The Economics of Crowding in Public Transport

André de Palma, Robin Lindsey, and Guillaume Monchambert*

September 21, 2015

– Working Paper –

Abstract

We analyze trip-timing decisions of public transit users who trade off crowding costs and disutility from traveling early or late. Considering fixed and then endogenous demand, we derive the equilibrium distribution of users across trains for three fare regimes: no fare, an optimal uniform fare, and an optimal train-dependent fare that supports the social optimum. We also derive the optimal number of trains and train capacity, and compare them across fare regimes. Finally we calibrate the model to a segment of the Paris RER A mass transit system and estimate the potential welfare gains from train-dependent fares.

Keywords: public transport; crowding; pricing; optimal capacity

JEL Codes: D62; R41; R48

*de Palma: Département Economie et Gestion, École Normale Supérieure de Cachan, 61 avenue du Président Wilson, 94235 Cachan, France (e-mail:andre.depalma@ens-cachan.fr); Lindsey: Sauder School of Business, University of British Columbia, Henry Angus Building, 2053 Main Mall Vancouver, BC V6T 1Z2, Canada (e-mail: robin.lindsey@sauder.ubc.ca); Monchambert: Département Economie et Gestion, École Normale Supérieure de Cachan, 61 avenue du Président Wilson, 94235 Cachan, France, and Center for Economic Studies, University of Leuven, Naamsestraat 69 box 3500, B 3000, Leuven, Belgium (e-mail:guillaume.monchambert@kuleuven.be). For useful comments we would like to thank participants at: the Annual Conference of the International Transportation Economics Association (ITEA) in Toulouse, June 2014; the Department of Civil and Environmental Engineering, The Hong Kong Univerity of Science and Technology, July 2014; the XVI Conference of the SIET Italian Association of Transport Economics and Logistics (SIET) in Florence, October 2014; the Department of Spatial Economics, Free University of Amsterdam, February 2015; the 50th Annual Conference of the Canadian Transportation Research Forum (CTRF), Montreal, May 2015.

1 Introduction

Since the pioneering work of Pigou (1920) and Knight (1924), economists have made major strides in studying road traffic congestion and devising cost-effective ways to alleviate it. By comparison, they have devoted relatively little attention to congestion in public transportation. Yet the costs of travel delays and crowding on transit systems are growing in cities around the world. A roundtable report by the International Transport Forum identifies crowding as a major source of inconvenience that increases the cost of travel (OECD, 2014).

Transit crowding imposes disutility on riders in several ways.[1] It increases waiting time and in-vehicle travel time, and reduces travel time reliability. Psychological studies find that crowding causes stress and feelings of exhaustion (Mohd Mahudin, Cox and Griffiths, 2012). A number of studies document how disutility from in-vehicle time increases with the number of users (Wardman and Whelan, 2011; Haywood and Koning, 2015). In a meta-analysis, Wardman and Whelan (2011) find that the monetary valuation of the disutility from public transport travel time is, on average, multiplied by a factor of 2.32 if a rider has to stand. Discomfort also occurs while entering and exiting transit vehicles, accessing stations on walkways and escalators, and so on. By discouraging travelers from taking transit, crowding also contributes indirectly to traffic congestion on roads.

Several recent studies have documented the aggregate cost of crowding. For example, Prud'homme et al. (2012) estimate that the eight percent increase in densities in the Paris subway between 2002 and 2007 imposed a welfare loss in 2007 of at least €75 million.[2] Veitch, Partridge and Walker (2013) estimate the annual cost of crowding in Melbourne metropolitan trains in 2011 at €208 million.

The costs of crowding are likely to grow as usage of public transit grows faster than capacity. Ongoing urbanization in both developed and developing countries is raising the number of city residents who rely on transit for mobility. Younger people are obtaining a driver's license at a later age or not at all, and choosing to live in high-density areas where transit service can provide most of their travel needs. Though the automobile still dominates in the US and Canada, public transit ridership is rising there too.[3] Cities, meanwhile, struggle to obtain adequate funding for capacity expansion and operations. Shortage of money is especially severe in countries such as Canada that lack long-term, dedicated transit funding mechanisms.

City planners are now recognizing that crowding should be considered in cost-benefit analysis of transit projects as well as travel-demand management policies (Parry and Small, 2009). For example, bus service was improved in London prior to introduction of the Congestion Charge in 2003. Similarly, bus, metro, and rail service were expanded in Stockholm before the Congestion Tax trial in 2006. Nevertheless, crowding and other dimensions of public transit quality are still often undervalued in project evaluation relative to more easily measured metrics such

[1] See Tirachini, Hensher and Rose (2013) for a review.

[2] Measured in passengers per square meter aboard trains.

[3] Transit ridership in the US has been growing since 1995 (`http://www.apta.com/mediacenter/pressreleases/2015/Pages/150309_Ridership.aspx`.) In all ten of the largest Canadian cities share of morning commutes by public transit increased from 2006 to 2011 (`http://www12.statcan.gc.ca/nhs-enm/2011/as-sa/99-012-x/2011003/tbl/tbl1b-eng.cfm`; `http://www12.statcan.gc.ca/nhs-enm/2011/as-sa/99-012-x/2011003/tbl/tbl1a-eng.cfm`). Allen and Levinson (2014) document the rapid growth in usage of commuter rail services in both countries.

as in-vehicle speed (OECD, 2014).

Expanding capacity is an obvious way to alleviate transit crowding, but it is expensive and time-consuming and it is often opposed by residents and businesses located near transit routes. An alternative is to use transit fares as a rationing mechanism. Vickrey (1963) was one of the first to advocate peak-load fares, and to note the common economic principles underlying congestion pricing of transit systems and congestion pricing of roads.

A number of transit agencies do practice some degree of time-of-day differentiation.[4] However, flat fares are still common in many large urban areas including Paris, Hong Kong, and Toronto. Opinions differ as to whether peak-period pricing is cost-effective.[5] One argument against it is that travelers such as morning commuters lack the flexibility to change their trip times. However, only a fraction of users need to shift in order to obtain congestion relief. Several surveys (e.g., of London and Melbourne) have found that an appreciable fraction of travelers are willing to shift travel time by 15 minutes, and in some cases more, if they are compensated in some way (e.g., by fare reductions, faster trains, or less crowding). Off-peak discounts have been implemented in some cities, and they are popular with travelers.

This paper has three goals. The first is to develop a general model of trip-timing decisions on crowded public transport systems and analyze equilibrium usage patterns. The second is to derive optimal time-of-day (TOD) varying fares and examine how TOD fares affect the timing of trips and the extent to which the costs of crowding can be alleviated relative to a flat (i.e., time-independent) fare scheme. The third is to derive optimal transit capacity for both flat and TOD fare regimes to assess how the efficiency of the pricing scheme affects the benefits of capacity investments. More specifically, we address two questions. First, how does the welfare gain from optimal TOD fares depend on the severity of crowding? Is it the case that, as in road traffic congestion models, the gains from congestion pricing increase more than proportionally with the number of users? Second, is it true that as is widely assumed congestion pricing is a substitute for investment in the sense that introducing TOD fares reduces the urgency of building new transit capacity?

Beginning with Mohring (1972), an economics literature has developed on public transit capacity investments, service frequency, and optimal pricing and subsidy policy. However, most studies have employed static models that cannot account for travelers' time-of-use decisions and the large daily variations in ridership and crowding typical of major transit systems. To model time-of-use decisions we adopt the demand side of Vickrey's (1969) classical bottleneck model of driving in which congestion takes the form of queuing behind a bottleneck. In this model motorists prefer to minimize the time they spend driving and to arrive at a particular time such as 8:30 a.m. to start work. In equilibrium, queuing delay grows smoothly to a peak at the preferred arrival time and then decreases smoothly back to zero. Individual travelers face a trade-off between traveling at the peak and suffering a long trip,

[4] These include the London subway (`http://www.tfl.gov.uk/cdn/static/cms/documents/tube-dlr-lo-adult-fares.pdf`), the Long Island Rail Road (`http://web.mta.info/lirr/about/TicketInfo/`), and the Washington, D.C. metro (`http://www.wmata.com/fares/`). In Singapore, commuting to the downtown core is free before 7:45 a.m. and a 50 cent discount applies when arriving between 7:45 a.m. and 8:00 a.m. (Singapore Land Transport Authority, 2013). Similarly, in Melbourne, weekday trips on the electrified train network before 7am are free for users with the myki travel card (`http://ptv.vic.gov.au/tickets/myki/myki-money/`).

[5] Informal discussions for Washington, D.C., and Toronto are found in Walker (2010) and Yauch (2015) respectively.

and traveling off-peak and incurring a *schedule delay cost* (i.e.; the cost of arriving earlier or later than desired).[6] Since Small (1982), many empirical studies have estimated the functional form and magnitude of *schedule delay costs* as well as how individuals adapt their schedules over time (e.g., Peer et al., 2015). Arnott, de Palma and Lindsey (1993) provide an extended theoretical analysis of the basic bottleneck model. Various extensions and applications of the model are reviewed in Small (2015).

The bottleneck model cannot be applied unchanged to public transit because of differences between transit service and driving on the supply side. Drivers can start their trips whenever they want, while transit users are constrained by the service timetable. Departure-time choices by transit are thus discrete, rather than continuous. Road capacity is determined by the infrastructure and throughput is essentially a continuous variable that is the same 24 hours day. By contrast, transit is supplied as a batch service and its capacity depends on the number of transit vehicles (i.e., buses or train sets) and vehicle capacities which are both choice variables. As discussed below, congestion also manifests on roads and transit in different ways.

Kraus and Yoshida (2002) were the first to apply the bottleneck model to public transit. They consider a rail service between a single origin and destination. The number of people who board a train is limited by its capacity, and congestion takes the form of queuing delay. Service discipline is first-come-first-served, and users traveling at the peak have to wait in line for several trains to pass before they can board. Kraus and Yoshida (2002) use their model to analyze optimal pricing and capacity decisions and compare the results with those obtained by Mohring (1972) using a simpler model.

Train capacity in Kraus and Yoshida's (2002) model is "hard" in the sense that it has no effect on users' costs until the capacity constraint is reached, but the number of passengers who board a train cannot exceed capacity at any cost. On most transit systems, congestion does not develop as abruptly as this, but rather increases smoothly or incrementally with passenger loads as crowding develops on walkways, escalators, and platforms as well as in vehicles themselves.

A few studies have taken steps towards modeling transit congestion in the form of crowding. Huang, Tian and Gao (2005) assume that travelers board trains in random order. Everyone waiting for a train is able to get on, but the discomfort incurred while aboard increases with the passenger load. Huang et al. (2007) and Tian, Huang and Yang (2007) build on Huang, Tian and Gao (2005), but take an engineering and/or operational research view of crowding and do not systematically explore the economic aspects of the problem or investigate optimal capacity decisions. de Palma, Kilani and Proost (2015) focus on the functional form of the crowding cost function for seated and standing passengers. They also derive an optimal timetable and pricing scheme for several stylized settings. However, they do not investigate analytically the welfare gains from TOD pricing or solve for optimal service capacity as measured by the number of trains and individual train capacity.

[6] Henderson (1974) adopted the same demand-side specification as Vickrey (1969), but instead of queuing assumed that travel delay manifests as flow congestion. In Henderson's model, the speed at which a vehicle travels throughout its trip is determined solely by the departure flow of vehicles when it starts its trip. The model therefore has the unrealistic feature that vehicles departing at different times do not interfere with each other at all.

Our paper parallels Kraus and Yoshida (2002) in considering transit service between a single origin and destination. A fixed number of transit vehicles departs the origin station according to a timetable. Like Kraus and Yoshida (2002) we assume that transit service operates on a separate right of way so that it neither affects or is affected by traffic flow on roads. The model is therefore applicable to train service or buses that run on dedicated bus lanes. We also assume that travelers do not choose between taking transit and driving so that the number of road users is exogeneous to the model and road traffic congestion can be ignored.

Our analysis builds on antecedent work in three ways. First, we explore in some depth the welfare gain from transit congestion pricing and how it depends on the functional form of the crowding cost function. Second, we derive the optimal number of trains and train capacity for three fare regimes: no-fare, optimal uniform-fare, and optimal TOD fares, and compare service supply across regimes. Third, we compare the properties of the model with those of the original bottleneck model and highlight some of the differences.

To preview results, the answers to the two questions we posed above are as follows. First, when capacity is fixed the welfare gain from implementing optimal TOD (i.e., train-dependent) fares may not increase with the total number of users or, therefore, the severity of crowding. Indeed, if the cost of crowding aboard a train grows at an increasing rate with the passenger load the welfare gain decreases with the total number of users. Second, even if the total number of users is fixed the optimal number of trains and train capacity can be higher with optimal TOD fares than when fares are uniform for all trains. Thus, while congestion pricing can improve utilization of a given transit service, it can actually increase the benefits of expanding capacity.

Section 2 describes the general model and derives the equilibrium trip-timing decisions of users for the flat-fare and optimal TOD fare regimes. Section 3 analyzes a case with linear crowding costs. Section 4 considers the long run in which the number of trains and train capacity are endogenous. Section 5 presents a numerical example based on the Paris RER A line, and Section 6 concludes.

2 The general model with inelastic demand

In this section we introduce a general model of public transit crowding which we call the "*PTC*" model. A transit line connects two stations without intermediate stops. The line runs on a timetable to which the operator adheres precisely. There are m trains, indexed in order of departure. Train k leaves the origin station at time t_k, $k = 1, ..., m$. Travel time aboard a train is independent of both departure time and train occupancy, and without loss of generality it is normalized to zero.

Each morning a fixed number, N, of identical users take the line to work. They know the timetable and the crowding level on each train, and choose which train to take. By assumption, they cannot increase their chances of securing a good seat by arriving at the origin station early. Users choose between trains based on the expected crowding disutility, $g(n)$, where n is the number of users taking the same train.[7] Crowding disutility is assumed

[7] Function $g(n)$ is an average over possible states: securing a good seat, getting a bad seat, having to stand in the middle of the

to be zero on an empty train (i.e., $g(0) = 0$), strictly increasing with n (i.e., $g'(.) > 0$), and twice continuously differentiable. Several properties of the model derived later depend on the curvature of $g(n)$ which will be described by the elasticity of $g'(n)$ with respect to n: $\varepsilon(n) \equiv g''(n) n/g'(n)$.[8]

Because trains are costly to procure and operate, it is natural to assume that all m trains are used. Letting n_k denote the number of users on train k we thus assume that for all $k = 1...m$, $n_k > 0$ which implies that $g(n_k) > 0$: users incur a crowding disutility on every train.

Since travel time is normalized to zero, an individual is either at home or at work. Following Vickrey (1969) and Small (1982), time at home yields an instantaneous time-varying utility $u_h(t)$, and time at work an instantaneous time-varying utility $u_w(t)$. Let (t_B, t_E) denote the time interval during which all travel takes place. It is assumed that during this interval, $u_h(t)$ is weakly decreasing and $u_w(t)$ is weakly increasing. The functions intersect at time t^* which is the *desired arrival time* (i.e., $u_h(t^*) = u_w(t^*)$). A user taking train k gains a total utility of $U(t_k) = \int_{t_B}^{t_k} u_h(t)\,dt + \int_{t_k}^{t_E} u_w(t)\,dt - g(n_k)$. If a train with unlimited capacity left at t^*, the user could travel from home to work at t^* without suffering crowding disutility. As a consequence, his utility would be maximal and equal to $U^{\max} = \int_{t_B}^{t^*} u_h(t)\,dt + \int_{t^*}^{t_E} u_w(t)\,dt$. We define the *user travel cost*, c_k, as the difference between this hypothetical maximal utility and the actual utility of taking train k: $c(t_k) \equiv U^{\max} - U(t_k) = g(n_k) + \delta(t_k)$, where $\delta(t_k)$ is the *schedule delay cost* such that

$$\delta(t_k) = \begin{cases} \int_{t_k}^{t^*} (u_h(t) - u_w(t))\,dt & \text{if } t_k < t^* \\ \int_{t^*}^{t_k} (u_w(t) - u_h(t))\,dt & \text{if } t_k \geq t^* \end{cases}.$$

Note that maximizing $U(t_k)$ is equivalent to minimizing $c(t_k)$. The schedule delay cost is the disutility accumulated while an individual is not where his utility is greatest. When the individual arrives at work before t^*, disutility is incurred because utility from being at home before t^* is higher than utility at work. Similarly, utility is foregone when arriving at work after t^* because time is more valuable at work than at home. Function $\delta(t)$ is weakly decreasing for $t < t^*$ and weakly increasing for $t > t^*$. Trains that arrive close to t^* have small values of $\delta(t)$, and will sometimes be called *timely trains*. As shown in the next subsection, timely trains are more heavily used than other trains.

In Section 3, it is assumed that $\delta(t)$ has a piecewise linear form: $\delta(t_k) = \beta(t^* - t_k)$ if $t_k < t^*$, and $\delta(t_k) = \gamma(t_k - t^*)$ if $t_k \geq t^*$, where β and γ are respectively marginal disutilites from arriving early and late.[9] This specification, called "step preferences", is used in most studies of road traffic congestion and public transit crowding.

In the general case, a user taking train k with n_k users incurs a combined schedule delay and crowding disutility $c(t_k) = \delta(t_k) + g(n_k)$, $k = 1, ..., m$. To economize on writing, henceforth $\delta(t_k)$ is written δ_k and $c(t_k)$ is written c_k unless time dependence is required for clarity.

corridor, standing close to the door, etc..

[8] The elasticity is respectively positive, zero, or negative as $g(n)$ is convex, linear, or concave.

[9] This piecewise linear form arises when the utility flows from being at home and at work satisfy $u_h(t) - u_w(t) = \beta$ if $t_k < t^*$, and $u_h(t) - u_w(t) = -\gamma$ if $t_k \geq t^*$. This property is discussed by Tseng and Verhoef (2008).

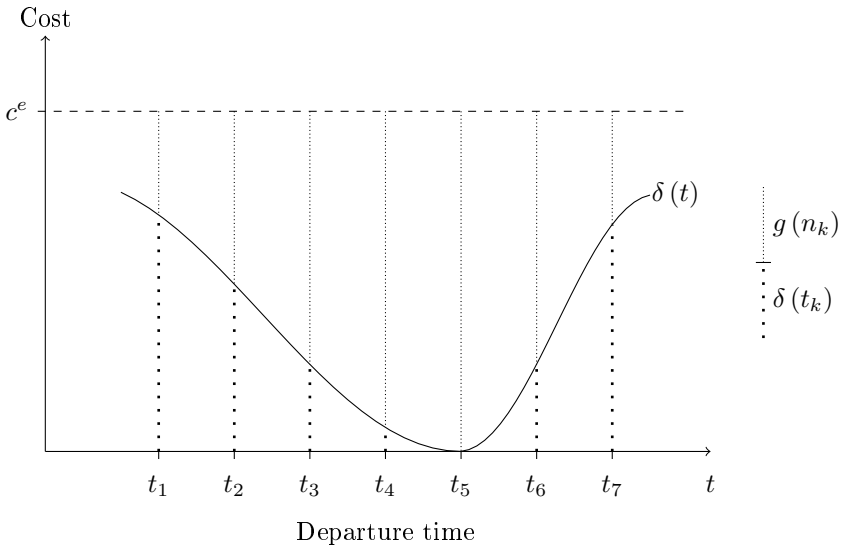

Figure 1: Schedule delay $\delta(t_k)$, crowding cost $g(n_k)$ and equilibrium cost c_e for seven trains, $t_5 = t^*$

2.1 User equilibrium

In this subsection we characterize user equilibrium when there is no fare. With N fixed, a fare would not affect either the division of users between trains or crowding costs. A uniform fare (i.e., independent of k) is introduced at the end of the subsection for the analysis of user equilibrium with elastic demand in Section 3.3.

Let superscript "e" denote the no-fare or user equilibrium (UE), and c^e the equilibrium trip cost which is to be determined. In UE, users distribute themselves between trains so that the user cost on every train is c^e. Hence, $\delta_k + g(n_k^e) = c^e$, $k = 1, ..., m$. Given $g'(.) > 0$, the inverse function $g^{-1}(.)$ exists, with $g^{-1}(0) = 0$ and $\left(g^{-1}\right)'(.) > 0$. The UE can therefore be solved for the n_k^e as a function of c^e:

$$n_k^e = g^{-1}(c^e - \delta_k), \quad k = 1, ..., m. \tag{1}$$

Since every user has to take some train, $\sum_{k=1}^m n_k^e = N$, or $\sum_{k=1}^m g^{-1}(c^e - \delta_k) - N = 0$. This equation implicitly determines a unique value of c^e. Figure 1 depicts a UE for seven trains ($m = 7$). Train $k = 5$ arrives on time and carries the most users.

Comparative statics properties of UE with respect to N are easily derived[10]

Proposition 1. *In equilibrium, user cost is an increasing function of N. It is convex, linear, or concave if $g(.)$ is respectively convex, linear, or concave.*

Similar to the the bottleneck model, user cost in the *PTC* model is an increasing function of total patronage, but the curvature of $c^e(N)$ differs. In the bottleneck model the curvature of $c^e(N)$ matches that of the schedule

[10] Equilibrium cost increases with the total number of passengers: $\partial c^e/\partial N = 1/\sum_{k=1}^m \left(g'(u_k)\right)^{-1} > 0$, where $u_k \equiv g^{-1}(c^e(N) - \delta_k)$. The second derivative, $\partial^2 c^e/\partial N^2$, has the same sign as $\sum_{k=1}^m g''(u_k)\left(g'(u_k)\right)^{-3}$ which depends on whether $g(.)$ is convex or concave.

delay cost function. With step preferences, the schedule delay cost function is linear and $c^e(N)$ is also linear. If the schedule delay cost function is convex (resp. concave), then $c^e(N)$ is convex (resp. concave).

By contrast, in the PTC model the curvature of $c^e(N)$ depends on the crowding cost function rather than the schedule delay cost function. This is because the train timetable is fixed in the short run, and users cannot travel earlier or later in response to growing demand. Furthermore, since each train's arrival time is fixed, the schedule delay cost incurred when taking a given train does not depend on N. The only way the service can accommodate additional demand is for trains to carry more passengers. Equilibrium user cost therefore increases at an increasing (resp. decreasing) rate with N if the marginal cost of crowding aboard a train increases (resp. decreases) with ridership.

User equilibrium in the PTC model is inefficient because users impose an external crowding cost on each other. The marginal social cost of a trip, MSC, is determined by differentiating the equilibrium total cost function, $TC^e = c^e \times N$, with respect to N: $MSC^e \equiv \partial TC^e/\partial N = c^e + (\partial c^e/\partial N)\, N$. The average marginal external cost of a trip is therefore $MEC^e \equiv MSC^e - c^e = (\partial c^e/\partial N)\, N$. With elastic demand (Section 3.3), transit is overused with a zero fare.[11] If the fare system is restricted to uniform fares, the fare should be set equal to the average external cost:

$$\tau^u = \frac{\partial c^e}{\partial N} N, \tag{2}$$

where superscript "u" denotes the optimal uniform fare. Total revenue from this fare is $R^u = \tau^u N$. The optimal uniform fare does not support the social optimum because the marginal external cost of crowding varies with train occupancy and it is larger on timely trains. As explained below, the social optimum can be achieved by levying train-specific fares.

2.2 Social Optimum

The social optimum (SO) differs from the UE because users are distributed between trains to equalize the marginal social costs of trips rather than their private costs. The marginal social cost of using train k is $MSC_k = \partial(c_k n_k)/\partial n_k = \delta_k + v(n_k)$, $k = 1, ..., m$, where $v(n_k) \equiv g(n_k) + g'(n_k)\, n_k$ is the marginal social crowding cost on train k. Let superscript "o" denote the SO. Total costs in the SO are $TC^o = \sum_{k=1}^{m} c_k n_k$, and the marginal social cost of a trip is $MSC^o = \partial TC^o/\partial N$. At the optimum, users are distributed across trains so that $MSC_k = MSC^o$ for every train:

$$\delta_k + v(n_k^o) = MSC^o,\ k = 1, ..., m. \tag{3}$$

Since $g'(.) > 0$ for $n > 0$, the marginal social crowding cost is always positive. In practice, it may not increase monotonically at all levels of ridership.[12] To facilitate analysis, however, we assume that $v'(.) > 0$. This is equivalent

[11] This might not be the case if transit is an alternative to driving and traffic congestion is severe.

[12] For example, $v(.)$ may drop when all seats are occupied and additional riders have to stand; see de Palma, Kilani and Proost (2015).

to assuming that $\varepsilon(n) > -2$.

Assumption 1. *The elasticity of $g'(n)$ with respect to n exceeds -2: $\varepsilon(n) > -2$.*

Assumption 1 is satisfied for all convex $g(.)$ functions and for all power function $g(n) \propto n^r$, $r > 0$. Given Assumption 1, the inverse function $v^{-1}(.)$ exists and it is increasing. Eq. (3) yields

$$n_k^o = v^{-1}(MSC^o - \delta_k). \tag{4}$$

Since all users must take some train in the SO, $\sum_{k=1}^{m} n_k^o = N$. Given Eq. (4), $\sum_{k=1}^{m} v^{-1}(MSC^o - \delta_k) - N = 0$, which implicitly determines a unique value of MSC^o. A counterpart to Prop. 1 then follows:

Proposition 2. *In the social optimum, the marginal social cost of a trip is an increasing function of N. It is convex, linear, or concave if $v(.)$ is respectively convex, linear, or concave.*

Comparing Prop. 2 with Prop. 1 it is clear that $v(.)$ plays the same role in shaping the SO as $g(.)$ does for the UE.[13] Prop. 2 contrasts again with the corresponding properties of the SO in the bottleneck model. For example, with linear schedule delay costs the marginal social cost of a trip in the bottleneck model is a linear function of N. In the PTC model it instead depends on the crowding cost function.

We now consider the distribution of ridership over trains. Intuition suggests that passenger loads are spread more evenly in the SO than the UE because smoother loads should reduce the total costs of crowding as discussed in de Palma, Kilani and Proost (2015). In fact, this is not invariably true but depends on how the marginal external crowding cost varies with usage. For any train, the marginal external crowding cost is

$$\frac{d(g'(n)\, n)}{dn} = g'(n)(1 + \varepsilon(n)).$$

The marginal external crowding cost increases with usage if $\varepsilon(n) > -1$, and decreases with usage if $\varepsilon(n) < -1$. The load patterns in the SO and UE are compared in[14]

Proposition 3. *If $\varepsilon(n) > -1$ ($\varepsilon(n) < -1$, respectively) the social optimum distribution of users across trains is a mean-preserving spread (respectively contraction) of the user equilibrium distribution of users across trains.*

The SO load pattern is a mean-preserving spread of the UE load pattern if the SO load pattern has more weight in the tails than the UE load pattern.[15] If the marginal external crowding cost increases monotonically with passenger load then $\varepsilon(n) > -1$.[16] If so, the marginal social costs of trips on two trains with unequal loads differ by more than their user costs. Consequently, the SO balance between crowding costs and schedule delay costs calls for

[13] Note that $v''(n) = 3g''(n) + ng'''(n)$. The marginal social cost of a trip can therefore be a convex function of N even if the user cost function is concave in N, and vice versa.

[14] Proofs of Prop. 3 and other results not established in the text are provided in the appendix.

[15] The relevant definition of MPS for discrete distributions is found in Rothschild and Stiglitz (1970), Subsection II.2, p. 229.

[16] Similar to Assumption 1, which is weaker, $\varepsilon(n) > -1$ is satisfied for all convex crowding cost functions, and crowding cost functions that belong to the class of power functions: $g(n) \propto n^r$, $r > 0$.

a smaller range of train loads than in the UE. Conversely, if $\varepsilon(n) < -1$, which is possible only if $g(.)$ is sufficiently concave,[17] then passenger loads are more peaked in the SO than the UE.

In summary, the difference between the SO and UE train loads depends on the curvature of the crowding cost function. According to most empirical studies, $g(.)$ is linear or convex (Wardman and Whelan, 2011; Haywood and Koning, 2015). If so, $\varepsilon(n) \geq 0$ and ridership in the UE is too concentrated on timely trains and should be spread out.

Regardless of whether the SO is more or less peaked than the UE, the SO usage pattern can be decentralized by charging a fare on train k equal to the marginal external cost of usage.[18] We will call the fare pattern the *SO-fare*. Given $MSC_k - c_k = g'(n_k) n_k$, the SO-fare is:

$$\tau_k^o = g'(n_k^o) n_k^o, \ k = 1, ..., m. \tag{5}$$

With this fare structure in place, users of train k incur a private cost equal to the social cost of a trip: $p_k^o = c_k^o + \tau_k^o = MSC^o, \ k = 1, ..., m$. The SO is more efficient than the UE because users are better distributed between trains. However, inclusive of the SO-fare users incur a higher private cost in the SO. To see this, note that at least one train is more crowded in the SO than the UE. Compared to the UE, in the SO a rider of that train incurs the same schedule delay cost but a higher crowding cost and a positive fare. Since all users incur the same private cost in the UE, and all users incur the same private cost in the SO, private costs are higher in the SO.[19]

Unless fare revenues are used to improve service in some way, charging fares to price crowding costs in the *PTC* model leaves users worse off.[20] By contrast, in the bottleneck model congestion pricing leaves private costs unchanged because the travel period is not affected. The bottleneck model therefore differs in the incidence of tolling costs.

Another property of the bottleneck model is that congestion toll revenue increases with N. This is because the average congestion externality increases with N, and hence so does the average toll. To determine how fare revenue in the *PTC* model varies with N, let R^o denote total revenue from the SO-fare. Now $R^o = \sum_{k=1}^{m} \tau_k^o n_k^o$, with n_k^o given in Eq. (4) and τ_k^o in Eq. (5). Revenue from the optimal uniform fare is $R^u = \tau^u N$. We have

Proposition 4. *Let $i = u, o$ index the pricing regime. Then,*

$$\frac{\partial R^i}{\partial N} = \frac{\partial MSC^i}{\partial N} N.$$

[17] For example, inequality $\varepsilon(n) < -1$ holds for the function $g(n) = c_0 + c_1 \ln(n) - kn$ for $c_0 > k$ and over the range $n \in [1, c_1/k)$.

[18] The fare is set according to Pigouvian principles. Revenue generation or other goals are ruled out.

[19] The difference in private cost is, however, smaller than the average fare paid because the social (i.e., resource) costs of travel are lower in the SO.

[20] This is also true of pricing road traffic congestion in Henderson's (1974) model although the physical effects of tolling differ. In his model, tolling causes the departure period to spread out, and the first and last users incur higher schedule delay costs than in the UE. Because the first and last users incur no congestion delay in either the UE or the SO, their costs are higher in the SO. Since all users incur the same private costs in the UE and SO, equilibrium private user costs are increased by tolling.

Prop. 4 reveals that in each fare regime fare revenue increases if the marginal social cost of a trip increases with total usage. This will be the case unless the crowding cost function is sufficiently concave.

Next, we examine how the welfare gain from implementing the SO-fare varies with usage. Let $G^{eo} \equiv TC^e - TC^o$ denote the welfare gain in shifting from the UE to the SO. Intuition suggests that G^{eo} increases with N: first because crowding becomes more onerous for each user, and second because more users suffer the increased cost. However, we have shown that the rate at which the cost of crowding increases with load depends on the curvature of the crowding cost function. It turns out that properties of the crowding cost function also govern how G^{eo} depends on N.

Consider the following assumptions:

Assumption 2. *The marginal external cost of crowding increases with load:* $\varepsilon(n) > -1$.

Assumption 3a. *The marginal social cost of crowding is a strictly convex function of load (i.e.,* $v''(n) > 0$*), and* $\varepsilon(n)$ *is a nonincreasing function of load (i.e.,* $\frac{d\varepsilon(n)}{dn} \leq 0$*).*

Assumption 3b. *The marginal social cost of crowding is a strictly concave function of load (i.e.,* $v''(n) < 0$*), and* $\varepsilon(n)$ *is a nondecreasing function of load (i.e.,* $\frac{d\varepsilon(n)}{dn} \geq 0$*).*

Assumption 3a holds if $g(n) \propto n^r$, $r \geq 1$, and Assumption 3b holds if $g(n) \propto n^r$, $0 < r < 1$. The effect of total ridership on the welfare gain from the SO-fare is described in

Proposition 5. *Let Assumption 2 hold. The welfare gain from the SO-fare,* G^{eo}*, decreases with* N*, increases with* N*, or is independent of* N *if Assumption 3a holds, Assumption 3b holds, or if* $g(\cdot)$ *is linear, respectively.*

Proposition 5 identifies conditions under which G^{eo} increases, decreases, or is independent of total ridership. Since the conditions are not collectively exhaustive, Prop. 5 does not establish the direction of change for all cases. Nevertheless, the conditions span a broad set of functions.

As noted earlier, most empirical studies find that $g(.)$ is linear or convex. According to Prop. 5, G^{eo} is then either constant or (if $d\varepsilon(n)/dn \leq 0$) a decreasing function of N. This is a surprise since it goes against the intuition described above. To understand why, note that the welfare gain derives from reallocating users between trains as in Prop. 3. If $g(.)$ is convex, users are reallocated more evenly. Since the difference in crowding costs between two trains equals the difference in schedule delay costs, the marginal benefit from starting to reallocate users is independent of N. However, as N increases the marginal crowding cost on each train becomes higher and the UE and SO train loads become more similar. Consequently, the *amount* of user reallocation decreases, and the total welfare gain therefore falls as well. The argument runs in reverse if $g(.)$ is concave since the optimal amount of reallocation then increases with N.[21]

[21] Another way to view Prop. 5 is in terms of the marginal social cost of usage, which is MSC^e in the UE and MSC^o in the SO. If $MSC^o < MSC^e$, an additional user causes total costs to rise by less in the SO than the UE, and G^{eo} rises. Conversely, if $MSC^o > MSC^e$, total costs rise more in the SO and G^{eo} falls. Thus, if $g(.)$ is convex an additional user is, paradoxically, more costly

Several empirical studies have found that schedule delay and crowding costs are close to linear (see Wardman et al., 2012, for scheduling cost, and Wardman and Whelan, 2011; Haywood and Koning, 2015, for crowding cost). Taking advantage of this evidence, for most of the balance of the paper we focus on a particular instance of the model in which the crowding cost function is linear. In Sections 4 and 5 we assume that the schedule delay cost function is linear as well. With linearity, the model can be readily extended to allow elastic demand and the optimal number of trains and train capacity can be characterized as well. Linearity also facilitates comparisons with the bottleneck model.

3 Linear crowding costs

Assume that $g(n) = \lambda n/s$, where $s > 0$ is a measure of train capacity, and $\lambda > 0$. The marginal social crowding cost function is then $v(n) = 2\lambda n/s$: twice the private crowding cost. The cost of taking train k is therefore $c_k = \delta_k + \lambda n_k/s$, $k = 1, ..., m$.

3.1 User equilibrium

Define $\overline{\delta} \equiv \frac{1}{m}\sum_{k=1}^{m} \delta_k$ as the unweighted average scheduling cost for trains. Eq. (1) and linear crowding costs lead to

Proposition 6. *In the uniform-fare equilibrium with linear crowding costs, train k carries a load $n_k^e = N/m + s\left(\overline{\delta} - \delta_k\right)/\lambda$, and user cost is $c^e = \overline{\delta} + \lambda N/(ms)$.*

For given values of m and s, equilibrium trip cost is a linear increasing function of ridership, N, as in the bottleneck model. This is a consequence of the assumptions here that the train timetable is fixed and crowding costs are linear.

As in the general model, timely trains carry more users than other trains. The difference in loads between two successive trains is proportional to parameter s, and inversely proportional to λ. Because the first (or last) train carries the fewest passengers, the solution satisfies all the non-negativity constraints $n_k^e > 0$ if $n_1^e > 0$ and $n_m^e > 0$:

$$N > \frac{ms}{\lambda}\left(\max\left[\delta_1, \delta_m\right] - \overline{\delta}\right). \tag{6}$$

Since service is costly to provide, condition (6) is satisfied when m and s are chosen optimally as in Section 4.

Aggregate travel costs are described in

to accommodate in the SO than in the UE even though users are distributed optimally between trains in the SO. If $g(.)$ is linear, $MSC^o = MSC^e$ and the difference in total costs between UE and SO is independent of N. In effect, the benefits of internalizing the crowding cost externality are exhausted once total usage is large enough for all trains to be used. We will illustrate this case diagrammatically in Section 3.

Proposition 7. *In the uniform-fare equilibrium with linear crowding costs, total schedule delay costs, SDC, total crowding costs, TCC, and total travel costs net of the fare, TC, are given by*

$$SDC^e = \overline{\delta}N - 4RV^o, TCC^e = \frac{\lambda N^2}{ms} + 4RV^o,\ TC^e = \overline{\delta}N + \frac{\lambda N^2}{ms},$$

where $RV^o \equiv \frac{s}{4\lambda}\left(\sum_{k=1}^{m}\delta_k^2 - \left[\sum_{k=1}^{m}\delta_k\right]^2/m\right)$.

As shown below, RV^o is the variable revenue from the SO-fare. Note that $RV^o > 0$ by the Cauchy-Schwarz inequality. Crowding costs are analogous to travel time costs (TTC) in traffic congestion models. In the bottleneck model equilibrium, $SDC^e = TTC^e$ for all values of N. In the PTC model the behavior of SDC^e and TCC^e is more complicated. Total schedule delay costs are lower than if users were equally distributed across trains (in which case $SDC^e = \overline{\delta}N$), and total crowding costs are higher by the same amount. This is because users crowd onto timely trains that arrive closer to t^*. For small values of N, only one train is used. Schedule delay costs are zero (if $t_1 = t^*$) while crowding costs are proportional to N^2. For a given value of m, $m > 1$, SDC^e is a linear increasing function of N with a negative intercept, while TCC^e increases with N^2 and has a positive intercept.

3.2 Social Optimum

The social optimum is readily derived using results for the general model in subsection 2.2. Given $v(n) = 2\lambda n/s$, $v^{-1}(x) = sx/(2\lambda)$. Eqs. (4) and (5) give

Proposition 8. *In the social optimum with linear crowding costs, train k caries a load of* $n_k^o = N/m + s\left(\overline{\delta} - \delta_k\right)/(2\lambda)$. *The optimum can be decentralized by charging a fare for train k of* $\tau_k^o = \lambda n_k^o/s$. *The marginal social costs of trips are the same for all trains and equal to* $\overline{\delta} + 2\lambda N/(ms)$.

According to Prop. 8, the marginal social cost of a trip is a linear increasing function of N. Assumption 2 holds for the linear crowding cost function. Hence, by Prop. 3 train loads are more evenly distributed in the social optimum than the uniform-fare equilibrium. The difference in loads between successive trains is only half as large. The non-negativity constraint on usage of all trains is satisfied if $N > ms\left(\max\left[\delta_1, \delta_m\right] - \overline{\delta}\right)/(2\lambda)$. This condition is satisfied if condition (6) is satisfied for the no-fare equilibrium. Compared to the uniform fare in Eq. (2),[22] the SO-fare is lower on the earliest and latest trains with $\delta_k > \overline{\delta}$, and higher on timely trains with $\delta_k < \overline{\delta}$.

Given Eqs. (4) and (5), total revenue from the SO-fare is $R^o = \lambda N^2/(ms) + RV^o$, where

$$RV^o \equiv \frac{s}{4\lambda}\left(\sum_{k=1}^{m}\delta_k^2 - m\overline{\delta}^2\right). \tag{7}$$

The first term in R^o matches revenue from an optimal uniform fare, R^u. The second term, RV^o, is extra revenue (when $m > 1$) due to variation of the fare. As noted in subsection 3.1 this is called *variable revenue*. A notable

[22] With linear crowding costs, $\tau^u = \lambda N/(ms)$, and $R^u = \lambda N^2/(ms)$.

feature of (7) is that variable revenue is independent of N. This property is discussed below in connection with the welfare gain from imposing the SO-fare. Aggregate costs in the SO are described in

Proposition 9. *In the decentralized social optimum with linear crowding costs, total schedule delay costs, SDC, total crowding costs, TCC, and total travel costs net of the fare, TC, are given by*

$$SDC^o = SDC^e + 2RV^o,\ TTC^o = TTC^e - 3RV^o,\ TC^o = TC^e - RV^o.$$

Total schedule delay costs are higher in the social optimum than the no-fare equilibrium, but crowding costs are smaller and total costs are lower by an amount equal to variable fare revenue. With linear crowding costs, the welfare gain from imposing the SO-fare is therefore equal to variable revenue: $G^{eo} = RV^o$. Prop. 9 can be compared with the corresponding formulas in the bottleneck model, denoted with a subscript Bn: $SDC^o_{Bn} = SDC^e_{Bn}$, $TTC^o_{Bn} = 0$, $RV^o_{Bn} = TTC^e_{Bn}$, and $TC^o_{Bn} TC^e_{Bn} - RV^o_{Bn}$.

Tolling in the bottleneck model eliminates queuing (the counterpart to crowding in the PTC model) without increasing total schedule delay costs. Variable revenue matches total queuing costs in the UE, and total costs are reduced by variable revenue. Thus, in both models variable revenue measures the welfare gain from tolling, but tolling is more effective in the bottleneck model because it eliminates the external cost of congestion without causing schedule delay costs to increase.

The numerical example in Section 5 features linear schedule delay costs and a constant headway, h, between trains. Given $G^{eo} = RV^o$, it is straightforward to show that for large values of m,

$$G^{eo} \simeq \frac{s}{48\lambda} \left(\frac{\beta\gamma}{\beta+\gamma} \right)^2 h^2 m \left(m^2 - 1 \right). \tag{8}$$

Eq. (8) reveals how the welfare gain from the SO-fare varies with parameters. First, as noted in discussing Prop. 5 above, G^{eo} is independent of total usage, N. To see why, consider a simple case with two trains.[23] The cost of using train k is $c_k = \delta_k + g n_k$ where $g \equiv \lambda / s$ measures the rate at which crowding costs increase with train load. Figure 2 depicts the UE and SO using a diagram with two vertical axes separated by N.[24] Usage of train 1 is measured to the right from the left-hand axis, and usage of train 2 to the left from the right-hand axis. By assumption, $\delta_2 > \delta_1$ so that train 1 is overused in the UE. The welfare gained in shifting users from train 1 to train 2 is shown by the triangular shaded area. The height of the triangle is $\delta_2 - \delta_1$, and the width of the triangle is $(\delta_2 - \delta_1) / (2g)$. The area of the triangle is therefore $(\delta_2 - \delta_1)^2 / (4g)$. It does not depend on N because neither dimension of the triangle depends on N. The height of the triangle equals the difference in marginal external costs of using the two trains in the UE. This is determined by the difference in their attractiveness, $\delta_2 - \delta_1$, not N. The width of the triangle is the optimal number of users to redistribute between trains which is proportional to $\delta_2 - \delta_1$, and inversely proportional

[23] This part of the discussion does not depend on Eq. (8), which is accurate only for large values of m.

[24] Figure 2 is analogous to Figure 3 in Arnott, de Palma and Lindsey (1990) which depicts a setting in which drivers choose between two routes that differ in free-flow travel time costs.

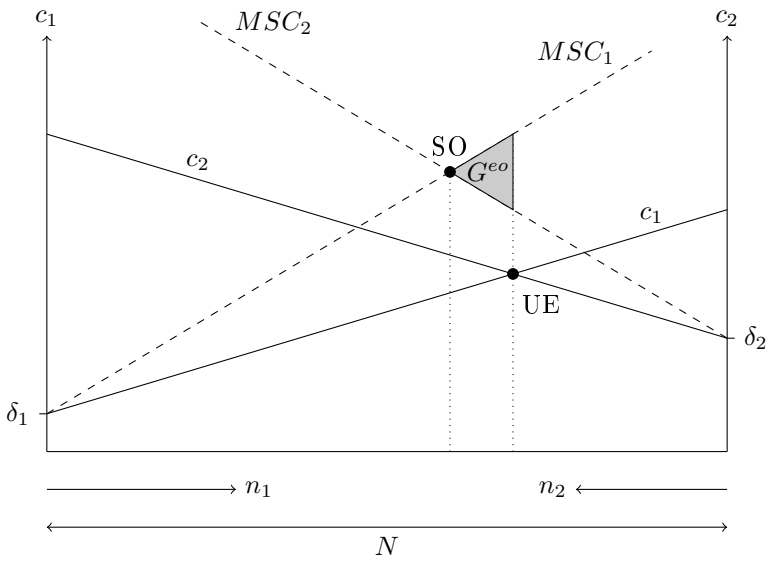

Figure 2: User equilibrium (UE), social optimum (SO) and welfare gain (G^{eo}) with two trains

to g. This, too, is independent of N.

For large values of m, Eq. (8) reveals that G^{eo} varies with the square of β and γ together. This is consistent with the quadratic dependence of G^{eo} on the schedule delay costs, δ_1 and δ_2, in the example with $m = 2$. G^{eo} varies with the square of the headway, h, for the same reason. G^{eo} varies inversely with the ratio $g = \lambda/s$ because the scope to alleviate crowding by redistributing riders between trains decreases if trains become crowded more quickly.

Finally, G^{eo} varies approximately with the cube of the number of trains. This highly nonlinear dependence is due to two multiplicative factors. First, the average schedule delay cost of trains is proportional to m. The average difference in schedule delay costs is therefore proportional to m, and the welfare gain from redistributing passengers between two trains varies with m^2. Second, the number of trains between which passenger loads can gainfully be redistributed is approximately proportional to m. Hence, the overall welfare gain varies approximately with m^3.

In the introduction to the paper we noted that the distribution of passengers between trains is governed by the trade-off users face between scheduling costs and crowding costs. It is therefore surprising that the parameters measuring the strength of these two costs have contrasting effects on the welfare gain from congestion pricing. According to Eq. (8), doubling the unit costs of schedule delay, β and γ, increases the welfare gain four-fold. By contrast, doubling the crowding cost parameter, λ, reduces the gain by half. In assessing the potential benefits from implementing congestion pricing, it is therefore important to predict how the parameter values will evolve over time. If parameters β, γ and λ all grow at a rate r, G^{eo} will grow at rate r too. By contrast, if work hours become more flexible in the future, β and γ could stagnate while λ continues to rise. Other things equal, G^{eo} would then decline.

3.3 Elastic demand

So far it has been assumed that transit ridership, N, is exogenous. In practice, travelers can often use other transport modes and they may choose to forego travel if it is too costly. To admit these possibilities we now assume that demand for public transport trips is a smooth and decreasing function of the private cost:

$$N = N(p), \ \frac{\partial N}{\partial p} < 0. \tag{9}$$

Consumers' surplus from trips is $CS(p) = \int_p^\infty N(u)\,du$, and social surplus (gross of capacity costs) is the sum of consumers' surplus and fare revenue: $SS(p,\tau) = CS(p) + R$.

The analysis parallels the comparisons in Arnott et al. (1993) for the bottleneck model. Let superscript n denote the no-fare regime and a hat ($\hat{}$) denote an equilibrium value with elastic demand. The equilibrium values in the no-fare, uniform-fare, and SO-fare regimes are compared in

Proposition 10. *For given values of m and s, linear crowding costs, and elastic demand, equilibrium private costs are the same in the SO-fare and optimal uniform-fare regimes, and lower in the no-fare regime: $\widehat{p}^o = \widehat{p}^u > \widehat{p}^n$.*

Equilibrium usage is the same in the SO-fare and optimal uniform-fare regimes, and higher in the no-fare regime: $\widehat{N}^o = \widehat{N}^u < \widehat{N}^n$.

Consumers' surplus is the same in the SO-fare and optimal uniform-fare regimes, and higher in the no-fare regime: $\widehat{CS}^o = \widehat{CS}^u < \widehat{CS}^n$.

Social surplus is highest in the SO-fare regime, intermediate in the optimal uniform-fare regime, and lowest in the no-fare regime: $\widehat{SS}^o > \widehat{SS}^u > \widehat{SS}^n$. Consequently, $\widehat{G}^{no} > \widehat{G}^{nu} > 0$, and $\widehat{G}^{no} > \widehat{G}^{uo} > 0$.

The results in Prop. 10 differ from those in the bottleneck model (Arnott et al., 1993). In the bottleneck model, the equilibrium price of a trip for a given N is the same in the social optimum and no-toll user equilibrium, and higher in the uniform-toll equilibrium. Consequently, $\widehat{p}^u_{Bn} > \widehat{p}^o_{Bn} = \widehat{p}^n_{Bn}$ which contrasts with $\widehat{p}^u = \widehat{p}^o > \widehat{p}^n$ in Prop. 10. The rankings of usage and consumers' surplus also differ, and only the rankings of social surplus and the welfare gain from pricing are the same.

4 Optimal transit service

We now turn attention to the long run when the transit authority can choose m, s, and the timetable for the m trains. For tractability, we assume that schedule delay costs are linear. We also assumed that the headway between trains is a given constant, h, which is reasonable if the headway is set to the shortest technologically feasible interval consistent with safe operations. First we derive the optimal timetable for given values of m and s, and then the derive properties of the optimal m and s for a general capacity cost function. Finally, we adopt a specific capacity function and derive analytical formulas for the optimal m and s while treating m as a continuous variable.

4.1 Optimal timetable

The optimal timetable is derived by minimizing users' total costs. In general, the optimal timetable for given m and s is not the same for the UE and the SO because their load patterns differ. However, the timetables are equal given linear schedule delay costs and a uniform headway.

Since the timetable consists of m successive trains with a constant headway h, the timetable is fully described by the arrival time of the last train, t_m. Let 1_x be the indicator function with $1_x = 1$ if x is true, and $1_x = 0$ otherwise.The optimal value of t_m is described in

Proposition 11. *With the optimal timetable the last train leaves at time* $t_m^o = t^* + h\left(m - \varphi_m - \mathbb{1}_{\frac{\gamma m}{\beta+\gamma} > \varphi_m}\right)$ *where* $\varphi_m \equiv \left\lfloor \frac{\gamma m}{\beta+\gamma} + \frac{1}{2} \right\rfloor$. *Train* k, *with* $k = \varphi_m + \mathbb{1}_{\frac{\gamma m}{\beta+\gamma} > \varphi_m}$ *arrives on time at* t^*. *The unweighted average schedule delay cost is* $\overline{\delta} \simeq \frac{\beta\gamma}{\beta+\gamma}\frac{mh}{2}$.

According to Prop. 11, the higher is the unit cost of late arrival (γ) relative to early arrival (β) the earlier train service begins. The fraction of trains that arrive before t^* is approximately $\gamma/(\beta+\gamma)$. This formula is approximate because the number of trains is integer-valued. For the same reason, the formula for average schedule delay cost, $\overline{\delta}$, is approximate too.

4.2 General capacity function

Let $K(m,s)$ denote the cost of providing service including capital, operations, and maintenance.[25] To facilitate analysis, for the remainder of this section we treat m as a continuous variable. (The formula for $\overline{\delta}$ in Prop. 11 is then exact.) Function $K(m,s)$ is assumed to be a strictly increasing and differentiable function of m and s. As in Section 3.3, we first consider the uniform-fare regimes and then the social optimum.

Uniform-fare regimes.- Let superscript e denote a generic uniform-fare regime which includes the no-fare and optimal uniform-fare regimes as special cases. Let $p(N)$ denote the inverse demand curve corresponding to demand function (9). With a uniform fare, social surplus net of capacity costs is

$$SS^e = \int_{n=0}^{N} p(n)\,dn - \left(\bar{\delta}N + \frac{\lambda N^2}{ms} + K(m,s)\right).$$

The transit authority chooses m, s, and the toll, τ, to maximize SS^e. For generality we allow the toll to depend on N, m, and s in an arbitrary way. To economize on notation, let K_m and K_s denote the derivatives of $K(m,s)$ with respect to m and s respectively. In addition, define the composite variables

$$D^e \equiv \frac{p_N N - \tau - \frac{d\tau}{dN}N}{p_N N - \frac{\lambda N}{ms} - \frac{d\tau}{dN}N},\ A_s^e \equiv \frac{\left(\tau - \frac{\lambda N}{ms}\right)\frac{d\tau}{ds}N}{p_N N - \frac{\lambda N}{ms} - \frac{d\tau}{dN}N}, \text{ and } A_m^e \equiv \frac{\left(\tau - \frac{\lambda N}{ms}\right)\frac{d\tau}{dm}N}{p_N N - \frac{\lambda N}{ms} - \frac{d\tau}{dN}N}.$$

[25] System costs are assumed to be independent of usage. Adding N as an argument of the service cost function would not affect results of interest.

First-order conditions for a maximum of SS^e are

$$\text{For } s \quad : \quad \frac{\lambda N^2}{ms^2} \cdot D^e + A_s^e = K_s, \tag{10a}$$

$$\text{For } m \quad : \quad \left(\frac{\lambda N}{m^2 s} - \frac{\partial \bar{\delta}}{\partial m}\right) N \cdot D^e + A_m^e = K_m. \tag{10b}$$

The right-hand side of (10a) is the marginal cost of expanding train capacity, and the left-hand side is the marginal benefit. The first term of the product on the left-hand side is the marginal benefit from expanding train capacity if usage remained fixed. The cost of crowding would decrease by $\lambda N/\left(ms^2\right)$ for each of the N users. If $\tau < \lambda N/(ms)$ and $p_N < \infty$, $D^e < 1$ and the actual reduction in crowding cost is smaller than this because the improved service quality attracts new users who value trips less than their marginal social cost. This is the induced or latent demand effect familiar in the context of road traffic congestion (e.g., Duranton and Turner, 2011). The other term on the left-hand side of (10a), A_s^e, describes the effect of a change in toll. If $\tau < \lambda N/(ms)$, and expanding train capacity induces a reduction in the toll (i.e., $d\tau/ds < 0$), then $A_s^e < 0$. Lowering the toll exacerbates the effect of latent demand, and the marginal benefit from expanding train is reduced further.

Equation (10b) has a similar interpretation to Eq. (10a). The right-hand side of (10b) is the marginal cost of adding a train, and the left-hand side is the marginal benefit.[26] The first term inside the brackets on the left-hand side is the marginal benefit per user from less crowding. The second term inside the brackets is the marginal disbenefit due to greater average schedule delay costs. The derivative $\partial\bar{\delta}/\partial m$ given in Prop. 11 is constant. This net benefit is diluted by the same factor, D^e, as in Eq. (10a). Term A_m^e again describes the effect of a change in toll. If $\tau < \lambda N/(ms)$, and expanding the number of trains induces a reduction in the toll (i.e., $d\tau/dm < 0$), then $A_m^e < 0$.

In the no-fare regime, $\tau = 0$, $D^e = \frac{p_N N}{p_N N - \frac{\lambda N}{ms}} < 1$, $A_s^e = 0$, and $A_m^e = 0$. Eqs. (10a) and (10b) lead to

Proposition 12. *In the no-fare regime, optimal train capacity, s, and number of trains, m, are defined by*

$$\frac{\lambda N^2}{ms^2} \cdot \frac{p_N N}{p_N N - \frac{\lambda N}{ms}} = K_s, \tag{11a}$$

$$\left(\frac{\lambda N}{m^2 s} - \frac{\partial \bar{\delta}}{\partial m}\right) N \cdot \frac{p_N N}{p_N N - \frac{\lambda N}{ms}} = K_m. \tag{11b}$$

With no toll, expanding capacity has no secondary effect in reducing the toll but latent demand acts in full force. Indeed, in the limit of perfectly elastic demand (i.e., $p_N \to 0$), the potential benefit from expanding s or m is completely dissipated.

With the optimal uniform fare, $\tau = \lambda N/(ms)$, $D^e = 1$, $A_s^e = 0$, and $A_m^e = 0$. Eqs. (10a) and (10b) give

[26] Since $K_m > 0$, the RHS of Eq. (10b) is positive. Provided $(\tau - \lambda N/(ms))\, d\tau/dm > 0$, this guarantees that $\lambda N/\left(m^2 s\right) - \partial\bar{\delta}/\partial m > 0$, or $N > \partial\bar{\delta}/\partial m \times sm^2/\lambda$. It is not clear that this condition is sufficient to guarantee condition (6), $n_k^e \geq 0$, for all k. However, m is treated in this section as a continuous variable. If m is restricted to integer values, the optimal number of trains is derived by increasing m in steps of one until the incremental net benefit becomes negative. With such a procedure, condition (6) is satisfied at the optimum.

Proposition 13. *In the optimal uniform-fare regime, optimal train capacity, s, and number of trains, m, are defined by the conditions*

$$\begin{aligned} \frac{\lambda N^2}{ms^2} &= K_s, \\ \left(\frac{\lambda N}{m^2 s} - \frac{\partial \bar{\delta}}{\partial m}\right) N &= K_m. \end{aligned}$$

In contrast to Prop. 12, in Prop. 13 the marginal benefits from expanding train capacity and the number of trains are not diluted by additional usage because usage is priced efficiently. This might suggest that the optimal values of s and m, s_*^u and m_*^u, are larger than their counterparts with a zero fare, s_*^n and m_*^n. However, at least for given values of s and m, usage is higher in the no-fare regime as per Prop. 10. This leaves the rankings of s_*^u and s_*^n, and m_*^u and m_*^n, ambiguous in general. Moreover, unlike in the bottleneck model it is not possible as in Arnott et al. (1993) to derive simple rankings in terms of the elasticity of demand. This is because capacity has two dimensions (m and s) rather than one, and also because the user cost in Prop. 6 has a fixed component that is independent of usage.

Social optimum.- With the SO-fare, social surplus net of capacity costs is given by

$$SS^o = \int_{n=0}^{N} p(N) - \left(\bar{\delta} N + \frac{\lambda N^2}{ms} + K(m,s)\right) + RV^o(m,s).$$

SS^o is the same as SS^e except for the last term, RV^o, which is a function of m and s, but does not depend on usage. In effect, net financial system costs in the social optimum are $K(m,s) - RV^o(m,s)$. Since usage is priced efficiently in both the SO-fare and optimal uniform-fare regimes, the first-order conditions for s_*^o and m_*^o are the same as in Prop. 13 for s_*^u and m_*^u, with the derivatives of $K(m,s) - RV^o(m,s)$ in place of the derivatives of $K(m,s)$. We have

Proposition 14. *In the SO-fare regime, optimal train capacity, s, and number of trains, m, are defined by the conditions*

$$\frac{\lambda N^2}{ms^2} = K_s - RV_s^o, \tag{12a}$$

$$\left(\frac{\lambda N}{m^2 s} - \frac{\partial \bar{\delta}}{\partial m}\right) N = K_m - RV_m^o, \tag{12b}$$

where RV_s^o and RV_m^o denote the derivatives of $RV^o(m,s)$ with respect to s and m respectively.

The right-hand sides of Eqs. (12a) and (12b) are smaller than their counterparts for the optimal uniform fare displayed in Prop. 13. The generation of variable revenue from the SO-fare effectively reduces the marginal financial cost of expanding either s or m. In the case of Eq. (12a) this implies that optimal train capacity conditional on the values of m and N is larger in the social optimum: $s_*^o(m,N) > s_*^u(m,N)$. Similarly, Eq. (12b) implies

that the optimal number of trains conditional on the values of s and N is also larger in the social optimum: $m_*^o(s,N) > m_*^u(s,N)$.

These rankings may seem surprising given that total system costs are lower in the social optimum than the uniform-fare regime. Inequality $m_*^o(s,N) > m_*^u(s,N)$ is explained by the fact that ridership is distributed more evenly across trains in the social optimum. More users take the earliest and latest trains in the social optimum which makes adding extra trains more beneficial. To understand the inequality $s_*^o(m,N) > s_*^u(m,N)$, recall from Eq. (8) that in the uniform-fare regime the deadweight loss from imbalanced ridership between trains *increases* with s. Expanding capacity is therefore more valuable in the social optimum.

Despite the inequalities $m_*^o(s,N) > m_*^u(s,N)$ and $s_*^o(m,N) > s_*^u(m,N)$, there is no guarantee that the unconditionally optimal values (s_*^o, m_*^o) in the social optimum are both larger than their counterparts (s_*^u, m_*^u). One reason is that $s_*^u(m,N)$ is a decreasing function of m, and $m_*^u(s,N)$ is a decreasing function of s, and one function can shift much more than the other. The other reason is that usage generally differs in the two regimes; i.e. $N_*^o \neq N_*^u$. To proceed further, we now adopt a specific capacity function.

4.3 A specific capacity function

Kraus and Yoshida (2002) distinguish in their model between the number of train runs and the number of train sets (a train set can make more than one run). They also account for the time required for a train set to make a round trip. These variables are absent from our model, and we adopt a simpler service cost function for transit of the form:

$$K(m,s) = (\nu_0 + \nu_1 s)\, m + \nu_2 s, \tag{13}$$

where ν_0, ν_1, and ν_2 are all non-negative parameters. The term $\nu_0 + \nu_1 s$ in (13) is the incremental capital and operating costs of running an additional train. It is a linear increasing function of train capacity. If $\nu_0 > 0$, there are scale economies with respect to train size. The second term in (13), $\nu_2 s$, accounts for costs that depend on train capacity but not the number of trains. Kraus and Yoshida (2002) interpret this term as capital costs for terminals.[27] In this subsection we focus on the optimal uniform fare and SO-fare because in these regimes the slope of the demand function does not affect the optimal values of m or s, and properties of the solution can be derived while treating N parametrically.

Optimal uniform fare.- With the optimal uniform fare, the first-order conditions for s and m are given by Prop. 13. Given the service cost function (13), these equations become

$$\frac{\lambda N^2}{m s^2} = \nu_1 m + \nu_2, \tag{14a}$$

$$\left(\frac{\lambda N}{m^2 s} - \frac{\partial \bar{\delta}}{\partial m}\right) N = \nu_0 + \nu_1 s. \tag{14b}$$

[27] They note that the linear specification is applicable if terminal cost is proportional to terminal area, and terminal area is proportional to train capacity.

Before solving (14a) and (14b) simultaneously, it is instructive to consider each equation by itself. Eq. (14b) can be solved for the conditionally-optimal number of trains considered in the previous subsection: $m_*^u(s,N) = \sqrt{\lambda/\left[s\left(N\partial\bar{\delta}/\partial m + \nu_0 + \nu_1 s\right)\right]}N$. This expression is of practical interest if the transit authority cannot adjust train capacity - perhaps because train platforms cannot be lengthened.[28] As expected, the optimal number of trains decreases with train capacity. The number of trains increases with demand at a rate faster than $\sqrt{N}$, but slower than N. Service quality degrades because both the duration of the travel period and average train occupancy increase.

First-order condition (14a) can be solved to obtain a formula for conditionally-optimal train capacity: $s_*^u(m,N) = \sqrt{\lambda/(\nu_1 m^2 + \nu_2 m)}N$. Optimal train capacity decreases with m at a rate faster than $m^{-1/2}$. It also varies proportionally with N. Thus, if the transit authority cannot add trains but can introduce bigger trains, it adds sufficient capacity to maintain a given level of crowding on each train. Service quality remains constant. In this sense, users fare better if the transit authority can only expand train capacity than if it can only add more trains.[29]

Eqs. (14a) and (14b) can be solved jointly to obtain quartic equations for the unconditionally optimal values, s_*^u and s_*^u.[30] The characteristics of the solution depend on the relative magnitudes of parameters ν_0 and ν_1.[31] If $\nu_1 = 0$, the cost of a train is independent of its capacity:

$$\begin{aligned} s_*^u(N) &= \left(\frac{\partial\bar{\delta}}{\partial m}\frac{\lambda}{\nu_2^2}N^3 + \frac{\lambda\nu_0}{\nu_2^2}N^2\right)^{1/3}, \\ m_*^u(N) &= \frac{\lambda}{\nu_2 \times s_*^{u2}(N)}N^2. \end{aligned}$$

According to Mohring's (1972) square-root rule, both optimal service frequency and the number of passengers carried per train (or bus) increase with $\sqrt{N}$. In the PTC model, service frequency is constant because headway is fixed. s_*^u rises with N at a rate faster than $N^{2/3}$,[32] and m_*^u grows at a rate slower than $N^{2/3}$, but since it does increase with N the duration of the travel period increases. As N becomes very large, m_*^u approaches a constant value and s_*^u increases approximately linearly with N.[33, 34] With $\nu_1 = 0$, it is possible to show that equilibrium user cost is a U-shaped function of N with a minimum at $N = \nu_0\left(\partial\bar{\delta}/\partial m\right)^{-1}$. However, both the equilibrium price, p, and the average system cost, $c^u(m_*^u, s_*^u) + K(m_*^u, s_*^u)/N$, decline monotonically with N. This is attributable to the fact that, with $\nu_1 = 0$, the service cost function has constant returns to scale while the user cost function has

[28] Train platforms may also have to be adjusted to accommodate wider trains: a problem that the French rail network, SNCF, has overlooked (Willsher, 2014).

[29] This might not be true if the headway between trains can be reduced.

[30] The equations are $\nu_1(s_*^u)^4 + Z(s_*^u)^3 - \lambda Z^2 N^2/\nu_2^2 = 0$, and $\nu_1(m_*^u)^4 + \nu_2(m_*^u)^3 - \lambda\nu_2^2 N^2/Z^2 = 0$, where $Z \equiv N\partial\bar{\delta}/\partial m + \nu_0$.

[31] If $\nu_2 = 0$, there would be no fixed costs of expanding train capacity such as train station infrastructure. Costs would be minimized by reducing m toward zero while increasing s proportionally to maintain ms constant. This would imply operating one train large enough to accommodate all passengers and setting the timetable so that everyone arrives on time which is implausible.

[32] In Kraus and Yoshida's (2002) model the effect of N on s is ambiguous. Neverthless, they remark (p.178) that "with realistic parameters" s is likely to increase with N.

[33] Eventually a physical limit to train capacity would be reached due to constraints on platform size or tractive power.

[34] If $\nu_0 = 0$ as well as $\nu_1 = 0$, m_u^* is independent of N, and s_u^* rises proportionally with N for all values of N. This is a highly unrealistic case since it means that procuring and operating trains is costless.

increasing returns.

The limiting case $\nu_0 = 0$ applies if there are no scale economies with respect to train size:

$$\begin{aligned} \nu_1 + \frac{\partial \bar{\delta}}{\partial m}\frac{N}{s_*^u} - \frac{\lambda}{\nu_2^2}\left(\frac{\partial \bar{\delta}}{\partial m}\right)^2\left(\frac{N}{s_*^u}\right)^4 &= 0, \\ \nu_1 \left(m_*^u\right)^4 + \nu_2 \left(m_*^u\right)^3 - \lambda\nu_2^2\left(\frac{\partial \bar{\delta}}{\partial m}\right)^{-2} &= 0. \end{aligned}$$

The first equation in the system solves for a unique value of N/s_*^u which implies that train capacity is chosen proportional to ridership. The second equation solves for a unique value of m_*^u which implies that the number of trains is independent of ridership. These properties imply that equilibrium user cost, c^u, price, p^u, and average system cost are all constant. Hence, unlike in Mohring's model (in which $\nu_1 = \nu_2 = 0$, $\nu_0 > 0$) there are no scale economies with respect to traffic density. However, the case $\nu_0 = 0$ is similar to the bottleneck model in which optimal capacity is proportional to usage, and equilibrium user cost, price, and average system cost are constants (see Arnott et al, 1993).

The degree of cost recovery from fare revenue is easily derived. Fare revenue is $R^u = \lambda N^2/\left(m_u s_u\right)$. Given first-order condition (14a) this implies $R_*^u = \left(\nu_1 m_*^u + \nu_2\right) s_*^u$. The cost recovery ratio, ρ, is therefore

$$\rho = \frac{R_*^u}{K\left(m_*^u, s_*^u\right)} = \frac{\left(\nu_1 m_*^u + \nu_2\right) s_*^u}{\nu_0 m_*^u + \left(\nu_1 m_*^u + \nu_2\right) s_*^u} \leq 1.$$

If there are no scale economies with respect to train size (i.e., $\nu_0 = 0$), fare revenue fully covers capacity costs. Otherwise, costs are only partially recovered and the service runs a deficit.

Social optimum.- For the social optimum, the first-order conditions for s and m are given in Eqs. (12a) and (12b). With the service cost function (13), the equations reduce to

$$\frac{\lambda N^2}{m s^2} = \nu_1 m + \nu_2 - R V_s^o, \tag{15a}$$

$$\left(\frac{\lambda N}{m^2 s} - \frac{\partial \bar{\delta}}{\partial m}\right) N = \nu_0 + \nu_1 s - R V_m^o. \tag{15b}$$

Unlike Eqs. (14a) and (14b), (15a) and (15b) cannot be solved to obtain useful expressions for s_*^o and m_*^o. As noted above, there is no guarantee in general that service quality is better in the social optimum than the uniform-fare regime in the sense that $s_*^o(N) > s_*^u(N)$ and $m_*^o(N) > m_*^u(N)$. Indeed, in the numerical example of Section 5 it turns out that $s_*^o(N) < s_*^u(N)$. However, with capacity function (13), $m_*^o(N) > m_*^u(N)$:

Proposition 15. *For a given usage level, the optimal number of trains is greater in the social optimum than in the optimal uniform-fare regime.*

Unlike for the optimal uniform-fare regime, there is no simple formula for the degree of cost recovery from SO-fare revenue. To derive further insights, and to rank m, s, and N for the three fare regimes, we now consider a

Table 1: Comparison of no-fare, optimal uniform fare, and SO-fare (i.e., social optimum) regimes: base-case parameter values

	Fare regime		
	No-fare (n)	Optimal uniform fare (u)	Social optimum (o)
m	25.26	24	26.70
s	1,762	1,733	1,710
N	37,173	32,600	32,907
p	6.40	9.48	9.22
$Rev/user$	0	3.45	3.39
TCC	161,558	133,499	111,520
SDC	76,210	63,244	80,376
TC	237,768	196,743	191,896
K	138,270	134,889	136,528
R	0	112,407	111,520
ρ	0	0.833	0.817
CS	1,873,288	1,766,213	1,774,816
SS	1,735,018	1,743,732	1,749,807
$Total\ gain$		8,714	14,789
$Gain/user$		0.27	0.45
$Rel.eff$	0	0.59	1

numerical example.

5 A numerical example

The numerical example draws on recent empirical estimates of crowding costs, and it is calibrated to describe service on the Paris RER A line during the morning peak.[35] Base-case parameter values are: $\beta = 7.4$ [€/(hr·user)], $\gamma = 17.2$ [€/(hr·user)], $\lambda = 4.4$ [€/user], and $h = 2.5$ [min/train]. The demand function (9) is assumed to have a constant-elasticity form $N = N_0 p^{\eta}$ with $\eta = -1/3$.[36] Parameter N_0 and parameters ν_0, ν_1, and ν_2 of the capacity cost function are chosen to yield equilibrium values for the optimal uniform-fare equilibrium of $N^u = 32,600$, $m^u_* = 24$, $s^u_* = 1,733$, and a cost recovery rate of 5/6. The resulting values are: $N_0 = 69,003$ [users], $\nu_0 = 936.7$ [€/train], $\nu_1 = 0.1344$ [€/user], and $\nu_2 = 61.63$ [€·train/user]. Results for the three fare regimes are reported in Table 1.[37]

5.1 No fare

With no fare, the equilibrium private cost (which equals the equilibrium user cost) is €6.40. There are $N^n = 37,173$ users who are accommodated in $m^n_* = 25.26$ trains with nominal capacities of $s^n_* = 1,762$. Total crowding costs (TCC^n) are more than double total schedule delay costs (SDC^n). Capital costs (K^n) are about 58 percent as large

[35] Parameter values are explained in online Appendix H.

[36] An elasticity of $-1/3$ is in the mid-range of empirical estimates (Oum, Waters II and Fu, 2008, p.249). Consumers' surplus is infinite with $\eta > -1$. To enable comparisons of consumers' surplus between regimes, the area to the left of the demand curve is computed only for $p \leq$ €100.

[37] Throughout the numerical example m is treated as a continuous variable. The results change very little if m is restricted to integer values (see online Appendix I).

as total user costs (TC^n). Given no fare, the degree of cost recovery is zero.

5.2 Optimal uniform fare

The optimal uniform fare works out to $\tau^u =$€3.45. It boosts the equilibrium private cost to $p^u =$ €9.48 which is €3.08 above the no-fare equilibrium price. Ridership drops to $N^u = 32,600$: about 12 percent below the no-fare level. Both the number of trains and train capacity are lower than with no fare although capacity costs are reduced by only 2.4 percent. Total crowding costs and total schedule delay costs are also lower than with no fare. By design, fare revenue of $R^u = 112,407$ covers a fraction $\rho^u = 0.833$ of capacity costs. Consumers' surplus is lower than with no fare, but social surplus is higher by €8,714 or about €0.27 per rider in the uniform-fare equilibrium. The relative efficiency of the optimal uniform fare can be measured by taking the no-fare and social optimum regimes as polar benchmarks and using the index

$$Eff_u = \frac{\widehat{SS}^u - \widehat{SS}^n}{\widehat{SS}^o - \widehat{SS}^n}.$$

With the base-case parameter values, $Eff_u \simeq 0.59$ so that the optimal uniform fare yields nearly 3/5 of the efficiency gain from the SO-fare.

5.3 Social optimum

The social optimum calls for more trains than either the no-fare or the uniform-fare regime. This is consistent with the result $m_*^o(N) > m_*^u(N)$ established for fixed demand in Prop. 15. However, train capacity is slightly lower than in the other two regimes. Ridership and consumers' surplus are slightly higher than with a uniform fare. Price, revenue per user, and cost recovery are slightly lower. Crowding costs are significantly lower than in the other regimes, but schedule delay costs are higher because the SO-fare spreads usage more evenly over trains. Capacity costs are intermediate between the other regimes. Social surplus is higher than with no fare by about €0.45 per rider.

5.4 Short-run versus long-run welfare gain from pricing

In Table 1, capacity is chosen optimally for each fare regime. Because rail transit capacity can take years to adjust, it is of interest to compare fare regimes in the "short run" when capacity is fixed. If pricing is assumed to become more efficient over time, there are three cases to consider: regime u with capacity fixed at (m_*^n, s_*^n), regime o with capacity fixed at (m_*^n, s_*^n), and regime o with capacity fixed at (m_*^u, s_*^u). Let G_x^{xy} denote the welfare gain in shifting from regime x to regime y when capacity remains fixed at its optimal level for regime x. With the base-case parameters one obtains $G_n^{nu} =$ €8,336, $G_u^{uo} =$ €5,273 and $G_n^{no} =$ €14,589. By comparison, from Table 1 the long-run welfare gains when capacity is adjusted optimally are $G^{nu} =$ €8,714, $G^{uo} =$ €6,076 and $G^{no} =$ €14,788. The long-run gains are higher by 4.5 percent, 15.2 percent and 1.4 percent respectively. The difference between

Table 2: Effects of increasing parameters β and γ (or parameter h) by 10 percent

	Fare regime		
	No-fare (n)	Opt. unif. fare (u)	Soc. opt. (o)
m	-4.98%	-4.98%	-4.44%
s	$+1.69\%$	$+1.70\%$	$+1.64\%$
N	-1.07%	-0.99%	-0.96%
Welf. gain		$+0.7\%$	$+6.3\%$

Table 3: Effects of increasing parameter λ by 10 percent

	Fare regime		
	No-fare (n)	Opt. unif. fare (u)	Soc. opt. (o)
m	$+3.02\%$	$+3.02\%$	$+2.92\%$
s	$+2.14\%$	$+2.14\%$	$+2.15\%$
N	-1.06%	-1.08%	-1.08%
Welf. gain		$+2.4\%$	$+1.4\%$

short-run and long-run gains is appreciable only for G^{uo}. This is mainly because regimes u and o differ the most in terms of optimal number of trains[38]

In all three fare regimes the equilibrium price is an increasing function of parameters β, γ, λ and h. Equilibrium usage thus decreases if these parameters increase in value. Because capacity is endogenous in this section, varying parameters β, γ, λ or h induces changes in s and m, and to determine the size of the effects it is necessary to solve for the new equilibria.

As a first experiment, parameters β and γ were both increased by 10 percent. The results are shown in Table 2. In each fare regime the number of trains drops by nearly 5 percent because users incur higher costs from schedule delay, which reduces demand for travel. Partly to compensate, train capacity increases by about 1.7 percent. Equilibrium prices rise, and usage drops slightly. Welfare gain G^{nu} increases by 0.7 percent, and welfare gain G^{uo} increases by 6.3 percent. An increase in headway, h, has exactly the same effect as an equal percentage increase in β and γ. Thus, the consequences of a 10 percent increase in h are as shown in Table 2.

As a second experiment, parameter λ was increased by 10 percent. The results are shown in Table 3. In all fare regimes the number of trains rises by about 3 percent while train capacity increases by just over 2 percent. Usage drops by about 1 percent. Welfare gains G^{nu} and G^{uo} both increase slightly.

Tables 2 and 3 depict long-run effects of changes in parameter values. These effects can differ significantly from the short-run effects when capacity is given. Consider, for example, welfare gain G^{uo}. In the short run with s and m fixed, G^{uo} is given by Eq. (8). With a 10 percent increase in β and γ, G^{uo} rises by a factor of $(1.1)^2$, or 21 percent. This is more than triple the 6.3 percent long-run increase shown in Table 2. A 10 percent increase in λ causes G^{uo} to fall in the short run by a factor $(1.1)^{-1}$ or about 9 percent. Yet Table 3 shows that the long-run gain actually rises by 1.4 percent.

[38] Note that by Prop. 10, usage with the SO-fare and capacity fixed at (m^u_*, s^u_*) is the same as usage with the optimal uniform fare. Thus, in the short run regimes u and o differ only in how passengers are distributed between trains.

The large differences between the short-run and long-run effects highlight the importance of the time horizon that is adopted for planning. For example, recent empirical research has led to improved estimates of the costs of public transport crowding (OECD, 2014). A rise in the estimated unit cost of crowding (i.e., parameter λ) might dissuade a planner with a short-run perspective from implementing train-dependent fares. By contrast, a planner with a long-run perspective could be spurred to go ahead. This illustrates the well-known lesson that pricing and capacity investment decisions are interdependent, and should be considered jointly (Lindsey, 2012).

6 Conclusion

In this article we have analyzed the time pattern of usage and crowding on a commuter rail line using a model (the *PTC* model) of trip-timing preferences. Users face a trade-off between riding a crowded train that arrives at a convenient time, and riding a less crowded train that arrives earlier or later than desired. We solve user equilibrium for three fare regimes: no fare, an optimal uniform fare that controls the total number of users, and an optimal train-dependent fare that controls the distribution of users between trains as well. We also solve for the optimal long-run number and capacities of trains for the three fare regimes.

In all fare regimes timely trains are more popular and correspondingly more crowded. Under plausible assumptions, passenger loads are distributed more evenly across trains in the social optimum than in the user equilibrium. Arrivals at the destination therefore occur at a more even rate, whereas in the bottleneck model the arrival rate is constant (and equal to bottleneck capacity) throughout the arrival period. Because crowding is assumed to occur at all levels of train occupancy, it is impossible to eliminate crowding costs even if fares can be varied freely. Consequently, imposing Pigouvian fares makes users worse off – at least before accounting for how the revenue is used.

Perhaps the most striking result is that if the crowding cost function is convex, the short-run welfare gain from introducing optimal train-dependent fares decreases with total ridership. The marginal social cost of accommodating an additional passenger is actually higher in the social optimum than with a uniform fare even though passengers are distributed optimally across trains in the social optimum. This finding constrasts with both conventional wisdom and models of road traffic flow including the bottleneck model.

Solving for optimal transit supply in the *PTC* model is complicated by the fact that capacity has two dimensions: the number of trains and the capacity of each train. We treat a special case with linear crowding and schedule delay cost functions, and a uniform headway between trains. The ranking of optimal capacity in the no-fare and optimal uniform-fare regimes is ambiguous in general. More users take transit in the no-fare regime, but the benefit from expanding capacity is diluted by latent demand. Expanding capacity is more valuable in the social optimum than the optimal uniform-fare regime because capacity is used more efficiently. The optimal number of trains is unambiguously higher in the social optimum because more users take additional trains. Optimal train capacity is

also higher in the social optimum if the number of trains is equal, but the ranking of capacity is ambiguous when the number of trains is optimized as well. The result that capacity investments tend to yield higher benefits in the social optimum again constrasts with conventional wisdom that efficient pricing and capacity investments are substitutes for relieving congestion. Since the result holds when demand is fixed, this property of the *PTC* model also differs from road traffic congestion models including the bottleneck model (compare Arnott et al., 1993, Section III).

For illustration we calibrate the model to describe the Paris RER A line during morning-peak conditions. With the base-case parameter values the welfare gain from implementing efficient pricing is €0.27 per user for the optimal uniform fare, and €0.45 for the optimal train-dependent fare. While these amounts may seem modest, the system-wide gain could be large. The RER A line carries more than 300 million users per year, and on average more than 1.5 million individuals used public transport in the Île-de-France region during the morning peak (7am-9am) in 2010.[39] Given 250 working days per year, a welfare gain of about €0.50 per trip, and doubling the number of trips to account (roughly) for evening travel, the annual total welfare gain from optimal pricing amounts to nearly €400 million per year. This figure is comparable to the social saving from a road traffic cordon congestion pricing scheme. Treating the Île-de-France region, and applying the bottleneck model to the road network, De Lara et al. (2013) estimated an annual social saving of €606 million from a cordon toll.

The analysis in this paper could be extended in various directions. One is to allow travelers to differ in their trip-timing preferences and disutility from crowding. Doing so would allow consideration of the equity implications of alternative fare regimes and service investment policies. Another extension is to consider rewards as a means of redistributing passengers across trains. An alternative to penalizing peak-period users with high fares is to reduce off-peak users with low, or possibly even negative, fares. As noted in the introduction, cities such as Singapore and Melbourne have implemented such schemes.[40] Pricing usage below marginal social cost is inefficient when it induces excessive travel, but the induced deadweight loss may be an acceptable price to pay if discounting fares helps overcome opposition to time-of-day pricing. The relative merits of alternative fare-reward pricing schemes depend on the extent to which lowering transit fares can alleviate road traffic congestion. If peak-period traffic congestion is severe and road usage is underpriced, second-best peak-period fares could actually be lower than off-peak fares. However, since cross-price elasticities of demand between driving transit are generally found to be small (Hensher, 1998; Litman, 2004) it is widely believed that manipulating transit fares (or other dimensions of service quality) is likely to have little effect on driving.[41]

A third extension is to combine crowding costs with queuing delay as in Kraus and Yosida (2002). Both forms of congestion are often manifest in transit systems. If a distinction is also made between seated and standing

[39] See p.11 in `http://www.lvmt.fr/IMG/pdf/RAPA_Chaire_Stif_2013-2014_v1.pdf`

[40] The Spitsmijden experiment in the Netherlands indicates that rewarding motorists for driving off-peak can also be succesful (Knockaert et al., 2012).

[41] An exception is Anderson (2014) who uses data from a 2003 transit strike in Los Angeles and finds that cessation of transit service has a large effect on traffic speeds on heavily congested roads.

passengers, as in de Palma, Kilani and Proost (2015), passengers can experience congestion in a number of ways: delays when accessing stations and waiting on the platform, delays when trains are too full to board, delays while boarding, discomfort while seated, greater discomfort and possibly fatigue while standing, and delays while alighting at the destination and exiting stations. The analysis of such a system is likely to be insightful but challenging.

References

Allen, John, and Herbert Levinson. 2014. "Accommodation of Long-Term Growth on North America's Commuter Railroads." Transportation Research Record, 2419: 40–49.

Arnott, Richard, André de Palma, and Robin Lindsey. 1990. "Departure time and route choice for the morning commute." Transportation Research Part B: Methodological, 24(3): 209–228.

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

De Lara, Michel, André de Palma, Moez Kilani, and Serge Piperno. 2013. "Congestion Pricing and Long Term Urban Form: Application to Paris Region." Regional Science and Urban Economics, 43(2): 282–295.

de Palma, André, Moez Kilani, and Stef Proost. 2015. "Discomfort in Mass Transit and its Implication for Scheduling and Pricing." Transportation Research Part B: Methodological, 71(0): 1–18.

Duranton, Gilles, and Matthew A. Turner. 2011. "The Fundamental Law of Road Congestion: Evidence from US Cities." American Economic Review, 101(6): 2616–52.

Haywood, Luke, and Martin Koning. 2015. "The Distribution of Crowding Costs in Public Transport: New Evidence from Paris." Transportation Research Part A: Policy and Practice, 77(0): 182–201.

Henderson, J. Vernon. 1974. "Road Congestion: A Reconsideration of Pricing Theory." Journal of Urban Economics, 1(3): 346–365.

Hensher, David A. 1998. "Establishing a Fare Elasticity Regime for Urban Passenger Transport." Journal of Transport Economics and Policy, 32(2): 221–246.

Huang, Hai-Jun, Qiong Tian, and Zi-You Gao. 2005. "An Equilibrium Model in Urban Transit Riding and Fare Polices." Algorithmic Applications in Management, 3521: 112–121.

Huang, Hai-Jun, Qiong Tian, Hai Yang, and Zi-You Gao. 2007. "Modal Split and Commuting Pattern on a Bottleneck-Constrained Highway." Transportation Research Part E: Logistics and Transportation Review, 43(5): 578–590.

Knight, Frank H. 1924. "Some Fallacies in the Interpretation of Social Cost." The Quarterly Journal of Economics, 38(4): 582–606.

Knockaert, Jasper, Yin-Yen Tseng, Erik T. Verhoef, and Jan Rouwendal. 2012. "The Spitsmijden Experiment: A Reward to Battle Congestion." Transport Policy, 24: 260–272.

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

Lindsey, Robin. 2012. "Road Pricing and Investment." Economics of Transportation, 1(1): 49–63.

Litman, Todd. 2004. "Transit Price Elasticities and Cross-Elasticities." Journal of Public Transportation, 7(2): 37–58.

Mohd Mahudin, Nor Diana, Tom Cox, and Amanda Griffiths. 2012. "Measuring Rail Passenger Crowding: Scale Development and Psychometric Properties." Transportation Research Part F: Traffic Psychology and Behaviour, 15(1): 38–51.

Mohring, Herbert. 1972. "Optimization and Scale Economies in Urban Bus Transportation." American Economic Review, 62(4): 591–604.

OECD. 2014. "Valuing Convenience in Public Transport." ITF Round Tables. Retrieved from http://www.internationaltransportforum.org/jtrc/DiscussionPapers/DP201402.pdf.

Oum, Tae Hoon, W. G. Waters II, and Xiaowen Fu. 2008. "Transport Demand Elasticities." In Handbook of Transport Modelling, 2nd edition. , ed. D.A. Hensher and K.J. Button. Oxford: Elsevier.

Parry, Ian W.H., and Kenneth A. Small. 2009. "Should Urban Transit Subsidies be Reduced?" American Economic Review, 99(3): 700–724.

Peer, Stefanie, Erik Verhoef, Jasper Knockaert, Paul Koster, and Yin-Yen Tseng. 2015. "Long-Run Versus Short-Run Perspectives on Consumer Scheduling: Evidence from a Revealed-Preference Experiment among Peak-Hour Road Commuters." International Economic Review, 56(1): 303–323.

Pigou, Arthur C. 1920. The Economics of Welfare. London: Macnillam.

Prud'homme, Rémy, Martin Koning, Luc Lenormand, and Anne Fehr. 2012. "Public Transport Congestion Costs: The Case of the Paris Subway." Transport Policy, 21: 101–109.

Rothschild, Michael, and Joseph E. Stiglitz. 1970. "Increasing risk: I. A definition." Journal of Economic Theory, 2(3): 225–243.

Singapore Land Transport Authority. 2013. "Travel Early, Travel Free on the MRT." Retrieved from http://www.lta.gov.sg/apps/news/page.aspx?c=2&id=c3983784- 2949-4f8d-9be7-d095e6663632.

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

Small, Kenneth A. 2015. "The Bottleneck Model: An Assessment and Interpretation." Economics of Transportation, 4(1-2): 110–117.

Tian, Qiong, Hai-Jun Huang, and Hai Yang. 2007. "Equilibrium Properties of the Morning Peak-Period Commuting in a Many-to-One Mass Transit System." Transportation Research Part B: Methodological, 41(6): 616–631.

Tirachini, Alejandro, David A Hensher, and John M Rose. 2013. "Crowding in Public Transport Systems: Effects on Users, Operation and Implications for the Estimation of Demand." Transportation Research Part A: Policy and Practice, 53: 36–52.

Tseng, Yin-Yen, and Erik T. Verhoef. 2008. "Value of Time by Time of Day: A Stated-Preference Study." Transportation Research Part B: Methodological, 42(7-8): 607–618.

Veitch, Tim, James Partridge, and Lauren Walker. 2013. "Estimating the Costs of Over-crowding on Melbourne's Rail System." Retrieved from http://atrf.info/papers/2013/2013_veitch_partridge_walker.pdf.

Vickrey, William S. 1963. "Pricing in Urban and Suburban Transport." American Economic Review, 53(2): 452–465.

Vickrey, William S. 1969. "Congestion Theory and Transport Investment." American Economic Review, 59(2): 251–260.

Walker, Jarrett. 2010. "Should Fares be Higher During Peak Hours?" Human Transit. Retrieved from http://www.humantransit.org/2010/05/should-fares-behigher- during-peak-hours.html.

Wardman, Mark, and Gerard Whelan. 2011. "Twenty Years of Rail Crowding Valuation Studies: Evidence and Lessons from British Experience." Transport Reviews, 31(3): 379–398.

Wardman, Mark, Phani Chintakayala, Gerard de Jong, and Diego Ferrer. 2012. "European Wide Meta-Analysis of Values of Travel Time." Significance. Retrieved from http://www.significance.nl/papers/2012-European

Willsher, Kim. 2014. "French Railway Operator SNCF Orders Hundreds of New Trains that are Too Big." The Guardian. Retrieved from http://www.theguardian.com/world/2014/may/21/french-railway-operator-sncf-orders-trains-too-big.

Yauch, Brady. 2015. "The Time is Now for Toronto to Look at Off-Peak Transit Fares." Consumer Policy Institute. Retrieved from http://cpi.probeinternational.org/2015/01/28/the-time-is-now-for-toronto-to-lookat-off-peak-transit-fares/.

Appendix

A Proof of Proposition 3

Let j index trains in order of decreasing schedule delay cost so that $\delta_1 > \delta_2 > ... > \delta_m$. (Because trains arrive early and late, the index does not correspond to the temporal sequence in which trains are run.) Since in the UE $n_j^e = g^{-1}\left[c^e - \delta_j\right]$ and $g'(.) > 0$, n_j^e increases with j: $n_1^e < n_2^e < ... < n_m^e$.

We show that $n_j^e \lessgtr n_j^o \Longleftrightarrow n_j^e g'\left(n_j^e\right) \lessgtr MSC^o - c^e$. Given $n_j^o = v^{-1}[MSC^o - \delta(t_j)]$, it follows that

$$\begin{aligned}
n_j^e &\lessgtr n_j^o \\
\Longleftrightarrow g^{-1}\left[c^e - \delta_j\right] &\lessgtr v^{-1}\left[MSC^o - \delta_j\right] \\
\Longleftrightarrow v\left\{g^{-1}\left[c^e - \delta_j\right]\right\} &\lessgtr MSC^o - \delta_j \\
\Longleftrightarrow c^e - \delta_j + g^{-1}\left[c^e - \delta_j\right] \times g'\left\{g^{-1}\left[c^e - \delta_j\right]\right\} &\lessgtr MSC^o - \delta_j \\
\Longleftrightarrow g^{-1}\left[c^e - \delta_j\right] \times g'\left\{g^{-1}\left[c^e - \delta_j\right]\right\} &\lessgtr MSC^o - c^e \\
\Longleftrightarrow n_j^e g'\left(n_j^e\right) &\lessgtr MSC^o - c^e.
\end{aligned}$$

Variables n_j^e and n_j^o have the same ranking as $n_j^e g'\left(n_j^e\right)$, the marginal external cost of crowding in the UE, and $MSC^o - c^e$, which is constant. Because total patronage, N, is fixed, some trains are more more heavily loaded in the UE, and the others are more heavily loaded in the SO. Consequently, if $ng'(n)$ is a strictly increasing function of n (i.e., $\varepsilon(n) > -1$), there exists a unique train $\hat{\jmath}$ such that $n_j^e < n_j^o$ when $j < \hat{\jmath}$, $n_{\hat{\jmath}}^e \geq n_{\hat{\jmath}}^o$, and $n_j^e > n_j^o$ when $j > \hat{\jmath}$. Conversely, if $ng'(n)$ is a strictly decreasing function of n (i.e., $\varepsilon(n) < -1$), there exists a unique train $\hat{\jmath}$ such that $n_j^e > n_j^o$ when $j < \hat{\jmath}$, $n_{\hat{\jmath}}^e \leq n_{\hat{\jmath}}^o$, and $n_j^e < n_j^o$ when $j > \hat{\jmath}$.

B Proof of Proposition 4

Total fare revenue from the optimal uniform fare is $R^u = \tau^u N$. Hence $\frac{\partial R^u}{\partial N} = \tau^u + \frac{\partial \tau^u}{\partial N} N$. Now

$$MSC^u = \frac{\partial TC^u}{\partial N} = \frac{\partial\left(c^u N\right)}{\partial N} = c^u + \frac{\partial c^u}{\partial N} N = c^u + \tau^u.$$

Thus

$$\frac{\partial MSC^u}{\partial N} N = \left(\frac{\partial c^u}{\partial N} + \frac{\partial \tau^u}{\partial N}\right) N = \tau^u + \frac{\partial \tau^u}{\partial N} N = \frac{\partial R^u}{\partial N}.$$

Total fare revenue from the SO-fare is $R^o = \sum_{k=1}^m \tau_k^o n_k^o$. Hence

$$\frac{\partial R^o}{\partial N} = \sum_{k=1}^m \left(\tau_k^o + \frac{\partial \tau_k^o}{\partial n_k^o} n_k^o\right) \frac{\partial n_k^o}{\partial N}.$$

The marginal social cost of a trip is the same for all trains that are used: $MSC^o = c_k^o + \tau_k^o$. Hence:

$$\begin{aligned}
\frac{\partial MSC^o}{\partial N} &= \left(\frac{\partial c_k^o}{\partial n_k^o} + \frac{\partial \tau_k^o}{\partial n_k^o}\right)\frac{\partial n_k^o}{\partial N}, \\
\frac{\partial MSC^o}{\partial N} n_k^o &= \left(\frac{\partial c_k^o}{\partial n_k^o} n_k^o + \frac{\partial \tau_k^o}{\partial n_k^o} n_k^o\right)\frac{\partial n_k^o}{\partial N} = \left(\tau_k^o + \frac{\partial \tau_k^o}{\partial n_k^o} n_k^o\right)\frac{\partial n_k^o}{\partial N}, \\
\frac{\partial MSC^o}{\partial N} N &= \sum_{k=1}^{m} \frac{\partial MSC^o}{\partial N} n_k^o = \sum_{k=1}^{m} \left(\tau_k^o + \frac{\partial \tau_k^o}{\partial n_k^o} n_k^o\right)\frac{\partial n_k^o}{\partial N} = \frac{\partial R^o}{\partial N}.
\end{aligned}$$

C Proof of Proposition 5

We prove the case for which the welfare gain G^{eo} decreases with N. The proof for the case in which G^{eo} decreases follows the same steps, and is omitted. As mentioned in the text, G^{eo} decreases if the marginal social cost of an additional user is higher in the SO than the UE. Thus, it suffices to show that $MSC^o > MSC^e$.

As in Appendix A, let k index trains in order of decreasing schedule delay cost so that in the no-fare equilibrium, $n_1^e < n_2^e < ... < n_m^e$. Equilibrium cost with no fare, c^e, is determined implicitly by Eq. (1):

$$\sum_{k=1}^{m} g^{-1}\left[c^e - \delta_k\right] - N = 0. \tag{C.1}$$

This equation can be written

$$\sum_{k=1}^{m} f\left[g\left(n_k^e\right) + g'\left(n_k^e\right) n_k^e\right] = N, \tag{C.2}$$

where $f(n) \equiv v^{-1}(n)$. Since $f(v(n)) = n$,

$$f'(n) = \frac{1}{v'(n)} = \frac{1}{2g'(n) + g''(n)\, n}. \tag{C.3}$$

The marginal social cost of a trip in the no-fare equilibrium is $MSC^e = \frac{\partial (c^e N)}{\partial N} = c^e + \frac{\partial c^e}{\partial N} N$. Using Eq. (C.1) to derive $\frac{\partial c^e}{\partial N}$ one obtains

$$MSC^e = c^e + \frac{N}{\sum_{k=1}^{m} \frac{1}{g'\left(n_k^e\right)}}. \tag{C.4}$$

The marginal social cost of a trip in the social optimum is defined implicitly by:

$$\sum_{k=1}^{m} f\left[MSC^o - \delta_k\right] = N. \tag{C.5}$$

By Assumption 1, the left-hand side of Eq. (C.5) is a strictly increasing function of MSC^o. Suppose we substitute eqn. (C.4) for MSC^e in place of MSC^o in Eq. (C.5). If the resulting left-hand side is less than N, then $MSC^o > MSC^e$ and the proof is complete. To economize on notation, let g_k denote $g\left(n_k^e\right)$, g_k' denote $g'\left(n_k^e\right)$, and

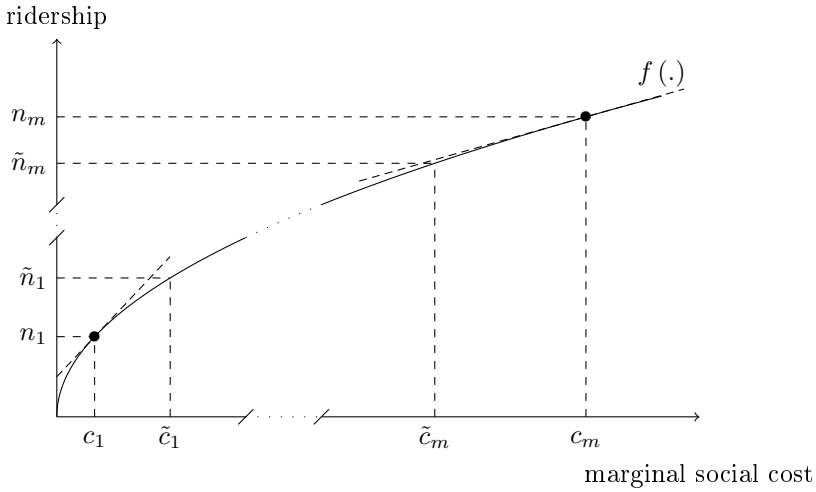

Figure 3: Ridership and marginal social cost

n_k denote n_k^e. After a few substitutions one can write

$$\sum_{k=1}^{m} f\left[MSC^e - \delta_k\right] = \sum_{k=1}^{m} f\left[g_k + \frac{N}{m}\frac{m}{\sum_{k=1}^{m}\frac{1}{g_k'}}\right].$$

Define

$$mec_k \equiv g_k + g_k' n_k, \tag{C.6}$$

and

$$\widetilde{mec}_k \equiv g_k + \frac{m}{\sum_{k=1}^{m}\frac{1}{g_k'}}\frac{N}{m}. \tag{C.7}$$

Given Eq. (C.2), we need to prove that the following expression is negative:

$$\Delta F \equiv \sum_{k=1}^{m} \underbrace{f\left[g_k + \frac{m}{\sum_{k=1}^{m}\frac{1}{g_k'}}\frac{N}{m}\right]}_{\tilde{n}_k} - \sum_{k=1}^{m} \underbrace{f\left[mec_k\right]}_{n_k}.$$

Given Assumption 1, $\tilde{n}_k > n_k$ for small k, and $\tilde{n}_k < n_k$ for large k. The rankings of $\tilde{n}_k$ and n_k, and of $\tilde{c}_k$ and c_k, are shown in Figure 3.

Function $f()$ is concave by Assumption 3a. Clearly, for all trains $\tilde{n}_k - n_k < \left(\tilde{c}_k - c_k\right) f'\left[c_k\right], \quad k = 1...m$. Using Eqs. (C.6), (C.7) and (C.3) this implies

$$\begin{aligned}
\Delta F &= \sum_{k=1}^{m}\tilde{n}_k - \sum_{k=1}^{m} n_k < \sum_{k=1}^{m}\left(\tilde{c}_k - c_k\right) f'\left[c_k\right] \\
&= \sum_{k=1}^{m}\left(\frac{m}{\sum_{k=1}^{m}\frac{1}{g_k'}}\frac{N}{m} - g_k' n_k\right)\frac{1}{2g_k' + g_k'' n_k}.
\end{aligned}$$

Now, $\sum_{k=1}^{m}\frac{1}{g_k'}=\frac{\sum_{j=1}^{m}\prod_{i\neq j}g_i'}{\prod_{i=1}^{k}g_i'}$. Hence

$$\begin{aligned}
\Delta F &= \sum_{k=1}^{m}\left(N\frac{\prod_{i=1}^{m}g_i'}{\sum_{j=1}^{m}\prod_{i\neq j}g_i'}-g_k'n_k\right)\frac{1}{2g_k'+n_kg_k''}\\
&= \sum_{k=1}^{m}\left(N\frac{\prod_{i\neq k}g_i'}{\sum_{j=1}^{m}\prod_{i\neq j}g_i'}-n_k\right)\frac{g_k'}{2g_k'+n_kg_k''}\\
&= \sum_{k=1}^{m}\left(\frac{\left(\sum_{l\neq k}n_l\right)\prod_{i\neq k}g_i'}{\sum_{j=1}^{m}\prod_{i\neq j}g_i'}+\left(\frac{\prod_{i\neq k}g_i'}{\sum_{j=1}^{m}\prod_{i\neq j}g_i'}-1\right)n_k\right)\frac{g_k'}{2g_k'+n_kg_k''}\\
&= \sum_{k=1}^{m}\left(\underbrace{\left(\sum_{l\neq k}n_l\right)}_{(1)}\underbrace{\frac{\prod_{i\neq k}g_i'}{\sum_{j=1}^{m}\prod_{i\neq j}g_i'}}_{(2)}-\underbrace{\frac{\sum_{j\neq k}\prod_{i\neq j}g_i'}{\sum_{j=1}^{m}\prod_{i\neq j}g_i'}}_{(3)}\underbrace{n_k}_{(4)}\right)\frac{g_k'}{2g_k'+n_kg_k''}
\end{aligned}\tag{C.8}$$

In the second line of eqn. (C.8),

$$\begin{aligned}
&\sum_{k=1}^{m}\left(N\frac{\prod_{i\neq k}g_i'}{\sum_{j=1}^{m}\prod_{i\neq j}g_k'}-n_k\right)\\
&= N\sum_{k=1}^{m}\left(\frac{\prod_{i\neq k}g_i'}{\sum_{j=1}^{m}\prod_{i\neq j}g_i'}\right)-\sum_{k=1}^{m}n_k\\
&= N-N=0.
\end{aligned}$$

Terms (1) and (2) in the last line of Eq. (C.8) are decreasing functions of k. Terms (3) and (4) are increasing functions of k. Hence Eq. (C.8) is negative if $\frac{g_k'}{2g_k'+n_kg_k''}$ is a non-decreasing function of k, or equivalently if $\varepsilon(n)=\frac{g_k''n_k}{g_k'}$ is a non-increasing function of k which is guaranteed by Assumption 3a.

D Proof of Proposition 10

We first consider uniform fares (which include no fare and the optimal uniform fare as special cases), and then the SO-fare. The goal of this section is to rank equilibrium prices and numbers of trips in the pricing regimes.

D.1 Uniform-fare regimes

With a uniform fare, the equilibrium private cost of a trip, p^e, equals the user cost plus the fare:

$$p^e=\bar{\delta}+\frac{\lambda N}{ms}+\tau.\tag{D.1}$$

Eq. (D.1) serves as a supply function for trips. Solving (D.1) and the demand function (9) yields the equilibrium private cost and number of trips, $\widehat{p}^e$ and $\widehat{N}^e$. If the fare is zero, the equilibrium price is

$$\widehat{p}^n = \bar{\delta} + \frac{\lambda \widehat{N}^n}{ms}. \tag{D.2}$$

Social surplus equals consumers' surplus: $\widehat{SS}^n = \widehat{CS}^n = \int_{\widehat{p}^n}^{\infty} N(u)\, du$. The optimal uniform fare is given by Eq. (2): $\widehat{\tau}^u = \frac{\lambda \widehat{N}^u}{ms}$, and fare revenue is $\widehat{R}^u = \tau^u \widehat{N}^u = \frac{\lambda \left(\widehat{N}^u\right)^2}{ms}$. The efficient price of a trip equals marginal social cost:

$$\widehat{p}^u = \widehat{MSC}^n = \widehat{c}^u + \widehat{\tau}^u = \bar{\delta} + \frac{2\lambda \widehat{N}^u}{ms}. \tag{D.3}$$

Social surplus is equal to $\widehat{SS}^u = \int_{\widehat{p}^u}^{\infty} N(u)\, du + \tau^u \widehat{N}^u$. Finally, the welfare gain in switching from no fare to the optimal uniform fare is $G^{eu} = \widehat{SS}^u - \widehat{SS}^n$.

D.2 Social optimum

The social optimum can be supported by imposing train-specific fares as described in Prop. 8. Total travel costs are derived by substituting Eq. (7) into the expression for TC^o given in Prop. 9: $\widehat{TC}^o = \bar{\delta} \widehat{N}^o + \frac{\lambda \left(\widehat{N}^o\right)^2}{ms} - RV^o$. Since variable revenue in Eq. (7) does not depend on the number of trips, the marginal social cost of a trip is $\widehat{MSC}^o = \bar{\delta} + \frac{2\lambda \widehat{N}^o}{ms}$. Similar to the optimal uniform-fare regime, the efficient price of a trip equals marginal social cost:

$$\widehat{p}^o = \widehat{MSC}^o \left(\widehat{N}^o\right) = \bar{\delta} + \frac{2\lambda \widehat{N}^o}{ms}. \tag{D.4}$$

Eqs. (D.3) and (D.4) reveal that the optimal price is the same function of usage in regimes u and o. This is consistent with the observation that, if the crowding cost function is linear, the marginal social cost of trips is the same in the SO and UE. Social surplus is equal to

$$\widehat{SS}^o = \int_{\widehat{p}^o}^{\infty} N(u)\, du + R^o \left(\widehat{N}^o\right) = \int_{\widehat{p}^o}^{\infty} N(u)\, du + \frac{\lambda \left(\widehat{N}^o\right)^2}{ms} + RV^o.$$

The welfare gain in switching from no fare to the SO-fare is $G^{no} = \widehat{SS}^o - \widehat{SS}^n$, and the welfare gain in switching from the optimal uniform fare to the SO-fare is $G^{uo} = \widehat{SS}^o - \widehat{SS}^u$.

D.3 Comparison of the regimes

Private costs in regimes n, u and o are given by Eqs. (D.2), (D.3), and (D.4) respectively. For given values of m, s, and N, it is clear that private costs are the same in regimes u and o, and lower in regime n. With elastic demand this implies that equilibrium usage is the same in regimes u and o, and higher in regime n. Correspondingly, the equilibrium private cost and consumers' surplus are the same in regimes u and o, and higher in regime n. Social

surplus is highest in regime o, lowest in regime n, and intermediate in regime u.

E Proof of Proposition 11

We first derive the optimal timetable for the UE, and then show that this timetable is also optimal for the SO.

E.1 Optimal timetable for user equilibrium

The optimal timetable is chosen to minimize total user costs. For the UE, total costs are given by Prop. 7: $TC^e = \overline{\delta}N + \frac{\lambda N^2}{ms}$. The timetable should therefore be chosen to minimize average schedule delay cost, $\overline{\delta}$. The timetable can be defined by the arrival time of the last train, t_m. It is clearly not optimal to set $t_m < t^*$, and have all trains arrive early, since $\overline{\delta}$ could be reduced by setting $t_m = t^*$. Similarly, it is not optimal to set $t_m > t^* + (m-1)h$, and have all trains arrive late, since $\overline{\delta}$ could be reduced by setting $t_m = t^* + (m-1)h$. Thus, one train must arrive during the interval $(t^* - h, t^*]$. Call it train $\hat{k}$. Train $\hat{k}$ is the last train to arrive at or before t^*. Average schedule delay cost is

$$\begin{aligned}
\overline{\delta} &= \frac{1}{m}\left(\sum_{k=1}^{\hat{k}} \beta\left(t^* - t_k\right) + \sum_{k=\hat{k}+1}^{m} \gamma\left(t_k - t^*\right)\right) \\
&= \frac{1}{m}\left(\sum_{k=1}^{\hat{k}} \beta\left(t^* - t_{\hat{k}} + h\left(\hat{k} - k\right)\right)\right. \\
&\quad \left. + \sum_{k=\hat{k}+1}^{m} \gamma\left(t_{\hat{k}} - t^* + h\left(k - \hat{k}\right)\right)\right) \\
&= \frac{1}{m}\left(\left(t^* - t_{\hat{k}}\right)\left[(\beta+\gamma)\hat{k} - \gamma m\right] + (\beta+\gamma)h\frac{\hat{k}\left(\hat{k}-1\right)}{2}\right. \\
&\quad \left. + \gamma h\frac{m\left(m+1-2\hat{k}\right)}{2}\right).
\end{aligned} \tag{E.1}$$

The first component of the right-hand side of Eq. (E.1), $\left(t^* - t_{\hat{k}}\right)$, is the time between the arrival time of train $\hat{k}$ and t^*. If $t_{\hat{k}} < t^*$ we can differentiate Eq. (E.1):

$$\frac{\partial \overline{\delta}}{\partial\left(t^* - t_{\hat{k}}\right)} = \frac{(\beta+\gamma)\hat{k}}{m} - \gamma.$$

If $\hat{k} > \gamma m/(\beta+\gamma)$, then $\partial\overline{\delta}/\partial\left(t^* - t_{\hat{k}}\right) > 0$ and $\overline{\delta}$ is minimized by setting $t^* - t_{\hat{k}}$ to its minimal value, i.e $t^* - t_{\hat{k}} = 0$. Conversely, if $\hat{k} < \gamma m/(\beta+\gamma)$, then $\partial\overline{\delta}/\partial\left(t^* - t_{\hat{k}}\right) < 0$ and $\overline{\delta}$ is minimized by setting $t^* - t_{\hat{k}} = h$. Hence it is optimal to schedule one train at t^*. Call it train k^*. Replacing $\hat{k}$ in Eq. (E.1) with k^* one obtains

$$\overline{\delta} = (\beta+\gamma)h\frac{k^*\left(k^*-1\right)}{2m} + \gamma h\frac{m+1-2k^*}{2}.$$

Treating k^* as a continuous variable for the moment, the first-order condition for minimizing $\overline{\delta}$ with respect to k^* is $k^{*o} = \frac{\gamma m}{\beta+\gamma} + \frac{1}{2}$. Since k^* is an integer, we have to compare $\overline{\delta}$ when $k^* = \lfloor k^{*o} \rfloor$ and when $k^* = \lfloor k^{*o} \rfloor + 1$. We find

$$\overline{\delta}_{k^*=\lfloor k^{*o} \rfloor} - \overline{\delta}_{k^*=\lfloor k^{*o} \rfloor+1} \lessgtr 0 \Longleftrightarrow \frac{\gamma m}{\beta+\gamma} \lessgtr \left\lfloor \frac{\gamma m}{\beta+\gamma} + \frac{1}{2} \right\rfloor.$$

Hence,

$$\begin{aligned} k^* &= \left\lfloor \frac{\gamma m}{\beta+\gamma} + \frac{1}{2} \right\rfloor + 1 \times 1_{\frac{\gamma m}{\beta+\gamma} > \left\lfloor \frac{\gamma m}{\beta+\gamma} + \frac{1}{2} \right\rfloor} \\ t_m &= t^* + h\left(m - \left\lfloor \frac{\gamma m}{\beta+\gamma} + \frac{1}{2} \right\rfloor - 1 \times 1_{\frac{\gamma m}{\beta+\gamma} > \left\lfloor \frac{\gamma m}{\beta+\gamma} + \frac{1}{2} \right\rfloor} \right) \end{aligned}$$

In summary, if $\gamma m/(\beta+\gamma) > \lfloor \gamma m/(\beta+\gamma) + 1/2 \rfloor$, then

$$\begin{aligned} k^* &= \lfloor \gamma m/(\beta+\gamma) + 1/2 \rfloor + 1, \text{ and} \\ t_m &= t^* + h(m - 1 - \lfloor \gamma m/(\beta+\gamma) + 1/2 \rfloor). \end{aligned}$$

Conversely, if $\gamma m/(\beta+\gamma) < \left\lfloor \frac{\gamma m}{\beta+\gamma} + \frac{1}{2} \right\rfloor$, then

$$\begin{aligned} k^* &= \lfloor \gamma m/(\beta+\gamma) + 1/2 \rfloor, \text{ and} \\ t_m &= t^* + h\left(m - \lfloor \gamma m/(\beta+\gamma) + 1/2 \rfloor\right). \end{aligned}$$

E.2 Optimal timetable for social optimum

Total costs in the social optimum are given by Prop. 9

$$TC^o = \overline{\delta} N + \frac{\lambda N^2}{ms} - \frac{s}{4\lambda}\left(\Delta - m\overline{\delta}^2\right).$$

TC^o differs from TC^e in including the third term on the right-hand side. Recall that

$$\Delta - m\overline{\delta}^2 = \sum_{k=1}^{m} \delta_k^2 - \frac{1}{m}\left[\sum_{k=1}^{m} \delta_k\right]^2, \tag{E.2}$$

where

$$\delta_k = \beta\left[t^* - t_m + h(m-k)\right]^+ + \gamma\left[t_m - t^* - h(m-k)\right]^+. \tag{E.3}$$

As above, let $\hat{k}$ be the last train to arrive at or before t^*. Differentiating (E.2) with respect to t_m, and using (E.3),

it is possible to show after considerable algebra that

$$\frac{\partial\left(\Delta - m\bar{\delta}^2\right)}{\partial t_m} = \frac{\gamma}{\beta+\gamma}m + 1 - \hat{k}.$$

The term $\Delta - m\bar{\delta}^2$ therefore reaches an extreme point for the same $\hat{k}$ as does $\bar{\delta}$. Hence TC^o reaches a minimum for the same timetable as TC^e.

F Derivatives of SS^e with respect to m and s

First-order conditions for a maximum of SS^e are[42]

$$\frac{\partial SS^e}{\partial s} = p\left(N\right)\frac{\partial N}{\partial s} - \left(-\frac{\lambda N^2}{ms^2} + \left(\bar{\delta} + \frac{2\lambda N}{ms}\right)\frac{\partial N}{\partial s} + K_s\right) = 0, \tag{F.1}$$

$$\frac{\partial SS^e}{\partial m} = p\left(N\right)\frac{\partial N}{\partial m} - \left(\frac{\partial\bar{\delta}}{\partial m}N - \frac{\lambda N^2}{m^2 s} + \left(\bar{\delta} + \frac{2\lambda N}{ms}\right)\frac{\partial N}{\partial m} + K_m\right) = 0. \tag{F.2}$$

The private cost of usage is given by Eq. (D.1) which can be written

$$p\left(N\right) - \left(\bar{\delta} + \frac{2\lambda N}{ms}\right) = \tau - \frac{\lambda N}{ms}. \tag{F.3}$$

The fare, τ, depends on the pricing regime. To maintain generality we assume for the moment that τ can depend on N, m, and s. Substituting (F.3) into (F.1) and (F.2) yields:

$$\frac{\lambda N^2}{ms^2} + \left(\tau - \frac{\lambda N}{ms}\right)\frac{\partial N}{\partial s} - K_s = 0, \tag{F.4}$$

$$\frac{\lambda N^2}{m^2 s} - \frac{\partial\bar{\delta}}{\partial m}N + \left(\tau - \frac{\lambda N}{ms}\right)\frac{\partial N}{\partial m} - K_m = 0. \tag{F.5}$$

The demand derivatives are obtained by totally differentiating (D.1):

$$\frac{\partial N}{\partial s} = \frac{-\frac{\lambda N}{ms^2} + \frac{d\tau}{ds}}{p_N - \frac{\lambda}{ms} - \frac{d\tau}{dN}} > 0, \tag{F.6}$$

$$\frac{\partial N}{\partial m} = \frac{\frac{\partial\bar{\delta}}{\partial m} - \frac{\lambda N}{m^2 s} + \frac{d\tau}{dm}}{p_N - \frac{\lambda}{ms} - \frac{d\tau}{dN}} > 0. \tag{F.7}$$

Substituting (F.6) and (F.7) into (F.4) and (F.5), the first-order conditions become

$$\frac{\lambda N^2}{ms^2}\cdot\frac{p_N N - \tau - \frac{d\tau}{dN}N}{p_N N - \frac{\lambda N}{ms} - \frac{d\tau}{dN}N} + \frac{\left(\tau - \frac{\lambda N}{ms}\right)\frac{d\tau}{ds}N}{p_N N - \frac{\lambda N}{ms} - \frac{d\tau}{dN}N} = K_s,$$

$$\left(\frac{\lambda N}{m^2 s} - \frac{\partial\bar{\delta}}{\partial m}\right)N\cdot\frac{p_N N - \tau - \frac{d\tau}{dN}N}{p_N N - \frac{\lambda N}{ms} - \frac{d\tau}{dN}N} + \frac{\left(\tau - \frac{\lambda N}{ms}\right)\frac{d\tau}{dm}N}{p_N N - \frac{\lambda N}{ms} - \frac{d\tau}{dN}N} = K_m.$$

[42] Given $\bar{\delta} = \beta\gamma/\left(\beta+\gamma\right)hm/2$ as per Prop. (11), $\partial\bar{\delta}/\partial m = \beta\gamma/\left(\beta+\gamma\right)h/2$ which is a constant.

G Proof of Proposition 15

Conditional on m and N, s_*^u is given by $s_*^u(m,N) = \sqrt{\frac{\lambda}{\nu_1 m^2 + \nu_2 m}}N$. First-order condition (15a) can be rearranged to obtain an analogous equation for s_*^o. Recall from Eq. (7) that $RV^o = \frac{s}{4\lambda}\left(\Delta - m\overline{\delta}^2\right)$ where $\Delta = \sum_{k=1}^m \delta_k^2$. Define

$$Y_j \equiv \sqrt{\frac{\lambda}{m\left[\nu_1 m + \nu_2 - X_j\left(\Delta - m\overline{\delta}^2\right)\right]}},$$

where $X_u = 0$ and $X_o = \frac{1}{4\lambda} > 0$. The two equations for s_*^u and s_*^o can be written together as

$$s_*^j(m,N) = Y_j N, \tag{G.1}$$

Substituting (G.1) into the first-order conditions for m_*^u and m_*^o respectively, one obtains

$$\nu_0 + \nu_1 Y_j N + \frac{\partial \overline{\delta}}{\partial m}N - \frac{\lambda N}{m^2}Y_j^{-1} - X_j Y_j N \frac{\partial\left(\Delta - m\overline{\delta}^2\right)}{\partial m} = 0. \tag{G.2}$$

Function (G.2) is negative for small values of m, and over the relevant range it is increasing in m. Hence, if (G.2) is decreasing in X, $m_*^o > m_*^u$. Retaining only terms in (G.2) that depend on X, and multiplying through by $m^2 Y_j^{-1}/N$, one obtains

$$\begin{aligned} & \nu_1 m^2 - m\left[\nu_1 m + \nu_2 - X\left(\Delta - m\overline{\delta}^2\right)\right] - Xm^2\frac{\partial\left(\Delta - m\overline{\delta}^2\right)}{\partial m} \\ = \; & -m\nu_2 - Xm\left[m\frac{\partial\left(\Delta - m\overline{\delta}^2\right)}{\partial m} - \left(\Delta - m\overline{\delta}^2\right)\right]. \end{aligned} \tag{G.3}$$

This expression is decreasing in X if $\Delta - m\overline{\delta}^2$ is a convex function of m. Setting $k^* = \frac{\gamma}{\beta+\gamma}m$, we find

$$\begin{aligned} \Delta - m\overline{\delta}^2 &= \frac{\beta\gamma m h^2}{12}\left[\frac{\beta\gamma m^2}{(\beta+\gamma)^2} + 2\right]; \\ \frac{\partial\left(\Delta - m\overline{\delta}\right)}{\partial m} &= \frac{\beta\gamma h^2}{12}\left[3\frac{\beta\gamma m^2}{(\beta+\gamma)^2} + 2\right] > 0; \\ \frac{\partial^2\left(\Delta - m\overline{\delta}\right)}{\partial m^2} &= \frac{\beta\gamma h^2}{12}\left[6\frac{\beta\gamma m}{(\beta+\gamma)^2}\right] > 0. \end{aligned}$$

Hence $\Delta - m\overline{\delta}^2$ is a convex function of m, Eq. (G.3) is decreasing in X, and $m_*^o > m_*^u$.

H Parameter values for numerical example

The numerical example requires base-case parameter values for β, γ, λ and h, and target values for N, m and s. The operating period was set to one hour, and target values were chosen for the optimal uniform-fare regime. This regime is intermediate in efficiency between the no-fare and SO-fare regimes, and it is arguably the most descriptive of public transit service in Paris where fares are positive and constant throughout the day.

Consider first the supply-side parameters m, h and s. According to the document "Schéma Directeur du RER A" written in June 2012 by the STIF (Syndicat des Transport d'ÃŽle-de-France), 30 trains per hour are supposed to operate during the morning peak in the East-West direction on the RER A line. However, the frequency actually achieved over the 4-year period February 2008 to February 2012 was only 24.4 trains per hour.[43] The target value for number of trains was thus set to $m = 24$, and the headway was set to $h = \frac{60}{24} = 2.5$ mins.

Two types of bi-level train sets are operated during the morning peak:[44]

- MI2N train sets with 904 seats and standing room for 1,636 users (4 users/$m\hat{A}^2$) for a total capacity of 2,540
- MI09 train sets with 948 seats and standing room for 1,683 users (4 users/$m\hat{A}^2$) for a total capacity of 2,614 users

This suggests a value for capacity of about $s = 2,600$. However, in the model users are assumed to travel from a single origin to a single destination whereas the RER A line serves many stations. La Défense is the most popular destination, but a substantial fraction of users pass through it. Only part of train capacity is thus effectively devoted to users who exit at La Défense. After experimentation with alternative values of s, and other parameters described below, we settled on a capacity equal to two-thirds of nominal train capacity so that $s = \frac{2}{3} \cdot 2,600 = 1,733$.

Consider now the demand-side parameters. According to a January 2011 document "Étude La Défense Analyse des Trafics" prepared by the DRIEA (Direction Régionale et Interdépartementale de l'Équipement et de l'Aménagement), in 2009, 32,600 users arrived at La Défense by RER A between 8:25am and 9:25am.[45] This count includes users traveling in both East-West and West-East directions, but it excludes users who are passing through. Including travel in both directions results in overestimation of traffic in one direction, whereas excluding users who pass through La Défense results in underestimation this traffic. Lacking an indication as to which bias dominates, we set $N = 32,600$.

Wardman et al. (2012) conduct a meta-analysis of estimates of β, γ and the value of travel time; call it α. They report point estimates of $\beta = 0.74 \cdot \alpha$ and $\gamma = 1.72 \cdot \alpha$ (see Table 19, p.25). For commuters in France, $\alpha =$€15/hr (see Table 15, p.21) which is consistent with the government-recommended value. This suggests setting $\beta = 0.74 \cdot 15$ = €11.1/hr, and $\gamma = 1.72 \cdot 15 =$ €25.8/hr. However, in the model it is assumed that users have the same desired

[43] See p.36 in http://www.stif.org/IMG/pdf/Deliberation_no2012-0163_relative_au_schema_directeur_du_RER_A.pdf.
[44] See p.54 in http://www.stif.org/IMG/pdf/Deliberation_no2012-0163_relative_au_schema_directeur_du_RER_A.pdf.
[45] See Figure 2 on page 8 in http://cpdp.debatpublic.fr/cpdp-grandparis/site/DEBATPUBLIC_GRANDPARIS_ORG/_SCRIPT/NTSP_DOCUMENT_FILE_DOWNLOADCB59.PDF.

arrival time, t^*. In reality, trip-timing preferences vary. The assumption of a common t^* leads to overestimation of schedule delay costs. In addition, with $\beta = €11.1$/hr and $\gamma = €25.8$/hr., condition (6) that all trains are used was violated given plausible values for other parameters. After experimentation with alternative values of β, γ and s (noted above) we scaled down β and γ by one-third to $\beta = €7.4$/hr and $\gamma = €17.2$/hr.

Empirical studies of public transit crowding often report crowding costs as time multipliers. This is consistent with evidence that disutility from crowding is proportional to amount of time spent in crowded conditions. The crowding cost parameter can then be written

$$\lambda = \alpha \cdot tt \cdot (tm - 1)\,, \tag{H.1}$$

where tt is travel time and tm is the time multiplier.

According to the survey "Étude mobilité transports à la Défense - Profils, usages et modes de déplacements des salariés et habitants du quartier d'affaires" by the EPAD (Établissement Public de la Région Pour l'Aménagement de la Défense), in 2006, the average travel time incurred by public transport riders who used only one transport mode to reach La Défense was 40 mins.[46] This is consistent with a study by the EnquÃªte Global Transport in 2010 which found an average travel time for commuters of 41 mins.[47] We thus set $tt = 40$ mins or 2/3 hrs.

Haywood and Koning (2015) have estimated time multipliers for Paris. They obtain a linear approximation of the time multiplier (see Eq. (10), p.194) of $tm = 1 + 0.11 \cdot d$, where d is the density of passengers per square metre. Substituting the estimates of α, tt and tm into Eq. (H.1) one obtains $\lambda = 15 \cdot 2/3 \cdot 0.11 \cdot d$. With a density of 4 users/$m\hat{A}^2$ for standing room on the train sets used on the RER A line (see above), this yields $\lambda = 4.4$.

I Sensitivity analysis

I.1 Integer-valued number of trains

The number of trains, m, has been treated as a continuous variable although it is discrete in reality. An integer constraint can be imposed by fixing m, and then choosing s using first-order conditions given in Prop. 12, 13 and 14 for regimes n, u and o respectively. To assess how the integer constraint affects results, m was first set to the largest integer smaller than the real-valued solution and then the next integer larger. Thus, for the no-fare regime m was first set to $\lfloor m_*^n \rfloor$ and then $\lfloor m_*^n \rfloor + 1$. Since m_*^u was calibrated to be an integer value, this was unnecessary for regime u. The integer value yielding the higher social surplus was then selected. The results changed very little, and social surplus was virtually unchanged. Integer constraints also had little effect for a range of other parameter values.

[46] See p.11 in `http://www.ladefense-seine-arche.fr/fileadmin/site_internet/user_upload/8-ENLIEN/etudes/etude-mobilite-transports.pdf`.

[47] See p.3 in `http://www.driea.ile-de-france.developpement-durable.gouv.fr/IMG/pdf/Fiche_Actifs__cle0cecb9.pdf`

Table 4: Comparison of no-fare, optimal uniform fare, and SO-fare (i.e., social optimum) regimes: $\eta = -2/3$

	Fare regime		
	No-fare (n)	Optimal uniform fare (u)	Social optimum (o)
m	26.34	24	26.75
s	1,764	1,733	1,725
N	41,006	32,600	33,220
p	6.72	9.48	9.22
$Rev/user$	0	3.45	3.39
TCC	187,604	133,499	112,503
SDC	88,044	63,244	81,248
TC	275,648	196,743	193,751
K	139,632	134,889	137,558
R	0	112,407	112,503
ρ	0	0.833	0.818
CS	1,206,851	1,106,343	1,115,033
SS	1,067,219	1,083,862	1,089,978
$Totalgain$		16,643	22,759
$Gain/user$	0	0.51	0.70
$Rel.eff$	0	0.73	1

I.2 Demand elasticity

If the price elasticity of demand is reduced to $\eta = 0$, ridership is the same in the three fare regimes. With $p_N = -\infty$, the first-order conditions (10a) and (10b) for s and m are the same for regimes n and u so that $s^u_* = s^n_*$, and $m^u_* = m^n_*$. Imposing the uniform fare yields no welfare gain at all, and merely transfers money from users to the transit authority. The SO-fare does yield a welfare gain although (with ridership fixed at 32,600) it is only €0.185 compared to €0.45 in the base case.

To examine the effects of a higher price elasticity, η was doubled in magnitude to $-2/3$.[48] To maintain equilibrium ridership at 32,600 in the optimal uniform-fare regime, parameter N_0 was increased to 146,056. The results are shown in Table 4. With the higher price elasticity, consumers' surplus and social surplus in each regime are lower than with the base-case parameters. Regime u is otherwise unaffected. However, the welfare gain per rider nearly doubles from €0.27 to €0.51. The welfare gain per rider in the social optimum increases from €0.45 to €0.70, but by a smaller percentage so that the relative efficiency of regime u increases.

J Glossary

J.1 Latin characters

c : user cost of a trip [€/user]

CS : total consumers' surplus [€]

e : superscript for uniform-fare regime

[48] Few transit services are likely to face such a high elasticity – especially during peak travel times.

$g(n)$: expected crowding cost function [€/user]

G^{xy} : welfare gain in shifting from pricing regime x to y

h : time interval between successive trains [hr/train]

k : index of train

K : capacity cost function [€]

m : number of trains used [trains]

MEC : marginal external cost of a trip [€/user]

MSC : marginal social cost of a trip [€/user]

n : superscript for no-fare regime

n : number of users on a train [users]

n_k : number of users taking train k [users/train]

N : total number of users [users]

o : subscript for socially-optimal fare regime

p : private trip cost including fare [€/user]

R : total fare revenue [€]

RV : variable fare revenue from socially optimal fare schedule [€]

s : measure of train capacity [users/train]

SDC : total schedule delay costs [€]

SS : social surplus [€]

t : departure time from origin station [clock time]

t^* : desired arrival time at destination [clock time]

TC : total user costs [€]

TCC : total crowding costs [€]

u : superscript for optimal uniform-fare regime

$v(n)$: marginal social crowding cost function [€/user]

J.2 Greek characters

β : cost per minute of arriving early [€/(hr·user)]

γ : cost per minute of arriving late [€/(hr·user)]

δ : schedule delay cost function [€/user]

ε : elasticity of $g'(n)$

η : elasticity of demand

λ : crowding cost parameter [€/user]

τ : fare [€/user]

ν_0 : Capacity cost function coefficient on m [€/train]

ν_1 : Capacity cost function coefficient on $m \cdot s$ [€/user]

ν_2 : Capacity cost function coefficient on s [€·train/user]