UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

PUBLIC TRANSPORT OPTIMIZATION UNDER DYNAMIC CONGESTION AND MODE CHOICE

TESIS PARA OPTAR AL GRADO DE
DOCTOR EN SISTEMAS DE INGENIERÍA

FERNANDO DAVID FERES TORREBLANCA

PROFESOR GUÍA:
LEONARDO JAVIER BASSO SOTZ

PROFESOR CO-GUÍA:
HUGO EMILIO SILVA MONTALVA

MIEMBROS DE LA COMISIÓN:
JOSÉ RAFAEL CORREA HAEUSSLER
SERGIO RODOLFO JARA DÍAZ
ERIK TEODOOR VERHOEF

SANTIAGO DE CHILE
2020

RESUMEN DE LA TESIS PARA OPTAR
AL GRADO DE DOCTOR EN SISTEMAS DE INGENIERÍA
POR: FERNANDO DAVID FERES TORREBLANCA
FECHA: 2020
PROF. GUIA: LEONARDO JAVIER BASSO SOTZ
PROF. CO-ADVISOR: HUGO EMILIO SILVA MONTALVA

PUBLIC TRANSPORT OPTIMIZATION UNDER DYNAMIC CONGESTION AND MODE CHOICE

La demanda de transporte público tiene un lugar esencial en los desplazamientos urbanos. En Europa, el transporte público concentra el 31% de los viajes en las 28 ciudades más grandes del continente, en América Latina este valor alcanza el 45% de los viajes. El creciente desarrollo económico y tecnológico ha contribuido a que un mayor número de personas tenga acceso al automóvil. Esta situación genera la posibilidad de que la gente elija entre el automóvil particular y el transporte público. Como resultado, existe una competencia por el espacio vial, el cual siempre es limitado. Resulta fundamental buscar políticas públicas que permitan el uso eficiente del espacio vial y reduzcan los costos asociados a estos viajes.

El primer paper de esta tesis es un análisis teórico de la eficiencia de los sistemas *Bus Rapid Transit* (BRT). Los sistemas BRT proporcionan vías segregadas a los buses que aumentan la velocidad. En este trabajo, utilizamos un enfoque de la congestión dinámica, lo cual considera que tanto la congestión vial como la congestión en las estaciones de BRT son endógenas al modelo. Mostramos analíticamente que, si la capacidad es perfectamente divisible, la implementación de un BRT es siempre eficiente (disminuye el costo social total). El análisis numérico permite demostrar que si la capacidad no es perfectamente divi, un BRT es eficiente en la mayoría de los casos. Además, el BRT puede inducir una mejora de Pareto en la que disminuyan tanto los costos de tiempo como los costos operativos del transporte público. Comparado con el óptimo cuando los buses circulan en tráfico mixto, el sistema BRT óptimo tiene: i) un período más corto de funcionamiento de los buses y los automóviles, ii) una mayor frecuencia y, lo que es muy importante, iii) más demoras para abordar los buses, es decir, colas más largas las paradas. El punto ii) implica que, aunque para cierto nivel de demanda, puede ser óptimo no prestar servicios de transporte público en tráfico mixto, con un BRT puede ser eficiente.

El segundo paper analiza la provisión eficiente de un sistema de transporte público operado por autobuses, los que comparten la capacidad víal con el automóvil. Proponemos un modelo de congestión dinámico con partición modal, con el transporte público y los autómoviles como sustitutos. El período de congestionado, el patrón de salida y la cola en la parada de buses son endógenos. Definimos frecuencias diferentes para el período congestionado y el no congestionado con el objetivo de capturar explícitamente los efectos de la congestión. Mostramos bajo condiciones plausibles que la frecuencia eficiente en el período no congestionado es mayor que la frecuencia en el período congestionado. Si comparamos nuestros resultados, usando análisis numérico, con el caso en que la frecuencia es constante, tenemos (i) menor umbral donde es eficiente ofrecer transporte público, (ii) costos de los usuarios menores, (iii) las frecuencias óptimas son mayores, (iv) mayores gastos de operación, y (v) mayores demoras para abordar el bus. Los puntos ii) y iii) implican que el patrón de frecuencia eficiente reduce los costos de los usuarios al aumentar la frecuencia y los gastos operativos.

ABSTRACT
AUTHOR: FERNANDO DAVID FERES TORREBLANCA
DATE: 2020
ADVISOR: LEONARDO JAVIER BASSO SOTZ
CO-ADVISOR: HUGO EMILIO SILVA MONTALVA

PUBLIC TRANSPORT OPTIMIZATION UNDER DYNAMIC CONGESTION AND MODE CHOICE

In large cities, the demand for public transport has an essential place in urban commuting. In the case of Europe, public transport concentrates 31% of all journeys in the 28 largest cities, while in Latin America this number reaches 45%. In the context of big cities, growing economic and technological development has contributed to a higher number of people having access to cars. This situation generates the possibility of people being able to choose between the car and public transport to make their commute. As a result, there is competition for road space, which is always limited. It is, therefore, essential to seek public policies aimed at an efficient use of road space and a reduction in the costs associated with these journeys.

The first paper in this thesis is a theoretical analysis of the efficiency of Bus Rapid Transit Systems (BRT). BRT systems provide segregated road capacity to buses to increase their speed. In this paper, we propose a dynamic congestion approach that endogenously models queuing both on the road and at BRT stations, which are at the center of our interest. We show analytically that, if capacity is perfectly divisible, implementing a BRT is always efficient (it decreases total social cost). We show numerically that if capacity is not perfectly divisible, a BRT is efficient in most cases. Moreover, BRT can induce a Pareto Improvement where both time costs and public transport operating costs decrease. Compared to the optimum when buses run in mixed traffic, the optimal BRT system has: (i) shorter period of bus operation and car-peak period, (ii) greater frequency and, very importantly, (iii) more boarding delays, i.e. longer queues at bus stops. Point (ii) implies that while it may be optimal not to provide any public transport service under mixed traffic for some levels of demand, it may well be worthwhile with a BRT.

The second paper analyzes the efficient provision of a public transport system operated by buses that share the road capacity with cars. We propose a dynamic congestion model with mode choice, where public transport and cars are substitutes modes. The congested period, the departure pattern, and the queuing at the bus stop are endogenous to the model. We define different frequencies for the congested and the uncongested period to explicitly capture the effects of congestion on the optimal frequency pattern. We show under plausible conditions that the efficient frequency during the uncongested period is higher than the frequency during the congested period. If we compare our results, using numerical analysis, versus cases where the frequency is constant, we can ascertain (i) the demand threshold for which it becomes efficient to provide public transport is lower, (ii) user costs are lower, (iii) optimal frequencies in both congested and uncongested periods are higher, (iv) the operational expenditure is higher, and (v) boarding delays at the bus stop increases. Points (ii) and (iii) imply that the efficient frequency pattern reduces user costs by increasing the frequency and operational expenditure.

A Inti
To Inti

Acknowledgements

Writing these acknowledgments maybe is the most challenging part of this thesis, and have no doubt that writing this thesis was a great effort. I would like to sincerely thank my family. In particular, I would like to thank Inti, my wife. Without a doubt, you are the person who made this possible. Without your support, I could not have completed this stage. Of course, I would like to thank my two research advisors, Professor Leonardo Basso and Professor Hugo Silva. I appreciate their outstanding contributions and dedication to this project, which, many times, was more than the strictly academic issues. I would also like to thank Professor Robin Lindsey for his dedication and kindness when I was in Vancouver.

I want to thank all my doctoral colleagues, they have been a significant part of this process. I especially want to acknowledge: Cristiam, Dana, Eduardo L. Eduardo Z, Renny, Vale, and Vero. Additionally, I would like to thank Giuseppe for all the conversations, academic and non-academic, and Vancouver's good times. I would like to thank Martina for all the willingness she always had to talk about transportation economics and to hear about my project.

Finally, thank you very much to all who were part of this stage.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The objective of this thesis is to study public transport optimization using a classic bottleneck model in two different cases. First, we analyze the efficiency of an infrastructure measure, a bus rapid transit system (BRT) (Chapter 2). Secondly, we consider that it is not possible to change infrastructure and analyze a management measure (Chapter 3).Both chapters have a common framework: dynamic congestion through a bottleneck model. The dynamic congestion modeling has several relevant points that are considering in this thesis. First, the departure time scheduling to and the arrival pattern to destination are endogenous, and second, queuing at the bus station is also endogenous. Consequently, all optimization made in public transport will change the arrival pattern and modify the optimization results.

In a dynamic congestion environment, only a limited amount of studies have been carried out on two-mode systems considering buses in mixed traffic. Moreover, the only previous paper that analyzed this area is that Huang et al. (2007). From a methodological point of view, this thesis contributes to both the literature on public transit and the bottleneck model.

This thesis is comprised of four chapters, including the Introduction and conclusions. Chapter 2 has been published, while Chapter 3 is a draft manuscript written in a consistent format, then each one is self-contained. We now offer a brief explanation of the contribution of each Chapter.

While the literature on BRTs is extensive regarding their best operation, design, or their urban and financial impact, there seldom have been analysed from a transport economics point of view. In Chapter 2 we propose a dynamic congestion approach, which is equipped to model queuing endogenously, both on the road and at BRT stations, which are the center of our interest. Commuters travel from a single residential area to the city center and have to choose to either drive or take public transportation, together with the departure time, which makes schedule delays important. If a commuter decides to travel by car, she may face road congestion. If she chooses to use public transport instead, she will need to go to a station where she may face boarding delays caused by queues. The bus will then go into the road, where it may join the queue of cars if traffic is mixed (without BRT), or it may not face road queuing if part of the capacity is devoted to a BRT (which decreases the capacity for cars). We focus on second-best policies, both for mixed traffic conditions and BRT, meaning that we consider that fares and tolls are time-invariant, and the public transport system operates at a constant headway. This framework makes the problem tractable and, we believe, at the same time, better represents what is and can be implemented in most cities.

We provide microeconomic analysis of BRTs in the context of dynamic congestion, where queuing and congestion delays are endogenous as a result of individual schedule of departures. The main difference with previous literature that looks into bimodal (car and public transport) systems is that, rather than focusing on crowding and assuming from the outset that capacities of each mode are independent, we focus on modeling boarding delays in equilibrium, and comparing mixed traffic conditions with what would arise from dedicating part of the road capacity to a BRT.

Our primary, most policy-relevant result is that in a second best-world where fares cannot vary perfectly with time, BRTs are efficient and have the potential to provide a Pareto improvement of the transport system: in equilibrium both bus users and car users can be better off, while the costs of providing public transport decrease. With a BRT, fares will be lower because the peak hours of operation of the system are shorter. The car peak-period is also shorter, despite the fact that capacity was taken from private transport. Importantly, this better-for-all situation features more boarding delays, that is, queues at bus stops will be longer than under mixed-traffic conditions. All these results provides strong support for the BRT surge observed around the globe while providing one –possibly not the only one– explanation for observed longer queues at stations. In the numerical analysis, we show that BRT is efficient. In most cases of indivisible capacity, a BRT is efficient.

In Chapter 3, we take into consideration that an efficient social design of a public transport system depends on the demand structure and especially on its time-of-day variation. We also consider that commuters can decide between using public transport and private cars, a modal choice that impacts the congestion level. All of this poses a severe challenge. This pattern not only affects the efficiency of public transport but also alters the cost for car users, modifying the system equilibrium. We deal with this problem using a dynamic congestion pattern that allows us to find the optimal frequency pattern considering modal elasticity, temporal elasticity, and congestion.

The main result of Chapter 3 is that, under plausible assumptions, the efficient frequency during the endogenous uncongested period is higher than the efficient frequency during the endogenous congested period. We show numerically that two-frequency optimization is efficient for users, reducing user cost, mainly from congestion , while increasing operator expenditure. Our numerical analysis shows that by only using a non-constant frequency optimization, without any additional road facilities for public transportation, we obtain a reduction of up to 14% in the social cost compared to the constant frequency case. Moreover, if we extend our analysis and compare the two-frequency (management measure) optimization against a BRT system (infrastructure measure), we find that the gains of a BRT system are less than 8% of the total cost.

Chapter 2

The efficiency of bus rapid transit (BRT) systems: a dynamic congestion approach

Abstract: The penetration of BRT systems has been increasing fast, although there have been many reports of heavy queuing to board the buses. We propose a dynamic congestion approach that endogenously models queuing both on the road and at BRT stations, which are the center of our interest. We show analytically that, if capacity is perfectly divisible, implementing a BRT is always efficient (it decreases total social cost), while we show numerically that if capacity is not perfectly divisible, a BRT is in most cases efficient. Moreover, BRT can induce a Pareto Improvement where both time costs and public transport operating costs decrease. Compared to the optimum when buses run in mixed traffic, the optimal BRT system has: (i) shorter period of bus operation and car-peak period, (ii) larger frequency and, very importantly, (iii) more boarding delays, i.e. longer queues at bus stops. Point (ii) implies that, while for some level of demands it may be optimal not to provide any public transport service under mixed traffic, with a BRT it may well be worthwhile.

Keywords: Bus Rapid Transit; Dynamic congestion; Bottleneck model; Bus stop delays

Diffusion: A version of this Chapter has been published as: Basso, Leonardo J. Feres, Fernando Silva, Hugo E. (2019). The efficiency of bus rapid transit (BRT) systems: A dynamic congestion approach. *Transportation Research Part B: Methodological,* 127:47 – 71.

2.1 Introduction

According to The Economist, city dwellers lose nearly U$S1,000 a year while sitting in traffic. Considering Great Britain, Germany, and the United States, congestion costs were estimated at US$461,000MM in 2017[1]. Of course, congestion affects commuters directly, but also other social costs are imposed on the rest of the population, such as pollution and noise. However, an additional cost, schedule delay costs, usually receives less attention. This cost occurs because, when there is congestion, people have to leave their homes earlier than desired or, in many cases, they arrive later than what they should.

One way to curb congestion is by taking measures to lure people into the public transport system since a transit vehicle moves a large number of people using less road capacity per person than a car. However, when public transport vehicles share road capacity with cars (think buses), they face a rather grim scenario: buses will always be slower than cars, while in addition they are usually considered less comfortable. Light- and heavy-rail may overcome the problem of speed, but are usually quite expensive.

There is an additional alternative, however, which is far less demanding financially: using part of the existing road capacity exclusively for buses. When buses are physically separated, through some investment, from cars, this has become to be known as a Bus Rapid Transit (BRT) system. A BRT system is defined, for example by the Institute for Transportation and Development Policy as "...a high-quality bus-based transit system that delivers fast, comfortable, and cost-effective services at metro-level capacities. It does this through the provision of dedicated lanes, with busways and iconic stations typically aligned to the center of the road, off-board fare collection, and fast and frequent operations"[2]. The same organization states that "Because BRT contains features similar to a light rail or metro system, it is much more reliable, convenient and faster than regular bus services. With the right features, BRT can avoid the causes of delay that typically slow regular bus services, like being stuck in traffic and queuing to pay on board"[3].

From the first system in Curitiba, Brazil, in 1977, the penetration of BRT systems has been increasing fast, mostly because of the promise of better, faster and cheaper public transport at a fraction of the cost of what a subway or heavy rail would cost. According to Global BRT Data Report from October 2018[4], in the year 2000 there were 40 cities with BRT systems, for a total constructed length of 1,100 kilometers. By 2018, the numbers exploded to 170 cities around the world, for a total of 376 corridors and 5,046 kilometers, while 121 additional cities are either building or have plans to build BRT systems. A regional panorama shows that Latin America leads with 171 corridors in 55 cities, followed by Asia, which has 94 BRT corridors in 43 cities, and Europe that has 58 corridors in 44 cities. North America has 37 corridors in 19 cities.

Despite this, not all BRTs have had a quiet life. There have been many reports of excess

[1] https://www.economist.com/blogs/graphicdetail/2018/02/daily-chart-20. Accessed on December 2018.

[2] Hensher et al. (2014) report that the mean peak-hour frequency, for 121 BRT systems on 12 countries is 116.1 buses per hour.

[3] https://www.itdp.org/library/standards-and-guides/the-bus-rapid-transit-standard/what-is-brt/. Accessed on December 2018.

[4] https://brtdata.org/. December 2018.

demand for the systems, which have taken the form of heavy queuing to board the buses. Well known are the cases of Transmilenio, in Bogota, Colombia, and Metrobus, in Istanbul, Turkey, but the Inter-American Development Bank has also reported heavy queuing in BRT systems in Lima (Peru), Montevideo (Uruguay) and Cali (Colombia); see for example Scholl et al. (2015, 2016). Lagos, in Nigeria, has also shown queuing problems in its BRT system[5].

While the literature on BRTs is extensive regarding their best operation, design, or their urban and financial impact, there seldom have been analysed from a transport economics point of view, which is what we do in this paper. The paper that comes closest to ours is Basso and Silva (2014), who study the efficiency and substitutability of bus lanes and pricing measures –such as congestion pricing– but do so in a static congestion framework. Here we propose a dynamic congestion approach, which is equipped to model queuing endogenously, both on the road and at BRT stations, which are the center of our interest. Commuters travel from a single residential area to the city center and have to choose to either drive or take public transportation, together with the departure time, which makes schedule delays important. If a commuter decides to travel by car, she may face road congestion. If she chooses to use public transport instead, she will need to go to a station where she may face boarding delays caused by queues. The bus will then go into the road, where it may join the queue of cars if traffic is mixed (without BRT), or it may not face road queuing if part of the capacity is devoted to a BRT (which decreases the capacity for cars). We focus on second-best policies, both for mixed traffic conditions and BRT, meaning that we consider that fares and tolls are time-invariant, and the public transport system operates at a constant headway. This framework makes the problem tractable and, we believe, at the same time, better represents what is and can be implemented in most cities.

The main result of our paper is that if capacity is perfectly divisible, implementing a BRT is always efficient in that it decreases total social cost. We show numerically that if capacity is not perfectly divisible, a BRT is, in most cases, efficient. Moreover, implementing BRT can induce a Pareto improvement where both users cost and public transport cost decrease. Compared to the optimum when buses run in mixed traffic, with a BRT, the transport system has (i) shorter hours of bus operation and car-peak period (ii) larger frequency and, very importantly, (iii) more boarding delays, i.e., longer queues at bus stops. Point (ii) implies that, while for some level of demands, it may be optimal not to provide any public transport service under mixed traffic, with a BRT, it may well be worthwhile. Point (iii) indicates that boarding delays are not necessarily a manifestation of poor operations.

As mentioned above, the paper that comes closest to ours in terms of the analyses sought is Basso and Silva (2014). Other papers that have analyzed dedicated bus lanes in a static framework are Mohring (1979), Small (1983), Kutzbach (2009), Basso et al. (2011) and Börjesson et al. (2017). Regarding dynamic congestion models, the literature is ample and followed the seminal papers by Vickrey (1969) and Arnott et al. (1990, 1993). But most of this literature focuses only on the case of private transport. The number of papers that deal with two modes is much slimmer. The two modes problem in a dynamic congestion framework was introduced in Tabuchi (1993), who considered a heavy rail, never congested, alternative to the road. Huang (2000) analyzed a similar setting but adding crowding costs,

[5]https://guardian.ng/features/executive-motoring/lagos-to-decongest-queues-at-brt-busstops/. Accessed on December 2018.

so that the tradeoff was not only between schedule and queuing delay and the transit fare, but also with crowding discomfort. Kraus and Yoshida (2002) and Yoshida (2008) add waiting time, specifically modeling the intermittent nature of the bus service. In this paper, we opt for a continuous frequency modeling, but use Krauss and Yoshida's approach to prove that this is indeed a very good approximation for the intermittent case, with quite small and bounded differences. Kraus (2003) optimizes the number of trains and the capacity of an individual train, thus affecting crowding and waiting times, while de Palma et al. (2017) add the analysis of optimal dynamic pricing of individual trains. van den Berg and Verhoef (2014) considers the effects of user heterogeneity on the car bottleneck – crowded train problem, while Wang et al. (2017) consider bottleneck capacity expansions and train subsidies.

Two modes systems but considering buses has been analyzed by Huang et al. (2007) and Gonzales and Daganzo (2012). In both papers, public transport is provided by buses that take up part of the bottleneck capacity. But Gonzales and Daganzo (2012) only consider the case when buses operate on a bus lane, and Huang et al. (2007) only model mixed traffic. They, therefore, do not analyze whether it is advantageous to separate buses from the car traffic; moreover, they focus on crowding costs while we focus on the boarding delays that may arise, in equilibrium, as a result of faster road travel by bus.

The structure of the paper is as follows. In Section 2.2 we describe the model and characterize the equilibrium under mixed-traffic conditions for any given public transport frequency and fare. Section 2.3 describes the first-best which entails no queuing delays and shows that time-variant fares and car tolls can decentralize it. We then, in Section 2.4, study the optimum under mixed traffic conditions, optimizing the time-invariant car toll, the bus fare, and the bus frequency. Section 2.5 studies the effects and efficiency of BRT systems both when capacity is perfectly divisible as when it is not. It also provides numerical examples to complement the analytical results. Finally, Section 2.6 summarizes the policy implications and concludes.

2.2 The model and equilibrium under mixed-traffic conditions

2.2.1 Basics

There are N identical users that travel from a residential area (H) to the Central Business District (CBD). All users have an identical desired arrival time to the CBD equal to t^*, and choose whether to travel by car or bus. As most of the analyses of dynamic road congestion, we follow the approach of Vickrey (1969) and Arnott et al. (1993) and use the bottleneck model. Travel by car is uncongested except for a bottleneck of capacity s, in which a queue develops if the combined arrival rate of vehicles exceeds the capacity. As the road bottleneck capacity is shared by both modes, it is the combined arrival rate of cars and buses that matters.

Public transport users, on the other hand, also have to walk to a station to access the bus and then wait for the bus. Waiting time is constant except for queuing to board the bus. A queue develops at the bus stop if the arrival rate of bus users is higher than the capacity of the system $k \cdot f$, where k is the fixed bus capacity and f the frequency. Throughout the paper, we assume that buses operate continuously with a constant headway given by $1/f$, so that the bus stop can be modeled as a bottleneck of capacity $k \cdot f$. The buses also have to pass through the bottleneck to reach the CBD, joining the cars since there is no bus lane and using a fraction $\lambda \cdot f/s < 1$ of the bottleneck capacity, where λ is an equivalence factor between buses and cars[6]. Figure 2.1 illustrates our setting.

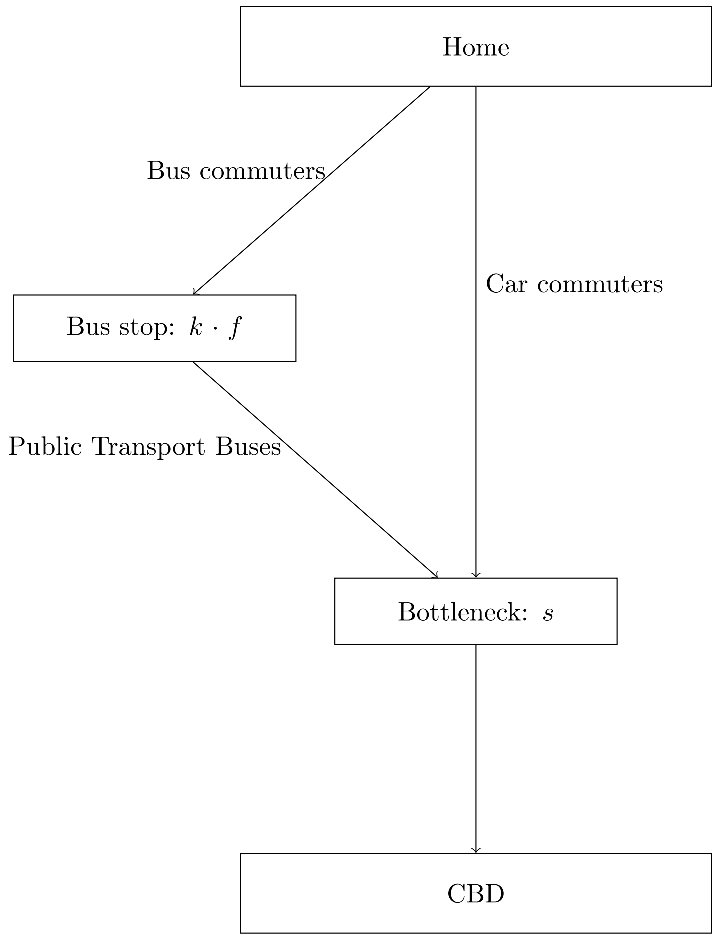


Figure 2.1: Two bottleneck diagram of the mixed traffic transport system.

It is essential to note that our model is a continuous approximation of a service of intermittent nature in which an integer number of buses of capacity k is dispatched at a constant headway (h, the time between each successive departure). This simplification, which has been used before (see, e.g., Huang et al., 2007), brings large advantages in terms of exposition and graphical analysis while only losing moderate generality. After all, frequencies are quite high in real BRT systems: Hensher et al. (2014) report a mean frequency in peak periods of 116 buses/hour for 121 systems over 12 different countries.

In Appendix A.1 we show, following Kraus and Yoshida (2002) and Yoshida (2008), that when the intermittent nature is modeled and passengers wait at bus stops to board the buses, there exists an equilibrium and it differs only slightly in terms of equilibrium costs from our continuous approximation (which is developed in the next subsection). In fact,

[6]This assumption ensures that the flow of buses alone is not large enough to generate road congestion

our model overestimates the user time equilibrium costs and this overestimation is inversely related to the frequency. Therefore, as frequency grows, our continuous model approaches the intermittent model. Moreover, in our numerical examples, we compute the overestimation of the aggregate equilibrium costs and find that it is always below 0.65%. Finally, the continuous approximation has been used in the transit provision literature (see, e.g., de Palma et al., 2017, Section 8) where the number of trains is treated as a continuous variable rather than restricted to integer values[7].

As it is customary in the literature, we follow Small (1982) and consider linear schedule delay costs. Therefore, individuals care about the difference between the actual arrival time and the desired arrival time, and every minute is valued equally. The individuals' time valuations are α for the value of (in-vehicle) travel time savings[8] and β and γ for the value of schedule delay early and late respectively. This is also referred to in the literature as $\alpha-\beta-\gamma$ preferences. The individuals' value of waiting time is denoted by α_2.

Denote N_c the number of car users. Following the standard bottleneck model Arnott et al. (1990, 1993), the generalized cost of a car user that departs from home at time t and arrives at the CBD at time t_a is:

$$c_c(t) = p_c + r_c + \alpha \cdot T_w(t) + \begin{cases} \beta \cdot (t^* - t_a) & \text{if } t_a \leq t^* \\ \gamma \cdot (t_a - t^*) & \text{if } t_a > t^* \end{cases} \tag{2.1}$$

where p_c is the car congestion price that the planner can set and r_c represents the resource (constant) costs of a trip which include fuel and parking costs, vehicle depreciation and constant travel times among others. α is the value of travel time savings, $T_w(t)$ is the travel time through the bottleneck, and the third term on the right-hand side of Eq. (2.1) is the schedule delay cost, which depends on whether the user arrives early or late. As usual in this model, we normalize travel times from the origin to the bottleneck and from the bottleneck to the destination to zero. Therefore, a user that departs at time t arrives at $t+T_w(t)$ to the CBD, i.e. $t_a = t + T_w(t)$.

The generalized cost of a bus trip follows the same logic. The difference is that waiting time is valued differently, at α_2, and that there are two sources of delays: road congestion and bus stop queuing. Denote N_b the number of bus users. The cost of a bus user that departs from home at time t and arrives at the CBD at time t_a is given by:

$$c_b(t) = p_b + r_b + \alpha_2 \cdot T_q(t) + \alpha \cdot T_w\left(t + T_q(t)\right) + \begin{cases} \beta \cdot (t^* - t_a) & \text{if } t_a \leq t^* \\ \gamma \cdot (t_a - t^*) & \text{if } t_a > t^* \end{cases} \tag{2.2}$$

where p_b is the fare and r_b is the resource (constant) cost of a trip which in this case includes access time costs, discomfort and constant travel time among others. α_2 is the value of waiting time, and $T_q(t)$ is the waiting time due to queuing at the bus stop. As a bus user that departs at t from home arrives at $t+T_q(t)$ at the bottleneck, the travel time through the bottleneck is $T_w\left(t + T_q(t)\right)$, which for simplicity is valued at α as well. Finally, the schedule

[7] We thank one anonymous referee and the editor for suggesting to study the relationship between the two approaches.

[8] We could consider different travel time cost between buses and cars, however including that complicates the expressions without conceptual gains

delay cost is the fourth term on the right-hand side of Eq. (2.2). A bus user that departs at time t from home arrives at the CBD at $t_a = t + T_q(t) + T_w(t + T_q(t))$.

Note that in the bottleneck model it is common to normalize to zero the time costs that are constant. In the case of a single mode this is without loss of generality, but for two modes this may not be the case. To avoid the loss of generality, as we explain above, we assume that the (resource) costs of each mode in the absence of tolls, p_c for cars and p_b for buses, include the costs of the constant travel times. For example, r_b includes the walking time to the station and in-vehicle travel times other than through the bottleneck. Furthermore, any difference that may make these two modes vertically differentiated, such as comfort, is also captured in the parameters r_c and r_b.

In general, equilibrium in our model must imply three conditions happening simultaneously: first, all car drivers must face the same total cost irrespective of their departure time; second, all bus users (if there is any) must also face the same total cost irrespective of their departure time; and, third, a car driver and bus user (if there is any) departing at the same time must face the same total cost. The first two conditions are dynamic equilibrium conditions. The third is the modal split condition.

2.2.2 Equilibrium with time-invariant prices

We now turn to the equilibrium in which prices are time-invariant, i.e. p_c and p_b do not depend on t and there are no bus lanes. Let t_c^s and t_c^e be the times of the first and last departure of an individual by car; this is what we define as the start and end of the car peak period. Analogously, let t_b^s and t_b^e be the times of the first and last departure of an individual by bus, or equivalently, the start and end of the bus peak period. The difference $[p_c + r_c] - [p_b + r_b]$ is key for modal equilibrium, and, for this reason, we characterize the equilibrium in three cases depending on whether the difference is negative, zero or positive. We refer to $p_m + r_m$ as the time-invariant full price of mode m. We begin by studying the case where $[p_c + r_c] - [p_b + r_c] > 0$. In all cases we assume a strictly positive frequency.

Lemma 2.1 *If the time-invariant full price of the car is higher than that of the bus, i.e., $p_c + r_c > p_b + r_b$, and both modes are used, the bus peak period starts earlier and ends later than the car peak hour. This is $t_b^s < t_c^s < t_c^e < t_b^e$.*

PROOF. We proceed by contradiction. Suppose that $p_c + r_c > p_b + r_b$, and the first bus user departs such that $t_b^s \geq t_c^s$; we will show that this cannot be an equilibrium. First, the total cost of a car user departing at t_b^s is $p_c + r_c + \beta \cdot (t^* - t_b^s) + \alpha \cdot T_w(t_b^s)$. The total cost of a bus user departing at t_b^s is $p_b + r_b + \beta \cdot (t^* - t_b^s) + \alpha \cdot T_w(t_b^s)$ because she will face congestion on the road (as $t_b^s \geq t_c^s$) but will face no queuing delay (as she is the first bus user). From the modal split equilibrium both total costs must be equal, implying $p_c + r_c = p_b + r_b$, which is a contradiction. Therefore, $t_b^s < t_c^s$. The proof that $t_c^e < t_b^e$ is analogous, since the last bus user does not face queuing delay either. Note that the argument applies for intermittent bus departures since the first-bus run users suffer no queuing delay. □

We now describe the equilibria for this two-modes system under mixed traffic conditions, starting with the situation when Lemma 2.1 holds, i.e., when $p_c + r_c > p_b + r_b$.

Buses and cars under mixed flow conditions share the bottleneck of capacity s, and, as the frequency of buses is constant (f), car users face a decreased capacity of $s - \lambda \cdot f$. Conditional on N_c, the start and end of the car peak period are given by equating the schedule delay costs for the first and last departure, as in Eq. (2.3). On the other hand, the length of the car peak must be such that all car drivers pass through the bottleneck, as in Eq. (2.4). Note that by Lemma 1, buses are in operation for all the car peak-period.

$$\beta(t^* - t_c^s) = \gamma(t_c^e - t^*) \tag{2.3}$$
$$(t_c^e - t_c^s) \cdot (s - \lambda \cdot f) = N_c \tag{2.4}$$

Following the same reasoning, the first and last departure by bus must only face schedule delay costs as in Eq. (2.5). On the other hand, the length of bus operations must be such that all commuters can actually go trough the bottleneck at the bus stop, as in Eq. (2.6).

$$\beta(t^* - t_b^s) = \gamma(t_b^e - t^*) \tag{2.5}$$
$$(t_b^e - t_b^s) \cdot k \cdot f = N_b \tag{2.6}$$

Solving the system of Eqs. (2.3) to (2.6) we obtain, conditional on N_b and N_c, the equilibrium times for car ans bus operations and the cost of traveling by each mode[9]. Defining $\delta = \beta \cdot \gamma / (\beta + \gamma)$, these are:

$$t_c^s = t^* - \frac{\delta}{\beta} \frac{N_c}{s - \lambda \cdot f} \tag{2.7}$$
$$t_c^e = t^* + \frac{\delta}{\gamma} \frac{N_c}{s - \lambda \cdot f} \tag{2.8}$$
$$t_b^s = t^* - \frac{\delta}{\beta} \frac{N_b}{k \cdot f} \tag{2.9}$$
$$t_b^e = t^* + \frac{\delta}{\gamma} \frac{N_b}{k \cdot f} \tag{2.10}$$
$$c_c = p_c + r_c + \delta \frac{N_c}{s - \lambda \cdot f} \tag{2.11}$$
$$c_b = p_b + r_b + \delta \frac{N_b}{k \cdot f} \tag{2.12}$$

The equilibrium modal split is obtained by equalization of costs across modes $(c_c = c_b)$ and using that $N_c + N_b = N$. The solution is unique and given by:

$$N_c = (s - \lambda \cdot f) \cdot \frac{N - \frac{(p_c + r_c - p_b - r_b) \cdot k \cdot f}{\delta}}{s - \lambda \cdot f + k \cdot f} \tag{2.13}$$
$$N_b = k \cdot f \cdot \frac{N + \frac{(p_c + r_c - p_b - r_b) \cdot (s - \lambda \cdot f)}{\delta}}{s - \lambda \cdot f + k \cdot f} \tag{2.14}$$

[9]The cost of traveling by car can be calculated from the first car departure, by multiplying β times $(t^* - t_c^s)$ from Eq. (2.7), since the dynamic equilibrium of each mode requires that the total cost of traveling in each mode is constant irrespective of the departure time. Using Eq. (2.9) the cost of traveling by bus can be computed.

Finally, the equilibrium cost c is:

$$c = p_c + r_c + \delta \frac{N - \frac{(p_c+r_c)-(p_b+r_b)\cdot k\cdot f}{\delta}}{s - \lambda \cdot f + k \cdot f} = p_b + r_b + \delta \frac{N + \frac{(p_c+r_c)-(p_b+r_b)\cdot(s-\lambda\cdot f)}{\delta}}{s - \lambda \cdot f + k \cdot f} \tag{2.15}$$

With the previous results we can now state our first proposition regarding equilibrium under mixed traffic condition:

Proposition 2.2 *If the time-invariant full price of the car is higher than of the bus, i.e., $p_c + r_c > p_b + r_b$, and $p_c + r_c - p_b - r_b < \delta N/(k \cdot f)$, then there is a unique equilibrium in which both modes are used. The bus peak period starts earlier and ends later than the car peak period. A queue at the bus stops starts to develop until the moment of departure of the first bus user that faces road congestion. During the car peak period, the length of the queue at the bus stop remains constant, and it begins to dissipate after the departure of the last bus user that faces road congestion. During the car peak hour, a queue at the bottleneck on the road begins to develop at the moment of the first car departure and grows linearly for early arrivals and shrinks linearly for late arrivals. The equilibrium is depicted in Figure 2.2.*

PROOF. See Appendix A.2 □

The intuition for this Proposition is simple. Following Lemma 1, the first bus user departs before there are cars on the road: she only faces schedule delay cost. Later bus departures trade schedule delay for queuing delay at bus stops. Then, when the first car user departs, she faces no congestion on the road but only schedule delay costs. Later departures by car trade congestion delays for schedule delay. At the same time, as bus users experience the same congestion delays than car users, the modal split equilibrium requires that queuing delays at the bus stop remain constant.

Figure 2.2 summarizes the equilibrium when $0 < p_c + r_c - p_b - r_b < \delta N/(k \cdot f)$. The lower panel shows the cumulative arrivals of vehicles in PCU to the road bottleneck and cumulative arrivals to the CBD. When car users are not departing, i.e., in $[t_b^s, t_c^s]$ and $[t_c^e, t_b^e]$, there is no road congestion and the inflow and outflow of vehicles occurs at a rate λf. Road congestion begins when the first car user departs and the queue builds up linearly from t_c^s to the time of an on-time arrival, and then dissipates linearly until it disappears at t_c^e. The rates are the same as in the classic bottleneck model of Arnott et al. (1993), as those are the ones that make the sum of queuing delay and schedule delay costs constant over time.

The upper panel shows the cumulative departures of bus users from home and cumulative arrivals to the road bottleneck. A queue at the bus stops starts to develop at t_b^s with bus users departing at a rate higher than capacity ($\frac{\alpha_2}{\alpha_2-\beta}kf$). The vertical distance between the cumulative departures schedule and the cumulative arrivals schedule is queue length, and the horizontal distance is travel time. The queue grows linearly until the moment of departure of the first bus user that faces road congestion, i.e., the user who arrives at the road bottleneck at t_c^s. From that moment the sum of road queuing delay and schedule delay costs are constant over time, so that user depart at a rate equal to the capacity such that bus stop delays are constant over time and thus user time costs are constant over time. The queue begins to dissipate when the first bus user does not face road congestion and it disappears at t_b^e.

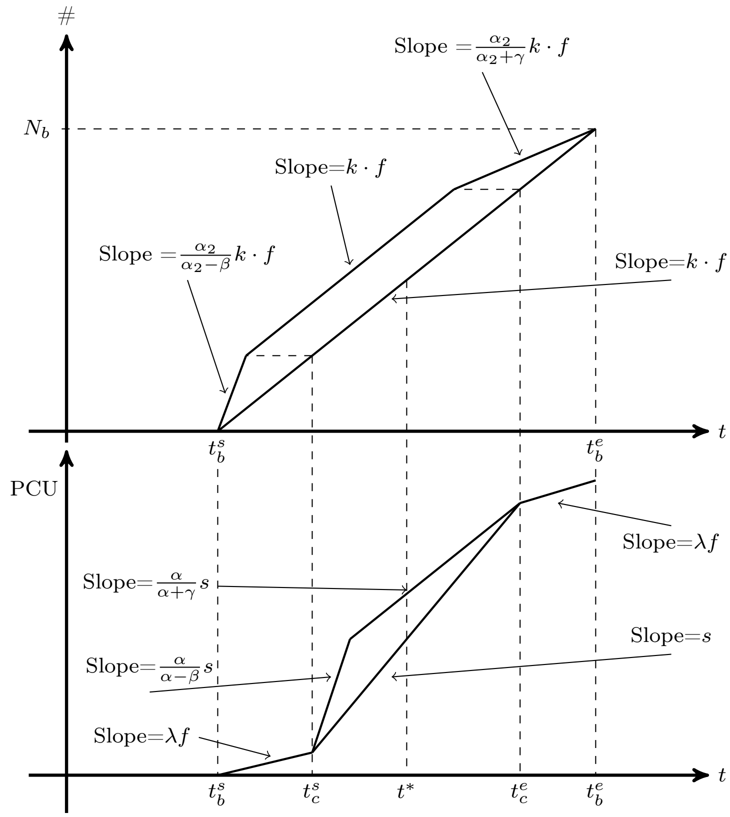


Figure 2.2: Interior equilibrium under mixed traffic when $0 < p_c + r_c - p_b - r_b < \delta N/(k \cdot f)$. The upper panel displays cumulative departures and arrivals of passengers at the bus stop. The lower panel displays cumulative arrivals of vehicles (in PCU) to the road bottleneck and to the CBD.

We now deal with the cases where $p_c + r_c - p_b - r_b \leq 0$.

Proposition 2.3 *If the time-invariant full price of the car is lower than that of the bus, i.e., $p_c + r_c < p_b + r_b$, there is a unique equilibrium in which all individuals travel by car and that mirrors the simple bottleneck model.*

PROOF. See Appendix A.3 □

The intuition is as follows: it cannot be an equilibrium that the first bus departure occurs before the first car departure because the first car user will face lower full price, lower schedule delay cost, and no congestion (since buses do not produce road congestion by themselves). But if the first bus user departs at the same time of after the first car user, then she will face a higher total cost than the car user that departed at that same time, because they will both face the same schedule delay cost, the same congestion cost, but the time-invariant full price of the bus is higher, so this cannot be an equilibrium either. It follows that only cars are used and the unique equilibrium is the same as the one in the simple bottleneck model.

Proposition 2.4 *If the time-invariant full prices are equal, i.e., $p_c + r_c = p_b + r_b$, there are multiple equilibria. These are a continuum of equilibria that range from an equilibrium in which the car is the only mode that is used to an equilibrium in which the peak period of both modes are the same. The latter is the equilibrium with the highest modal share for buses, the lowest total user travel cost, and has no queuing at bus stops.*

PROOF. See Appendix A.4 □

The full proof is provided in the appendix but, to understand why everything goes consider that $p_c+r_c = p_b+r_b+\varepsilon$. From Proposition 2.2, if $\varepsilon > 0$, the car period contains the bus period and both modes are used. On the other hand, from Proposition 2.3, if $\varepsilon < 0$, then buses are not used. Therefore, if ε approaches zero from a positive value, one obtains identical periods, with buses being used but without queuing at bus stops, just as what happens in Figure 2.2 when periods coincide. But, if ε approaches zero from a negative value, buses are not used in equilibrium. It follows that when $\varepsilon = 0$, both equilibria (from positive and negative limits) may occur, but also everything that is in between may be an equilibrium.

2.3 First best

Road congestion and boarding delays are pure deadweight loss because they can be reduced without increasing schedule delay costs. Also, at the social optimum trips must occur over a continuous time interval; otherwise, there would be wasteful schedule delay costs. Therefore, at the social optimum, there should be no road congestion or bus stop queuing, and both capacities must be fully utilized. This is a standard result in deterministic bottleneck models (see e.g. Arnott et al. (1993)).

To avoid congestion on the road bottleneck, cars must arrive at a rate equal to $s - \lambda \cdot f$ if the public transport frequency is strictly positive, while they must arrive at rate s otherwise. On the other hand, to avoid boarding delays, the arrival rate to the bus stop must be $k \cdot f$ while $f > 0$ and 0 otherwise.

If resource costs r_c and r_b where equal, the minimum user cost would be reached when schedule delay cost is minimized; this happens when the period of operation of the public transport system matches the cars' peak period, as this will ensure the best utilization of the transport capacity, namely $s + (k - \lambda) \cdot f$. Moreover, the schedule delay cost of the first and last departure has to be the same. If one considers that resource costs are not the same –most likely with $r_c > r_b$– then the minimum user cost would not be reached when schedule delay cost is minimal and, therefore, the cars' peak period would no longer be the same that the bus hours of operation, but it will shorter.

However, in our model, we also consider the operational cost of providing public transport, and this has an impact on the social optimum. We model these costs as a function of the bus fleet, to capture the capital expenses and part of the labor expenses, and as a function of the number of buses that are dispatched in the peak period, to capture operational expenses. As the kilometers driven for each dispatch is the same, the second term includes the total vehicle-kilometers. The expenditure of providing a frequency f is therefore given by:

$$E(f) = c_1 \cdot f \cdot T + c_2 \cdot f \cdot \Delta t_b \tag{2.16}$$

where the first term is the expenditure related to the bus fleet and the second to dispatches. c_1 is the constant cost per bus and $f \cdot T$ is the fleet required to provide a frequency f when the cycle time is T. This cycle time can be decomposed into two times, the free-flow cycle time, T^0, and the time spent passing the road bottleneck. c_2 is the constant cost per dispatch and $f \cdot \Delta t_b$ is the number of dispatches in the entire period where buses operate, which we denote by $t_b^e - t_b^s = \Delta t_b$.

As we argue above, at the social optimum there is no road congestion and therefore the cycle time is constant and equal to the round trip free flow time, named T^0. The expenditure then collapses to:

$$E(f) = c_1 \cdot f \cdot T^0 + c_2 \cdot f \cdot \Delta t_b \tag{2.17}$$

With this formulation in which the public transport expenditure is an increasing function of the length of the operation period and the frequency, there are two tradeoffs between user costs and operating costs. First, conditional on the period of operation, increasing frequency induces decreased user costs but increased expenditure. Second, conditional on a given frequency, decreasing the operational period of buses leads to lower public transport costs; however, this is something that leads to a decreased capacity of the public transport system so that it can only be achieved by moving some bus users to cars. This, in turn, increases user costs through increased schedule delay costs. Therefore, car and bus hours of operation which minimizes user cost will only be socially optimal when there are no operating cost advantages of decreasing the operation period of buses ($c_2 = 0$). As c_2 becomes positive, the car peak period increases, while bus operating hours decrease. In balance, it is not clear whether in the first best the car peak period is included in or includes bus operating hours. It depends on the relative strength of the effect of $r_c > r_b$ and $c_2 > 0$.

In summary, the first best is characterized by the absence of congestion on the road and bus stop queuing, by continuous hours of operations of both buses and cars, and:

- If $r_c = r_b$ and $c_2 = 0$, then buses and cars have identical peak hours hours.
- If $r_c = r_b$ and $c_2 > 0$, then bus operations hours are included in the car peak period
- If $r_c > r_b$ and $c_2 = 0$, then car peak hours are included in the bus operations hours.
- If $r_c > r_b$ and $c_2 > 0$, the ordering of peak periods is uncertain.

The actual optimal frequency and operating periods can be calculated from the social cost minimization problem, yet the actual expressions are rather uninformative, so we abstain from deriving them here.

As it is characteristic of these models, the first best can be decentralized with perfectly time-variant prices. By mirroring the queuing delay costs of the untolled equilibrium, these time-varying prices decentralize the optimal arrival rates to both bottlenecks (the road and the bus stop), and the optimal operational period of buses. As we argue in the Introduction, our interest is on second-best policies rather than on the decentralization of the first-best. However, before turning to second-best analysis, it is worth noting that in the first best whether there is a BRT system or not is irrelevant because there is no road congestion.

2.4 Optimum in mixed traffic with time-invariant prices

In Section 2.2 we characterized equilibria under mixed traffic conditions for the two modes system. We show how to optimize it and what are the features of this optimum. Start by recalling that, according to Proposition 2.2 an interior unique equilibrium appears only when $0 < p_c + r_c - p_b - r_b < \delta N/(k \cdot f)$. If, on the other hand, $p_c + r_c - p_b - r_b = 0$ there are multiple equilibria (Proposition 2.4), while if $p_c + r_c - p_b - r_b < 0$ equilibrium has no bus users (Proposition 2.3). The way to proceed to optimize the system, then, is the following: we will minimize the social cost function that is valid for an interior equilibrium (detailed below) over $p_c - p_b$ and f. If these two values fulfill the condition of Proposition 2.2, then this is the optimum, and it will feature a car peak hour that is included in the bus hours of operation, and congestion at the bottleneck and queuing at the bus stop, as in figure 2.2.

If, on the other hand, the result of this minimization (denoted by *) leads to $f^* < 0$, this means that it is optimum not to provide public transportation at all. If, however,$f^* > 0$ but $p_c^* + r_c - p_b^* - r_b < 0$, one need to compares two things: if it is better, social welfare wise, not to offer any public transport, or if it is better to impose $p_c^* + r_c - p_b^* - r_b$, together with a new optimal f.

An issue to address is that when $p_c + r_c - p_b - r_b = 0$, there are multiple interior equilibria. We solve this by focusing on the equilibrium in which the peak period of both modes are the same because it is the equilibrium with the lowest total user travel cost. Although the prices cannot decentralize this particular equilibrium, it is reasonable to focus on this one as a sufficiently small price difference (for example, $p_c + r_c - p_b - r_b = \varepsilon > 0$) would induce an equilibrium that is arbitrarily close. In other words, as ε approaches zero, the unique equilibrium for a positive price difference approaches the equilibrium in which the peak period of both modes are the same with a zero price difference and, in this case, the public transport system has no queuing at the bus stop.

We can now write the user costs and the public transport costs as a function of the time-invariant full price difference and the bus frequency, for the case when equilibrium is interior, in order to start building the social cost function. Subtracting the price paid for each user and adding the total time costs of bus and car users using Eq. (2.15), we obtain the total user cost (UC):

$$UC = \delta \frac{N^2 + \left(\frac{p_c + r_c - p_b - r_b}{\delta}\right)^2 k \cdot f \cdot (s - \lambda \cdot f)}{s - \lambda \cdot f + k \cdot f} + r_b \cdot N_b + r_c \cdot N_c \tag{2.18}$$

where N_c and N_b are equilibrium values.

Since in mixed traffic there is always road congestion, we need to account for this travel time in the transit costs. We assume that the cycle time that determines the fleet size is the maximum travel time over the period plus the round trip free flow time. Noting that the maximum travel time is simply the user travel cost divided by α and using Eq. (2.15), we can write the cycle time as:

$$T = T^0 + \frac{\delta}{\alpha} \frac{N - \frac{(p_c + r_c - p_b - r_b)\cdot k \cdot f}{\delta}}{s - \lambda \cdot f + k \cdot f} \tag{2.19}$$

Where the second term is the maximum travel time which is determined by the maximum road queue.

Replacing the cycle time of Eq. (2.19) into the public transport expenditure function in Eq. (2.17) and calculating the operating period of public transport, Δt_b, from Eqs. (2.9), (2.10) and (2.14) we obtain:

$$E = c_1 \cdot f \cdot \left(T^0 + \frac{\delta}{\alpha} \frac{N - \frac{(p_c + r_c - p_b - r_b)\cdot k \cdot f}{\delta}}{s - \lambda \cdot f + k \cdot f} \right) + c_2 \cdot f \cdot \frac{N + \frac{(p_c + r_c - p_b - r_b)\cdot (s - \lambda \cdot f)}{\delta}}{s - \lambda \cdot f + k \cdot f} \tag{2.20}$$

We define the social cost (SC) directly as the sum of the total user cost (Eq. (2.18)) and the public transport system cost (Eq. (2.20)). Due to the non-linearity, it is not possible to obtain closed-form solutions of the problem. From the first-order condition with respect to $p_c - p_b$, we can write the optimal time-invariant full price difference, conditional on the frequency, as

$$p_c^* + r_c - p_b^* - r_b = \frac{c_1 \cdot \delta \cdot f \cdot k - c_2 \cdot \alpha \cdot (s - \lambda \cdot f)}{2\alpha \cdot k \cdot (s - \lambda \cdot f)} + \frac{r_c - r_b}{2} \tag{2.21}$$

Regarding frequency, the first-order condition is not particularly informative, as it leads to a nonlinear function of f and $p_c - p_b$. What may be shown though is that $\mathrm{d}f^*/\mathrm{d}N > 0$. This is quite natural but, in terms of the optimized system, it implies that for low total demands it may be better not to provide public transportation at all, as it will show up in the numerical simulation in the next section.

The expression for the optimal time invariant full price difference –conditional on f– Eq. (2.21) may be positive or negative, depending on the resources cost of each mode, and the relative values of c_1, the capital cost of the fleet, and c_2, the operational cost. If the right-hand side of Eq. (2.21) is negative, then the optimal solution may be a corner solution, where $p_c^* + r_c - p_b^* - r_b = 0$ and frequency is positive, or a solution where transit service is not provided.

2.5 Bus rapid transit (BRT)

2.5.1 Analytical results

We model the implementation of a Bus Rapid Transit (BRT) system as dedicating a fraction ϕ of the road capacity to buses. Since the road capacity that the buses use, for a given frequency f, is $\lambda \cdot f$, it would be socially optimal to dedicate exactly the capacity they need, i.e., $\phi = \lambda \cdot f/s$. This is because dedicating less would generate road congestion for buses in the form of queuing delays, which is inefficient[10], and dedicating more capacity to buses would be wasteful. However, if capacity is not perfectly divisible, it may not be possible to set the dedicated capacity to its optimum (because, for example, it is less than a lane). We will refer to the perfect divisibility capacity case as the Optimal BRT (or BRT-DC) while, when ϕ can only be set to a value larger than λf^*, we will call it BRT-IC, for indivisible capacity.

We first analyze the effects of implementing an optimal BRT, that is, when capacity is perfectly divisible and it is set exactly to $\lambda \cdot f^*$. We carry out this analysis by steps in order to isolate effects: we first consider that the mixed-traffic optimal frequency and prices are maintained, we then let frequency change, and finally look at the full-fledged unweighted social cost minimization values. Let us start then by describing equilibrium, conditional on prices and frequency, under an optimal BRT system. All the calculations needed for this Section that are not in previous Sections or part of other proofs are in Appendix A.5.

Consider first that the time-invariant full price of the car is higher than for the bus, i.e., $p_c + r_c > p_b + r_b$. This is the case analogous to the one in Proposition 2.2. Under an optimal BRT system, the effective capacity for cars is set at $s - \lambda \cdot f$, and therefore, the timing of car departures mirrors the basic and classic bottleneck model of Arnott et al. (1993), but with decreased capacity. Because of this, we will describe equilibria but omit many routine steps. The first and last car user to depart face only schedule delay costs and a queue at the bottleneck on the road begins to develop at the moment of the first car departure and grows linearly for early arrivals and shrinks linearly for late arrivals. Under mixed traffic, bus users face two successive bottlenecks: the bus stop where boarding delays may arise and the road bottleneck. Under a BRT system, they face no road congestion delays and the only delays they could face are boarding delays at the bus stop. Therefore, as we model the bus stop as a bottleneck of capacity $k \cdot f$, the timing of the departures of bus users also mirrors a basic bottleneck model. The first and last bus user to depart face only schedule delay cost and a queue at the bus stops develops linearly for early arrivals and dissipates linearly for late arrivals.

Given a time-invariant full price of the car higher than that of bus, the bus peak period starts earlier and ends later than the car peak period. The proof is identical to the one for

[10] If the capacity dedicated to buses is less than $\lambda \cdot f$, a queue will start developing from the beginning of the period in which buses are running. Moreover, a queue will start developing and will continue to grow until buses stop running. This will imply that for late arrivals the sum of road queuing delay and schedule delay costs will grow and departures that arrive late cannot be part of an equilibrium. Therefore, if the capacity dedicated to buses is less than $\lambda \cdot f$ not only road queuing delays are imposed, but the capacity of the system drastically drops as late arrivals are not possible in equilibrium. As a result it is straightforward to show that it is inefficient to induce road congestion to buses.

Lemma 1. It follows that, the equilibrium costs for each mode, conditional on the number of passengers, are exactly the same as the ones under mixed traffic in Eqs. (2.11) and (2.12):

$$c_c = p_c + r_c + \delta \frac{N_c}{s - \lambda \cdot f} \tag{2.22}$$

$$c_b = p_b + r_b + \delta \frac{N_b}{k \cdot f} \tag{2.23}$$

As a result, conditional on prices and frequency, the equilibrium modal split will also be the same under a BRT system than under mixed traffic. The only difference is that the reduction in road congestion delays for bus users due to the implementation of the BRT system is transformed into bus boarding delays. In the timing equilibrium, the full price must be constant over the entire peak period, and the only way for this to hold is that boarding delays exactly compensate schedule delay savings. Figure 2.3 summarizes the equilibrium when the time-invariant full price of the car is higher than that of the bus. Comparing this with Figure 2.2, it is evident how the mixed-traffic and BRT equilibria are different when $pc + r_c > p_b + r_b$. In particular, note that the lower graph shows congestion only for cars and not for all vehicles, as opposed to the case in Figure 2.2. This is because with a BRT, only cars suffer from road congestion.

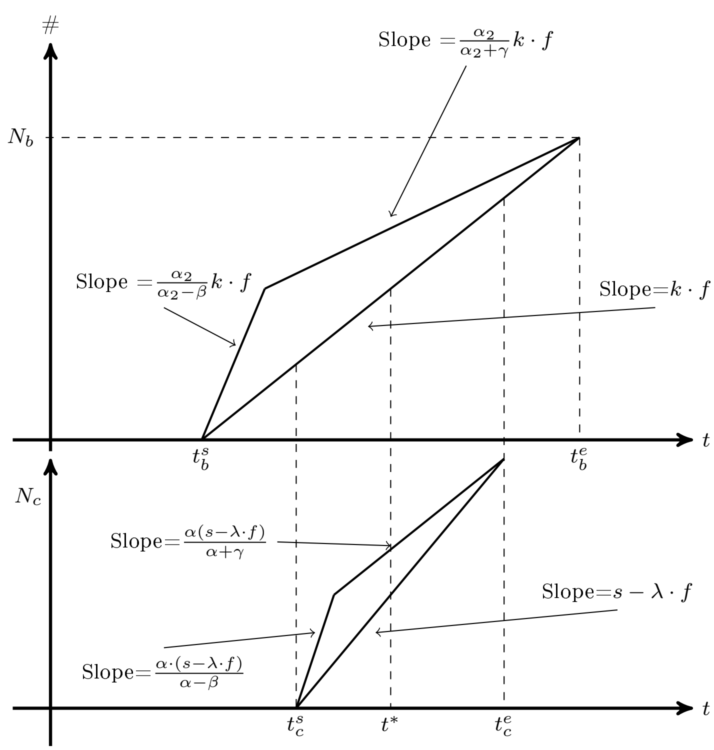


Figure 2.3: Equilibrium under BRT when $0 < p_c + r_c - p_b - r_b < \delta N/(k \cdot f)$

Under mixed traffic and equal time-invariant full prices for cars and buses, there are multiple equilibria (see Proposition 2.4). This happens because, since road congestion affects buses, congestion and schedule delay costs are the same for all modes at all times, making a unilateral switch from one mode to the other fruitless. With a BRT system implemented, this does not longer hold. Buses do not experience road congestion and, therefore, the first and last user to depart by bus only faces schedule delays. As the first car user to depart also faces only schedule delays, there is a unique equilibrium in which the peak period of both modes are the same and which is as in Figure 2.3, but with the operation periods coinciding. From that moment, the timing of departures for each mode mirrors that of two parallel basic bottleneck models, as explained above. The implications of this are strong: in mixed-traffic conditions, when time-invariant full prices are equal, there are no boarding delays, but bus users face congestion delays. With a BRT, bus users face no congestion delays but, in equilibrium, these costs are converted into queuing delays.

The third case, when the time-invariant full price of the car is lower than the bus, is also very different when a BRT system is implemented. Under mixed traffic, the unique equilibrium has all the individuals traveling by car, because buses suffer from the same road congestion as cars and it is not possible for a bus user to experience a lower generalized cost than a car user that departs from home exactly at the same time. Under a BRT system, this does not hold because buses do not suffer from road congestion. The unique equilibrium is similar to the one depicted in Figure 2.3, but with the difference that the car peak period starts earlier and ends later than the bus peak period: As the time-invariant full price of the car is lower than that the bus, in equilibrium, the schedule delay cost of the first user should be higher for cars than for buses.

The following proposition summarizes the equilibria when a BRT system is implemented.

Proposition 2.5 *With time-invariant full prices are not too different, i.e.,* $-\delta N/(s-\lambda \cdot f) < p_c + r_c - p_b - r_b < \delta N/(k \cdot f)$*, there is a unique equilibrium in which both modes are used. Queuing delays at bus stops occur for the whole duration of the operation of the BRT system, while road congestion delays for cars occur for the whole duration of their peak period. Moreover:*

- *i if* $0 < p_c + r_c - p_b - r_b < \delta N/(k \cdot f)$*, the bus peak period starts earlier and ends later than the car peak period.*
- *ii if* $p_c + r_c - p_b - r_b = 0$*, the peak period of both modes are the same.*
- *iii if* $-\delta N/(s-\lambda \cdot f) < p_c + r_c - p_b - r_b < 0$*, the bus peak period starts later and ends earlier than the car peak period.*

PROOF. See Appendix A.5.1 □

The description of equilibrium –conditional on prices and frequency– shows that there is no gain **for users** from implementing a BRT system. If prices and frequencies are not changed from the ones from mixed traffic conditions, car users do not experience any change, and public transport users, in equilibrium, exactly compensate reduced road congestion for increased boarding delays at the bus stop. However, as buses do not face road congestion, the bus speed is higher which induces cost savings because the required fleet to provide the same frequency is lower. Thus, the gains from implementing a BRT system if frequency and prices are not adjusted and capacity is perfectly divisible come from operating cost savings:

conditional on prices and frequencies a BRT system is efficient as it induces a decrease in unweighted social cost.

We now turn to discuss how frequency and prices should be adjusted together with the implementation of the BRT system. First, it is straightforward to understand that the implementation of a BRT together with a frequency increase is a Pareto improvement in the sense that both user and operators costs are reduced. To see this, consider the case described above in which the implementation of a BRT without frequency or price adjustments reduces non-marginally the operators' costs as it fully eliminates road congestion for buses. Now, if a share of these cost savings is spent on increasing the frequency, the capacity of the public transport system will be increased, the peak period will be shorter, and the user costs will decrease. This shows that implementing a BRT may induce a Pareto Improvement: both users' time cost and public transport costs decrease. Importantly, this better situation –which provides strong support for the observed BRT surge– will have, compared to the optimum when buses run in mixed traffic: (i) shorter hours of bus operation and car-peak period (ii) larger frequency, and (iii) more boarding delays, i.e. longer queues at bus stops. Note that point (ii) implies that, while for some level of demands it may be optimal not to provide any public transport service under mixed traffic, with a BRT it may well be worthwhile. This in fact shows up clearly on the numerical analyses below.

The last stage of the analysis of the optimal BRT is to consider a planner that minimizes social cost. As we showed above, the user costs are the same as under mixed traffic conditional on the frequency and time-invariant full price difference, and given by Eq. (2.18), which now also holds for negative price differences:

$$UC = \delta \frac{N^2 + \left(\frac{p_c + r_c - p_b - r_b}{\delta}\right)^2 k \cdot f \cdot (s - \lambda \cdot f)}{s - \lambda \cdot f + k \cdot f} + r_b \cdot N_b + r_c \cdot N_c \tag{2.24}$$

where N_c and N_b are equilibrium values.

The difference with mixed traffic, though is that, under a BRT system, buses do not face road congestion, thus the public transport expenditure function is the same as in the first-best, which substituting the bus peak period $\Delta t_b = t_b^e - t_b^s$, from Eqs (2.9), (2.10), and (2.14), is:

$$E = c_1 \cdot f \cdot T^0 + c_2 \cdot f \cdot \frac{N + \frac{(p_c + r_c - p_b - r_b) \cdot (s - \lambda \cdot f)}{\delta}}{s - \lambda \cdot f + k \cdot f} \tag{2.25}$$

As a result, when choosing frequency and prices, the planner faces the trade-offs discussed in Section 2.3. From the first-order condition of the sum of user costs and public transport costs (Eqs (2.24) and (2.25)) with respect to $p_c - p_b$, we obtain:

$$p_c^* + r_c - p_b^* - r_b = -\frac{c_2}{2k} + \frac{r_c - r_b}{2} \tag{2.26}$$

It is immediate to note that the optimal time-invariant full price difference is constant in this case, and does not depend on the frequency (which has yet to be calculated) or demand N. This is in stark analytical contrast with the mixed traffic conditions case and comes from the fact that buses do not suffer from congestion on the road under BRT. It also clear that the sign of the left hand side depends on the relative values of c_2/k and $r_c - r_b$ and, therefore, those values determine whether the optimal hours of operation of the BRT include, or are included, in the car peak period, according to Proposition 2.5. This is, actually, perfectly in line with what happens in the first best, something that can be seen by looking at the summary of its properties in Section 2.3. To be clear, the optimal BRT will not recover first best hours of operations, but the ordering of the initial and final operation time of each mode will be consistent wit the first best.

The expression for the optimal frequency is rather complex, but can be shown to be a linear function of demand, N; see Appendix A.5.2. In Appendix A.5.3, we prove that $N_b/(k \cdot f)$, which determines the length of the bus operations hours does not depend on N. It thus follows directly that the car peak period does not depend on N either. Then, if a positive frequency is optimally provided, operations periods will not change as demand N increases: frequencies will be adjusted to accommodate the new demand (generating bus stop queuing), but without affecting the rush hours. Of course, if frequencies are so high that they cannot, technologically grow more, then operations periods will increase. The intuition for this result is simple. As N (demand) grows, under a BRT system the transport capacity can be adjusted through changes in frequencies. An additional unit of frequency will increase public transport capacity by k, while reducing private transport capacity by $\lambda < k$. If N increases, then, it induces f to increase but without changing prices (see Eq. (2.26)). And since prices did not change, the optimal timing of the first departure, by either car or bus, has no reason to be modified because the change in f, for those initial times, does not affect their queuing or road congestion costs. The same applies for final times.

We now turn to the the case when a BRT is implemented but capacity is not perfectly divisible. What may happen is that it would be optimal to take $\lambda \cdot f^*$ from the capacity s of the bottleneck, but that actually corresponds to less than a lane and buses cannot actually circulate. What reality may dictate is that the capacity dedicated for the BRT must be a constant ϕ, which corresponds, for instance, to one out of three or four lanes. It is very evident that if $\lambda \cdot f^*$ is very close to the exogenous ϕ, where f^* is the optimal capacity of the divisible capacity BRT, then implementing a BRT will be efficient and may be a Pareto improvement, as discussed above. But if that is not the case, it may happen that there is too much unused capacity with a BRT, making it less efficient than letting mixed traffic conditions prevail. It is, then something that depends on the specifics of each case and, therefore, we relegate more insights to the numerical analysis that follows.

Still, we show in Appendix A.5.4 that, for the case when the BRT takes an exogenous capacity given by ϕ', the optimal time-invariant full price is the same as in the case of divisible capacity and the periods of operation are all constant with respect to N (yet different than in the case of divisible capacity).

2.5.2 Numerical analyses

In this section we illustrate our analytical results using values for the parameters that may represent an actual situation. The values for the numerical simulations are presented in Table

2.1[11]. We consider that there is no (time-invariant) congestion charge for cars, i.e. $p_c = 0$. This way, we report below, directly, the bus fare (per trip).

Table 2.1: Parameters for numerical examples

Parameter	Units	Value
c_1	[US$/bus]	290
c_2	[US$/hour]	130
T_0	[hour]	0.33
N	[Conmuters]	6000 to 13000
s	[PCU/hour]	6000
ΔR	[US$/Pax]	2.0
α	[US$/hour]	2.6
β	[US$/hour]	1.95
γ	[US$/hour]	3.9
k	[Pax/bus]	80
α_2	[US$/hour]	5.2
t^*	[hh:mm]	08:00
λ	[PCU/bus]	3.5

We solve for optimality under mixed traffic conditions, BRT with perfect divisibility of capacity and a BRT with indivisible capacity. Results are reported in Table 2.2, Figure 2.4 and Figure 2.5 panels (a) to (f).

We highlight some of the most important results next:

- As analytically predicted, when BRTs provide positive frequencies, they both have fares and periods of operations that are constant with respect to N.
- It is very clear that an optimal BRT is efficient in terms of the total cost, and then even with imperfectly divisible capacity, a BRT is still a better choice for many demand levels (starting at demands somewhere between 8,000 and 8,500 commuters).
- Without BRT, there is a large interval of demand where it is simply optimal not to provide public transport (up to 10,000 commuters). Yet, if part of the road capacity was dedicated exclusively for buses, then a very frequent bus service would optimally emerge. This means that it would not be the case that, as demand increases and congestion builds, planners need to incorporate public transport little by little, increasing frequency up to a point where dedicated bus lanes "becomes" a need. On the contrary, as car congestion builds because of increased demand, the planner should, from the beginning, implement BRT systems.
- Under mixed traffic conditions, when providing public transport is optimal, it is much more infrequent and more expensive fare wise for consumers, than what a BRT provides. Yet, despite being a large improvement, the equilibrium with BRT will feature much more bus stop queuing.

[11]Public transport costs c_1 and c_2, the difference in resource costs ΔR and the equivalence factor between cars and buses, λ, come from Basso and Silva (2014). The values for α, β, γ and α_2 are consistent with Huang et al. (2007) and Gonzales and Daganzo (2012)

Table 2.2: Numerical examples

	N	f	P_b	N_c	N_b	t_b^s	t_c^s	t_c^e	t_b^e	TC	OC	UC	CC	SDC	QC
Mixed traffic	6000	0	-	6000	0	-	07:20	08:20	-	19800	0	19800	3900	3900	0
	6500	0	-	6500	0	-	07:16	08:21	-	22154	0	22154	4577	4577	0
	7000	0	-	7000	0	-	07:13	08:23	-	24617	0	24617	5308	5308	0
	7500	0	-	7500	0	-	07:10	08:25	-	27188	0	27188	6094	6094	0
	8000	0	-	8000	0	-	07:06	08:26	-	29867	0	29867	6933	6933	0
	8500	0	-	8500	0	-	07:03	08:28	-	32654	0	32654	7827	7827	0
	9000	0	-	9000	0	-	07:00	08:30	-	35550	0	35550	8775	8775	0
	9500	0	-	9500	0	-	06:56	08:31	-	38554	0	38554	9777	9777	0
	10000	0	-	10000	0	-	06:53	08:33	-	41667	0	41667	10833	10833	0
	10500	5	1.75	9797	703	06:46	06:54	08:32	08:36	44867	2748	42120	11099	11262	164
	11000	10	2.00	9677	1323	06:55	06:55	08:32	08:32	48089	5534	42555	11601	11601	0
	11500	18	2.00	9296	2204	06:57	06:57	08:31	08:31	51272	9279	41994	11701	11701	0
	12000	25	2.00	8976	3024	06:59	06:59	08:30	08:30	54436	12803	41632	11840	11840	0
	12500	32	2.00	8702	3798	07:00	07:00	08:29	08:29	57583	16161	41422	12009	12009	0
	13000	39	2.00	8464	4536	07:02	07:02	08:28	08:28	60716	19387	41329	12200	12200	0
BRT - DC	6000	13	1.81	4975	1025	07:20	07:26	08:16	08:19	19589	2929	16660	2298	2951	1460
	6500	20	1.81	4956	1544	07:20	07:26	08:16	08:19	21675	4413	17262	2127	3110	2114
	7000	26	1.81	4937	2063	07:20	07:26	08:16	08:19	23762	5898	17864	1977	3291	2723
	7500	33	1.81	4917	2583	07:20	07:26	08:16	08:19	25848	7382	18465	1844	3489	3297
	8000	40	1.81	4898	3102	07:20	07:26	08:16	08:19	27934	8867	19067	1727	3703	3842
	8500	46	1.81	4878	3622	07:20	07:26	08:16	08:19	30020	10351	19669	1622	3929	4361
	9000	53	1.81	4859	4141	07:20	07:26	08:16	08:19	32107	11836	20271	1528	4166	4859
	9500	59	1.81	4840	4660	07:20	07:26	08:16	08:19	34193	13320	20872	1443	4411	5339
	10000	66	1.81	4820	5180	07:20	07:26	08:16	08:19	36279	14805	21474	1366	4665	5803
	10500	73	1.81	4801	5699	07:20	07:26	08:16	08:19	38365	16289	22076	1296	4925	6253
	11000	79	1.81	4781	6219	07:20	07:26	08:16	08:19	40452	17774	22678	1231	5192	6692
	11500	86	1.81	4762	6738	07:20	07:26	08:16	08:19	42538	19258	23280	1172	5463	7120
	12000	93	1.81	4743	7257	07:20	07:26	08:16	08:19	44624	20743	23881	1118	5740	7539
	12500	99	1.81	4723	7777	07:20	07:26	08:16	08:19	46710	22227	24483	1067	6020	7949
	13000	106	1.81	4704	8296	07:20	07:26	08:16	08:19	48797	23712	25085	1020	6304	8353
BRT - IC	6000	35	1.81	3279	2721	07:21	07:27	08:16	08:19	21295	7831	13464	1748	3453	1705
	6500	42	1.81	3279	3221	07:21	07:27	08:16	08:19	23360	9270	14091	1748	3766	2018
	7000	48	1.81	3279	3721	07:21	07:27	08:16	08:19	25426	10709	14717	1748	4079	2332
	7500	55	1.81	3279	4221	07:21	07:27	08:16	08:19	27492	12148	15344	1748	4393	2645
	8000	61	1.81	3279	4721	07:21	07:27	08:16	08:19	29558	13587	15971	1748	4706	2958
	8500	68	1.81	3279	5221	07:21	07:27	08:16	08:19	31624	15026	16597	1748	5019	3272
	9000	74	1.81	3279	5721	07:21	07:27	08:16	08:19	33690	16466	17224	1748	5333	3585
	9500	81	1.81	3279	6221	07:21	07:27	08:16	08:19	35755	17905	17851	1748	5646	3898
	10000	87	1.81	3279	6721	07:21	07:27	08:16	08:19	37821	19344	18477	1748	5959	4212
	10500	94	1.81	3279	7221	07:21	07:27	08:16	08:19	39887	20783	19104	1748	6273	4525
	11000	100	1.81	3279	7721	07:21	07:27	08:16	08:19	41953	22222	19731	1748	6586	4838
	11500	107	1.81	3279	8221	07:21	07:27	08:16	08:19	44019	23661	20357	1748	6899	5152
	12000	113	1.81	3279	8721	07:21	07:27	08:16	08:19	46085	25101	20984	1748	7213	5465
	12500	120	1.81	3279	9221	07:21	07:27	08:16	08:19	48150	26540	21611	1748	7526	5778
	13000	126	1.81	3279	9721	07:21	07:27	08:16	08:19	50216	27979	22237	1748	7839	6092

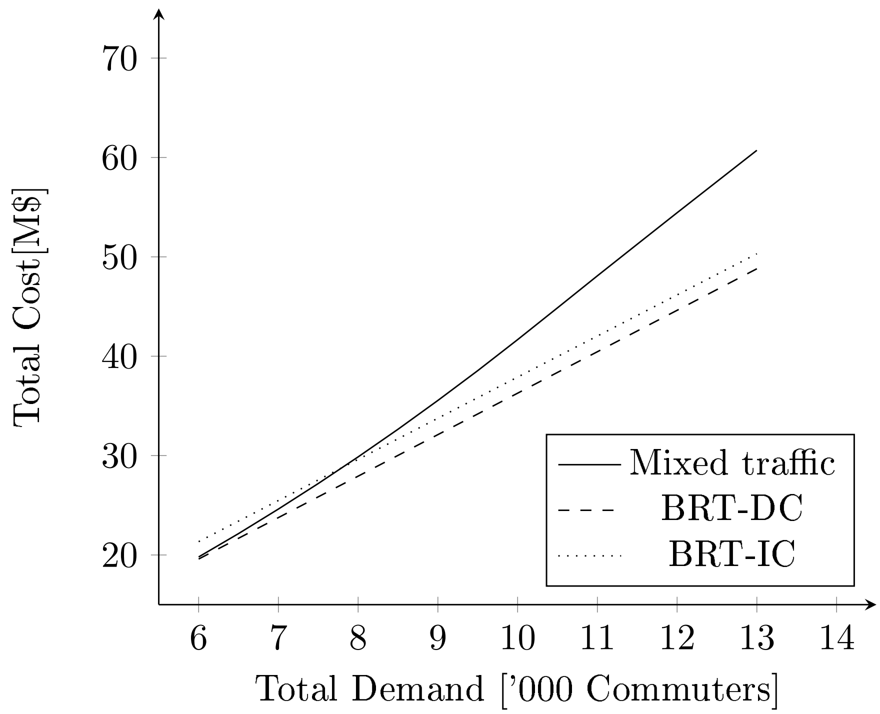


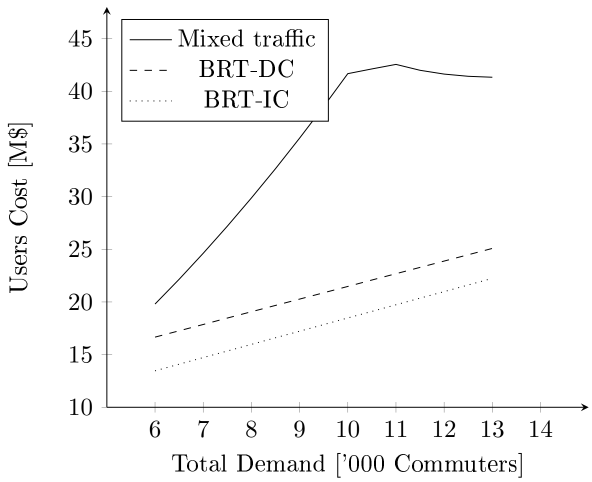
Figure 2.4: Total cost

- The bus and car hours of operation with BRT are significantly shorter than under mixed traffic conditions. Thus, a BRT, despite taking capacity from cars, will decrease the car peak period, not increase it.

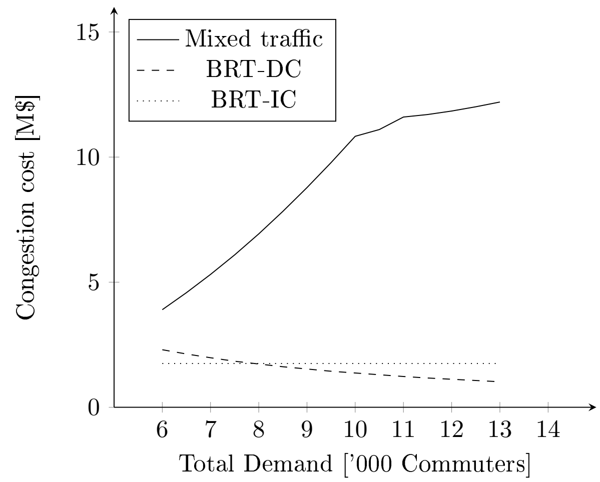
2.6 Conclusions

Bus rapid transit systems are a public transport development that has seen an exponential grow around the world. Its philosophy is quite simple: by taking capacity from cars, and dedicating it exclusively for buses, the transport capacity increases, leading to a modal change towards a system that is fast, and that costs only a fraction of what rail or subway would. A global overview of the number of projects being built or that have been considered is astonishing. But at times it seems that BRTs have been victims of its success; in several cases, severe queuing at BRT stations has been reported.

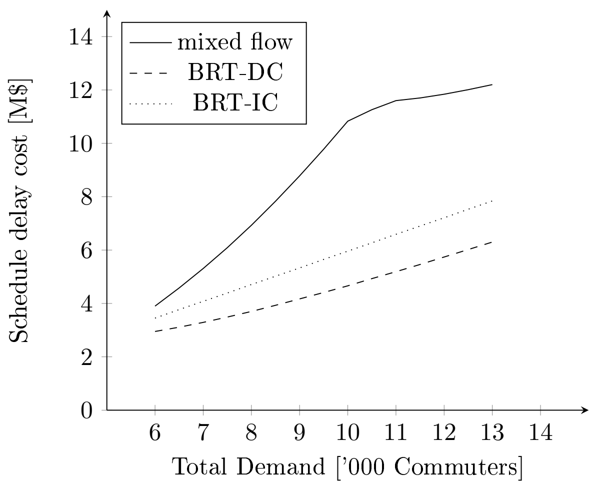
In this paper, we provide what we believe is the first microeconomic analysis of BRTs in the context of dynamic congestion, where queuing and congestion delays are endogenous as a result of individual schedule of departures. The main difference with previous literature that looks into bimodal (car and public transport) systems is that, rather than focusing on crowding and assuming from the outset that capacities of each mode are independent, we focus on modeling boarding delays in equilibrium, and comparing mixed traffic conditions with what would arise from dedicating part of the road capacity to a BRT. This is, we believe, our main methodological contribution. In terms of conclusions of the analyses, our primary, most policy-relevant result is that in a second best-world where fares cannot vary perfectly with time, BRTs are efficient and have the potential to provide a Pareto improvement of the transport system: in equilibrium both bus users and car users can be better off, while



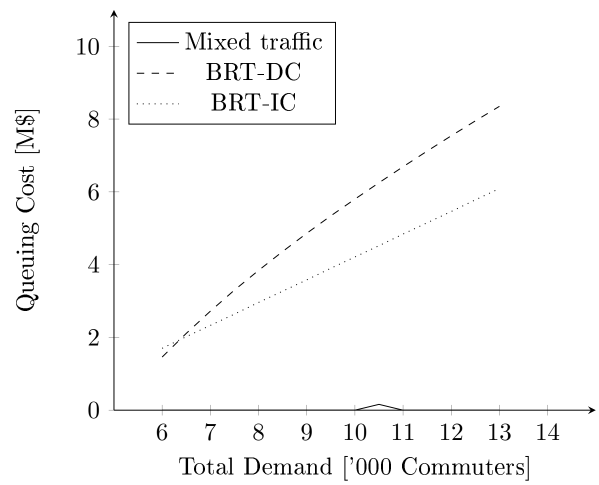
(a) User cost



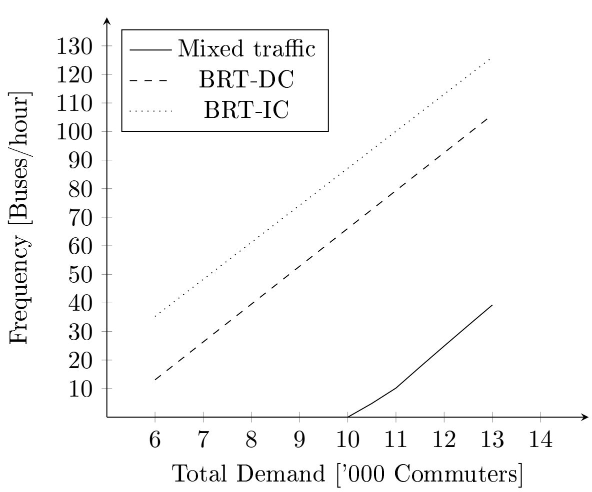
(b) Congestion cost



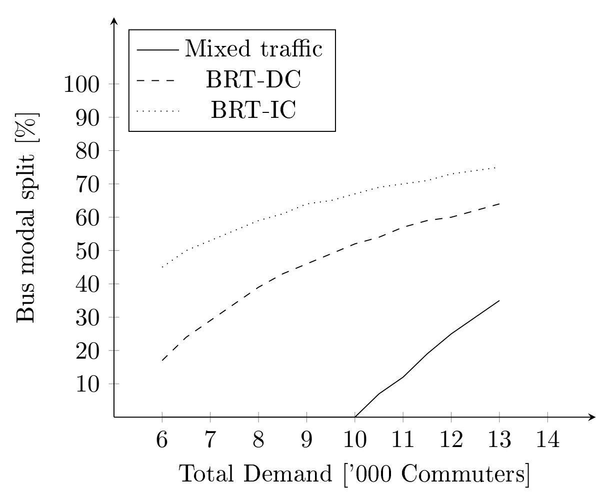
(c) Schedule delay cost



(d) Queuing cost



(e) Frequency



(f) Bus modal split

Figure 2.5: Sumary report

the costs of providing public transport decrease. With a BRT, fares will be lower because the peak hours of operation of the system are shorter. The car peak-period is also shorter, despite the fact that capacity was taken from private transport. Importantly, this better-for-all situation features more boarding delays, that is, queues at bus stops will be longer than under mixed-traffic conditions. All these results provides strong support for the BRT surge observed around the globe while providing one –possibly not the only one– explanation for observed longer queues at stations.

Concerning future research, we made some simplifying assumptions that may be lifted to assess their importance in our results. To name a few, inelastic demand, homogenous users, full access to cars and no crowding costs. We believe that pushing a research agenda along these lines is both challenging from a methodological point of view and relevant given the continuing importance of BRT systems in real transport policies.

Chapter 3

Rush hour frequency management: A dynamic congestion approach

Abstract:

This paper analyzes the efficient provision of a public transport system operated by buses that share the road capacity with cars. We propose a dynamic congestion model with mode choice, where public transport and cars are substitutes modes. The congested period, the departure pattern, and the queuing at the bus stop are endogenous to the model. We define different frequencies for the congested and the uncongested period to explicitly capture the effects of congestion on the optimal frequency pattern. We show under plausible conditions that the efficient frequency during the uncongested period is higher than the frequency during the congested period. If we compare our results, using numerical analysis, versus cases where the frequency is constant, we can ascertain (i) the optimal provision of public transport is efficient for lower demand, (ii) user costs are lower, (iii) optimal frequencies are higher, (iv) the operational expenditure is higher, and (v) boarding delays at the bus stop increases. Points (ii) and (iii) imply that the efficient frequency pattern reduces users costs by increasing the frequency and operational expenditure.

Keywords: Public transport; Optimal frequencies; Dynamic congestion; Bottleneck model

3.1 Introduction

It is no surprise that congestion is a major source of inefficiencies. According to Citylab[1], in the U.S, congestion cost U$S305 billion during 2017, an increase of U$S10 billion compared to the previous year. Of course, congestion affects commuters directly but also the urban system. Public transportation is an essential part of the urban system; it moves a large number of people using less road capacity per person than by car. When public transport buses share road capacity with cars, they are affected by congestion on at least the same level as cars. The typical situation is that the public transport vehicles share the road capacity with cars. Dedicated infrastructure such as bus lanes or Bus Rapid Transit (BRT) systems are not the general rule. In its study 'The identification and management of bus priority schemes' [2] published in a April 2017, Imperial College London summarized the percentage of bus priority in 14 cities around the world, which showed the maximum percentage of bus priority was attained in Brussels, with less than 25% of the road network, followed by Barcelona, Dublin and Seattle with less than 20% of bus priority kilometers, with the remaining ten cities under 10%[3] of bus priority kilometers. In terms of Latin American, in Santiago[4], the percentage of bus priority was around a 10% of the total road network by 2018.

An efficient social design of a public transport system depends on the demand structure and especially on its time-of-day variation. This optimal design has to take into consideration the fact that commuters can decide between using public transport and private cars, a modal choice that has an impact on the congestion level. All of this poses a severe challenge. An optimal design needs to define a frequency pattern. This pattern not only affects the efficiency of the public transport but also alters the cost for car users, modifying the system equilibrium. In this paper, we deal with this problem using a dynamic congestion pattern that allows us to find the optimal frequency pattern taking into consideration modal elasticity, temporal elasticity, and congestion. Optimization of the public transport system has to consider that public transport vehicles run during periods with and without congestion, considering that model bus frequency depends on the congestion level to induce social efficiencies.

Focusing on public transport optimization literature, we can highlight the papers by Jansson (1980, 1984), in wich he formulates a two-period problem without congestion, where the off-peak operation has no capital cost because it considers the incremental cost of an already acquired bus. It does not solve the general case problem, with optimal frequency, and argues that a single frequency is optimal in most cases. Following Jansson (1980, 1984), Jara-Díaz et al. (2017) analyzed the optimal fleet, frequency, and vehicle capacity for two periods with

[1] See https://www.bloomberg.com/news/articles/2018-02-07/new-study-of-global-traffic-reveals-that-traffic-is-bad (Access on Aug 2020)

[2] See The identification and management of bus priority schemes, page 10 https://www.imperial.ac.uk/media/imperial-college/research-centres-and-groups/centre-for-transport-studies/rtsc/The-Identification-and-Management-of-Bus-Priority-Schemes—RTSC-April-2017_ISBN-978-1-5262-0693-0.pdf (Access on Aug 2020)

[3] The cities considered in the study are: Barcelona (Spain), Brussels (Belgium), Dublin (Ireland), Istanbul (Turkey), Kuala Lumpur (Malaysia), Lisbon (Portugal), London (UK), Montréal (Canada),New York (USA), Paris (France), Seattle (USA), Singapore, Sydney (Australia), and Vancouver (Canada)

[4] See *Informe de Gestion 2018*, page 22. http://www.dtpm.cl/index.php/documentos/informes-de-gestion (Access on Aug 2020)

known demand and shows that taking the lower demand period into consideration modifies the social optimum. They consider an inelastic public transport demand, congestion is modeled only by differences in cycle time between periods, and there is no demand elasticity between periods.

If we place the focus on dynamic congestion models, the literature is extensive, following the seminal papers by Vickrey (1969) and Arnott et al. (1990, 1993). Almost all of this literature focuses solely on the case of private transport. Fewer papers consider two modes .The first introduction of a two modes problem in a dynamic congestion framework was Tabuchi (1993), who models a heavy rail as a congestion-free alternative. Huang (2000) analyzed a similar setting but adding crowding costs, so that the trade-off was not only between schedule and queuing delay and the transit fare, but also with crowding discomfort. Kraus and Yoshida (2002) optimizes the number of trains and the capacity of an individual train, thus affecting waiting times, and optimizes the timetable creating clusters of dispatches. de Palma et al. (2017) add the analysis of optimal dynamic pricing of individual trains and optimize frequency considering a constant headway. van den Berg and Verhoef (2014) examines the effects of user heterogeneity on the car bottleneck-crowded train problem, while Wang et al. (2017) consider bottleneck capacity expansions and train subsidies.

Public transport provided by buses in a two-mode system has been analyzed in Huang et al. (2007), Gonzales and Daganzo (2012), and Basso et al. (2019). In all three papers, buses use a part of the road capacity. Huang et al. (2007) only models mixed traffic, Gonzales and Daganzo (2012) only consider the case when buses operate on a bus lane, and Basso et al. (2019) analyze and compare both scenarios. In this paper, we use the approach of Basso et al. (2019)[5], where they model a continuous frequency and prove that it is an excellent approximation for the intermittent case. We optimize two constant frequencies considering two periods[6], one when there is road congestion and a different one where there is no road congestion. in both cases we use a continuous approximation.

The main result of our paper is that, under plausible assumtions, the efficient frequency during the uncongested period is higher than the efficient frequency during the congested period. We show numerically that two-frequency optimization is efficient for users, reducing user cost, mainly from congestion , while increasing operator expenditure. If we compare our results with the optimization with one constant frequency[7], the two-frequency optimization is efficient to provide public transport for lower demand, (ii) user costs are lower, (iii) operational costs are higher and (iv) boarding delays at the bus stop increase. Points (ii) and (iii) imply that two-frequency optimization reduces user costs through increased operational expenditure. Point (iv) means that boarding delays are not a sign of poor operation. We highlight that our results achieve up to 14% of cost reduction. Finally, we compare the two-frequency optimization against an efficient BRT system to understand how efficient it as an infrastructure measure as opposed to with a management measure.

The structure of the paper is as follows. In Section 3.2 we describe the model and char-

[5]A version of Chapter 2 has published as: Basso, Leonardo J. Feres, Fernando. Silva, Hugo E. (2019) The efficiency of bus rapid transit (BRT) systems: A dynamic congestion approach. *Transportation Research Part B: Methodological,* 127:47 – 71.

[6]We call it: Two-frequency optimization

[7]Results from Basso et al. (2019)

acterize the equilibrium for two different known frequencies and a time-invariant fare. In Section 3.3, we then study the optimization of the frequency pattern. Section 3.4 has numerical examples to complement our analytical results. Finally, Section 3.5 summarizes the results and concludes.

3.2 Two-frequency equilibrium

We consider two areas connected by a highway: a residential zone (H) and the Central Business District (CBD). Every morning, N identical users travel from H to CBD. All individuals have the same work start time t^* and choose whether to travel by car or bus. Following the approach of Vickrey (1969) and Arnott et al. (1990, 1993), we consider that travel by car is uncongested except for a bottleneck of capacity s, in which car users face congestion effects (queue) if and only if the combined arrival rate (comprehending private vehicles and public transport service) exceeds the given capacity s.

Access to public transport requires walking to the nearest bus stop. At the bus stop, the waiting time is divided into two parts: constant waiting time and a variable waiting time according to the presence of a queue at the bus stop. A queue develops if the arrival rate to the bus stop is higher than the system capacity.

As we show bellow, the interesting case is where buses run for a more extended period than cars. If there is no road congestion, bus frequency is f_u, and the period is called *uncongested* period; when there is road congestion the period is called *congested* period and the frequency is f_j. We consider a fixed bus capacity equal to k passengers per bus. In each period, buses operate continuously with a constant headway given by $1/f_u$ and $1/f_j$, for the uncongested and congested respectively. In order to ensure that the buses flow alone is not large enough to generate road congestion is necessary that $\lambda \cdot f_i/s < 1$ with $i = u$ or j, where λ is an equivalence factor between buses and cars.

Basso et al. (2019) model the bus service using a continuous approximation of a service with an intermittent nature. Figure 3.1 illustrates our setting, considering $i = u$ or j.

Following Small (1983), as it is customary in the literature, a user who arrives at time $t < t^*$ incurs a time early cost of $\beta \cdot (t^* - t)$; on the other hand, if she arrives at time $t > t^*$ incurs a time late cost of $\gamma \cdot (t - t^*)$. β and γ are schedule delay cost parameters. The travel time valuation is α for the value of (in-vehicle) travel time . Our parameter setting is usually referred in the literature as $\alpha - \beta - \gamma$ preferences. The individual's value of waiting time is denoted by α_2.

We denote with N_c the number of car users and N_b the number of bus users. The generalized cost of a user that departs from home at time t and arrives at the CBD at time t_a is $c_c(t)$ if going by car, and $c_b(t)$ if using the bus. $c_c(t)$ and $c_b(t)$ are defined by:

$$c_c(t) = p_c + r_c + \alpha \cdot T_w(t) + \begin{cases} \beta \cdot (t^* - t_a) & \text{if } \; t_a \leq t^* \\ \gamma \cdot (t_a - t^*) & \text{if } \; t_a > t^* \end{cases} \tag{3.1}$$

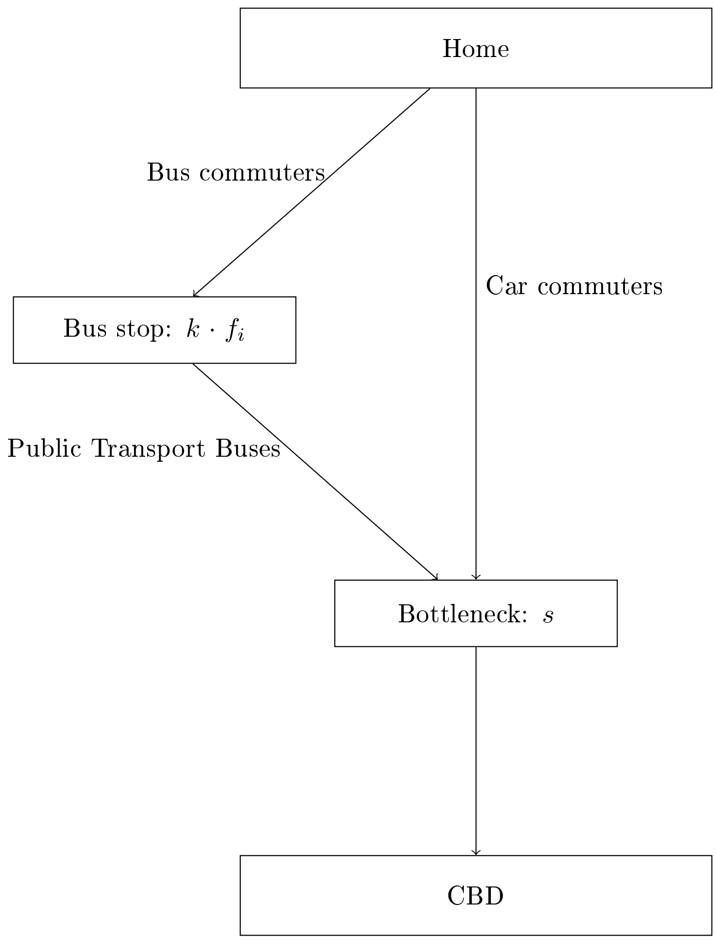


Figure 3.1: Two-bottleneck diagram of the mixed traffic transport system.

$$c_b(t) = p_b + r_b + \alpha_2 \cdot T_q(t) + \alpha \cdot T_w\left(t + T_q(t)\right) + \begin{cases} \beta \cdot (t^* - t_a) & \text{if } t_a \leq t^* \\ \gamma \cdot (t_a - t^*) & \text{if } t_a > t^* \end{cases} \tag{3.2}$$

For car users, p_c is the car congestion price that the planner can set, and r_c represents the resource (constant) costs of a trip, which include fuel and parking costs, vehicle depreciation, and constant travel times, among others. Bus users face a fare p_b and a constant resource cost of a trip r_b, which includes access time costs, discomfort, free-flow travel time, and constant waiting time, among others. $T_w(t)$ is the travel time through the bottleneck if the user arrives at the bottleneck at time t. $T_q(t)$ represents the queuing time for the user who arrives at the queue at time t. The last term on the right-hand side of Eqs. (3.1) and (3.2) represent the schedule delay cost, where t_a is the arrival time to the CBD.

Any difference that may make these two modes vertically differentiated, such as comfort is captured in the parameters r_c and r_b. Therefore, a car user who departs at time t arrives at $t + T_w(t)$ to the CBD, i.e., $t_a = t + T_w(t)$. For a bus user who departs at time t from home arrives at the CBD at $t_a = t + T_q(t) + T_w(t + T_q(t))$. In (3.2), $T_q(t)$ is the waiting time due to queuing at the bus stop, $\alpha_2 \cdot T_q(t)$ is the waiting cost.

We consider that prices are time-invariant, i.e. p_c and p_b do not depend on t; buses and cars share the road capacity without any road facility exclusively dedicated to public transport.

According to Basso et al. (2019), if the time-invariant full price of the car is higher than that of the bus, i.e., $p_c + r_c > p_b + r_b$, and both modes are used, the bus peak period starts earlier and ends later than the car peak hour. This is $t_b^s < t_c^s < t_c^{\mathrm{e}} < t_b^{\mathrm{e}}$. That result does not depend on the frequency level; therefore, it is valid if we consider a frequency as is defined in Eq. (3.3).

The times t_c^s and t_c^{e} are the times of the first and last departures of an individual by car; t_b^s and t_b^{e} are the times of the first and last departures of an individual by bus. In this context we call *congested* period the interval between t_c^s and t_c^{e}, and *uncongested* period to the following periods: t_b^s to t_c^s ; t_c^{e} to t_b^{e}[8]. We define the following frequency function:

$$f(t) = \begin{cases} f_u \text{ if } t \in [t_b^s, t_c^s) \cup (t_c^{\mathrm{e}}, t_b^{\mathrm{e}}] \\ f_j \text{ if } t \in [t_c^s, t_c^{\mathrm{e}}] \\ 0 \text{ if Any other case} \end{cases} \tag{3.3}$$

The equilibrium, in this context, considers that all users have the same total cost irrespective of their departure time or their travel mode. In (3.4) and (3.5), we calculate the cost for the fist car user and the first bus user, who depart at t_c^s and t_b^s, respectively. Congestion and queuing at the bus stop are transitory phenomena, since we notice that the first bus user who departs at t_b^s neither faces queuing at the bus stop nor road congestion. For the first car user, the situation is similar, and does not face road congestion. Considering that $c_c(t_c^s) = c_b(t_b^s)$, we obtain (3.8).

$$c_c(t_c^s) = p_c + r_c + \beta(t^* - t_c^s) \tag{3.4}$$
$$c_b(t_b^s) = p_b + r_b + \beta(t^* - t_b^s) \tag{3.5}$$

Using the same argument that we use earlier, we calculate the total cost for the last car user and the last bus user, they depart at times t_c^{e} and t_b^{e} respectively. In this case, car user does not face road congestion, because, she is the last user and the congestion is dissipated. Last public transport user does not face queuing at the bus stop. Considering that $c_c(t_c^{\mathrm{e}}) = c_b(t_b^{\mathrm{e}})$, we obtain (3.9).

$$c_c(t_c^s) = p_c + r_c + \gamma(t_c^{\mathrm{e}} - t^*) \tag{3.6}$$
$$c_b(t_b^s) = p_b + r_b + \gamma(t_b^{\mathrm{e}} - t^*) \tag{3.7}$$

Eq. (3.8) represents the period of time of the early bus users, who use the public transport system before the first car user. On the other hand, Eq. (3.9) represents the period of later bus users, those who use the public transportation after the last car user. Combining both equations, we obtain Eq. (3.10), which represents the duration of the *uncongested* period. Whether the *uncongested* period exists or not depends on the full price difference. If the full

[8]The period definition only make sense if $t_b^s < t_c^s$ and $t_c^{\mathrm{e}} < t_b^{\mathrm{e}}$.

prices are more differentiated *uncongested* period is going to be longer.

$$t_c^s - t_b^s = \frac{(p_c + r_c) - (p_b + r_b)}{\beta} \tag{3.8}$$

$$t_b^e - t_c^e = \frac{(p_c + r_c) - (p_b + r_b)}{\gamma} \tag{3.9}$$

$$t_c^s - t_b^s + t_b^e - t_c^e = \frac{\beta + \gamma}{\beta \cdot \gamma} \cdot ((p_c + r_c) - (p_b + r_b)) \tag{3.10}$$

Eq. (3.10) shows us that the *uncongested* duration depends directly on car and public transport full price.

Mixed traffic conditions imply that buses and cars share road capacity s. The frequency of buses is constant over each period. In particular, frequency is constant and equal to f_j in the interval $[t_c^s, t_c^e]$. For that interval, car users face a decreased capacity of $s - \lambda \cdot f_j > 0$, it means that buses are modeled as a continuous flow, and the buses by themselves do not generate congestion. Eq. (3.11) represents the equalization of cost for the first and last car user; we consider those users only face schedule delay cost. Since they are the first and last users there is no road congestion. Throughout the entired congested period it is necessary that all vehicles pass through the bottleneck, which is summarized in Eq. (3.12). It is clear that Eqs. (3.11) and (3.12) do not depend on N_b, t_b^s or t_b^e,

$$\beta(t^* - t_c^s) = \gamma(t_c^e - t^*) \tag{3.11}$$

$$(t_c^e - t_c^s) \cdot (s - \lambda \cdot f_j) = N_c \tag{3.12}$$

For bus user equilibrium, it is possible to follow the same reasoning we use below. Eq. (3.13) shows that the first and last departures by bus must only face schedule delay cost. Using Eq. (3.14) we set other important condition; the operational period of the public transport system must be such that all commuters can go through the bottleneck at the bus stop. As bus frequency depends on the *congested* and *uncongested* period, then Eq. (3.14) contains t_c^s and t_c^e.

$$\beta(t^* - t_b^s) = \gamma(t_b^e - t^*) \tag{3.13}$$

$$(t_c^e - t_c^s) \cdot k \cdot f_j + (t_c^s - t_b^s) \cdot k \cdot f_u + (t_b^e - t_c^e) \cdot k \cdot f_u = N_b \tag{3.14}$$

Solving the system of equations defined by Eqs. (3.11) to (3.14), we obtain, conditional on N_b and N_c, the equilibrium times for car and bus operations and the equilibrium cost of traveling by each mode. Defining $\delta = \beta \cdot \gamma / (\beta + \gamma)$, these are:

$$t_c^s = t^* - \frac{\delta}{\beta} \frac{N_c}{s - \lambda \cdot f_j} \tag{3.15}$$

$$t_c^e = t^* + \frac{\delta}{\gamma} \frac{N_c}{s - \lambda \cdot f_j} \tag{3.16}$$

$$t_b^s = t^* - \frac{\delta}{\beta} \frac{(f_u - f_j) \frac{N_c}{s - \lambda \cdot f_j} + \frac{N_b}{k}}{f_u} \tag{3.17}$$

$$t_b^e = t^* + \frac{\delta}{\gamma} \frac{(f_u - f_j) \frac{N_c}{s - \lambda \cdot f_j} + \frac{N_b}{k}}{f_u} \tag{3.18}$$

$$c_c = p_c + r_c + \delta \frac{N_c}{s - \lambda \cdot f_j} \tag{3.19}$$

$$c_b = p_b + r_b + \delta \frac{(f_u - f_j) \frac{N_c}{s - \lambda \cdot f_j} + \frac{N_b}{k}}{f_u} \tag{3.20}$$

Analyzing Eqs. (3.15) to (3.20), we recover the results from Basso et al. (2019), if f_u goes to f_j.

Equalizing costs across modes give us the equilibrium modal split ($c_c = c_b$) and using that $N_c + N_b = N$, we get the unique solution given by:

$$N_c = \frac{(s - \lambda \cdot f_j)\left(\delta \cdot N - k \cdot f_u \cdot ((p_c + r_c) - (p_b + r_b))\right)}{\delta(s - \lambda \cdot f_j + k \cdot f_j)} \tag{3.21}$$

$$N_b = k \cdot \frac{\delta \cdot f_j \cdot N + f_u \cdot ((p_c + r_c) - (p_b + r_b))(s - \lambda \cdot f_j)}{\delta(s - \lambda \cdot f_j + k \cdot f_j)} \tag{3.22}$$

It seems natural to consider that $N_c, N_b > 0$. Analyzing Eqs. (3.21) and (3.22), the interior solution requires:

$$\delta \cdot N - k \cdot f_u \cdot ((p_c + r_c) - (p_b + r_b)) > 0 \tag{3.23}$$

Condition (3.23) indicates that f_u is not enough large to captures all demand during the *uncongested* period. It is important to remember that $((p_c + r_c) - (p_b + r_b))/\delta$ represents the total duration of the *uncongested* period. In all cases, we use the analysis N_c and N_b that satisfies Condition (3.23). In Section 3.3 we optimize f_u and f_j, however we will assume that f_u satisfies Condition (3.23).

Finally, the equilibrium cost c is:

$$c = \frac{(p_c + r_c) \cdot (s - \lambda \cdot f_j + k \cdot f_j) - f_u \cdot k \cdot (p_c + r_c - p_b - r_b) + \delta \cdot N}{s - \lambda \cdot f_j + k \cdot f_j} \tag{3.24}$$

Considering all the previous results, we can state the first proposition.

Proposition 3.1 *If the full car full price (congestion price plus resource costs) is higher than the public transport full price (bus fare plus resource cost), and the difference between car full price and bus full price is not large enough, ie.* $p_c + r_c > p_b + r_b$ *and* $p_c + r_c - p_b - r_b < \delta N/(k \cdot f_u)$, *then there is a unique equilibrium in which both modes are used. The public transport demand starts earlier and ends later than the car demand. Queues at bus stops start to develop up until the moment of departure of the first bus user that faces road congestion. During the car peak period, the length of the queue at the bus stop remains constant, and it begins to dissipate after the departure of the last bus user that faces road congestion. During the car peak hour, a queue at the bottleneck on the road begins to develop at the moment of the first car departure and grows linearly for early arrivals and shrinks linearly for late arrivals.*

PROOF. See appendix B.1 □

The intuition for this Proposition is as follows. Since the car full price is higher than the bus full price, in order to keep the total cost constant, there are incentives for bus users to start their journeys earlier, because bus users trade schedule delay cost for price (monetary and resources), it implies that the first bus user departs when there are no cars on the road. Since the total cost between modes is constant, subsequent bus departures trade schedule delay for queuing delays at bus stops. Then, when the first car user departs, they face no road congestion, only schedule delay costs. Later departures by car trade congestion delays for schedule delays. As we know, car users time cost is given by congestion cost plus schedule delay cost. A bus user who takes the bus at the same moment as a car user starting their trip will experiences the same congestion and schedule delays as car users; then there is no possibility of any variation in the bus users cost during this period, consequently queuing delays at the bus stop remain constant during the *congested* period.

In Section 3.3, we optimize f_u and f_j. However it is important to show a few extreme cases that will be useful in the following section.

It is not difficult to consider an equilibrium without public transportation; we set $f_u = f_j = 0$, and we recover the results from Arnott et al. (1990, 1993). In this case the equilibrium cost (c_c^o) and the user time cost ($\bar{c}_c$) are :

$$c_c^o = p_c + r_c + \frac{\delta \cdot N}{s} \tag{3.25}$$

$$\bar{c}_c = \frac{\delta \cdot N}{s} \tag{3.26}$$

On the other hand, an equilibrium without cars on the road is much harder to conceive. First, we would require a situation where cars are not used. For this, we need to consider a bus-only situation, it means, if a commuter wishes to switch modes unilaterally, the commuter has to face a cost higher or equal than the current cost. If there are no cars on rad and frequency is f, then equilibrium cost is the following.

$$c_b^f = p_b + r_b + \frac{\delta \cdot N}{k \cdot f} \tag{3.27}$$

$$\bar{c_b}^f = \frac{\delta \cdot N}{k \cdot f} \tag{3.28}$$

where c_b^f and $\bar{c_b}^f$ are the average cost and the user time cost of the only-bus situation for a frequency f.

A bus user who wants to switch to car will face a cost $c_c^s = p_c + r_c$, then it is necesary that f satisfies:

$$f \geq f^o = \frac{\delta \cdot N}{k \cdot ((p_c + r_c) - (p_b + r_b))} \tag{3.29}$$

where f^o is the minimum frequency that allows us to have an equilibrium without cars on the road.

3.3 Frequency optimization

We characterized one equilibrium under mixed traffic conditions for two modes, and two frequencies for public transport. The main idea of frequency differentiation is to consider the changes in operational expenditure in relation to the congestion level. As users are homogeneous, those who start their trip at the beginning of the *uncongested* period show the same behavior in comparison with any other user during the *congested* period. During the *congested* period, we simplify the modeling considering a flat frequency throughout the period. It would then seem natural to also consider a constant frequency during the *uncongested* period.

In Section 3.2, we describe an interior equilibrium, requiring that $0 < p_c + r_c - p_b - r_b < \delta N/(k \cdot f_j)$. We proceed to optimize frequency following the next procedure: we will minimize the social cost function that is valid for an interior equilibrium (detailed next) over $\Delta p = p_c - p_b$ and f_u, and f_j. If these three values fulfill the conditions of Proposition 3.1 and 3.2, then this is the optimum.

If, on the other hand, the result for the first-order conditions (denoted by *) do not satisfy the conditions of Propositions 3.1 and 3.2, then it is optimum not to provide public transportation at all. The *uncongested* period does not exist if $p_c^* + r_c - p_b^* - r_b = 0$, and the analysis is the same as in Basso et al. (2019).

Defining $\Delta r = r_c - r_b$ and $\Delta p = p_c - p_b$ we calculate the total user cost (UC) for the interior equilibrium as N times the equilibrium users cost determined by Eq. (3.24) and subtracting the price paid[9] by each user.

$$UC = N \cdot \frac{(p_c + r_c) \cdot (s - \lambda \cdot f_j + k \cdot f_j) - f_u \cdot k \cdot (\Delta p + \Delta r) + \delta \cdot N}{s - \lambda \cdot f_j + k \cdot f_j} - N_c \cdot p_c - N_b \cdot p_b \tag{3.30}$$

[9]We consider congestion charges or bus fares as resource transfers.

Reordering the terms and using the equilibrium modal split, N_c and N_b, from Eqs. (3.21) and (3.22), we get the following result:

$$\begin{aligned} UC =& N \cdot \frac{(p_c + r_c) \cdot (s - \lambda \cdot f_j + k \cdot f_j) - f_u \cdot k \cdot (\Delta p + \Delta r) + \delta \cdot N}{s - \lambda \cdot f_j + k \cdot f_j} \\ & - p_c \cdot \frac{(s - \lambda \cdot f_j)\,(\delta \cdot N - k \cdot f_u \cdot (\Delta p + \Delta r))}{\delta (s - \lambda \cdot f_j + k \cdot f_j)} \\ & - p_b \cdot k \cdot \frac{\delta \cdot f_j \cdot N + f_u \cdot (\Delta p + \Delta r)\,(s - \lambda \cdot f_j)}{\delta (s - \lambda \cdot f_j + k \cdot f_j)} \end{aligned} \tag{3.31}$$

The operational cost has an impact on the social optimum. It is, therefore, necessary to consider this in our model. We model the operational cost considering two components: fleet size and the number of buses that are dispatched. The number of dispatched buses at rush hour captures the operational and labor expenses. Also, as shown in our model, since all buses run the same distance, they capture the total vehicle-kilometers. Considering this, we define the expenditure of providing a frequency described in Eq. (3.3) by:

$$E(f) = \underbrace{c_1 \cdot \max\{f_j \cdot T_j; f_u \cdot T_u\}}_{E_1(f)} + \underbrace{c_2 \cdot (f_j \cdot (t_c^{e} - t_c^{s}) + f_u \cdot ((t_c^{s} - t_b^{s}) + (t_b^{e} - t_c^{e})))}_{E_2(f)} \tag{3.32}$$

where $E_1(f)$ represents the expenditure related to the bus fleet and $E_2(f)$ to dispatche and c_1 is the constant cost per bus. Initially, we are unaware what the frequency levels are, therefore the fleet size required to provide a frequency $f(t)$ is expressed by $\max\{f_j \cdot T_j; f_u \cdot T_u\}$, T_j is the cycle time during the *congested* period and T_u is the cycle time during the *uncongested* period. The cycle time can be split into two terms; the free-flow cycle time (that is also the cycle time in the *uncongested* period, T^0) and the time spent passing through the road bottleneck. The constant c_2 is the cost per bus dispatched and $f_j \cdot (t_c^{e} - t_c^{s}) + f_u \cdot ((t_c^{s} - t_b^{s}) + (t_b^{e} - t_c^{e}))$ is the number of dispatches during the entire operational period.

Since mixed traffic always suffers road congestion, we need to account for this travel time in transit expenditures. We consider that the cycle time that determines the fleet size during the peak period is a fraction of the maximum travel time over the period and round trip free flow time. The maximum travel time is simply the individual user cost minus the car full price and divided by α. Using Eq. (3.19) minus $p_c + r_c$ and considering N_c from Eq. (3.21). We write the cycle times as:

$$T_u = T^0 \tag{3.33}$$

$$T_j = T^0 + z \cdot \frac{\delta}{\alpha} \frac{N - \frac{(\Delta p + \Delta r) \cdot k \cdot f_u}{\delta}}{s - \lambda \cdot f_j + k \cdot f_j} \tag{3.34}$$

where the second term of Eq. 3.34 is a fraction of the maximum travel time which is determined by the maximum road queue length, and $z \in (0, 1]$[10].

Replacing the cycle time of Eq. (3.34) into E_1, and Eqs. (3.15) to (3.21), into E_2 we

[10] It is easy to prove that if the frequency is constant during the *congested* period then $z = 1/2$.

obtain:

$$E_1 = c_1 \cdot \max\left\{ f_j \cdot \left(T^0 + z \cdot \frac{\delta}{\alpha} \frac{N - \frac{(\Delta p + \Delta r)\cdot k \cdot f_u}{\delta}}{s - \lambda \cdot f_j + k \cdot f_j} \right); f_u \cdot T^0 \right\} \tag{3.35}$$

$$E_2 = c_2 \cdot \left(\frac{\Delta p + \Delta r}{\delta} \cdot f_u + \frac{(\delta \cdot N - k \cdot f_u \cdot (\Delta p + \Delta r))}{\delta(s - \lambda \cdot f_j + k \cdot f_j)} \cdot f_j \right) \tag{3.36}$$

We define the social cost (SC) directly as the sum of the total user cost (Eq. (3.31)) and the public transport system expenditure (Eqs. (3.35) and (3.36)).

$$\begin{aligned} SC = & N \cdot \frac{(p_c + r_c) \cdot (s - \lambda \cdot f_j + k \cdot f_j) - f_u \cdot k \cdot (\Delta p + \Delta r) + \delta \cdot N}{s - \lambda \cdot f_j + k \cdot f_j} \\ & - p_c \cdot \frac{(s - \lambda \cdot f_j)\,(\delta \cdot N - k \cdot f_u \cdot (\Delta p + \Delta r))}{\delta(s - \lambda \cdot f_j + k \cdot f_j)} \\ & - p_b \cdot k \cdot \frac{\delta \cdot f_j \cdot N + f_u \cdot ((\Delta p + \Delta r))\,(s - \lambda \cdot f_j)}{\delta(s - \lambda \cdot f_j + k \cdot f_j)} \\ & + c_1 \cdot \max\left\{ f_j \cdot \left(T^0 + z \cdot \frac{\delta}{\alpha} \frac{N - \frac{(\Delta p + \Delta r)\cdot k \cdot f_u}{\delta}}{s - \lambda \cdot f_j + k \cdot f_j} \right); f_u \cdot T^0 \right\} \\ & + c_2 \cdot \left(\frac{\Delta p + \Delta r}{\delta} \cdot f_u + \frac{(\delta \cdot N - k \cdot f_u \cdot (\Delta p + \Delta r))}{\delta(s - \lambda \cdot f_j + k \cdot f_j)} \cdot f_j \right) \end{aligned} \tag{3.37}$$

Let us define T_u as the duration of the *uncongested* period, the sum of the early and late periods. Using Eqs. (3.17) and (3.18) and the definitions of Δp, Δr and δ.

$$T_u = \frac{\Delta p + \Delta r}{\delta} \tag{3.38}$$

We assume for all the following analysis that $\Delta p + \Delta r = p_c - p_b + r_c - r_b$ is strictly positive and constant.

We show in Section 3.2, that there are three possible equilibria. After the frequency optimization, it is possible to reach one of those. We focus on the equilibrium where both modes have positive demand. However, it is essential to set conditions for the other two cases: all demand using public transportation, and all demand using a private car.

It seems natural that if public transportation expenditure is low, it is efficient to a provide frequency high enough to transport all demand by bus. In Appendix B.3, we show that Eqs (3.39) and (3.40) are sufficient conditions for c_1 and c_2 to reach an optimal frequency in which all commuters use public transportation.

$$c_2 < k \cdot \frac{\delta \cdot N}{s} - c_1 \cdot \frac{T^0}{T_u} - k \cdot \Delta p \tag{3.39}$$

$$c_1 < \frac{N \cdot \delta \cdot \lambda}{s} \cdot \frac{T_u}{T^0} \tag{3.40}$$

On the other hand, if public transportation expenditure is high, it comes as no surprise that it is inefficient to offer a bus service. In Appendix B.3, we prove that Eq. (3.41) is a sufficient condition to ensure that the optimal frequencies are $f_j^* = f_u^* = 0$.

$$c_2 > \max\left\{k \cdot \frac{\delta \cdot N}{s} - c_1 \cdot \frac{T^0}{T_u} - k \cdot \Delta p; \frac{\delta \cdot N}{s} \cdot (k - \lambda) - k \cdot \Delta p\right\} \tag{3.41}$$

To achieve positive demand in both modes, it is necessary that c_1 and c_2 satisfy the following conditions.

$$c_2 < \max\left\{k \cdot \frac{\delta \cdot N}{s} - c_1 \cdot \frac{T^0}{T_u} - k \cdot \Delta p; \frac{\delta \cdot N}{s} \cdot (k - \lambda) - k \cdot \Delta p\right\} \tag{3.42}$$

$$c_1 > \frac{N \cdot \delta \cdot \lambda}{s} \cdot \frac{T_u}{T^0} \tag{3.43}$$

Keeping these conditions in mind , we state the following proposition.

Proposition 3.2 *If both modes have positive demand, then the total fleet is used completely in both period, i.e.* $f_j^* \cdot \left(T^0 + z \cdot \frac{\delta}{\alpha} \frac{N - \frac{(\Delta p + \Delta r) \cdot k \cdot f_u^*}{\delta}}{s - \lambda \cdot f_j^* + k \cdot f_j^*}\right) = f_u^* \cdot T^0$*, and we conclude that* $f_u^* > f_j^*$.

PROOF. See Appendix B.2 □

The intuition for Proposition 3.2 is as follows. Let us to suppose that fleet size is fixed, and the dispatched cost is equal to zero ($c_2 = 0$). In this case, the optimal solution is to move as many buses as possible; it means the optimal solution considers a higher frequency during the *uncongested* period (due to the lower cycle time) and a lower frequency during the *congested* period. If we consider $c_2 > 0$ and small, the intuition still considers that it is efficient to use the total fleet as much as possible. This case also makes a link with the existence of positive demand in both modes, c_2 has to meet the conditions from Eqs. (3.42) and (3.43). Eq. (3.42) creates an upper boundary for c_2.

High frequency is concentrated during the *uncongested* period; if we compare our result with the previous literature, it is possible to notice a similarity with the efficient timetable by Kraus (2003); who creates clusters that start at the desired arrival time with each cluster using all the entire available fleet. The next cluster starts when the fleet is available again (considering a fixed cycle time). Our optimization also creates clusters located during the *uncongested* period using all the available fleet.

3.4 Numerical analyses

In this section we illustrate our analytical results using values for the parameters that may represent an actual situation. The values for the numerical simulations are presented in Table 3.1. The parameters are the same are those used by Basso et al. (2019)[11] in their numerical

[11] We consider $z = 1$ to have comparable results with Basso et al. (2019).

analyses. It is important to remember that we consider that there is no (time-invariant) congestion charge for cars, i.e. $p_c = 0$. Therefore, we report below, directly, the bus fare (per trip).

Table 3.1: Parameters for numerical examples

Parameter	Units	Value
c_1	[US$/bus]	290
c_2	[US$/hour]	130
T_0	[hour]	0.33
N	[Conmuters]	6000 to 13000
s	[PCU/hour]	6000
Δr	[US$/Pax]	2.0
α	[US$/hour]	2.6
β	[US$/hour]	1.95
γ	[US$/hour]	3.9
k	[Pax/bus]	80
α_2	[US$/hour]	5.2
t^*	[hh:mm]	08:00
λ	[PCU/bus]	3.5
z		1

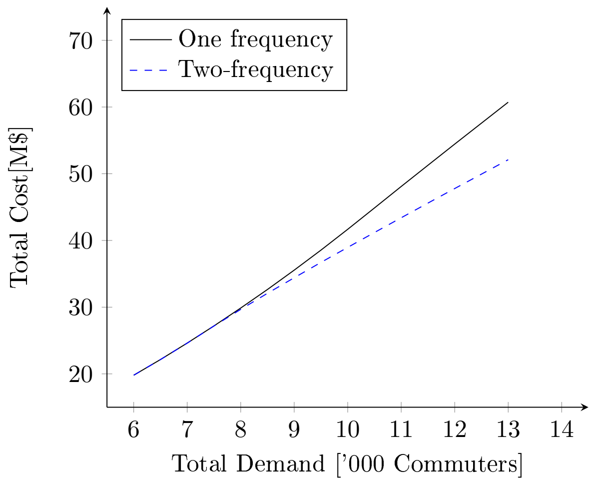
In our numerical analysis, we optimize under mixed traffic a model with two frequencies; one during the uncongested period and other during the congested period. We also optimize the fare for public transportation. Frequencies during each period are constant and there are no road facilities for public transportation.

First, we compare our numerical results with one of the cases developed in Basso et al. (2019), the optimization under mixed traffic with a constant frequency. Our results use the legend ‘Two-frequency’ and results for Basso et al. (2019) use the legend ‘One frequency’. It is important to point out that results are under mixed traffic conditions in both cases.

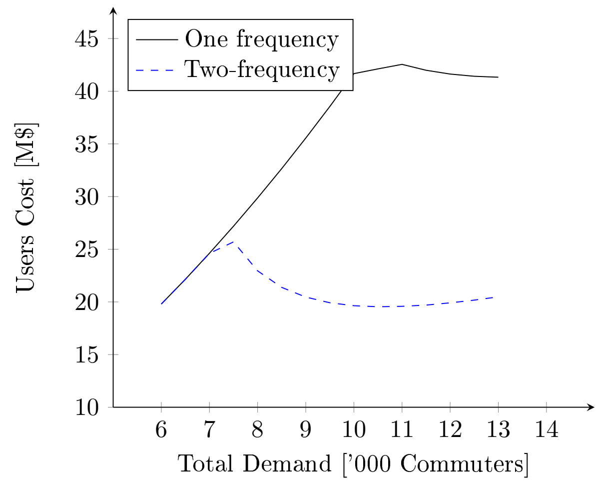
The ‘One frequency’ optimization is not efficient for demands under 10,000 commuters; therefore, for lower demand, public transportation is not provided. The ‘Two-frequency’ optimization provides public transportation when demand is around 7,500 commuters, and in all cases the ‘One frequency’ optimization provides lower frequencies than the ‘Two-frequency’ optimization. Analyzing panel (a) from Figure 3.2, the demand in which the two curves separate, it indicates the moment when it is efficient to provide public transportation for the ‘Two-frequency’ optimization. Also, the ‘Two-frequency’ optimization generates an improvement in the social cost, with a reduction up to 14% of the total cost (for 13,000 commuters). This efficiency comes from a significant reduction of the user costs shown in panel (b) from Figure 3.2.

For more clarity, we show in Table 3.2 the optimization results for a demand of $N = 11,000$ commuters. For the complete results, see Table B.1 from Appendix B.4.

The ‘Two-frequency’ optimization attracts more demand for public transportation than the ‘One frequency’ optimization. Public transport demand for the mixed traffic optimization



(a) Total cost



(b) User cost

Figure 3.2: 'One frequency' versus 'Two-frequency'

Table 3.2: Numerical Results for $N = 11,000$ commuters

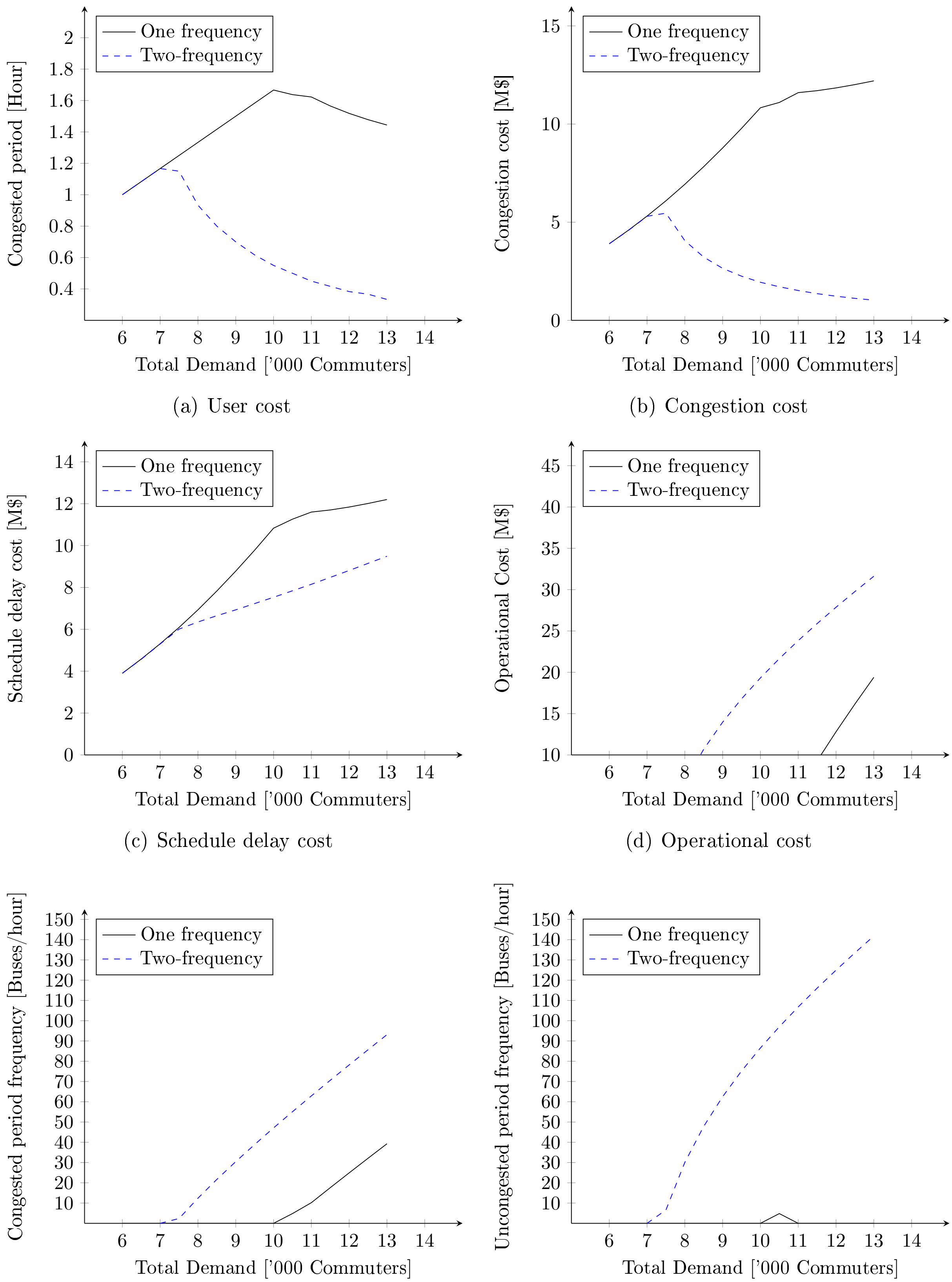
	f_j	f_u	p_b	N_c	N_b	t_b^s	t_c^s	t_c^e	t_b^e	TC	OC	UC	CC	SDC	QC
One frequency	10	0	2.00	9677	1323	6:55	6:55	8:32	8:32	48089	5534	42555	11601	11601	0
Two-Frequency	63	107	1.09	2690	8310	7:13	7:41	8:09	8:23	43409	23829	19581	1522	8159	4521

is $N_c^{MT} = 1,323$ commuters and for the 'Two-frequency' is $N_c^{TF} = 8,310$ ($\Delta N = 6,987$). The rise of public transport demand implies an increase in operation cost ($\Delta OC = 18,295$), but total cost decreases in $\Delta TC = -4,680$, mostly explained by an important reduction in users cost ($\Delta UC = -22,974$). The reduction of user cost is directly related to the reduction of rush-hour duration, starting 18 minutes later (from $6:55$ to $7:13$) and ending 9 minutes early (from $8:32$ to $8:23$).

Focusing on Figure 3.3, panels (b) and (c) allow us to analyze the user costs reduction as a reduction of the schedule delay cost and a decrease of the congestion cost. From panel (a) is clear that *congested* period reduces quickly with the demand, consequently reducing the congestion cost.

If we compare panels (e) and (f), the *uncongested* period frequency is higher than the *congested* period frequency, which is consistent with the theoretical result from Proposition 3.2. The frequency increasing impacts on the operational cost. However, this impact is restrained because the frequency during the *uncongested* period reduces the car's modal split, thus reducing the total congestion, and this frequency does not require any additional fleet.

Our final step is to compare how efficient the 'Two-frequency' optimization is in comparison with an efficient BRT system. The main idea of this point is to understand how efficient an infrastructure measure is in comparison with a management measure. The optimized BRT system is divided into two types according to the results from Basso et al. (2019), one where the road capacity is completely divisible (BRT-DC), making it is possible to reserve the exact capacity for buses that public transportation needs, and a second case where the road capacity



(a) User cost

(b) Congestion cost

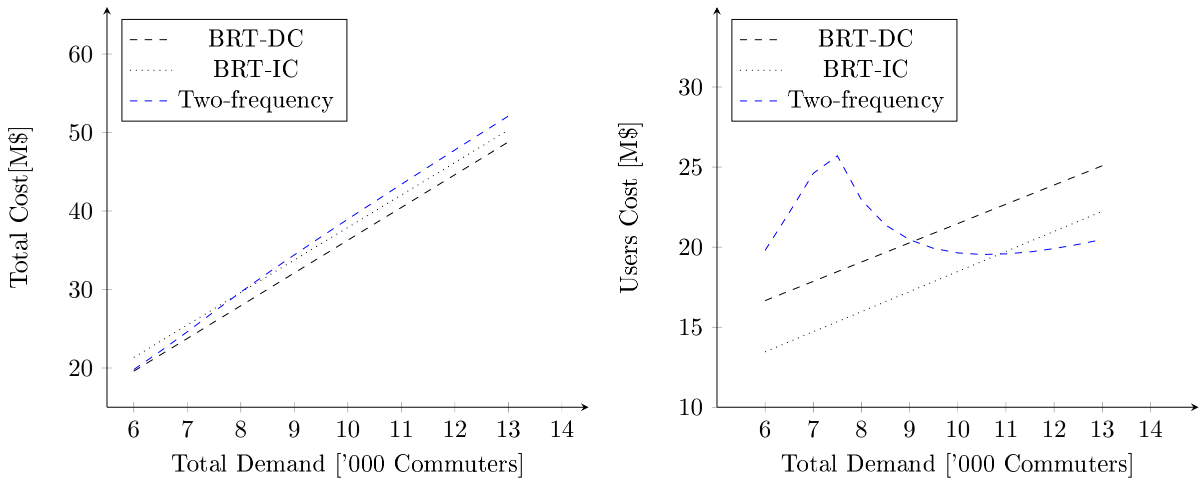
(c) Schedule delay cost

(d) Operational cost

(e) Congested period frequency

(f) Uncongested period frequency

Figure 3.3: 'One frequency' versus 'Two-frequency'



(a) Total cost

(b) User cost

Figure 3.4: BRT versus 'Two-frequency'

is indivisible (BRT-IC); hence we reserve a fraction of the road capacity. For example, with one full lane, in all cases, the reserved capacity is enough to avoid road congestion for buses.

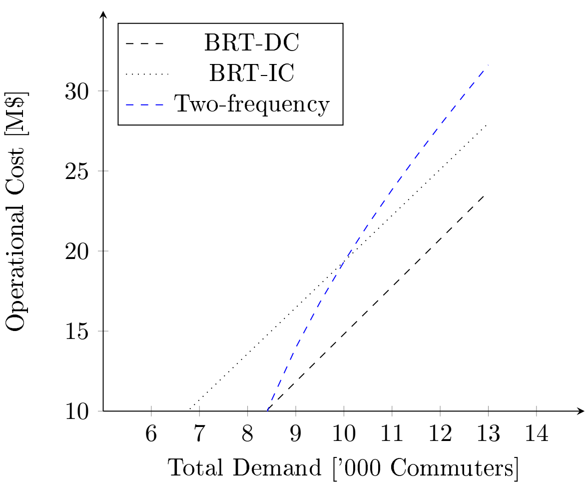
It is clear that an efficient BRT system is more efficient than the 'Two-frequency' optimization. However, the differences in efficiency are relatively small. In our analysis, the best-case scenario pegs the total cost for the efficient BRT-DC at less than 8% lower than the 'Two-frequency' optimization. If we compare the 'Two-frequency' optimization versus the efficient BRT-IC, the total cost for the BRT-IC is less than 4% lower than the 'Two-frequency' optimization.

The difference between the total cost reported in Figure 3.4 is a consequence of the operational cost. If we analyze panel (a) from Figure 3.5, the operational cost differences come from the *uncongested* period frequency. For a demand close to 10,000 commuters, the *uncongested* period frequency for the 'Two-frequency' optimization is higher than the optimal BRT frequency (BRT-DC and BRT-IC). The *congested* period frequency is always lower than the optimal BRT frequency.

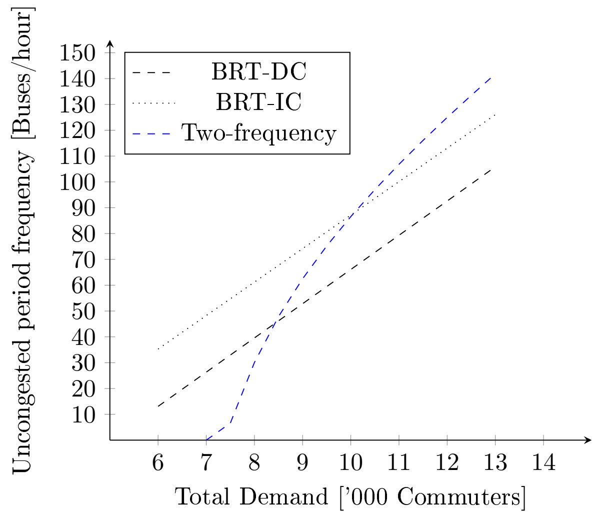
3.5 Conclusions

In this paper we have analyzed the problem of finding the optimal frequency whenthe rush hour can be divided into two parts: one when there is no congestion (no cars on-road) and one when there is congestion on the road. We use a dynamic congestion model in a mixed traffic environment (buses and cars share the road capacity) with inelastic total demand. Modal split and temporal distribution of demand are endogenous to the model as part of the full equilibrium. Rush-hour length, *congested* and *uncongested* periods, and cycle time during the *congested* period are endogenous to the model.

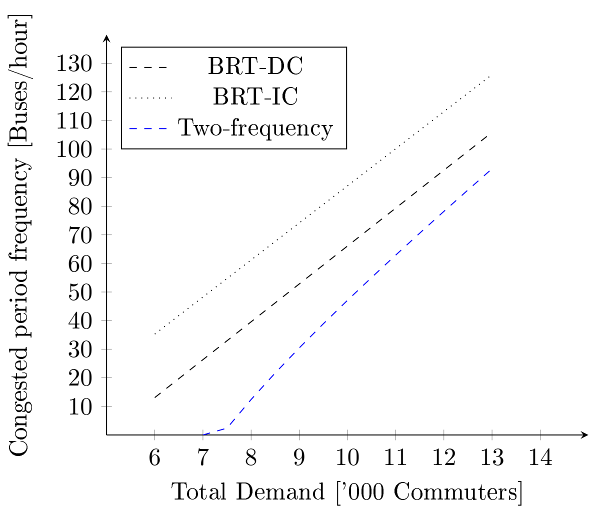
We provide a microeconomic analysis of the rush-hour non-constant frequency optimization, where queuing and congestion delays are endogenous as a result of individual schedules of departures. The main difference with previous literature that researches bi-modal (car



(a) Operational cost



(b) Uncongested period frequency



(c) Congested period frequency

Figure 3.5: BRT versus 'Two-frequency'

and public transport) systems is that they consider that public transport vehicles are free of road congestion and in those cases where they do model road congestion for buses, they do not use more than one frequency during rush hour.

It is essential to keep in mind that in our paper we model only one type of commuter, who can choose freely between car or bus, and as the system is in equilibrium, all users have the same total cost. We show analytically and numerically that, under plausible conditions, the optimal frequency during the *congested* period is lower than the frequency during the *uncongested* period. Our results hold up in the reduction of users costs with a decrease in congestion and schedule delay cost by moving people from cars to buses and scheduling those buses during the *uncongested* period. Our numerical analysis shows that by only using a non-constant frequency optimization, without any additional road facilities for public transportation, we obtain a reduction of up to 14% in the social cost compared to the constant frequency case. Moreover, if we extend our analysis and compare the two-frequency (management measure) optimization against a BRT system (infrastructure measure), we find that the gains of a BRT system are less than 8% of the total cost.

Regarding future research, we made assumptions that may be modified to test our results. To consider heterogeneous users and their effect on the optimal frequency, especially during the uncongested period, it is plausible that the inclusion of captive demand for public transport modifies optimal frequency. We consider that focusing the research on these topics is methodologically relevant and contributes to the development public policies.

Chapter 4

Conclusions

We studied public transport optimization using a classic bottleneck model, considering second-best policies. In Chapter 2 we provide a microeconomic analysis of BRTs, while Chapter 3 analyzes rush-hour non-constant frequency optimization.

Chapter 2 shows that even frequency and operational time are optimized. A BRT system will generate queuing at the bus station, while fares cannot vary perfectly with time. With a BRT, fares will be lower because the peak hours of operation of the system are shorter. The car peak-period is also shorter, even though capacity was taken from private transport. Importantly, this better-for-all situation features more boarding delays, that is to say, queues at bus stops will be longer than under mixed-traffic conditions.

In Chapter 3 we analyzed the problem of finding the optimal frequency when we able to divide the rush hour into two parts: one when there is no congestion (no cars on-road) and one when there is congestion on the road. We use a dynamic congestion model in a mixed traffic environment (buses and cars share the road capacity) with inelastic total demand. We show analytically and numerically that, under reasonable assumptions, the optimal frequency during the congested period is lower than the frequency during the uncongested period.

Regarding future research, we made assumptions that may be modified to test our results. To name a few, inelastic demand, homogeneous users, full access to cars, no crowding costs, and homogeneous users. We consider that focus the research on these topics is methodologically relevant and contributes to the development of public policies.

Bibliography

Arnott, R., de Palma, A., and Lindsey, R. (1990). Economics of a bottleneck. *Journal of Urban Economics*, 27(1):111–130.

Arnott, R., de Palma, A., and Lindsey, R. (1993). A Structural Model of Peak-Period Congestion : A Traffic Bottleneck with Elastic Demand. *The American Economic Review*, 83(1):161–179.

Basso, L., Feres, F., and Silva, H. (2019). The efficiency of bus rapid transit (BRT) systems: A dynamic congestion approach. *Transportation Research Part B: Methodological*, 127.

Basso, L. J., Guevara, C. A., Gschwender, A., and Fuster, M. (2011). Congestion pricing, transit subsidies and dedicated bus lanes: Efficient and practical solutions to congestion. *Transport Policy*, 18(5):676–684.

Basso, L. J. and Silva, H. E. (2014). Efficiency and substitutability of transit subsidies and other urban transport policies. *American Economic Journal: Economic Policy*, 6(4):1–33.

Börjesson, M., Fung, C. M., and Proost, S. (2017). Optimal prices and frequencies for buses in stockholm. *Economics of transportation*, 9:20–36.

de Palma, A., Lindsey, R., and Monchambert, G. (2017). The economics of crowding in rail transit. *Journal of Urban Economics*, 101:106–122.

Gonzales, E. J. and Daganzo, C. F. (2012). Morning commute with competing modes and distributed demand: User equilibrium, system optimum, and pricing. *Transportation Research Part B: Methodological*, 46(10):1519–1534.

Hensher, D. A., Li, Z., and Mulley, C. (2014). Drivers of bus rapid transit systems–influences on patronage and service frequency. *Research in Transportation Economics*, 48:159–165.

Huang, H. J. (2000). Fares and tolls in a competitive system with transit and highway: The case with two groups of commuters. *Transportation Research Part E: Logistics and Transportation Review*, 36(4):267–284.

Huang, H. J., Tian, Q., Yang, H., and Gao, Z. Y. (2007). Modal split and commuting pattern on a bottleneck-constrained highway. *Transportation Research Part E: Logistics and Transportation Review*, 43(5):578–590.

Jansson, J. (1980). Simple bus line model for optimisation of service frequency and bus size.

Journal of Transport Economics and Policy, 14(1):53–80.

Jansson, J. (1984). *Transport system optimization and pricing*. Wiley, Chichester.

Jara-Díaz, S., Fielbaum, A., and Gschwender, A. (2017). Optimal fleet size, frequencies and vehicle capacities considering peak and off-peak periods in public transport. *Transportation Research Part A: Policy and Practice*, 106(August):65–74.

Kraus, M. (2003). A new look at the two-mode problem. *Journal of Urban Economics*, 54(3):511–530.

Kraus, M. and Yoshida, Y. (2002). The Commuter's Time-of-Use Decision and Optimal Pricing and Service in Urban Mass Transit. *Journal of Urban Economics*, 51(1):170–195.

Kutzbach, M. J. (2009). Motorization in developing countries: Causes, consequences, and effectiveness of policy options. *Journal of Urban Economics*, 65(2):154–166.

Mohring, H. (1979). The benefits of reserved bus lanes, mass transit subsidies, and marginal cost pricing in alleviating traffic congestion. In Mieszkowski, P. and Straszheim, M., editors, *Current Issues in Urban Economics*, pages 165–195. The Johns Hopkins University Press, Baltimore.

Scholl, L., Bouillon, C. P., Oviedo, D., Corsetto, L., and Jansson, M. (2016). Urban transport and poverty: Mobility and accessibility effects of idb-supported brt systems in cali and lima. Technical report, Inter-American Development Bank.

Scholl, L., Guerrero, A., Quintanilla, O., Celse L'Hoste, M., and Sadeghi, P. (2015). Comparative case studies of three idb-supported urban transport projects. Technical report, Inter-American Development Bank.

Small, B. K. A. (1982). The Scheduling of Consumer Activities : Work Trips. *The American Economic Review,*, 72(3):467–479.

Small, K. A. (1983). Bus priority and congestion pricing on urban expressways. *Research in Transportation Economics*, 1:27–74.

Tabuchi, T. (1993). Bottleneck congestion and modal split. *Journal of Urban Economics*, 34:414–431.

van den Berg, V. A. and Verhoef, E. T. (2014). Congestion pricing in a road and rail network with heterogeneous values of time and schedule delay. *Transportmetrica A: Transport Science*, 10(5):377–400.

Vickrey, W. S. (1969). Congestion theory and transport investment. *The American Economic Review*, 59(2):251–260.

Wang, W. W., Wang, D. Z., Zhang, F., Sun, H., Zhang, W., and Wu, J. (2017). Overcoming the downs-thomson paradox by transit subsidy policies. *Transportation Research Part A: Policy and Practice*, 95:126–147.

Yoshida, Y. (2008). Commuter arrivals and optimal service in mass transit: Does queuing behavior at transit stops matter? *Regional Science and Urban Economics*, 38(3):228–251.

Appendix A

A.1 Equilibrium with intermittent bus capacity

The analysis and exposition of the public transport modeling in this section closely follows Kraus and Yoshida (2002) and Yoshida (2008). Literal excerpts are not marked as such, and they are taken to be acknowledged by this sentence. The demand side of the model is the same as in Section 2.2 except that instead of bus passengers experiencing boarding delays at a unit time cost of α_2, bus passengers arrive to the bus stop and might have to wait for the bus. We denote the waiting unit time cost by α_2 so that the models are analogous.

Suppose that there are R bus runs, i.e. R bus departures from the only bus stop where passengers can board, with a departure every h minutes. Each train has a strict capacity of k passengers, so that overloading is impossible. R is an integer but k is continuously-valued. For simplicity, R and k are assumed to be such that the total capacity provided, $R \cdot k$, is just enough to accommodate the demand of N_b bus passengers:

$$R \cdot k = N_b \tag{A.1}$$

We denote the times that the R runs depart from the bus stop by t_b^s,t_2,...,t_{R-1},t_b^{e} respectively, from the first to the last run. These are known to commuters with certainty. A bus user that departs at time t and arrives at time t_a faces the following costs:

$$c_b(t) = p_b(t) + r_b + \alpha_2 \cdot T_q(t) + \alpha \cdot T_w\left(t + T_q(t)\right) + \begin{cases} \beta \cdot (t^* - t_a) & \text{if } \; t_a \leq t^* \\ \gamma \cdot (t_a - t^*) & \text{if } \; t_a > t^* \end{cases} \tag{A.2}$$

which is the same expression as Eq. (2.2), but has to be interpreted differently. The difference is that $T_q(t)$ is the waiting time due to the fact that buses arrive intermittently at the stop. If a bus rapid transit system is in place, then the costs of departing at t are given by Eq. (A.2) but with $T_w = 0$. Also, following Kraus and Yoshida (2002) and Yoshida (2008), there are no boarding delays.

We begin by studying the case in which buses do not interact with cars on the road, which corresponds to the analysis in the presence of a BRT system. The focus of this section is to prove that there is an equilibrium when bus service is intermittent and that as frequency grows, it approaches the continuous approximation that we use in the main text.

To prove that there is an equilibrium we need to prove that there is a departure rate of bus and car users that satisfies the three equilibrium conditions: (i) all car drivers must face the same total cost irrespective of their departure time; (ii) all bus users must also face the same total cost irrespective of their departure time; and, (iii), a car driver and a bus user departing at the same time must face the same total cost. The first two conditions are dynamic equilibrium conditions. The third is the modal split equilibrium condition.

The existence of a departure pattern that satisfies the dynamic equilibrium condition for bus users is derived in Yoshida (2008). We, therefore, explain the intuition and derive the equilibrium costs, but do not prove its existence. The proofs of the equilibrium costs and its underlying departure pattern are in Section 2 of Yoshida (2008)[1]. To satisfy the dynamic user equilibrium condition for bus users, cost must be the same for all passengers on a particular run. For this to happen, they must all arrive at the origin stop *en masse*, as otherwise some would wait more than others. Also, they need to be the same for all runs. This can only be possible if the sum of waiting cost and the schedule delay cost is the same for all commuters. Thus, equilibrium requires that a commuter who incurs higher schedule delay cost is compensated by lower waiting cost. The highest schedule delay cost is incurred by either the passengers on the first run or the passengers on the last run. To minimize the schedule delays, conditional on the number of runs, the planner sets the times such that the first and last bus departure experience the same scheduling costs. This is:

$$\beta(t^* - t_b^s) = \gamma(t_b^{\mathrm{e}} - t^*) \tag{A.3}$$

Passengers in the first and last departure do not incur any waiting (they arrive *en masse* just at t_b^s and t_b^{e}). All other passengers incur waiting by arriving at the bus stop strictly earlier than the departure times of the bus that they board, so that everyone has the same user cost. Naturally, the passengers who arrive closer to t^* must wait longer.

As R runs operate in the period $t_b^{\mathrm{e}} - t_b^s$ at a constant headway of h, the following condition provides a relationship between the number of runs, headway and the duration of the period in which public transport services are provided:

$$t_b^{\mathrm{e}} - t_b^s = (R-1) \cdot h \tag{A.4}$$

Combining Eqs. (A.3) and (A.4) we obtain the equilibrium user time costs of bus users, UTC_b:

$$UTC_b = \delta(R-1) \cdot h \tag{A.5}$$

The equilibrium user time costs are the same as in Yoshida (2008). What is interesting is the close relationship with our continuous approach. To see this, note that $h = 1/f$ and that $R \cdot k = N_b$. Replacing R and h from these expressions into Eq. (A.5), we obtain:

$$UTC_b = \delta \cdot \frac{N_b}{k \cdot f} - \frac{\delta}{f} \tag{A.6}$$

[1]For more intuition on the departure pattern of bus users, see Section 4.1 of Kraus and Yoshida (2002).

The first term on the right-hand side of Eq. (A.6) is exactly equal to the user time costs of our continuous model (see Eq. (2.23)) so that the only difference is the second term (δ/f) that decreases with frequency. Therefore, our continuous approach overestimates the equilibrium costs for bus users, conditional on the frequency, capacity and demand, and this overestimation decreases with frequency.

We now turn to the car dynamic equilibrium. As we model the BRT system as dedicating a fraction ϕ of the capacity for buses and $1-\phi$ for cars, the analysis for car users, given that they have to go through a bottleneck of capacity $(1-\phi)\cdot s$, does not change. The equilibrium costs are the same as derived in Section 2.5 and the same as in any simple bottleneck model. Denoting t_c^s and t_c^{e} the first and last departure by car users, the conditions that define the equilibrium are that the equalization of schedule delay costs for the first and last user and that the length of the car peak must be such that the N_c car drivers pass through the bottleneck:

$$\beta(t^* - t_c^s) = \gamma(t_c^{\mathrm{e}} - t^*) \tag{A.7}$$

$$(t_c^{\mathrm{e}} - t_c^s)\cdot(1-\phi)\cdot s = N_c \tag{A.8}$$

Combining these two equations we can write the equilibrium user time costs, conditional on N_c:

$$UTC_c = \delta \cdot \frac{N_c}{(1-\phi)\cdot s} \tag{A.9}$$

The remaining step is to obtain the equilibrium modal split and compare equilibrium total user costs of this model with ours. The conditions that define the modal split equilibrium are equalization of costs across modes and that total demand is the sum of the demand of each mode. Adding the prices (p_m) and resource costs (r_m) for each mode m, and using Eqs. (A.6) and (A.9), the conditions can be written as:

$$p_c + r_c + \delta\cdot\frac{N_c}{(1-\phi)\cdot s} = p_b + r_b + \delta\cdot\frac{N_b}{k\cdot f} - \frac{\delta}{f} \tag{A.10}$$

$$N_b + N_c = N \tag{A.11}$$

Solving this system of equations, we obtain the equilibrium modal split:

$$N_b = \frac{k\cdot f\cdot((\Delta R + \Delta p)\cdot(1-\phi)\cdot s + \delta\cdot N) + \delta\cdot k\cdot(1-\phi)\cdot s}{\delta\cdot((1-\phi)\cdot s + k\cdot f)} \tag{A.12}$$

$$N_c = N - N_b \tag{A.13}$$

To simplify the comparison with the continuous model and to make the exposition clearer, we rewrite the modal splits to relate them with the equilibrium modal split of the continuous model. Letting "int" be the index that denotes the intermittent model and "cont" the continuous, and assuming that the same fraction ϕ of the capacity is dedicated to buses in the continuous model, we obtain:

$$N_b^{\mathrm{int}} = N_b^{\mathrm{cont}} + \frac{k\cdot(1-\phi)\cdot s}{(1-\phi)\cdot s + k\cdot f} \tag{A.14}$$

$$N_c^{\mathrm{int}} = N_c^{\mathrm{cont}} - \frac{k\cdot(1-\phi)\cdot s}{(1-\phi)\cdot s + k\cdot f} \tag{A.15}$$

Using these expressions we can obtain the equilibrium total user cost in the intermittent model, $UC_{\text{int}} = N_c \cdot (UTC_c + r_c) + N_b \cdot (UTC_b + r_b)$:

$$UC_{\text{int}} = \delta \frac{k \cdot (1-\phi)\, s \cdot \frac{(\Delta p + \Delta R)}{\delta} - N \cdot k}{k \cdot f + (1-\phi)\, s} + \delta \frac{k \cdot f \cdot (1-\phi)\, s \cdot (\frac{\Delta p + \Delta R}{\delta})^2 + N^2}{k \cdot f + (1-\phi)\, s} \tag{A.16}$$
$$+ R_b \cdot N + \Delta R \cdot \left(N_c^{\text{cont}} - \frac{k \cdot (1-\phi) \cdot s}{(1-\phi) \cdot s + k \cdot f} \right)$$

Again the user equilibrium costs are closely related to the user equilibrium costs of the continuous approximation. To see this, we substitute the equilibrium values of the modal split in the continuous model, which can be obtained by solving Eq. (2.23), into the expression for the total user cost of Eq. (2.24). This gives the following expression for the equilibrium total user cost in the continuous model, UC_{cont}:

$$UC_{cont} = \delta \frac{N^2 + \left(\frac{\Delta p + \Delta R}{\delta}\right)^2 ((1-\phi) \cdot s \cdot k \cdot f)}{(1-\phi) \cdot sf + k \cdot f} + \Delta R \cdot N_c^{\text{cont}} + R_b \cdot N \tag{A.17}$$

Combining Eqs. (A.17) and (A.16) we obtain an expression for the equilibrium total user cost in the intermittent model as a function of UC_{cont}:

$$UC_{\text{int}} = UC_{\text{cont}} + \frac{k \cdot (1-\phi)\, s \cdot \Delta p - N \cdot k \cdot \delta}{(1-\phi)\, s + k \cdot f} \tag{A.18}$$

These two expressions allows us to obtain the aggregate overestimation of user costs of our continuous model as a percentage of the total user cost in the intermittent model, $\Omega \equiv \frac{UC_{cont} - UC_{int}}{UC_{int}}$:

$$\Omega = \frac{\delta \cdot k \cdot (\delta \cdot N - \Delta p((1-\phi) \cdot sf))}{\Delta p \cdot k \cdot ((1-\phi)\, sf) \cdot [f \cdot (\Delta p + \Delta R) + \delta] + \delta \cdot N \cdot [\delta(N-k) + r_c(s - \lambda f) + r_b f k]} \tag{A.19}$$

While the expression is not very informative, what is important to note is that Ω approaches zero as the frequency increases[2]. Therefore, we have proven that when a BRT is in place, there exists an equilibrium if the service is intermittent and that the equilibrium user costs of our continuous model slightly overestimate this cost. This overestimation, using the parameters of our numerical example, is between 0.2% and 0.6%.

We now turn to the case of mixed traffic. We showed that the two approaches of modeling public transport were very similar in terms of bus user's time cost. The same is true under mixed traffic as the waiting at bus stops occurs before buses and cars interact on the road, and the potential difference is whether the equilibrium costs for car users are too different due to the intermittent nature of buses. We begin by proving that an equilibrium exists, and then show that the intermittent model approaches our continuous model as frequency increases.

[2]The numerator is a function of f and the denominator of f^2.

The conditions that are needed in equilibrium are the dynamic equilibrium conditions for the departures of car and bus users and the modal split condition. In other words, equalization of costs across departure times and modes. We assume that the departure rates for car and bus users that are able to make the cost constant over departure times exist and analyze the equilibrium costs. We show that the solution in mixed traffic is analogous to the solution derived above for a BRT system and then prove that the equilibrium rates exist. Therefore, we conclude that also in the case of mixed traffic the continuous approximation slightly overestimates user costs and that the overestimation approaches zero as frequency increases.

We focus on the relevant case of our analysis: when the condition in Proposition 1 holds, i.e., $0 < \Delta p + \Delta R < \delta N/(kf)$, and, thus, there is a unique interior equilibrium in mixed traffic conditions in which both modes are used.

By Lemma 1, which holds in this case, the bus peak period starts earlier and ends later than the car peak hour. This is, $t_b^s < t_c^s < t_c^e < t_b^e$. For car users, just as under a BRT, it must be true that, conditional on N_c, the start and end of the car peak period are given by equating the schedule delay costs for the first and last departure, as in equation (2.3). On the other hand, the length of the car peak must be such that all car drivers pass through the bottleneck, as in equation (2.4). Eq. (2.4) holds because when buses are intermittent the bus frequency is assumed to be constant and all buses move with a fixed speed. Therefore, the arrival frequency of buses to the road bottleneck equals f and, as each bus is λ PCU, the capacity used by buses is λf. Therefore, conditional on N_c, equilibrium time costs for car users are given by Eq. (2.11):

$$UTC_c = \delta \frac{N_c}{s - \lambda \cdot f} \tag{A.20}$$

For bus users, the highest schedule delay cost is incurred by either the passengers on the first run or the passengers on the last run. To minimize the schedule delays, conditional on the number of runs, the planner sets the times such that the first and last bus departure experience the same scheduling costs. This is:

$$\beta(t^* - t_b^s) = \gamma(t_b^e - t^*) \tag{A.21}$$

Under mixed traffic, by Lemma 1, the first and last bus user do not face road congestion, so passengers in the first and last departure cannot incur any waiting in equilibrium. As R runs operate in the period $[t_b^s, t_b^e]$ at a constant headway of $h = 1/f$, the conditions in Eqs. (A.4) and (A.5) hold and, therefore, bus user time cost, conditional on N_b, are given by (A.6):

$$UTC_b = \delta \cdot \frac{N_b}{k \cdot f} - \frac{\delta}{f} \tag{A.22}$$

Thus, the modal split given by equating the full price of each mode, i.e., $p_c + r_c + UTC_c = p_b + r_b + UTC_b$ is:

$$N_b = \frac{k \cdot f \cdot ((\Delta R + \Delta p) \cdot (s - \lambda f) + \delta \cdot N) + \delta \cdot k \cdot (s - \lambda f)}{\delta \cdot ((s - \lambda f) + k \cdot f)} \tag{A.23}$$

$$N_c = N - N_b \tag{A.24}$$

Therefore, just as in our continuous approximation, conditional on f, Δp and on dedicating exactly the capacity buses need under a BRT ($\phi = \lambda f$), the modal split is the same under mixed traffic than under BRT (see the intuition of this result in Section 5). As a consequence, the overestimation of our continuous approximation under mixed traffic is the same as the overestimation derived under a BRT system above. $\Omega \equiv \frac{UC_{cont}-UC_{int}}{UC_{int}}$ is given by Eq. (A.19), which approaches zero as the frequency increases.

The only remaining step to prove that there exists an equilibrium under mixed traffic with intermittent bus operation is to prove that there exist departure rates for car and bus users that can make time costs constant over departure times.

We first prove this for car users. As the bus frequency is assumed to be constant and all buses move with a fixed speed, the arrival frequency of buses to the road bottleneck equals f. Letting $\mathrm{d}_c(t)$ be the departure rate from home of car users, a queue develops if $\mathrm{d}_c(t) + \lambda f$ exceeds the bottleneck capacity s. Let $\hat{t_c}$ denote the most recent time at which there was no queue at the road bottleneck, then the queue length (in PCU) at time t, $V(t)$, is:

$$V(t) = \int_{\hat{t_c}}^{t} (\mathrm{d}_c(x) + \lambda f)\mathrm{d}x - s \cdot (t - \hat{t_c}) \tag{A.25}$$

Note that this is the same as in Eq. (A.32) and also matches the mixed traffic analysis of Huang et al. (2007). As the cost of a departure by car is given by Eq. (2.1), the analysis in Eqs. (A.31)-(A.33) hold and we obtain the same equilibrium departure rate for car users as in Proposition 1 (Eq. (A.26)):

$$\mathrm{d}_c(t) = \begin{cases} \frac{\alpha s}{\alpha-\beta} - \lambda f & \text{if } t \leq \tilde{t_c} \\ \frac{\alpha s}{\alpha+\gamma} - \lambda f & \text{if } t > \tilde{t_c} \end{cases} \tag{A.26}$$

The result is intuitive and it is the same as in our continuous model: the sum of the departure rates by cars, $\mathrm{d}_c(t)$, and buses, λf, matches the departure rate of the classic bottleneck model, which makes road congestion to exactly compensate changes in schedule delay costs across departure times.

The final step is to derive the equilibrium departure rate for bus users. As the buses are intermittent, to satisfy the dynamic user equilibrium condition, cost must be the same for all passengers on a particular run. For this to happen, they must all arrive at the origin stop *en masse*, as otherwise some would wait more than others. Also, they need to be the same for all runs. This can only be possible if the sum of waiting cost, the schedule delay cost and the road queuing costs is the same for all commuters. Passengers in the first and last departure do not incur any waiting (they arrive *en masse* just at t_b^s and t_b^e). All the other passengers incur waiting by arriving at the bus stop strictly earlier than the departure times of the bus that they board, so that everyone has the same user cost. The passengers in the runs that face road congestion (i.e., arrive to the bottleneck between t_c^s and t_c^e) must experience the same waiting cost as the sum of road congestion and schedule delay cost is already constant over time (which is necessary for car users to be in equilibrium). For the remaining passengers, who do not face road congestion and are not on the first and last run, the waiting time has to be longer for those who arrive closer to t^*.

In summary, the departures by bus users are also intermittent and such that: (i) passengers on the first and last run do not face waiting costs, (ii) passengers in runs that do not face road congestion depart such that the wait is longer the closer to t^* they arrive, and, (iii) passengers in runs that face road congestion arrive such that the waiting time is the same for all those runs. This is analogous to our continuous approximation depicted in Figure 2.

We have proven that under mixed traffic conditions, there exists an equilibrium if the service is intermittent and that the equilibrium user costs of our continuous model slightly overestimate this cost. This overestimation, given by Ω in Eq. (A.19), is between 0.2% and 0.3% according to our numerical examples.

A.2 Proof of Proposition 2.2

If $p_c + r_c > p_b + r_b$, by Lemma 2.1, the bus peak period starts earlier and ends later than the car peak period.

We now prove that there is a unique equilibrium in which both modes are used by contradiction. We first show that equilibria with all individuals traveling by either car or buses do not exist, and then show that there exists only one interior equilibrium.

Suppose that there is an equilibrium in which all individuals travel by bus. As we model public transport as a bottleneck of capacity kf, the usual equilibrium conditions hold: the first and last departure face no queuing, their schedule delay cost must be the same, and the period of operation must be such that all commuters can actually go through the bottleneck at the bus stop. These conditions are given by Eqs. (2.5) and (2.6) but considering that $N_b = N$. This implies that the user time cost in this candidate equilibrium is $\delta N/(kf)$ and the full price of a departure at any time is $p_b + r_b + \delta N/(kf)$. Consider a deviation from a bus user to departing by car at t^*. As the bus frequency f is not enough to create road congestion, the time costs of the deviating user would be zero and she would experience a full price of $p_c + r_c$. This is profitable if and only if $p_c + r_c < p_b + r_b + \delta N/(k \cdot f)$, which is exactly the condition stated in Proposition 2.2. Therefore, we have proven that an equilibrium with no car trips does not exist when $p_c + r_c < p_b + r_b + \delta N/(k \cdot f)$.

We now prove that an equilibrium with no bus trips does not exist either. Consider now that all users travel by car. This case is exactly the same regarding equilibrium time costs as a simple bottleneck model with capacity $s - \lambda f$. Therefore, for it to be an equilibrium, the conditions in Eqs. (2.3) and (2.4) must hold for $N_c = N$. Thus, user time costs are in this case equal to $\delta N/(s - \lambda f)$. Consider now a deviation from a car user that departs at time t to departing by bus at the same time. As road congestion exists and buses share capacity with cars, the bus user experiences exactly the same road congestion and schedule delay costs. But as the time-invariant full price of a car is higher than of the bus (the condition $p_c + r_c > p_b + r_b$ in the Proposition), the deviation would be profitable. Therefore, an equilibrium with all users traveling by car does not exist when $p_c + r_c > p_b + r_b$.

The remaining step is to show that there exist a unique equilibrium in which both modes are used. If there exists an equilibrium in which both modes are used, as we show in Section 2.2, it must satisfy conditions in Eqs. (2.3)-(2.12). The modal split that satisfies these equilibrium conditions is given by Eqs. (2.13) and (2.14). The solution is unique and the only remaining step is to show that both N_c and N_b are positive. Recall from Eqs. (2.13)-(2.14) that:

$$N_c = (s - \lambda \cdot f) \cdot \frac{N - \frac{(p_c + r_c - p_b - r_b) \cdot k \cdot f}{\delta}}{s - \lambda \cdot f + k \cdot f} \tag{A.27}$$

$$N_b = k \cdot f \cdot \frac{N + \frac{(p_c + r_c - p_b - r_b) \cdot (s - \lambda \cdot f)}{\delta}}{s - \lambda \cdot f + k \cdot f} \tag{A.28}$$

It follows directly from Eqs. (A.27) and (A.28) that N_c is positive if and only if $N - \frac{(p_c+r_c-p_b-r_b)\cdot k \cdot f}{\delta} > 0$ and N_b is positive if $(p_c + r_c - p_b - r_b) > 0$. As the conditions stated in the Proposition are that $0 < p_c + r_c - p_b - r_b < \delta N/(k \cdot f)$ it follows that there exists a unique equilibrium in which both modes are used.

The last part of the proof is to prove that the equilibrium departing patterns are such that:

(i) During the car peak hour, a queue at the bottleneck on the road begins to develop at the moment of the first car departure and grows linearly for early arrivals and shrinks linearly for late arrivals.

(ii) A queue at the bus stops starts to develop until the moment of departure of the first bus user that faces road congestion. During the period in which buses and cars share the road capacity and there is road congestion, the length of the queue at the bus stop remains constant, and it begins to dissipate after the departure of the last bus user that faces road congestion.

And that those patterns are as in Figure 2.2.

To prove these two conditions consider the cost for a departure at time t for each mode in Eq. (2.1) and (2.2) replacing the corresponding arrival times:

$$c_c(t) = p_c + r_c + \alpha \cdot T_w(t) + \begin{cases} \beta \cdot (t^* - t - T_w(t)) & \text{if } t + T_w(t) \le t^* \\ \gamma \cdot (t + T_w(t) - t^*) & \text{if } t + T_w(t) > t^* \end{cases} \tag{A.29}$$

$$\begin{gathered} c_b(t) = p_b + r_b + \alpha_2 \cdot T_q(t) + \alpha \cdot T_w\left(t + T_q(t)\right) + \\ \begin{cases} \beta \cdot (t^* - t - T_q(t) - T_w(t + T_q(t))) & \text{if } t + T_q(t) + T_w(t + T_q(t)) \le t^* \\ \gamma \cdot (t + T_q(t) + T_w(t + T_q(t)) - t^*) & \text{if } t + T_q(t) + T_w(t + T_q(t)) > t^* \end{cases} \end{gathered} \tag{A.30}$$

It must be true that the time-derivative of these two expressions is zero, so that costs are constant over departure times. We begin deriving the equilibrium car departure rate and prove (i). Denote $\tilde{t_c}$ the departure time for an on-time arrival by car. Differentiating Eq. (A.29) and equating to zero we obtain:

$$\frac{\partial T_w(t)}{\partial t} = \begin{cases} \frac{\beta}{\alpha - \beta} & \text{if } t \le \tilde{t_c} \\ \frac{\gamma}{\alpha + \gamma} & \text{if } t > \tilde{t_c} \end{cases} \tag{A.31}$$

which is the same result as in the classic bottleneck model of Arnott et al. (1993). It states that queuing delays must exactly compensate the changes in schedule delay costs. The difference lies in how a queue develops in our model of mixed traffic. As the road bottleneck capacity is shared by both modes, it is the combined arrival rate of cars and buses that matters. Letting $\mathrm{d}_c(t)$ be the departure rate from home of car users, a queue develops if $\mathrm{d}_c(t) + \lambda f$ exceeds the bottleneck capacity s. Let $\hat{t_c}$ denote the most recent time at which there was no queue at the road bottleneck, then the queue length (in PCU) at time t, $V(t)$, is:

$$V(t) = \int_{\hat{t_c}}^{t} (\mathrm{d}_c(x) + \lambda f)\mathrm{d}x - s \cdot (t - \hat{t_c}) \tag{A.32}$$

Finally, an individual's queuing time is simply the length of the queue divided by the capacity of the bottleneck:

$$T_w(t) = \frac{V(t)}{s} \tag{A.33}$$

Combining Eqs. (A.31)-(A.33) we obtain the equilibrium departure rate for car users:

$$\mathrm{d}_c(t) = \begin{cases} \frac{\alpha s}{\alpha - \beta} - \lambda f & \text{if } t \leq \tilde{t_c} \\ \frac{\alpha s}{\alpha + \gamma} - \lambda f & \text{if } t > \tilde{t_c} \end{cases} \tag{A.34}$$

The result is intuitive: the sum of the departure rates by cars, $\mathrm{d}_c(t)$, and buses, λf, matches the departure rate of the classic bottleneck model, which makes road congestion to exactly compensate changes in schedule delay costs across departure times. The rate d_c also matches the result of Huang et al. (2007), who model mixed traffic in a similar fashion as us but with intermittent departures.

We now proceed analogously to prove (ii). Denote $\tilde{t_b}$ the departure time for an on-time arrival by bus. Differentiating Eq. (A.30) and equating to zero we obtain:

$$\frac{\partial T_q(t)}{\partial t} = \begin{cases} \dfrac{\beta - \frac{\partial T_w}{\partial t}(\alpha - \beta)}{\alpha_2 - \beta + \frac{\partial T_w}{\partial t}(\alpha - \beta)} & \text{if } t \leq \tilde{t_b} \\ \\ \dfrac{-\gamma - \frac{\partial T_w}{\partial t}(\alpha + \gamma)}{\alpha_2 - \beta + \frac{\partial T_w}{\partial t}(\alpha + \gamma)} & \text{if } t > \tilde{t_b} \end{cases} \tag{A.35}$$

This indicates that the delays at the bus stop must exactly compensate the changes in schedule delay costs and road congestion delays. Note that if $\frac{\partial T_w}{\partial t}$ is zero, then the rate is analogous to a classic bottleneck model. Letting $\mathrm{d}_b(t)$ be the departure rate from home of bus users, a queue develops at the bus stop if $\mathrm{d}_b(t)$ exceeds the capacity kf. Let $\hat{t_b}$ denote the most recent time at which there was no queue at the bus stop, then the queue length (in passengers) at time t, $V_b(t)$, is:

$$V_b(t) = \int_{\hat{t_b}}^{t} \mathrm{d}_b(x)\mathrm{d}x - kf \cdot (t - \hat{t_b}) \tag{A.36}$$

Finally, an individual's queuing time at the bus stop is simply the length of the queue divided by the capacity of the bus stop:

$$T_q(t) = \frac{V_b(t)}{kf} \tag{A.37}$$

which is analogous to the classic bottleneck model, but with capacity kf. Note that the the equilibrium departure rate for bus users depends on the road congestion pattern. Denote $\tilde{t}_b^s$ the departure time for a bus user that arrives at t_c^s, so that she is the first to face road congestion, and $\tilde{t}_b^e$ the departure time for a bus user that arrives at t_c^e, so that she is the last to face road congestion.

Replacing Eqs. (A.36) and (A.37) into Eq. (A.35) we obtain the equilibrium departure rate for bus users:

$$\mathrm{d}_b(t) = \begin{cases} \frac{\alpha_2 kf}{\alpha_2 - \beta} \text{ if } t \leq \tilde{t}_b^s \\ kf \text{ if } \tilde{t}_b^s < t \leq \tilde{t}_b^e \\ \frac{\alpha_2 kf}{\alpha_2 + \gamma} \text{ if } \tilde{t}_b^e < t \end{cases} \tag{A.38}$$

The result is intuitive. At times when there is no road congestion and users arrive early, the departure rates of bus users must be such that the queuing at bus stops grows such that it compensates the reductions in schedule delay early. The same must be true for late arrivals that do not face road congestion, bus stops delays must decrease as schedule delay late increases. This is why the rates at $t \leq \tilde{t}_b^s$ and $t > \tilde{t}_b^e$ are fully analogous to the classic bottleneck model. On the other hand, at times when road congestion is compensating the changes in schedule delay costs (a condition that is necessary for car users to be in equilibrium), bus stops delays must be constant over time and the departure rate is exactly equal to the capacity kf.

A.3 Proof of Propostion 2.3

To prove that when $p_c + r_c < p_b + r_b$ there is a unique equilibrium in which all individuals travel by car, we start by this candidate equilibrium and show that there is no gainful deviation. Suppose then that $N_c = N$ and that buses operate at frequency f. The candidate equilibrium is analogous to the classic bottleneck model with equilibrium user time costs equal to $\delta N/(s - \lambda f)$[3], and the equilibrium full price of traveling by car is therefore $r_c + p_c + \delta N/(s - \lambda f)$.

Consider a deviation to a departure at any time t by bus. As buses face the same congestion as cars, if this departure occurs at a time in which there is road congestion, the deviating user will face the same time costs. The result will be a full price of $r_b + p_b + \delta N/(s - \lambda f)$. As $p_c + r_c < p_b + r_b$, the deviation increases the user costs. Consider now that the deviation is to a time in which there is no queuing, i.e. before or after there is road congestion. As time costs in the candidate equilibrium are constant and equal to the schedule delay cost of the first and last departure, schedule delay costs before or after are higher than the candidate equilibrium time costs. This, together with the fact that $p_c + r_c < p_b + r_b$, makes the deviation not gainful and the candidate equilibrium to be the unique equilibrium.

[3]See Eqs. (2.3) and (2.4) and the discussion below those equations for the derivation of the equilibrium costs. The derivation of the departure pattern is the same as the rate derived in Appendix A.2.

A.4 Proof of Proposition 2.3

To prove that when the time-invariant full price of buses and cars are equal (i.e., $p_c + r_c = p_b + r_b$) there are multiple equilibria, we consider a candidate equilibrium with an arbitrary number of car users N_c and show that the candidate equilibrium is indeed an equilibrium for multiple values of N_c. We begin by proving that $N_c = 0$ cannot be an equilibrium, we then prove that $N_c = N$ is an equilibrium and finally that there is a threshold $\bar{N}_c$ with $0 < \bar{N}_c < N$ for which any $N_c \in [\bar{N}_c, N]$ is an equilibrium.

First, in the proof of Proposition 1 in Appendix A.2 we demonstrate that an equilibrium in which all individuals travel by bus does not exist if $p_c + r_c < p_b + r_b + \delta N/(kf)$. As $p_c + r_c = p_b + r_b$, the inequality holds and $N_c = 0$ cannot be an equilibrium.

Now consider a candidate equilibrium in which all individuals travel by car, i.e. $N_c = N$. In Appendix A.4 we showed that a deviation from this candidate is profitable if and only if $p_c + r_c > p_b + r_b$. Therefore, $N_c = N$ is an equilibrium when $p_c + r_c = p_b + r_b$. The intuition is simple and similar as the one provided for Proposition 2. When all individuals travel by car and costs are constant over time, road congestion is such that it exactly compensates changes in schedule delay costs. As car and buses face the same road congestion, a user that switches mode and takes a bus cannot decrease its travel time costs. As the time-invariant full prices are the same for both modes, the deviation leaves her indifferent. Thus, all individual traveling by car is a (weak) Nash equilibrium. This is also the intuition why there are multiple equilibria as long as bus users do not face delays at bus stops. We formalize this argument below.

Consider a candidate equilibrium in which both modes are used and denote N_c and N_b the number of car and bus users respectively, with $N_c + N_b = N$. Denote t_c^s the time of the first departure by car and t_c^e the last. As shown in Section 2.2.2, for this candidate to be an equilibrium, it must be true that the start and end of the car peak period are given by equating the schedule delay costs for the first and last departure and that all N_c users pass through the bottleneck in that period. This leads to user time costs for cars equal to $\delta N_c/(s - \lambda f)$. But time costs for car users can be constant over time only if there is road congestion that compensates changes in schedule delay costs and, as shown, in Appendix A.2, this can only occur when departure rates for car users are given by Eq. (A.34). Given this road congestion pattern of the candidate equilibrium, together with $p_c + r_c = p_b + r_b$, the time cost of a bus user that departs at $t \in [t_c^s; t_c^e]$ is exactly the same as a car user that departs at the same time if and only if she does not face delays at the bus stop. Furthermore, a bus user that departs earlier than t_c^s or later than t_c^e, faces higher costs as schedule delay costs at those times exceed the time costs in the period $[t_c^s; t_c^e]$. Therefore, as long as the departure rate of bus users does not exceed the capacity kf, all departures in $[t_c^s; t_c^e]$ face the same costs regardless of the mode.

It is clear then that starting from the candidate equilibrium in which all individuals travel by car, multiple equilibria exist in which a fraction of the users travel by bus at times in which car users depart. This requires that bus users do not face bus stop queuing delays, which can only occur if their departure rate is not higher than the capacity. It is also needed that car users depart in such a way that the start and end of the car peak period are given by equating the schedule delay costs for the first and last departure and that the departure rates of car users follow Eq. (A.34) such that the sum of road congestion and schedule delays is constant over time.

The equilibrium with the highest share of bus users and, thus, lower share of car users is given by a full utilization of the public transport capacity. This is, with a departure rate of bus users equal to kf and the car and bus period exactly overlapping. Denote the number of car and bus users of this case by $\bar{N}_c$ and $\bar{N}_b$ respectively. The equilibrium conditions are the same as those in equations (2.3)–(2.6) but with the same start and end times for both modes. The derivations are analogous and the equilibrium cost and modal share are the same but evaluated at $p_c + r_c = p_b + r_b$:

$$\bar{N}_c = (s - \lambda \cdot f) \cdot \frac{N}{s - \lambda \cdot f + k \cdot f} \tag{A.39}$$

$$\bar{N}_b = k \cdot f \cdot \frac{N}{s - \lambda \cdot f + k \cdot f} \tag{A.40}$$

$$c = p + \delta \frac{N}{s - \lambda \cdot f + k \cdot f} \tag{A.41}$$

where p is the time-invariant full price of any mode ($p = p_c + r_c = p_b + r_b$). Note that in this case, the period around t^* is the shortest possible, and thus this equilibrium induces the lower total schedule delay costs.

Therefore for any $N_c \in [\bar{N}_c; N]$ there is an equilibrium. The first and last departure by car are such that the following holds:

$$\beta(t^* - t_c^s) = \gamma(t_c^{\mathrm{e}} - t^*) \tag{A.42}$$

$$(t_c^{\mathrm{e}} - t_c^s) \cdot (s - \lambda \cdot f) = N_c \tag{A.43}$$

the departure rates are the ones in Eq. (A.34) and the remaining $N_b = N - N_c$ bus users depart in the period $[t_c^s; t_c^{\mathrm{e}}]$ at a rate lower than kf such that there are no bus stop queuing delays.

A.5 Calculations and results for Section 2.5

A.5.1 Proof of Proposition 2.5

As we show in Section 5, conditional on N_c and N_b, for a given frequency and time-invariant prices the equilibrium costs for each mode are those in Eq. (2.23):

$$c_c = p_c + r_c + \delta \frac{N_c}{s - \lambda \cdot f} \tag{A.44}$$

$$c_b = p_b + r_b + \delta \frac{N_b}{k \cdot f} \tag{A.45}$$

By equating the costs and using that $N_c + N_b = N$, we obtain the equilibrium demand for each mode:

$$N_b = \frac{k \cdot f \cdot ((\Delta R + \Delta p) \cdot (s - \lambda f) + \delta \cdot N)}{\delta \cdot ((s - \lambda f) + k \cdot f)} \tag{A.46}$$

$$N_c = \frac{(s - \lambda f) \cdot (\delta \cdot N - k \cdot f \cdot ((\Delta R + \Delta p))}{\delta \cdot ((s - \lambda f) + k \cdot f)} \tag{A.47}$$

where $\Delta R = r_c - r_b$ and $\Delta p = p_c - p_b$. From these expressions it follows that:

$$N_b > 0 \Longleftrightarrow \Delta R + \Delta p > \frac{-\delta \cdot N}{s - \lambda f} \tag{A.48}$$

$$N_c > 0 \Longleftrightarrow \Delta R + \Delta p < \frac{\delta \cdot N}{k \cdot f} \tag{A.49}$$

which proves that when $-\delta N/(s - \lambda \cdot f) < p_c + r_c - p_b - r_b < \delta N/(k \cdot f)$, there is a unique equilibrium in which both modes are used.

The departure patterns, as in the classic bottleneck of Arnott et al. (1993), must be such that the queue length evolves over the peak period such that the sum of schedule delay costs and queuing delay costs are constant over time. As schedule delay costs decrease linearly for early arrivals and increase linearly for late arrivals, queuing delays for each mode must increase for early arrivals and decrease for late arrivals. Thus, only the first and las user to depart do not face queuing delays. This proves that queuing delays at bus stops occur for the whole duration of the operation of the BRT system, while road congestion delays for cars occur for the whole duration of their peak period.

We now prove turn to prove results (i)–(iii) in Proposition 2.5. Let t_b^s and t_c^s the time of the first departure by bus and car, respectively. As both individuals face only schedule delay costs, the full prices of each departure are:

$$c_c(t_c^s) = p_c + r_c + \beta(t^* - t_c^s) \tag{A.50}$$

$$c_b(t_b^s) = p_b + r_b + \beta(t^* - t_b^s) \tag{A.51}$$

In equilibrium, the two full prices should be equal, so $p_c+r_c+\beta(t^*-t_c^s) = p_b+r_b+\beta(t^*-t_b^s)$ holds. From this we can write an expression for $t_c^s - t_b^s$:

$$t_c^s - t_b^s = \frac{p_c + r_c - p_b - r_b}{\beta} \tag{A.52}$$

Analogously, for the last arrival of each mode we obtain:

$$c_c(t_c^{\mathrm{e}}) = p_c + r_c + \gamma(t_c^{\mathrm{e}} - t^*) \tag{A.53}$$

$$c_b(t_b^{\mathrm{e}}) = p_b + r_b + \gamma(t_b^{\mathrm{e}} - t^*) \tag{A.54}$$

And as also these two full prices should be equal, $p_c + r_c + \gamma(t^* - t_c^s) = p_b + r_b + \beta(t^* - t_b^s)$ holds and we can write an expression for $t_b^{\mathrm{e}} - t_c^{\mathrm{e}}$:

$$t_b^{\mathrm{e}} - t_c^{\mathrm{e}} = \frac{p_c + r_c - p_b - r_b}{\gamma} \tag{A.55}$$

It follows directly from Eqs. (A.52) and (A.55) that:

(i) if $0 < p_c + r_c - p_b - r_b$, the bus peak period starts earlier and ends later than the car peak period.

(ii) if $p_c + r_c - p_b - r_b = 0$, the peak period of both modes are the same.

(iii) if $p_c + r_c - p_b - r_b < 0$, the bus peak period starts later and ends earlier than the car peak period.

A.5.2 Calculations for the optimal price difference and frequency with perfectly divisible capacity (BRT-DC)

The social cost function (SC) is given by the sum of operators costs (OC) and users costs (UC). Substituting the equilibrium values of N_c and N_b in (2.24) we obtain these costs as a function of the frequency and the price difference:

$$UC = \frac{\delta\left(\frac{k\cdot f\cdot(\Delta p+\Delta R)^2\cdot(s-\lambda\cdot f)}{\delta^2}+N^2\right)}{s-\lambda\cdot f+k\cdot f} + \frac{\Delta R\cdot(s-f\lambda)\left(N-\frac{k\cdot f\cdot(\Delta p+\Delta R)}{\delta}\right)}{(1-\phi)\,sf+k\cdot f} + N\cdot R_b \tag{A.56}$$

$$OC = c_1\cdot f\cdot T_0 + c_2\cdot f\frac{\left(\frac{(\Delta R+\Delta p)(s-f\lambda)}{\delta}+N\right)}{s-\lambda\cdot f+k\cdot f} \tag{A.57}$$

$$SC = UC + OC \tag{A.58}$$

The first order condition with respect to the price difference, $\partial SC/\partial\Delta P$, results in:

$$\frac{f\cdot((1-\phi)\,sf)\,(\mathrm{c}2+k(\Delta R+2\Delta p))}{\delta\cdot(s-\lambda\cdot f+k\cdot f)} = 0 \tag{A.59}$$

which yields the optimal price difference of Eq. (2.26):

$$\Delta p^* = -\frac{c_2+k\cdot\Delta R}{2\cdot k} \tag{A.60}$$

The first order condition with respect to frequency $\partial SC/\partial f$, evaluated at the optimal price difference, gives the expression for the optimal frequency:

$$\begin{aligned}
&\frac{k\,(4\cdot\delta\cdot a\,(c_1\cdot T_0\cdot(s-\lambda\cdot f+k\cdot f)^2+\delta N^2(\lambda-k)))}{4\cdot\delta\cdot k\cdot(s-\lambda\cdot f+k\cdot f)^2}\\
&+\frac{\left(\Delta R^2\cdot k\cdot(\lambda\cdot f^2(k-\lambda)+2\cdot\lambda\cdot f\cdot s-s^2)\right)}{4\cdot\delta\cdot k\cdot(s-\lambda\cdot f+k\cdot f)^2}\\
&-\frac{(4\delta\Delta Rk^2Ns)-c_2{}^2\,(\lambda\cdot f^2\cdot(\lambda-k)-2\cdot\lambda\cdot f\cdot s+s^2)}{4\cdot\delta\cdot k\cdot(s-\lambda\cdot f+k\cdot f)^2}\\
&+\frac{2c_2k\,(\Delta R\,(\lambda\cdot f^2\cdot(\lambda-k)-2\cdot\lambda\cdot fs+s^2)+2\delta\cdot N\cdot s)}{4\cdot\delta\cdot k\cdot(s-\lambda\cdot f+k\cdot f)^2} = 0
\end{aligned} \tag{A.61}$$

As Eq. (A.61) is quadratic, there are two possible solutions. They can be written as follows:

$$f_1^* = \frac{A-\sqrt{B}}{D} \tag{A.62}$$

$$f_2^* = \frac{A+\sqrt{B}}{D} \tag{A.63}$$

with:

$$A = -4 \cdot c_1 \cdot \delta \cdot k \cdot s \cdot T^0 \cdot (k-\lambda) - \lambda \cdot s(c_2 - \Delta R \cdot k)^2 \tag{A.64}$$

$$B = k \cdot \left(4c_1 \cdot \delta \cdot k \cdot T^0 \cdot (k-\lambda) + \lambda \cdot (c_2 - \Delta R k)^2\right)(c_2 \cdot s - \Delta R \cdot k \cdot s + 2\delta \cdot N \cdot (\lambda - k))^2 \tag{A.65}$$

$$D = (k-\lambda)\left(4c_1 \cdot \delta \cdot k \cdot T^0 \cdot (k-\lambda) + \lambda \cdot (c_2 - \Delta R \cdot k)^2\right) \tag{A.66}$$

It is straightforward to show that $D > 0$ and $A < 0$ hold, so the only positive solution is f_2^* of Eq. (A.63). Moreover, it follows that the optimal frequency depends linearly on N, as the last term in parenthesis on the right-hand side of the expression for B is quadratic and depends linearly on N. As frequency depends on $\sqrt{B}$, and this is the only dependency on N, f_2^* is linear in N.

A.5.3 Calculations for the optimal period of operation for buses with perfectly divisible capacity (BRT-DC)

The period of operation for buses is defined by:

$$N_b/(k \cdot f) \tag{A.67}$$

Substituting the optimal price difference of Eq. (A.60) and the optimal frequency of Eq. (A.63) in the expression for N_b of Eq. (2.14), we obtain:

$$N_b(f_2^*, \Delta p^*)/(k \cdot f_2^*) = \frac{\lambda \cdot c_2 - \Delta R \cdot k \cdot \lambda + \sqrt{k \cdot (4c_1 \cdot \delta \cdot k \cdot T_0(k-\lambda) + \lambda \cdot (c_2 - \Delta R \cdot k)^2)}}{2 \cdot \delta \cdot k \cdot (k-\lambda)} \tag{A.68}$$

Which does not depend on the total demand. This proves that the length of the bus operations hours does not depend on N.

A.5.4 Calculations for the case of indivisible capacity (BRT-IC)

For the calculations of the indivisible capacity case (BRT-IC) we follow the exact same steps as above, but assuming that the capacity for cars is $s_1 < s - \lambda f$. The costs become:

$$UC = \frac{\delta\left(\frac{k \cdot f \cdot (\Delta p + \Delta R)^2 \cdot s_1}{\delta^2} + N^2\right)}{s_1 + k \cdot f} + \frac{\Delta R \cdot s_1\left(N - \frac{k \cdot f \cdot (\Delta p + \Delta R)}{\delta}\right)}{s_1} + N \cdot R_b \tag{A.69}$$

$$OC = c_1 \cdot f \cdot T_0 + c_2 \cdot f \frac{\left(\frac{(\Delta R + \Delta p)(s - f\lambda)}{\delta} + N\right)}{s_1 + k \cdot f} \tag{A.70}$$

and the first order condition with respect to the price difference yields:

$$\Delta p^* = -\frac{c_2 + k \cdot \Delta R}{2 \cdot k} \tag{A.71}$$

The first order condition for the frequency, evaluated at the optimal price difference, gives one positive root:

$$f^* = \frac{-2c_1k^2s_1T_0\delta + (-c_2s_1 + k\Delta R s_1 + 2kN\delta)\sqrt{c_1k^3T_0\delta}}{2c_1k^3T_0\delta} \tag{A.72}$$

Substituting these optimal values in the demand for cars and buses we obtain:

$$N_c = \frac{s_1\left(c_2 \cdot k - k^2 \cdot \Delta R + 2\sqrt{c_1 \cdot k^3 \cdot T_0 \cdot \delta}\right)}{2k^2 \cdot \delta} \tag{A.73}$$

$$N_b = N - \frac{s_1\left(c_2 \cdot k - k^2 \cdot \Delta R + 2\sqrt{c_1 \cdot k^3 \cdot T_0 \cdot \delta}\right)}{2k^2 \cdot \delta} \tag{A.74}$$

which imply the following car and bus peak duration:

$$N_c/s_1 = \frac{\left(c_2 \cdot k - k^2 \cdot \Delta R + 2\sqrt{c_1 \cdot k^3 \cdot T_0 \cdot \delta}\right)}{2k^2 \cdot \delta} \tag{A.75}$$

$$N_b/(k \cdot f^*) = \sqrt{\frac{c_1 \cdot k \cdot T_0}{\delta}} \tag{A.76}$$

Both expressions do not two depend on the total number of passengers N. Therefore, we have proven that the length of the bus operations hours also does not depend on N in the case of imperfect indivisibility of capacity.

Appendix B

B.1 Proof of Proposition 3.1

If $p_c + r_c > p_b + r_b$, by Basso et al. (2019), the bus peak period starts earlier and ends later than the car peak period.

We now prove by contradiction that there is a unique equilibrium in which both modes are used. First, we show that equilibria with all individuals traveling either by car or bus do not exist, and then show that there exists only one interior equilibrium.

Let us suppose that there is an equilibrium in which all individuals travel by bus. As we model public transport as a bottleneck of capacity $k \cdot f_u$, the usual equilibrium conditions hold: the first and last departures face no queuing, their schedule delay cost must be the same, and the period of operation must be such that all commuters can actually go through the bottleneck at the bus stop. These conditions are given by:

$$\beta(t^* - t_b^s) = \gamma(t_b^{\mathrm{e}} - t^*) \tag{B.1}$$

$$(t_b^{\mathrm{e}} - t_b^s) \cdot k \cdot f_u = N \tag{B.2}$$

This implies that the user time cost in this candidate equilibrium is $\delta N/(k \cdot f_u)$ and the full price of a departure at any time is $p_b + r_b + \delta N/(k \cdot f_u)$. Consider a deviation from a bus user to departing by car at t^*. As the bus frequency f_u is not enough to create road congestion, the time costs of the deviating user would be zero and they would experience a full price of $p_c + r_c$. This is profitable if and only if $p_c + r_c < p_b + r_b + \delta N/(k \cdot f_u)$, which is exactly the condition stated in Proposition 1. Therefore, we have proven that an equilibrium with no car journeys does not exist when $p_c + r_c < p_b + r_b + \delta N/(kf)$.

We now prove that an equilibrium with no bus journeys also does not exist either. Consider now that all users travel by car. This case is exactly the same regarding equilibrium time costs as a simple bottleneck model with capacity $s - \lambda \cdot f_j$. Therefore, for it to be an equilibrium, the conditions are:

$$\beta(t^* - t_c^s) = \gamma(t_c^{\mathrm{e}} - t^*) \tag{B.3}$$

$$(t_c^{\mathrm{e}} - t_c^s) \cdot (s - \lambda \cdot f_j) = N_c \tag{B.4}$$

Consider now a deviation from a car user that departs at time t to departing by bus at the same time. As road congestion exists and buses share capacity with cars, the bus user experiences exactly the same road congestion and schedule delay costs. But as the time-invariant full price of a car is higher than of the bus (the condition $p_c + r_c > p_b + r_b$ in the Proposition), the deviation would be profitable. Therefore, an equilibrium with all users traveling by car does not exist when $p_c + r_c > p_b + r_b$.

The remaining step is to show that there exists a unique equilibrium in which both modes are used. If an equilibrium exists in which both modes are used, as we show in Section 3.2, it must satisfy the conditions in Eqs. (3.8)-(3.20). The modal split that satisfies these equilibrium conditions is given by Eqs. (3.21) and (3.22). The solution is unique and the only remaining step is to show that both N_c and N_b are positive. Recall from Eqs. (3.21)-(3.22) that:

$$N_c = \frac{(s - \lambda \cdot f_j)\,(\delta \cdot N - k \cdot f_u \cdot ((p_c + r_c) - (p_b - r_b)))}{\delta(s - \lambda \cdot f_j + k \cdot f_j)} \tag{B.5}$$

$$N_b = k\frac{\delta \cdot f_j \cdot N + f_u \cdot ((p_c + r_c) - (p_b + r_b))\,(s - \lambda \cdot f_j)}{\delta(s - \lambda \cdot f_j + k \cdot f_j)} \tag{B.6}$$

Following directly from Eqs. (B.5) and (B.6) that N_c is positive if and only if $N - \frac{(p_c + r_c - p_b - r_b)\cdot k \cdot f_u}{\delta} > 0$ and N_b is positive if $(p_c + r_c - p_b - r_b) > 0$. As the conditions stated in the Proposition are that $0 < p_c + r_c - p_b - r_b < \delta N/(k \cdot f_u)$, as a consequence there exists a unique equilibrium in which both modes are used.

The last part is to prove that the equilibrium departing patterns are such that:

(i) During the car peak hour, a queue at the bottleneck on the road begins to develop at the moment of the first car departure and grows linearly for early arrivals and shrinks linearly for late arrivals.

(ii) A queue at the bus stops starts to up until the moment of departure of the first bus user that faces road congestion. During the period in which buses and cars share the road capacity and there is road congestion, the length of the queue at the bus stop remains constant and begins to dissipate after the departure of the last bus user that faces road congestion.

To prove these two conditions consider the cost for a departure at time t for each mode in Eq. (3.1) and (3.2) replacing the corresponding arrival times:

$$c_c(t) = p_c + r_c + \alpha \cdot T_w(t) + \begin{cases} \beta \cdot (t^* - t - T_w(t)) & \text{if } t + T_w(t) \leq t^* \\ \gamma \cdot (t + T_w(t) - t^*) & \text{if } t + T_w(t) > t^* \end{cases} \tag{B.7}$$

$$c_b(t) = p_b + r_b + \alpha_2 \cdot T_q(t) + \alpha \cdot T_w\left(t + T_q(t)\right) + \\ \begin{cases} \beta \cdot (t^* - t - T_q(t) - T_w(t + T_q(t))) & \text{if } t + T_q(t) + T_w(t + T_q(t)) \leq t^* \\ \gamma \cdot (t + T_q(t) + T_w(t + T_q(t)) - t^*) & \text{if } t + T_q(t) + T_w(t + T_q(t)) > t^* \end{cases} \tag{B.8}$$

It must be true that the time-derivative of these two expressions is zero, so that costs are constant over departure times. We begin deriving the equilibrium car departure rate and prove (i). Denote $\tilde{t_c}$ the departure time for an on-time arrival by car. Differentiating Eq. (B.7) and equating to zero we obtain:

$$\frac{\partial T_w(t)}{\partial t} = \begin{cases} \frac{\beta}{\alpha-\beta} & \text{if } t \leq \tilde{t_c} \\ \frac{\gamma}{\alpha+\gamma} & \text{if } t > \tilde{t_c} \end{cases} \tag{B.9}$$

which is the same result as in the classic bottleneck model of Arnott et al. (1993). It states that queuing delays must exactly compensate the changes in schedule delay costs. The difference lies in how a queue develops in our model of mixed traffic. As the road bottleneck capacity is shared by both modes, it is the combined arrival rate of cars and buses that matters. Letting $\mathrm{d}_c(t)$ be the departure rate from home of car users, a queue develops if $\mathrm{d}_c(t) + \lambda f$ exceeds the bottleneck capacity s. Let $\hat{t_c}$ denote the most recent time at which there was no queue at the road bottleneck, then the queue length (in PCU) at time t, $V(t)$, is:

$$V(t) = \int_{\hat{t_c}}^{t} (\mathrm{d}_c(x) + \lambda \cdot f(x))\mathrm{d}x - s \cdot (t - \hat{t_c}) \tag{B.10}$$

as $x \in [t_c^s, t_c^e]$ we know that $f(x) = f_j$.

Finally, an individual's queuing time is simply the length of the queue divided by the capacity of the bottleneck:

$$T_w(t) = \frac{V(t)}{s} \tag{B.11}$$

Combining Eqs. (B.9)-(B.11) we obtain the equilibrium departure rate for car users:

$$\mathrm{d}_c(t) = \begin{cases} \frac{\alpha s}{\alpha-\beta} - \lambda \cdot f_j & \text{if } t \leq \tilde{t_c} \\ \frac{\alpha s}{\alpha+\gamma} - \lambda \cdot f_j & \text{if } t > \tilde{t_c} \end{cases} \tag{B.12}$$

The result is intuitive: the sum of the departure rates by cars, $\mathrm{d}_c(t)$, and buses, λf, matches the departure rate of the classic bottleneck model, which makes road congestion to exactly compensate changes in schedule delay costs across departure times. The rate d_c also matches the result of Huang et al. (2007), who model mixed traffic in a similar approach as us, but with intermittent departures.

We now proceed analogously to prove (ii). Denote $\tilde{t_b}$ the departure time for an on-time arrival by bus. Differentiating Eq. (B.8) and equating to zero we obtain:

$$\frac{\partial T_q(t)}{\partial t} = \begin{cases} \dfrac{\beta - \frac{\partial T_w}{\partial t}(\alpha-\beta)}{\alpha_2 - \beta + \frac{\partial T_w}{\partial t}(\alpha-\beta)} & \text{if } t \leq \tilde{t_b} \\ \\ \dfrac{-\gamma - \frac{\partial T_w}{\partial t}(\alpha+\gamma)}{\alpha_2 - \beta + \frac{\partial T_w}{\partial t}(\alpha+\gamma)} & \text{if } t > \tilde{t_b} \end{cases} \tag{B.13}$$

This indicates that the delays at the bus stop must exactly compensate the changes in schedule delay costs and road congestion delays. Note that if $\frac{\partial T_w}{\partial t}$ is zero, then the rate is analogous to a classic bottleneck model. Letting $\mathrm{d}_b(t)$ be the departure rate from home of bus users, a queue develops at the bus stop if $\mathrm{d}_b(t)$ exceeds the capacity $k \cdot f_j$ or $k \cdot f_u$. Let $\hat{t_b}$ denote the most recent time at which there was no queue at the bus stop, then the queue length (in passengers) at time t, $V_b(t)$, is:

$$V_b(t) = \int_{\hat{t_b}}^{t} (\mathrm{d}_b(x) - k \cdot f(t)) \, \mathrm{d}x) \tag{B.14}$$

where $f(t)$ is defined in Eq. (3.3).

Finally, an individual's queuing time at the bus stop is simply the length of the queue divided by the capacity of the bus stop:

$$T_q(t) = \frac{V_b(t)}{k \cdot f(t)} \tag{B.15}$$

which is analogous to the classic bottleneck model, but with capacity $k \cdot f(t)$. Note that the equilibrium departure rate for bus users depends on the road congestion pattern. Denote $\tilde{t_b^s}$ the departure time for a bus user that arrives at t_c^s, so that they are the first to face road congestion, and $\tilde{t_b^{\mathrm{e}}}$ the departure time for a bus user that arrives at t_c^{e}, so that she is the last to face road congestion.

By replacing Eqs. (B.14) and (B.15) with Eq. (B.13) we obtain the equilibrium departure rate for bus users:

$$\mathrm{d}_b(t) = \begin{cases} \frac{\alpha_2 \cdot k \cdot f(t)}{\alpha_2 - \beta} & \text{if } \ t \leq \tilde{t_b^s} \\ k \cdot f(t) & \text{if } \ \tilde{t_b^s} < t \leq \tilde{t_b^{\mathrm{e}}} \\ \frac{\alpha_2 \cdot k \cdot f(t)}{\alpha_2 + \gamma} & \text{if } \ \tilde{t_b^{\mathrm{e}}} < t \end{cases} \tag{B.16}$$

The result described in Eq. (B.16) can be rewritten using the definition of $f(t)$.

$$\mathrm{d}_b(t) = \begin{cases} \frac{\alpha_2 \cdot k \cdot f_u}{\alpha_2 - \beta} & \text{if } \ t \leq \tilde{t_b^s} \\ k \cdot f_j & \text{if } \ \tilde{t_b^s} < t \leq \tilde{t_b^{\mathrm{e}}} \\ \frac{\alpha_2 \cdot k \cdot f_u}{\alpha_2 + \gamma} & \text{if } \ \tilde{t_b^{\mathrm{e}}} < t \end{cases} \tag{B.17}$$

The result is intuitive. At times when there is no road congestion and users arrive early, the departure rates of bus users must be such that queuing at bus stops increases in such a way that it compensates the reductions in early schedule delay. The same must be true for late arrivals that do not face road congestion, bus stops delays must decrease as late schedule delay increases. This is why the rates at $t \leq \tilde{t_b^s}$ and $t > \tilde{t_b^{\mathrm{e}}}$ are fully analogous to the classic bottleneck model. On the other hand, at times when road congestion is compensating the changes in schedule delay costs (a condition that is necessary for car users to be in equilibrium), bus stop delays must be constant over time and the departure rate is exactly equal to the capacity $k \cdot f_j$.

B.2 Proof of Propositions 3.2

We define the social cost (SC) directly as the sum of the total user cost (equation (3.31)) and public transport system expenditure (equations (3.35) and (3.36)).

$$UC = \frac{\delta \cdot N^2}{s - \lambda \cdot f_j + k \cdot f_j} + \frac{k \cdot (\Delta p + \Delta r)\left(\frac{(\Delta p+\Delta r)\cdot(s-\lambda \cdot f_j)\cdot f_u}{\delta} + (f_j - f_u)\cdot N\right)}{s - \lambda \cdot f_j + k \cdot f_j} + r_b \cdot N_b + r_c \cdot N_c \tag{B.18}$$

where N_c and N_b are equilibrium values from equations (3.21) and (3.22).

Public transport expenditure is as follows

$$E = E_1 + E_2 \tag{B.19}$$

where:

$$E_1 = c_1 \cdot \max\left\{ f_j \cdot \left(T^0 + z \cdot \frac{\delta}{\alpha} \frac{N - \frac{(\Delta p+\Delta r)\cdot k \cdot f_u}{\delta}}{s - \lambda \cdot f_j + k \cdot f_j} \right); f_u \cdot T^0 \right\} \tag{B.20}$$

$$E_2 = c_2 \cdot \left(\frac{\Delta p + \Delta r}{\delta} \cdot f_u + \frac{(\delta \cdot N - k \cdot f_u \cdot (\Delta p + \Delta r))}{\delta(s - \lambda \cdot f_j + k \cdot f_j)} \cdot f_j \right) \tag{B.21}$$

We consider that both modes have positive demand. It is sufficient that c_1 and c_2 satisfies:

$$c_2 < \max\left\{ k \cdot \frac{\delta \cdot N}{s} - c_1 \cdot \frac{T^0}{T_u} - k \cdot \Delta p; \frac{\delta \cdot N}{s} \cdot (k - \lambda) - k \cdot \Delta p \right\} \tag{B.22}$$

$$c_1 > \frac{N \cdot \delta \cdot \lambda}{s} \cdot \frac{T_u}{T^0} \tag{B.23}$$

We proceed to define the first-order condition for the minimization problem, due to the non-differentiability of the social cost function. In order to solve the optimization problem, it is necessary to consider two cases:

$$f_j \cdot \left(T^0 + z \cdot \frac{\delta}{\alpha} \frac{N - \frac{(\Delta p+\Delta r)\cdot k \cdot f_u}{\delta}}{s - \lambda \cdot f_j + k \cdot f_j} \right) \geq f_u \cdot T^0 \tag{B.24}$$

$$f_j \cdot \left(T^0 + z \cdot \frac{\delta}{\alpha} \frac{N - \frac{(\Delta p+\Delta r)\cdot k \cdot f_u}{\delta}}{s - \lambda \cdot f_j + k \cdot f_j} \right) \leq f_u \cdot T^0 \tag{B.25}$$

We have to assume one of the two above conditions. First,we assume that Eq. (B.24) holds and we calculate and analyze how f_u affects SC.

$$\frac{\partial SC}{\partial f_u} = \frac{(\Delta p + \Delta r)(\alpha \cdot c_2(s - \lambda \cdot f_p) - k(z \cdot c_1 \cdot \delta \cdot f_j + \alpha \cdot \Delta p \cdot f_j \cdot \lambda - \alpha \cdot \Delta p \cdot s + \alpha \cdot \delta \cdot N))}{\alpha \cdot \delta \cdot (s - \lambda \cdot f_j + k \cdot f_j)} \tag{B.26}$$

We assume that both modes have positive demand, then $k \cdot f_u \cdot (\Delta p + \Delta r)/\delta < N$, otherwise all demand would be satisfied during the *un-congested* period, contradicting that both modes have positive demand. SC is decreasing on f_u if condition (B.27) is satisfied.

$$c_2 \cdot (s - \lambda \cdot f_j) \leq z \cdot k \frac{c_1 \cdot f_j \cdot \delta}{\alpha} + k \cdot (\delta \cdot N - \Delta p \cdot (s - \lambda \cdot f_j)) \tag{B.27}$$

It is not hard to probe that the right hand of Cond. (B.22) implies Cond. (B.27). Then, TC is decreasing on f_u, then:

$$f_j \cdot \left(T^0 + z \cdot \frac{\delta}{\alpha} \frac{N - \frac{(\Delta p + \Delta r) \cdot k \cdot f_u}{\delta}}{s - \lambda \cdot f_j + k \cdot f_j} \right) = f_u \cdot T^0 \tag{B.28}$$

Now we have to analyze the case where Eq. (B.25) holds. For this, we calculate the first-order condition for f_j.

$$\frac{\partial SC}{\partial f_j} = -\frac{((\Delta p + \Delta r) \cdot k \cdot f_u - \delta \cdot N) \cdot (c_2 \cdot s + \Delta p \cdot k \cdot s - \delta \cdot N \cdot (k - \lambda))}{\delta \cdot (s - \lambda \cdot f_j + k \cdot f_j)^2} \tag{B.29}$$

It is clear that $\frac{\partial SC}{\partial f_j}$ does not change its sign, then the optimal solution will be a corner solution. As both modes have positive demand, we know that $k \cdot f_u \cdot (\Delta p + \Delta r)/\delta < N$, and it is our hypothesis that: $c_2 \cdot s + \Delta p \cdot k \cdot s - \delta \cdot N \cdot (k - \lambda) < 0$, hence $\frac{\partial SC}{\partial f_j} < 0$ implying that, TC is decreasing on f_j. Then, the optimal condition is:

$$f_j \cdot \left(T^0 + z \cdot \frac{\delta}{\alpha} \frac{N - \frac{(\Delta p + \Delta r) \cdot k \cdot f_u}{\delta}}{s - \lambda \cdot f_j + k \cdot f_j} \right) = f_u \cdot T^0 \tag{B.30}$$

A necessary condition, for having positive demand in both modes is that f_u has an upper boundary. It is necessary that $\partial SC/\partial f_u > 0$.

$$\frac{\partial SC}{\partial f_u} = \frac{(\Delta p + \Delta r) \cdot (c_2 \cdot (s - \lambda \cdot f_j) + k \cdot \Delta p \cdot (s - \lambda \cdot f_j) - \delta \cdot k \cdot N)}{\delta(s - \lambda \cdot f_j + k \cdot f_j)} + c_1 \cdot T^0 \tag{B.31}$$

Eq. (B.31) depends on f_j, then we study the shape of $\partial SC/\partial f_u > 0$ over f_j. It easy to show using Eq. (B.22) that $\partial^2 SC/\partial f_u \partial f_j > 0$.

$$\frac{\partial^2 SC}{\partial f_u \partial f_j} = -\frac{k \cdot (\Delta p + \Delta r) \left(c_2 \cdot s + k \cdot s \Delta p - \delta \cdot N \cdot (k - \lambda)\right)}{\delta(s - \lambda \cdot f_j + k \cdot f_j)^2} \tag{B.32}$$

As $\partial^2 SC/\partial f_u \partial f_j > 0$ then $\partial SC/\partial f_u$ is an increasing function over f_j, it means that if $\partial SC/\partial f_u$ evaluated in $f_j = 0$ is positive, then the function will be positive for all $f_j > 0$.

$$\frac{\partial SC}{\partial f_u}|_{f_j=0} = \frac{c_1 \cdot \delta \cdot s \cdot T^0 + (\Delta p + \Delta r)(c_2 \cdot s + \Delta p \cdot k \cdot s - \delta \cdot k \cdot N)}{\delta \cdot s} \tag{B.33}$$

Considering Eqs. (B.22) and (B.23), Eq. (B.33) is positive. This directly implies a necessary condition for the parameters, otherwise, f_u would be enough higher to satisfy all demand during the *uncongested* period contradicting the hypothesis that both modes have positive demand.

B.3 Proposition B.2

Proposition B.1 *If the fleet and dispatch costs satisfy the following condition*

$$c_2 < k \cdot \frac{\delta \cdot N}{s} - c_1 \cdot \frac{T^0}{T_u} - k \cdot \Delta p \tag{B.34}$$

$$c_1 < \frac{N \cdot \delta \cdot \lambda}{s} \cdot \frac{T_u}{T^0} \tag{B.35}$$

then, all demand is satisfied by public transportation.

PROOF. We define the social cost (SC) directly as Eq. (3.37).

$$\begin{aligned} SC =& N \cdot \frac{(p_c + r_c) \cdot (s - \lambda \cdot f_j + k \cdot f_j) - f_u \cdot k \cdot (\Delta p + \Delta r) + \delta \cdot N}{s - \lambda \cdot f_j + k \cdot f_j} \\ & - p_c \cdot \frac{(s - \lambda \cdot f_j)\,(\delta \cdot N - k \cdot f_u \cdot (\Delta p + \Delta r))}{\delta(s - \lambda \cdot f_j + k \cdot f_j)} \\ & - p_b \cdot k \cdot \frac{\delta \cdot f_j \cdot N + f_u \cdot ((\Delta p + \Delta r))\,(s - \lambda \cdot f_j)}{\delta(s - \lambda \cdot f_j + k \cdot f_j)} \\ & + c_1 \cdot \max\left\{ f_j \cdot \left(T^0 + z \cdot \frac{\delta}{\alpha} \frac{N - \frac{(\Delta p + \Delta r) \cdot k \cdot f_u}{\delta}}{s - \lambda \cdot f_j + k \cdot f_j} \right); f_u \cdot T^0 \right\} \\ & + c_2 \cdot \left(\frac{\Delta p + \Delta r}{\delta} \cdot f_u + \frac{(\delta \cdot N - k \cdot f_u \cdot (\Delta p + \Delta r))}{\delta(s - \lambda \cdot f_j + k \cdot f_j)} \cdot f_j \right) \end{aligned} \tag{B.36}$$

Due to the non-differentiability of the social cost function, it is necessary to consider two cases to solve the optimization problem:

$$f_j \cdot \left(T^0 + z \cdot \frac{\delta}{\alpha} \frac{N - \frac{(\Delta p + \Delta r) \cdot k \cdot f_u}{\delta}}{s - \lambda \cdot f_j + k \cdot f_j} \right) \geq f_u \cdot T^0 \tag{B.37}$$

$$f_j \cdot \left(T^0 + z \cdot \frac{\delta}{\alpha} \frac{N - \frac{(\Delta p + \Delta r) \cdot k \cdot f_u}{\delta}}{s - \lambda \cdot f_j + k \cdot f_j} \right) \leq f_u \cdot T^0 \tag{B.38}$$

We have to assume one of the two conditions above. First,we assume that Eq. (B.37) holds and we calculate and analyze how f_u affects SC.

$$\frac{\partial SC}{\partial f_u} = \frac{(\Delta p + \Delta r)(c_2(s - \lambda \cdot f_j) - k(z \cdot c_1 \cdot \delta \cdot f_j/\alpha + \Delta p \cdot f_j \cdot \lambda - \Delta p \cdot s + \delta \cdot N))}{\delta \cdot (s - \lambda \cdot f_j + k \cdot f_j)} \tag{B.39}$$

$$\frac{\partial^2 SC}{\partial f_u \partial f_j} = -\frac{(\Delta p + \Delta r) \cdot k \cdot (c_2 \cdot s + \Delta p \cdot k \cdot s + \delta \cdot c_1 \cdot s/\alpha - \delta \cdot N \cdot (k - \lambda))}{\delta \cdot (s - \lambda \cdot f_j + k \cdot f_j)} \tag{B.40}$$

If c_2 also satisfies that:

$$c_2 < \frac{\delta \cdot N}{s} \cdot (k - \lambda) - k \cdot \Delta p \tag{B.41}$$

then, it is easy to probe that $\partial SC/\partial f_u < 0$ then f_u is the maximum possible.

On the other hand if c_2 satisfies

$$\frac{\delta \cdot N}{s} \cdot (k-\lambda) - k \cdot \Delta p < c_2 < k \cdot \frac{\delta \cdot N}{s} - c_1 \cdot \frac{T^0}{T_u} - k \cdot \Delta p \tag{B.42}$$

It is easy to show that $\partial^2 SC/\partial f_u \partial f_j < 0$, we calculate

$$\frac{\partial SC}{\partial f_u}(0, f_u) = \frac{(\Delta p + \Delta r)(c_2 \cdot s + \Delta p \cdot k \cdot s - \delta \cdot k \cdot N)}{\delta \cdot s} \tag{B.43}$$

As $c_2 < k \cdot \frac{\delta \cdot N}{s} - c_1 \cdot \frac{T^0}{T_u} - k \cdot \Delta p < k \cdot \frac{\delta \cdot N}{s} - k \cdot \Delta p$, then $c_2 \cdot s + \Delta p \cdot k \cdot s - \delta \cdot k \cdot N < 0$ it implies that:

$$\frac{\partial SC}{\partial f_u}(0, f_u) = \frac{(\Delta p + \Delta r)(c_2 \cdot s + \Delta p \cdot k \cdot s - \delta \cdot k \cdot N)}{\delta \cdot s} < 0 \tag{B.44}$$

As $\partial^2 SC/\partial f_u \partial f_j < 0$ then Eq. (B.44) implies that f_u is the maximum possible. If we assume that

$$f_j \cdot \left(T^0 + z \cdot \frac{\delta}{\alpha} \frac{N - \frac{(\Delta p + \Delta r) \cdot k \cdot f_u}{\delta}}{s - \lambda \cdot f_j + k \cdot f_j} \right) \geq f_u \cdot T^0$$

then it is necessary that:

$$f_j \cdot \left(T^0 + z \cdot \frac{\delta}{\alpha} \frac{N - \frac{(\Delta p + \Delta r) \cdot k \cdot f_u}{\delta}}{s - \lambda \cdot f_j + k \cdot f_j} \right) = f_u \cdot T^0$$

Now, we assume that Eq. (B.38) holds. We calculate $\partial SC/\partial f_u$

$$\frac{\partial SC}{\partial f_u} = \frac{(\Delta p + \Delta r) \cdot (c_2 \cdot (s - \lambda \cdot f_j) + k \cdot \Delta p \cdot (s - \lambda \cdot f_j) - \delta \cdot k \cdot N) + c_1 \cdot T^0 \cdot \delta \cdot (s - \lambda \cdot f_j + k \cdot f_j)}{\delta \cdot (s - \lambda \cdot f_j + k \cdot f_j)} \tag{B.45}$$

$$A = c_2 \cdot (s - \lambda \cdot f_j) + k \cdot \Delta p \cdot (s - \lambda \cdot f_j) - \delta \cdot k \cdot N + \frac{c_1 \cdot T^0 \cdot (s - \lambda \cdot f_j + k \cdot f_j)}{T_u} \tag{B.46}$$

Reordering the terms we have

$$A = (c_2 + k \cdot \Delta p + c_1 \cdot \frac{T^0}{T_u}) \cdot (s - \lambda \cdot f_j) - \delta \cdot k \cdot N + c_1 \cdot \frac{T^0}{T_u} \cdot k \cdot f_j \tag{B.47}$$

First, we assume that

$$\frac{c_1 \cdot T^0}{T_u} < \frac{N \cdot \delta \cdot \lambda}{s} \tag{B.48}$$

Using Eq. (B.34) we modify the condition

$$A < k \cdot \frac{\delta \cdot N}{s} \cdot (s - \lambda \cdot f_j) - \delta \cdot k \cdot N + c_1 \cdot \frac{T^0}{T_u} \cdot k \cdot f_j = f_j \cdot k \cdot \left(\frac{c_1 \cdot T^0}{T_u} - \frac{N \cdot \delta \cdot \lambda}{s} \right) \tag{B.49}$$

We conclude that $\partial SC / \partial f_u < 0$ then f_u does not have an upper boundary.

□

Proposition B.2 *If the fleet and dispatch costs satisfy*

$$c_2 > \max \left\{ \frac{\delta \cdot N}{s} \cdot (k - \lambda) - k \cdot \Delta p; k \cdot \frac{\delta \cdot N}{s} - c_1 \cdot \frac{T^0}{T_u} - k \cdot \Delta p \right\} \tag{B.50}$$

then, the provision of public transport is not socially efficient and all demand is met by cars.

PROOF. We define the social cost (SC) directly as Eq. (3.37).

$$\begin{aligned} SC = & N \cdot \frac{(p_c + r_c) \cdot (s - \lambda \cdot f_j + k \cdot f_j) - f_u \cdot k \cdot (\Delta p + \Delta r) + \delta \cdot N}{s - \lambda \cdot f_j + k \cdot f_j} \\ & - p_c \cdot \frac{(s - \lambda \cdot f_j) \left(\delta \cdot N - k \cdot f_u \cdot (\Delta p + \Delta r) \right)}{\delta (s - \lambda \cdot f_j + k \cdot f_j)} \\ & - p_b \cdot k \cdot \frac{\delta \cdot f_j \cdot N + f_u \cdot \left((\Delta p + \Delta r) \right) (s - \lambda \cdot f_j)}{\delta (s - \lambda \cdot f_j + k \cdot f_j)} \\ & + c_1 \cdot \max \left\{ f_j \cdot \left(T^0 + z \cdot \frac{\delta}{\alpha} \frac{N - \frac{(\Delta p + \Delta r) \cdot k \cdot f_u}{\delta}}{s - \lambda \cdot f_j + k \cdot f_j} \right); f_u \cdot T^0 \right\} \\ & + c_2 \cdot \left(\frac{\Delta p + \Delta r}{\delta} \cdot f_u + \frac{\left(\delta \cdot N - k \cdot f_u \cdot (\Delta p + \Delta r) \right)}{\delta (s - \lambda \cdot f_j + k \cdot f_j)} \cdot f_j \right) \end{aligned} \tag{B.51}$$

Due to the non-differentiability of the social cost function, to solve the optimization problem, it is necessary to consider two cases:

$$f_j \cdot \left(T^0 + z \cdot \frac{\delta}{\alpha} \frac{N - \frac{(\Delta p + \Delta r) \cdot k \cdot f_u}{\delta}}{s - \lambda \cdot f_j + k \cdot f_j} \right) \geq f_u \cdot T^0 \tag{B.52}$$

$$f_j \cdot \left(T^0 + z \cdot \frac{\delta}{\alpha} \frac{N - \frac{(\Delta p + \Delta r) \cdot k \cdot f_u}{\delta}}{s - \lambda \cdot f_j + k \cdot f_j} \right) \leq f_u \cdot T^0 \tag{B.53}$$

We assume that $k \cdot f_u \cdot (\Delta p + \Delta r) / \delta < N$, because we suppose that f_u is not large enough to satisfy all demand. At the end we check that our assumption is correct.

We analyze the first-order condition assuming that Eq. (B.52) holds. Considering that $k \cdot f_u \cdot (\Delta p + \Delta r)/\delta < N$, then we calculate f_j using the first-order condition.

$$f_j = \frac{\sqrt{\cdot c_1 \cdot \delta \cdot T^0 \cdot (k \cdot f_u \cdot (\Delta p + \Delta r) - \delta \cdot N)(c_1 \cdot \delta \cdot s \cdot z/\alpha + c_2 \cdot s + \Delta p \cdot k \cdot s - \delta \cdot \mathrm{N} \cdot (k - \lambda))}}{\alpha \cdot c_1 \cdot \delta \cdot T^0 (k - \lambda)} - s \tag{B.54}$$

Eq. (B.54) requires for a real solution that:

$$A = c_1 \cdot \delta \cdot s \cdot z/\alpha + c_2 \cdot s + \Delta p \cdot k \cdot s - \delta \cdot \mathrm{N} \cdot (k - \lambda) < 0 \tag{B.55}$$

As Condition (B.50) holds and $c_1 \cdot \delta \cdot s \cdot z/\alpha > 0$, it is easy to show that $A > 0$, as there exists no real solution for the first-order condition, then optimal the solution is $f_j = 0$ or f_j that satisfies:

$$f_j \cdot \left(T^0 + z \cdot \frac{\delta}{\alpha} \frac{N - \frac{(\Delta p + \Delta r) \cdot k \cdot f_u}{\delta}}{s - \lambda \cdot f_j + k \cdot f_j} \right) = f_u \cdot T^0$$

On the other hand, we calculate:

$$\frac{\partial^2 SC}{\partial f_j^2} = 2 \frac{((\Delta p + \Delta r) \cdot k \cdot f_u - \delta \cdot N) \cdot (k - \lambda) \cdot (c_2 \cdot s \cdot \alpha + \Delta p \cdot k \cdot s \cdot \alpha - \delta \cdot N(k - \lambda) + \delta \cdot c_1 \cdot s)}{\alpha \cdot \delta \cdot (s - \lambda \cdot f_j + k \cdot f_j)^3} \tag{B.56}$$

$$\frac{\partial^2 SC}{\partial f_j \partial f_u} = - \frac{(\Delta p + \Delta r) \cdot k \cdot (c_2 \cdot s \cdot \alpha + \Delta p \cdot k \cdot s \cdot \alpha - \delta \cdot N(k - \lambda) + \delta \cdot c_1 \cdot s)}{\alpha \cdot \delta \cdot (s - \lambda \cdot f_j + k \cdot f_j)^2} \tag{B.57}$$

It is not difficult to notice that $\partial^2 SC/\partial f_j \partial f_u < 0$ and $\partial^2 SC/\partial f_j^2 < 0$. If we evaluate $\partial SC/\partial f_j$ in $(f_j = 0, f_u = 0)$ and $(f_j = 0, f_u = \delta \cdot N/(k \cdot (\Delta p + \Delta r)))$ we get:

$$\frac{\partial SC}{\partial f_j}(f_j = 0, f_u = 0) = c_1 \cdot \left(z \cdot \frac{\delta \cdot N}{\alpha \cdot s} + T^0 \right) + \frac{c_2 \cdot N}{s} + \frac{\Delta p \cdot k \cdot N}{s} + N \cdot \left(\frac{\delta \cdot \lambda \cdot N}{s^2} - \frac{\delta \cdot k \cdot N}{s^2} \right) \tag{B.58}$$

$$\frac{\partial SC}{\partial f_j} \left(f_j = 0, f_u = \frac{\delta \cdot N}{k \cdot (\Delta p + \Delta r)} \right) = c_1 \cdot T^0 \tag{B.59}$$

As $\partial^2 SC/\partial f_j^2 < 0$ and $\partial SC/\partial f_j > 0$, we conclude that $f_j^* = 0$

We obtain the expression for the first-order condition for f_u and use that $f_j^* = 0$.

$$\frac{\partial TC}{\partial f_u}(f_j^* = 0, f_u) = c_1 T^0 + (\Delta p + \Delta r) \left(\frac{(c_2 + \Delta p \cdot k)}{\delta} - \frac{k \cdot N}{s} \right) \tag{B.60}$$

Using Eq. (B.50), we conclude that $\frac{\partial TC}{\partial f_u}(f_j^* = 0, f_u) > 0$, implying that $f_u^* = 0$.

Now we analyze the first-order condition assuming that Eq. (B.53) holds. Considering that $k \cdot f_u \cdot (\Delta p + \Delta r)/\delta < N$, then we calculate the first-order condition for f_j.

$$\frac{\partial SC}{\partial f_j} = -\frac{(\Delta p + \Delta r) \cdot f_u \cdot k - \delta \cdot N) \cdot (c_2 \cdot s + \Delta p \cdot k \cdot s - \delta \cdot N(k - \lambda))}{\delta \cdot (s - \lambda \cdot f_j + k \cdot f_j)} \tag{B.61}$$

Using Condition (B.50), we show that $\partial SC / \partial f_j > 0$, then the optimal solution is $f_j = 0$. Now we calculate $\partial SC \partial f_u$ considering that $f_j = 0$.

$$\frac{\partial SC}{\partial f_u}(f_j = 0, f_u) = \frac{c_1 \cdot \delta \cdot s \cdot T^0 + (\Delta p + \Delta r)(c_2 \cdot s + \Delta p \cdot k \cdot s - \delta \cdot k \cdot N)}{\delta \cdot s} \tag{B.62}$$

Using Eq. (B.50) we conclude that $\partial SC \partial f_u > 0$, then $f_j = 0$. Then the optimal frequencies are $f_j^* = f_u^* = 0$.

$\square$

B.4 Numerical analysis results

Table B.1: Numerical Analysis

	N	f_j	f_u	p_b	N_c	N_b	t_b^s	t_c^s	t_c^e	t_b^e	TC	OC	UC	CC	SDC	QC
Mixed traffic	6000	0	0	-	6000	0	-	7:20	8:20	-	19800	0	19800	3900	3900	0
	6500	0	0	-	6500	0	-	7:16	8:21	-	22154	0	22154	4577	4577	0
	7000	0	0	-	7000	0	-	7:13	8:23	-	24617	0	24617	5308	5308	0
	7500	0	0	-	7500	0	-	7:10	8:25	-	27188	0	27188	6094	6094	0
	8000	0	0	-	8000	0	-	7:06	8:26	-	29867	0	29867	6933	6933	0
	8500	0	0	-	8500	0	-	7:03	8:28	-	32654	0	32654	7827	7827	0
	9000	0	0	-	9000	0	-	7:00	8:30	-	35550	0	35550	8775	8775	0
	9500	0	0	-	9500	0	-	6:56	8:31	-	38554	0	38554	9777	9777	0
	10000	0	0	-	10000	0	-	6:53	8:33	-	41667	0	41667	10833	10833	0
	10500	5	5	1.75	9797	703	6:46	6:54	8:32	8:36	44867	2748	42120	11099	11262	164
	11000	10	0	2.00	9677	1323	6:55	6:55	8:32	8:32	48089	5534	42555	11601	11601	0
	11500	18	0	2.00	9296	2204	6:57	6:57	8:31	8:31	51272	9279	41994	11701	11701	0
	12000	25	0	2.00	8976	3024	6:59	6:59	8:30	8:30	54436	12803	41632	11840	11840	0
	12500	32	0	2.00	8702	3798	7:00	7:00	8:29	8:29	57583	16161	41422	12009	12009	0
	13000	39	0	2.00	8464	4536	7:02	7:02	8:28	8:28	60716	19387	41329	12200	12200	0
BRT - DC	6000	13	0	1.81	4975	1025	7:20	7:26	8:16	8:19	19589	2929	16660	2298	2951	1460
	6500	20	0	1.81	4956	1544	7:20	7:26	8:16	8:19	21675	4413	17262	2127	3110	2114
	7000	26	0	1.81	4937	2063	7:20	7:26	8:16	8:19	23762	5898	17864	1977	3291	2723
	7500	33	0	1.81	4917	2583	7:20	7:26	8:16	8:19	25848	7382	18465	1844	3489	3297
	8000	40	0	1.81	4898	3102	7:20	7:26	8:16	8:19	27934	8867	19067	1727	3703	3842
	8500	46	0	1.81	4878	3622	7:20	7:26	8:16	8:19	30020	10351	19669	1622	3929	4361
	9000	53	0	1.81	4859	4141	7:20	7:26	8:16	8:19	32107	11836	20271	1528	4166	4859
	9500	59	0	1.81	4840	4660	7:20	7:26	8:16	8:19	34193	13320	20872	1443	4411	5339
	10000	66	0	1.81	4820	5180	7:20	7:26	8:16	8:19	36279	14805	21474	1366	4665	5803
	10500	73	0	1.81	4801	5699	7:20	7:26	8:16	8:19	38365	16289	22076	1296	4925	6253
	11000	79	0	1.81	4781	6219	7:20	7:26	8:16	8:19	40452	17774	22678	1231	5192	6692
	11500	86	0	1.81	4762	6738	7:20	7:26	8:16	8:19	42538	19258	23280	1172	5463	7120
	12000	93	0	1.81	4743	7257	7:20	7:26	8:16	8:19	44624	20743	23881	1118	5740	7539
	12500	99	0	1.81	4723	7777	7:20	7:26	8:16	8:19	46710	22227	24483	1067	6020	7949
	13000	106	0	1.81	4704	8296	7:20	7:26	8:16	8:19	48797	23712	25085	1020	6304	8353
BRT - IC	6000	35	0	1.81	3279	2721	7:21	7:27	8:16	8:19	21295	7831	13464	1748	3453	1705
	6500	42	0	1.81	3279	3221	7:21	7:27	8:16	8:19	23360	9270	14091	1748	3766	2018
	7000	48	0	1.81	3279	3721	7:21	7:27	8:16	8:19	25426	10709	14717	1748	4079	2332
	7500	55	0	1.81	3279	4221	7:21	7:27	8:16	8:19	27492	12148	15344	1748	4393	2645
	8000	61	0	1.81	3279	4721	7:21	7:27	8:16	8:19	29558	13587	15971	1748	4706	2958
	8500	68	0	1.81	3279	5221	7:21	7:27	8:16	8:19	31624	15026	16597	1748	5019	3272
	9000	74	0	1.81	3279	5721	7:21	7:27	8:16	8:19	33690	16466	17224	1748	5333	3585
	9500	81	0	1.81	3279	6221	7:21	7:27	8:16	8:19	35755	17905	17851	1748	5646	3898
	10000	87	0	1.81	3279	6721	7:21	7:27	8:16	8:19	37821	19344	18477	1748	5959	4212
	10500	94	0	1.81	3279	7221	7:21	7:27	8:16	8:19	39887	20783	19104	1748	6273	4525
	11000	100	0	1.81	3279	7721	7:21	7:27	8:16	8:19	41953	22222	19731	1748	6586	4838
	11500	107	0	1.81	3279	8221	7:21	7:27	8:16	8:19	44019	23661	20357	1748	6899	5152
	12000	113	0	1.81	3279	8721	7:21	7:27	8:16	8:19	46085	25101	20984	1748	7213	5465
	12500	120	0	1.81	3279	9221	7:21	7:27	8:16	8:19	48150	26540	21611	1748	7526	5778
	13000	126	0	1.81	3279	9721	7:21	7:27	8:16	8:19	50216	27979	22237	1748	7839	6092
Two-Frequncy	6000	0	0	-	6000	0	-	7:20	8:20	-	19800	0	19800	3900	3900	0
	6500	0	0	-	6500	0	-	7:16	8:21	-	22154	0	22154	4577	4577	0
	7000	0	0	-	7000	0	-	7:13	8:23	-	24617	0	24617	5308	5308	0
	7500	2	7	1.28	6986	514	6:51	7:13	8:23	8:34	27181	1472	25709	5464	6010	263
	8000	12	30	1.24	5654	2346	6:58	7:22	8:18	8:30	29679	6719	22960	4070	6356	1226
	8500	22	48	1.21	4768	3732	7:03	7:27	8:16	8:28	32080	10690	21390	3225	6650	1978
	9000	30	62	1.18	4129	4871	7:06	7:31	8:14	8:26	34418	13956	20462	2658	6938	2607
	9500	39	75	1.15	3643	5857	7:09	7:35	8:12	8:25	36710	16784	19926	2253	7230	3156
	10000	47	87	1.13	3260	6740	7:10	7:37	8:11	8:24	38968	19320	19648	1948	7531	3649
	10500	55	97	1.11	2948	7552	7:12	7:39	8:10	8:23	41199	21650	19549	1711	7841	4101
	11000	63	107	1.09	2690	8310	7:13	7:41	8:09	8:23	43409	23829	19581	1522	8159	4521
	11500	71	116	1.08	2471	9029	7:14	7:42	8:08	8:22	45603	25893	19710	1367	8483	4918
	12000	78	125	1.06	2283	9717	7:15	7:44	8:07	8:22	47782	27868	19914	1238	8815	5294
	12500	86	134	1.05	2121	10379	7:15	7:45	8:07	8:22	49950	29772	20177	1130	9151	5655
	13000	93	142	1.03	1978	11022	7:16	7:46	8:06	8:21	52107	31619	20488	1037	9492	6002