Document de travail du LEM / Discussion paper LEM 2019-23

Urban transport in polycentric cities

Quentin DAVID

LEM UMR 9221 / quentin.david@univ-lille.fr

Moez KILANI

LEM UMR 9221 / moez.kilani@univ-littoral.fr

Urban transport in polycentric cities

Quentin David* Moez Kilani†

September 30, 2019

Abstract

We develop a model of urban transport in polycentric cities and evaluate the impact of different policies on modal split and welfare. We consider a city composed of a city center and two suburban areas. The transit network connects each suburban area to the city center. To go from one suburban area to the other, commuters have either to transit through the city center or to use a private car. Cars are thus overused, in particular in the outskirts. First, we show that an optimal pricing scheme can restore efficiency. Second, we compare three administration regimes (public, semi-public and private) and discuss their relative efficiencies. Third, we show that opening a new transit line linking the suburbs increases total welfare, but is not Pareto-improving unless crowding is high on existing transit lines.

Keywords: Polycentric city; Urban transport; External costs; Transport investment

*Université de Lille, Lille, France. LEM - Lille Économie Management (CNRS - UMR 9221) email: quentin.max.david@gmail.com

†Université du Littoral, Côte d'Opale (ULCO), Dunkerque, France. LEM - Lille Économie Management (CNRS - UMR 9221) email: moez.kilani@univ-littoral.fr

1 Introduction

Urban growth generally involves the development of economic activities in the outskirts, markedly changing the structure of traffic flows. For instance, Aguilera et al. (2009), describing the evolution of traffic flows in the metropolitan region of Paris, show that the proportion of standard commuting from the suburbs to the city center has declined, while reverse commuting from the center to the suburbs, and commuting between suburbs, have both increased. This pattern is seen in many large urban areas.

In older cities, the urban public transport systems were initially designed with radial lines when urban planners were dealing with commutes between suburbs and city centers. As a result, commuting by public transit between suburban areas usually requires an inconvenient transit via the city center, encouraging the use of private cars for these trips. Growing concerns about congestion, environmental issues and public health prompted several local authorities to explore reforms of urban transport to curb the use of private cars and reduce the external costs generated. Road pricing is an efficient tool to reduce external costs, but in practice, users tend to oppose it. When road pricing is unfeasible, other reforms, such as discounted fares on public transport, are frequently explored (cf. Parry and Small, 2009). Instead of centralized decision making, the privatization of some transport activities may enable different pricing schemes (cf. de Palma et al., 2007). Improving public transport provision is also a way to reduce the use of private cars. Improving service quality or investment in new lines connecting suburbs are likely to make public transport more attractive by reducing waiting time and obviating tedious transit through the city center.

In this paper, we study urban commuting in a polycentric city composed of one city center and two suburban areas.[1] We consider two transport modes: public transit linking each suburban area to the city center, and private cars. We study the modal choice made by exogenously located workers for all origin-destination (OD) pairs and address two important questions related to commuting efficiency. First, we examine the efficiency of various administration regimes for each mode. We consider unpriced equilibrium, fully public, semi-public and private regimes. In semi-public regimes, transit is managed by the public sector, and roads are managed by a private operator. Second, we consider the investment in a new circular transit line directly connecting suburbs, and evaluate its impact on average user cost. Our analysis takes some account of service quality in public transport as measured by service frequency.

We show that the unpriced equilibrium is not optimal (it does not minimize aggregate

[1]In the literature, the city center is sometimes called the central business district (CBD) and suburban areas suburban business districts (SBDs). Employment patterns do not concern us here, and so we prefer to use the more general terms city center and suburban areas.

user cost) because it leads to an overuse of private cars, especially for trips between the suburbs. The optimum can be achieved in a fully public regime either by imposing a road toll or by subsidizing public transport. We show that there is always an optimal fare-toll gap that yields optimal modal split as an equilibrium. If the administration of the roads is delegated to a private operator, then this operator sets the tolls at a high level, leading to an equilibrium where public transport is overused and crowded. When the regulator can choose the level of public transport fares but the roads are administered by the private operator, we show that the optimum can be reached. We also discuss the case of a duopoly where the two transport modes are managed by separate private operators. In this case, tolls and fares are strategic complements, and we obtain an equilibrium where pricing of the two modes is set too high.

Developing a transit network directly connecting the suburbs is always of benefit to commuters traveling between those suburbs, but then service frequency is likely to be reduced on radial lines following decrease in demand. This can lengthen waiting time for some users and so can limit the overall benefit of the investment. On crowded lines, however, a moderate decrease in demand will not cause any significant drop in service. The net impact of the new line is then likely to be positive on aggregate user cost.

Our analysis is extended by numerical illustrations of several configurations. For this purpose, we wrote a standalone Fortran program specifically adapted to our setup and available online.[2] In the numerical illustration, we show that under standard conditions, opening a new transit line between the suburbs increases total welfare, but the improvement is driven by the effect on suburb-to-suburb commuters. For the other passengers, commuting costs increase as the service frequency on their lines decreases. Hence if such users could vote on constructing the new line, it is likely that they would oppose it unless crowding was high. A decrease in the number of passengers results in lower service frequency. Kilani et al. (2017) characterized this outcome, though in a framework with a single transport mode.

The analysis of the two-mode problem developed in this paper has to allow for the network structure adopted. In the previous literature on modal choice, the network and city structure are either not explicitly specified (as in Mohring (1972), David and Foucart (2014)), monocentric (see Hamilton (1982), Kilani et al. (2014)) or composed of a single OD pair (Tabushi (1993), Verhoef and Small (2004), de Palma et al. (2008)). To our knowledge, ours is the first analytical model of modal choice in a city with multiple business centers, and thus a relatively complex OD matrix. We use a general transport cost function and adopt a methodology that keeps the model tractable while deriving most results analytically. The solution to the model is derived using a mathematical

[2] `https://sites.google.com/site/quentinmaxdavid/research`

representation of traffic equilibrium initially proposed by Beckman and McGuire (1956). Network equilibrium flows correspond to a set of equations and inequalities that are not easy to solve directly. Instead, the problem is formulated as a nonlinear convex programming problem.[3]

Our framework implicitly assumes that not all workers locate efficiently. This is out of line with the prediction of the monocentric model. There are various reasons for making this assumption, echoing the debate raised by Hamilton (1982), who empirically observed that a large share of home-to-work trips occurs between different business centers. We refer the interested reader to Giuliano and Small (1993) who have already discussed job-housing balance/imbalance and over-commuting, defined as the difference between actual commute and the commute required to access jobs when people are efficiently located. The issues and emergence of spatial mismatch and multicentricity are not directly addressed here, as we focus on commuting behaviors between existing business districts.

The paper is organized as follows. In the next section, we set up the theoretical model. The equilibrium and the optimality conditions are derived in Section 3 where we also consider various administrative regimes for roads and transit. Section 4 analyzes public transport provision. We examine the determinants of transit frequencies and consider a modification of the transit network: investment in a new transit line. To illustrate the main findings, we develop a linear version of the model and a numerical example in Section 5. Section 6 concludes.

2 The model

We consider a city with a single city center and two suburban areas. There are both radial and circular roads directly connecting suburban areas and the central area but there are only radial transit lines connecting suburban areas to the city center. To go from one suburban area to the other, a transit user must therefore travel first into the city center and then back out to their destination. This is a simplification of realistic situations with a central main square and multiple suburban areas. A number of cities have developed a radial transit network where transit lines run out from the city center to suburban districts. The Paris region is one such case.

This paper evaluates transport costs in this context and explores policy reforms that can increase urban welfare (pricing, extension of transit network and changes in transit

[3] With a single OD network it is possible to solve the equilibrium equation directly and check whether or not it yields an interior solution. Since the number of possible cases is limited, most authors do this to solve their simple models. However, with a slightly more complex network, the enumeration of all possible cases becomes impractical, and our procedure is both useful for analytical tractability and efficient for numerical computation.

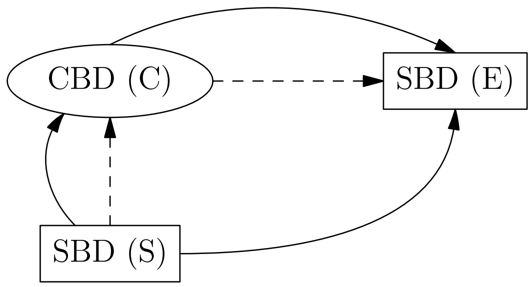

Figure 1: The network structure: roads are in solid lines, rails are in dashed lines

frequencies). These issues are of great importance for many urban areas where transit developed first through radial lines. As a result, most trips between suburban areas are made by car (see Aguilera (2005), for instance). These are relatively long-distance trips, potentially responsible for high CO_2 emissions and congestion. The objective of many reforms in the transport sector is to reduce the use of cars in metropolitan areas through a better provision of public transport services. Some regions such as Paris have ongoing projects for the development of circular transit lines.

The main district is denoted C, for "Center", the first suburban area is denoted S for "South" and the second E for "East". Fig. 1 depicts the geometry of this city. Dashed lines show public transit, continuous lines roads. As indicated by the arrows, we look only at commuting from the South to the Center and to the East, and from the Center to the East, and not the reverse. We assume that the flows of commuters are symmetrical.

The modal choice of commuters is deciding whether to use either a car on the road (private mode, denoted R) or transit (public mode, denoted T). There are three groups of commuters, differing in their residential location (origin) and workplace (destination). The group sizes are respectively denoted N_{sc}, N_{ce} and N_{se}, where the subscripts (ij) refer to the OD pairs. Thus $i = C, S$ and $j = S, E$. We define n_{ij}^M as the number of users going from i to j, using mode M, with $M = R, T$. We have $n_{ij}^R + n_{ij}^T = N_{ij}$. As shown in Fig. 1 there is a total of five links, two transit links and three road links. On each link (ij), the total number of users of mode M, is denoted u_{ij}^M. More precisely, we have $u_{ij}^R = n_{ij}^R$ for $(ij) \in \{(sc), (ce), (se)\}$ and $u_{ij}^T = n_{ij}^T + n_{se}^T$ for $(ij) \in \{(sc), (ce)\}$.

These trips conflict with the prediction of the monocentric city, whereby each worker commutes to the nearest business area. The term "wasteful commuters" was introduced by Hamilton (1982) to characterize these trips. Hamilton found that wasteful commuting was very common in the metropolitan areas he observed (Los Angeles). The empirical conclusion was first criticized by White (1988), before it was confirmed by Small and Song (1992). It is now considered as a major shortcoming of the monocentric city model (cf. Brueckner, 2011). We do not address this question, and do not consider why a household might locate "inefficiently", but accept the empirical evidence that "wasteful commuting" does exist.

Transport is costly for all modes. Road users drive directly to their destinations. It is assumed that those who drive between suburbs do not use radial roads. Congestion on roads depends on (and is increasing in) the number of users on them. Crowding in transit depends on the number of users of transit, and the cost function depends on the number of users and the service frequency. Transport cost functions are defined for each link connecting pairs of business locations i and j. They are denoted c_{ij}^M and depend on the number of users of the same link, denoted u_{ij}^M. For a given link (i, j), the generalized transport cost using the road is given by

$$C_{ij}^R(u_{ij}^R) = \tau_{ij} + c_{ij}^R(u_{ij}^R) \qquad \text{for} \qquad (ij) = \{(sc), (ce) \text{ or } (se)\}, \tag{1}$$

where τ_{ij} denotes a road toll imposed on the users and $c_{ij}^R(u_{ij}^R) = F_{ij}^R + \tilde{c}_{ij}^R(u_{ij}^R)$ is the monetary value of the time spent for the commute. It is the sum of a free-flow travel cost, F_{ij}^R, and the user cost due to congestion, $\tilde{c}_{ij}^R(u_{ij}^R)$. The free-flow travel cost encompasses the monetary value of the travel time and the vehicle operating cost. We assume that $\tilde{c}^R(u^R)$ is twice-differentiable with $\partial\tilde{c}^R(u^R)/\partial u^R > 0$ and make no specific assumption on the second-order derivative at this stage. For public transport, the generalized transport cost on a single link (ij)[4] for the users is given by

$$C_{ij}^T(u_{ij}^T, f_{ij}) = p_{ij} + c_{ij}^T(u_{ij}^T, f_{ij}) \qquad \text{for} \qquad (ij) = \{(sc),\ \text{or } (ce)\} \tag{2a}$$

where p_{ij} is the transport fares paid by the users of the public mode and $c_{ij}^T(u_{ij}^T, f_{ij}) = \frac{c_w}{2f_{ij}} + F_{ij}^T + \tilde{c}_{ij}^T(u_{ij}^T, f_{ij})$ is the monetary value of the time spent commuting by train. It is composed of the waiting time at the train station, the crowding-free travel cost, F_{ij}^T, and the monetary value of crowding when there are u_{ij}^T passengers in the train and a train frequency of f_{ij} on the line. Assuming that transit passengers arrive uniformly at the station (i.e. they do not use timetables), the waiting time is given by $c_w/2f_{ij}$ where c_w is the monetary value of the maximum waiting time between two trains. We assume that crowding costs increase with passengers and decrease with frequency, i.e.

$$\frac{\partial\tilde{c}^T(u^T, f)}{\partial f} < 0 \text{ and } \frac{\partial\tilde{c}^T(u^T, f)}{\partial u^T} > 0,$$

and make no specific assumptions on the second-order partial derivatives. On the (se) link, there is no direct transit line. Commuters must transit through C. Their travel cost

[4] A single link is defined as either (sc) or (ce). When using transit, the link (se) is composed of two single links.

is given by

$$C^T_{se}(u^T_{sc}, u^T_{ce}, f_{sc}, f_{ce}) = \beta(p_{sc} + p_{ce}) + c^T_{sc}(u^T_{sc}, f_{sc}) + c^T_{ce}(u^T_{ce}, f_{ce}) - \Gamma(\alpha). \tag{2b}$$

Transit users of the (se) line are assumed to pay a fare proportional to $(p_{sc}+p_{ce})$. In the real world, fares are generally higher than $\max\{p_{sc}, p_{ce}\}$ and smaller or equal to $(p_{sc}+p_{ce})$. Values of parameter β can be set to capture these and other situations. For example, $\beta < 1$ would correspond to a subsidy provided to users (se). The last term in (2b) reflects the switching cost at station C that depends on the synchronization between the arrivals of trains (sc) and the departures of trains (ce). Although transit users bear the monetary value of the time spent on each line, we add a term that captures the possibility of improving the connectivity between the two lines. We assume that the higher α, the better the coordination between the two lines. If the same train goes from S to E with a stop in C, $\alpha = 1$ and the transit user does not incur the waiting time in C. If $\alpha = 0$, there is no coordination between the two lines and the transit user incurs the full cost of waiting at the two train stations. An additional waiting cost (c_w) may be considered. The value of α ranges between 0 and 1 and corresponds to the quality of synchronization. It is possible that the same train starts at S, goes to C and continues to E. In this case there is no switching cost and perfect synchronization between the two trains. A small value of these parameters may reflect the intermediate stop at station C. Parameter α can thus be chosen to describe a variety of situations. Timetables for the services are assumed to adopt uniform schedules, and train loadings to be equal on all vehicles, so it is straightforward to compute both waiting time and crowding cost. All costs are expressed for a whole trip.

The transport sector is administered by one or two operators (one for each mode) who can be either private (profit-maximizing) or public (welfare-maximizing). The choice variable for the commuters is the transport mode. The operator of the roads can decide to impose a toll on a given link, and the operator of public transport decides the fares and service frequencies. For public transport, there is also the possibility of extending the transit network to make a direct connection between suburban areas S and E. We first consider that there is no cost for administering the roads or the railways. In this context, frequencies (f_{ij}) and coordination of the transit system (α) are considered exogenous. We relax these assumptions in an extension and use a cost of providing operating vehicles and of synchronizing the two lines.

3 Equilibrium, optimum and administration regimes

In this section, we use a general cost function to characterize the equilibrium and the optimum. We compare them and provide the conditions for the decentralization of the optimum. We then consider various administration regimes. Under a public regime, both roads and rail transport are assumed to be managed by the social planner. Under a semi-public regime, roads are administered by a private operator, and rail transport is administered by the social planner. In a duopoly scenario, both public transit and roads are administered by private operators. We compare pricing schemes and welfare in the three scenarios.

In this section, service frequencies (f_{ij}) and synchronization (α) are assumed to be exogenous and cost-free.[5] As a consequence, we focus on the social optimum for commuters. In the next section, we relax these assumptions and look at transport provision.

3.1 Equilibrium

Workers are assumed to commute from their dwellings to their workplaces using either private cars or public transit. As demand is perfectly inelastic, we can rewrite n_{ij}^R as functions of n_{ij}^T:

$$n_{sc}^R = N_{sc} - n_{sc}^T, \qquad n_{ce}^R = N_{ce} - n_{ce}^T, \qquad \text{and} \qquad n_{se}^R = N_{se} - n_{se}^T,$$

respectively, and we are left with three endogenous variables: n_{sc}^T, n_{ce}^T, n_{se}^T. This problem meets the Wardrop equilibrium conditions: for a given set of commuters, the user cost in the private mode is equal to the user cost in the public mode. If a mode is not used it must be associated with higher cost. In an interior solution, when both the public and the private modes are used, we have, in equilibrium: $C_{ij}^T(u_{ij}^T) = C_{ij}^R(u_{ij}^R)$ for all $(ij) = (sc), (ce)$ and (se), i.e.

$$\tau_{sc} + c_{sc}^R(u_{sc}^R) = p_{sc} + c_{sc}^T(u_{sc}^T), \tag{3a}$$

$$\tau_{ce} + c_{ce}^R(u_{ce}^R) = p_{ce} + c_{ce}^T(u_{ce}^T), \text{ and} \tag{3b}$$

$$\tau_{se} + c_{se}^R(u_{se}^R) = \beta(p_{sc} + p_{ce}) + c_{sc}^T(u_{sc}^T) + c_{ce}^T(u_{ce}^T) - \Gamma(\alpha). \tag{3c}$$

We note that in a corner solution, some modes (or links) may not be used, and the above conditions associated with these modes (or links) do not hold. We have the following

[5] As f_{ij} is assumed to be exogenous in this section. For ease of reading, it will be removed from the expressions. For instance, $C_{ij}^T(u_{ij}^T, f_{ij})$ will be denoted $C_{ij}^T(u_{ij}^T)$. This assumption will be relaxed in the next section, and f_{ij} will be reintroduced in the equations.

result:

Proposition 1 (Equilibrium) *The problem of modal choice has at least one equilibrium: (i) If the set of equations $C^T_{ij}(u^T_{ij}) = C^R_{ij}(u^R_{ij})$, for $(ij) = (sc), (ce)$ and (se), has a feasible solution then it is the sole interior equilibrium (where each group uses both transport modes). (ii) In all other cases, the problem has at least a corner solution and some groups do not use both transport modes.*

Proofs are in Appendix A. To obtain this result, we transform the equilibrium problem into a minimization problem for commuters, with constraints for the variables. We show that the objective function is convex. The stability of the equilibrium is straightforward in this problem since an additional user will always increase the user cost. Any deviation is costly, and at the equilibrium, every user pays their minimum travel cost.

3.2 Optimum

The total cost is the sum of the users' and the operators' costs. An optimum is reached when the total cost is minimized. As there are no operating costs for the roads, the total cost is the sum of users' costs and the cost of operating the transit system. This latter depends on the frequency of the services and the effort made to coordinate the two lines. As these values are exogenously fixed for the time being, the operator's cost is fixed. Since transit fares and road tolls are redistributed to the population, they are welfare-neutral. The objective is therefore to minimize the following social-cost function for n^T_{sc}, n^T_{ce} and n^T_{se}:

$$\sum_{ij=sc,ce,se} u^R_{ij}\, c^R_{ij}(u^R_{ij}) + \sum_{ij=sc,ce} u^T_{ij} c^T_{ij}(u^T_{ij}) - n^T_{se}\Gamma(\alpha). \tag{4}$$

The endogenous variables should satisfy the usual constraints, i.e. $0 \le n^T_{sc} \le N_{sc}, 0 \le n^T_{ce} \le N_{ce}$ and $0 \le n^T_{se} \le N_{se}$. We have the following result.

Proposition 2 (Optimum) *There exist traffic flows n^M_{ij} that satisfy the constraints and minimize the total cost function in (4). If there is an interior solution and if $\partial^2 c^M_{ij}/\partial u^2 \ge 0$ for $M = T, R$, then there is a single solution.*

There are various standard formulations of congestion (e.g. the BPR or the quadratic formulations) that satisfy the second-order condition for $c^T_{ij}(u)$. The convexity of the cost function in public transport may not be satisfied if the availability of seats is taken into account. The MAS formulation (cf. de Palma et al., 2015), for example, has a second order derivative, which is not always positive. In the general case we can have multiple solutions. We show in the proof that the above condition on the second-order derivative is a sufficient condition for a single solution because it guarantees the convexity of the objective function.

The equilibrium is generally distinct from the optimal solution. The first-order conditions with respect to the objective function in (4) yield

$$c_{ij}^T(u_{ij}^T) + u_{ij}^T \, \frac{\partial c_{ij}^T(u_{ij}^T)}{\partial n_{ij}^T} = c_{ij}^R(u_{ij}^R) - u_{ij}^R \, \frac{\partial c_{ij}^R(u_{ij}^R)}{\partial n_{ij}^T} \tag{5a}$$

for groups $ij = \{sc\}, \{ce\}$, and

$$\sum_{ij=sc,ce} \left(c_{ij}^T(u_{ij}^T) + u_{ij}^T \frac{\partial c_{ij}^T(u_{ij}^T)}{\partial n_{se}^T} \right) - \Gamma(\alpha) = c_{se}^R(u_{se}^R) - u_{se}^R \frac{\partial c_{se}^R(u_{se}^R)}{\partial n_{se}^T} \tag{5b}$$

for $ij = \{se\}$, which is a usual statement that the optimum distribution of users is such that the social marginal costs, not the private marginal costs, are equal for all the alternatives used.

3.3 Administration regimes

In most cases, urban transit systems are controlled by public authorities because of their cost structure, and because this activity is not in general profitable. To what extent the public operator should set tolls or fares to induce a decrease in transport cost is a subject of debate, and should take into account the externalities produced by the transport system.

In this section, we consider various scenarios for the management of the transport system. We start by considering a fully public administration where the social planner administers both the roads and the transit system. We then turn to the case where the administration of roads is delegated to a private operator. Finally, we look at a fully private administration where each transport system is administered by a private operator. We call this case a duopoly administration.

3.3.1 Public administration

By comparing (5) with the equilibrium conditions (3), we see that the optimum can be decentralized if

$$p_{ij}^{op} - \tau_{ij}^{op} = u_{ij}^{T} \frac{\partial c_{ij}^{T}}{\partial n_{ij}^{T}} + u_{ij}^{R} \frac{\partial c_{ij}^{R}}{\partial n_{ij}^{T}} \text{ for } (ij) = (sc) \text{ and } (ce) \tag{6a}$$

$$\beta^{op}(p_{se}^{op} + p_{ce}^{op}) - \tau_{se}^{op} = \sum_{ij=sc,ce} u_{ij}^{T} \frac{\partial c_{ij}^{T}}{\partial n_{se}^{T}} + u_{se}^{R} \frac{\partial c_{se}^{R}}{\partial n_{se}^{T}}, \tag{6b}$$

where the superscript "*op*" denotes the optimum. We have the following result.

Corollary 1 (Decentralisation of the optimum) *Any pricing scheme such that the differences between the traffic fares and the road tolls correspond to the differences between the marginal social cost of crowding and of congestion ensures the optimum is reached.*

Having control over one of the two tools (tolls or fares) is sufficient to reach the optimum so long as the social planner can set tolls or fares such that (6a) and (6b) are satisfied. A typical regime for the decentralization of the optimum is when public transport is unpriced and roads are tolled according to (6) with $p_{ij} = 0$. We note that pricing public transport only can also lead to the optimum, but it is a little more difficult to implement in practice because we have to distinguish users (SE) from the other two groups. Generally, flat pricing of public transport with similar fares for all groups, which is used in several cities (and is being debated for the Paris region), will not yield the optimum without road pricing. The external cost considered could also reflect emissions of pollutants. This would lead to greater distortions in equilibrium.

3.3.2 Semi-public administration

In this section, roads are assumed to be operated by a private operator whose objective function is to maximize profit. The transit system is operated by a public agent whose objective function is to minimize total transport cost.

Generally speaking, a private operator will increase its revenues by imposing higher tolls on road users. For the social planner, two scenarios are of interest. In the first one, the public operator sets fares at zero. If tolls are higher than the optimal ones, then compared with the first-best situation, the private mode will be underused. In the second scenario we let the public operator increase the public transport fares to return to the optimum.

Operating the roads is assumed to be costless. The private operator earns the sum of the toll revenues collected on the three roads. It is given by

$$\pi^R(\tau_{sc}, \tau_{ce}, \tau_{se}, p_{sc}, p_{ce}) = \sum_{ij=sc,ce,se} \tau_{ij}\, n_{ij}^R. \tag{7}$$

The operator is constrained by the equilibrium choice of users described in equations (3). The first-order conditions with respect to the tolls yields $\lambda_{ij} = -(N_{sc} - n_{sc}^T) \leq 0$ for $\{ij\} = \{sc\}, \{ce\}, \{se\}$, where λ_{ij} is the multiplier of constraint (3). Substituting in the first-order conditions with respect to n_{ij}^T shows that the tolls imposed by the private operator satisfy

$$\tau_{ij}^{sp}(\cdot) = (u_{ij}^R + u_{se}^R)\frac{\partial c_{ij}^T}{\partial n_{ij}^T} - u_{ij}^R\, \frac{\partial c_{ij}^R}{\partial n_{ij}^T} \geq 0 \tag{8a}$$

for $(ij) = (sc), (ce)$ and

$$\tau_{se}^{sp}(\cdot) = \sum_{ij=sc,ce} (u_{ij}^R + u_{se}^R)\frac{\partial c_{ij}^T}{\partial n_{se}^T} - u_{se}^R\, \frac{\partial c_{se}^R}{\partial n_{se}^T} \geq 0. \tag{8b}$$

We note that $\partial c_{ij}^R / \partial n_{ij}^T < 0$. Comparing (8) with (6), we therefore see that for the same level of transit fares, the road operator imposes tolls that are higher than the optimum level. The following proposition states that when public transport is unpriced, the private operator imposes tolls that are higher than the optimum tolls.

Proposition 3 (Unpriced public transport) *If the social planner makes the transit free ($p_{ij} = 0$), road tolls that are imposed by the private operator are higher than those that decentralize the optimum, i.e. $\tau_{ij}^{sp}(p_{ij} = 0) \geq \tau_{ij}^{op}(p_{ij} = 0)$ for $(ij) = (sc), (ce), (se)$. In this case, public transport is overused by all user groups, i.e. $n_{sc}^{T,sp} \geq n_{sc}^{T,O}$, $n_{ce}^{T,sp} \geq n_{ce}^{T,sp}$ and $n_{se}^{T,sp} \geq n_{se}^{T,O}$.*[6]

Tolling roads and subsidizing public transport for economic efficiency is a prevalent idea. Proposition 3 confirms that road tolls will reduce the use of roads but lead to overuse of public transport. A similar conclusion can be found in Kraus (2012) and Kilani et al. (2014). With unpriced roads it is optimal to reduce fares below the marginal social cost, but when Pigovian tolls are imposed on road users it is optimal to raise fares so that crowding costs are endogenized. The next proposition shows that it is always possible for the public operator to reach the optimum.

[6]Superscript "O" refers to the optimum.

Proposition 4 (Pricing scheme in the semi-public regime) *The public operator can reach the optimum by setting full fares for users* (se)*, i.e.* $\beta = 1$*, and fares*

$$p_{ij}^{sp}(\cdot) = (N_{ij} + N_{se})\frac{\partial c_{ij}^T}{\partial n_{ij}^T} \geq 0 \tag{9}$$

for the other two groups $(ij) = (sc), (ce)$*. The private operator then imposes roads tolls that are higher than those it would impose if public transit was free.*

The semi-public regime induces a strategic competition between the public and private operators. The strategic variables in this duopoly, fares (choice variable for the public operator) and road tolls (choice variable for the private operator), are strategic complements. As a result, the equilibrium is characterized by overpricing, but the public operator is able to reach the optimum, as shown in Proposition 4. In practice, this may raise the issue of acceptability, because users of the modes pay higher fares and higher tolls. The reaction function of the public operator is obtained from Eq. (9), and the reaction function for the private operator from Eq. (8).

From the expressions of the fares in Proposition (4) it is clear that user price is proportional to marginal social cost. If there is no crowding in public transport, then the road operator will impose optimal tolls. The next proposition states this result.

Corollary 2 (No crowding) *If there is no crowding in public transport* $(\partial c_{ij}^T / \partial n_{ij}^T = 0)$ *the private operator imposes socially optimum road tolls when public transport is kept unpriced.*

In this case, the optimum can be easily achieved through the privatization of roads and by charging no fares in public transit. An illustration in a simplified network and a brief discussion of this proposition is provided in Appendix B. This result dates back to Knight (1924) and is usually quoted for the advocation of privatized road management. The scope of this result, however, is not general. It is well known that it is sensitive to two main assumptions (cf. Lindsey, 2012), both adopted in our framework. The first is the elastic demand in the network, and the second is the homogeneity of the users. Even if users are distinguished by their OD pairs, they have the same time values and they perceive the same magnitude of discomfort. If one of these two assumptions is not satisfied, then the private operator will impose a non-optimal road toll.

3.3.3 Private administration

In this section, we consider two distinct competing private operators free to set road prices and transit fares. The first one is operating the roads, as in the preceding section, and the second one is operating public transport. The operators do not incur any cost, and their profit is equal to their revenue. The road operator therefore maximizes objective function given in (7), as in the semi-public regime, and the public transport operator chooses transit fares to maximize

$$\pi^T = p_{sc} n_{sc}^T + p_{ce} n_{ce}^T + \beta (p_{sc} + p_{ce}) n_{se}^T \tag{10}$$

In both cases, and since we assume an interior solution, equilibrium conditions in (3) constrain each operator. The first-order conditions for the road operator are still given by (8) and the first order conditions for the transit operator are (where the superscript "du" stands for "duopoly"):

$$\tau_{ij}^{du} + c_{ij}^R + n_{ij}^T \frac{\partial c_{ij}^R}{\partial n_{ij}^T} = c_{ij}^T + u_{ij}^T \frac{\partial c_{ij}^T}{\partial n_{ij}^T} \text{ for } (ij) = (sc), (ce) \text{ and} \tag{11a}$$

$$\tau_{se}^{du} + c_{se}^R + n_{se}^T \frac{\partial c_{se}^R}{\partial n_{se}^T} = c_{sc}^T + c_{ce}^T + u_{sc}^T \frac{\partial c_{sc}^T}{\partial n_{se}^T} + u_{ce}^T \frac{\partial c_{ce}^T}{\partial n_{se}^T} - \Gamma(\alpha). \tag{11b}$$

Combining these first-order conditions with the equilibrium conditions, we have an expression for the fares:

$$p_{ij}^{du} = u_{ij}^T \frac{\partial c_{ij}^T}{\partial n_{ij}^T} - n_{ij}^T \frac{\partial c_{ij}^R}{\partial n_{ij}^T} \tag{12a}$$

$$\beta = \frac{u_{sc}^T \frac{\partial c_{sc}^T}{\partial n_{se}^T} + u_{ce}^T \frac{\partial c_{ce}^T}{\partial n_{se}^T} - n_{se}^T \frac{\partial c_{se}^R}{\partial n_{se}^T}}{p_{sc} + p_{ce}}. \tag{12b}$$

Optimal tolls set by the private operator on roads are given by equations (8). These reaction functions between two operators competing to attract an inelastic demand lead to a strategic complementarity between transit fares and road tolls.

4 Public transport provision

In the previous section, we looked at the pricing scheme for urban transport under various administration regimes. We now turn to the analysis of transport provision and address two key questions. First, we study public transport provision at given network structure. More precisely, we look at the frequencies of the service provided and the coordination between the two trains. Second, we discuss the consequences of modifying the network by building a new transit line between the two suburban areas.

For many cities whose public transport network is radial, the question of circular lines is of prime importance. Following the example of Paris, some cities have projects to add new peripheral transit lines to their urban transit networks.

4.1 Endogenous service frequency and synchronization

Consider that the operator chooses the frequency of the services and the level of the synchronization between the two trains (α) at the central station. The cost of providing operating vehicles and their synchronization is given by

$$\kappa \sum_{ij=sc,ce} f_{ij} + \upsilon(\alpha), \tag{13}$$

where $\kappa > 0$ denotes the unit cost of operating a vehicle (the summation is done over all links), and $\upsilon(\alpha)$ (with $\upsilon'(\cdot) > 0$) is the cost of deploying an effort to synchronize the two trains. The above term has to be added to the social cost function (4).

The first order conditions (5) remain unchanged, but additional conditions for frequencies and coordination emerge. At optimum, we must have:

$$\kappa = -u_{ij}^T \frac{\partial c_{ij}^T(u_{ij}^T, f_{ij})}{\partial f_{ij}} \text{ for } ij = sc \text{ and } ce, \text{ and} \tag{14a}$$

$$\upsilon'(\alpha) = n_{se}^T \Gamma'(\alpha), \tag{14b}$$

whose interpretation is straightforward. The marginal cost of increasing frequencies or improving synchronization (left hand sides) must equal their social marginal benefits (right hand sides). These expressions implicitly display positive links between service frequencies or the level of coordination and the number of transit users ($f_{ij}^{*'}(u_{ij}^T) > 0$ and $\alpha^{*'}(u_{ij}^T) > 0$). If a public operator is in charge of public transport, optimal service frequencies and coordination are easily achieved.

If a private operator is in charge of public transport, its objective function described in (10) becomes

$$\pi^T = p_{sc}n_{sc}^T + p_{ce}n_{ce}^T + \beta(p_{ce} + p_{ce})n_{se}^T - \kappa \sum_{ij=sc,ce} f_{ij} - \upsilon(\alpha),$$

subject to the same equilibrium conditions (3).

Lemma 1 (Decentralization of service frequencies and coordination) *The decision rule of the private operator with respect to service frequencies (f_{ij}) and with respect to the level of coordination (α) between the two transit lines is optimal.*

The solutions of the Lagrangian associated with the maximization problem of the private operator leads to the same first order conditions (14) as those associated with the optimum.

4.2 Investment in a circular transit line SE

The purpose of this section is to study the consequences of building a new, direct transit line between S and E. For simplicity, and without loss of generality, we will assume that building such a line costs nothing. Although this assumption is non-realistic, we will show that such a line can, in some cases, decrease social welfare. Such cases would be more likely to be found if we considered positive building costs. In addition, we will consider interior solutions and assume that every commuter uses the direct OD connection. In other words, we assume that any commuter going from i to j uses either the direct transit line or the direct road.

In this new setup, looking at the modal choice on each link is extensively discussed by David and Foucart (2014). The novelty of our approach consists in looking at the welfare consequences of a modification to the transportation network. Theoretically, the main difference from the previous sections is the equilibrium condition (3c), which becomes (we neglect road prices and transit fares, as our focus is on social welfare):

$$c^R_{ij}(u^R_{ij}) = c^T_{ij}(u^T_{ij}, f_{ij}), \forall\{ij\} \in \{sc, ce, se\}$$

with $u^M_{ij} = n^M_{ij}$, $\forall\{ij\} \in \{sc, ce, se\}$ and $M \in \{T, R\}$.

The total (social) cost function can be reorganized by mode and OD pairs:

$$C = n^T_{sc}c^T_{sc} + n^T_{ce}c^T_{ce} + n^T_{se}c^T_{se} + n^R_{sc}c^R_{sc} + n^R_{ce}c^R_{ce} + n^R_{se}c^R_{se} - \kappa \sum_{ij=sc,ce} f_{ij}$$

where the third term replaces $n^T_{se}(c^T_{sc} + c^T_{ce}) - \Gamma(\alpha)$ and $\upsilon(\alpha)$ drops as there is no longer any coordination issue for SE commuters using the transit lines.

Comparison between the two networks is difficult to perform formally without defining explicit functions for the cost parameters. Instead, we discuss the welfare effects on commuters of a new transit line under three scenarios. First, we assume congestion and crowding with fixed frequencies. Second, we assume congestion, no crowding in public transport, and endogenous frequencies. Third and last, we discuss the most general case, considering together congestion, crowding and endogenous frequencies. The main results for the six groups of commuters are summarized in Table 1.

The first scenario is easy to understand and easily identifies the forces in play. With fixed frequencies, we do not consider network externalities in transit. In this case, opening the new transit line reduces the number of commuters on the two other transit lines (SE commuters no longer use the SC or CE transit lines). These commuters are assumed to be better off (by revealed preferences). Commuters from S to C and C to E that were using the transit are also better off as there is less crowding. Some road users will therefore change their modal choice, reducing congestion on roads up to the point where the cost of using roads and the cost of using transit is equal again for each OD pair. In this scenario, opening a new transit line is Pareto-improving. Every commuter is better off.

In the second scenario, we consider endogenous service frequency and congestion on roads, but no crowding in transit. This scenario corresponds to the case of a congested city where transit is underused. With the new transit line, all SE commuters are assumed to be better off. Transit users get to E faster and without crowding, and congestion decreases because more users in this group choose public transport. On the other two lines, there are fewer transit users. Since there is no crowding, this does not increase their welfare. On the contrary, because there are fewer commuters using the transit, frequencies decrease, increasing the cost of using transit for both SC and CE commuters. Some commuters will change their modal choice and use the road. Congestion increases on roads, and frequencies decrease again. We are back at equilibrium when the costs of the two modes are equal. Compared with the original network, there are more road users and fewer users of transit on the SC and CE links. They all face higher costs. Only the SE commuters are better off. Under this scenario, the opening of a new transit line is very likely to decrease social welfare.

In our last scenario, we assume congestion, crowding and endogenous frequencies. As in the two previous scenarios, the impact on SE commuters is assumed to be positive. The impact on the SC and CE commuters depends on the relative forces described above. The new line decreases both crowding and frequencies (there are fewer commuters in these transit lines). If the former dominates on both transit lines, it would be Pareto-improving: if there is less crowding and the impact on frequencies is weak, we expect the new line to be Pareto-improving. If the latter dominates, it is not Pareto-improving and could even be welfare-decreasing. If the negative impact on SC and CE commuters is greater than

Table 1: Impact of opening a new SE transit line for commuters

		Congestion, crowding, fixed frequencies	Congestion, no crowding, endogenous frequencies	Congestion, crowding, endogenous frequencies
SC	$c^R_{sc} = c^T_{sc}$	↓	↑	↓ or ↑*
	$n^R_{sc} = 1 - n^T_{sc}$	↓	↑	↑ or ↓*
CE	$c^R_{ce} = c^T_{ce}$	↓	↑	↓ or ↑*
	$n^R_{ce} = 1 - n^T_{ce}$	↓	↑	↑ or ↓*
SE	$c^R_{se} = c^T_{se}$	↓	↓	↓
	$n^R_{se} = 1 - n^T_{se}$	↓	↓	↓
Conclusions on transport costs:		Pareto improving (transport costs decrease for all commuters)	Not Pareto improving Welfare increases if impact on SC and CE is lower than impact on SE.	Pareto improving if crowding effects overcome frequencies effects for SC and CE. Impact on welfare depends on relative impact on crowding, congestion and frequencies

*Arrow on the left if crowding effect dominates network effect (on frequencies), arrow on the right otherwise

the positive effect on SE commuters, welfare decreases.

To conclude, so long as frequencies are endogenous (a reasonable assumption), opening a new transit line between suburban areas may be desirable only if crowding in public transport is an issue. Otherwise, it is likely to be welfare-decreasing by reducing frequencies and increasing congestion on roads.

5 Linear formulation and numerical illustration

In this section, we adopt a simplified linear formulation to illustrate the properties of the model and the main results of our analysis. This formulation is then used for the numerical illustration.

5.1 The linear model

We use a specific (linear) cost function to derive analytical solutions and study various administration regimes for roads and public transit. Given the notation above, the specific formulation concerns functions $\tilde{c}^m_{ij}$. For roads, the cost on link (ij) is assumed to be

$$\tilde{c}^R_{ij}(u^R_{ij}) = a^R_{ij}\, u^R_{ij} \tag{15a}$$

where a_{ij}^R denotes the marginal external congestion cost on roads. For public transport on link (ij), the cost function is assumed to be

$$\tilde{c}_{ij}^T(u_{ij}^T, f_{ij}) = \frac{a_{ij}^T u_{ij}^T}{f_{ij}}. \tag{15b}$$

In an interior solution, the number of users of public transport on each line (u_{sc}^T and u_{ce}^T, with $u_{sc}^T = n_{sc}^T + n_{se}^T$ and $u_{ce}^T = n_{ce}^T + n_{se}^T$) is obtained from the equilibrium conditions (3). From these solutions, it is possible to derive reaction functions between the three groups of transit users in equilibrium. We have:

$$n_{sc}^T = \rho_{sc}(n_{se}^T), \quad n_{ce}^T = \rho_{ce}(n_{se}^T) \quad \text{and} \quad n_{se}^T = \rho_{se}(n_{ce}^T, n_{sc}^T). \tag{16}$$

These reaction functions are linear and decreasing in their arguments, reflecting the fact that users n_{sc}^T and n_{ce}^T compete with users n_{se}^T for transit lines. Under the linear formulation, both the equilibrium and the optimum conditions can be written in matrix form. This will be useful for discussing the decentralization of the equilibrium. Let matrix A be given by

$$A = \begin{pmatrix} a_{sc}^R + \frac{a_{sc}^T}{f_{sc}} & 0 & \frac{a_{sc}^T}{f_{sc}} \\ 0 & a_{ce}^R + \frac{a_{ce}^T}{f_{ce}} & \frac{a_{ce}^T}{f_{ce}} \\ \frac{a_{sc}^T}{f_{sc}} & \frac{a_{ce}^T}{f_{ce}} & a_{se}^R + \frac{a_{sc}^T}{f_{sc}} + \frac{a_{ce}^T}{f_{ce}} \end{pmatrix},$$

vector $x' = (n_{sc}^T, n_{ce}^T, n_{se}^T)$, and vectors b^F, b^A and b^τ be given by

$$b^\tau = \begin{pmatrix} \tau_{sc} - p_{sc} \\ \tau_{ce} - p_{ce} \\ \tau_{se} - \beta\left(p_{sc} + p_{ce}\right) \end{pmatrix}, \qquad b^A = \begin{pmatrix} a_{sc}^R N_{sc} \\ a_{ce}^R N_{ce} \\ a_{se}^R N_{se} \end{pmatrix}$$

and

$$
b^F = \begin{pmatrix} F_{sc}^R - F_{sc}^T - \frac{c_w}{2f_{sc}} \\ \\ F_{ce}^R - F_{ce}^T - \frac{c_w}{2f_{ce}} \\ \\ F_{se}^R - F_{sc}^T - F_{ce}^T - \frac{c_w}{2f_{sc}} - \frac{c_w}{2f_{ce}} - \Gamma(\alpha) \end{pmatrix}.
$$

We can show that an interior solution for the equilibrium problem can be set as a system of three equations of the form

$$
Ax = b^F + b^\tau + b^A. \tag{17}
$$

On the other hand, in the linear formulation described here, the first-order conditions for the minimization problem presented in (5) can be put in the matrix form $2Ax = b^F + 2b^A$. It follows that the optimum tolls are given by $b^\tau = -\left(\frac{1}{2}\right) b^F$.

The problem with the linear cost, no toll and no transit fare is illustrated on Fig. 2. The three planes correspond to the reaction functions shown above. At their intersection we find the equilibrium point P^e. The other point, denoted P^O, corresponds to the optimum. In this example, public transit is underused and the optimum requires imposing positive tolls on the three roads. The decentralization of the optimum can be explained on the basis of this illustration. The three planes denoted $\mathscr{P}_{ij}$ correspond to the three reaction functions shown above. From equation (17), we see that road tolls enter additively. As a consequence, variations in the tolls result in parallel movements of the planes $\mathscr{P}_{ij}$. Clearly, by appropriately moving the three planes, the equilibrium point P^e can be located anywhere, and in particular at the same location as P^O. We will discuss a numerical example with details in Section 5.2.

With the linear formulation, some static comparatives are possible by directly differentiating the solutions. Let us consider users $\{sc\}$. An increase in N_{sc} increases n_{sc}^T, while an increase in N_{se} leads to a decrease in n_{sc}^T. A less direct effect is that N_{ce} has the same impact as N_{sc}, but with smaller magnitude. An increase in N_{ce} will increase the use of public transit by this group (n_{ce}^T increases) and discourage users $\{se\}$ from using the same mode (causing a decrease in n_{se}^T). Overall, it has a positive impact on n_{sc}^T. An increase in the switching cost $\Gamma(\alpha)$ reduces n_{se}^T and thereby increases n_{sc}^T and n_{sc}^T. An increase in the waiting cost reduces the attractiveness of the public mode. The reduction in the number of users offers greater comfort, which can attract some additional users. This rebound effect is small in this case. For $\alpha = 1$, an increase in c_w reduces the attractiveness of

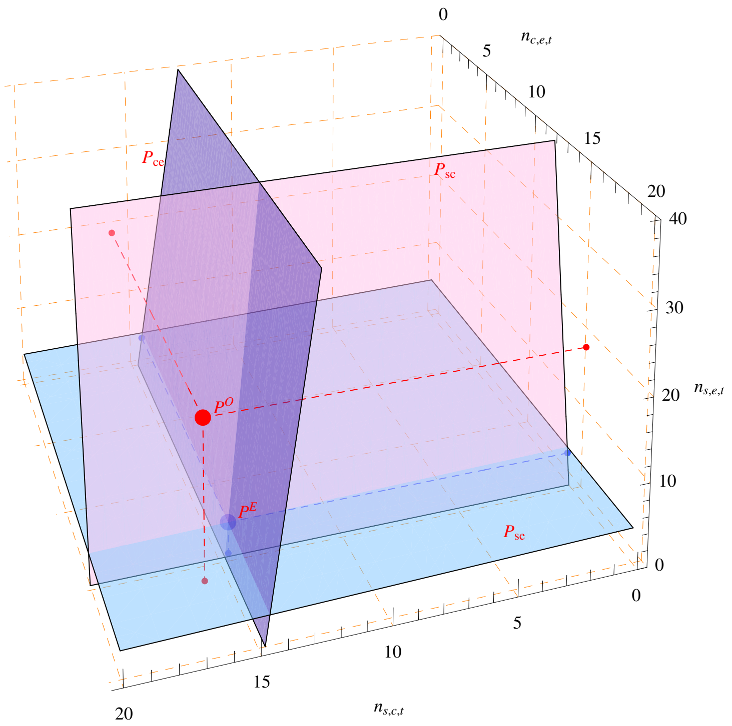

Figure 2: An illustration of the equilibrium.

public transit for all three groups of users. For a smaller value of α, an increase in c_w still has a negative impact on n_{se}^T and n_{ce}^T, but may lead to an increase in n_{se}^T.

Let us consider some specific situations. Writing the conditions for an interior solution is complex when there is crowding in transit and congestion on roads. The case of no crowding in transit is useful because it leads to simple expressions as stated in this result. A straightforward computation shows that without crowding in public transport, if $F_{ij}^R + a_{ij}^R N_{ij} > \frac{c_w}{2f_{ij}} + F_{ij}^T - \tau_{ij} + p_{ij} > F_{ij}^R$ for $\{ij\} = \{sc\}$ or $\{ce\}$, then group $\{ij\} = \{sc\}$ or $\{ce\}$ uses both transport modes (a similar condition holds for group (se)). When this condition is not satisfied, group (ij) uses only one mode.

If the difference in the free-flow travel costs in the two modes is very large, then all users will choose the same mode. Roads can be the selected mode when congestion costs produced by all users do not compensate for the high cost in public transport. This can occur when public transport has a very slow speed or very low frequency. It may also occur when connections are complex (for users se in this model). Since crowding is not considered here, public transport can be the only selected mode when roads are too long or of poor quality, inducing a large generalized cost compared with public transport. This situation can occur only in some particular real situations. High congestion can motivate car owners to use public transport. For example, in downtown Paris more than 90% of trips are made by public transport. The roads are very congested, but this is more often due to users traveling between Paris and the outer parts of the city. Those traveling inside Paris have to choose between very congested roads and frequent, fast, underground transit services. Despite the high crowding in public transit at peak hours, most users prefer public transport.

5.2 A numerical illustration

In this section, we provide a numerical illustration where we consider the linear formulation of Section 5.1. As a starting point, we mimic the observed modal choice described in Table 6 of Kilani et al. (2014) for Paris. They consider both a small and a large periphery for Paris, and differentiate between journeys from the center to the periphery, and between suburbs. It turns out that 22% to 41% of commuters going from the center to the periphery (or the reverse) use a private car. When considering periphery-to-periphery journeys, we find that 56% to 80% of commuters use a private car. Our theoretical model cannot be as subtle and will be calibrated such that about 25% of commuters from the center to the periphery and 80% of those commuting between peripheral areas use a private car.

To reach these orders of magnitude, we use the parameter values reproduced in Table 2. We wrote a program to perform the computations.[7] We use a right-angled isosceles triangle for the city structure. Each line connecting the city center to the suburban areas measures 5 km and the tangent measures 7.07 km. In terms of population, although there are more commuters going from the periphery to the center than from any one peripheral area to any other, there is usually a single city center and many suburban areas (more than two). As a result, the congestion between two suburbs is caused not only by commuters traveling between them, but also by those commuting between other suburbs but having to use the same peripheral routes. To overcome this problem and because we make no a priori assumptions, we decided to set the number of commuters on each link at 100 individuals. Free-flow travel speed is set at 15 km/h for the user of public transport between the city center and suburban areas and 20 km between them. Free-flow driving speed is set at 30 km/h when the origin or the destination is the city center, but at 40 km/h between suburban areas. Congestion parameter is set at 0.2. Crowding in transit is set at 0.05. The opportunity cost of time is 10 euros per hour and the time spent waiting or switching between two trains is set 50% higher, at 15 euros per hour. Train synchronization (α) multiplies the switching cost. It is set at 0.1, reflecting very good train coordination at the city center. The operating cost of public transit is

[7] This program is available at https://sites.google.com/site/quentinmaxdavid/research so that anyone can test the impact of a change in one of these parameters on any of the scenarios described in the paper.

Table 2: Parameters used for the numerical example

Parameters	$S \to C$	$C \to E$	$S \to E$
Population (N_{ij})	100	100	100
Distances (km)	5	5	7.07
Free-flow travel speed (km/h)			
on roads	30	30	40
in public transit	15	15	20
Congestion parameter (c^R_{ij})	0.2	0.2	0.2
Crowding parameter (c^T_{ij})	0.05	0.05	0.05
Other cost parameter			
Opportunity cost of time (€/h)	10		
Waiting cost (€/h)	15		
Switching cost (€/h)	15		
Train syncronization (α)	0.1		
Transit operating cost (ϕ)	22.5		

used as a scale parameter and takes the value 22.5.

The output of the numerical computations are given in Table 3. Each line corresponds to a case discussed in the paper. The base scenario is the unpriced equilibrium where train frequency is set endogenously to minimize operating costs. With these frequencies of 6 trains per hour, about 80% of the population use the transit for radial journeys, and this figure falls to 27% for commuters between suburban areas. The social transport costs in this city reaches 2,900. We start comparing this base scenario with different situations with the same transit structure (no SE transit line) and hold frequencies constant and exogenous. To reach the optimum (line 2), we need to increase the use of public transport for all user groups. This can be achieved by road tolls (as displayed in the Table) or with the equivalent subsidies for public transit users. We note that this increase is modest for radial commuters, while the users of public transit must be more than doubled among the SE commuters.

We then consider the privatization of one or both transport modes, holding the other mode free of charge. Although we only considered the privatization of roads in our model, we also consider the case where transit is managed by a private operator in this numerical exercise. When roads are managed by a private operator, tolls are higher than optimum, but the social cost of mobility is close to optimum. When public transit is privatized, its use decreases dramatically and social cost of transport soars. Finally, if both transport modes are privatized, then road tolls and transit fares become huge (due to the strategic complementarity of operator pricing). Besides being socially undesirable (social cost is above the unpriced equilibrium), it is likely to be unacceptable to the population.

Let us now consider the opening of a new SE transit line. For the first scenario, we consider that frequencies on original radial lines remain constant, and we assume the same frequency on the new line. Of course, ignoring the building cost of the new line lowers the social transport cost. Interestingly, the unpriced equilibrium is now closer to the optimum, and optimal road tolls are low. It is also noteworthy that although it mainly reduces transport costs for SE users, this situation is Pareto-improving, as transport costs decrease for all user groups. The use of the new transit line by the SE commuters reduces crowding in the original transit lines, increasing their use by the radial commuters. This result is valid whatever the frequency considered on the new transit line: irrespective of the frequency on this new line, these commuters could, at worst, continue to commute as they did without the SE line. When adjusting the frequencies to the number of users, the message remains the same from a social point of view: social cost decreases compared with the base scenario. In this example, it is of note that when frequencies are endogenous, opening the SE transit line is no longer Pareto-improving. The smaller number of transit users in the SC and CE lines (SE commuters no longer use these lines) leads to a decrease in frequencies, and so an increase in the cost of using transit for the SC and CE transit users. The gap between the unpriced equilibrium and the optimum is also wider because the endogenous choice of frequency acts as an amplifying factor.

Table 3: Results for the numerical example, symmetric case

	Transit users		Transit fares		Road tolls		User cost		Frequency		Total social
	$n^T_{sc} = n^T_{ce}$	n^T_{se}	$p_{sc} = p_{ce}$	p_{se}	$\tau_{sc} = \tau_{ce}$	τ_{se}	$C^k_{sc} = C^k_{ce}$	C^k_{se}	$f_{sc} = f_{ce}$	f_{se}	transport cost
Base model (with endogenous frequencies)											
Unpriced equilibrium	80.9%	26.8%	–	–	–	–	5.48	15.34	6*	–	2,900
Exogenous frequencies											
Optimum	86.8%	56.0%	–	–	1.46	6.42	5.77	15.92	6	–	2,695
Semi-public regimes - exogenous frequencies											
Private roads	90.5%	63.4%	–	–	2.29	8.09	5.87	16.11	6	–	2,714
Private public transit	40.5%	13.4%	8.54	3.58	–	–	13.57	18.02	6	–	4,047
Private regime											
Duopoly	60.3%	42.3%	12.92	10.17	9.6	12.25	18.36	25.42	6	–	3,040
The opening of a SE transit line - exogenous frequencies											
Unpriced equilibrium	82.0%	86.6%	–	–	–	–	5.27	3.38	6	6	1,796
Optimum	89.0%	91.3%	–	–	1.46	0.98	5.33	3.42	6	6	1,771
The opening of a SE transit line with endogenous frequencies											
Unpriced equilibrium	80.5%	85.5%	–	–	–	–	5.56	3.64	5.18*	5.34*	1,826
Optimum	88.3%	90.7%	–	–	1.52	1.03	5.53	3.61	5.43*	5.5*	1,788

* Endogenous frequencies (chosen to minimize the sum of operating cost and wait time cost)

Finally, we have not discussed the building costs of the SE transit infrastructure in this paper. We know that these costs are very high (see Bono et al., 2019, for a discussion of these costs). In our numerical example, the difference between the social costs with and without such infrastructure can be interpreted as the opportunity cost of building the line. Nevertheless, whatever the building costs, such a line seems socially desirable, though not for all commuters. If commuters could vote for or against the building of such a line, most should oppose it, as both SC and CE commuters would be worse off.

6 Conclusion

We have developed a model of urban transport in a polycentric city explicitly including commuting between suburban areas. This framework extends earlier literature on mode and route choice (Parry and Small, 2009; de Palma et al., 2007) through the adoption of a more complex and realistic network. We show that the unpriced equilibrium is in general not optimal, and cars are overused, particularly in the outskirts. The optimum can be decentralized through road pricing, or under a semi-public administration where roads are managed by a private operator and rail by a public, welfare-maximizing operator. When the two operators compete, road tolls and public transport fares are strategic complements, and monetary transport costs are significantly high.

Cities are expanding, and both reverse commuting and commuting between suburban areas are growing. In this context, many cities whose original public transit infrastructures were radial are considering investment in circular rail lines directly connecting the suburbs. The metropolitan area of Paris is a case in point, with the "Grand Paris Express" project whose estimated cost is 35 billion euros over the next decade. We show that such network expansion is socially desirable, but some users may become worse off unless the current network is overcrowded. In practice, it seems that the Parisian underground network is overused and the Grand Paris Express is likely to improve welfare for all citizens by reducing the crowding externality without affecting frequencies. One of the objectives of the model we have developed is to provide a tool to evaluate and compare several transport reforms that can provide some insights on policy choices.

The numerical example developed in the last section illustrates and quantifies our main findings. The worst scenario is obtained under the semi-public regime where roads are unpriced while public transport is administered by a private, profit-maximizing operator. The privatization of roads is quite efficient because it achieves a high use of public transit. The fully private regime reaches a social cost in-between the two semi-public regimes, but it is likely to be unacceptable to the population because it implies very high tolls and fares for commuters (we note that these tolls and fares do not enter the social cost function as they are transfers).

The impact of a network expansion that consists in opening a direct transit line between the suburbs largely depends on how service frequencies adjust. When frequencies are exogenous, providing a new rail line reduces the user costs for all commuters. This is the consequence of the lower level of crowding in the radial lines with the same service quality. With endogenous frequencies, the opening of the SE transit line reduces the number of users in the radial lines, reducing the frequencies and increasing the user cost of radial commuters accordingly.

In practice, several big cities have transit systems that are operating at their capacity limits. Service frequencies are then set at the maximum possible level and would not decrease if demand decreased slightly. In our model, this situation corresponds to the exogenous frequency scenario. In that case, building a new transit line would increase welfare and is likely to be Pareto-improving. Nevertheless, when crowding in the radial lines is moderate, commuters might vote against it if asked. Although the line increases total welfare, the effect is driven by commuters at the outskirts of the city, and the effect is negative (though relatively small) for commuters located downtown.

Finally, we emphasize that we have made some simplifying assumptions to keep the model analytically tractable. The case of inelastic demand and fixed origin-destination matrix are particularly important and can be relaxed for future extensions. It is also important to note that we did not incorporate vehicle emissions in the model. This important variable was disregarded to keep the model as simple as possible and because in many cases environmental constraints are not expected to change the signs of the impacts we have obtained. Emissions differ from congestion and crowding externalities in that they have broader impacts. Traffic on road SE, for example, will generate negative externalities not only for the group involved, but also for the city's entire population. It is thus obvious that if the model takes emissions into account, the benefit from line SE will be higher. An extension in this direction is straightforward and will be useful to quantify the impact of the new line. We leave this task for future empirical research.

Acknowledgements

This research received financial support from the "Société du Grand Paris" and from CNRS under project PEPS "Produire un modèle permettant d'évaluer la variabilité des risques de dépasser les seuils critiques". The usual disclaimer applies.

References

Aguilera, A. (2005), 'Growth in commuting distances in French polycentric metropolitan areas: Paris, lyon and marseille', *Urban studies* **42**(9), 1537–1547.

Aguilera, A., Wenglenski, S. and Proulhac, L. (2009), 'Employment suburbanisation, reverse commuting and travel behaviour by residents of the central city in the Paris metropolitan area', *Transportation Research Part A: Policy and Practice* **43**(7), 685–691.

Beckman, M. and McGuire, C. B. (1956), *Studies in the economics of transportation*, Yale University Press.

Brueckner, J. (2011), *Lectures on Urban Economics*, The MIT Press.

David, Q. and Foucart, R. (2014), 'Modal choice and optimal congestion', *Regional Science and Urban Economics* **48**, 12 – 20.

de Palma, A., Kilani, M. and Lindsey, R. (2007), 'Maintenance, service quality and congestion pricing with competing roads', *Transportation Research Part B: Methodological* **41**(5), 573–591.

de Palma, A., Kilani, M. and Lindsey, R. (2008), 'The merits of separating cars and trucks', *Journal of Urban Economics* **64**(2), 340–361.

de Palma, A., Kilani, M. and Proost, S. (2015), 'Discomfort in mass transit and its implication for scheduling and pricing', *Transportation Research Part B: Methodological* **71**, 1–18.

Giuliano, G. and Small, K. A. (1993), 'Is the journey to work explained by urban structure?', *Urban studies* **30**(9), 1485–1500.

Hamilton, B. W. (1982), 'Wasteful commuting', *Journal of Political Economy* **90**(5), 1035–51.

Kilani, M., de Palma, A. and Proost, S. (2017), 'Are users better-off with new transit lines?', *Transportation Research Part A: Policy and Practice* **103**, 95–105.

Kilani, M., Proost, S. and van der Loo, S. (2014), 'Road pricing and public transport pricing reform in Paris: complements or substitutes?', *Economics of Transportation* **3**(2), 175–187.

Knight, F. H. (1924), 'Some fallacies in the interpretation of social cost', *The Quarterly Journal of Economics* pp. 582–606.

Kraus, M. (2012), 'Road pricing with optimal mass transit', *Journal of Urban Economics* **72**(2), 81–86.

Lindsey, R. (2012), 'Road pricing and investment', *Economics of transportation* **1**(1), 49–63.

Mohring, H. (1972), 'Optimization and scale economies in urban bus transportation', *American Economic Review* **62**(4), 591–604.

Parry, I. W. and Small, K. A. (2009), 'Should urban transit subsidies be reduced?', *The American Economic Review* **99**(3), 700–724.

Small, K. A. and Song, S. (1992), '"Wasteful" commuting: A resolution', *Journal of Political Economy* **100**(4), 888–98.

Smith, M. (1979), 'The existence, uniqueness and stability of traffic equilibria', *Transportation Research Part B: Methodological* **13**(4), 295–304.

Tabushi, T. (1993), 'Bottleneck congestion and modal split', *Journal of Urban Economics* **34**, 414–431.

Verhoef, E. and Small, K. (2004), 'Product differentiation on roads', *Journal of Transport Economics and Policy* **38**(1), 127–156.

White, M. J. (1988), 'Urban commuting journeys are not "wasteful"', *Journal of Political Economy* **96**(5), 1097–110.

A Proofs

A.1 Proof of proposition 1

We state the equilibrium problem as a constrained minimization one and then characterize its solutions. Following Smith (1979), the equilibrium problem can be cast as a minimization of the following objective function[8]

$$\sum_{ij=sc,ce,se} \int_0^{u_{ij}^R} c_{ij}^R(z)dz + \sum_{ij=sc,ce} \int_0^{u_{ij}^T} c_{ij}^T(z, f_{ij})dz + \int_0^{n_{se}^T} \left(c_s + (\alpha - 1)\frac{c_w}{f_{ce}} \right) dz, \quad (18)$$

[8]Detailed expression of this problem is given in the Appendix A.1.

under the constraints $0 \leq n_{sc}^T \leq N_{sc}$, $0 \leq n_{ce}^T \leq N_{ce}$ and $0 \leq n_{se}^T \leq N_{se}$. The existence of a solution is seen by observing that the objective function is continuous and is it defined on a compact set (we apply the Weierstrass theorem). The first-order conditions for an interior solution is a set of three equalities. The advantage of using the minimization problem is that it handles corner solutions as well. Twice-differentiation of the objective function given by Eq.(18) yields the Hessian matrix

$$H = \begin{pmatrix} \frac{\partial c_{sc}^R}{\partial n_{sc}^T} + \frac{\partial c_{sc}^T}{\partial n_{sc}^T} & 0 & \frac{\partial c_{sc}^T}{\partial n_{sc}^T} \\ 0 & \frac{\partial c_{ce}^R}{\partial n_{ce}^T} + \frac{\partial c_{ce}^T}{\partial n_{ce}^T} & \frac{\partial c_{ce}^T}{\partial n_{ce}^T} \\ \frac{\partial c_{sc}^T}{\partial n_{sc}^T} & \frac{\partial c_{ce}^T}{\partial n_{ce}^T} & \frac{\partial c_{ce}^R}{\partial n_{ce}^T} + \frac{\partial c_{sc}^T}{\partial n_{sc}^T} + \frac{\partial c_{ce}^T}{\partial n_{ce}^T} \end{pmatrix}.$$

It is easy to see that the first two minors of H are positive so long as the first-order derivatives of the cost functions are positive. The computation of the determinant (the third minor) is easy. We check that it is also positive. Hence the matrix H is definitely positive, and it follows that the minimization problem is convex. To state the stability, consider an initial equilibrium. For a given user group all routes used have the same generalized costs. If one user decides to change their decision unilaterally, the generalized cost on the new road can only be higher.

A.2 Proof of proposition 2

As in the first proof, the existence of a solution is seen by observing that the objective function is continuous and is it defined on a compact set. For the case of a single interior equilibrium, we have to prove that the objective function in Eq. (4) is convex. The 3×3-Hessian matrix is too large and we provide its components one by one. The first two elements in the diagonal of this matrix have

$$2\frac{\partial c_{ij}^R}{\partial n_{ij}^R} + n_{ij}^R \frac{\partial^2 c_{ij}^R}{\partial (n_{ij}^R)^2} + 2(C_{ij}^T)_{[1,0]} + (n_{ij}^T + n_{se}^T)(C_{ij}^T)_{[2,0]} \qquad \text{for} \quad ij = sc, ce,$$

and where $[n, m]$ subscript denotes the partial derivatives of orders n and m, respectively, for the first and second derivatives. The third and last element in the diagonal is

$$2\frac{\partial c_{ce}^R}{\partial n_{se}^R} + n_{se}^R \frac{\partial c_{ij}^R}{\partial (n_{se})^2} + 2(c_{sc}^T)_{[1,0]} + 2(c_{ce}^T)_{[1,0]} + (n_{sc}^T + n_{se}^T)(C_{sc}^T)_{[2,0]} + (n_{ce}^T + n_{se}^T)(C_{ce}^T)_{[2,0]}.$$

Elements $(2,1)$ and $(1,2)$ are zero. The other four elements are equal to

$$2(c_{ij}^T)_{[1,0]} + (n_{ij}^T + n_{se}^T)(c_{ij}^T)_{[2,0]} \quad \text{for} \quad ij = sc, ce,$$

where $ij = sc$ for elements $(1,3)$ and $(3,1)$, and $ij = ce$ for elements $(2,3)$ and $(3,2)$.

When the second-order partial derivatives are positive, the condition in Proposition 2, it is clear that the first principal minor is positive. The second principal minor is a determinant of a diagonal matrix and is also positive. The computation of the third minor, the determinant of matrix H, involves tedious computations that also show that it is positive. We put the details of this computation in the accompanying MATHEMATICA notebook.

This proves that the total cost, the objective function, is convex. Together with the bound constraints, this yields a convex program with a single solution where the objective function reaches its minimum. We note that the conditions on the second-order partial derivatives are sufficient but not necessary. Even with negative second-order derivatives the Hessian matrix remains positive definite as long as the first-order derivatives dominate.

A.3 Proof of corollary 1

Consider an interior solution $(n_{sc}^*, n_{ce}^*, n_{se}^*)$. Let us ignore the second argument, the frequencies, in function $\tilde{c}_{ij}^T$ and write them as $\tilde{c}_{ij}^T(n_{sc}^T + n_{se}^T)$. The set of the three equalities can be written as

$$\begin{aligned}
\tau_{sc} &= \underbrace{\tilde{c}_{sc}^T(n_{sc}^T + n_{se}^T) - c_{sc}^R(n_{sc}^R)}_{h_{sc}(n_{sc}, n_{se})} \\
\tau_{ce} &= \underbrace{\tilde{c}_{ce}^T(n_{ce}^T + n_{se}^T) - c_{ce}^R(n_{ce}^R)}_{h_{ce}(n_{ce}, n_{se})} \\
\tau_{se} &= \underbrace{\tilde{c}_{ce}^T(n_{ce}^T + n_{se}^T) + \tilde{c}_{se}^T(n_{sc}^T + n_{se}^T) - c_{se}^R(n_{sc}^R)}_{h_{se}(n_{sc}, n_{ce}, n_{se})}.
\end{aligned}$$

Functions h_{ij}^T compute the toll levels on each road as a function of the number of users in transit and on roads. Note that by the construction of the cost functions, cf. Eqs. (1)-(2), the right hand members above are monotonously increasing in n_{ij}^T, so functions h_{ij} are uniquely defined.

A.4 Proof of Proposition 3

The profit made by the private operator is

$$\tau_{sc}n_{sc}^R + \tau_{ce}n_{ce}^R + \tau_{se}n_{se}^R.$$

Substituting for the tolls from (3), and using $n_{ij}^R = N_{ij} - n_{ij}^T$, in the profit made by the private operator, the latter is written as $\left[c_{sc}^T - c_{sc}^R\right](N_{sc} - n_{sc}^T) + \left[c_{ce}^T - c_{ce}^R\right](N_{ce} - n_{ce}^T) + \left[c_{sc}^T + c_{ce}^T - c_{se}^R\right](N_{se} - n_{se}^T)$, when public transport fares are zero. This expression of the profit can be rearranged to obtain

$$\begin{aligned} &N_{sc}c_{sc}^T - \left(n_{sc}^T c_{sc}^T + n_{sc}^R c_{sc}^R\right) + N_{ce}c_{ce}^T - \left(n_{ce}^T c_{ce}^T + n_{ce}^R c_{ce}^R\right) + \\ &N_{se}(c_{sc}^T + c_{ce}^T) - \left(n_{se}^T(c_{sc}^T + c_{ce}^T) + n_{se}^R c_{se}^R\right) \\ &\qquad = (N_{sc} + N_{se})c_{sc}^T + (N_{se} + N_{ce})c_{se}^T - TC, \end{aligned} \tag{19}$$

where TC is the total social cost given in (4). Now consider the first-order condition with respect to n_{sc}^T in (19). It yields

$$(N_{sc} + N_{se})\frac{\partial c_{sc}^T}{\partial n_{sc}^T} = \frac{\partial TC}{\partial n_{sc}^T}. \tag{20}$$

The left-hand side member in (20) is clearly positive, and so the right-hand member is. It is straightforward to check that the first-order conditions with respect to n_{ce}^T and n_{se}^T yields a similar conclusion. Hence the private operator sets the tolls so that the derivative of the social cost is positive. Since the cost function is convex, this is possible only when public transport is overused. The tolls set by the private operator are then higher than the optimum tolls.

A.5 Proof of Proposition 4

Consider the profit function of the private operator. We make the same substitutions for the toll but with non-zero fares. We obtain a first-order condition

$$(N_{sc} + N_{se})\frac{\partial c_{sc}^T}{\partial n_{sc}^T} - p_{sc} = \frac{\partial TC}{\partial n_{sc}^T},$$

which is slightly different from (20). A similar condition is obtained for n^T_{ce}. To prove that $\beta = 1$, the first-order conditions yield

$$\beta = \frac{\sum_{ij=sc,ce}(N_{ij} + N_{se})\frac{\partial c^T_{ij}}{\partial n^T_{se}}}{p_{sc} + p_{ce}}.$$

If we develop the term in the summation and substitute for fares, we obtain

$$\beta = \frac{(N_{sc} + N_{se})\frac{\partial c^T_{sc}}{\partial n^T_{se}} + (N_{ce} + N_{se})\frac{\partial c^T_{ce}}{\partial n^T_{se}}}{(N_{sc} + N_{se})\frac{\partial c^T_{sc}}{\partial n^T_{sc}} + (N_{ce} + N_{se})\frac{\partial c^T_{ce}}{\partial n^T_{ce}}}$$

which is equal to one when the impact of the users on the travel cost does not depend on their group (assuming we have a homogeneous population).

Hence when the public operator sets fares given by (9), the first-order conditions for the optimum are met.

A.6 Proof of Corollary 2

It follows by setting partial derivatives to zero in (7) and if there is no crowding in public transport ($\partial c^T_{ij}/\partial n^T_{ij} = 0$).

B Discussion of proposition 2

This result is better illustrated in a simplified network with two links between a given OD pair. Let the free-flow travel cost on the longer link be fixed at f_0,[9] the free-flow travel cost on the shorter link be $f + a\,n$ where f and a are given positive numbers and n the number of the users of the shorter link. An interior solution requires $f < f_0 < f + a\,N$, a condition that we assume. At equilibrium, the user costs are equal on the links, and simple algebra shows that $n^e = (f_0 - f)/a$. The total number of the users who travel from the origin to the destination is N (inelastic demand), and we denote the users of the shorter link n, so $N-n$ is the number of the longer link. At equilibrium, it is clear that the total cost is $n(f+a\,n)+f_0(N-n)$ which reaches its minimum at $n^O = (f_0-f)/2a$. A toll $\tau^O = (f_0 - f)/2$ imposed on the users of the shorter route will decentralize the optimum. At the same time, if that route is managed by a private operator who maximizes its profit τn under the constraint that user costs are equal, then it will maximize its profit by imposing the optimum toll, i.e. $\tau^{sp} = \tau^O = (f_0 - f)/2$.

[9]The notation is specific to this example.

Hence we can state that the road toll imposed by a private operator is the sum of two parts. The first one is the optimum toll that would have been imposed by a public manager, and the second part corresponds mainly to crowding in transit. The second part is zero if there is no crowding and increases as the impact of crowding increases. Analytically, we write the road toll for link (ij) as

$$\tau_{ij}^{sp} = \tau_{ij}^{O} + (\text{crowding-induced part}) .$$

The existence of crowding in public transit reduces the magnitude of the demand elasticity of the users of the road. The private operator gains from this rigidity in user choice by increasing the road toll.

In the linear case, and focusing on an interior solution, we can obtain the tolls imposed by the private operator. Since the derivative with respect to the number of users is constant, we can easily express these tolls as a function of the first-best tolls. Doing so, we obtain

$$\begin{aligned}
\tau_{sc}^{sp} &= \tau_{sc}^{O} + \frac{1}{2}\left[p_{sc} + a_{sc}^{T}\frac{N_{sc}+N_{se}}{f_{sc}}\right] \\
\tau_{ce}^{sp} &= \tau_{ce}^{O} + \frac{1}{2}\left[p_{ce} + a_{ce}^{T}\frac{N_{ce}+N_{se}}{f_{ce}}\right] \\
\tau_{se}^{sp} &= \tau_{sc}^{sp} + \tau_{ce}^{sp} + \frac{1}{2}\left[c_s + (\alpha-1)\frac{c_w}{2} + (\beta-1)p_{ce} + F_{ce}^{R} + F_{sc}^{R} - F_{se}^{R}\right].
\end{aligned}$$

Hence for the same fares, $\tau_{ij}^{sp} > \tau_{ij}^{O}$ for all (ij). Roads are underused and public transport is overused, inducing high crowding levels. In response, the public transit operator will reduce total travel cost by imposing positive fares. Compared with the first-best optimum considered above, and where transit fares were fixed at zero, we find that second-best fares when public transport is operated by a private operator are higher than first-best fares. The case where $N_{sc} = N_{ce}$ provides a particularly simple expression, and the public operator can reach the first-best mode choice by imposing fares given by

$$\begin{aligned}
p_{ce}^{O} &= \frac{N_{sc}+N_{ce}}{\frac{1-\beta}{a_{sc}^{R}} + \frac{f_{ce}}{a_{ce}^{T}}} \\
p_{sc}^{O} &= \frac{f_{ce}a_{sc}^{T}}{f_{sc}a_{ce}^{T}}\, p_{ce}^{O}
\end{aligned}$$

We note that this result is in line with the conclusions of Kilani et al. (2014); Kraus (2012) for the case of the model studied here.