0.148)	(0.126)
income	-0.049*	-0.042*	-0.108***
	(0.019)	(0.018)	(0.020)
age	0.046***	0.047***	0.041***
	(0.009)	(0.009)	(0.007)
age squared	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)
peak	-0.253*	-0.254*	-0.553***

	(0.121)	(0.119)	(0.093)
downtown	-0.535*	-0.581**	-0.065
	(0.221)	(0.220)	(0.198)
car out	-5.559***	-5.353***	-4.985***
	(1.028)	(1.019)	(1.038)
Taxi			
female	-0.703**	-0.716**	-0.822***
	(0.265)	(0.238)	(0.135)
income	0.060***	0.053***	0.007
	(0.014)	(0.014)	(0.014)
age	-0.208***	-0.185***	-0.201***
	(0.010)	(0.008)	(0.008)
age squared	0.002***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)
peak	-1.799***	-1.589***	-2.286***
	(0.267)	(0.258)	(0.193)
downtown	1.673***	1.060***	1.889***
	(0.243)	(0.241)	(0.198)
car out	-30.768***	-21.758***	-38.071***
	(2.266)	(0.220)	(2.077)

5.2   Value of Time

To know the welfare cost of congetsion in São Paulo City, we need to know the valuation of the time lost in traffic by people. We made a structural model of utilities and estimated its parameters. From that, we can estimate the value of time as we have already presented by equations 7 and 9. By our model, we have a distribution of VOT since it varies according to income and to varying-parameters of time and expenditure cost. Figure 11 shows a kernel distribution of VOT, Table 6 display the stats for the VOTs estimated according to the three models and Table 7 compare it to possible proxies

literature recommends.

In Figure 11, we exhibit the VOT distribution from the Main Specification in red, while the distributions for the same observations but using the estimated parameters from the alternative models are in hatched lines. We call the standard logit's ascVOT, and the additional mixed logit AVOT. There is a considerable overlap between the density peaks and the shapes of the Main Specification's and the Alternative Mixed's VOT distributions. In regards to the Std. Logit's VOT distribution, there is only partial overlap with the Main Specification's one. The standard logit one is displaced to left, which means larger absolute VOTs. Another interesting fact is that VOT distribution has more variation when parameters are fixed and only income varies as in Standard Logit, while in varying-parameters models the distribution is smoothed.

Figure 11 – Utility Effect of Additional Minute in Traffic (R$/min)

In Table 6, we see the statistics in numbers. Average VOTs of the Main Specification and the Alternative Mixed have a really small difference. Yet, the Std. Logit average has a 10 cents difference, which is consistent with its distribution yielding higher losses to utility. Our approach is conservative at not estimating the welfare costs with the highest value we

literature recommends.

In Figure 11, we exhibit the VOT distribution from the Main Specification in red, while the distributions for the same observations but using the estimated parameters from the alternative models are in hatched lines. We call the standard logit's ascVOT, and the additional mixed logit AVOT. There is a considerable overlap between the density peaks and the shapes of the Main Specification's and the Alternative Mixed's VOT distributions. In regards to the Std. Logit's VOT distribution, there is only partial overlap with the Main Specification's one. The standard logit one is displaced to left, which means larger absolute VOTs. Another interesting fact is that VOT distribution has more variation when parameters are fixed and only income varies as in Standard Logit, while in varying-parameters models the distribution is smoothed.

Figure 11 – Utility Effect of Additional Minute in Traffic (R$/min)
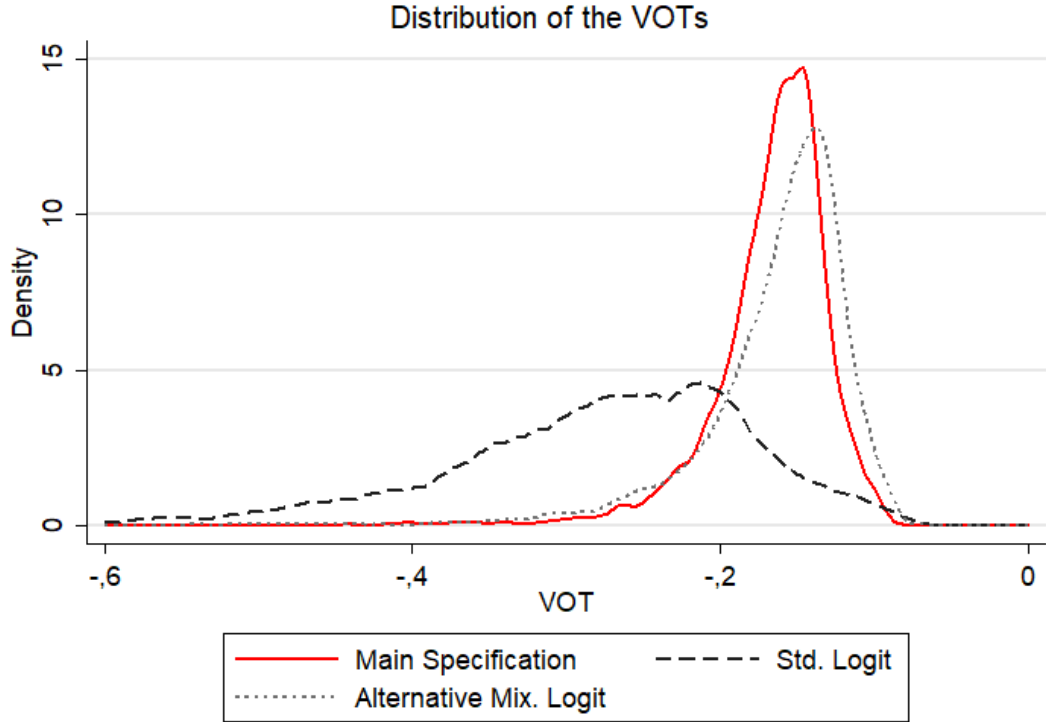
In Table 6, we see the statistics in numbers. Average VOTs of the Main Specification and the Alternative Mixed have a really small difference. Yet, the Std. Logit average has a 10 cents difference, which is consistent with its distribution yielding higher losses to utility. Our approach is conservative at not estimating the welfare costs with the highest value we

found. In fact, in the next table, we see some alternative methods that use higher values than ours.

Table 6 – Average VOT Stats for the different models (R$/min[2012])

	Mean	Std. Err.	[95% Conf. Interval]	
Main VOT	**-0.164**	0.000	-.0164	-.163
Std. Logit VOT	-0.270	0.001	-0.271	-0.268
Alt. Mixed VOT	-0.160	0.000	-0.160	-0.159

A diversity of methods has been employed to compute the value of time in economic analysis and in the appraisal of transportation projects (CONCAS; KOLPAKOV, 2009; DALBEM; BRANDãO; MACEDO-SOARES, 2010).Most of them make use of prescribed formulas related to wages or economic productivity as is shown by Santos (2012)'s literature review. Our approach has advantages over methods that use values adapted from other studies, or that use prescribed formulas. We employ a value that is estimated from an econometric structural model, using an individual-level survey that is representative of the local trips. Moreover, our VOT is trip-specific, so we avoid noise from the aggregation and importation of values.

Notwithstanding our model's advantages, it has some important limitations and simplifications. First, it is susceptible to econometric mistakes, for example, there can be an omitted variable we have not thought about inducing bias. Second, it assumes an intangible measure of welfare, as utility is a theoretical concept that justifies choices. Third, it relies on utility's linearity regarding time and money, as we assume VOT is constant for one or more extra minutes in traffic. And fourth, even if all the previous assumptions are true, a direct transition between time valuation on transport choice and opportunity cost can still be contested(CONCAS; KOLPAKOV, 2009; HADDAD; VIEIRA, 2015).

We now turn to present alternative ways of estimating the value of time lost in traffic as an exercise of robustness check.One way to estimate the opportunity cost of time is to look at the economic output that individuals could potentially produce by using GDP/worker/hour in São Paulo (similar to Cintra (2014)). From a firm and labor force perspective, we could instead use wage/hour, which reflects the monetary cost that the firm is paying for a business trip and is the amount the worker accepts for a non-leisure

activity. Some studies suggest that employers should account for an additional 33% of the hourly wage, or for the wage plus applicable benefits, when estimating the costs of sending an employee on a working trip(SANTOS, 2012, pp.61-65). Although in a commuting trip it may be unfair to assume that people are effectively working.

The World Bank is one of the main funders of transport projects in the world and it has many studies and handbooks on the theme. It commissions economic analyses and appraisals to approve projects and, based on the European Surveys of valuation of time, it recommends that 30% of the average gross household income is used as rule of thumb for the VOT (MACKIE; NELLTHORP; LAIRD, 2005; DALBEM; BRANDãO; MACEDO-SOARES, 2010). Finally, we compare our values with the results from similar methods applied to the São Paulo Metropolitan Area. As part of a study on urban tolls, Pacheco & Chagas (2016) has estimated the VOT working with the 2007 Origin/Destination Survey, not using Google Maps data and choosing a different specification for the mixed logit model. We adjust the results in Pacheco and Chagas by the IPCA (Indíce de Preços ao Consumidor Amplo) inflation index for comparability to our 2012 results. We also assume that people work 8 hours per day and 22 days per month on average, to a total of 176 hours per month.

Table 7 – Our VOT estimate compared to potential substitutes

Different Approaches	Adjusted Value* (R$/h)		Comparison
	2012	2018	Δ %
Main Model	**10.27**	**14.88**	
Individual Income	8.14	11.79	+ 26
1,33 x Individual Wage (Santos, 2012)	10.83	15.68	-5
30% Household Income (World Bank)	7.4	10.72	+39
GDP / worker (Cintra, 2014)	44.31	64.19	-77
Mix. Logit by Pacheco & Chagas 2016	11.17	16.18	-8

*Estimates adjusted to 2012 and 2018 by Brazilian Official Inflation Index (IPCA-IBGE)

Our mean estimate is close to Pacheco & Chagas (2016)'s. It is also only approximately one fourth higher than the individual income average and two fifths higher than World Bank's rule of thumb. It is very close to the business value of travel time recommended in Mackie, Nellthorp & Laird (2005) and Concas & Kolpakov (2009). On

the other hand, our result is much lower than the amount employed by (CINTRA, 2014). This last difference is due to the methodological differences explained in the introduction (we are estimating welfare personal costs as opposed to business costs) and to the fact that we adopt a conservative approach that avoids overestimating the loss values. From this comparison, we are comfortable that our structural estimates in lines with the literature, so we proceed to calculating the congestion cost estimates.

6 Congestion Costs

6.1 The Free Flow Counterfactual

We know that São Paulo is heavily congested, but which counterfactual can we adopt? How would São Paulo be if there was no congestion? Here once more, we take advantage of information from the Google Maps API to simulate a counterfactual scenario for the trips in our survey dataset. We assume that any trip whose main mode is bus, driving a car, taxi or motorcycle could be affected by traffic gridlocks and we would like to know the travel duration when congestion is absent. Google provides an estimated duration for car trips under different traffic conditions and we can query for different times. We request the trip duration by the recorded time in which the trips took place and also at 5:30 a.m - when São Paulo is about to be lighted by sunrise but the journey has not yet started - and compare the two.

Unfortunately, the API Directions could not return the duration of travel under traffic for bus trips, probably because of public transit timing constraints. To deal with that, we use Google's car trip duration to estimate a retardation rate and from that we produce the counterfactual (Eq.10 and Eq.11). We use this same method for all trips in the simulation, including the ones by car. Otherwise, we could simply take recorded travel duration for car trips and subtract it from Google's given duration at 5:30 a.m. By first calculating a ratio between Google's duration at departure time and at dawn thereafter multiplying it by recorded duration is a way to allow for two trips at same origin/ destination having different counterfactuals durations whether their factual durations

are different. So we account for intrinsic characteristics of the trip rather than constrain counterfactuals to a fixed duration in a given route. In other words, we assume that faster drivers go even faster in free-ways and that slower drivers keep driving slowly.

In the equations below, we clarify the method. r is the retardation rate for a trip i, resulting from the division of g_i by h_i, the car trip duration at the time of the trip and at 5:30 a.m (no traffic), respectively, both provided by Google. We multiply the inverse of this rate by the recorded trip duration in the survey to estimate the potential free flow duration. The delay (d_i) that we attribute to congestion is the difference between the factual (t_i) and counterfactual durations $(\hat{t}_i)$. From the travel demand model, we take the VOT for each trip and multiply it by the delay to compute individualized travel losses. Finally, we sum up the losses to get the total congestion cost.

$$r_i = \frac{g_i}{h_i} \tag{10}$$

$$\hat{t}_i = 1/r_i \times t_i \tag{11}$$

$$d_i = t_i - \hat{t}_i \tag{12}$$

$$Loss_i = VOT_i \times d_i \tag{13}$$

$$\text{Congestion Cost} = \sum_{i=1}^{n} Loss_i \tag{14}$$

In the cost analysis, we assess only a subgroup of trips. Given that we have structural estimates for work-related trips, we estimate costs only for these. Quantitatively, recall that work-motivated trips are 49% of the 43.713mi trips happening each day (21,419mi trips). And, more importantly, from the 25.337mi trips that we consider as vulnerable to congestion (using a bus, car or motorcycle), the work-related subset represents only 53% (13,428mi trips). Therefore, we are excluding a large volume of trips that could potentially increase our loss estimate. This said, we proceed to the results analysis.

6.2   The Delays Caused by Traffic

First, we examine the retardation rate estimates in terms of magnitude and compare them to average retardation rate from four other sources (Figure 12). The first is the TomTom Traffic Index, a measure of the increase in overall travel times compared to a free-flow situation(TomTom, 2016). TomTom is a global company producing GPS-based navigation gadgets and watches and they use their applications to measure travel speed. The travel speed historical data are used to calculate travel time on individual road segments and entire networks and they build up the index by giving a higher weight to busier than to quieter roads. They have an overall index and specific morning peak time and evening peak time indexes (TomTom, 2016). The second source are CET's reports about traffic conditions. CET agents drive in São Paulo's main roads with a test vehicle twice a year and measure the time spent in each pre-determined route and how much time the vehicle was retarded by congestion or by traffic lights. They do that during the morning and evening peaks and from that they calculate the average rate of retardation (CET, 2013b; CET, 2017). Here we use CET's numbers for 2012, the survey year, and for 2016 (the last report available to date), as we collected Google's information during 2017. At last, we show the retardation rate estimated by Haddad & Vieira (2015) for São Paulo. They estimate it by using a regression that models average commuting duration per city according to its urban characteristics, in which the commuting duration comes from the 2010 Brazilian Census. The final retardation rate is calculated from the Census' observed value of São Paulo minus the fitted value by the regression.

The average retardation rates across the studies are all around 30%, no more than 5 percentage points above or below. When all the retardation rates are considered, including the peak time-specific ones, only the TomTom Index shows a larger variation, with the others being consistently very close to our estimates. This shows that our approach remains within the expected range. Moreover, by using specific rates for each trip instead of general averages we also keep a more realistic distribution.

In Figure 13, the distribution of estimated delays is examined. 89% of the trips
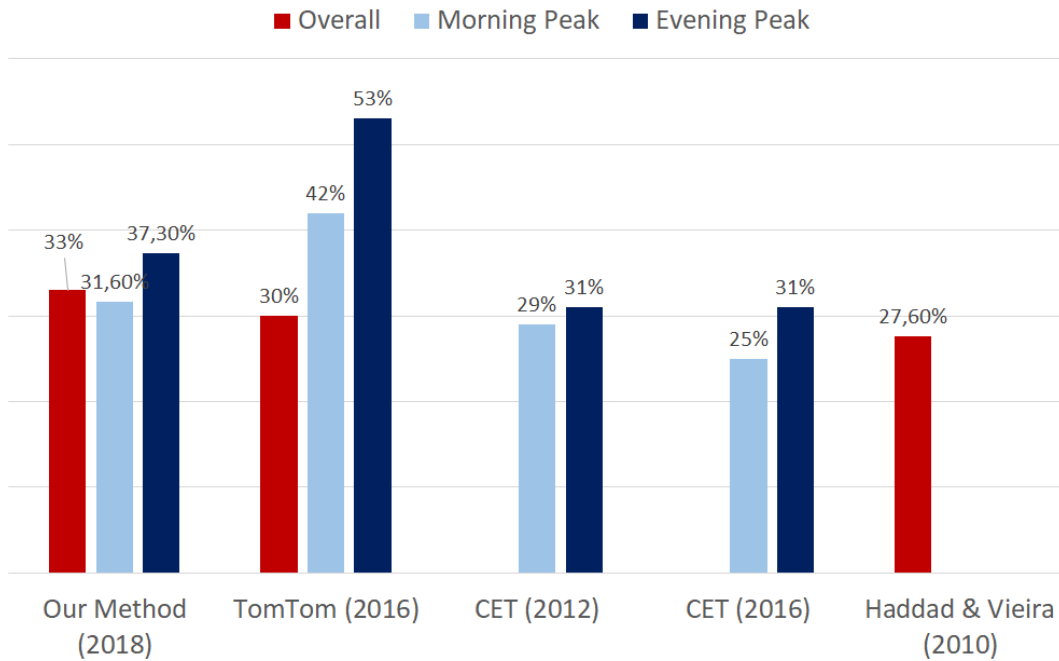
Figure 12 – Comparing retardation rate estimates

in our valid sample[9] have their duration increased by traffic. Why the other 11% is not impacted? Taking a look at the departure time distribution of the non-delayed trips, it can be seen most of it happens in late night (Figure 14). The rest of this non-delayed possibly have set place where traffic is not a problem or also can represent errors in our estimates coming from Google's system. Notwithstanding these details, the important message here is almost all the trips in analysis get extra time due to externalities arising from people transporting themselves simultaneously in urban environment.

6.3 The Welfare Costs of Congestion

We estimate a total labor force welfare cost of 29.119 mi R$/day[10] in São Paulo. This is equivalent to 7.338 billion Reais in one year .[11] per year excluding weekends and a cost of R$747/year per worker. This means that, on average, workers accumulate almost one monthly Brazilian minimum wage (currently R$954) in lost opportunity costs over a year. While this looks like a sizeable loss, we need to further explore its relative

---

[9] observations inputted in the estimation, for which we have predicted parameters
[10] values of 2012 adjusted by IPCA
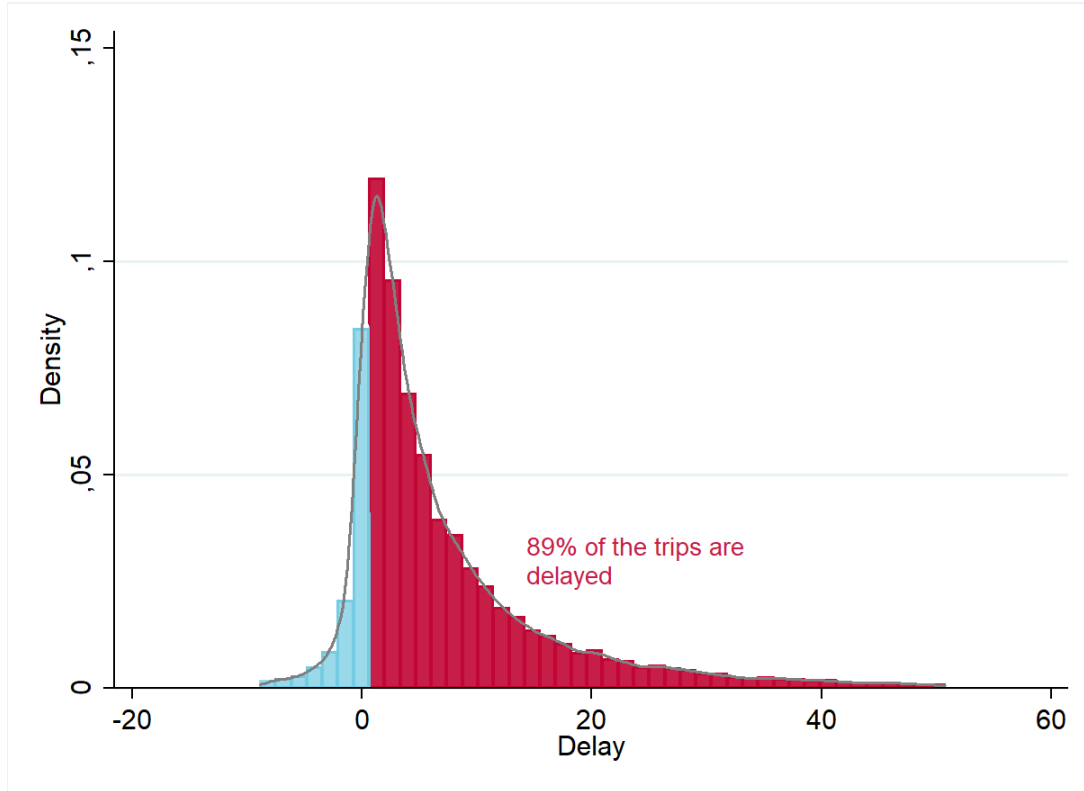[11] 252 working days per year accounted

Figure 13 – Distribution of estimated delays on travel duration (minutes)

proportion. To do that, we confront our welfare cost estimates to two amounts related to the opportunity cost of the road system in the São Paulo Metropolitan Area. First, in Table 8, we present an alternative way in which welfare costs are estimated. In the sequence, we get closer to the policy perspective by examining the revenues from taxation over automotive vehicles in the SPMA (9).

Haddad & Vieira (2015) have proposed a different approach to estimating opportunity costs of traffic frictions, using a general equilibrium model approach. They estimate the average commuting time using a regression at the municipality level, where the sample are the Brazilian municipalities, and conclude the duration of trips in SPMA is 27.63% larger than its expected value. Note that in their aggregated and global approach, they define a unique rate of delay while we estimate one for each trip. From this value, they simulate what would happen to the economy if all the travel durations were reduced by this amount. There is a general reduction of production costs, which benefit the whole economy and in consequence increases family consumption. People derive utility from
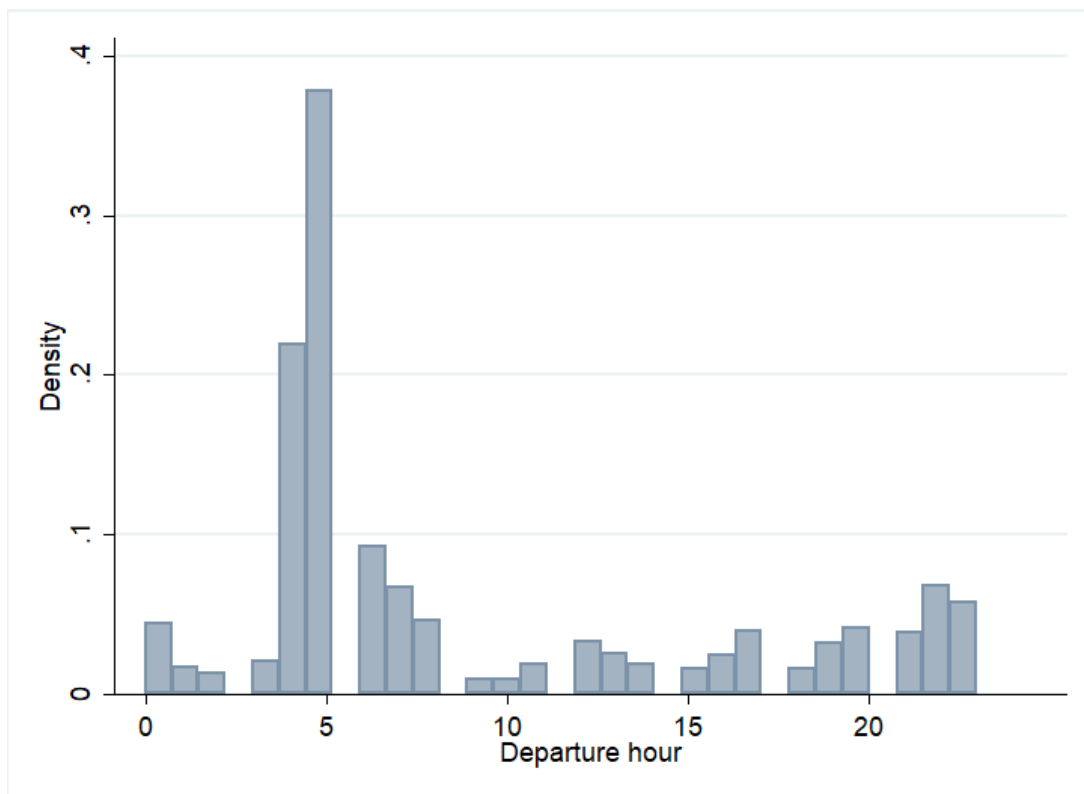
Figure 14 – Histogram of departure hour for trips that were not estimated as delayed

consumption and this can be seen as one of the dimensions of welfare.

Haddad & Vieira (2015) compute values[12] that refer to three time horizons: very short run, short run and long run. While these scenarios do not have an exact temporal correspondence[13] in months or years, and although the outcomes that they measure are not the same as our welfare estimates, we confront them to our linear projection of one-year and five-year welfare costs using a billion R$ scale. Their 'very short run' estimates is the lowest measure of per capita costs, but after incorporating agglomeration effects to consider 'short run' results, their average opportunity cost per capita reaches R$2,322, which is approximately two thirds of our estimate of five years cost. Again, these amounts are not directly comparable to our numbers, but the objective is to show that we arrive at comparable magnitudes using totally different methods and thus we introduce an alternative way to measure welfare.

---

[12] We update their values from 2010 R$ using the IPCA.

[13] Haddad & Vieira (2015) define 'very short run' result as the one that do not consider agglomeration effects and variation over capital stocks, 'short run' includes agglomeration effects through accessibility changes, and 'long-run' allows to capital stocks to be endogenous

Table 8 – The estimated welfare costs from traffic retardation (1000 R$ / 1000 heads)

Welfare costs	Total	Population	R$/capita	Source:
Utility (year)	**7,337,906**	9,813[1]	747.77	Author
Utility (5 years)	**36,689,529**	9,813	3738.87	Author
Consumption (very short run)	2,458,410	21,392[2]	114.92	Haddad & Vieira
Consumption (short run)	49,676,549	21,392	2322.24	Haddad & Vieira

[1]Number of workers in SP Metropolitan Area. Source: METRÔ-SP (2013)

[2]Population in SP Metropolitan Area. Source:EMPLASA (2018)

The Automotive Vehicles Ownership Tax (Portuguese acronym IPVA) is charged annually by the São Paulo State Government. From the total amount collected, 40% remains with the state, 40% goes to the Municipal Government according to the vehicle's origin and 20% is channeled to a fund devoted to basic education (FUNDEB)(FAZENDA, 2017b). Accordingly to the state official information, the IPVA revenue is destined to transportation, health, education, security and others.(FAZENDA, 2017b). Though the IPVA funds are not earmarked for transportation policy exclusively, we understand that it should be an important source of resources for transportation policy since it comes from the automotive fleet. In 9, we present the total amount collected by the municipal governments of SPMA through IPVA and the total originated in SPMA by summing the amount belonging to municipal and state governments.

The result is that even if the Municipal and State Government would partner in a theoretical redistribution of the IPVA revenue across workers to compensate for welfare costs, it would fall short by 18% (or R$ 1.291 billion). If a more pro-government citizen might conclude that the government should increase the IPVA taxation in order to compensate citizens for congestion-related losses, in another hand, a more liberal citizen might instead claim that taxes are paid even though the government is unable to prevent losses larger than the total IPVA budget. Yet the goal of this thesis is not to discuss the funding of transport policy or the optimum tax rate, what can be stated from the welfare cost analysis is that it is effectively high and that governments should be concerned with making efforts to reduce those costs.

Although we do not know the best policy to solve current traffic problems, we can

Table 9 – Budget from Automotive Vehicles Ownership Tax (IPVA) - (1000 R$ / 1000 vehicles)

IPVA 2017	Tax Revenue[1]	Fleet[2]	R$ / vehicle	Welfare Cost	% Wf. Cost
Municipalities	3,023,454	12,931	233.82	7,337,906	41%
Municipalities + State	6,046,907	12,931	467.64		82%

[1]Government's Available Budget from IPVA. Source: (FAZENDA, 2017a)

[2]Number of Automotive Vehicles in SP Metropolitan Area. Source:DENATRAN (2018)

observe it will have to deal with technology transformations and also make smart use of them. E-commerce and telecommuting are examples of new economic patterns made possible by technology revolution and that can shift travel demand. As well, technology directly changed how people transport themselves across the space with the insurgence of applications of carpooling, ride services and navigation that accompanied 'smart phone' popularization. In this group, navigation applications is of special interest to transport management, since it can reallocate vehicle to better routes in terms of travel time. At best, real-time communication between gadgets and a central program could manage traffic to minimize total congestion in cities, resulting in maximum aggregated welfare.

Among the navigators, Waze is the most popular and Brazil was among the countries with the highest number of subscribers(Waze, 2016). The effects of this popularization were specially noticed in October 23rd, 2017, when malfunctions in Waze led to increased gridlocks in São Paulo. In the next section, we explore the consequences of this event, estimate its welfare cost and discuss the meaning of the estimation.

7   Welfare effects of Waze use in São Paulo

7.1   Technology, Mobility and Waze

The rapid development of Information and Communication Technologies (ICT) has significantly modified the ways in which people have leisure time, work, and behave. Mobility is a fundamental part of this movement as the demand for travel is affected by e-commerce and online communication, the transport supply is altered by carpooling and ride-service applications, and driving decisions are influenced by widespread GPS

navigation.

Amid all the changes, navigation applications were popularized through the prolif-eration of Internet and GPS access by smartphones. According to Meirelles (2017), there were 1.4 Internet-enabled mobile gadgets[14] in Brazil per inhabitant in May of 2017 (292mi), and from those 71% were smartphones, which means one device per inhabitant.

Once a GPS-enabled mobile gadget is acquired, the marginal cost of using GPS navigation depends upon the Internet data consumption, but this usually represents only a small fraction of the typical data packages offered by telecoms. Moreover, it is often possible to upload the data using a wi-fi connection and still have access to many of the navigation functionalities while in offline mode. As a consequence, navigation through applications such as Waze or Google Maps has become very popular, notably in larger cities where it is difficult to memorize all the routes and traffic jams are a problem.

Waze is the leader in the navigation app market having 65 million monthly active users in 185 countries. The company, founded in 2007 in Israel, was acquired by Google in 2013, which was already the owner of Waze's biggest rival, Google Maps. The differential feature in Waze that can largely explain its success is the interactivity, where a network of live information creates the feeling that users are community members helping each other. For example, a driver can warn and be warned about traffic levels, accidents on the route, location of gas station. More dedicated 'Wazers' can even edit maps.

Waze's booster is thus a type of 'collective intelligence' , which can lead to further social utilization. For example, Waze was one of the main tools used to locate people during the 2013 Austin (Texas-USA) floods and to alert people during the 2013 Oklahoma (USA) tornadoes (FARIA, 2014). Yet Waze users are not necessarily purely altruistic in their engagement, as they are in search of the best solution for themselves (FARIA, 2014). Also, Waze explores social participation as a marketing strategy and incentivizes it through the ranking of drivers by engagement levels.

Technically, Waze is classified as a Participatory Sensing Systems (PSS) by infor-

---

[14] Notebooks, tablets and mobile phones considered as mobile gadgets.

maticians, as it allows people to share data about the environment (or context) they are inserted in at any time and place. PSS are seen as a strong means to make ubiquitous computing a reality (SILVA et al., 2013), providing useful inputs to public services, such as traffic management. Recently, engineers have been discussing how to use GPS-enabled mobile phones to monitor and manage traffic more efficiently, pointing out that it can be further improved in the coming years (WANG et al., 2012; VLAHOGIANNI; KARLAFTIS; GOLIAS, 2014). Here we explore a sample of the power of GPS-enabled smartphones using an atypical day in 2017 when Waze induced unusual traffic jams in São Paulo.

7.2 The Natural Experiment

By the big number of active users (Waze, 2016), and the large quantity of 'wazers' groups[15] within SPMA (FARIA, 2014, pp.82-85), we could expect Waze had a significant impact in São Paulo's traffic. But this became evident on October 23rd 2017.

On that day (henceforth referred to as 'the Waze day'), another traffic jam disturbed commuters in São Paulo. Yet none of the usual causes seemed to explain what was going on – there was no adverse weather, no big accidents, nor crowded games and concerts – except for the fact that many drivers were reporting that Waze was directing them to the congested areas they usually avoided. The unexpected traffic conditions were widely reported and pointed out to a Waze bug as a key cause (KLEINA, 2017; CAPELAS; RIBEIRO; MENGUE, 2017b; SANTINO, 2017).

The newspaper "O Estado de S. Paulo" told the stories and reproduced comments of some users that took up to three hours to accomplish the route they usually do in around one hour (CAPELAS; RIBEIRO; MENGUE, 2017a). In most of the narratives a pattern could be identified: Waze was suggesting drivers the direction of the *23 de Maio* Avenue to access the airport, in spite of helping travelers to arrive at the desired destination as fastest as possible. Most of these drivers are used to get good advices by the application, but latter realized Waze was not really helping at that time.

---

[15] Waze allows people to create private communities inside the application interface. Inside the group environment, people can chat and exchange information. They are usually composed by people with a common interest, for example, living in the same neighborhood or owning similar cars.

In addition, the newspaper disclosed Waze has admitted the bug, which was classified as an 'incident' that was solved in the afternoon (CAPELAS; RIBEIRO; MENGUE, 2017a). In the intent to confirm the information and to obtain data and details, we have also contacted the Waze's staff. The company responsible for Waze's public relations in Brazil gave us an answer by email admitting there was a bug and emphasizing it was a rare incident that was already fixed (The official message is reproduced as stated originally in Portuguese in the Annex II of this Thesis).

The airport's neighborhood is already loaded with intense traffic due to the large volume of passengers arriving to or leaving the local. The fact the zone was attracting more vehicles due to Waze looks to have triggered spillover effects over the city's traffic. Capelas, Ribeiro & Mengue (2017a) shows numbers of CET's Slowness Index claiming the number of congested kilometers within the city on the Waze day were "highly above the mean upper limit during the afternoon".

The Slowness Index by CET measures how many kilometers are congested in the Municipality of São Paulo. CET's agents are strategically placed with tablets and smartphones broadcasting real time data to the central (CET, 2013a). They monitor the 175 most important traffic corridors of the city from 7a.m. to 8p.m, which sum up to a total of 867.673 kilometers. The central office aggregates the information and provides traffic behavior trend by zone each 10 minutes and congested kilometers by corridor each 30 minutes on its public website. As the number of congested kilometers is based on agents' observation, CET calculates an upper limit and a lower limit to it through statistical methods. The calculations take into account historical data of the same day of the week for the past 12 months (CET, 2013a).

Despite accessing CET's real time data is easy, retrieving detailed historical data requires bureaucratic procedures. You have to make official requests that can be answered within 21 days and demand an official appeal if your first order was not fully answered. In the Annex II of this thesis, the chart published on the "O Estado de S. Paulo"(CAPELAS; RIBEIRO; MENGUE, 2017a) is shown containing the Slowness Index along October 23rd and the historical mean upper and lower limits based on Mondays of the past 12 months.

To identify effects that are likely to be caused by Waze bug rather than unidentified temporal factors, we narrow the analysis to the period between 20 days before and 20 days after the incident. In Figure 15 the evolution of the Slowness Index along Waze Day is showed by the red line, while the black line is the average for the period and the hatched lines are the maximum and minimum. All the statistics are conditioned on the time of the day. From 10:30 a.m., the Slowness Index went into an increasing trend compared to the average, breaking the maximum records from 1:30 to 2:30 p.m. At 4:00 p.m. the jams were apparently dissolved as the Index got close to the average, and, at 4:30 p.m., there were less congested kilometers than average.
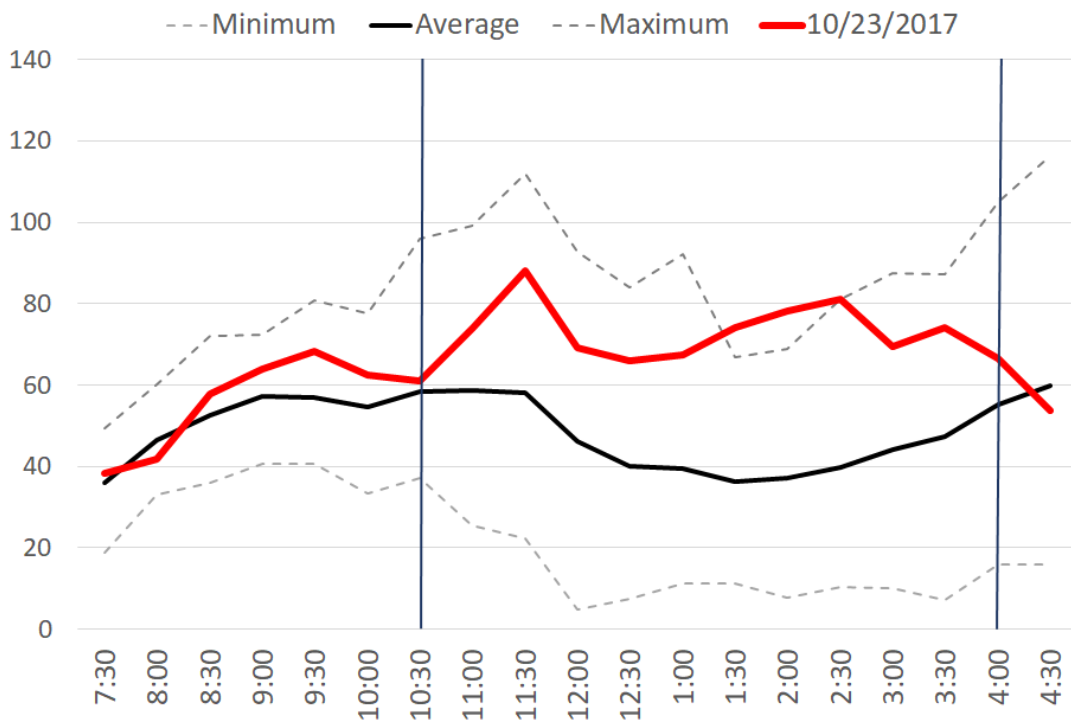

Figure 15 – Comparing Slowness Index of atypical Waze day to statics of close period - Source: CET

From the media (KLEINA, 2017; SANTINO, 2017; Canaltech, 2017; CAPELAS; RIBEIRO; MENGUE, 2017a; CAPELAS; RIBEIRO; MENGUE, 2017b) and the company official's communication, we could not identify anyone expecting for Waze's malfunction. Furthermore, to the best of our knowledge, there was no other day in which a similar extensive bug was reported. Because of that, we can assume the incident was not expected beforehand, being exogenous to any commuting decision. It means we had an experiment from which we can derive Waze effect on consumer surplus.

Our PMU sample is representative of an average day in São Paulo, so we compare recorded travel durations in the survey to estimated travel times on the Waze day. This is similar to how we estimated congestion effects. There, the average travel duration was compared to a hypothetical situation without traffic friction. Now, estimated travel durations for the Waze day are compared to the average travel durations. Instead of being shorter than the average travel durations, as in the congestion analysis, assuming the bug increased congestion implies that the counterfactual durations are now longer than the average (Figure 16).
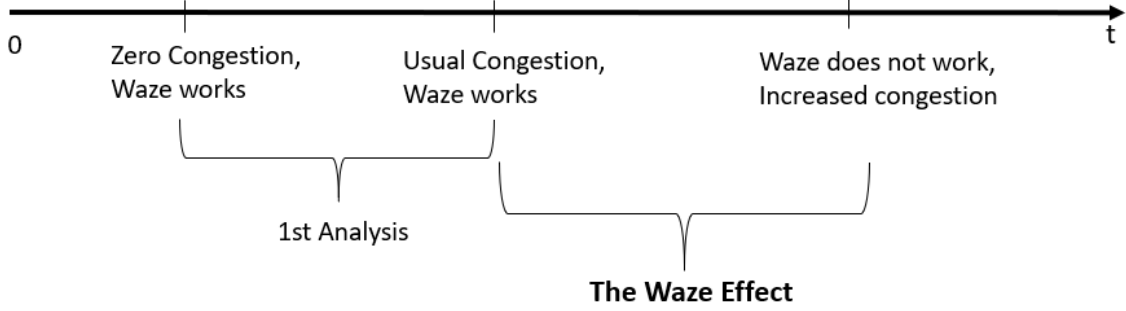

Figure 16 – Different effects under analysis: Waze malfunction led to increased travel duration

Thereafter, we estimate the welfare loss due to the Waze malfunction using a common day as the counterfactual. Our approach is similar to that described by equations 10 to 14, but now the trip i delay provoked by Waze (w_i) is given by the estimated travel duration on the Waze day ($\tilde{t}_i$) minus the recorded travel duration in the survey (t_i). To estimate $\tilde{t}_i$, we first compute a new delay rate ($\tilde{d}_i$) from the Slowness Index (s_i) on the Waze day at the specific time period (p) divided by the average Slowness Index ($\bar{s}_i$) for that period. Then, $\tilde{d}_i$ multiplies t_i to result in $\tilde{t}_i$. For example, for a driver who went out at 11:30a.m. on October 23rd, the Slowness Index was 1.5 times the average, so we multiply t_i by 1.5 to obtain $\tilde{t}_i$. For another driver who left home at 15:30 p.m., we use the 1.2 multiplication factor, and so on. Finally, the delay caused by Waze is multiplied by the Value of Time estimated from the travel demand model to yield an estimated welfare loss for each trip. To get the total welfare effect from the Waze day, we sum up the losses for

70

all trips.

$$\tilde{d}_i = \frac{s_{ip}}{\bar{s}_{ip}} \tag{15}$$

$$\tilde{t}_i = \tilde{d}_i \times t_i \tag{16}$$

$$w_i = \tilde{t}_i - t_i \tag{17}$$

$$WLoss_i = VOT_i \times \tilde{d}_i \tag{18}$$

$$\text{Waze Effect} = \sum_{i=1}^{n} Loss_i \tag{19}$$

There are some issues in estimating the Waze effect from the Slowness Index. The most important is that we do not have congestion data disaggregated across space, and this can create noise. For example, the VOT is not homogeneously distributed across the population and it can have spatial correlation. Therefore, the affected commuters could have a different VOT distribution than those unaffected, yielding less accurate estimates. Another problem is that we do not know how much the congested roads represent proportionally to the complete extension of each trip. While CET could be missing some congested roads, it is likely to capture the gridlocks at the main avenues that should also represent the largest fraction of trips. In future work we expect to obtain and model the disaggregated CET data.

Given those limitations, the following measures were taken to avoid further inaccuracy. First, the trips considered vulnerable to congestion are those taken by cars, buses or motorcycles, similar to the former analysis . Second, the incidence of delays was considered only for trips begging and ending inside the city of São Paulo. This improves data consistency as Waze has not provided details about the geographic and temporal extension of the problem (despite having formally acknowledge the malfunctioning), and

as the Slowness Index measures congested kilometers only on avenues within the limits of the Municipality of São Paulo (not in the Metropolitan Area).

Furthermore, the estimation is restricted to trips beginning between 10:30a.m. and 3:59p.m., corresponding to the part of the Waze day when the Index diverged from the average trend. To define the beginning of this interval we considered the Slowness Index average starts to decrease from 10:30 a.m to 11:30 a.m., while on the Waze day it increased from 10:30 a.m. and arrived to the maximum of that day at 11:30 a.m.(Figure 15). Also, we take into account people's narrative that there was uncommon traffic jam at least since 11:00a.m.(CAPELAS; RIBEIRO; MENGUE, 2017a). To define the end of the analysis interval, we considered that from 4:00 p.m. to 4:30 p.m. the Slowness Index on the Waze day went into a decreasing trend that put it under the average number at 4:30 p.m (Figure 15). As well, reported Waze's official speech was that the problem was solved 'by the end of the afternoon'(CAPELAS; RIBEIRO; MENGUE, 2017a; SANTINO, 2017).

The last measures taken to avoid inaccuracy refer to methods and results from previous sections. From the Section 5, we retrieve the structural Main Model Specification included only trips motivated by work, thus, the simulation of this Section comprise only these same 'working trips' to be consistent. From the Subsection 6.2, results of estimating the delay caused by traffic using the PMU and Google data pointed out that 89% of the trips are affected (Look at Figure 13). Since the complementary 11% are not susceptible to any traffic friction on typical days, we would rather consider they are not disturbed by the Waze bug either.

7.3 Welfare Effects

The total amount of lost utility due to the Waze malfunction is shown on Table 10. This is 1.8 times the daily congestion cost estimated for the regular congestion in section 6. It implies that the total welfare loss with respect to a free flow situation almost tripled during the Waze day (as the Waze effect is added to the average traffic cost). This reveals Waze's long reach as well as the large potential of welfare loss when an event substantially disturbs the traffic in São Paulo.

Table 10 – Estimated Welfare Effect of the Waze malfunction on a single day (1000 R$)

Welfare Effect	Loss
Waze Malfunction Day	**52,298**
Usual Congestion (Day)	29,119
Total	81,416

The loss associated to the malfunction represents the combination of two factors whose subtle difference deserves attention. On the one hand, travel durations went up partially because people trusted Waze and were directed to overloaded routes that they would have otherwise avoided. This is the perverse effect of the malfunctioning, which was due to a system error and is exclusive to such exceptional situations. On the other hand, absent the malfunction, Waze optimizes the users' choices with regards to the routes to follow, taking drivers to ways they may not know or would not consider driving through. The latter is Waze's standard benefit, which saves valuable travel minutes. As the system was not working properly on October 23rd, part of the increased duration of trips can be attributed to absence of the benefit of Waze rather than to the presence of the perverse effect. In fact, it could be speculated that the absence of the standard benefit was the dominant driver as rational users can ignore the instructions once they realize that it is not working correctly.

The Waze effect analysis sheds further light on the fact that the use of GPS navigation has reached adoption levels that cannot be ignored when thinking about the economics and management of transportation systems. This conclusion may be especially relevant in the coming years as the number of users may continue to rise due to the increased accessibility of smartphones and technology-based habits.

8 Conclusion

Traffic jams are an enormous problem in big cities and have motivated sizable public discussions in the past few years. From an individual perspective, driving a car is a comfortable mean to move between two points. This preference can become even stronger when streets are congested and the alternative is to face hours of traffic standing up inside

a public vehicle. Yet even if the car may usually seem like the best option, *ceteris paribus*, everyone could hypothetically get better off in a coordinated movement of using public buses or active transport, in which space would be more efficiently occupied.

Economic theory can be applied to understanding travel behavior and traffic congestion. Travelers are taking into account their own immediate costs when choosing their mean of transport and cannot have control over other agents. As such, each traveler minimizes their individual costs at the time of choice and imposes externalities on the others. Social arrangements generate travel demand shocks, the peak times, when people need to go to school or to office. In turn, scarcity of space for all the society demands results in a shift of costs, generating a deadweight loss.

Following an economic logic, some policies propose congestion tolls to make agents internalize the social costs of their chosen behavior. Other policy recommendations suggest that fuel taxes can be used as tools to control traffic, while others still insist on the need to widen avenues. Whatever the preferred strategy, trade-offs and side effects are always present. Different types of costs are generated in trying to solve the traffic issues and, as a consequence, disputes with the potential to block the implementation of policies can emerge. In this thesis, we alert to the cost of not solving the traffic problem.

First, a distribution of the value of the time spent at trips motivated by work within São Paulo Metropolitan Area is estimated. From a mode choice demand model, we have inferred commuters' value of time in traffic was R\$0.16 on average. Updating the value for the current year, it represents a loss of R\$14.88 per hour in traffic.

After that, it is estimated that the delay rate of travel durations in confront to a hypothetic free flow situation is 33% on average. Moreover, through the combination of each trip recorded in the Mobility Survey to Google data, it was possible to draw the distribution of the added travel minutes, what has not been presented by other works using alternative approaches to estimate the delay rate (TomTom, 2016; CET, 2013b, 2013b; HADDAD; VIEIRA, 2015). We have estimated 89% of the trips taken by any kind of car, bus or motorcycle and whose reason is work have the duration increased by traffic.

Joining the VOT with the delay estimates, we have estimated a total amount of R\$7.338 billion in *Marshallian* welfare annual costs due to the time lost as compared to the free flow counterfactual. As it is derived from a sample that includes only trips motivated by work, we can say that yields an average annual cost of R\$747.77 per worker. This is a large effect, and one that does not account for broader consequences for the economy or caused by air pollution. To show how sizable the losses are, we have calculated that a potential redistribution to the affected commuters of the annual budget raised through the automobile ownership tax in São Paulo would be insufficient to compensate for the congestion-related losses.

The robustness checks provide evidence that the estimates are consistent. The VOT distribution does not change significantly when the sample is changed from work-motivated trips to all trips, and the average VOT is consistent with the different approaches. In estimating the delay provoked by traffic on trips, the overall average rate does not show large discrepancies with estimates obtained using other methods and data. Furthermore, the use of Google Maps API data to fill up missing data left by the surveys of Origin-Destination has had a satisfactory performance. The API has provided answers for most of the queries based on the trips recorded in the Mobility Survey and enabled the estimation of econometric models with parameters towards the expected direction.

Therefore, we can confidently state that the monetary annual loss caused by overloaded traffic is in the scale of billions and governments as well as civil organizations should look for solutions to reduce this loss. Best efforts and tools ought to be made available, and technological devices should not be neglected.

We have estimated the welfare effects elicited by Waze, the most popular GPS navigation application. By combining the same VOTs estimated to the cost of the congestion analysis to additional traffic congestion data, the utility loss of the Waze malfunction on October 23rd, 2017 is gauged. We estimate a total loss of R\$52.298 million for the five and a half hours that the bug lasted in São Paulo. This amounts to 1.8 times the value estimated for the daily cost on a usual day of congestion with regards to a free flow situation, implying that the total cost when the regular congestion and the Waze effect

are compounded is almost three times the one without the Waze bug.

This shows the extent peer-to-peer navigation can impact traffic. Many users trust Waze and have good experience at using it, otherwise there would not be such an impact caused by its bad suggestions of ways on that day. This is a recent phenomenon that lacks of better understanding yet.

On the one hand, the social navigation is likely to improve traffic flow as data is shared among drivers and the program advises better routes. In most optimistic predictions for the near future, a central program receiving real-time data could perfectly return instructions sprawling traffic across the streets to minimize congestion. This is more likely to be done with the emerging development of automated vehicles. The computers in charge to drive could receive the routes traced by the central.

On the other hand, new risks emerge. Bugs on technologies whose usage has become a widespread habit can lead to mass impact, as seemed to be the case on 'the Waze day'. Besides that, malfunction can be caused as an incident of unpredicted errors of programming, as well as generated by hackers' attacks doing it intentionally, thus, users should be aware technology is imperfect.

More studies can contribute to the comprehension of broader effects of the GPS-based social navigation, where there is specially lack of empirical analysis. Specifically talking about the study of this thesis, it could be improved in the future with the incorporation of disaggregated data and with the addition of social network data. Silva et al. (2013), for example, collected a dataset of Waze alerts directly from Twitter that could be adapted to the welfare costs analysis.

Regarding the general traffic congestion, we have provided an estimate to its costs but have not in-depth examined any policy proposed to solve that. Solutions for the urban congestion are demanded and may take in consideration interdisciplinary analysis of urban planning and traffic management, involving studies of architecture, sociology, engineering, economics and others.

Economic studies can particularly contribute for that in some topics. First, more

studies quantifying the impacts of different policies, specially comparing alternatives, are useful to orientate policy-makers towards better decisions. However, even if the best policies from the economical perspective are known by the policy-makers, it is likely those cannot be implemented because of private interests able to block that. Economic studies can further help to understand the political economy behind traffic policies either by theoretical formulations or by empirical applications. Comprehending that, policies can be more accurately designed, possibly moving from an optimum plan to a second-best feasible one with the aim of diminishing the costs of congestion.

Bibliography

ADLER, T.; BEN-AKIVA, M. A theoretical and empirical model of trip chaining behavior. *Transportation Research Part B: Methodological*, v. 13, n. 3, p. 243–257, set. 1979. ISSN 0191-2615. Disponível em: <http://www.sciencedirect.com/science/article/pii/019126157990016X>.

AKBAR, P.; DURANTON, G. Measuring the cost of congestion in highly congested city: Bogotá. CAF, 2017.

Anfavea. *Estatísticas*. 2017. Disponível em: <http://www.anfavea.com.br/estatisticas.html>.

ANP. *Levantamento de preços e de margens de comercialização de combustíveis*. 2017. Disponível em: <http://www.anp.gov.br/precos-e-defesa-da-concorrencia/precos/levantamento-de-precos>.

ARNOTT, B. R.; PALMA, A. D. E.; LINDSEY, R. American Economic Association A Structural Model of Peak-Period Congestion : A Traffic Bottleneck with Elastic Demand Author (s): Richard Arnott , André de Palma and Robin Lindsey Source : The American Economic Review , Vol . 83 , No . 1 (Mar ., 1993). v. 83, n. 1, p. 161–179, 1993.

BARCELLOS, T. M. *Não são só 20 centavos: efeitos sobre o tráfego da Região Metropolitana de São Paulo devido a redução na tarifa de ônibus financiada pelo aumento da CIDE nos combustíveis da cidade de São Paulo*. Tese (text) — Universidade de São Paulo, jun. 2014. Disponível em: <http://www.teses.usp.br/teses/disponiveis/96/96131/tde-23092014-135716/>.

BECKMANN, M. J.; MCGUIRE, C. B.; WINSTEN, C. B. *Studies in the Economics of Transportation*. [s.n.], 1955. ISBN 0-444-85222-0. Disponível em: <http://www.rand.org/pubs/research_memoranda/RM1488.html%5Cnhttp://cowles.econ.yale.edu/archive/reprints/specpub-BMW.pdf>.

BEN-AKIVA, M. E.; LERMAN, S. R. *Discrete Choice Analysis: Theory and Application to Travel Demand*. [S.l.]: MIT Press, 1985. Google-Books-ID: oLC6ZYPs9UoC. ISBN 978-0-262-02217-0.

BHAT, C. R.; KOPPELMAN, F. S. Activity-Based Modeling of Travel Demand. In: *Handbook of Transport Science*. [S.l.: s.n.], 2003. p. 39–65. ISBN 978-1-4020-7246-8.

BLISS, L. *The Case (Again) for Congestion Pricing in New York City*. 2018. Disponível em: <https://www.citylab.com/transportation/2018/01/could-congestion-pricing-finally-work-for-new-york-city/550958/>.

BOWMAN, J. L.; BEN-AKIVA, M. E. Activity-based disaggregate travel demand model system with activity schedules. *Transportation Research Part A: Policy and Practice*, v. 35, n. 1, p. 1–28, 2000. ISSN 09658564.

CABRAL, P. The city with 180km traffic jams. *BBC News*, set. 2012. Disponível em: <http://www.bbc.com/news/magazine-19660765>.

CAITANO, A. Metrô em Higienópolis se transformou na maior polêmica da semana. *EXAME*, São Paulo-SP, Brasil, maio 2011. Disponível em: <https://exame.abril.com.br/brasil/metro-em-higienopolis-se-transformou-na-maior-polemica-da-semana/>.

CAMERON, A. C.; TRIVEDI, P. K. *Microeconometrics: methods and applications*. [S.l.]: Cambridge university press, 2005.

Canaltech. *SP: Waze admite falha que levou motorista a congestionar via*. 2017. Disponível em: <https://www.terra.com.br/noticias/tecnologia/canaltech/waze-admite-falha-que-direcionou-motoristas-a-avenida-congestionada-em-sao-paulo, 2aa76030ab0897c108f347a30fd188e4abeonyxp.html>.

CAPELAS, B.; RIBEIRO, B.; MENGUE, P. Newspaper, *Com lentidão triplicada, São Paulo tem segunda-feira atípica*. 2017. Disponível em: <http://sao-paulo.estadao.com.br/noticias/geral,com-lentidao-triplicada-sao-paulo-tem-segunda-feira-atipica, 70002057753>.

CAPELAS, B.; RIBEIRO, B.; MENGUE, P. *Falha no Waze pode ter ajudado a triplicar trânsito em São Paulo*. 2017. Disponível em: <https://exame.abril.com.br/tecnologia/falha-no-waze-por-ter-ajudado-a-triplicar-transito-em-sao-paulo/>.

CET. *CET APRESENTA "TRÂNSITO AGORA"*. 2013. Disponível em: <http://cetsp1.cetsp.com.br/noticias/detalhes/novositetransito.htm>.

CET. *Desempenho do Sistema Viário: Volumes e Velocidades - 2012*. São Paulo-SP, Brasil, 2013. Disponível em: <http://www.cetsp.com.br/media/228058/2012%20%20volumes%20e%20velocidades.pdf>.

CET. *Desempenho do Sistema Viário: Volumes e Velocidades - 2016*. São Paulo-SP, Brasil, 2017.

CINTRA, M. Os custos dos congestionamentos na cidade de São Paulo. p. 35, 2014. Disponível em: <http://bibliotecadigital.fgv.br/dspace/handle/10438/11576>.

COHEN, P.; HAHN, R.; HALL, J.; LEVITT, S.; METCALFE, R. *Using Big Data to Estimate Consumer Surplus: The Case of Uber*. [S.l.], 2016. Disponível em: <http://www.nber.org/papers/w22627>.

CONCAS, S.; KOLPAKOV, A. Synthesis of Research on Value of Time and Value of Reliability. jan. 2009. Disponível em: <https://trid.trb.org/view/884762>.

Consumo Combustível. *Consumo Combustível | Consumo Carros, Consumo Estrada, Consumo Cidade*. 2017. Disponível em: <https://consumocombustivel.com.br/>.

CPTM, C. P. d. T. M. *Relatório da Administração 2017*. São Paulo, 2017. Disponível em: <http://www.cptm.sp.gov.br/a-companhia/Pages/Relat%C3%B3rio-da-Administra%C3%A7%C3%A3o.aspx>.

CPTM, C. P. d. T. M. Institutional, *A Companhia | CPTM*. 2018. Disponível em: <http://www.cptm.sp.gov.br/a-companhia/Pages/a-companhia.aspx>.

CRAWFORD, G. S.; YURUKOGLU, A. The Welfare Effects of Bundling in Multichannel Television Markets. *American Economic Review*, v. 102, n. 2, p. 643–685, abr. 2012. ISSN 0002-8282. Disponível em: <https://www.aeaweb.org/articles?id=10.1257/aer.102.2.643>.

DALBEM, M. C.; BRANDãO, L.; MACEDO-SOARES, T. D. L. v. A. d. Economic appraisal of transportation projects: best practices and recommendations for Brazil. *Revista de Administração Pública*, v. 44, n. 1, p. 87–117, fev. 2010. ISSN 0034-7612. Disponível em: <http://www.scielo.br/scielo.php?script=sci_abstract&pid=S0034-76122010000100005&lng=en&nrm=iso&tlng=pt>.

DENATRAN. Ministério das Cidades, *Frota de Veículos*. 2018. Disponível em: <http://www.denatran.gov.br/estatistica/237-frota-veiculos>.

DUTZIK, T.; INGLIS, J.; BAXANDALL, P. Millennials in Motion: Changing Travel Habits of Young Americans and the Implications for Public Policy. out. 2014. Disponível em: <https://trid.trb.org/view/1328287>.

EMPLASA. *Região Metropolitana de São Paulo*. 2018. Disponível em: <https://www.emplasa.sp.gov.br/RMSP>.

EMTU. *Relatório Tarifário - RMSP*. [S.l.], 2012.

FARIA, H. A. d. Por onde vamos? : o engajamento, a colaboração e o crowdsourcing no aplicativo Waze. abr. 2014. Disponível em: <http://bdm.unb.br/handle/10483/7337>.

FAZENDA, G. do Estado de São Paulo Secretaria da. Portal da Transparência, *Repasse de Tributos a Municípios*. 2017. Disponível em: <https://www.fazenda.sp.gov.br/RepasseConsulta/Consulta/repasse.aspx>.

FAZENDA, G. do Estado de São Paulo Secretaria da. *Sobre o IPVA*. 2017. Disponível em: <https://portal.fazenda.sp.gov.br:443/servicos/ipva/Paginas/Sobre.aspx>.

Fenabrave. *Emplacamentos | Fenabrave*. 2017. Disponível em: <http://www3.fenabrave.org.br:8082/plus/modulos/listas/index.php?tac=indices-e-numeros&idtipo=1&layout=indices-e-numeros>.

FERES, L. J. *Diferencial de tarifa entre pico e vale como ferramenta de suavização da demanda no sistema de transporte público da cidade de São Paulo*. Tese (Doutorado), 2015. Disponível em: <http://dspace.insper.edu.br/xmlui/handle/11224/701>.

FIX, N. Y. C. *Advisory Panel Report*. [S.l.], 2018.

FODERARO, L. W. Navigation Apps Are Turning Quiet Neighborhoods Into Traffic Nightmares. *The New York Times*, Leonia, NJ, USA, dez. 2017. Disponível em: <https://www.nytimes.com/2017/12/24/nyregion/traffic-apps-gps-neighborhoods.html>.

FODERARO, L. W. New Jersey Town Aims to Keep App-Guided Outsiders Off Its Streets. *The New York Times*, jan. 2018. ISSN 0362-4331. Disponível em: <https://www.nytimes.com/2018/01/22/nyregion/leonia-gps-navigation-apps.html>.

Folhapress. São Paulo registra o maior congestionamento da história. *Folha de S.Paulo*, São Paulo-SP, Brasil, maio 2014. Disponível em: <http://www1.folha.uol.com.br/cotidiano/2014/05/1459269-sao-paulo-registra-o-maior-congestionamento-da-historia.shtml>.

HADDAD, E. A.; VIEIRA, R. S. Mobilidade, acessibilidade e produtividade: nota sobre a valoração econômica do tempo de viagem na região metropolitana de São Paulo. *Núcleo de Economia Regional e Urbana, Universidade de São Paulo*, 2015.

HAGHANI, A.; HAMEDI, M.; SADABADI, K.; YOUNG, S.; TARNOFF, P. Data Collection of Freeway Travel Time Ground Truth with Bluetooth Sensors. *Transportation Research Record: Journal of the Transportation Research Board*, v. 2160, p. 60–68, set. 2010. ISSN 0361-1981. Disponível em: <http://trrjournalonline.trb.org/doi/abs/10.3141/2160-07>.

HANDY, S. L.; YANTIS, T. *The Impacts of Telecommunications Technologies on Nonwork Travel Behavior*. Austin, Texas, 1997. 62 p.

HENSHER, D. A.; ROSE, J. M.; GREENE, W. H. *Applied Choice Analysis: A Primer*. [S.l.]: Cambridge University Press, 2005. Google-Books-ID: 8yZrtCCABAgC. ISBN 978-0-521-84426-0.

HERRERA, J. C.; WORK, D. B.; HERRING, R.; BAN, X. J.; JACOBSON, Q.; BAYEN, A. M. Evaluation of traffic data obtained via GPS-enabled mobile phones: The Mobile Century field experiment. *Transportation Research Part C: Emerging Technologies*, v. 18, n. 4, p. 568–583, ago. 2010. ISSN 0968-090X. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0968090X09001430>.

HäGERSTRAND, T. What about people in Regional Science? *Papers of the Regional Science Association*, v. 24, n. 1, p. 6–21, dez. 1970. ISSN 0486-2902, 1435-5957. Disponível em: <https://link.springer.com/article/10.1007/BF01936872>.

IBGE, I. B. d. G. e. E. *IBGE Cidades*. 2017. Disponível em: <https://cidades.ibge.gov.br/>.

IBGE, I. B. d. G. e. E. *Produto Interno Bruto dos Municípios | Estatísticas | IBGE*. 2017. Disponível em: <https://www.ibge.gov.br/estatisticas-novoportal/economicas/contas-nacionais/9088-produto-interno-bruto-dos-municipios.html?=&t=o-que-e>.

IDEIA, B. D.; SOCIEDADE, I. C. e.; ESCOLHAS, I. *Mobilidade Urbana e Baixo Carbono - IDEIA/iCS/Escolhas: Relatório Quantitativo*. [S.l.], 2018. Disponível em: <http://climaesociedade.org/wp-content/uploads/2018/01/IDEIA-Relat%C3%B3rio-Quantitativo-ICS-14DEZ17.pdf>.

Inmetro. *Inmetro - Tabelas de consumo/eficiência energética*. 2017. Disponível em: <http://www.inmetro.gov.br/consumidor/tabelas.asp>.

Institute of Transportation Studies UC Berkeley. *Driving Under the Influence of Apps*. 2018. Disponível em: <https://www.youtube.com/watch?v=NOkYK2yuyeM>.

JAVANMARDI, M.; LANGERUDI, M. F.; SHABANPOUR, R.; MOHAMMADIAN, K. Mode Choice Modelling Using Personalized Travel Time and Cost Data. In: *14th International Conference on Travel Behaviour Research*. [S.l.: s.n.], 2015.

KLEINA, N. *Erro no Waze deixa trânsito de São Paulo (ainda mais) congestionado*. 2017. Disponível em: <https://www.tecmundo.com.br/mobilidade-urbana-smart-cities/123450-erro-waze-deixa-transito-paulo-ainda-congestionado.htm>.

LOBEL, F. Acidentes com mortes nas marginais contrariam discurso da gestão. *Folha de S.Paulo*, São Paulo-SP, Brasil, dez. 2017. Disponível em: <http://www1.folha.uol.com.br/cotidiano/2017/12/1940174-acidentes-com-mortes-nas-marginais-contrariam-discurso-da-gestao-doria.shtml>.

LUCINDA, C. R.; MOITA, R. M. S.; MEYER, L. G.; LEDO, B. A. The Economics of Sub-optimal Policies for Traffic Congestion. *Journal of Transport Economics and Policy (JTEP)*, v. 51, n. 4, p. 225–248, out. 2017.

LUZ, H. d. M. *Três ensaios sobre contribuições econômicas para a escolha de modal*. Tese (Doutorado), 2017. Disponível em: <http://acervodigital.ufpr.br/handle/1884/47776>.

MACKIE, P.; NELLTHORP, J.; LAIRD, J. *Valuation of Time Savings*. Washington, DC, 2005. Disponível em: <https://openknowledge.worldbank.org/handle/10986/11800>.

MAS-COLELL, A.; WHINSTON, M. D.; GREEN, J. R.; others. *Microeconomic theory*. [S.l.]: Oxford university press New York, 1995. v. 1.

MCFADDEN, D. The measurement of urban travel demand. *Journal of public economics*, v. 3, n. 4, p. 303–328, 1974.

MCFADDEN, D.; TRAIN, K. Mixed MNL Models for Discrete Response. *Journal of Applied Econometrics*, v. 15, n. 5, p. 447–470, 2000. ISSN 0883-7252. Disponível em: <http://www.jstor.org/stable/2678603>.

MCFADDEN, D.; TYE, W. B.; TRAIN, K. *An application of diagnostic tests for the independence from irrelevant alternatives property of the multinomial logit model*. [S.l.]: Institute of Transportation Studies, University of California, 1977.

MEIRELLES, F. d. S. 28º pesquisa anual do uso de ti. Centro de Tecnologia de Informação Aplicada (GVcia), 2017.

MENA, F.; VELOSO, L. Pedágio urbano é visto como solução inescapável para o trânsito. *Folha de S.Paulo*, São Paulo-SP, Brasil, out. 2013. Disponível em: <http://www1.folha.uol. com.br/cotidiano/2013/10/1355285-pedagio-urbano-e-visto-como-solucao-inescapavel. shtml>.

METRÔ-SP. *Pesquisa de Mobilidade 2012 - Síntese das Informações*. [S.l.], 2013.

METRÔ-SP. *Metrô São Paulo | Quem somos - saiba sobre transporte metropolitano*. 2018. Disponível em: <http://www.metro.sp.gov.br/metro/institucional/quem-somos/index. aspx>.

MOITA, R. M. S.; LOPES, C. E. M. Demanda por meios de transporte na grande São Paulo : uma análise de políticas públicas. *Pesquisa e Planejamento Econômico*, Instituto de Pesquisa Econômica Aplicada (Ipea), ago. 2016. Disponível em: <http://repositorio.ipea.gov.br/handle/11058/7293>.

MOKHTARIAN, P. L.; HANDY, S. L.; SALOMON, I. Methodological issues in the estimation of the travel, energy, and air quality impacts of telecommuting. *Transportation Research Part A: Policy and Practice*, v. 29, n. 4, p. 283–302, jul. 1995. ISSN 0965-8564. Disponível em: <http://www.sciencedirect.com/science/article/pii/096585649400029A>.

MOKHTARIAN, P. L.; ORY, D. T. Don't Work, Work at Home, or Commute? Discrete Choice Models of the Decision for San Francisco Bay Area Residents. mar. 2005. Disponível em: <https://escholarship.org/uc/item/5cs0q85s>.

MORTIMER, J. H. Price Discrimination, Copyright Law, and Technological Innovation: Evidence from the Introduction of DVDs. *The Quarterly Journal of Economics*, v. 122, n. 3, p. 1307–1350, ago. 2007. ISSN 0033-5533. Disponível em: <https://academic.oup.com/qje/article/122/3/1307/1879537>.

MotoClube. *Moto Clube.* 2017. Disponível em: <http://motoclube.com/>.

Motonline. *Motonline | Motos no Portal Mais Completo do Brasil.* 2017. Disponível em: <http://www.motonline.com.br/>.

PACHECO, T. S.; CHAGAS, A. *Transport's demand in the Metropolitan Region of São Paulo and urban toll policy as a measure to reduce congestion.* [S.l.], 2016.

PARRY, I. W. Pricing urban congestion. *Annu. Rev. Resour. Econ.*, v. 1, n. 1, p. 461–484, 2009.

PARRY, I. W. H.; WALLS, M.; HARRINGTON, W. Automobile Externalities and Policies. *Journal of Economic Literature, 45 (2), 373-399.*, v. 45, n. 2, p. 373–399, 2007. ISSN 0022-0515.

PATIRE, A. D.; WRIGHT, M.; PRODHOMME, B.; BAYEN, A. M. How much GPS data do we need? *Transportation Research Part C: Emerging Technologies*, v. 58, p. 325–342, set. 2015. ISSN 0968-090X. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0968090X15000662>.

Prefeitura da Cidade de São Paulo. *Secretaria Municipal de Mobilidade e Transportes | Prefeitura da Cidade de São Paulo.* 2012. Disponível em: <http://www.prefeitura.sp.gov.br/cidade/secretarias/transportes/noticias/?p=45950>.

Prefeitura da Cidade de São Paulo. *Institucional | Secretaria Municipal de Mobilidade e Transportes | Prefeitura da Cidade de São Paulo.* 2018. Disponível em: <http://www.prefeitura.sp.gov.br/cidade/secretarias/transportes/institucional/sptrans/acesso_a_informacao/index.php?p=179590>.

QUAN, T.; WILLIAMS, K. *Product Variety, Across-Market Demand Heterogeneity, and the Value of Online Retail.* Rochester, NY, 2017. Disponível em: <https://papers.ssrn.com/abstract=2993236>.

SANTINO, R. *Falha do Waze agrava congestionamento em vias de São Paulo.* 2017. Disponível em: <https://olhardigital.com.br/noticia/falha-do-waze-agrava-congestionamento-em-vias-de-sao-paulo/71912>.

SANTOS, C. C. d. O valor do tempo na avaliação de projetos de transportes. 2012. Disponível em: <http://www.lume.ufrgs.br/handle/10183/61365>.

SCARINGELLA, R. S. a Crise Da Mobilidade Urbana Em São Paulo. *São Paulo em Perspectiva*, v. 15, n. 1, p. 55–59, 2001. ISSN 0102-8839. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-88392001000100007&lng=pt&nrm=iso&tlng=pt>.

SHAHEEN, S.; COHEN, A. Is It Time for a Public Transit Renaissance?: Navigating Travel Behavior, Technology, and Business Model Shifts in a Brave New World. *Journal of Public Transportation*, v. 21, n. 1, jan. 2018. ISSN 1077-291X. Disponível em: <http://scholarcommons.usf.edu/jpt/vol21/iss1/8>.

SILVA, T. H.; MELO, P. O. S. V. d.; VIANA, A. C.; ALMEIDA, J. M.; SALLES, J.; LOUREIRO, A. A. F. Traffic Condition Is More Than Colored Lines on a Map: Characterization of Waze Alerts. In: *Social Informatics*. Springer, Cham, 2013. (Lecture Notes in Computer Science), p. 309–318. ISBN 978-3-319-03259-7 978-3-319-03260-3. Disponível em: <https://link.springer.com/chapter/10.1007/978-3-319-03260-3_27>.

SPTRANS, S. P. T. S. *SPTrans Institucional*. 2018. Disponível em: <http://www.sptrans.com.br/a_sptrans/>.

TomTom. *About | TomTom Traffic Index*. 2016. Disponível em: <https://www.tomtom.com/en_gb/trafficindex/about>.

TRAIN, K. E. *Discrete Choice Methods with Simulation*. [S.l.]: Cambridge University Press, 2009. Google-Books-ID: 4yHaAgAAQBAJ. ISBN 978-1-139-48037-6.

UC Berkeley. *The Impact of Routing Apps on Traffic: Alexandre Bayen*. 2017. Disponível em: <https://www.youtube.com/watch?v=OTYHrozkazg>.

UOL Carros. *Motos: notícias, lançamentos e avaliações sobre carros - UOL Carros*. 2017. Disponível em: <https://carros.uol.com.br/motos>.

VICKREY, W. L. American Economic Association Pricing in Urban and Suburban Transport Author (s): William S . Vickrey Source : The American Economic Review , Vol . 53 , No . 2 , Papers and Proceedings of the Seventy-Fifth Annual Meeting of the American Economic Associa. *American Economic Review*, v. 53, n. 2, p. 452–465, 1963.

VICKREY, W. L. American Economic Association The Process of Financial Innovation Author (s): William L . Silber Source : The American Economic Review , Vol . 73 , No . 2 , Papers and Proceedings of the Ninety-Fifth Annual Meeting of the American Economic Association (. *American Economic Review*, v. 59, n. 2, p. 251–260, 1969.

VLAHOGIANNI, E. I.; KARLAFTIS, M. G.; GOLIAS, J. C. Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C: Emerging Technologies*, v. 43, p. 3–19, jun. 2014. ISSN 0968-090X. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0968090X14000096>.

WALTERS, A. a. a. The Theory and Measurement of Private and Social Cost of Highway Congestion Published by : Econometric Society Stable URL : http://www.jstor.org/stable/1911814 Accessed : 05-03-2016 08 : 06 UTC Your use of the JSTOR archive indicates your acceptance of th. v. 29, n. 4, p. 676–699, 1961.

WANG, P.; HUNTER, T.; BAYEN, A. M.; SCHECHTNER, K.; GONZáLEZ, M. C. Understanding Road Usage Patterns in Urban Areas. *Scientific Reports*, v. 2, p. 1001, dez. 2012. ISSN 2045-2322. Disponível em: <https://www.nature.com/articles/srep01001>.

Waze. *2016 Driver Satisfaction Index*. [S.l.], 2016. Disponível em: <https://inbox-static.waze.com/driverindex.pdf>.

WEE, B. van; GEURS, K.; CHORUS, C. Information, communication, travel behavior and accessibility. *Journal of Transport and Land Use*, v. 6, n. 3, p. 1–16, 2013. ISSN 1938-7849. Disponível em: <http://www.jstor.org/stable/26202660>.

WESTERMAN, M.; LITJENS, R.; LINNARTZ, J.-p. Integration Of Probe Vehicle And Induction Loop Data: Estimation Of Travel Times And Automatic Incident Detection. *UC Berkeley: California Partners for Advanced Transportation Technology*, jan. 1996. Disponível em: <https://escholarship.org/uc/item/8mh629c3>.

WRIGHT, J.; JOY, D. Using Vehicles Equipped with Toll Tags as Probes for Providing Travel Times. *UC Berkeley: California Partners for Advanced Transportation Technology*, abr. 2001. Disponível em: <https://escholarship.org/uc/item/9f17h2j0>.

Appendix A: Descriptive statistics of estimated model's variables

The original Mobility Survey dataset has only trips people actually did. To estimate the mode choice demand model, we have to complete the dataset with potential trips that the same individuals could have done from the same origin to the same destination but using alternative modes of transport. Additionally, literature recommends not to estimate the structural demand models using trips with different motivations, for example, mode choice for trips done because of work and for trips whose reason was leisure should be estimated separately. Our main specification in this thesis is only for trips motivated by work.

Because of the reasons above and for completeness, we provide descriptive statistics of the variables used in the models for the trips from the original dataset, for the trips of the expanded dataset and for the trips inputted into the main specification estimation, respectively.

Observations: Only really undertaken trips (actual chosen modes)

variable	N	mean	p50	min	max	sd
age	45,774	34.01	33.00	1.00	95.00	18.42
age squared	45,774	1,495.76	1,089.00	1.00	9,025.00	1,421.42
car out	45,774	0.13	0.00	0.00	1.00	0.34
downtown	45,774	0.13	0.00	0.00	1.00	0.34
female	45,774	0.50	0.00	0.00	1.00	0.50
financial costs	45,774	2.23	0.96	0.00	189.62	4.34
income (R$1000)	45,774	4.10	2.90	0.00	80.00	4.40
cost/income	45,470	0.89	0.19	0.00	189.76	2.18
time x income	45,774	147.63	76.00	0.00	6,600.00	214.81
peak	45,774	0.57	1.00	0.00	1.00	0.49
travel time	45,774	38.16	25.00	1.00	270.00	36.39
working trip	45,774	0.50	1.00	0.00	1.00	0.50

Observations: All the trips, all the modes that could be chosen

variable	N	mean	p50	min	max	sd
age	181,515	34.40	33.00	1.00	95.00	18.10
age squared	181,515	1513.00	1,089.00	1.00	9,025.00	1402.00
car out	181,515	0.14	0.00	0.00	1.00	0.35
downtown	181,515	0.12	0.00	0.00	1.00	0.33
female	181,515	0.49	0.00	0.00	1.00	0.50
financial costs	181,515	11.20	3.00	0.00	360.00	25.40
income (R$1000)	181,515	4.24	3.00	0.00	80.00	4.37
cost/income	180,490	4.75	0.85	0.00	2,681.00	16.80
time x income	181,515	200.00	84.80	0.00	12,509.00	376.00
peak	181,515	0.56	1.00	0.00	1.00	0.50
travel time	181,515	49.10	27.40	0.02	1058.00	67.80
working trip	181,515	0.52	1.00	0.00	1.00	0.50

Unchosen trips by car and by motorcycle are excluded if the traveler's household has no car or motorcycle

Observations: All the modes that could be chosen, only for trips motivated by work

variable	N	mean	p50	min	max	sd
age	95,004	37.90	36.00	1.00	93.00	13.20
age squared	95,004	1612.00	1296.00	1.00	8649.00	1094.00
car out	95,004	0.17	0.00	0.00	1.00	0.38
downtown	95,004	0.14	0.00	0.00	1.00	0.34
female	95,004	0.43	0.00	0.00	1.00	0.50
financial costs	95,004	13.80	3.00	0.00	360.00	29.40
income (R$1000)	95,004	4.46	3.30	0.00	80.00	4.42
cost/income	94,956	5.42	0.94	0.00	439.00	15.20
time x income	95,004	260.00 & 128.00	0.00	12509.00	445.00	
peak	95,004	0.55	1.00	0.00	1.00	0.50
travel time	95,004	62.90	39.20	0.02	1058.00	77.70
working trip	95,004	1.00	1.00	1.00	1.00	0.00

Annex I: Waze's official declaration about the malfunction

This message regarding the events of October 23rd, 2017 was received by email from the Public Relations Company on behalf of Waze in Brazil:

"O Brasil é uma das maiores e mais ativas redes de usuários do Waze no mundo. Temos ciência da importância da plataforma no dia a dia da nossa comunidade brasileira e trabalhamos para manter a excelência esperada do nosso serviço. Passamos por uma situação pontual e bastante rara que afetou alguns usuários, porém a questão já está resolvida.

Investimos constantemente em inovação da plataforma com o lançamento de novas funcionalidades - incluindo recursos solicitados pelos nossos usuários no Brasil, como o Rodízio Veicular e o Lembrete de Faróis para as estradas e rodovias.

Também firmamos parceiras com orgãos públicos locais para oferecer maior utilidade nos nossos serviços. Por exemplo, recentemente, anunciamos a nossa parceira com a cidade de São Paulo em uma iniciativa que permite que os nossos usuários reportem semáforos apagados diretamente para a central de controle da CET, facilitando o processo de monitoramento e manutenção."

Annex II: 10/23/2017 chart plotting the Slowness Index

This chart was published by "O Estado de S. Paulo" newspaper (CAPELAS; RIBEIRO; MENGUE, 2017a) at 10/23/2017 11:23 p.m. The Slowness Index evolution along October 23rd, 2017 is represented by the red line. The dark green line is the mean upper limit from Mondays of the 12 past months and the light green is the lower limit. The limits are calculated through statistical methods by CET. (CET, 2013a)
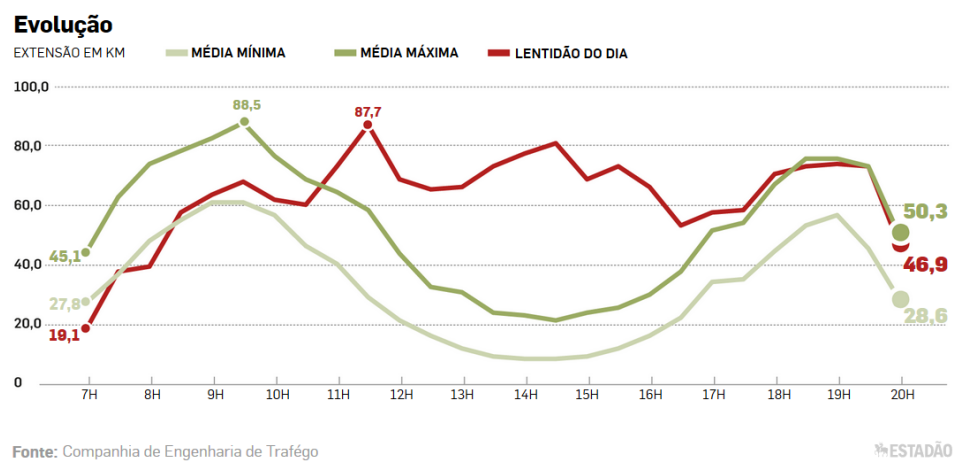
The Slowness Index chart from 10/23/2017. Source: CET *apud* Capelas, Ribeiro & Mengue (2017a)