Transportation Research Part B 127 (2019) 47–71

Contents lists available at ScienceDirect

Transportation Research Part B

journal homepage: www.elsevier.com/locate/trb

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

Leonardo J. Basso[a], Fernando Feres[b], Hugo E. Silva[c,d,*]

[a] Departamento de Ingeniería Civil, Universidad de Chile, Chile
[b] Facultad de Ciencias Físicas y Matemáticas, PhD in Engineering Systems program, Universidad de Chile, Chile
[c] Departmento de Ingeniería de Transporte y Logística, Pontificia Universidad Católica de Chile, Chile
[d] Instituto de Economía, Pontificia Universidad Católica de Chile, Chile

ARTICLE INFO

Article history:
Received 31 December 2018
Revised 26 June 2019
Accepted 27 June 2019

Keywords:
Bus rapid transit
Dynamic congestion
Bottleneck model
Bus stop delays

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.

© 2019 Elsevier Ltd. All rights reserved.

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

☆ We thank the Editor, two referees, and attendants at the International Transport Economics Association meetings, Jos van Ommeren, Erik Verhoef, Andrés Gómez-Lobo and Luis Rizzi for helpful and insightful comments which materially helped improve this paper. The authors gratefully acknowledge financial support from the Complex Engineering Systems ISCI (grant CONICYT FB0816) and FONDECYT-Chile, grant 1161032. Silva also acknowledges financial support from the Center of Sustainable Urban Development CEDEUS (grant CONICYT/FONDAP 15110020) and the Bus Rapid Transit Centre of Excellence (BRT+) funded by the Volvo Research and Educational Foundations (VREF) and Feres from CONICYT-PCHA/Doctorado Nacional/2016-21160707.

* Corresponding author.
E-mail addresses: lbasso@ing.uchile.cl (L.J. Basso), fernando.feres@ing.uchile.cl (F. Feres), husilva@uc.cl (H.E. Silva).

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

https://doi.org/10.1016/j.trb.2019.06.012
0191-2615/© 2019 Elsevier Ltd. All rights reserved.

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 km. By 2018, the numbers exploded to 170 cities around the world, for a total of 376 corridors and 5,046 km, 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 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, 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

---

[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.

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

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 we describe the model and characterize the equilibrium under mixed-traffic conditions for any given public transport frequency and fare. Section 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 4, study the optimum under mixed traffic conditions, optimizing the time-invariant car toll, the bus fare, and the bus frequency. Section 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 6 summarizes the policy implications and concludes.

2. The model and equilibrium under mixed-traffic conditions

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] Fig. 1 illustrates our setting.

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 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, 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:

$$c_c(t) = p_c + r_c + \alpha \cdot T_w(t) + \begin{cases} \beta \cdot (t^* - t_a) & \text{if } t_a \le t^* \\ \gamma \cdot (t_a - t^*) & \text{if } t_a > t^* \end{cases} \tag{1}$$

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

[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.

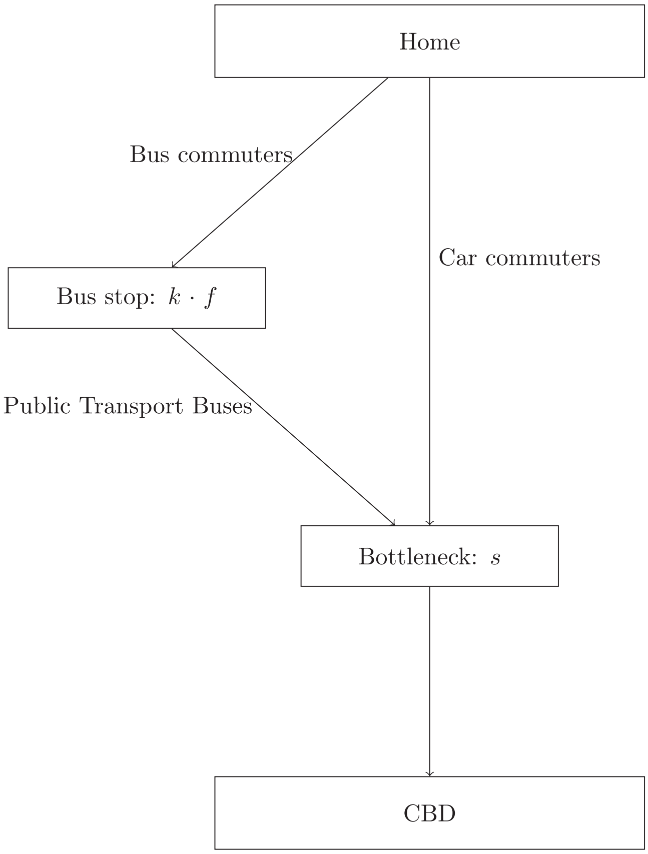

Fig. 1. Two bottleneck diagram of the mixed traffic transport system.

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. (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:

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

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 $T_w\big(t + T_q(t)\big)$, which for simplicity is valued at α as well. Finally, the schedule delay cost is the fourth term on the right-hand side of Eq. (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. 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 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 $p_c + r_c + \beta \cdot \left(t^* - t_b^s\right) + \alpha \cdot T_w\left(t_b^s\right)$. The total cost of a bus user departing at t_b^s is $p_b + r_b + \beta \cdot \left(t^* - t_b^s\right) + \alpha \cdot T_w\left(t_b^s\right)$ because she may 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 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. (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. (4). Note that by Lemma 1, buses are in operation for all the car peak-period.

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

$$(t_c^e - t_c^s) \cdot (s - \lambda \cdot f) = N_c \tag{4}$$

Following the same reasoning, the first and last departure by bus must only face schedule delay costs as in Eq. (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. (6).

$$\beta \cdot (t^* - t_b^s) = \gamma \cdot (t_b^e - t^*) \tag{5}$$

$$(t_b^e - t_b^s) \cdot k \cdot f = N_b \tag{6}$$

Solving the system of Eqs. (3)–(6) we obtain, conditional on N_b and N_c, the equilibrium times for car and bus operations and the cost of traveling by each mode.[9] Defining $\delta = \beta \cdot \gamma / (\beta + \gamma)$, these are:

$$t_c^s = t^* - \frac{\delta}{\beta} \frac{N_c}{s - \lambda \cdot f} \tag{7}$$

$$t_c^e = t^* + \frac{\delta}{\gamma} \frac{N_c}{s - \lambda \cdot f} \tag{8}$$

$$t_b^s = t^* - \frac{\delta}{\beta} \frac{N_b}{k \cdot f} \tag{9}$$

$$t_b^e = t^* + \frac{\delta}{\gamma} \frac{N_b}{k \cdot f} \tag{10}$$

$$c_c = p_c + r_c + \delta \frac{N_c}{s - \lambda \cdot f} \tag{11}$$

$$c_b = p_b + r_b + \delta \frac{N_b}{k \cdot f} \tag{12}$$

[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. (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. (9) the cost of traveling by bus can be computed.

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:

$$N_c = (s - \lambda \cdot f) \cdot \frac{N - \frac{(p_c + r_c - p_b - r_b) \cdot k \cdot f}{\delta}}{s - \lambda \cdot f + k \cdot f} \tag{13}$$

$$N_b = k \cdot f \cdot \frac{N + \frac{(p_c + r_c - p_b - r_b) \cdot (s - \lambda \cdot f)}{\delta}}{s - \lambda \cdot f + k \cdot f} \tag{14}$$

Finally, the equilibrium cost c is:

$$c = p_c + r_c + \delta \frac{N - \frac{(p_c + r_c) - (p_b + r_b) \cdot k \cdot f}{\delta}}{s - \lambda \cdot f + k \cdot f} = p_b + r_b + \delta \frac{N + \frac{(p_c + r_c) - (p_b + r_b) \cdot (s - \lambda \cdot f)}{\delta}}{s - \lambda \cdot f + k \cdot f} \tag{15}$$

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

Proposition 1. *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 Fig. 2.*

Proof. See Appendix B □

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.

Fig. 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

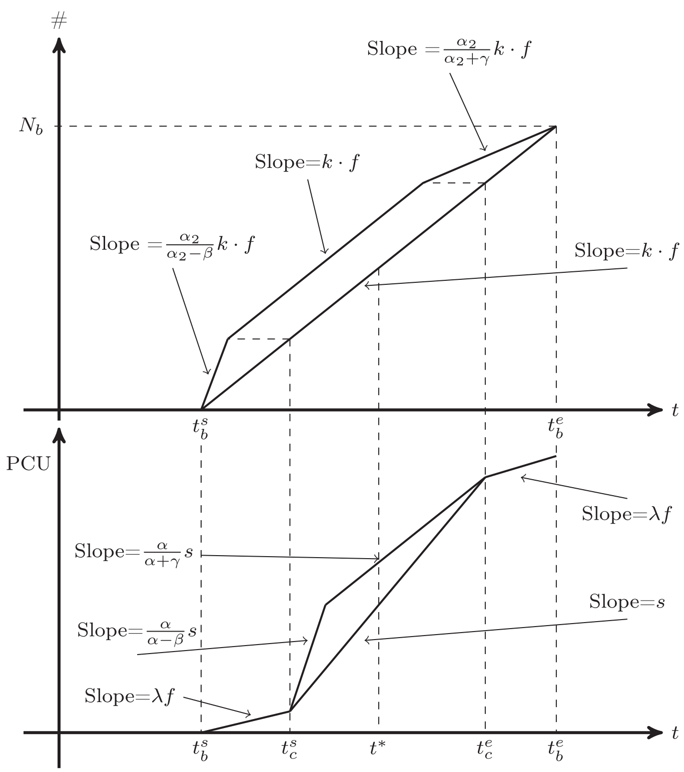

Fig. 2. Interior equilibrium under mixed traffic when $0 < p_c + r_c - p_b - r_b < \delta N/(k \cdot f)$. The upper panel displays cumulative departures and arrivals of passengers at the bus stop. The lower panel displays cumulative arrivals of vehicles (in PCU) to the road bottleneck and to the CBD.

$[t_b^s, t_c^s]$ 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 waiting 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.

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

Proposition 2. *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 C □

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 time-invariant 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 or 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 3. *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 D □

The full proof is provided in the appendix but, to understand why everything goes consider that $p_c + r_c = p_b + r_b + \epsilon$. From Proposition 1, if $\epsilon > 0$, the car period contains the bus period and both modes are used. On the other hand, from Proposition 2, if $\epsilon < 0$, 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 Fig. 2 when periods coincide. But, if ϵ approaches zero from a negative value, buses are not used in equilibrium. It follows that when $\epsilon = 0$, both equilibria (from positive and negative limits) may occur, but also everything that is in between may be an equilibrium.

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 of passengers to the bus stop must be $k \cdot f$ while $f > 0$ and 0 otherwise.

If resource costs r_c and r_b were 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:

$$E(f) = c_1 \cdot f \cdot T + c_2 \cdot f \cdot \Delta t_b \tag{16}$$

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, T^0. The expenditure then collapses to:

$$E(f) = c_1 \cdot f \cdot T^0 + c_2 \cdot f \cdot \Delta t_b \tag{17}$$

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.
- 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 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.

4. Optimum in mixed traffic with time-invariant prices

In Section 2 we characterized equilibria under mixed traffic conditions for the two-modes system. In this Section we show how to optimize it and what are the features of this optimum. Start by recalling that, according to Proposition 1 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 3), while if $p_c + r_c - p_b - r_b < 0$ the equilibrium has no bus users (Proposition 2). 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 1, 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 Fig. 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, $p_c + r_c - p_b - r_b = \epsilon > 0$) 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. (15), we obtain the total user cost (UC):

$$UC = \delta \frac{N^2 + \left(\frac{p_c + r_c - p_b - r_b}{\delta}\right)^2 k \cdot f \cdot (s - \lambda \cdot f)}{s - \lambda \cdot f + k \cdot f} + r_b \cdot N_b + r_c \cdot N_c \tag{18}$$

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. (15), we can write the cycle time as:

$$T = T^0 + \frac{\delta}{\alpha}\frac{N - \frac{(p_c+r_c-p_b-r_b)\cdot k\cdot f}{\delta}}{s-\lambda\cdot f+k\cdot f} \tag{19}$$

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

Replacing the cycle time of Eq. (19) into the public transport expenditure function in Eq. (17) and calculating the operating period of public transport, Δt_b, from Eqs. (9), (10) and (14) we obtain:

$$E = c_1\cdot f\cdot\left(T^0 + \frac{\delta}{\alpha}\frac{N - \frac{(p_c+r_c-p_b-r_b)\cdot k\cdot f}{\delta}}{s-\lambda\cdot f+k\cdot f}\right) + c_2\cdot f\cdot\frac{N + \frac{(p_c+r_c-p_b-r_b)\cdot(s-\lambda\cdot f)}{\delta}}{s-\lambda\cdot f+k\cdot f} \tag{20}$$

We define the social cost (SC) directly as the sum of the total user cost (Eq. (18)) and the public transport system cost (Eq. (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

$$p_c^* + r_c - p_b^* - r_b = \frac{c_1\cdot\delta\cdot f\cdot k - c_2\cdot\alpha\cdot(s-\lambda\cdot f)}{2\alpha\cdot k\cdot(s-\lambda\cdot f)} + \frac{r_c - r_b}{2} \tag{21}$$

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 $df^*/dN > 0$. 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. (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. (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.

5. Bus rapid transit (BRT)

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 the 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 E.

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 1. 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,

[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.

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 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. (11) and (12):

$$c_c = p_c + r_c + \delta \frac{N_c}{s - \lambda \cdot f} \tag{22}$$

$$c_b = p_b + r_b + \delta \frac{N_b}{k \cdot f} \tag{23}$$

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. Fig. 3 summarizes the equilibrium when the time-invariant full price of the car is higher than that of the bus. Comparing this with Fig. 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 Fig. 2. This is because with a BRT, only cars suffer from road congestion.

Under mixed traffic and equal time-invariant full prices for cars and buses, there are multiple equilibria (see Proposition 3). 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 Fig. 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

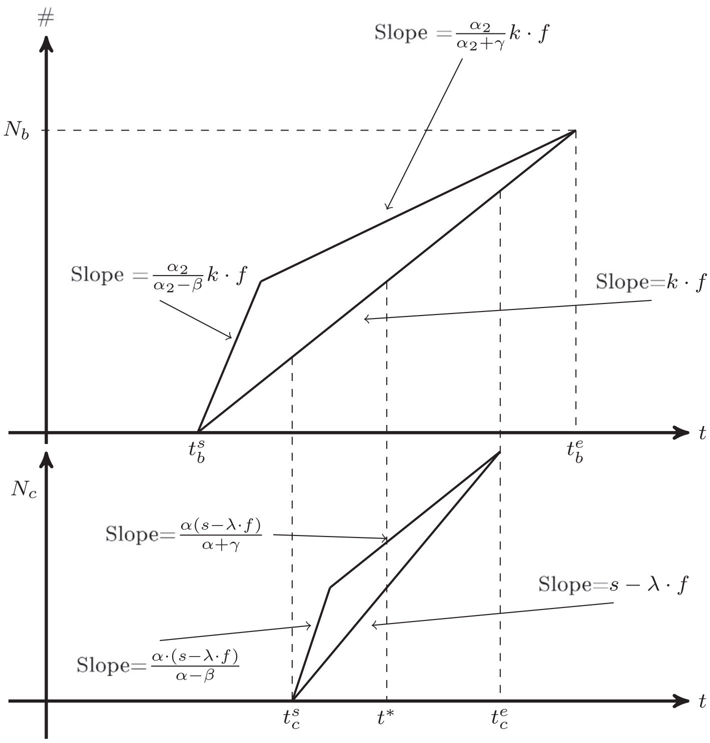

Fig. 3. Equilibrium under BRT when $0 < p_c + r_c - p_b - r_b < \delta N/(k \cdot f)$.

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 Fig. 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 4. *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 E.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. (18), which now also holds for negative price differences:

$$UC = \delta\frac{N^2 + \left(\frac{p_c+r_c-p_b-r_b}{\delta}\right)^2 k\cdot f\cdot(s-\lambda\cdot f)}{s-\lambda\cdot f+k\cdot f} + r_b\cdot N_b + r_c\cdot N_c \tag{24}$$

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. (9), (10), and (14), is:

$$E = c_1\cdot f\cdot T^0 + c_2\cdot f\cdot\frac{N + \frac{(p_c+r_c-p_b-r_b)\cdot(s-\lambda\cdot f)}{\delta}}{s-\lambda\cdot f+k\cdot f} \tag{25}$$

Table 1
Parameters for numerical examples.

Parameter	Units	Value
c_1	[US$/bus]	290
c_2	[US$/hour]	130
T_0	[hour]	0.33
N	[Commuters]	6000 to 13,000
s	[PCU/hour]	6000
ΔR	[US$/Pax]	2.6
α	[US$/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

As a result, when choosing frequency and prices, the planner faces the trade-offs discussed in Section 3. From the first-order condition of the sum of user costs and public transport costs (Eqs. (24) and (25)) with respect to $p_c - p_b$, we obtain:

$$p_c^* + r_c - p_b^* - r_b = -\frac{c_2}{2k} + \frac{r_c - r_b}{2} \tag{26}$$

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 is 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 4. 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 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 E.2. In Appendix E.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 sim-

Table 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	–	19,800	0	19,800	3900	3900	0
	6500	0	–	6500	0	–	07:16	08:21	–	22,154	0	22,154	4577	4577	0
	7000	0	–	7000	0	–	07:13	08:23	–	24,617	0	24,617	5308	5308	0
	7500	0	–	7500	0	–	07:10	08:25	–	27,188	0	27,188	6094	6094	0
	8000	0	–	8000	0	–	07:06	08:26	–	29,867	0	29,867	6933	6933	0
	8500	0	–	8500	0	–	07:03	08:28	–	32,654	0	32,654	7827	7827	0
	9000	0	–	9000	0	–	07:00	08:30	–	35,550	0	35,550	8775	8775	0
	9500	0	–	9500	0	–	06:56	08:31	–	38,554	0	38,554	9777	9777	0
	10,000	0	–	10,000	0	–	06:53	08:33	–	41,667	0	41,667	10,833	10,833	0
	10,500	5	1.75	9797	703	06:46	06:54	08:32	08:36	44,867	2748	42,120	11,099	11,262	164
	11,000	10	2.00	9677	1323	06:55	06:55	08:32	08:32	48,089	5534	42,555	11,601	11,601	0
	11,500	18	2.00	9296	2204	06:57	06:57	08:31	08:31	51,272	9279	41,994	11,701	11,701	0
	12,000	25	2.00	8976	3024	06:59	06:59	08:30	08:30	54,436	12,803	41,632	11,840	11,840	0
	12,500	32	2.00	8702	3798	07:00	07:00	08:29	08:29	57,583	16,161	41,422	12,009	12,009	0
	13,000	39	2.00	8464	4536	07:02	07:02	08:28	08:28	60,716	19,387	41,329	12,200	12,200	0
BRT-DC	6000	13	1.81	4975	1025	07:20	07:26	08:16	08:19	19,589	2929	16,660	2298	2951	1460
	6500	20	1.81	4956	1544	07:20	07:26	08:16	08:19	21,675	4413	17,262	2127	3110	2114
	7000	26	1.81	4937	2063	07:20	07:26	08:16	08:19	23,762	5898	17,864	1977	3291	2723
	7500	33	1.81	4917	2583	07:20	07:26	08:16	08:19	25,848	7382	18,465	1844	3489	3297
	8000	40	1.81	4898	3102	07:20	07:26	08:16	08:19	27,934	8867	19,067	1727	3703	3842
	8500	46	1.81	4878	3622	07:20	07:26	08:16	08:19	30,020	10,351	19,669	1622	3929	4361
	9000	53	1.81	4859	4141	07:20	07:26	08:16	08:19	32,107	11,836	20,271	1528	4166	4859
	9500	59	1.81	4840	4660	07:20	07:26	08:16	08:19	34,193	13,320	20,872	1443	4411	5339
	10,000	66	1.81	4820	5180	07:20	07:26	08:16	08:19	36,279	14,805	21,474	1366	4665	5803
	10,500	73	1.81	4801	5699	07:20	07:26	08:16	08:19	38,365	16,289	22,076	1296	4925	6253
	11,000	79	1.81	4781	6219	07:20	07:26	08:16	08:19	40,452	17,774	22,678	1231	5192	6692
	11,500	86	1.81	4762	6738	07:20	07:26	08:16	08:19	42,538	19,258	23,280	1172	5463	7120
	12,000	93	1.81	4743	7257	07:20	07:26	08:16	08:19	44,624	20,743	23,881	1118	5740	7539
	12,500	99	1.81	4723	7777	07:20	07:26	08:16	08:19	46,710	22,227	24,483	1067	6020	7949
	13,000	106	1.81	4704	8296	07:20	07:26	08:16	08:19	48,797	23,712	25,085	1020	6304	8353
BRT-IC	6000	35	1.81	3279	2721	07:21	07:27	08:16	08:19	21,295	7831	13,464	1748	3453	1705
	6500	42	1.81	3279	3221	07:21	07:27	08:16	08:19	23,360	9270	14,091	1748	3766	2018
	7000	48	1.81	3279	3721	07:21	07:27	08:16	08:19	25,426	10,709	14,717	1748	4079	2332
	7500	55	1.81	3279	4221	07:21	07:27	08:16	08:19	27,492	12,148	15,344	1748	4393	2645
	8000	61	1.81	3279	4721	07:21	07:27	08:16	08:19	29,558	13,587	15,971	1748	4706	2958
	8500	68	1.81	3279	5221	07:21	07:27	08:16	08:19	31,624	15,026	16,597	1748	5019	3272
	9000	74	1.81	3279	5721	07:21	07:27	08:16	08:19	33,690	16,466	17,224	1748	5333	3585
	9500	81	1.81	3279	6221	07:21	07:27	08:16	08:19	35,755	17,905	17,851	1748	5646	3898
	10,000	87	1.81	3279	6721	07:21	07:27	08:16	08:19	37,821	19,344	18,477	1748	5959	4212
	10,500	94	1.81	3279	7221	07:21	07:27	08:16	08:19	39,887	20,783	19,104	1748	6273	4525
	11,000	100	1.81	3279	7721	07:21	07:27	08:16	08:19	41,953	22,222	19,731	1748	6586	4838
	11,500	107	1.81	3279	8221	07:21	07:27	08:16	08:19	44,019	23,661	20,357	1748	6899	5152
	12,000	113	1.81	3279	8721	07:21	07:27	08:16	08:19	46,085	25,101	20,984	1748	7213	5465
	12,500	120	1.81	3279	9221	07:21	07:27	08:16	08:19	48,150	26,540	21,611	1748	7526	5778
	13,000	126	1.81	3279	9721	07:21	07:27	08:16	08:19	50,216	27,979	22,237	1748	7839	6092

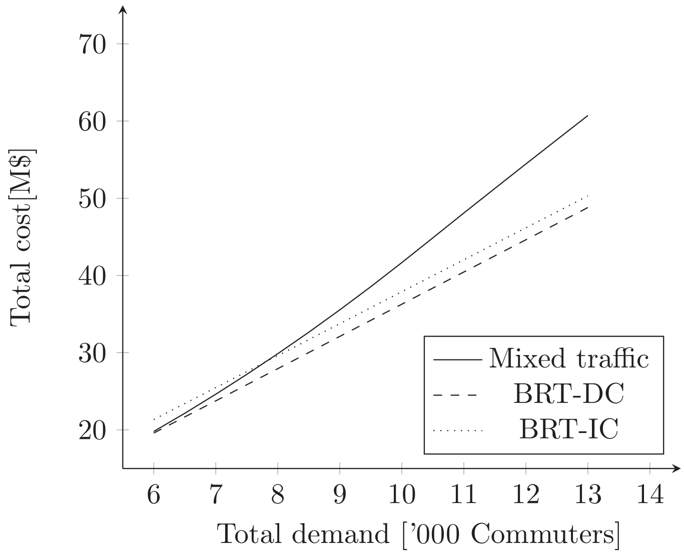

Fig. 4. Total cost.

ple. 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. (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 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 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 E.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).

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 1.[11] 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).

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, Figs. 4 and 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 8000 and 8500 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

[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).

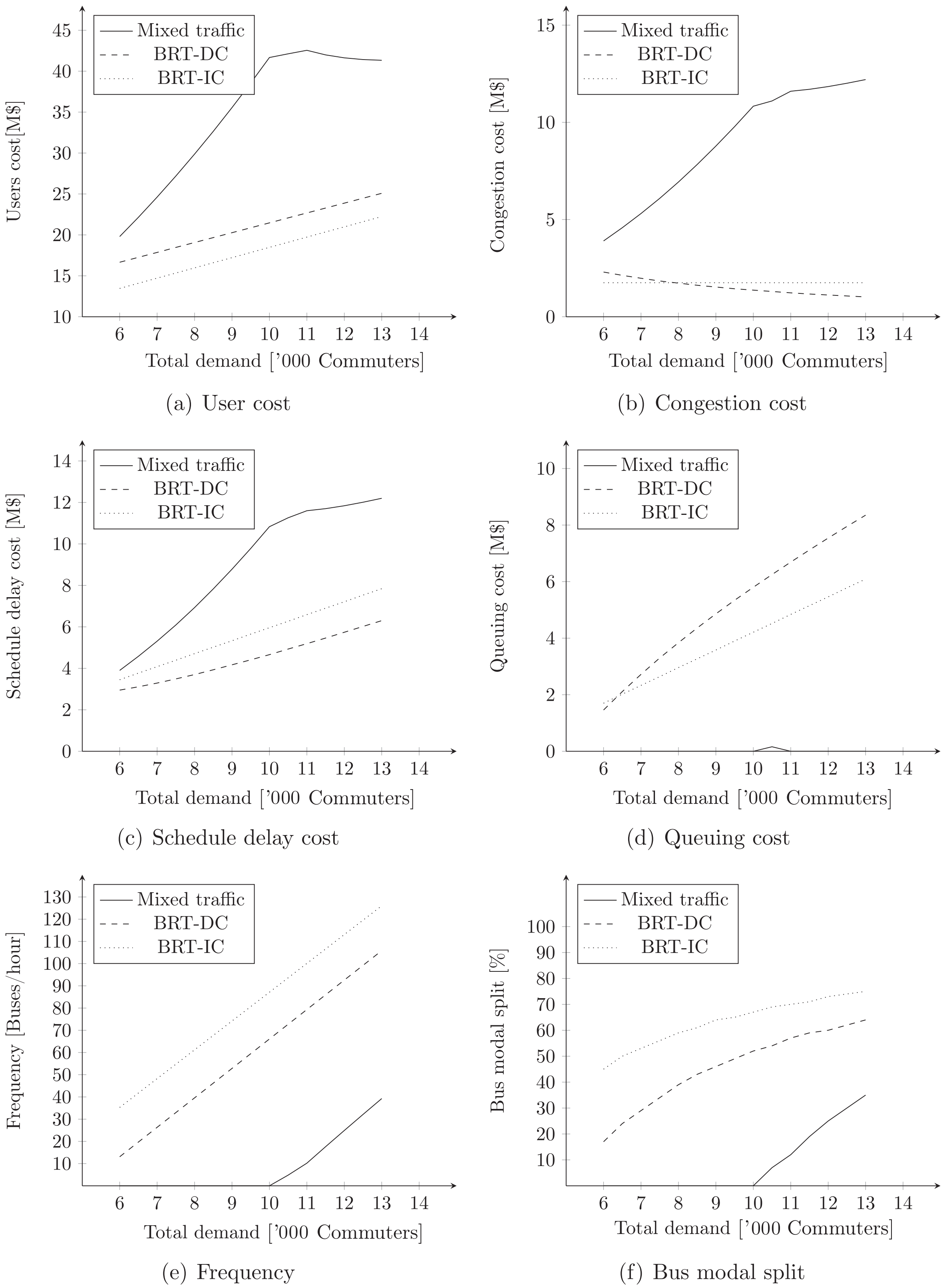

Fig. 5. Sumary report.

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.
- 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.

6. Conclusions

Bus rapid transit systems are a public transport development that has seen an exponential growth 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 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.

Appendix A. 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 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:

$$R \cdot k = N_b \tag{27}$$

We denote the times that the R runs depart from the bus stop by $t_b^s, t_2, \ldots, t_{R-1}, t_b^e$ 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:

$$c_b(t) = p_b(t) + r_b + \alpha_2 \cdot T_q(t) + \alpha \cdot T_w(t + T_q(t)) + \begin{cases} \beta \cdot (t^* - t_a) & \text{if } t_a \le t^* \\ \gamma \cdot (t_a - t^*) & \text{if } t_a > t^* \end{cases} \tag{28}$$

which is the same expression as Eq. (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. (28) 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).[12] 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:

$$\beta \cdot (t^* - t_b^s) = \gamma \cdot (t_b^e - t^*) \tag{29}$$

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 $t_b^e - t_b^s$ 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:

$$t_b^e - t_b^s = (R-1) \cdot h \tag{30}$$

Combining Eqs. (29) and (30) we obtain the equilibrium user time costs of bus users, UTC_b:

$$UTC_b = \delta(R-1) \cdot h \tag{31}$$

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. (31), we obtain:

$$UTC_b = \delta \cdot \frac{N_b}{k \cdot f} - \frac{\delta}{f} \tag{32}$$

The first term on the right-hand side of Eq. (32) is exactly equal to the user time costs of our continuous model (see Eq. (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 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:

$$\beta \cdot (t^* - t_c^s) = \gamma \cdot (t_c^e - t^*) \tag{33}$$

$$(t_c^e - t_c^s) \cdot (1-\phi) \cdot s = N_c \tag{34}$$

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

$$UTC_c = \delta \cdot \frac{N_c}{(1-\phi) \cdot s} \tag{35}$$

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. (32) and (35), the conditions can be written as:

$$p_c + r_c + \delta \cdot \frac{N_c}{(1-\phi) \cdot s} = p_b + r_b + \delta \cdot \frac{N_b}{k \cdot f} - \frac{\delta}{f} \tag{36}$$

$$N_b + N_c = N \tag{37}$$

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

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

$$N_b = \frac{k \cdot f \cdot ((\Delta R + \Delta p) \cdot (1-\phi) \cdot s + \delta \cdot N) + \delta \cdot k \cdot (1-\phi) \cdot s}{\delta \cdot ((1-\phi) \cdot s + k \cdot f)} \tag{38}$$

$$N_c = N - N_b \tag{39}$$

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:

$$N_b^{\text{int}} = N_b^{\text{cont}} + \frac{k \cdot (1-\phi) \cdot s}{(1-\phi) \cdot s + k \cdot f} \tag{40}$$

$$N_c^{\text{int}} = N_c^{\text{cont}} - \frac{k \cdot (1-\phi) \cdot s}{(1-\phi) \cdot s + k \cdot f} \tag{41}$$

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

$$UC_{\text{int}} = \delta \frac{k \cdot (1-\phi)s \cdot \frac{(\Delta p + \Delta R)}{\delta} - N \cdot k}{k \cdot f + (1-\phi)s} + \delta \frac{k \cdot f \cdot (1-\phi)s \cdot (\frac{\Delta p + \Delta R}{\delta})^2 + N^2}{k \cdot f + (1-\phi)s} + R_b \cdot N + \Delta R \cdot \left(N_c^{\text{cont}} - \frac{k \cdot (1-\phi) \cdot s}{(1-\phi) \cdot s + k \cdot f} \right) \tag{42}$$

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. (23), into the expression for the total user cost of Eq. (24). This gives the following expression for the equilibrium total user cost in the continuous model, UC_{cont}:

$$UC_{cont} = \delta \frac{N^2 + \left(\frac{\Delta p + \Delta R}{\delta}\right)^2 ((1-\phi) \cdot s \cdot k \cdot f)}{(1-\phi) \cdot sf + k \cdot f} + \Delta R \cdot N_c^{\text{cont}} + R_b \cdot N \tag{43}$$

Combining Eqs. (43) and (42) we obtain an expression for the equilibrium total user cost in the intermittent model as a function of UC_{cont}:

$$UC_{\text{int}} = UC_{\text{cont}} + \frac{k \cdot (1-\phi)s \cdot \Delta p - N \cdot k \cdot \delta}{(1-\phi)s + k \cdot f} \tag{44}$$

These two expressions allow 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, $\Omega \equiv \frac{UC_{cont} - UC_{int}}{UC_{int}}$:

$$\Omega = \frac{\delta \cdot k \cdot (\delta \cdot N - \Delta p((1-\phi) \cdot s \cdot f))}{\Delta p \cdot k \cdot ((1-\phi)s \cdot f) \cdot [f \cdot (\Delta p + \Delta R) + \delta] + \delta \cdot N \cdot [\delta(N-k) + r_c(s - \lambda f) + r_b f k]} \tag{45}$$

While the expression is not very informative, what is important to note is that Ω approaches zero as the frequency increases.[13] 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.

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

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

period are given by equating the schedule delay costs for the first and last departure, as in Eq. (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. (4). Eq. (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. (11):

$$UTC_c = \delta \frac{N_c}{s - \lambda \cdot f} \tag{46}$$

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:

$$\beta \cdot \left(t^* - t_b^s\right) = \gamma \cdot \left(t_b^e - t^*\right) \tag{47}$$

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. (30) and (31) hold and, therefore, bus user time cost, conditional on N_b, are given by (32):

$$UTC_b = \delta \cdot \frac{N_b}{k \cdot f} - \frac{\delta}{f} \tag{48}$$

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:

$$N_b = \frac{k \cdot f \cdot ((\Delta R + \Delta p) \cdot (s - \lambda f) + \delta \cdot N) + \delta \cdot k \cdot (s - \lambda f)}{\delta \cdot ((s - \lambda f) + k \cdot f)} \tag{49}$$

$$N_c = N - N_b \tag{50}$$

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. (45), 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 $d_c(t)$ be the departure rate from home of car users, a queue develops if $d_c(t) + \lambda f$ 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:

$$V(t) = \int_{\hat{t}_c}^{t} (d_c(x) + \lambda f)dx - s \cdot (t - \hat{t}_c) \tag{51}$$

Note that this is the same as in Eq. (58) and also matches the mixed traffic analysis of Huang et al. (2007). As the cost of a departure by car is given by Eq. (1), the analysis in Eqs. (57)–(59) hold and we obtain the same equilibrium departure rate for car users as in Proposition 1 (Eq. (52)):

$$d_c(t) = \begin{cases} \frac{\alpha s}{\alpha - \beta} - \lambda f & \text{if } t \le \tilde{t}_c \\ \frac{\alpha s}{\alpha + \gamma} - \lambda f & \text{if } t > \tilde{t}_c \end{cases} \tag{52}$$

The result is intuitive and it is the same as in our continuous model: the sum of the departure rates by cars, $d_c(t)$, 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 Fig. 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. (45), is between 0.2% and 0.3% according to our numerical examples.

Appendix B. Proof of Proposition 1

If $p_c + r_c > p_b + r_b$, by Lemma 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. (5) and (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 $p_c + r_c < p_b + r_b + \delta N/(kf)$, which is exactly the condition stated in Proposition 1. Therefore, we have proven that an equilibrium with no car trips does not exist when $p_c + r_c < p_b + r_b + \delta N/(kf)$.

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. (3) and (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, it must satisfy conditions in Eqs. (3)–(12). The modal split that satisfies these equilibrium conditions is given by Eqs. (13) and (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. (13) and (14) that:

$$N_c = (s - \lambda \cdot f) \cdot \frac{N - \frac{(p_c + r_c - p_b - r_b) \cdot k \cdot f}{\delta}}{s - \lambda \cdot f + k \cdot f} \tag{53}$$

$$N_b = k \cdot f \cdot \frac{N + \frac{(p_c + r_c - p_b - r_b) \cdot (s - \lambda \cdot f)}{\delta}}{s - \lambda \cdot f + k \cdot f} \tag{54}$$

It follows directly from Eqs. (53) and (54) 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 Fig. 2.

To prove these two conditions consider the cost for a departure at time t for each mode in Eqs. (1) and (2) replacing the corresponding arrival times:

$$c_c(t) = p_c + r_c + \alpha \cdot T_w(t) + \begin{cases} \beta \cdot (t^* - t - T_w(t)) & \text{if } t + T_w(t) \le t^* \\ \gamma \cdot (t + T_w(t) - t^*) & \text{if } t + T_w(t) > t^* \end{cases} \tag{55}$$

$$c_b(t) = p_b + r_b + \alpha_2 \cdot T_q(t) + \alpha \cdot T_w(t + T_q(t)) + \begin{cases} \beta \cdot (t^* - t - T_q(t) - T_w(t + T_q(t))) & \text{if } t + T_q(t) + T_w(t + T_q(t)) \le t^* \\ \gamma \cdot (t + T_q(t) + T_w(t + T_q(t)) - t^*) & \text{if } t + T_q(t) + T_w(t + T_q(t)) > t^* \end{cases} \tag{56}$$

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. (55) and equating to zero we obtain:

$$\frac{\partial T_w(t)}{\partial t} = \begin{cases} \frac{\beta}{\alpha-\beta} & \text{if } t \leq \tilde{t_c} \\ \frac{\gamma}{\alpha+\gamma} & \text{if } t > \tilde{t_c} \end{cases} \tag{57}$$

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 $d_c(t)$ be the departure rate from home of car users, a queue develops if $d_c(t) + \lambda f$ 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:

$$V(t) = \int_{\hat{t_c}}^{t} (d_c(x) + \lambda f)dx - s \cdot (t - \hat{t_c}) \tag{58}$$

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

$$T_w(t) = \frac{V(t)}{s} \tag{59}$$

Combining Eqs. (57)–(59) we obtain the equilibrium departure rate for car users:

$$d_c(t) = \begin{cases} \frac{\alpha s}{\alpha-\beta} - \lambda f & \text{if } t \leq \tilde{t_c} \\ \frac{\alpha s}{\alpha+\gamma} - \lambda f & \text{if } t > \tilde{t_c} \end{cases} \tag{60}$$

The result is intuitive: the sum of the departure rates by cars, $d_c(t)$, 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. (56) and equating to zero we obtain:

$$\frac{\partial T_q(t)}{\partial t} = \begin{cases} \dfrac{\beta - \frac{\partial T_w}{\partial t}(\alpha-\beta)}{\alpha_2 - \beta + \frac{\partial T_w}{\partial t}(\alpha-\beta)} & \text{if } t \leq \tilde{t_b} \\ \dfrac{-\gamma - \frac{\partial T_w}{\partial t}(\alpha+\gamma)}{\alpha_2 - \beta + \frac{\partial T_w}{\partial t}(\alpha+\gamma)} & \text{if } t > \tilde{t_b} \end{cases} \tag{61}$$

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 $d_b(t)$ be the departure rate from home of bus users, a queue develops at the bus stop if $d_b(t)$ 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:

$$V_b(t) = \int_{\hat{t_b}}^{t} d_b(x)dx - kf \cdot (t - \hat{t_b}) \tag{62}$$

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:

$$T_q(t) = \frac{V_b(t)}{kf} \tag{63}$$

which is analogous to the classic bottleneck model, but with capacity kf. 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 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. (62) and (63) into Eq. (61) we obtain the equilibrium departure rate for bus users:

$$d_b(t) = \begin{cases} \frac{\alpha_2 kf}{\alpha_2-\beta} & \text{if } t \leq \tilde{t_b^s} \\ kf & \text{if } \tilde{t_b^s} < t \leq \tilde{t_b^e} \\ \frac{\alpha_2 kf}{\alpha_2+\gamma} & \text{if } \tilde{t_b^e} < t \end{cases} \tag{64}$$

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*.

Appendix C. Proof of Proposition 2

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 $\delta N/(s - \lambda f)$,[14] 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.

Appendix D. Proof of Proposition 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 B 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 D 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, 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 B, this can only occur when departure rates for car users are given by Eq. (60). 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. (60) 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

[14] See Eqs. (3) and (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 B.

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 Eqs. (3)–(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 $p_c + r_c = p_b + r_b$:

$$\bar{N}_c = (s - \lambda \cdot f) \cdot \frac{N}{s - \lambda \cdot f + k \cdot f} \tag{65}$$

$$\bar{N}_b = k \cdot f \cdot \frac{N}{s - \lambda \cdot f + k \cdot f} \tag{66}$$

$$c = p + \delta \frac{N}{s - \lambda \cdot f + k \cdot f} \tag{67}$$

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:

$$\beta \cdot (t^* - t_c^s) = \gamma \cdot (t_c^e - t^*) \tag{68}$$

$$(t_c^e - t_c^s) \cdot (s - \lambda \cdot f) = N_c \tag{69}$$

the departure rates are the ones in Eq. (60) and the remaining $N_b = N - N_c$ bus users depart in the period $[t_c^s; t_c^e]$ at a rate lower than kf such that there are no bus stop queuing delays.

Appendix E. Calculations and results for Section 5

E1. Proof of Proposition 4

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. (23):

$$c_c = p_c + r_c + \delta \frac{N_c}{s - \lambda \cdot f} \tag{70}$$

$$c_b = p_b + r_b + \delta \frac{N_b}{k \cdot f} \tag{71}$$

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

$$N_b = \frac{k \cdot f \cdot ((\Delta R + \Delta p) \cdot (s - \lambda f) + \delta \cdot N)}{\delta \cdot ((s - \lambda f) + k \cdot f)} \tag{72}$$

$$N_c = \frac{(s - \lambda f) \cdot (\delta \cdot N - k \cdot f \cdot ((\Delta R + \Delta p))}{\delta \cdot ((s - \lambda f) + k \cdot f)} \tag{73}$$

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

$$N_b > 0 \iff \Delta R + \Delta p > \frac{-\delta \cdot N}{s - \lambda f} \tag{74}$$

$$N_c > 0 \iff \Delta R + \Delta p < \frac{\delta \cdot N}{k \cdot f} \tag{75}$$

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 turn to prove results (i)–(iii) in Proposition 4. 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:

$$c_c(t_c^s) = p_c + r_c + \beta \cdot (t^* - t_c^s) \tag{76}$$

$$c_b(t_b^s) = p_b + r_b + \beta \cdot (t^* - t_b^s) \tag{77}$$

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

$$t_c^s - t_b^s = \frac{p_c + r_c - p_b - r_b}{\beta} \tag{78}$$

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

$$c_c(t_c^e) = p_c + r_c + \gamma \cdot (t_c^e - t^*) \tag{79}$$

$$c_b(t_b^e) = p_b + r_b + \gamma \cdot (t_b^e - t^*) \tag{80}$$

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

$$t_b^e - t_c^e = \frac{p_c + r_c - p_b - r_b}{\gamma} \tag{81}$$

It follows directly from Eqs. (78) and (81) 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.

E2. 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 (24) we obtain these costs as a function of the frequency and the price difference:

$$UC = \frac{\delta\left(\frac{k \cdot f \cdot (\Delta p + \Delta R)^2 \cdot (s - \lambda \cdot f)}{\delta^2} + N^2\right)}{s - \lambda \cdot f + k \cdot f} + \frac{\Delta R \cdot (s - f\lambda)\left(N - \frac{k \cdot f \cdot (\Delta p + \Delta R)}{\delta}\right)}{(1 - \phi)sf + k \cdot f} + N \cdot R_b \tag{82}$$

$$OC = c_1 \cdot f \cdot T_0 + c_2 \cdot f \frac{\left(\frac{(\Delta R + \Delta p)(s - f\lambda)}{\delta} + N\right)}{s - \lambda \cdot f + k \cdot f} \tag{83}$$

$$SC = UC + OC \tag{84}$$

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

$$\frac{f \cdot ((1 - \phi)sf)(c2 + k(\Delta R + 2\Delta p))}{\delta \cdot (s - \lambda \cdot f + k \cdot f)} = 0 \tag{85}$$

which yields the optimal price difference of Eq. (26):

$$\Delta p^* = -\frac{c_2 + k \cdot \Delta R}{2 \cdot k} \tag{86}$$

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:

$$\frac{k\left(4 \cdot \delta \cdot a\left(c_1 \cdot T_0 \cdot (s - \lambda \cdot f + k \cdot f)^2 + \delta N^2(\lambda - k)\right)\right)}{4 \cdot \delta \cdot k \cdot (s - \lambda \cdot f + k \cdot f)^2} + \frac{\left(\Delta R^2 \cdot k \cdot \left(\lambda \cdot f^2(k - \lambda) + 2 \cdot \lambda \cdot f \cdot s - s^2\right)\right)}{4 \cdot \delta \cdot k \cdot (s - \lambda \cdot f + k \cdot f)^2}$$
$$- \frac{\left(4\delta \Delta R k^2 N s\right) - c_2{}^2\left(\lambda \cdot f^2 \cdot (\lambda - k) - 2 \cdot \lambda \cdot f \cdot s + s^2\right)}{4 \cdot \delta \cdot k \cdot (s - \lambda \cdot f + k \cdot f)^2} + \frac{2c_2k\left(\Delta R\left(\lambda \cdot f^2 \cdot (\lambda - k) - 2 \cdot \lambda \cdot fs + s^2\right) + 2\delta \cdot N \cdot s\right)}{4 \cdot \delta \cdot k \cdot (s - \lambda \cdot f + k \cdot f)^2} = 0 \tag{87}$$

As Eq. (87) is quadratic, there are two possible solutions. They can be written as follows:

$$f_1^* = \frac{A - \sqrt{B}}{D} \tag{88}$$

$$f_2^* = \frac{A + \sqrt{B}}{D} \tag{89}$$

with:

$$A = -4 \cdot c_1 \cdot \delta \cdot k \cdot s \cdot T^0 \cdot (k - \lambda) - \lambda \cdot s(c_2 - \Delta R \cdot k)^2 \tag{90}$$

$$B = k \cdot \left(4c_1 \cdot \delta \cdot k \cdot T^0 \cdot (k - \lambda) + \lambda \cdot (c_2 - \Delta R k)^2\right)(c_2 \cdot s - \Delta R \cdot k \cdot s + 2\delta \cdot N \cdot (\lambda - k))^2 \tag{91}$$

$$D = (k - \lambda)\left(4c_1 \cdot \delta \cdot k \cdot T^0 \cdot (k - \lambda) + \lambda \cdot (c_2 - \Delta R \cdot k)^2\right) \tag{92}$$

It is straightforward to show that $D > 0$ and $A < 0$ hold, so the only positive solution is f_2^* of Eq. (89). 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.

E3. Calculations for the optimal period of operation for buses with perfectly divisible capacity (BRT-DC)

The period of operation for buses is defined by:

$$N_b/(k \cdot f) \tag{93}$$

Substituting the optimal price difference of Eq. (86) and the optimal frequency of Eq. (89) in the expression for N_b of Eq. (14), we obtain:

$$N_b(f_2^*, \Delta p^*)/(k \cdot f_2^*) = \frac{\lambda \cdot c_2 - \Delta R \cdot k \cdot \lambda + \sqrt{k \cdot \left(4c_1 \cdot \delta \cdot k \cdot T_0(k-\lambda) + \lambda \cdot (c_2 - \Delta R \cdot k)^2\right)}}{2 \cdot \delta \cdot k \cdot (k-\lambda)} \tag{94}$$

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

E4. 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:

$$UC = \frac{\delta\left(\frac{k \cdot f \cdot (\Delta p + \Delta R)^2 \cdot s_1}{\delta^2} + N^2\right)}{s_1 + k \cdot f} + \frac{\Delta R \cdot s_1\left(N - \frac{k \cdot f \cdot (\Delta p + \Delta R)}{\delta}\right)}{s_1} + N \cdot R_b \tag{95}$$

$$OC = c_1 \cdot f \cdot T_0 + c_2 \cdot f \frac{\left(\frac{(\Delta R + \Delta p)(s - f\lambda)}{\delta} + N\right)}{s_1 + k \cdot f} \tag{96}$$

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

$$\Delta p^* = -\frac{c_2 + k \cdot \Delta R}{2 \cdot k} \tag{97}$$

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

$$f^* = \frac{-2c_1 k^2 s_1 T_0 \delta + (-c_2 s_1 + k\Delta R s_1 + 2kN\delta)\sqrt{c_1 k^3 T_0 \delta}}{2c_1 k^3 T_0 \delta} \tag{98}$$

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

$$N_c = \frac{s_1\left(c_2 \cdot k - k^2 \cdot \Delta R + 2\sqrt{c_1 \cdot k^3 \cdot T_0 \cdot \delta}\right)}{2k^2 \cdot \delta} \tag{99}$$

$$N_b = N - \frac{s_1\left(c_2 \cdot k - k^2 \cdot \Delta R + 2\sqrt{c_1 \cdot k^3 \cdot T_0 \cdot \delta}\right)}{2k^2 \cdot \delta} \tag{100}$$

which imply the following car and bus peak duration:

$$N_c/s_1 = \frac{\left(c_2 \cdot k - k^2 \cdot \Delta R + 2\sqrt{c_1 \cdot k^3 \cdot T_0 \cdot \delta}\right)}{2k^2 \cdot \delta} \tag{101}$$

$$N_b/(k \cdot f^*) = \sqrt{\frac{c_1 \cdot k \cdot T_0}{\delta}} \tag{102}$$

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.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.trb.2019.06.012.

References

Arnott, R., de Palma, A., Lindsey, R., 1990. Economics of a bottleneck. J. Urban Econ. 27 (1), 111–130. doi:10.1016/0094-1190(90)90028-L.

Arnott, R., de Palma, A., Lindsey, R., 1993. A structural model of peak-Period congestion: a traffic bottleneck with elastic demand. Am. Econ. Rev. 83 (1), 161–179.

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

Basso, L.J., Silva, H.E., 2014. Efficiency and substitutability of transit subsidies and other urban transport policies. Am. Econ. J. 6 (4), 1–33.

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

Börjesson, M., Fung, C.M., Proost, S., 2017. Optimal prices and frequencies for buses in Stockholm. Econ. Transp. 9, 20–36.

Gonzales, E.J., Daganzo, C.F., 2012. Morning commute with competing modes and distributed demand: user equilibrium, system optimum, and pricing. Transp. Res. Part B 46 (10), 1519–1534. doi:10.1016/j.trb.2012.07.009.

Hensher, D.A., Li, Z., Mulley, C., 2014. Drivers of bus rapid transit systems–influences on patronage and service frequency. Res. Transp. Econ. 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. Transp. Res. Part E 36 (4), 267–284. doi:10.1016/S1366-5545(00)00002-8.

Huang, H.J., Tian, Q., Yang, H., Gao, Z.Y., 2007. Modal split and commuting pattern on a bottleneck-constrained highway. Transp. Res. Part E 43 (5), 578–590. doi:10.1016/j.tre.2005.12.003.

Kraus, M., 2003. A new look at the two-mode problem. J. Urban Econ. 54 (3), 511–530. doi:10.1016/S0094-1190(03)00079-2.

Kraus, M., Yoshida, Y., 2002. The commuter's time-of-use decision and optimal pricing and service in urban mass transit. J. Urban Econ. 51 (1), 170–195. doi:10.1006/juec.2001.2242.

Kutzbach, M.J., 2009. Motorization in developing countries: causes, consequences, and effectiveness of policy options. J. Urban Econ. 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., Straszheim, M. (Eds.), Current Issues in Urban Economics. The Johns Hopkins University Press, Baltimore, pp. 165–195.

de Palma, A., Lindsey, R., Monchambert, G., 2017. The economics of crowding in rail transit. J. Urban Econ. 101, 106–122. doi:10.1016/j.jue.2017.06.003.

Scholl, L., Bouillon, C.P., Oviedo, D., Corsetto, L., 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., 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. Am. Econ. Rev. 72 (3), 467–479.

Small, K.A., 1983. Bus priority and congestion pricing on urban expressways. Res. Transp. Econ. 1, 27–74.

Tabuchi, T., 1993. Bottleneck congestion and modal split. J. Urban Econ. 34, 414–431. doi:10.1006/juec.1993.1044.

Vickrey, W.S., 1969. Congestion theory and transport investment. Am. Econ. Rev. 59 (2), 251–260.

Wang, W.W., Wang, D.Z., Zhang, F., Sun, H., Zhang, W., Wu, J., 2017. Overcoming the Downs–Thomson paradox by transit subsidy policies. Transp. Res. Part A 95, 126–147.

Yoshida, Y., 2008. Commuter arrivals and optimal service in mass transit: does queuing behavior at transit stops matter? Reg. Sci. Urban Econ. 38 (3), 228–251.