Paper proposed to ETRR, February 2019.

Microeconomics of a taxi service in a ring-shaped city

Fabien Leurent,[1]
Université Paris Est, Laboratoire Ville Mobilité Transport, École des Ponts ParisTech

Abstract

To a client, taxi quality of service involves not only the riding time and comfort, but also the access time between the instants of booking (or readying oneself) and pick-up. In turn, the access time depends on fleet size and the macroscopic patterns of service usage: demand volume and its spread in space, average ride time, transaction times.

In this article, we investigate the formation of the access time and derive its economic consequences for a taxi service in an idealized city with ring shape and spatial homogeneity, hence circular symmetry.

At the operational level, under given supply and demand conditions the access time stems from the number of busy vehicles, which obeys to a second-degree characteristic equation. This enables us to model fleet size as a function of target demand volume and access time. Taking then a broader perspective, demand is elastic to supply conditions including access time, ride time, transaction time and tariff fare. We model short-term traffic equilibrium and demonstrate the existence and uniqueness of an equilibrium state.

Next, at the tactical level the service supplier sets up the fleet size and the tariff fare in order to satisfy an economic objective. We model medium-term supply-demand equilibrium under three regulation patterns of, respectively, (i) service monopoly and the maximization of production profit, (ii) system optimum and the maximization of social surplus, (iii) second best system optimum subject to a budgetary constraint. In each pattern, both the tariff fare and the access time are linked by analytical formulas to exogenous conditions about the territory, the demand and the cost function of service provision.

Theoretical properties are obtained to compare the patterns under specific demand function with constant elasticity of volume to generalized cost: under constant elasticity of -2, the monopoly tariff and generalized cost are more than twice as large as their system optimum counterparts, and exact doubles of their second best optimum counterparts in the absence of fixed production costs.

At the strategic level, the model can be applied to assess decisions on vehicle technology (motor type, driving technology) and on service location by the service supplier, as well as the regulation policies by public authorities.

Keywords: Traffic Model; Stochastic Equilibrium; Availability Function; Supply-Demand Equilibrium; Monopoly Operation; Collective Optimum; Second-Best Optimum.

---

[1] Corresponding author: fabien.leurent@enpc.fr. École des Ponts ParisTech, Laboratoire Ville Mobilité Transport, 6-8 avenue Blaise Pascal, Champs sur Marne 77455 Marne la Vallée Cedex

1. Introduction

1.1 Background

Over two decades, a bunch of technological developments have considerably enhanced the attractiveness and efficiency of taxi services: GPS geolocation combined to geographic databases have made shortest paths readily available to taxis; platform technology has empowered the matching of vehicles and customer requests; the individual equipment both in smartphones and service mobile applications has eased transaction operations considerably, from trip planning and rendezvous arrangement to payment, passing by vehicle booking and on-screen tracking for reassurance; furthermore, on-board Wifi enables the client to re-use the ride time for another activity (Andreasson et al. 2016). The early adoption of these empowerments has enabled pioneering companies such as Uber and Lyft from the USA, Didi from China, Ola from India etc to grow strong and fast and take the market lead in many big cities (Boutueil et al. 2018).

In the academic literature, there are four main research streams targeted to taxi analysis. The first stream focuses on algorithms to enhance service efficiency, e.g. in customer-vehicle matching and the assignment of duties to vehicles (e.g. Dial (1995), Malucelli et al. (1999), Lioris (2010)).

The second stream is interested in spatial details and the network modeling of taxi services, so as to establish their quality performance at the trip level and derive the demand volume, possibly by comparison to alternative modes on the basis of their respective utilities. Most notable in this stream of Travel Demand Modeling is the lineage of contributions by Yang, Wong and co-workers at Hong-Kong Technical University over the years. Yang and Wong (1998) devised a network model of taxi operations, including occupied time to serve customers as well as vacant time employed to find the next customers: their model of taxi movements and demand trips is a combined model of traffic distribution and assignment, for which an efficient mathematical formulation and solution algorithm were provided by Wong et al. (2001), who also addressed road congestion and demand elasticity. Yang et al. (2005a) put forward a multi-period dynamic analysis with endogenous intensity of taxi services. Wong et al. (2008) extended the network model to multiple user classes and multiple travel modes. Yang et al. (2010) introduced a meeting function that relates the meeting rate of waiting customers and vacant taxis in a given sub-area to their respective local densities: using a Cobb-Douglas specification, they succeeded to embed it in the network model. The equilibrium properties were further studied by Yang and Yang (2011).

The third stream belongs to microeconomic theory: service supply and demand are modeled in a simplified way by means of fleet size and demand function, respectively, so as to focus on their relationship and investigate issues of service quality, modal share, fare policy, supplier competition and market regulation. Douglas (1972) established the basic theory by combining (i) a demand function with respect to price and wait time, (ii) a pricing rule linking the tariff fare to trip time, (iii) a production cost function proportional to taxi time occupied and vacant, (iv) a “delay distribution” i.e. a function relating the taxi unoccupied time from customer drop-off to next customer pick-up, to the density of vacant taxis and also the car speed. He emphasized the crucial role of wait time in the analysis of supply-demand equilibrium first under fixed fare then as a second best system optimum, before coming to scale effects and quality differentiation. De Vany (1975) compared a regulated monopoly targeted to social optimum and a competitive market: he showed that an increase in the regulated price has more effect on competitive supply than on a monopolistic firm. Manski and Wright (1976) provided a systems analysis of the taxi market and studied the allocation of taxi licenses: they showed that, under a wide range of variations, increasing their number would both reduce customer wait time and increase taxi occupancy rate. Arnott (1996) considered a homogeneous two-dimensional space in order to analyze Douglas model in a less aggregative perspective and provide guidance to mobility planners: in turn, he emphasized the crucial role of wait time in taxi economics. Yang et al. (2002) linked the network model to the economic analysis of market equilibrium under different regulation patterns: monopoly, competitive, 1st best and 2nd best social optimum. Yang et al. (2005b) studied congestion externalities due to the movements of both occupied and vacant taxi movements together with normal vehicular traffic. Flores-Guri (2003) applied the Douglas model to the New-York case: having defined a production function linking the demand volume to the service fare and the density of vacant vehicles, he observed a certain inelasticity of demand to the number of vacant vehicles. Indeed, the matching function is a key component to refine the modeling of wait time and better understand the linkage of demand to supply.

The fourth stream of taxi literature is devoted to behavioral studies of taxi customers and taxi drivers, on the basis of social surveys or taxi trip databases. Recent instances include a study of customer search strategies by taxi drivers by Wong et al. (2015), a Stated preference survey of electric vehicle adoption by taxi owner and drivers (Yang et al., 2018), among other works.

1.2 Objective

This paper has a twofold objective of, first, providing a physical function to theorize the mean access time between customer and taxi service and, second, deriving a series of economic properties concerning supply-demand equilibrium, supply management and its regulation.

The physical function for access time is based on a set of postulates: prominent among them is an idealized city shaped as a regular ring, along which spatial uniformity induces a property of circular symmetry. The idealization is clearly restrictive: however, it enables to model the positions of customers and vehicles and the distances between them. Upon it we build a stochastic model of taxi occupancy state (i.e. Busy vs. Vacant), in which the average number of busy vehicles satisfies a simple characteristic equation, from which stems the average access time.

Then, by availing ourselves of the analytical access time function, we investigate classical issues of economic theory in a specific way. We establish the increasingness and convexity of the access time function with respect to demand volume, as well as the existence and uniqueness of supply-demand equilibrium in the short run i.e. given fleet size N and tariff fare τ. Coming to supply management, we specify a simple function of production cost and model the supplier behavior either as a monopoly, or a regulated service purported to system optimum either first-best or second-best. In these developments, we lay the emphasis on the access time in order to reveal its influences and its relations to supply and demand characteristics.

1.3 Methodology and research contributions

As our supply-demand model involves a crude but effective description of space, together with abstract location of taxis as well as customers, it constitutes a bridge between the second and third research streams devoted to Travel demand modeling and Microeconomic theory, respectively. More precisely, the access time function is much analogous to a travel time function according to usage volume: it constitutes a piece of traffic science. The mathematical study of short-term supply-demand equilibrium is typical of the theory of traffic assignment to a transportation network (e.g. Beckmann et al. 1956).

The idealized shape of a ring city has been contemplated by some urbanists and urban planners: Maupu (2006) claimed that the “Hollow city” would be ideally suited to mass transit transportation as a single line could provide access to all places. Our stochastic model of a taxi service in such a spatial context is an application of queuing theory: the characteristic equation of the average number of busy vehicles is an approximation of Little’s law (cf. Kleinrock, 2975). We have also modeled an underlying Markov chain which is a bi-sided waiting queue: such queue has already been applied to taxi services in order to model one taxi rank (Conolly et al. (2002), based on Kashyap (1966)).

As for economic analysis, our contribution is to carry out the access time function throughout the well-known steps of economic theory: supply-demand interaction, supply cost function, supply behavior, monopoly, market regulation. While previous research emphasized the (N, Q) factors and the (N, τ) management levers, we shift to the (t, Q) and (t, τ) pairs of factors and management levers, respectively.

1.4 Paper structure

The rest of the article consists of eight sections. We begin by representing the territory as a set of places in which mobility demand is located (section 2). Then we model the transport supply, in terms of traffic infrastructure, transport service and operational processes (section 3). We can then establish traffic for the service as a process of interaction between the taxi operator and all customers, and characterize the mean state of the system in stationary regime (section 4).

We then study the traffic equilibrium between supply and demand, by combining the demand function with the characteristic supply functions: from this, we deduce the characteristic conditions of how the demand makes use of the supply (section 5). Next, we turn to supplier behavior and its regulation: starting from the monopoly problem (section 6), we then deal with system optimum either first best (section 7) or second best (section 8).

In addition, we make a short application case to illustrate system state variations depending on technological options and regulation patterns (section 9). Finally, we conclude with a summary combined with a discussion (section 10).

Notation

R city ring radius

O city center

Q number of individual trips requested during period H on a given day

$\lambda \equiv Q/\mathrm{H}$ time intensity of trip demand

ℓ in-vehicle level of comfort

v average car speed along the ring

L_{R} trip length

$t_{\mathrm{R}} \equiv L_{\mathrm{R}}/v$ trip run time

t_{T} transaction time on taxi side, $\tilde{t}_{\mathrm{T}}$ on customer side

$t_{\mathrm{A}} = t = t_{\mathrm{A}}^{\mathrm{V}}$ access time from taxi to customer

$\tilde{t}_{\mathrm{A}} \equiv t_{\mathrm{A}}^{\mathrm{V}} + t_{\mathrm{A}}^{\mathrm{W}}$ availability time to a customer includes access time and allocation wait time

N fleet size

h a particular moment within period H

$b \equiv \pi\mathrm{R}/v$ time parameter typical of urban territory

$k^{+} \geq 0$ number of vacant vehicles

$k^{-} \geq 0$ number of customers waiting for vehicle assignment

$k \equiv k^{+} - k^{-}$ state variable of taxi service Markov chain

$\mathrm{T}_{\mathrm{A}}^{\mathrm{V}}(Q, N)$ Access time function with respect to Q and N

$\hat{\mathrm{N}}(Q, t)$ fleet sizing function depending on Q and t

τ tariff fare per ride

D demand function with respect to τ, t, t_{R}, $\tilde{t}_{\mathrm{T}}$.

ε, ε_{τ}, ε_{t} Demand volume elasticity with respect to g, τ and t, respectively

α, β Value of time to service user, respectively for Ride and Access

g trip generalized cost to service user

C_{P} Production cost function with respect to N and Q ($\hat{\mathrm{C}}_{P}$ with respect to t and Q)

χ_{0} Fixed part of production cost and $\dot{\chi} \equiv \partial \mathrm{C}_{P}/\partial N$ derivative of cost with respect to fleet size

c_{u} variable supply cost per unit of vehicle time

<u>Derived parameters:</u>

$\mathrm{c}_{u}^{+} \equiv \mathrm{c}_{u} + \dot{\chi}/\mathrm{H}$,

$\eta \equiv \dot{\chi}/(\beta + \mathrm{c}_{u}^{+})$,

$y_{\mathrm{RT}} \equiv c_u^{+} t_{\mathrm{RT}} + \alpha \tilde{t}_{\mathrm{RT}}$

2. Representation of the territory

We represent the territory as a space that has been urbanized to form a certain shape (§ 2.1) and that contains a population of individuals (§ 2.2) whose mobility demand we model in a simplified way (§ 2.3).

2.1 Spatial configuration: a ring around a circle

About the spatial configuration, let us make the following postulates:

[C-1] that the city of Orbicity extends in space in the shape of a ring: we call the radius of the circle R and its geographical center O.

[C-2] Human activities (homes, jobs, amenities…) are located on a narrow strip all around the circle, hence over a circumference of $2\pi\mathrm{R}$.

Selecting one point M_0 on the circle as reference, we identify a particular position M by the angle $\theta \in]-\pi, \pi]$ between the vectors $\overrightarrow{\mathrm{OM}_0}$ and $\overrightarrow{\mathrm{OM}}$.

2.2 Location of individuals and spatial homogeneity

We denote the size of the population established in the territory by P, i.e. the number of individuals there. To establish circular symmetry, we also postulate:

[L-1] a homogeneous settlement pattern: each segment of the circle situated on the angular interval $[\theta, \theta + \delta\theta]$ carries a population $\psi.\delta\theta$, ratio $\psi \equiv \mathrm{P}/(2\pi\mathrm{R})$ being a linear density of population along the circle.

[L-2] More broadly, that activities are homogeneously located on the circle: homes, jobs, urban amenities (services, shops, leisure). At a given moment, therefore, the dynamic distribution of individuals between the places remains uniform.

2.3 Mobility demand

We study the mobility of the individuals in the territory by time period of duration H (e.g. an hour or a day or the period for which the transport service is open). During this period, we call Q the number of individual trips requested and completed by individuals. The postulates on mobility demand are:

[M-1] that the usage volume is distributed stochastically with a uniform structure in time: let us denote the time intensity of the overall demand by $\lambda \equiv Q/\mathrm{H}$;

[M-2] spatial uniformity with regard to the generation of trips by places;

[M-3] that individuals are homogenous as for trip generation: at any time, each individual has a certain probability of generating a trip from the position at which they are currently active. The time intensity of this probability is λ / P;

[M-4] each trip generated from a point M is characterized by the angle θ between M and the destination point. We postulate that θ follows a statistical distribution on $[-\pi, \pi]$ that is independent of the point of origin, with a cumulative distribution function denoted by F.

The postulate [M-3] of uniform distribution of starting points, combined with the postulate [M-4] on the spatial distribution of trips from each starting point, imply that the destination places have equal probabilities for all trips using the service.

3. The transport service

We consider here only one mode of transport: a taxi service running on the roadway infrastructure. The assumptions used to represent the service therefore concern firstly the infrastructure (§ 3.1), secondly the service as it appears to a customer (§ 3.2) and thirdly its operational procedures (§ 3.3).

3.1 Transport infrastructure and speed of travel

We postulate that a two-way road serves all the places in the territory, and therefore runs through all the points on the circle. The city's highway infrastructure consists exclusively of this road, which espouses in its shape the layout of activities. In this respect, Orbicity is very similar to the utopian "Hollow City" model proposed by Jean-Louis Maupu (2006).

For the problem tackled here, we make the following assumptions:

[I-1] in each direction of travel, vehicles move at a mean speed denoted by v;

[I-2] a taxi can stop anywhere to park, or to pick up or drop off a passenger, without disturbing the flow;

[I-3] a taxi can change direction, i.e. turn round, at any point;

[I-4] delays associated with manoeuvres (leaving or entering the traffic flow, stops, U-turns) are negligible, both for the vehicle concerned and for other vehicles (travelling in one direction or the other).

3.2 The taxi service and its quality of service

We assume that rides are provided by a single taxi operator, under the following conditions:

[S-1] one taxi ride per individual customer trip;

[S-2] a uniform vehicle type: the level of comfort is denoted by ℓ;

[S-3] a trip time, denoted by t_R, proportional to the distance covered L_R so that $t_\mathrm{R} \equiv L_\mathrm{R}/v$;

[S-4] the "customer path" in terms of the sequence Plan-Book-Ticket is managed by a web app: for each ride, the transaction time for the customer is $\tilde{t}_\mathrm{T}$, and for the taxi concerned t_T. Each of these durations includes the time taken for passenger entry and exit, door opening and closing, luggage handling, and, if necessary, for the taxi to change its direction of travel;

[S-5] an access time $\tilde{t}_\mathrm{A}$ (A for "Availability") between the moment the customer books and the time the taxi arrives to pick them up. This time encompasses the time the taxi takes to reach the customer, denoted by t_A without tilde or t_A^V (V for "Vehicle"), with the possible addition of the time for the customer to be allocated a vehicle, denoted by t_A^W (W for "Waiting").

To summarize, for the customer, quality of service is characterized in terms of comfort level ℓ and times t_R, $\tilde{t}_\mathrm{T}$ and $\tilde{t}_\mathrm{A}$. The trip time t_R is the same for the vehicle and the customer, while the access times t_A and $\tilde{t}_\mathrm{A}$, as well as the transaction times t_T and $\tilde{t}_\mathrm{T}$, have a common base but also a part that is specific to the customer or to the vehicle respectively.

3.3 Operational management of the service

Let N be the size of the taxi operator's fleet, i.e. the number of vehicles. We ignore the proportion of vehicles inoperative for maintenance or repair.

We postulate that, at any moment h:

[O-1] busy vehicles are vehicles that are allocated to a customer, whether the latter is already on board (cf. t_{R}), in the entry or exit process (cf. t_{T}), or in reserved access (cf. $t_{\mathrm{A}}^{\mathrm{V}}$): the number of such vehicles is denoted by n; the number of vacant vehicles is therefore $N-n$;

[O-2] when booking a vehicle, a new customer is allocated the vacant vehicle closest to him or her on the circle, in order to minimize t_{A};. When there are no vacant vehicles, the next taxi to become vacant will be assigned to the waiting customer with longest wait time, i.e. on a First-Come First-Serve basis.

[O-3] each customer's request is handled by the best route possible, i.e. the shortest in time according to the angle $\theta \in]-\pi, \pi]$ specific to the journey. So if $\theta > 0$, the trip takes place in the direct trigonometric direction, or in the opposite direction if $\theta < 0$. The trip time is therefore $t_{\mathrm{R}} = R|\theta|/v$.

On average,

$$\bar{t}_{\mathrm{R}} = \frac{\mathrm{R}}{v}\bar{f} \quad \text{where} \quad \bar{f} \equiv \int_{-\pi}^{\pi} |\theta| \, \mathrm{dF}(\theta). \tag{3.1}$$

This technical process is managed by a centralized control system that coordinates the vehicles in order to provide a unified service. This control system also manages the commercial relationship with customers.

The availability time, $\tilde{t}_{\mathrm{A}}$, includes a run time $t_{\mathrm{A}}^{\mathrm{V}}$ that depends on the angle between the customer's starting point and the starting point of the allocated vehicle. If there are no empty vehicles at the time of the request, then a customer waiting time is added, denoted by $t_{\mathrm{A}}^{\mathrm{W}}$.

4. Service operation with traffic in stochastic equilibrium

In this section, we model the dynamic operation of the service: vehicle occupancy, access to customers, and idle phases for each taxi between two paying rides. This dynamic operation is fundamentally random: individual requests are made without coordination and relate to a variety of trips. A first issue pertains to the distribution of the time taken for a taxi to reach a customer: assigning the nearest vacant vehicle gives rise to a simple model based on the number of vacant vehicles (§ 4.1). At any time, however, this number is itself a random variable: we model it with a Markov model, which is described in detail in Appendix B. In particular, we establish the statistical distribution of the dynamic state of the system in a stationary regime. In practice, the mean value can be approximated by a simple formula – a second-degree equation (§ 4.2). From this, we deduce the mean access time per taxi, which is roughly equivalent to the mean access time per customer if the fleet is sufficiently large (§ 4.3).

4.1 Vehicle access time as a random variable

At a given moment h, the number of vacant vehicles k^+ is the difference between the fleet size N and a number of busy vehicles n:

$$k^+ \equiv N - n. \tag{4.1}$$

For a customer requiring the service at that moment, if $k^+ = 0$ then the customer must wait for a vehicle to become available; otherwise i.e. if $k^+ > 0$ the service assigns them the nearest vacant vehicle. We identify the positions of the vacant vehicles relative to the starting position of the new customer by the angle $\theta \in]-\pi,\pi]$. When the system is in a stationary regime, all places are equivalent both for the customer's position and for the location of vacant vehicles, which is in principle at their last customer's destination point, given that this point is uniformly distributed from the point of origin, which is itself uniformly distributed. So $\forall i \in \{1,2..,k^+\}$, θ_i is uniformly distributed on $]-\pi,\pi]$.

The distance L_i between the vacant vehicle i and the customer is $L_i = \mathrm{R}.|\theta_i|$ where $|\theta_i|$ is uniformly distributed on $]0,\pi]$. Its cumulative distribution function is therefore

$$\mathrm{F}^{(i)}(x) \equiv \Pr\{L_i \le x\} = \mathrm{F}_{\mathrm{AR}}(x), \text{ where } \mathrm{F}_{\mathrm{AR}}(x) = \frac{\min\{x, \pi\mathrm{R}\}}{\pi\mathrm{R}} 1_{\{x \ge 0\}}. \tag{4.2}$$

The vacant vehicle closest to the customer is located at distance $L_{\min} \equiv \min\{L_i : i \in \{1,2,...k^+\}\}$.

According to the hypotheses regarding points of origin and trips, the destination points are equiprobable, so the vacant vehicles have positions that are independent and identically distributed according to $\mathrm{F}_{\mathrm{AR}}(x)$. The minimum distance therefore has a cumulative distribution function $\mathrm{F}_{\min}^{(k^+)}(x)$ with the property that

$$\begin{aligned} 1 - \mathrm{F}_{\min}^{(k^+)}(x) &= \Pr\{\min L_i > x\} \\ &= \bigcap\nolimits_{i \in \{1,2,...k^+\}} \Pr\{L_i > x\} \\ &= \prod\nolimits_{i \in \{1,2,...k^+\}} (1 - \mathrm{F}_{\mathrm{AR}}(x)) \end{aligned}$$

So
$$\mathrm{F}_{\min}^{(k^+)}(x) = 1 - [1 - \mathrm{F}_{\mathrm{AR}}(x)]^{k^+}. \tag{4.3}$$

The mean is easily calculated:

$$\begin{aligned} \mathrm{E}[L_{\min} \mid k^+] &= \int_0^{\pi\mathrm{R}} x \,\mathrm{d}\mathrm{F}_{\min}^{(k^+)}(x) = [x\,\mathrm{F}_{\min}^{(k^+)}(x)]_0^{\pi\mathrm{R}} - \int_0^{\pi\mathrm{R}} \mathrm{F}_{\min}^{(k^+)}(x)\,\mathrm{d}x \\ &= \pi\mathrm{R} - \pi\mathrm{R}\int_0^1 [1-(1-u)^{k^+}]\,\mathrm{d}u = \pi\mathrm{R}\int_0^1 (1-u)^{k^+}\,\mathrm{d}u \\ &= \frac{\pi\mathrm{R}}{k^+ + 1} \end{aligned}$$

So
$$\bar{L}_{\min}^{(k^+)} = \pi\mathrm{R}/(k^+ + 1). \tag{4.4}$$

And
$$\bar{t}_{\mathrm{A}}^{\mathrm{V}(k^+)} = \frac{b}{k^+ + 1} \text{ where } b \equiv \frac{\pi\mathrm{R}}{v}. \tag{4.5}$$

4.2 The characteristic equation of the mean number of busy vehicles

When the system is in dynamic operation, the number of vacant vehicles varies according to the number of requests: the stochasticity of the origin and destination points spreads to the number of vacant vehicles. In Appendix B, we model the state of the system as a relatively simple Markov chain, with a discrete state variable which, in the positive range, is the number of vacant taxis and, in the negative range, the number of customers waiting for a taxi to be assigned to them (differing only in the sign).

In stationary regime, the mean number of busy taxis is equal to the time frequency of rides multiplied by the mean time per ride (Little's law in queuing theory):

$$\bar{n} = \lambda(t_T + \bar{t}_R + \bar{t}_A^V). \tag{4.6}$$

We postulate that the mean access time can be closely approximated by the function $\bar{t}_A^{V(k^+)}$ applied to the value for the mean number of vacant vehicles, $\hat{k} = N - \bar{n}$: in other words,

$$\bar{t}_A^V \approx \frac{b}{N+1-\bar{n}}. \tag{4.7}$$

Let us combine (4.6) and (4.7) in order to characterize $\bar{n}$: denoting $t_{RT} \equiv t_R + t_T$, we have

$$\bar{n} = \lambda(t_{RT} + \frac{b}{N+1-\bar{n}}). \tag{4.8}$$

This condition is in fact a second-degree equation in $\bar{n}$. It can be restated as $\lambda b - (\bar{n} - \lambda t_{RT})(N+1-\bar{n}) = 0$, or equivalently

$$\bar{n}^2 - \bar{n}(N+1+\lambda t_{RT}) + \lambda(b + (N+1)t_{RT}) = 0. \tag{4.9}$$

The existence of a solution requires that

$$(N+1+\lambda t_{RT})^2 \geq 4\lambda(b+(N+1)t_{RT})$$

$$\Leftrightarrow \quad (N+1-\lambda t_{RT})^2 \geq 4\lambda b \tag{4.10}$$

$$\Leftrightarrow \quad N+1 \geq \lambda t_{RT} + 2\sqrt{\lambda b} \;\text{ or }\; N+1-\lambda t_{RT} \leq -2\sqrt{\lambda b}. \tag{4.11a,b}$$

For the service fleet to meet customers' requests, it must hold that $N+1 \geq \lambda t_{RT}$ so that only (4.11a) makes sense. Thus, coming back to (4.9), the solution must verify that

$$\bar{n} = \tfrac{1}{2}(N+1+\lambda t_{RT}) + \varepsilon\sqrt{X^2 - \lambda b} \quad \text{where} \quad X \equiv \tfrac{1}{2}(N+1-\lambda t_{RT}) \quad \text{and} \quad \varepsilon \in \{-1,+1\}.$$

Only the root with $\varepsilon = -1$ makes sense, since that with $\varepsilon = +1$ induces a too large $\bar{n}$: by setting $\lambda \to 0^+$ i.e. almost no requests, (4.11b) would yield $\bar{n} \approx N+1$ - indeed an absurd outcome. To sum up, the average number of busy vehicles verifies

$$\bar{n} = \tfrac{1}{2}(N+1+\lambda t_{RT}) - \sqrt{X^2 - \lambda b}. \tag{4.12}$$

The following proposition is demonstrated in Appendix A.

Proposition 1: Properties of the mean number of busy vehicles:

(i) Function $\bar{n}$ is positive and continuous with respect to fleet size N and demand flow λ.

(ii) With respect to N, it is a decreasing function.

(iii) With respect to λ, $\bar{n}$ is an increasing and convex function.

4.3 The technical availability function

From this, we can easily deduce the mean access time per taxi: according to (4.6), $\bar{t}_A^V = \frac{1}{\lambda}\bar{n} - t_{RT}$. Then, from (4.12) we get

$$\bar{t}_A^V = \frac{X}{\lambda} - \sqrt{(\frac{X}{\lambda})^2 - \frac{b}{\lambda}} \quad \text{where} \quad X \equiv \tfrac{1}{2}(N+1-\lambda t_{RT}). \tag{4.13}$$

As $\bar{t}_{\mathrm{A}}^{\mathrm{V}} = b/(N+1-\bar{n})$, Proposition 1 implies the following properties (Cf. Appendix A).

Proposition 2: Properties of the mean vehicle access time:

(i) The mean access time $\bar{t}_{\mathrm{A}}^{\mathrm{V}}$ is a positive function that is continuous with respect to N , λ, t_{RT} and b.

(ii) It decreases with respect to N and increases with respect to b.

(iii) With respect to λ, the mean access time is an increasing and convex function.

The last property implies that $\partial\bar{t}_{\mathrm{A}}^{\mathrm{V}}/\partial\lambda$ is positive and increasing with λ, making $\partial\bar{t}_{\mathrm{A}}^{\mathrm{V}}/\partial Q$ positive and increasing: in other words, the access time is subject to congestion. The taxi service is a congestible economic good due to rivalry between its consumers.

4.4 On fleet sizing to meet production objectives

We can invert the dependencies, in order to deduce $\bar{n}$ and N from the mean vehicle access time $\bar{t}_{\mathrm{A}}^{\mathrm{V}}$. Owing to the characteristic equation, the mean number of busy vehicles depends simply on this: denoting $t_{\mathrm{ART}} \equiv t_{\mathrm{A}} + \bar{t}_{\mathrm{R}} + t_{\mathrm{T}}$:

$$\bar{n} = \lambda t_{\mathrm{ART}}. \tag{4.14}$$

Moreover, according to the approximation of mean access time in (4.7),

$$N + 1 - \bar{n} = b/t_{\mathrm{A}}.$$

By rearranging and ignoring the -1 here, we obtain the fleet size as the following function of demand volume Q and vehicle access time t_{A} denoted simply as t:

$$\hat{\mathrm{N}}(Q,t) = \frac{Q}{\mathrm{H}}(t_{\mathrm{RT}} + t) + \frac{b}{t}. \tag{4.15}$$

Formulas (4.14-15) link the service conditions, fleet size and occupancy to demand and to quality of service represented by mean access time, as well as to the technical conditions of transaction time t_{T} and mean running time t_{R}.

Proposition 3: Properties of the Fleet sizing function:

(i) Target fleet size $\hat{N}$ is a decreasing, convex function of mean access time t.

(ii) It is an increasing function of demand volume Q and of the times t_{R}, t_{T} and b.

The Proof is provided in Appendix A. From $\partial\hat{\mathrm{N}}/\partial t \leq 0$, it follows that:

$$Q \leq b\mathrm{H}t^{-2}. \tag{4.16}$$

Proposition 4: Upper bound on service volume. *The demand volume is bounded from above by $b\mathrm{H}t^{-2}$.*

In other words, the access time exerts a quadratic influence to reduce the usage volume.

Illustration. Figure 1 depicts both the mean access time function (part a) and the fleet sizing function (part b) with respect to usage volume Q. The parameter values are set up to $\mathrm{R} =$ 5km, $\mathrm{H} =$ 14h, $v =$ 20km/h hence $b = .79$h, $t_{\mathrm{R}} = .25$h, $t_{\mathrm{T}} = \tilde{t}_{\mathrm{T}} = .03$h.

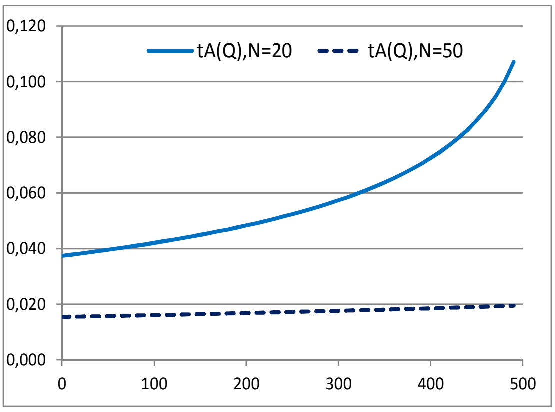

Fig. 1a. Access time according to usage volume, under given fleet size.

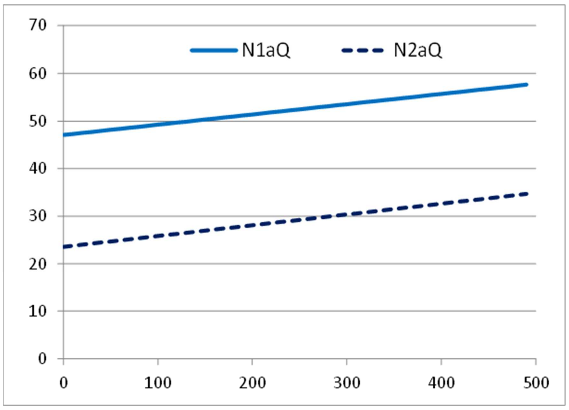

Fig. 1b. Fleet size according to usage volume, under given access time.

5. Demand function and traffic equilibrium

In a real-world service, the volume of demand is not exogenous, but is sensitive to the supply. We model demand on the basis of the price and quality of the service (§ 5.1), before studying the traffic equilibrium between supply and demand, in other words the relationship between volume of demand and access time, when the other conditions are exogenous (§ 5.2).

5.1 The demand function

We now postulate that demand for the service depends on the conditions available to the potential customers, expressed here mainly in terms of tariff τ and access time $t_{\mathrm{A}}^{\mathrm{V}} = t$. Here, we omit to express the run time t_{R} and the transaction time $\tilde{t}_{\mathrm{T}}$, in order to focus on τ and t.

As a first approach, we consider an aggregate demand function which links the volume of trips with the price and mean access time:

$$Q = \mathrm{D}(\tau, t)\,. \tag{5.1}$$

In principle, the volume decreases with price and with a reduction in quality of service, therefore with an increase in t. We can therefore invert the D function in relation to each of its arguments:

$$\tau = \mathrm{D}_{\tau}^{(-1)}(Q, t)\,. \tag{5.2}$$

$$t = \mathrm{D}_{t}^{(-1)}(Q, \tau)\,. \tag{5.3}$$

These two are also decreasing functions. In the aggregate demand function, we ignore the variability of t around its mean, even if in each particular situation the potential customer will avoid using taxis if the value of t is too high. Similarly, the influence of tariff is handled approximately: here, we use a mean price per ride, whereas in practice typical fares include a fixed pick-up component, combined with a variable component proportional to the distance or run time.

5.2 Traffic equilibrium: short-term interaction between supply and demand

In the theory of traffic assignment on a transport network, the supply and demand system is at traffic equilibrium when the interactions between supply and demand are jointly met:

$$t = \mathrm{T}_{\mathrm{A}}^{\mathrm{V}}(N, t_{\mathrm{T}}, t_{\mathrm{R}}, Q) \text{ according to (4.10)}$$

$$Q = \mathrm{D}(\tau, t) \text{ according to (5.1)}$$

The conjunction of these two conditions is a fixed point problem for the pair (t, Q). By replacing the second condition with (5.3), we obtain the following equation depending solely on the variable Q:

$$\mathrm{D}_t^{(-1)}(\tau, Q) = \mathrm{T}_{\mathrm{A}}^{\mathrm{V}}(N, t_{\mathrm{T}}, t_{\mathrm{R}}, Q). \tag{5.4}$$

Let us assume that service supply is fixed in the short run, i.e. τ and N are taken as exogenous. The following proposition is demonstrated in Appendix A.

Proposition 5: Existence and uniqueness of (short term) supply-demand equilibrium*. Given τ and N it holds that:*

(i) There is at most one supply-demand equilibrium;

(ii) If the demand function is continuous and verifies that $\mathrm{D}(\tau, b) < NH/b$, then there exists a supply-demand equilibrium.

We then express the solution as a function

$$Q = \mathrm{D}_{\mathrm{A}}(\tau, N, t_{\mathrm{T}}, t_{\mathrm{R}}, \tilde{t}_{\mathrm{T}}), \tag{5.5}$$

from which the mean access time is deduced by

$$t = \mathrm{T}_{\mathrm{A}}^{\mathrm{V}}(N, t_{\mathrm{T}}, t_{\mathrm{R}}, \mathrm{D}_{\mathrm{A}}(\tau, N, t_{\mathrm{T}}, t_{\mathrm{R}}, \tilde{t}_{\mathrm{T}})) \equiv \mathrm{T}_{\mathrm{AD}}(N, t_{\mathrm{T}}, t_{\mathrm{R}}, \tau, \tilde{t}_{\mathrm{T}}). \tag{5.6}$$

Proposition 6: Sensitivity analysis of (short term) supply-demand equilibrium*. The functions D_{A} and T_{AD} have monotonous variations in the same direction according to each of the factors N, τ, t_{T}, t_{R} and $\tilde{t}_{\mathrm{T}}$: (i) the two functions are increasing with respect to N; (ii) they are decreasing with respect to τ, t_{T}, t_{R} and $\tilde{t}_{\mathrm{T}}$.*

The proof is given in Appendix A.

Illustration. Figure 2 depicts the interaction of supply and demand in the short run, under given N and τ. Parameter values on the supply side are those already used for Figure 1. As for demand, the function has constant elasticity $\varepsilon = -2$ to generalized cost $g \equiv \tau + \alpha \tilde{t}_{\mathrm{RT}} + \beta t$, where $\alpha = 15$ €/h and $\beta = 10$ €/h. A reference point is (Q_0, g_0) with $g_0 =$ €10 per trip and Q_0 = 50,000 trips per day.

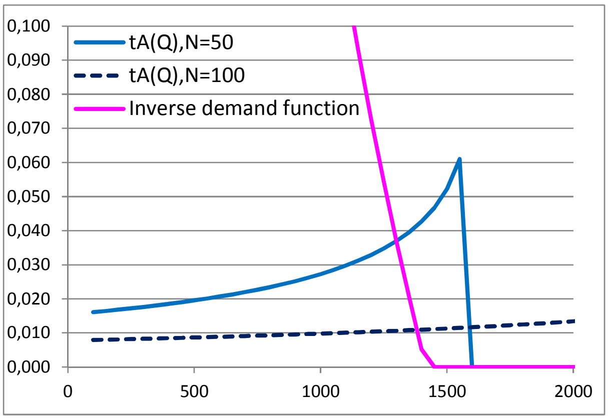

Fig. 2. Supply-demand interaction of taxi service in Orbicity, and the short-run traffic equilibrium.

6. Production cost and monopoly strategy

In traffic assignment on a transport network, despite the terminology usually employed, traffic equilibrium – even with elastic demand – is not a supply-demand equilibrium in the traditional sense of economic theory, since it does not say anything about the producer's strategic behavior as an economic agent (e.g. Ortuzar and Willumsen, 2011). In this section, we model production cost (§ 6.1), in order to study the strategic behavior of a producer in a monopoly (§ 6.2). The associated mathematical problem is easy to solve (§ 6.3), in particular for a volume of demand with constant elasticity (§ 6.4) or that depends only on generalized cost (§ 6.5). Under constant elasticity of demand to generalized cost, the optimality conditions yield explicit formulas for τ and t as action levers, hence for τ and N as supply factors: taking constant elasticity of -2 yields straightforward formulas (§ 6.6).

6.1 The supply cost function

We model the cost of production over a day as a function

$$\mathrm{C}_P(N,Q) \equiv \chi(N) + Q\,\mathrm{c}_u(t_{\mathrm{RT}} + t_{\mathrm{A}}), \tag{6.1}$$

in which the function χ is the daily cost of providing vehicles and operational staff (including vehicle rental or purchase amortization, maintenance and insurance, staff salaries and the associated charges, depots and headquarters leases), whereas c_u is the cost of use per trip time unit (including energy consumption, marginal wear…). The function χ increases with N more or less proportionally, above base cost χ_0. As a first approach, the derivative expressed as $\dot{\chi} \equiv d\chi / dN$ will be deemed constant. However, economies of scale can be expected, both in the purchase and maintenance of the vehicles, and in the allocation of people to drive them.

Using the relation between fleet size and a quality of service target $t_{\mathrm{A}}^{\mathrm{V}} = t$, we model the cost of production in relation to access time and volume of demand:

$$\hat{\mathrm{C}}_P(t,Q) \equiv \mathrm{C}_P(\hat{\mathrm{N}}(Q,t),Q) = \chi(\hat{\mathrm{N}}(Q,t)) + Q\,\mathrm{c}_u(t_{\mathrm{RT}} + t). \tag{6.2}$$

This formula shows that t (resp. Q) has a twofold influence on the cost of production, affecting both fixed and variable costs. Denoting $\mathrm{c}_u^+ \equiv \mathrm{c}_u + \dot{\chi}/H$ and considering the formula (4.15) for $\hat{N}(Q,t)$, the coefficient of sensitivity of $\hat{\mathrm{C}}_P$ to t is

$$\frac{\partial}{\partial t}\hat{\mathrm{C}}_P = \dot{\chi}\frac{\partial\,\hat{\mathrm{N}}}{\partial t} + Q\,\mathrm{c}_u = \dot{\chi}\left(\frac{Q}{\mathrm{H}} - \frac{b}{t^2}\right) + Q\,\mathrm{c}_u = Q\,\mathrm{c}_u^+ - \frac{\dot{\chi}b}{t^2}. \tag{6.3}$$

This is an increasing function: it takes negative values up to $t^* \equiv \sqrt{\dot{\chi}b/(Q\,\mathrm{c}_u^+)}$, then positive values.

Then, the sensitivity coefficient of $\hat{\mathrm{C}}_P$ to Q is

$$\frac{\partial}{\partial Q}\hat{\mathrm{C}}_P = \frac{\dot{\chi}}{\mathrm{H}}(t_{\mathrm{RT}} + t) + \mathrm{c}_u(t_{\mathrm{RT}} + t) = \mathrm{c}_u^+(t + t_{\mathrm{RT}}). \tag{6.4}$$

The coefficient is always positive. It does not depend directly on Q. It increases with t.

Moreover, the mean cost of service provision per ride is

$$\frac{\hat{\mathrm{C}}_P}{Q} = \frac{\chi}{Q} + \mathrm{c}_u(t_{\mathrm{RT}} + t) = \frac{\chi_0}{Q} + \frac{\dot{\chi}b}{Qt} + \mathrm{c}_u^{+}(t_{\mathrm{RT}} + t)\,. \tag{6.5}$$

This is therefore always higher than the marginal cost $\partial\hat{\mathrm{C}}_P/\partial Q$.

6.2 The monopolist's problem

If the provider is free to choose its fleet size and price, it will seek to maximize its profits by working on these two factors. Its strategic behavior is that of an unregulated monopoly.[2]

The producer's profit function is the difference between its revenues from sales and the costs of production:

$$\mathrm{P}_P(N,\tau,Q) = \tau.Q - \mathrm{C}_P(N,Q)\,. \tag{6.6}$$

The monopoly supplier's behavior is expressed by the following optimization program:

$$\max_{N,\tau}\ \mathrm{P}_P(N,\tau,Q)\ \text{ subject to }\ Q = \mathrm{D}(\tau,t) \text{ and } t = \mathrm{T}_{\mathrm{A}}^{\mathrm{V}}(N,t_{\mathrm{T}},t_{\mathrm{R}},Q)\,. \tag{6.7}$$

It is simpler to study profit maximization in terms of the price τ and access time t interpreted as a quality of service target. With the demand function "internalized", the profit function then becomes:

$$\hat{\mathrm{P}}_P(\tau,t) = \tau.\mathrm{D}(\tau,t) - \hat{\mathrm{C}}_P(t,\mathrm{D}(\tau,t))\,. \tag{6.8}$$

This varies depending on its arguments with the following sensitivity coefficients:

$$\frac{\partial}{\partial\tau}\hat{\mathrm{P}}_P(\tau,t) = \mathrm{D} + \tau\frac{\partial\mathrm{D}}{\partial\tau} - \frac{\partial\hat{\mathrm{C}}_P}{\partial Q}\frac{\partial\mathrm{D}}{\partial\tau}\,, \tag{6.9a}$$

$$\frac{\partial}{\partial t}\hat{\mathrm{P}}_P(\tau,t) = \tau\frac{\partial\mathrm{D}}{\partial t} - \frac{\partial\hat{\mathrm{C}}_P}{\partial t} - \frac{\partial\hat{\mathrm{C}}_P}{\partial Q}\frac{\partial\mathrm{D}}{\partial t}\,. \tag{6.9b}$$

Maximum profit is achieved at a point (τ,t), which satisfies the first-order optimality conditions:

$$\frac{\partial}{\partial\tau}\hat{\mathrm{P}}_P(\tau,t) = 0 \quad\text{and}\quad \frac{\partial}{\partial t}\hat{\mathrm{P}}_P(\tau,t) = 0\,. \tag{6.10a, b}$$

The system is equivalent to

$$(\tau - \frac{\partial\hat{\mathrm{C}}_P}{\partial Q})\frac{\partial\mathrm{D}}{\partial\tau} = -\mathrm{D}\,, \tag{6.11a}$$

$$(\tau - \frac{\partial\hat{\mathrm{C}}_P}{\partial Q})\frac{\partial\mathrm{D}}{\partial t} = \frac{\partial\hat{\mathrm{C}}_P}{\partial t}\,. \tag{6.11b}$$

By dividing each side in (6.11b) by the corresponding side in (6.11a), we get that at the monopoly optimum,

$$-\frac{\partial\mathrm{D}/\partial t}{\partial\mathrm{D}/\partial\tau}\mathrm{D} = \partial\hat{\mathrm{C}}_P/\partial t\,. \tag{6.12}$$

[2] Without intervention by a regulator or competition that would restrict the market price.

On the left side, we recognize demand volume Q multiplied by the monetary value of access time for a user, $\beta \equiv \frac{\partial D}{\partial t} / \frac{\partial D}{\partial \tau}$. Then (6.12) is equivalent to:

$$Q t^2 = b.\eta\,, \text{ where } \eta \equiv \frac{\dot{\chi}}{c_u^{+} + \beta}, \tag{6.13}$$

Coming back to equation (6.11a), it is equivalent to $\tau = \frac{\partial \hat{C}_P}{\partial Q} - \frac{D}{\partial D / \partial \tau}$.

Denoting by ε_τ the elasticity of volume of demand to price, then $\frac{D}{\partial D / \partial \tau} = \frac{\tau}{\varepsilon_\tau}$ and we get

$$\tau = \frac{\varepsilon_\tau}{1+\varepsilon_\tau} \frac{\partial \hat{C}_P}{\partial Q}. \tag{6.14}$$

At this stage, we can already determine whether the profit can be positive, which is the necessary condition for the supplier to provide a service (unless linked with additional resources). On average per ride, the condition of profitability is that $\tau \geq C_P / Q$. For a tariff that optimizes profit, this condition is equivalent to

$$\frac{\varepsilon_\tau}{1+\varepsilon_\tau} \frac{\partial \hat{C}_P}{\partial Q} \geq \frac{\hat{C}_P}{Q}. \tag{6.15}$$

Note also that the producer's profit cannot be positive if $\varepsilon_\tau > -1$. Assuming then that $\varepsilon_\tau < -1$ and using the formulas from §6.1, (6.15) means that

$\frac{c_u^{+} t_{ART}}{-\varepsilon_\tau - 1} \geq \frac{\dot{\chi} b}{Q t} + \frac{\chi_0}{Q}$, and also that

$$Q \geq (-\varepsilon_\tau - 1) \frac{\chi_0 + \dot{\chi} b / t}{c_u^{+} (t_{RT} + t)} \quad \text{if} \quad \varepsilon_\tau \leq -1. \tag{6.16}$$

Instance. Suppose that $\dot{\chi}$ = €400 per day for a taxi operating 10 hours per day, 7 days a week (i.e. 70 hours a week, therefore 2 employees per taxi, plus a 3rd to cover vacations, absences, cleaning…) and c_u = €5/h to cover 20 km with a diesel engine. Therefore c_u^{+} = €45/h. Then let us suppose that R = 3 km, hence $b \approx$ 14 min. If the access time is 1.5 minutes, i.e. 0.025 hours, the customer requirement must be in excess of 2400 rides a day to enable for service profitability. Under these conditions, for t_T = 2 min, the marginal cost is approximately €11.8, hence a price of almost €12. At this price, the service is likely to capture only a small proportion of trips around the city, let us say 5%: it would need a city providing at least 50,000 trips a day, i.e. a population of at least 15,000 (assuming 3.5 trips per person per day). If the modal share of the taxi is only 1%, then it would require an urban population of at least 75,000, even with our ideal geometry.

6.3 The solution to the monopolist's problem

In (6.12), let us replace the partial derivatives is by their respective counterparts in terms of elasticity, ε_τ and ε_t respectively:

$$-\frac{\tau}{\varepsilon_\tau} = \frac{t}{\varepsilon_t} \frac{1}{Q} \frac{\partial \hat{C}_P}{\partial t}. \tag{6.17}$$

Using the formulas for the partial derivatives of the production costs, conditions (6.14) and (6.17) are equivalent to the twofold system as follows:

$$\tau = \frac{\varepsilon_\tau}{1+\varepsilon_\tau} c_u^+ t_{\mathrm{ART}}, \tag{6.18a}$$

$$\frac{-\varepsilon_t}{1+\varepsilon_\tau} c_u^+ t_{\mathrm{ART}} = c_u^+ t - \frac{\dot{\chi} b}{Qt}. \tag{6.18b}$$

Considering τ to be linked to t by (6.18a) and Q to t both directly and via τ, through the demand function,[3] then (6.18b) constitutes a highly nonlinear equation solely in t.

It can be reformulated as:

$$(1+\frac{\varepsilon_t}{1+\varepsilon_\tau}) c_u^+ t + \frac{\varepsilon_t}{1+\varepsilon_\tau} c_u^+ t_{\mathrm{RT}} - \frac{\dot{\chi} b}{Qt} = 0 \ , \ \text{where} \ \ Q = \mathrm{D}(\tau_t, t). \tag{6.19}$$

Owing to (6.13), it is easy to recover production indicators as simple functions of the mean access time t only:

Fleet size: $$\hat{\mathrm{N}}(Q,t) = \frac{Q}{\mathrm{H}} t_{\mathrm{RT}} + \frac{b}{t} = Q(\frac{t_{\mathrm{RT}}+t}{\mathrm{H}} + \frac{t}{\eta}), \tag{6.20a}$$

Productivity per taxi: $$\frac{Q}{N} = (\frac{t_{\mathrm{RT}+t}}{\mathrm{H}} + \frac{t}{\eta})^{-1}, \tag{6.20b}$$

Per ride production cost: $$\frac{\hat{\mathrm{C}}_P}{Q} = \frac{\chi_0}{Q} + \frac{b\dot{\chi}}{Qt} + c_u^+ t_{\mathrm{ART}} = c_u^+ t_{\mathrm{RT}} + (\beta + 2c_u^+) t + \frac{\chi_0}{b\eta} t^2. \tag{6.20c}$$

6.4 Demand function with constant elasticities

Let us postulate here that the demand function has a product form with constant elasticities with respect to tariff and access time, respectively. Thus,

$$\mathrm{D}(\tau, t) = Q_0 (\frac{\tau}{\tau_0})^{\varepsilon_\tau} (\frac{t}{t_0})^{\varepsilon_t}, \tag{6.21}$$

Then $\dot{\mathrm{Q}}_\tau = \varepsilon_\tau Q/\tau$ and $\dot{\mathrm{Q}}_t = \varepsilon_t Q/t$, so that $\beta \equiv \frac{\dot{\mathrm{Q}}_t}{\dot{\mathrm{Q}}_\tau} = \frac{\tau}{t}\frac{\varepsilon_t}{\varepsilon_\tau}$.

Combining (6.19) and (6.21), we obtain

$$(1+\frac{\varepsilon_t}{1+\varepsilon_\tau}) c_u^+ t + \frac{\varepsilon_t}{1+\varepsilon_\tau} c_u^+ t_{\mathrm{RT}} = \frac{\dot{\chi} b}{Q_0 t_0 (\frac{\tau}{\tau_0})^{\varepsilon_\tau} (\frac{t}{t_0})^{\varepsilon_t + 1}}. \tag{6.22}$$

Using (6.18a), denoting $\varepsilon'_\tau \equiv \varepsilon_\tau/(1+\varepsilon_\tau)$ and defining $Z \equiv \frac{Q_0 c_u^+ (\varepsilon'_\tau c_u^+)^{\varepsilon_\tau}}{b\dot{\chi}\tau_0^{\varepsilon_\tau} t_0^{\varepsilon_t}}$, we get

[3] Hence the local elasticities ε_τ and ε_t depend on t

$$(t_{\mathrm{RT}}+t)^{\varepsilon_\tau} t^{\varepsilon_t+1} \frac{(t_{\mathrm{RT}}\varepsilon_t+(1+\varepsilon_t+\varepsilon_\tau)t)}{(1+\varepsilon_\tau)} = \frac{1}{Z}. \tag{6.23}$$

The fixed-point problem can be restated as

$$\frac{\varepsilon_t}{1+\varepsilon_\tau} t_{\mathrm{RT}} + \frac{1+\varepsilon_t+\varepsilon_\tau}{1+\varepsilon_\tau} t = \frac{1}{Z}(t_{\mathrm{RT}}+t)^{-\varepsilon_\tau} t^{-\varepsilon_t-1}. \tag{6.24}$$

The left-side function is a linear affine function that increases with respect to t if $\varepsilon_\tau < -1$, starting from origin ordinate $\frac{\varepsilon_t}{1+\varepsilon_\tau} t_{\mathrm{RT}} \geq 0$. As $-\varepsilon_\tau > 1$, the right-side function increases quicker than an affine function and it starts from ordinate value of zero at the origin. Thus there exists a unique solution.

Instance. Assuming $\varepsilon_\tau = -2$ and $\varepsilon_t = -1$, the FPP (6.24) becomes $t_{\mathrm{RT}} + 2t = Z^{-1}(t_{\mathrm{RT}}+t)^2$, i.e. $(t_{\mathrm{RT}}+t)^2 - 2(t_{\mathrm{RT}}+t)Z + t_{\mathrm{RT}}Z = 0$. Here, $Z = Q_0\tau_0^2 t_0/(4b\dot{\chi}c_u^+)$. Its solution is:

$$t = Z - t_{\mathrm{RT}} + \sqrt{Z^2 - t_{\mathrm{RT}}Z}.$$

Illustration. Figure 3 depicts the daily revenue, cost and profit functions with respect to demand volume. All curves are parameterized according to access time t, from which τ, Q and N, $\tau.Q$, C_P and P_P come out. Supply parameters are set up to $\chi_0 = 1000$€/day, $\dot{\chi} =$ 300€/day, $\mathrm{c}_u = 4$€/h. Concerning demand, constant elasticities $\varepsilon_\tau = -\frac{3}{2}$ and $\varepsilon_t = -\frac{1}{2}$ are considered, with reference point at Q_0 = 4,000 trips per day for $\tau_0 = 6$€ and $t_0 =$ 6min.

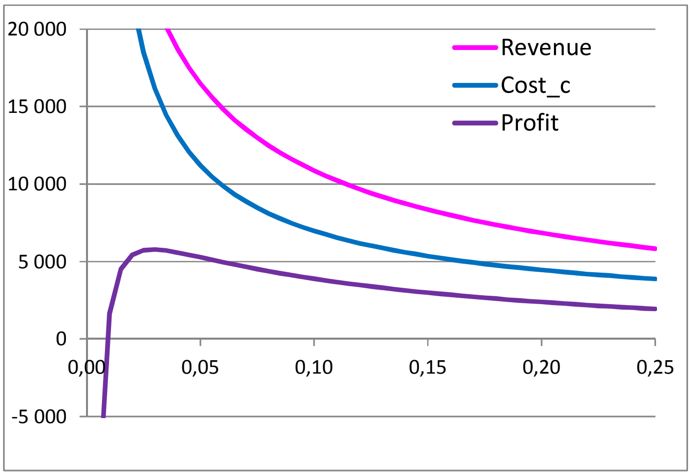

Fig. 3. Service revenues, costs and profits according to mean access time.

6.5 Demand function with respect to generalised cost

To extend the analysis, let us consider a demand function that is sensitive to price and access time through generalized cost. This is defined as follows, where α denotes the monetary value of ride time for a user (recall that β denotes the money value of access time):

$$g \equiv \tau + \alpha(t_{\mathrm{R}} + \tilde{t}_{\mathrm{T}}) + \beta t. \tag{6.25}$$

If $Q = \mathrm{D}(g)$ then $\partial \mathrm{D}/\partial\tau = \mathrm{dD}/\mathrm{d}g = \varepsilon Q/g$ where ε denotes the elasticity of demand to generalized cost, whereas $\partial \mathrm{D}/\partial t = \beta \mathrm{dD}/\mathrm{d}g = \varepsilon\beta Q/g$.

The optimality conditions for τ and t become respectively:

$$\tau - \frac{\partial \hat{\mathrm{C}}_P}{\partial Q} = -\frac{g}{\varepsilon}, \tag{6.26a}$$

$$\tau - \frac{\partial \hat{\mathrm{C}}_P}{\partial Q} = \frac{\partial \hat{\mathrm{C}}_P/\partial t}{\partial \mathrm{D}/\partial t} = \frac{\partial \hat{\mathrm{C}}_P}{\partial t}\frac{g}{\beta\varepsilon Q}. \tag{6.26b}$$

As these two conditions imply (6.13), given β there is a simple relationship linking t to Q:

$$t = \sqrt{\frac{b\eta}{Q}} \equiv \hat{\mathrm{t}}(Q). \tag{6.27a}$$

Combining (6.25) and (6.26b), we recover

$-\varepsilon\tau + \varepsilon \mathrm{c}_u^{\dagger}(t_{\mathrm{RT}} + t) = g = \tau + \alpha\tilde{t}_{\mathrm{RT}} + \beta t$, so that τ comes out as the following function of Q:

$$\hat{\tau}(Q) \equiv \frac{\varepsilon \mathrm{c}_u^{\dagger} t_{\mathrm{RT}} - \alpha\tilde{t}_{\mathrm{RT}}}{1+\varepsilon} + \frac{\varepsilon \mathrm{c}_u^{\dagger} - \beta}{1+\varepsilon}\hat{\mathrm{t}}(Q), \tag{6.27b}$$

By substitution, we recover $g = \alpha\tilde{t}_{\mathrm{RT}} + \dfrac{\varepsilon\, \mathrm{c}_u^{\dagger} t_{\mathrm{RT}} - \alpha\tilde{t}_{\mathrm{RT}}}{1+\varepsilon} + (\dfrac{\varepsilon \mathrm{c}_u^{\dagger} - \beta}{1+\varepsilon} + \beta)\sqrt{\dfrac{\dot{\chi} b}{(\beta + \mathrm{c}_u^{\dagger})Q}}$.

This makes the generalized cost a second, supply-related function of Q: letting $\varepsilon' \equiv \varepsilon/(1+\varepsilon)$ and $y_{\mathrm{RT}} \equiv \mathrm{c}_u^{\dagger} t_{\mathrm{RT}} + \alpha\tilde{t}_{\mathrm{RT}}$,

$$\hat{\mathrm{g}}(Q) \equiv \varepsilon'.(y_{\mathrm{RT}} + \sqrt{\frac{\dot{\chi} b(\mathrm{c}_u^{\dagger} + \beta)}{Q}}). \tag{6.27c}$$

As $g = \mathrm{D}^{(-1)}(Q)$ on the demand side, we finally obtain a problem with respect to Q:

$$\mathrm{D}^{(-1)}(Q) = \varepsilon'.(y_{\mathrm{RT}} + \sqrt{\frac{\dot{\chi} b(\mathrm{c}_u^{\dagger} + \beta)}{Q}}). \tag{6.28}$$

The relation (6.28) is a fixed point problem only in Q. The functions $\mathrm{D}^{(-1)}$ and $\hat{\mathrm{g}}$ decrease with Q. The existence and uniqueness of a solution depend on the form of the function D.

Let us hypothesize that [H-1] ε' is positive and non-decreasing with respect to Q, [H-2] that $\mathrm{D}^{(-1)}(Q).\sqrt{Q}$ is decreasing from a value Q_1 such that $\mathrm{D}^{(-1)}(Q_1) > \hat{\mathrm{g}}(Q_1)$.

Condition [H-1] requires that $\varepsilon < -1$. Condition [H-2] requires function $\mathrm{D}^{(-1)}(Q).\sqrt{Q}$ to have negative elasticity, i.e. $\frac{1}{\varepsilon} + \frac{1}{2} \leq 0$ hence $|\varepsilon| \leq 2$.

Multiplying both sides in (6.28) by $\sqrt{Q}$, an equivalent formulation of the fixed-point problem is:

$$\mathrm{D}^{(-1)}(Q).\sqrt{Q} = \varepsilon'.(y_{\mathrm{RT}}\sqrt{Q} + \sqrt{\dot{\chi} b(\mathrm{c}_u^{\dagger} + \beta)}). \tag{6.29}$$

Proposition 6: existence and uniqueness of monopoly solution. Under conditions [H-1] and [H-2], the FPP characterizing the monopoly strategy admits at least one solution, which is unique on $[Q_1,+\infty[$.

Proof. [H-1] ensures that the left-side function in (6.29) is decreasing, with value at Q_1 above that of the right-side function. The right-side function is increasing and has no upper limit. Thus the two curves must cross on $[Q_1,+\infty[$ and this happens only once due to the respective direction of variations.

If the elasticity is constant, we write $Q = Q_0(g/g_0)^{\varepsilon}$ and $\mathrm{D}^{(-1)}(Q) = g_0(Q/Q_0)^{1/\varepsilon}$. An equivalent, dual formulation of the fixed-point problem involves t as the unknown variable in place of Q: based on (6.13) and (6.29), the characteristic equation is

$$g_0\left(\frac{b\eta}{Q_0}\right)^{1/\varepsilon} t^{-2/\varepsilon} = \varepsilon'.[y_{\mathrm{RT}} + (\mathrm{c}_u^+ + \beta)t]. \tag{6.30}$$

6.6 Special case with demand elasticity of -2 to generalized cost

In the special case with $\varepsilon = -2$, then $\varepsilon' = 2$ and the previous equation becomes linear:

$$g_0\sqrt{\frac{Q_0}{b\eta}}\, t = 2[y_{\mathrm{RT}} + (\mathrm{c}_u^+ + \beta)t]. \tag{6.31}$$

From this we deduce that:

$$t^{\mathrm{MO}} = \frac{y_{\mathrm{RT}}}{\frac{1}{2}g_0\sqrt{Q_0/(b\eta)} - (\mathrm{c}_u^+ + \beta)}, \tag{6.32a}$$

$$Q^{\mathrm{MO}} = \left(\frac{\frac{1}{2}g_0\sqrt{Q_0} - \sqrt{b\dot{\chi}(\mathrm{c}_u^+ + \beta)}}{y_{\mathrm{RT}}}\right)^2, \tag{6.32b}$$

$$g^{\mathrm{MO}} = \frac{y_{\mathrm{RT}}}{\frac{1}{2} - x}, \text{ where } x \equiv \frac{1}{g_0}\sqrt{\frac{b\dot{\chi}(\mathrm{c}_u^+ + \beta)}{Q_0}}, \tag{6.32c}$$

$$\tau^{\mathrm{MO}} + \alpha\tilde{t}_{\mathrm{RT}} = g^{\mathrm{MO}} - \beta t^{\mathrm{MO}} = \frac{y_{\mathrm{RT}}}{\frac{1}{2} - x}\left(1 - \frac{\beta}{\beta + \mathrm{c}_u^+}x\right). \tag{6.32d}$$

Furthermore, from (6.26a) the tariff verifies $\tau = \frac{1}{2}g + \mathrm{c}_u^+(t_{\mathrm{RT}} + t)$, hence $\tau = y_{\mathrm{RT}} + (\mathrm{c}_u^+ + \beta)t + \mathrm{c}_u^+ t_{\mathrm{ART}}$. Subtracting from that the per trip production cost stated in (6.20c), we obtain the per trip profit as follows:

$$\tau - \frac{\hat{\mathrm{C}}_P}{Q} = y_{\mathrm{RT}} - \frac{\chi_0}{Q}.$$

The monopoly can be profitable in the absence of subsidy only if

$$Q \geq \chi_0 / y_{\mathrm{RT}}, \text{ i.e. } \tfrac{1}{2}g_0\sqrt{Q_0} \geq \sqrt{\chi_0 y_{\mathrm{RT}}} + \sqrt{b\dot{\chi}(\mathrm{c}_u^+ + \beta)}.$$

The threshold value for Q involves the fixed supply cost χ_0 in a proportional way, divided by a per trip cost that monetizes the run and transaction times on both the supplier and customer sides.

Illustration. Figure 4 depicts the profit maximizing behavior of the service provider in the Monopoly setting. With respect to demand volume Q, the users' demand function yields a demand-side generalized cost $\mathrm{D}^{(-1)}(Q)$, whereas the profit maximizing behavior yields a supply-side generalized cost $\hat{g}(Q)$. The intersection point characterizes a state of supply-demand equilibrium. Supply parameters are set up as in Figure 2. On the demand side, average user values of time are set to $\alpha = 12$ €/h and $\beta = 10$ €/h. A reference point on the demand function is (Q_0, g_0) with $g_0 = 10$ € and $Q_0 = 50{,}000$ trips per day. Constant elasticity is set to $\varepsilon = -2$, yielding $\varepsilon' = 2$ and curve labeled "gMO". The alternative supply curve labeled "gSO" will be shown in the next Section to model the first-best System Optimum: it obeys to the same formula as gMO except for setting up $\varepsilon' = 1$.

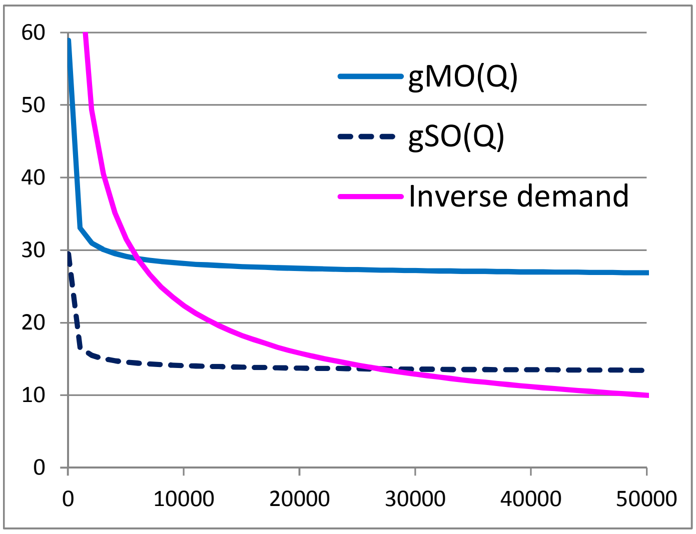

Fig. 4. Generalized cost functions according to demand volume: monopoly solution lies at the intersection point of the supply and demand curves.

7. System Optimum (SO, First-best)

In order to maximize its profit, a monopoly producer tends to limit demand, and therefore the market share of its service. Let us now change perspective in order to study the collective optimum for the service. After modeling the collective surplus (§ 7.1), we characterize an optimum state for the system in terms of tariff and access time (§ 7.2). We then address the specific demand function with constant elasticity to generalized cost (§ 7.3) and provide closed-form results for two special values of that elasticity: value -2 (§ 7.4) and value -1 (§ 7.5).

7.1 Definition of the collective surplus

To society as a whole, the general interest associated to an economic service involves not only the producer's profit but also and above all, the contribution of the service to the well-being of its customers. When demand depends solely on generalized cost, the customers' well-being is measured by the following function of net consumer surplus (e.g. Varian, 1992):

$$\mathrm{P}_D(g) = \int_g^{+\infty} \mathrm{D}(g')\,\mathrm{d}g'.$$

The collective surplus is then

$$\mathrm{P}_S \equiv \mathrm{P}_D(g) + \mathrm{P}_P = \int_g^{+\infty} \mathrm{D}(g')\,\mathrm{d}g' + \tau.\mathrm{D}(g) - \mathrm{C}_P\,. \tag{7.1}$$

Here the analysis is restricted to two categories of agents, the consumers and the producer of the service. Given that the issue is urban mobility, the scope could be widened by also considering residents affected by local impacts (noise, pollution), the environment at least in terms of energy consumption and greenhouse gas emissions, competing providers who produce other mobility services, other road users who travel and park on the streets alongside taxis, etc. We will broaden the scope in a subsequent article, which will deal specifically with regulation, and will consider a wide range of instruments.

Here, we consider the two variables of action: price and mean access time. The sensitivity coefficient of the collective surplus to price is

$$\frac{\partial}{\partial\tau}\hat{\mathrm{P}}_S(\tau,t) = -\mathrm{D} + \mathrm{D} + \tau\dot{\mathrm{D}} - \frac{\partial\hat{\mathrm{C}}_P}{\partial Q}\dot{\mathrm{D}} = (\tau - \frac{\partial\hat{\mathrm{C}}_P}{\partial Q})\dot{\mathrm{D}} \text{ where } \dot{\mathrm{D}} \equiv \mathrm{dD}/\mathrm{d}g\,. \tag{7.2a}$$

whereas the sensitivity coefficient of the collective surplus to access time is:

$$\frac{\partial}{\partial t}\hat{\mathrm{P}}_S(\tau,t) = -\beta\mathrm{D} + \tau\beta\dot{\mathrm{D}} - \frac{\partial\hat{\mathrm{C}}_P}{\partial t} - \beta\dot{\mathrm{D}}\frac{\partial\hat{\mathrm{C}}_P}{\partial Q}\,. \tag{7.2b}$$

These two coefficients are less than those of the producer's profit, because an increase in either tariff or access time will decrease the net consumer surplus.

7.2 The collective optimum and the associated optimality conditions

To society as a whole, the optimum state of the service as a supply-demand system is a state at which the collective surplus is maximized with respect to price and mean access time. Such a system optimum (SO) state can be stated as a solution to the following maximization program:

$$\max_{\tau,t}\ \hat{\mathrm{P}}_S(\tau,t) \quad \text{subject to} \quad g = \tau + \alpha\tilde{t}_{\mathrm{RT}} + \beta t \quad \text{and} \quad Q = \mathrm{D}(g)\,. \tag{7.3}$$

The mathematical maximization program presents the following first order optimality conditions:

$$\frac{\partial}{\partial\tau}\hat{\mathrm{P}}_S(\tau,t) = 0 \text{ and so } \tau = \frac{\partial\hat{\mathrm{C}}_P}{\partial Q}\,, \tag{7.4a}$$

$$\frac{\partial}{\partial t}\hat{\mathrm{P}}_S(\tau,t) = 0 \text{ and so } \frac{\partial\hat{\mathrm{C}}_P}{\partial t} = -\beta\mathrm{D}\,. \tag{7.4b}$$

Condition (7.4a) gives that

$$\tau = c_u^{+}(t_{\mathrm{RT}} + t)\,. \tag{7.5}$$

Condition (7.4b) is equivalent to relation (6.13) obtained for the monopoly:

$$Qt^2 = \frac{\dot{\chi}b}{\beta + \mathrm{c}_u^{+}}\,. \tag{7.6}$$

The collective optimization problem is handled in the same way as the monopolist's problem: it is reduced to a fixed point problem depending only on Q. From (7.5), we relate t to Q as follows:

$$t = \tilde{\mathrm{t}}(Q) \equiv \sqrt{\frac{b\eta}{Q}}\,. \tag{7.6a}$$

Combining this to (7.4a), we obtain

$$\tau = \tilde{\tau}(Q) \equiv \mathrm{c}_u^{+}(t_{\mathrm{RT}} + \sqrt{\frac{b\eta}{Q}})\,. \tag{7.6b}$$

So that the generalized cost now satisfies the condition that $\tilde{\mathrm{g}}(Q) = \tilde{\tau}(Q) + \alpha\tilde{t}_{\mathrm{RT}} + \beta\,\tilde{\mathrm{t}}(Q)$, therefore

$$g = \tilde{\mathrm{g}}(Q) \equiv y_{\mathrm{RT}} + \frac{\sqrt{b\dot{\chi}(\beta + c_u^{+})}}{\sqrt{Q}}\,. \tag{7.6c}$$

Which gives the final condition dependent only on Q,

$$\mathrm{D}^{(-1)}(Q) = y_{\mathrm{RT}} + \frac{\sqrt{b\dot{\chi}(\beta + c_u^{+})}}{\sqrt{Q}}\,. \tag{7.7}$$

As in Proposition 6, if $\mathrm{D}^{(-1)}(Q).\sqrt{Q}$ is a decreasing function from a value Q_1 such that $\mathrm{D}^{(-1)}(Q_1) > \tilde{\mathrm{g}}(Q_1)$, then as the right-hand side function is increasing and has no upper limit, the two curves necessarily cross only once, which ensures that there exists a unique solution beyond Q_1.

Since the left-hand member in (7.7) is lower than the one in the monopolist's problem, $\tilde{\mathrm{g}} < \hat{g}$ therefore $\tilde{Q} > \hat{Q}$ and hence $\tilde{t} < \hat{t}$ and $\tilde{\tau} < \hat{\tau}$: it is in the collective interest for the access time and tariff to be reduced, and consequently for the vehicle fleet to be larger than it is for a monopolist. Furthermore, based on (6.20), the reduction in t compared with the monopolist's problem implies a reduction in the unit cost of provision, $\hat{\mathrm{C}}_P/Q = \mathrm{c}_u^{+}\,t_{\mathrm{RT}} + (2\mathrm{c}_u^{+} + \beta)t + \chi_0/Q$, together with an increase in fleet size as well as in taxi productivity.

However, the most salient difference between the two states, monopoly versus collective optimum, lies in the tariff level. Comparing (7.6b) to (6.27b), the MO tariff is in excess of the SO counterpart not only by the increase in access time but also, and more significantly by far, by a $-\varepsilon g$ term that is greater than the full generalized cost if $\varepsilon < -1$.

As (6.13) holds true, so do the formulas (6.20) linking the production indicators to the access time. From (6.4), (6.5) and (7.4a), we have that

$$\frac{\hat{\mathrm{C}}_P}{Q} = \tau + \frac{\chi_0}{Q} + \frac{\dot{\chi}b}{Qt}\,.$$

As the average production cost is greater than the associated tariff fare, the service structurally runs at a loss. The producer's net profit is negative by the following amount, which needs be paid to the producer as a balancing subsidy:

$$-\hat{\mathrm{P}}_P = \chi_0 + \frac{\dot{\chi}b}{\tilde{t}} = \chi_0 + \sqrt{\dot{\chi}b(\beta + \mathrm{c}_u^+)\tilde{Q}}\,.$$

The shorter the access time, the greater the producer's deficit.

7.3 Demand function with constant elasticity to generalized cost

If the elasticity is constant, we write $Q = Q_0(g/g_0)^{\varepsilon}$ and $\mathrm{D}^{(-1)}(Q) = g_0(Q/Q_0)^{1/\varepsilon}$. An equivalent, dual formulation of the fixed-point problem involves t as the unknown variable in place of Q: based on (6.13) and (7.7), the characteristic equation is

$$g_0\left(\frac{b\eta}{Q_0}\right)^{1/\varepsilon} t^{-2/\varepsilon} = y_{\mathrm{RT}} + (\mathrm{c}_u^+ + \beta)t\,. \tag{7.8}$$

Illustration. Figure 5 depicts the influences of the access time on generalized cost functions either demand-related or supply-related. Each intersection point is a solution to a specific fixed point problem. Two different demand-related curves are modeled, first for $\varepsilon = -2$ and second for $\varepsilon = -1$. Three supply-related curves are modeled: for Monopoly (MO) under $\varepsilon = -2$, System Optimum (SO) and 2nd best System Optimum (S2), both of which do not depend on demand elasticity. All parameter values are set up as in Figures 3 and 4, save for ε, Q_0 = 20,000 trips/day and g_0 = 20€. The instance suggests first that only short access times are relevant, second that the different generalized costs associated to the different management strategies do exhibit significantly different values, even if the underlying access time values are about close. In other words, the management strategies differ mostly regarding tariff fares. The gap between MO and SO is much wider than that between SO and S2.

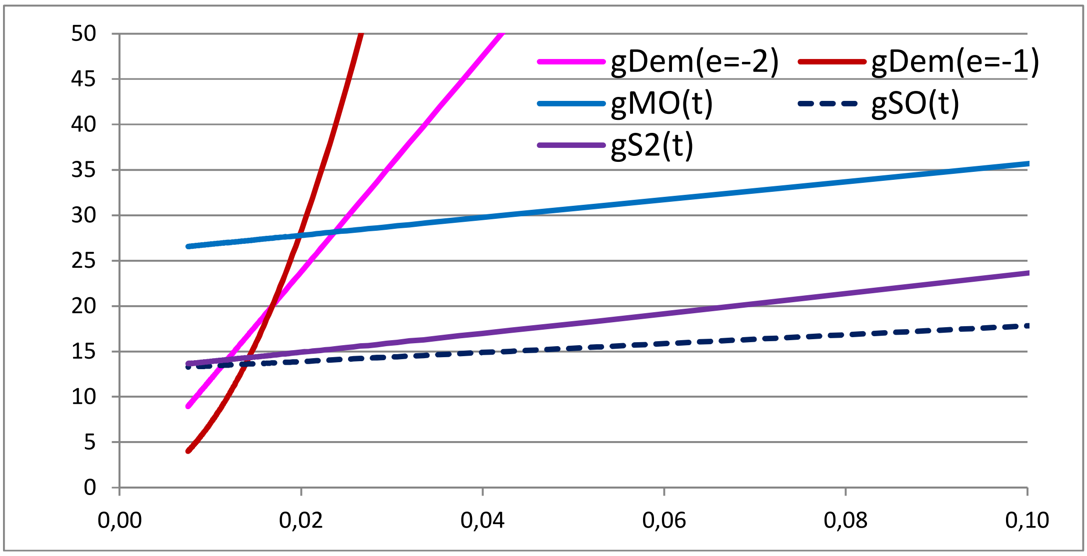

Fig. 5. Generalized cost with respect to access time: demand-side cost with -2 elasticity (resp. -1), supply-side costs under 1st best System Optimum (SO), 2nd best (S2), as compared to Monopoly (MO).

7.4 Special case with demand elasticity of -2 to generalized cost

In the special case with $\varepsilon = -2$, then the previous equation becomes linear:

$$g_0\sqrt{\frac{Q_0}{b\eta}}\,t = y_{\mathrm{RT}} + (\mathrm{c}_u^+ + \beta)t\,, \tag{7.9}$$

From which we deduce that:

$$t^{\mathrm{SO}} = \frac{y_{\mathrm{RT}}}{g_0 \sqrt{Q_0/(b\eta)} - (\mathrm{c}_u^+ + \beta)}\,, \tag{7.10a}$$

$$Q^{\mathrm{SO}} = \left(\frac{g_0\sqrt{Q_0} - \sqrt{b\dot{\chi}(\mathrm{c}_u^+ + \beta)}}{y_{\mathrm{RT}}}\right)^2, \tag{7.10b}$$

$$g^{\mathrm{SO}} = \frac{y_{\mathrm{RT}}}{1-x},\ \text{ where }\ x \equiv \frac{1}{g_0}\sqrt{\frac{b\dot{\chi}(\mathrm{c}_u^+ + \beta)}{Q_0}}\,, \tag{7.11c}$$

$$\tau^{\mathrm{SO}} + \alpha\tilde{t}_{\mathrm{RT}} = g^{\mathrm{SO}} - \beta t^{\mathrm{SO}} = \frac{y_{\mathrm{RT}}}{1-x}\left(1 - \frac{\beta}{\beta + \mathrm{c}_u^+}x\right). \tag{7.12d}$$

The values for access time, generalized cost and price, are to be brought together with the values for the monopoly's optimum profit. Clearly,

$$\frac{g^{\mathrm{MO}}}{g^{\mathrm{SO}}} = \frac{t^{\mathrm{MO}}}{t^{\mathrm{SO}}} = \frac{\tau^{\mathrm{MO}} + \alpha\tilde{t}_{\mathrm{RT}}}{\tau^{\mathrm{SO}} + \alpha\tilde{t}_{\mathrm{RT}}} = \frac{1-x}{\frac{1}{2}-x}.$$

Now $x \in\,]0, \frac{1}{2}[$ so $\frac{1}{2} \geq \frac{1}{2} - x$ and $\frac{1-x}{\frac{1}{2}-x} = 1 + \frac{1}{2}/(\frac{1}{2}-x) \geq 2$. Therefore $g^{\mathrm{MO}} \geq 2g^{\mathrm{SO}}$, from which it follows that $t^{\mathrm{MO}} \geq 2t^{\mathrm{SO}}$, $\tau^{\mathrm{MO}} \geq 2\tau^{\mathrm{SO}} + \alpha\tilde{t}_{\mathrm{RT}}$ and $Q^{\mathrm{MO}} \leq \frac{1}{4}Q^{\mathrm{SO}}$.

On these assumptions, the monopolist's optimum tariff is more than twice as high as the optimum price for the community.

An instance of comparison between SO and MO under $\varepsilon = -2$ is provided in Figure 4 (cf. § 6.6).

7.5 Special case with demand elasticity of -1 to generalized cost

In the special case of elasticity $\varepsilon = -1$, the equation (7.8) is quadratic in t,

$$\frac{g_0 Q_0}{b\eta}t^2 = y_{\mathrm{RT}} + (\mathrm{c}_u^+ + \beta)t\,. \tag{7.13}$$

Only the positive solution is valid. Putting the equation as $t^2 - \frac{b\dot{\chi}}{g_0 Q_0}t - \frac{b\eta\, y_{\mathrm{RT}}}{g_0 Q_0} = 0$, we get

$$t^{\mathrm{SO}} = \frac{b\dot{\chi}}{2g_0 Q_0} + \sqrt{\left(\frac{b\dot{\chi}}{2g_0 Q_0}\right)^2 + \frac{b\eta\, y_{\mathrm{RT}}}{g_0 Q_0}}\,, \tag{7.14a}$$

$$Q^{\mathrm{SO}} = \frac{b\eta}{(t^{\mathrm{SO}})^2}\,, \tag{7.14b}$$

$$g^{\mathrm{SO}} = g_0\sqrt{\frac{Q_0}{b\eta}}\,t^{\mathrm{SO}}\,, \tag{7.14c}$$

$$\tau^{\mathrm{SO}} + \alpha\tilde{t}_{\mathrm{RT}} = g^{\mathrm{SO}} - \beta t^{\mathrm{SO}} = \left(g_0\sqrt{\frac{Q_0}{b\eta}} - \beta\right)t^{\mathrm{SO}}. \tag{7.14d}$$

8. Second-best optimum with a budgetary constraint

As the first-best system optimum state results in an operating deficit for the producer, we will also study the second-best optimum that takes the condition of production profitability as a constraint. We shall firstly state the constrained maximization program and its optimality conditions (§ 8.1), secondly consider a demand function with constant elasticity with respect to generalized cost (§ 8.2) and thirdly address the two particular cases of value -2 (§ 8.3) or value -1 (§ 8.4).

8.1 Constrained optimization problem

When the public finances are tight, it is difficult to subsidize service production and to achieve the first-best system optimum. An alternative policy is to search for the conditions that maximize the collective surplus while ensuring service profitability. Such "second-best" system optimum is defined mathematically as the solution to the following program of constrained maximization. The economic program is then:

$$\max_{\tau,t}\ \hat{\mathrm{P}}_S(\tau,t,Q) \text{ subject to } Q=\mathrm{D}(\tau,t) \text{ and } \tau.Q\geq \hat{\mathrm{C}}_P(t,Q). \tag{8.1}$$

To solve this problem, we associate the parameter $\gamma \geq 0$ with the budget constraint and form the Lagrangian function for the system,

$\pounds(\tau,t,\gamma)\equiv \hat{\mathrm{P}}_S(\tau,t,Q)+\gamma(\tau.Q-\hat{\mathrm{C}}_P(t,Q))$, where $Q=\mathrm{D}(\tau,t)$, hence

$$\pounds(\tau,t,\gamma)=\int_g^{+\infty}\mathrm{D}+(1+\gamma)(\tau.Q-\hat{\mathrm{C}}_P(t,Q)). \tag{8.2}$$

From this, the partial derivatives can easily be calculated:

$$\partial\pounds/\partial\tau=\gamma\mathrm{D}+(1+\gamma)(\tau-\partial\hat{\mathrm{C}}_P/\partial Q).\dot{\mathrm{D}}, \tag{8.3a}$$

$$\partial\pounds/\partial t=-\alpha\mathrm{D}+(1+\gamma)[\alpha\dot{\mathrm{D}}.(\tau-\partial\hat{\mathrm{C}}_P/\partial Q)-\partial\hat{\mathrm{C}}_P/\partial t], \tag{8.3b}$$

$$\partial\pounds/\partial\gamma=\tau.Q-\hat{\mathrm{C}}_P(t,Q). \tag{8.3c}$$

The constrained optimization problem is the equivalent of a saddle point problem for the Lagrangian, with the following first-order optimality conditions:

$$\partial\pounds/\partial\tau=0, \tag{8.4a}$$

$$\partial\pounds/\partial t=0, \tag{8.4b}$$

$$\partial\pounds/\partial\gamma\geq 0 \text{ and } \gamma.\partial\pounds/\partial\gamma=0. \tag{8.4c, d}$$

By (8.3a) and (8.4a), at the constrained optimum

$$(1+\gamma)(\tau-\partial\hat{\mathrm{C}}_P/\partial Q).\dot{\mathrm{D}}=-\gamma\mathrm{D}, \tag{8.5a}$$

By combining this relation with (8.3b) and (8.4b), we obtain $\partial\hat{\mathrm{C}}_P/\partial t=-\beta\mathrm{D}$, hence

$$\frac{\dot{\chi}b}{Qt^2}=\beta+\mathrm{c}_u^+. \tag{8.5b}$$

So the relationship (6.13) between Q and t holds for the second-best optimum state (denoted S2) as well as for the MO and SO states. So the formulas (6.20) for production indicators still apply.

As the production runs at a loss under SO, the budgetary constraint is active under S2. Thus

$$\tau = \frac{\hat{\mathrm{C}}_P}{Q} = \mathrm{c}_u^+ t_{\mathrm{RT}} + \mathrm{c}_u^+ t + \frac{\dot{\chi} b}{Qt} + \frac{\chi_0}{Q} = \mathrm{c}_u^+ t_{\mathrm{RT}} + (\beta + 2\mathrm{c}_u^+)t + \frac{\chi_0}{Q}. \tag{8.6a}$$

$$\tau - \frac{\partial \hat{\mathrm{C}}_P}{\partial Q} = \frac{\dot{\chi} b}{Qt} + \frac{\chi_0}{Q} = (\beta + \mathrm{c}_u^+)t + \frac{\chi_0}{Q}. \tag{8.6b}$$

Using (8.5a), we obtain that $(\beta + \mathrm{c}_u^+)t = -\gamma' \mathrm{D}/\dot{\mathrm{D}}$, where we define $\gamma' \equiv \gamma/(1+\gamma)$. By replacing $\mathrm{D}/\dot{\mathrm{D}}$ with g/ε, we get

$$\gamma' = -\varepsilon \frac{(\beta + \mathrm{c}_u^+)t}{g}. \tag{8.6}$$

Equations (8.5b) and (8.6a) determine t and τ as functions of Q when the constraint is binding:

$$t = \sqrt{\frac{b\eta}{Q}} \equiv \tilde{\mathrm{t}}(Q). \tag{8.7a}$$

$$\tau = \mathrm{c}_u^+ t_{\mathrm{RT}} + (\beta + 2\mathrm{c}_u^+)\sqrt{\frac{b\eta}{Q}} + \frac{\chi_0}{Q}. \tag{8.7b}$$

Therefore the constrained generalized cost is

$$g = y_{\mathrm{RT}} + 2(\beta + \mathrm{c}_u^+)\sqrt{\frac{b\eta}{Q}} + \frac{\chi_0}{Q}. \tag{8.7c}$$

The S2 state satisfies the following fixed-point problem with respect to Q:

$$\mathrm{D}^{(-1)}(Q) = y_{\mathrm{RT}} + 2(\beta + \mathrm{c}_u^+)\sqrt{\frac{b\eta}{Q}} + \frac{\chi_0}{Q}. \tag{8.8}$$

As the right-side function in (8.8) is greater than that in (7.7), the intersection with the left-side function takes place at a lower value: $Q^{\mathrm{S2}} \le Q^{\mathrm{SO}}$, yielding then $g^{\mathrm{S2}} \ge g^{\mathrm{SO}}$ via the demand function and $t^{\mathrm{S2}} \ge t^{\mathrm{SO}}$ via relation (8.5b). From (8.7b) compared to (7.6), it also follows that $\tau^{\mathrm{S2}} \ge \tau^{\mathrm{SO}}$.

Let us compare now the S2 state to the Monopoly one. If an S2 state exists, then the producer’s profit can take non-negative values. Under MO it is greater than under S2, so that $\tau^{\mathrm{MO}} Q^{\mathrm{MO}} \ge \tau^{\mathrm{S2}} Q^{\mathrm{S2}}$.

Yet $\mathrm{P_S}(Q^{\mathrm{S2}}) \ge \mathrm{P_S}(Q^{\mathrm{MO}})$, so that necessarily $\mathrm{P}_D(Q^{\mathrm{S2}}) \ge \mathrm{P}_D(Q^{\mathrm{MO}})$, requiring that $Q^{\mathrm{S2}} \ge Q^{\mathrm{MO}}$. It also holds that:

- $\tau^{\mathrm{MO}} \ge \tau^{\mathrm{S2}}$ from the comparison of revenues,
- $g^{\mathrm{MO}} \ge g^{\mathrm{S2}}$ from the inverse demand function,
- $t^{\mathrm{MO}} \ge t^{\mathrm{S2}}$ from relations (6.13) and (8.5b).

8.2 Demand function with constant elasticity to generalized cost

If the demand function has constant elasticity ε to generalized cost, again we have that $Q = Q_0(g/g_0)^{\varepsilon}$ and $D^{(-1)}(Q) = g_0(Q/Q_0)^{1/\varepsilon}$. Owing to (8.5b), the S2 program admits a dual formulation with respect to access time t as follows:

$$g_0\left(\frac{b\eta}{Q_0}\right)^{1/\varepsilon} t^{-2/\varepsilon} = y_{\mathrm{RT}} + 2(c_u^{\dagger} + \beta)t + \frac{\chi_0}{b\eta}t^2. \tag{8.9}$$

This implies that if $-\varepsilon < 1$ then there is a unique solution to the S2 state. More generally, the existence of an S2 state is equivalent to the feasibility of supply profitability, i.e. to the GO condition for the Monopoly.

Illustration. Figure 6 depicts both the S2 and the SO fixed point problems according to usage volume. Two different demand curves are modeled, first for $\varepsilon = -2$ and second for $\varepsilon = -1$. All parameter values are set up as in Figure 5. The S2 supply-related function is greater than its SO counterpart, thereby reducing the demand volume while increasing the generalized cost at the solution point: yet in this instance the difference between the SO and S2 generalized costs is small.

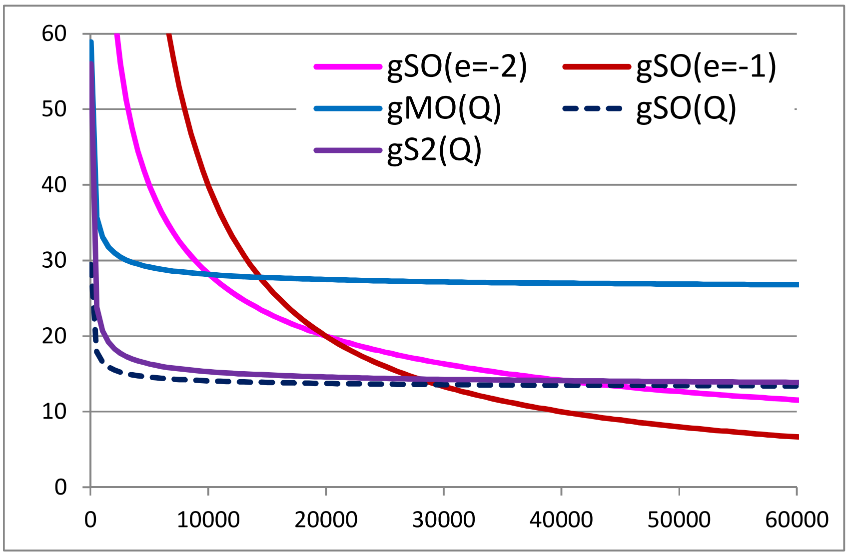

Fig. 6. Generalized cost with respect to usage volume: demand-side cost with -2 elasticity (resp. -1), supply-side costs under 1st best System Optimum (SO), 2nd best (S2), as compared to Monopoly (MO).

8.3 Special case with demand elasticity of -2 to generalized cost

When $\varepsilon = -2$, the dual fixed-point problem is a second-degree equation in t:

$g_0\sqrt{\dfrac{Q_0}{b\eta}}\,t = y_{\mathrm{RT}} + 2(c_u^{\dagger} + \beta)t + \dfrac{\chi_0}{b\eta}t^2$, which is equivalent to

$$g_0\sqrt{Q_0 b\eta}\,t = b\eta y_{\mathrm{RT}} + 2b\dot{\chi}t + \chi_0 t^2. \tag{8.10}$$

For a non-negative solution to exist, we must have that $g_0\sqrt{Q_0 b\eta} \geq 2b\dot{\chi}$, i.e. $Q_0 \geq 4\dfrac{b\dot{\chi}(\beta + c_u^{\dagger})}{g_0^2}$.

If $\chi_0 > 0$, we also require that $\frac{1}{2} g_0 \sqrt{Q_0 b\eta} \geq b\dot{\chi} + \sqrt{\chi_0 b\eta y_{\mathrm{RT}}}$. Then the two roots of equation (8.10) are positive and we keep to the smaller one which induces larger demand volume hence greater overall surplus: the solution is then

$$t^{\mathrm{S2}} = \frac{\frac{1}{2} g_0 \sqrt{Q_0 b\eta} - b\dot{\chi}}{\chi_0} - \sqrt{\left(\frac{\frac{1}{2} g_0 \sqrt{Q_0 b\eta} - b\dot{\chi}}{\chi_0}\right)^2 - \frac{b\eta\, y_{\mathrm{RT}}}{\chi_0}}, \text{ if } \chi_0 > 0, \tag{8.11a}$$

$$\text{otherwise } t^{\mathrm{S2}} = \frac{b\eta\, y_{\mathrm{RT}}}{g_0 \sqrt{Q_0 b\eta} - 2b\dot{\chi}} \text{ if } \chi_0 = 0, \tag{8.11a'}$$

$$Q^{\mathrm{S2}} = \frac{b\eta}{(t^{\mathrm{S2}})^2}, \tag{8.11b}$$

$$g^{\mathrm{S2}} = g_0 \sqrt{\frac{Q_0}{b\eta}}\, t^{\mathrm{S2}}, \tag{8.11c}$$

$$\tau^{\mathrm{S2}} + \alpha \tilde{t}_{\mathrm{RT}} = g^{\mathrm{S2}} - \beta t^{\mathrm{S2}} = \left(g_0 \sqrt{\frac{Q_0}{b\eta}} - \beta\right) t^{\mathrm{S2}}. \tag{8.11d}$$

We can bring the values for access time, generalized cost and price in relation to those in the monopoly optimum profit case. Obviously, if $\chi_0 = 0$,

$$\frac{g^{\mathrm{S2}}}{g^{\mathrm{MO}}} = \frac{t^{\mathrm{S2}}}{t^{\mathrm{MO}}} = \frac{\tau^{\mathrm{S2}} + \alpha \tilde{t}_{\mathrm{RT}}}{\tau^{\mathrm{MO}} + \alpha \tilde{t}_{\mathrm{RT}}} = \frac{\frac{1}{2} - x}{1 - 2x} = \frac{1}{2}.$$

Thus, the generalized cost and access time in the second-best optimum are half those of their equivalent for the monopolist. The tariff $\tau^{\mathrm{S2}} = \frac{1}{2}(\tau^{\mathrm{MO}} - \alpha \tilde{t}_{\mathrm{RT}})$ is even smaller. As for the volumes, $Q^{\mathrm{S2}} = 4Q^{\mathrm{MO}}$, which shows that the absence of service regulation reduces demand by a factor of 4.

8.4 Special case with demand elasticity of -1 to generalized cost

When $\varepsilon = -1$, the dual fixed-point problem is also a second-degree equation in t:

$g_0 \frac{Q_0}{b\eta} t^2 = y_{\mathrm{RT}} + 2(c_u^{+} + \beta) t + \frac{\chi_0}{b\eta} t^2$, which is equivalent to

$$(g_0 Q_0 - \chi_0) t^2 - 2b\dot{\chi} t - b\eta y_{\mathrm{RT}} = 0. \tag{8.12}$$

For a positive solution to exist, we must have that $g_0 Q_0 > \chi_0$. The existence of solutions also requires that $b\dot{\chi}^2 \geq (g_0 Q_0 - \chi_0)\eta y_{\mathrm{RT}}$, which holds true if $g_0 Q_0 > \chi_0$. The solution is then

$$t^{\mathrm{S2}} = \frac{b\dot{\chi}}{g_0 Q_0 - \chi_0} + \sqrt{\left(\frac{b\dot{\chi}}{g_0 Q_0 - \chi_0}\right)^2 + \frac{b\eta y_{\mathrm{RT}}}{g_0 Q_0 - \chi_0}}, \tag{8.13a}$$

$$Q^{\mathrm{S2}} = \frac{b\eta}{(t^{\mathrm{S2}})^2}, \tag{8.13b}$$

$$g^{\mathrm{S2}} = \frac{g_0 Q_0}{b\eta} (t^{\mathrm{S2}})^2, \tag{8.13c}$$

$$\tau^{S2} + \alpha\tilde{t}_{RT} = g^{S2} - \beta t^{S2} = (\frac{g_0 Q_0}{b\eta} t^{S2} - \beta) t^{S2}. \tag{8.13d}$$

We can relate these values to those for the first-best system optimum SO in (7.14): in fact, the S2 access time follows the same formula as the solution to the SO equation except for the denominator, $2g_0Q_0$ or g_0Q_0 as opposed to $g_0Q_0 - \chi_0$ in (8.13a). Thus t^{S2}, being constituted of larger terms than its SO counterpart t^{SO}, is greater than it.

Illustration. In Figure 7, the sensitivity of the mean access time under S2 policy is analyzed with respect to the fixed production cost χ_0. All parameter values are set up as in Figure 5. Two different demand curves are modeled, first for $\varepsilon = -2$ and second for $\varepsilon = -1$. In both cases the resulting access times are fairly close. Both functions increase with χ_0 in a smooth way that is quite slow. This suggests that the S2 policy is a robust one, which makes it all the more valuable as a fair compromise between customers' well-being and supply profitability.

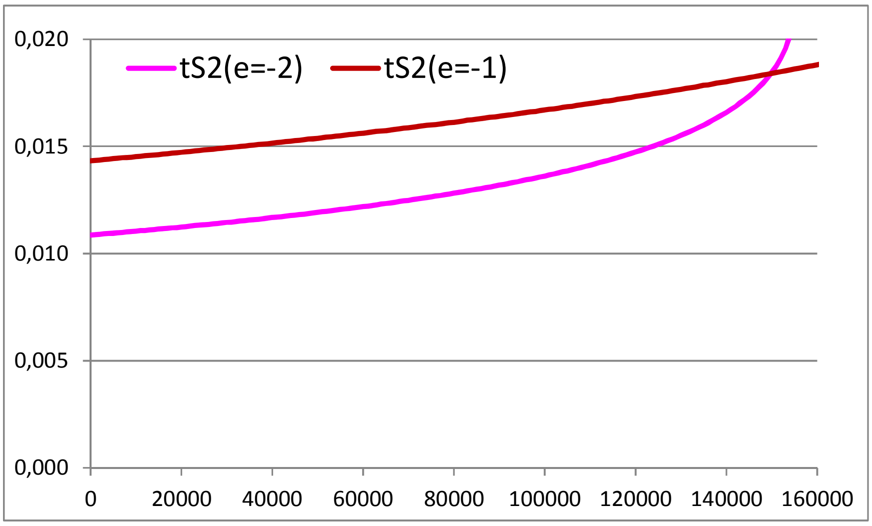

Fig. 7. Sensitivity analysis of Access time with respect to fixed production cost χ_0, under S2 policy. Two special cases of demand function are considered, with elasticity of either -2 or -1.

9. Prospective Application

Let us apply the model in a prospective way, in a classroom instance which is at best a sketch study but absolutely not a consultant's report.

9.1 Technological issues

Having laid the emphasis on regulation patterns and the associated service management, let us finally come to the technological issues. These pertain to:

i) the generalization of connectivity for all mobile entities, taxis as well as customers (with their smartphone) around a service platform that centralizes system management and eases the transactions: transaction efforts have drastically fallen down, from say 10 min to 2 min for t_T as well as for $\tilde{t}_T$.

ii) Elastic Vehicles (EVs) for taxi fleets: under typical French conditions (e.g. Leurent and Windisch, 2015) this enables to cut the per vehicle time-based variable cost

c_u from say 5€/h for conventional cars down to 1€/h. Concerning the fixed cost per taxi $\dot{\chi}$, while electric vehicles have higher selling prices, their technical life is longer; the combined effect may well be a decrease in the daily cost. In our numerical application we allow for a 5% decrease.

iii) Autonomous Driving (AD) technologies are developing. As for taxis, full automation at Level 5 is required in order to save on driving costs and reset the business model drastically. We model the AD option by reducing the daily fixed cost per taxi $\dot{\chi}$ from say €400 to €100, including vehicle leasing, insurance and maintenance, cleaning services.

9.2 Scenario design

Technological developments take place in a cumulative way. Nowadays, platform technology is implemented almost everywhere. The dissemination of EVs in taxi fleets in under way – still far from complete. AD at Level 5 is yet to come. So we identify four technological generations, numbered from 0 to 3:

- Generation 0, Pre-platform, is now obsolete but provides a reference for comparison.
- Generation 1 involves Platform only.
- Generation 2 combines Platform and EVs.
- Generation 3 involves Platform plus EVs plus AD.

Then, each technological generation can be combined to a regulation pattern, MO versus SO or S2, giving rise to 12 scenarios.

Table 1 provides the parameter values assigned to each technological generation.

Parameter	0/ Pre-platform	1/ Platform	2/ PF+EVs	3/ PF+EVs+AD
t_R min	24	24	24	24
t_T min	10	2	2	1
$\dot{\chi}$ €/day	400	400	380	100
c_u €/h	5	5	1	1

Tab.1: Parameter set-up according to technological generation.

9.3 Simulation outcomes

We applied the model to each scenario combining a technological generation with a regulation pattern. Demand function with constant elasticity -2 is considered from reference point at $Q_0 =$ 50,000 trips/day for $g_0 = 20$€ per trip. Average user's values of times are set to $\alpha = \beta = 25$€/h in order to reflect a population of taxi users. The fixed service cost is omitted (i.e. $\chi_0 = 0$) since the outcomes have reduced sensitivity to it (cf. Figure 7). The simulation outcomes are given in Table 2 in terms of service performance (quantity of usage, access time), production factors (fleet size, taxi daily productivity), price (tariff fare), level of service indicator (generalized cost), along with economic indicators of fare revenues, supply cost and profit, demand surplus and collective profit.

A couple of comments are in order here:

- From Pre-platform under Monopoly to PF+EV+AD under SO or S2, the demand volume is multiplied by 25, access time divided by 8, fleet size multiplied by 16, tariff fares divided by 13: such ratios represent one or two orders of magnitude for service development.
- A ratio of about 25 is rooted in two changes: from MO to S2 there is a factor of about 4 (a theoretical property for demand elasticity of -2), while from Pre-platform to PF+EV+AD the factor is about 6.
- Technology alone will not reduce the fare so drastically: under the next technological generation, a monopolistic service would still charge a relatively high fare (€17 per trip) whereas SO and S2 would reduce the trip fare to €3 or 4.
- Technological development and regulation are much beneficial to the demand side, and also to the supply side in terms of business activity, fleet size, commercial revenues. Production profit, however, is positive only in the Monopolistic configuration.

TECHNOLOGY	Pre-platform			Platform			PF+EV			PF+EV+AD		
REGULATION	MO	SO	S2	MO	SO	S2	MO	SO	S2	MO	SO	S2
Trips per day	947	4 327	3 787	1 571	7 181	6 283	1 926	8 712	7 703	6 200	26 001	24 801
Access time (h)	0,075	0,035	0,038	0,058	0,027	0,029	0,054	0,025	0,027	0,020	0,010	0,010
Gen trip cost	72,68	33,99	36,34	56,42	26,39	28,21	50,95	23,96	25,48	28,40	13,87	14,20
Fare	56,81	19,13	21,41	44,31	15,05	16,83	38,95	12,67	14,15	17,67	3,39	3,72
Fleet size	53,3	206,1	182,3	67,8	261,3	231,2	80,6	311,8	278,3	230,1	860,2	822,8
trips/day.veh	17,7	21,0	20,8	23,2	27,5	27,2	23,9	27,9	27,7	26,9	30,2	30,1
Fare revenues	53 779	82 771	81 085	69 598	108 096	105 740	75 016	110 391	109 016	109 587	88 259	92 287
Supply Cost	23 552	91 690	81 078	30 661	119 583	105 728	31 477	122 138	109 003	25 567	96 479	92 260
Supply Profit	30 226	-8 919	7	38 937	-11 487	12	43 539	-11 747	13	84 020	-8 220	27
Demand surplus	68 798	147 095	137 595	88 623	189 484	177 247	98 127	208 712	196 255	176 071	360 561	352 142
System Profit	99 024	138 176	137 603	127 561	177 997	177 259	141 666	196 965	196 268	260 091	352 341	352 168

Tab.2: Simulation outcomes for technological and regulation mixes.

10. Conclusion

The Orbicity model of Taxi supply, demand and regulation has a four-layer architecture that is exhibited in Fig. 8. On the first layer, the average access time t is modeled as an "Availability function" of fleet size N and demand volume Q, conditionally to running time parameter b specific to the city. We inverted the availability function into a relation $\hat{N}(t,Q)=\frac{1}{H}Q(t_{\mathrm{RT}}+t)+b/t$, which is a fleet sizing rule on the basis of target usage volume and access time.

On the second layer, demand volume Q is modeled as a function D according to tariff fare τ and time t of access, t_{R} of riding and t_{T} of transaction. Thus, both Q and t are jointly determined as implicit functions of N and τ in short-term demand-supply equilibrium.

On the third layer, given a regulation pattern either Monopoly or System optimum or Second best optimum, the service supplier determines N and τ so as to meet a specific management objective. Pattern-specific optimality conditions constitute characteristic equations with respect to τ and t, from which stem the other state variables. Whatever the pattern, a relation $Q.t^2=\dot{\chi}b/(c_u^{\dagger}+\beta)$ links Q and t to city run parameter b, the per vehicle fixed cost of service provision $\dot{\chi}$, the variable supply cost per unit of trip time $c_u^{\dagger}$ and the user's monetary value of access time β.

On the fourth layer, the service supplier takes strategic decisions on vehicle technology (in terms of engine type and driving mode) and on territorial implantation (ring identification), while the public authorities decide on the regulation pattern.

For a demand function with constant elasticity, additional theoretical properties were obtained. In particular, when the volume of demand has an elasticity to generalized cost of -2, then monopoly operation results in an optimum price more than twice that of the second-best optimum, an access time that is also more than double, and demand that is lower by a factor of more than four. The ratios between monopoly and first-best collective optimum are even higher. System optimization drives expansion in supply and demand, making it possible to reduce both tariffs and access times, to the greater benefit of users.

The ring postulate was instrumental to establish the availability function on the basis of specific territorial conditions: circle radius and average running speed. The composite parameter b can be taken as a standalone exogenous parameter to depict local conditions in less specific urban territories. Under this enlarged interpretation, the availability function will still hold as an approximation of Little's law in queuing theory.

The theoretical model may be further developed in a number of research directions: more complex pricing schemes including a distance-based component, demand segmentation with respect to value of time, temporal variations along the day, service hybridization between passenger taxi and freight delivery (cf. the Uber-Eats service), spatial composition involving several kinds of circles, less uniform spatial configuration involving local peaks of demand along the service ring, more complex regulation schemes involving price caps, fleet size limitations etc., service competition and its regulation, deeper composition of the collective welfare function to emphasize environmental and social stakes, etc.

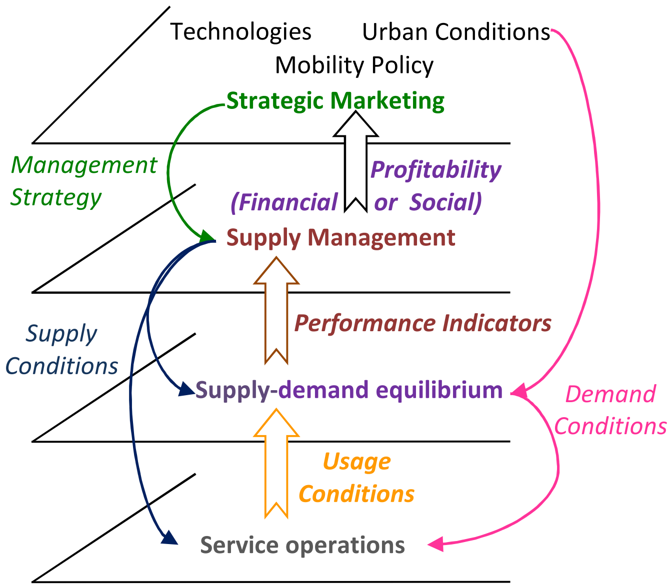

Fig. 8. Architecture of the Orbicity Taxi Service Model.

Acknowledgment. The author is indebted to Mr. Jean Grébert of the Renault Group for stimulating discussions on car-based shared mobility services, and for putting forward the image of a noria-like service. He is also grateful to Mr. Jaafar Berrada for careful reading, useful comments and an introduction to the academic literature.

11. References

Andreasson, I., Leurent, F., Corman, F. and dell'Olio, L. (2016) Modelling the Diversity and Integration of Transit Modes. In Gentile, G. & Noeckel, K. (eds): *Modelling Public Transport Passenger Flows in the Era of Intelligent Transport Systems*, Springer Tracts on Transportation and Traffic, Vol. 10: 485-520.

Arnott R. (1996) Taxi travel should be subsidized. *Journal of Urban Economics* 40, 316-333.

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

Boutueil V., Quillerier T. and Voskoboynikova A. (2018) The benefits and pitfalls of deregulating taxi markets. Can contrasted case studies help inform the debate? Paper presented to the 2019 TRB Annual meeting, forthcoming in *Transportation Research Record*.

Conolly, B. W., Parthasarathyb, P. R., Selvarajub, N. (2002) Double-ended queues with impatience. *Computers and Operations Research* 29: 2053–2072.

De Vany A.S. (1975). Capacity Utilization Under Alternative Regulatory Restraints: An Analysis of Taxi Markets. *Journal of Political Economy* 83(1): 83-94.

Dial R. (1995) Autonomous dial-a-ride transit introductory overview. *Transportation Research Part C* 5: 261–275.

Douglas G. W. (1972) Price regulation and optimal service standards: The taxicab industry, . *Journal of Transport Economics and Policy*, 116-127.

Flores-Guri, D. (2003) An economic analysis of regulated taxi markets. *Review of industrial organization* 23: 255-266.

Kashyap B.R.K. (1966) The double-ended queue with bulk service and limited waiting space, *Operations Research* 14: 822–834.

Kleinrock L (1975) *Queueing systems*. Volume I: Theory. Wiley, New York.

Leurent F., Windisch E. (2015) Benefits and costs of electric vehicles for the public finances: an integrated valuation model based on input-output analysis, with application to France. *Research in Transportation Economics* 415:51-62.

Lioris, E. (2010) *Évaluation et optimisation de systèmes de taxis collectifs en simulation. Modélisation et simulation*. PhD manuscript at Ecole des Ponts ParisTech. <pastel00565617>

Malucelli F., Nonato M., and Pallottino S. (1999) Demand adaptive systems : some proposals on flexible transit. *Operational Research in Industry* (T.A. Ciriani et al., ed.), McMillan Press, London.

Manski C.F. and Wright J.D. (1976) Nature of Equilibrium in the Market for Taxi Services. *Transportation Research Record* 619, 11-15.

Maupu, J.L. (2006) *La ville creuse pour un nouvel urbanisme durable : nouvel agencement des circulations et des rues*. Lharmattan, Paris.

Ortuzar J. de Dios and Willumsen L. (2011) *Modeling Transport*, 4th edition. Wiley.

Varian, H.R. (1992) *Microeconomic Analysis*, 3rd ed. W.W. Norton & Company, New York.

Wong, K.I., Wong, S.C. and Yang, H. (2001) Modeling urban taxi services in congested road networks with elastic demand. *Transportation Research Part B* 35 (9), 819–842.

Wong, K.I., Wong, S.C., Bell, M.G.H. and Yang, H. (2005) Modeling the bilateral micro-searching behavior for urban taxi services using the absorbing Markov chain approach. *Journal of Advanced Transportation* 39 (1), 81–104.

Wong, K.I., Wong, S.C., Yang, H. and Wu, J.H. (2008) Modeling urban taxi services with multiple user classes and vehicle types. *Transportation Research Part B* 42(10), 985–1007.

Yang H., Leung C.W.Y, Wong S.C. and Bell M.G.H. (2010) Equilibria of bilateral taxi–customer searching and meeting on networks. *Transportation Research Part B* 44: 1067–1083.

Yang, H. and Wong, S.C. (1998) A network model of urban taxi services. *Transportation Research Part B* 32 (4), 235–246.

Yang, H., Wong, S.C. and Wong, K.I. (2002) Demand-supply equilibrium of taxi services in a network under competition and regulation. *Transportation Research Part B* 36 (9), 799–819.

Yang, H., Ye, M., Tang, W.H. and Wong, S.C. (2005a.) A multi-period dynamic model of taxi services with endogenous service intensity. *Operations Research* 53 (3), 501–515.

Yang, H., Ye, M., Tang, W.H. and Wong, S.C., (2005b) Regulating taxi services in the presence of congestion externalities. *Transportation Research Part A* 39 (1), 17–40.

Yang W. H., Wong R.C.P. and Szeto W.Y. (2018) Modeling the acceptance of taxi owners and drivers to operate premium electric taxis: Policy insights into improving taxi service quality and reducing air pollution. *Transportation Research Part A*: 118: 581-593.

Appendix A: On Service Operations and Traffic Equilibrium

A.1 On the mean number of busy vehicles

Proof of Proposition 1. Point (i) stems from the decomposition of $\bar{n}$ in simple, continuous functions. Denoting $\sqrt{\bullet} \equiv \sqrt{X^2 - \lambda b}$, we get that $\partial\bar{n}/\partial N = \frac{1}{2} - X/\sqrt{\bullet} = (\sqrt{\bullet} - X)/2$ which is less than zero, hence (ii). We also obtain that $\partial\bar{n}/\partial\lambda = \frac{1}{2}t_{\mathrm{RT}} + (b + X.t_{\mathrm{RT}})/\sqrt{\bullet}$. Thus $\partial\bar{n}/\partial\lambda \geq 0$. Furthermore,

$$\begin{aligned}
\frac{\partial^2\bar{n}}{\partial\lambda^2} &= -\frac{t_{\mathrm{RT}}^2}{\sqrt{\bullet}} - \tfrac{1}{4}\sqrt{\bullet}^{-3}(b + X\,t_{\mathrm{RT}})(2X\frac{\partial X}{\partial\lambda} - b) \\
&= \tfrac{1}{4}\sqrt{\bullet}^{-3}((b + X\,t_{\mathrm{RT}})^2 - t_{\mathrm{RT}}^2(X^2 - b\lambda)) \\
&= \tfrac{1}{4}\sqrt{\bullet}^{-3}(b\lambda t_{\mathrm{RT}}^2 + b(b + 2X\,t_{\mathrm{RT}})) \;\; \text{which is positive}
\end{aligned}$$

This makes $\bar{n}$ a convex function of λ, yielding (iii), therefore also of demand volume Q.

A.2 On the mean vehicle access time

Proof of Proposition 2. Point (i) comes from relation $\bar{t}_{\mathrm{A}}^{\mathrm{V}} = b/(N + 1 - \bar{n})$ since $b > 0$ and $N > \bar{n}$.

(ii) As $\partial\bar{t}_{\mathrm{A}}^{\mathrm{V}}/\partial N = -b(N + 1 - \bar{n})^{-2}(1 - \partial\bar{n}/\partial N)$, it is negative, making $\bar{t}_{\mathrm{A}}^{\mathrm{V}}$ a decreasing function of N. With respect to b, we have that

$\frac{\partial \bar{t}_{\mathrm{A}}^{\mathrm{V}}}{\partial b} = \frac{1}{N+1-\bar{n}} + \frac{b}{(N+1-\bar{n})^2}\frac{\partial \bar{n}}{\partial b}$ which is positive: thus $\bar{t}_{\mathrm{A}}^{\mathrm{V}}$ is an increasing function of b.

(iii) With respect to λ, $\frac{\partial \bar{t}_{\mathrm{A}}^{\mathrm{V}}}{\partial \lambda} = \frac{b}{(N+1-\bar{n})^2}\frac{\partial \bar{n}}{\partial \lambda}$ which is positive. Indeed, $\bar{t}_{\mathrm{A}}^{\mathrm{V}}$ is a function chaining two positive increasing functions $\lambda \mapsto \bar{n}$ and $\bar{n} \mapsto t$, so it increases with λ. As for convexity, from Prop.1-(iii) function $\lambda \mapsto \bar{n}$ is increasing and convex, whereas function $\bar{n} \mapsto t$ is not only increasing but also convex since $\partial^2 t / \partial \bar{n}^2 = 2b(N+1-\bar{n})^{-3} \geq 0$. Thus function $\lambda \mapsto \frac{\partial \bar{t}}{\partial \lambda} = \frac{\partial \bar{t}}{\partial \bar{n}} . \frac{\partial \bar{n}}{\partial \lambda}$, being the product of two positive increasing functions, is positive and increasing, which makes $\bar{t}_{\mathrm{A}}^{\mathrm{V}}$ a convex function of λ.

A.3 On the fleet sizing function

<u>Proof of Proposition 3</u>. (i) As t decreases with N, then the inverse function $\hat{N}$ decreases with t. As $\partial \hat{N} / \partial t = Q / H - b / t^2$, for $\partial \hat{N} / \partial t \leq 0$ to hold it is required that $Q \leq H\, b t^{-2}$.

(ii) Given t, it is easy to compute

$\frac{\partial \hat{N}}{\partial Q} = \frac{t_{\mathrm{ART}}}{H} \geq 0$ hence $\hat{N}$ increases with Q.

$\frac{\partial \hat{N}}{\partial t_{\mathrm{R}}} = \frac{\partial \hat{N}}{\partial t_{\mathrm{T}}} = \frac{Q}{H} \geq 0$ hence $\hat{N}$ increases with t_{R} and t_{T}.

$\frac{\partial \hat{N}}{\partial b} = \frac{1}{t} \geq 0$ hence $\hat{N}$ increases with b.

A.4 Existence and uniqueness of traffic equilibrium

<u>Proof of Proposition 5</u>. (i) The left side of the fixed-point problem (5.4) decreases with Q while the right side increases. There is therefore at most one solution.

(ii) Function $\mathrm{T}_{\mathrm{A}}^{\mathrm{V}}$ increases continuously from $b/(N+1)$ to b when Q increases from 0 to NH / t_{RT}. Then function $\mathrm{f}(Q) \equiv \mathrm{D}_t^{(-1)}(\tau, Q) - \mathrm{T}_{\mathrm{A}}^{\mathrm{V}}(N, t_{\mathrm{T}}, t_{\mathrm{R}}, Q)$, being the sum of two smooth functions that decrease with Q, is also smooth and decreasing. As $0 \leq \mathrm{D}(\tau, b/(N+1))$ and $\mathrm{D}_t^{(-1)}$ is decreasing, then $\mathrm{D}_t^{(-1)}(\tau, 0^+) \geq b/(N+1)$ i.e. $\mathrm{f}(0^+) \geq 0$. Furthermore, under condition $\mathrm{D}(\tau, b) < NH / b$, then $b \geq \mathrm{D}_t^{(-1)}(\tau, NH / b)$ i.e. $\mathrm{f}(\frac{1}{b} NH) \geq 0$. So on interval $[0, \frac{1}{b} NH]$ function f is smooth and its sign changes from negative to positive; by the Bolzano-Weierstrass theorem, there is a point $Q*$ in the interval at which $\mathrm{f}(Q*) = 0$: this constitutes a supply-demand equilibrium.

A.5 Sensitivities of traffic equilibrium

<u>Proof of Proposition 6</u>. (i) an increase in N lowers the value of $\mathrm{T}_{\mathrm{A}}^{\mathrm{V}}$, which increases $Q*$ as well as $t*$.

(ii) As for τ, between τ_1 and $\tau_2 > \tau_1$, the demand function $t \mapsto \mathrm{D}(\tau_2, t)$ is lower than $t \mapsto \mathrm{D}(\tau_1, t)$, and therefore in turn the reciprocal function $Q \mapsto \mathrm{D}_t^{(-1)}(\tau_2, t)$ is lower than

$Q \mapsto \mathrm{D}_t^{(-1)}(\tau_1, t)$. So, the fixed point Q_2^* is lower than Q_1^*. Thus the price has the effect of lowering the equilibrium volume Q^*, and also the access time t^* since $\mathrm{T}_\mathrm{A}^\mathrm{V}$ increases with Q. Then, an increase in the transaction time on the production side, t_T, raises $\mathrm{T}_\mathrm{A}^\mathrm{V}$, which reduces Q^* and t^*. The transaction time on the customer side, $\tilde{t}_\mathrm{T}$, lowers D and therefore also $\mathrm{D}_t^{(-1)}$, which also lowers Q^* and t^*. Therefore a combined reduction of the two transaction times produces a twofold increase in Q^* as well as in t^*.

Lastly, a variation in the riding time t_R has an effect similar to a joint variation, in the same direction, in t_T and $\tilde{t}_\mathrm{T}$: an increase in t_R leads to a reduction in both Q^* and t^*.

Appendix B: a Markov model for traffic in a taxi service

Abstract. We consider a taxi service operating in an urban environment that is spatially homogeneous both in travel demand and in motoring conditions. The postulate of uniformity allows us to model the service as a Markov chain depending on the number of vehicles vacant at a given moment, if this is positive, or otherwise on the number of customers waiting. We combine the two cases into a single state variable in the set of integer differences, which amounts to a double ended queue, which is conventional in the analysis of a taxi rank. We model the rates of transition specific to our system and to its spatial extension. From this, we deduce the stationary regime law in stochastic equilibrium, and from this we establish the conditions of existence as a function of the parameters of the system.

In stationary regime, we obtain the distribution of customer waiting time and the distribution of the number of empty vehicles. We characterize the mode of the number of vacant vehicles and we show that it is equivalent to the mean arising from a formulation similar to Little's law. Finally, we show that the number of vacant vehicles is very close to a random Poisson variable.

Keywords: Taxi Service; Homogeneous City; Markov Chain; Bilateral Queue; Stochastic Equilibrium; Waiting Time; Empty Vehicles

Introduction. This appendix presents the stochastic model of the dynamic operation of the supply and demand system for a taxi service in an Orbicity. We model this system by a Markov model with states and transitions, in which the state variable is reduced to an integer difference: in the positive range, the variable models the number of empty taxis, whereas in the negative range, the variable models the number of customers waiting to be allocated a taxi (with only a difference in sign). We begin by describing the states and transitions in this Markov model (§ B.1). Then we established the probability distribution for the system in stationary regime (§ B.2). For the dynamic stationary regime, we can therefore calculate the mean waiting time for a customer until a vehicle is allocated (§ B.3).

We then model the time of access to the service, including waiting time before a taxi is allocated and the time required for the taxi to reach the customer (§ B.4). We can then characterize the availability function that links the mean access time to the technical characteristics of the supply and to the spatiotemporal characteristics of the demand (§ B.5). To facilitate the calculations, we posit additional simplifying hypotheses that constitute the "approximate model" (§ B.6). We study the mode (§ B.7) and dispersion (§ B.8) of the stationary distribution, showing that it can be approximated numerically by a Normal distribution. In addition, a Poisson distribution offers an excellent approximation (§ B.9). We conclude by recapitulating the simplifying assumptions in the model (§ B.10).

B.1 A state-transition model for the dynamics of the service

The customer's access time, $\tilde{t}_{\mathrm{A}}$, encompasses the time between the vehicle being allocated to and reaching the customer, denoted by $t_{\mathrm{A}}^{\mathrm{V}}$, and the time the customer waits before being assigned the vehicle, devoted by $t_{\mathrm{A}}^{\mathrm{W}}$. These two terms depend on the number of vehicles empty in real time, i.e. k^{+}.

If $k^{+} > 0$, then vehicle allocation is immediate, i.e.

$$k^{+} > 0 \Rightarrow \; t_{\mathrm{A}}^{\mathrm{W}} = 0 . \tag{B.1}$$

If there are no vacant vehicles, the customer must wait for a currently occupied vehicle to be allocated. This waiting time depends on the current occupancies of the different taxis, but also on the number of customers waiting, let us say k^{-}.

Obviously, states $k^{+} > 0$ and $k^{-} > 0$ are incompatible:

$$k^{+} > 0 \Rightarrow \; k^{-} = 0 , \tag{B.2a}$$

$$k^{-} > 0 \Rightarrow \; k^{+} = 0 . \tag{B.2b}$$

We can therefore model the current state of the system by a main variable, denoted by k, which links the number of taxis empty and the number of customers waiting as follows:

$$k \equiv k^{+} - k^{-} . \tag{B.3}$$

On its own, this variable does not account for the level of occupancy of the vehicles (i.e. the additional time needed to service waiting customers), but only for their current state of availability.

We model the occupancy level indirectly, via the system's transition rates between the different states constituted by the values of k. Provisionally, we denote μ_k as the rate of transition from state k to state $k+1$ and λ_{k+1} as the rate of transition from state $k+1$ to state k. During the dynamic changes to the system, there is no event that results in a shift from a current state k to any state other than its immediate neighbors $k+1$ or $k-1$.

Thus the states and transitions constitute a Markov chain illustrated in figure B1.

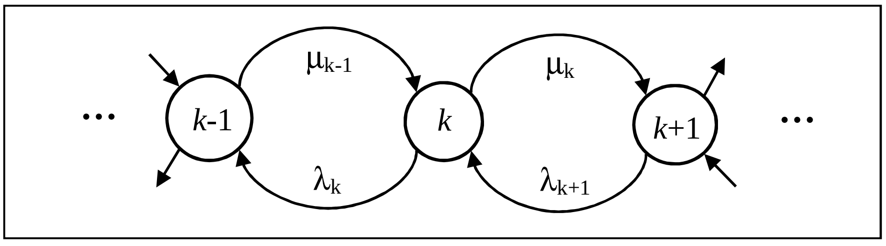

Fig. B1. State transition diagram.

The state of the system is bounded above by the value N, whereas for the negative values there is no lower bound (i.e. there is no upper bound for k^{-}, which is an approximation since k^{-} is in fact limited by population size).

Let us look at the rates of transition between states, recalling that the transition rate between states m and n is defined by $\xi_{m,n} = \lim_{\delta h \to 0^{+}} \delta \Pr\{k_{h+\delta h} = n \,|\, k_h = m\} / \delta h$.

For positive values of k, a transition from $k+1$ to $k \geq 0$ corresponds to the allocation of a vacant vehicle to a customer: as this vehicle is vacant and there is no other customer waiting, the assignment is immediate and the rate of transition depends only on the arrivals of customers in the system. Assuming that customers arrive independently according to time intensity λ, then precisely

$$\lambda_{k+1} \equiv \xi_{k+1,k} = \lambda \text{ for } k \geq 0\,. \tag{B.4a}$$

Dually, a transition from k to $k+1$ corresponds to the end of a taxi ride. We now model the service time for each taxi, including t_{T}, $t_{\mathrm{A}}^{\mathrm{V}}$ and t_{R}, as an exponential variable, with a parameter $\mu_k^{(1)}$ calibrated as follows. As the parameter of an exponential distribution is the inverse of its mean value, we know that

$$\frac{1}{\mu_k^{(1)}} = t_{\mathrm{T}} + \mathrm{E}[t_{\mathrm{A}}^{\mathrm{V}/k}] + \mathrm{E}[t_{\mathrm{R}}]\,.$$

As there are $N-k$ vehicles operating in state k^-, with enough independence between them, the rate of transition to state $k+1$ is the sum of their respective rates of service completion, hence

$$\mu_k \equiv \xi_{k,k+1} = \frac{N-k}{t_{\mathrm{T}} + \mathrm{E}[t_{\mathrm{A}}^{\mathrm{V}/k}] + \mathrm{E}[t_{\mathrm{R}}]} \quad \text{for} \quad k > 0\,. \tag{B.4b}$$

In fact $\mu_k^{(1)}$ is obtained by two approximations: not only the Markovian assumption of exponential distribution, but also that the access times until pickup, $t_{\mathrm{A}}^{\mathrm{V}/k}$, correspond to k vacant vehicles for all the busy vehicles, whereas in certain cases their current occupancy arises from a different state k'.

For negative values of k, the transition from $k+1$ to k also corresponds to the arrival of an additional customer in the system. We add here the hypothesis that the probability of a customer agreeing to wait is only r, such that

$$\lambda_{k+1} = \lambda.r \quad \text{for} \quad k < 0\,. \tag{B.4c}$$

The transition from k to $k+1$ also corresponds to the completion of a vehicle's service, so given that there are N busy vehicles

$$\mu_k = \frac{N}{t_{\mathrm{T}} + \mathrm{E}[t_{\mathrm{A}}^{\mathrm{V}/1}] + \mathrm{E}[t_{\mathrm{R}}]} \quad \text{for} \quad k \leq 0\,. \tag{B.4d}$$

The difference between (B.4d) and (B.4b) arises on the one hand from the number of occupied vehicles, which does not depend on k^- but only on k^+, which is zero if $k \leq 0$, and on the other hand from the time it takes for the taxi to reach its customer. Since this taxi is provisionally the only one empty, the mean for this time is $t_{\mathrm{A}}^{\mathrm{V}/1}$.

In this way, we were able to use transition rates to model vehicle occupancy and the arrival of customers in the system.

B.2 Stationary distribution of the system state

The system is in a stationary dynamic regime, in other words in stochastic equilibrium, if at a moment h the state k_h takes the values $k \in K$ for $K \equiv \{...,-n,-n+1,...,-1,0,1,2,...,N\}$ with respective probabilities p_k independent of h.

For a Markov chain with simple transitions, the stationary distribution $[p_k : k \in K]$ must satisfy local balance conditions for the probability flux between k and $k+1$:

$$p_k.\mu_k = p_{k+1}.\lambda_{k+1}. \tag{B.5}$$

In other words, the probability flux from state k to state $k+1$ equals the reverse flux from $k+1$ to k. From this, we deduce by recurrence that

$$p_k = p_0 \prod_{i=0}^{k-1} \frac{\mu_i}{\lambda_{i+1}} \quad \text{for} \quad k > 0, \tag{B.6a}$$

$$p_k = p_0 \prod_{i=0}^{-k-1} \frac{\lambda_{-i}}{\mu_{-i-1}} \quad \text{for} \quad k < 0. \tag{B.6b}$$

The normalization of the probability distribution (i.e. $\sum_{k \in K} p_k = 1$) then gives us p_0:

$$\frac{1}{p_0} = 1 + \sum_{k<0} \prod_{i=0}^{-k-1} \frac{\lambda_{-i}}{\mu_{-i-1}} + \sum_{k=1}^{N} \prod_{i=0}^{k-1} \frac{\mu_i}{\lambda_{i+1}}. \tag{B.7}$$

The negative part of the range is easy to calculate because the transition rates are uniform: as $\lambda_{-i} = \lambda$ and μ_{-i-1} is independent of i (cf. (B.4d)), by denoting

$$\rho \equiv \frac{\lambda_{-i}}{\mu_{-i-1}} = \frac{\lambda r (t_\mathrm{T} + \mathrm{E}[t_\mathrm{A}^{\mathrm{V/1}}] + \mathrm{E}[t_\mathrm{R}])}{N},$$

the relation (B.6b) can be interpreted as a geometric sequence:

$$p_{-k} = p_0 \rho^k \quad \text{for} \quad k \geq 0. \tag{A.8}$$

Example. Take a circular city with a radius R = 2km, population P = 5,000 people with 3 trips per person over the period H = 10h. Therefore customers arrive with an overall intensity of λ = 1,500 people per hour. We also set r = 80%. We assume that v = 20km/h therefore β = 6.37/h, and also that t_T = 1 min therefore ξ = 260. The derived parameter ρ equals 99%. The probability distribution is illustrated in figure B2: we observe that the probability is concentrated around the value $\hat{k}$ = 235, with a very high concentration. The discrete random variable is very close to a Gaussian continuous random variable.

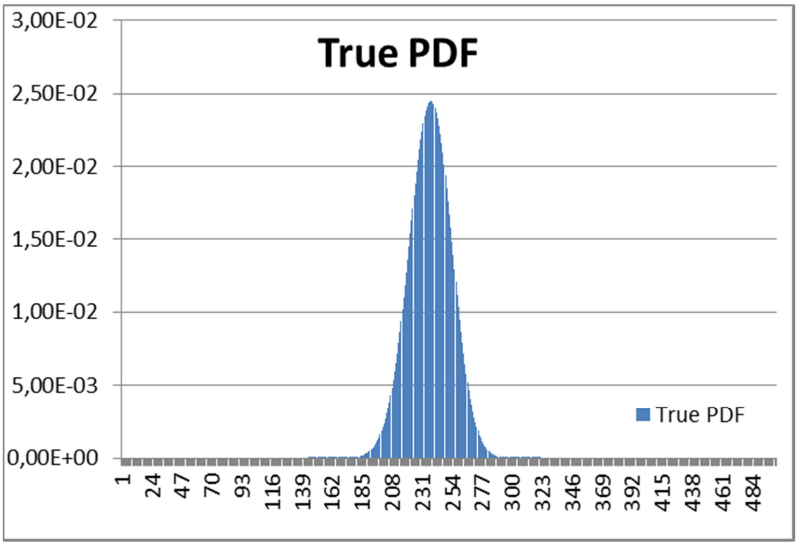

Fig. B2. Stationary distribution of the system's state variable.

B.3 Customer waiting time until allocation of a taxi

In stationary regime, the probability that all vehicles are occupied is

$$\Pr\{k \leq 0\} = \frac{p_0}{1-\rho}. \tag{B.9}$$

We can also obtain the mean customer waiting time until allocation of a vehicle by applying Little's law only to this phase: the mean number of customers waiting in the system equals the flow of arrivals (i.e. $\lambda(r.p_{k\leq 0} + p_{k>0})$) multiplied by the mean waiting time.

The mean number of waiting customers is defined by

$$\mathrm{E}[k^-] \equiv \sum_{k=0}^{+\infty} k.p_{-k} \ . \tag{B.10a}$$

Thanks to the properties of the geometric distribution,

$$p_0 \sum_{k=0}^{+\infty} k\rho^k = p_0\rho \sum_{k=1}^{+\infty} (k-1)\rho^k = p_0\rho \tfrac{\partial}{\partial\rho} \sum_{k=0}^{+\infty} \rho^k = p_0\rho \tfrac{\partial}{\partial\rho} \tfrac{1}{1-\rho} = p_0\rho/(1-\rho)^2 .$$

So
$$\mathrm{E}[k^-] = p_0 \frac{\rho}{(1-\rho)^2}. \tag{B.10b}$$

Thus the mean value for waiting time per customer before allocation of a taxi is:

$$\mathrm{E}[t_A^W] = \frac{\mathrm{E}[k^-]}{\lambda(rp_{k\leq 0} + p_{k>0})} = \frac{\rho}{(1-\rho)^2 \lambda(\frac{r}{1-\rho} + \sum_{k=1}^{N} \prod_{i=0}^{k-1} \frac{\mu_i}{\lambda_{i+1}})}. \tag{B.11}$$

Additionally, in the positive range, the mean number of empty vehicles is

$$\mathrm{E}[k \mid k > 0] = [\sum_{k=1}^{N} p_k k] \Big/ \sum_{k=1}^{N} p_k \ . \tag{B.12}$$

B.4 Vehicle access time

Between a taxi being assigned to a customer and the moment of pickup, the time spent $t_{\mathrm{A}}^{\mathrm{V}(k)}$ depends on the state of the system. We write $b \equiv (\pi \mathrm{R})/v$. For the negative part of the range, $k^{+} = 0$, therefore

$$\mathrm{E}[t_{\mathrm{A}}^{\mathrm{V}(k)} \mid k \leq 0] = \frac{b}{2}. \tag{B.13a}$$

For the positive part of the range,

$$\mathrm{E}[t_{\mathrm{A}}^{\mathrm{V}(k)} \mid k > 0] = \left[\sum_{k=1}^{N} p_k \frac{b}{k+1}\right] \Big/ \sum_{k=1}^{N} p_k \ . \tag{B.13b}$$

In all, by joining the negative and positive parts of the range,

$$\mathrm{E}[t_{\mathrm{A}}^{\mathrm{V}}] = \frac{b}{2}\left[\frac{p_0}{1-\rho} + 2\sum_{k=1}^{N} \frac{p_k}{k+1}\right]. \tag{B.13c}$$

Example (continued). Let us revisit the previous example. The probability of the negative states is negligible. For a customer requesting the service, allocation of a taxi is almost always immediate.

B.5 The exact availability function

In our Markovian analytical framework, we modeled on the one hand the mean waiting time per customer until allocation of a taxi, $t_{\mathrm{A}}^{\mathrm{W}}$, and on the other hand the mean time taken for the taxi to reach the customer, $t_{\mathrm{A}}^{\mathrm{V}}$.

The sum of these two durations constitutes the customer's mean access time to the service:

$$\mathrm{E}[\tilde{t}_{\mathrm{A}}] = \mathrm{E}[t_{\mathrm{A}}^{\mathrm{W}}] + \mathrm{E}[t_{\mathrm{A}}^{\mathrm{V}}]. \tag{B.14a}$$

Specifically,

$$\mathrm{E}[\tilde{t}_{\mathrm{A}}] = \frac{\rho}{(1-\rho)\lambda\left(r + (1-\rho)\sum_{k=1}^{N}\prod_{i=0}^{k-1}\frac{\mu_i}{\lambda_{i+1}}\right)} + \frac{b}{2}\,\frac{1 + 2(1-\rho)\sum_{k=1}^{N}\frac{1}{k+1}\prod_{i=0}^{k-1}\frac{\mu_i}{\lambda_{i+1}}}{1 + (1-\rho)\sum_{k=1}^{N}\prod_{i=0}^{k-1}\frac{\mu_i}{\lambda_{i+1}}}. \tag{B.14b}$$

This is effectively a function of the supply characteristics, $(\mathrm{R}, v, t_{\mathrm{T}})$, and of the demand characteristics, (r, λ). Figure B3 shows the linked sequence of influences in the model.

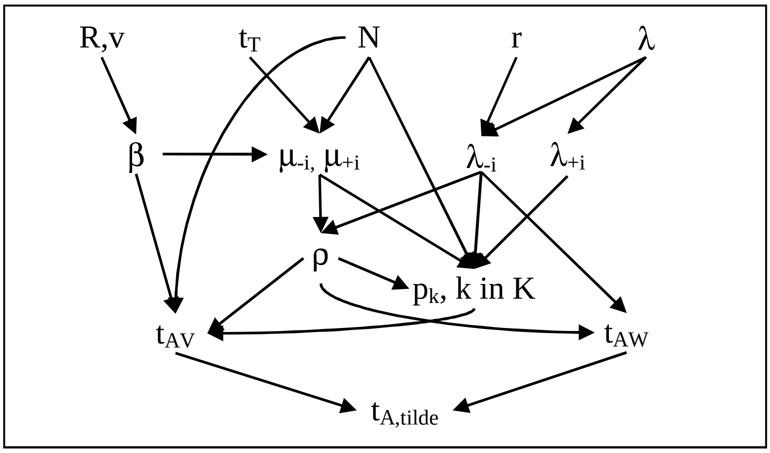

Fig. B3. Diagram of influences in the detailed model.

B.6 Approximate formulae

When the probability that all the vehicles are occupied is negligible, then the actual rate of arrival of customers is more or less equal to λ, since $\lambda(r.p_{k\le 0}+p_{k>0})\approx\lambda$.

This is also the rate at which customers are picked up by taxis, and therefore also the rate at which the taxis attain a busy state. The mean service time per ride is $t_{\mathrm{S}}=t_{\mathrm{T}}+\bar{t}_{\mathrm{R}}+\bar{t}_{\mathrm{A}}^{\mathrm{V}}$. The mean number of busy taxis satisfies Little's law: it is the rate of arrivals multiplied by the mean ride time, i.e.

$$\bar{n}=\lambda(t_T+\bar{t}_R+\bar{t}_{\mathrm{A}}^{\mathrm{V}}). \tag{B.15}$$

We model the mean access time by the following approximation, which applies the access time function $\bar{t}_{\mathrm{A}}^{\mathrm{V}(k^+)}$ to the mean number of vacant vehicles, $\bar{k}=N-\bar{n}$, in other words,

$$\bar{t}_{\mathrm{A}}^{\mathrm{V}}=\frac{b}{N+1-\bar{n}}. \tag{B.16}$$

In the example considered previously, the high concentration of k^+ around $\bar{k}$ fully justifies approximation of the mean. By combining the two previous relations, we obtain a characteristic equation for the mean number of busy taxis, $\bar{n}$: writing $t_{\mathrm{RT}}\equiv t_{\mathrm{T}}+\bar{t}_{\mathrm{R}}$, we have

$$\bar{n}=\lambda t_{\mathrm{RT}}+\frac{\lambda b}{N+1-\bar{n}}. \tag{B.17a}$$

This is a second degree equation, which is easy to solve by equivalence between successive conditions:

$$(N+1-\bar{n})(\bar{n}-\lambda t_{\mathrm{RT}})=\lambda b$$

$$\bar{n}^2-(N+1+\lambda t_{\mathrm{RT}})\bar{n}+\lambda t_{\mathrm{RT}}(N+1)+\lambda b=0$$

$$(\bar{n}-\tfrac{N+1+\lambda t_{\mathrm{RT}}}{2})^2=(\tfrac{N+1+\lambda t_{\mathrm{RT}}}{2})^2-\lambda t_{\mathrm{RT}}(N+1)-2\lambda/\beta=(\tfrac{N+1-\lambda t_{\mathrm{RT}}}{2})^2-\lambda b$$

$$\bar{n}=\tfrac{N+1+\lambda t_{\mathrm{RT}}}{2}\pm\sqrt{(\tfrac{N+1-\lambda t_{\mathrm{RT}}}{2})^2-\lambda b}$$

i.e.

$$\bar{n} = \frac{N+1+\lambda t_{\mathrm{RT}}}{2} \pm \frac{N+1-\lambda t_{\mathrm{RT}}}{2}\sqrt{1-\frac{4\lambda b}{(N+1-\lambda t_{\mathrm{RT}})^2}}\,. \tag{B.17b}$$

If the number $N+1-\lambda t_{\mathrm{RT}}$ is sufficiently large for $(N+1-\lambda t_{\mathrm{RT}})^2 >> 4\lambda b$, then we can approximate the expression under the root in (B.17b) by $1-2\lambda b/(N+1-\lambda t_{\mathrm{RT}})^2$, therefore

$$\bar{n} \approx \frac{N+1+\lambda t_{\mathrm{RT}}}{2} \pm \left(\frac{N+1-\lambda t_{\mathrm{RT}}}{2} - \frac{2\lambda/\beta}{N+1-\lambda t_{\mathrm{RT}}}\right), \tag{B.17c}$$

So again $\bar{n} \in \{N+1-\frac{\lambda b}{N+1-\lambda t_{\mathrm{RT}}},\ \lambda t_{\mathrm{RT}} + \frac{\lambda b}{N+1-\lambda t_{\mathrm{RT}}}\}$. The first solution is too large, so only the second one is suitable, which prompts us to reiterate the associated formulae.

$$\bar{n} = \frac{N+1+\lambda t_{\mathrm{RT}}}{2} - \frac{N+1-\lambda t_{\mathrm{RT}}}{2}\sqrt{1-\frac{4\lambda b}{(N+1-\lambda t_{\mathrm{RT}})^2}}\,. \tag{B.18a}$$

$$\bar{n} \approx \lambda t_{\mathrm{RT}} + \frac{\lambda b}{N+1-\lambda t_{\mathrm{RT}}}\,. \tag{B.18b}$$

Example (continued). The parameter values require that $\frac{1}{2}\lambda b =$ 235.62 and $\lambda t_{\mathrm{RT}} =$ 260.6, as well as a mean number of empty vehicles of $N-\bar{n}$ = 235.37 in the precise model. The characterization gives $\bar{n}$ = 262.58 and $N-\bar{n}$ = 237.42. The simplified approximation gives $\bar{n}$ = 262.60 and $N-\bar{n}$ = 237.40. The two approximations are therefore of very high quality, to within 1% of each other.

B.7 The mode of the state distribution

The mode of the distribution is the state that presents a maximum probability, which we denote by $\hat{k}$. Since the probabilities of the positive states grow in accordance with k up to the mode, and then decrease, we can look for the mode as the value that equalizes the ratio p_{k+1}/p_k at 1: the mode is therefore given by the equation

$$N-\hat{k} = \lambda\left(t_{\mathrm{RT}} + \frac{b}{\hat{k}+1}\right).$$

This condition is reformulated in accordance with $\hat{n} = N-\hat{k}$, which is a number of busy vehicles since $\hat{k}$ is a number of vacant vehicles. We then obtain the same equation as the one that characterizes the mean number of busy vehicles: this proves that if this mean equation is a good approximation, then $\hat{n}$ is the mean number of busy vehicles, in other words the mode and the mean coincide.

B.7 On the dispersion of the distribution of states

Our numerical experiment suggests a strong correlation of the probability of the states around the mean value. This phenomenon obviously arises from the formula of the ratio $r_k \equiv p_{k+1}/p_k$ among the positive states: low k have a high ratio, while as one approaches N that bounds the distribution, the ratio becomes very low. Moreover, between two distant states k_1 and $k_2 > k_1$, the probability ratio is the product of the ratios between consecutive states:

$$\frac{p_{k_2}}{p_{k_1}} = \prod_{k=k_1}^{k_2-1} r_k\,.$$

This implies that the states that are close to the mode of the distribution have a relatively high probability, whereas moving away from the mode, the probability quickly becomes negligible.

In order to quantify the dispersion of the distribution, we will characterize the states k according to the value of the ratio r_k, let us say α. The equation defining α is

$$N-k_\alpha=\alpha\lambda(t_{\mathrm{RT}}+\frac{b}{k_\alpha+1}).$$

In order to solve it, we simply reuse the mean equation, replacing λ with $\alpha.\lambda$ and therefore λt_{RT} with $\alpha\lambda t_{\mathrm{RT}}$. We then obtain the exact and the approximate solutions:

$$N-k_\alpha=\frac{N+1+\alpha\lambda t_{\mathrm{RT}}}{2}-\frac{N+1-\alpha\lambda t_{\mathrm{RT}}}{2}\sqrt{1-\frac{4\alpha\lambda b}{(N+1-\alpha\lambda t_{\mathrm{RT}})^2}}, \tag{B.19a}$$

$$N-k_\alpha\approx\alpha\lambda t_{\mathrm{RT}}+\frac{\alpha\lambda b}{N+1-\alpha\lambda t_{\mathrm{RT}}}. \tag{B.19b}$$

The states below the mode have $r_k>1$ and those above the mode, $r_k<1$. Through the interplay of increases and decreases:

$$k<k_{\alpha'}<\hat{k}\;\Rightarrow\; p_k<p_{k_{\alpha'}}/\alpha'^{k_{\alpha'}-k}=p_{k_{\alpha'}}(1/\alpha')^{k_{\alpha'}-k},$$

$$k>k_\alpha>\hat{k}\;\Rightarrow\; p_k<p_{k_\alpha}\alpha^{k-k_\alpha}.$$

With our notations, $\alpha'>1>\alpha$. This principle already enables us to increase the probability of the tails of the distribution: by the properties of the geometric laws

$$\Pr\{k<k_{\alpha'}\}\le\frac{p_{k_{\alpha'}}}{1-1/\alpha'}, \text{ and}$$

$$\Pr\{k>k_\alpha\}\le\frac{p_{k_\alpha}}{1-\alpha}.$$

For the $[\hat{k}_{\alpha'},\hat{k}_\alpha]$ part, we estimate the probabilities of the states by an approximation from the mode $\hat{k}$: letting $\xi\equiv\lambda t_{\mathrm{RT}}$,

$$p_k=p_{\hat{k}}\frac{(\hat{k}+1)(\hat{k}+2)}{(k+1)(k+2)}\prod_{i=0}^{k-\hat{k}-1}\frac{N-\hat{k}-i}{\xi+\frac{\lambda b}{1+\hat{k}+i}}\approx p_{\hat{k}}\prod_{i=0}^{k-\hat{k}-1}\frac{(\hat{n}-i)(1+\hat{k}+i)}{\xi(1+\hat{k}+i)+\lambda b}\quad\text{in the case where } k>\hat{k}$$

$$\approx p_{\hat{k}}(\frac{\hat{n}.(1+\hat{k})}{\lambda b+\xi(1+\hat{k})})^{\Delta k}\prod_{i=0}^{k-\hat{k}-1}(1-\frac{i}{\hat{n}})(1+\frac{i}{1+\hat{k}})(1+\frac{\xi.i}{\lambda b+\xi(1+\hat{k})})^{-1}\quad\text{denoting } \Delta k\equiv k-\hat{k}$$

$$\approx p_{\hat{k}}(\frac{\hat{n}.(1+\hat{k})}{\lambda b+\xi(1+\hat{k})})^{\Delta k}\exp\sum_{i=0}^{\Delta k-1}\ln(1-\frac{i}{\hat{n}})+\ln(1+\frac{i}{1+\hat{k}})-\ln(1+\frac{\xi.i}{\lambda b+\xi(1+\hat{k})})$$

As $\ln(1-x)\approx-x$ for low values of x, the sum in the exponential can be approximated by

$$(\sum_{i=0}^{\Delta k-1} i)(-\frac{1}{\hat{n}}+\frac{1}{1+\hat{k}}-\frac{\xi}{\lambda b+\xi(1+\hat{k})}) = \frac{1}{2}(\Delta k)(\Delta k-1)\frac{(\hat{n}-1-\hat{k})(\lambda b+\xi(1+\hat{k}))-\xi\hat{n}(1+\hat{k})}{\hat{n}(1+\hat{k})(\lambda b+\xi(1+\hat{k}))}$$

$$\approx \frac{(\Delta k)^2}{2}\frac{\lambda b(\hat{n}-1-\hat{k})-\xi(1+\hat{k})^2}{\hat{n}(1+\hat{k})(\lambda b+\xi(1+\hat{k}))}$$

$$\approx -\tfrac{1}{2}(\Delta k)^2\frac{1-\lambda b(\hat{n}-1-\hat{k})/(\xi(1+\hat{k})^2)}{\hat{n}(1+\lambda b/(\xi(1+\hat{k}))}$$

We suppose that $\xi(1+\hat{k})^2 > \lambda b(\hat{n}-1-\hat{k})$, or else similarly, since $\xi > \frac{1}{2}\lambda b$, that $(1+\hat{k})^2 > 2(\hat{n}-1-\hat{k})$, i.e. more or less that $\hat{k} > \sqrt{2(\hat{n}-1-\hat{k})}$. Under this condition,

$$p_k \approx p_{\hat{k}}(\frac{\hat{n}}{\xi+\lambda b/(1+\hat{k})})^{\Delta k}\exp(-\tfrac{1}{2\gamma}(\Delta k)^2) \quad \text{by denoting} \quad \gamma \equiv \frac{\hat{n}\,(1+\lambda b/(\xi(1+\hat{k}))}{1-\lambda b(\hat{n}-1-\hat{k})/(\xi(1+\hat{k})^2)} \approx \hat{n}$$

by the definition of the mode, $\frac{\hat{n}}{\xi+\lambda b/(1+\hat{k})} = 1$ approximately, hence

$$p_k \approx p_{\hat{k}}\exp(-\tfrac{1}{2\gamma}(\Delta k)^2)\,. \tag{B.20}$$

This relation is also true for the $k < \hat{k}$ part, by a similar demonstration in which the role of the product of the rates of transition between states and k and $\hat{k}$, is reversed. From this we deduce that:

$$\sum_{k=k_{\alpha'}}^{k_\alpha} p_k \approx p_{\hat{k}} + 2p_{\hat{k}}\sum_{i>0}\exp(-\tfrac{1}{2\gamma}i^2) \approx p_{\hat{k}}\int_{-\infty}^{\infty}\exp(-\tfrac{1}{2\gamma}x^2)\,\mathrm{d}x \approx p_{\hat{k}}\sqrt{2\pi\gamma}\text{, therefore} \tag{B.21}$$

$$p_{\hat{k}} \approx 1/\sqrt{2\pi\gamma}\,,$$

As well as: $\mathrm{V}[k] = \sum_k p_k(k-\hat{k})^2 \approx 2p_{\hat{k}}\sum_{i>0} i^2\exp(-\tfrac{1}{2\gamma}i^2) \approx p_{\hat{k}}\int_{-\infty}^{\infty} x^2\exp(-\tfrac{1}{2\gamma}x^2)\,\mathrm{d}x \approx \gamma$.

Ultimately,

$$\mathrm{V}[k] \approx \gamma \approx \hat{n}\,. \tag{B.22}$$

Therefore $\mathrm{SD}[k] \approx \sqrt{\hat{n}}$.

Example (continued). We directly calculate $\mathrm{SD}[k] \approx 16.28$. Since $\sqrt{\hat{n}} \approx 16.20$, the approximate formula provides an excellent estimation.

A.9 An approximate Poisson model

We showed earlier that the probability distribution is very close to a normal distribution of mean $\hat{k}$ and standard deviation $\sqrt{\hat{n}} = \sqrt{N-\hat{k}}$. So for the number of busy vehicles, the mean and variance are the same, equaling $\hat{n}$. This suggests that the number could follow a Poisson distribution. And in fact, if we posit $\tilde{p}_n \equiv p_{N-n}$ from the state { $n = 0$ } = { $k = N$ }, i.e. $\tilde{p}_0 = p_N$, these probabilities verify the recurrence relation

$$\tilde{p}_n.\frac{n}{t_{\mathrm{T}}+b(\frac{1}{2}+1/(N-n+1))} = \tilde{p}_{n-1}\lambda\text{, i.e. } \tilde{p}_n = \frac{\tilde{p}_{n-1}}{n}.\lambda(t_{\mathrm{RT}}+\frac{b}{N-n+1})\,.$$

By approximating $\lambda(t_{RT}+\frac{b}{N-n+1}) \approx \lambda(t_{RT}+\frac{b}{\hat{k}+1})$ denoted z, then $\tilde{p}_n \approx \frac{\tilde{p}_{n-1}}{n}.z$ and by recurrence:

$$\tilde{p}_n \approx \frac{\tilde{p}_0}{n!} z^n . \tag{B.23}$$

We recognize a Poisson distribution with parameter z, which incidentally gives an approximation $\tilde{p}_0 \approx \exp(-z)$. Since z satisfies the characteristic equation (B.16), $z = \hat{n}$, thus

$$\mathrm{E}[n] = \mathrm{V}[n] = z = \hat{n} . \tag{B.24}$$

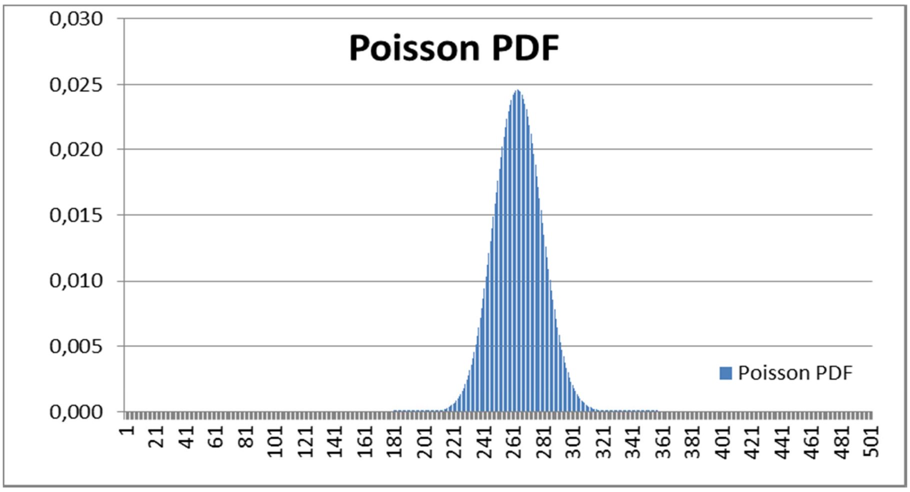

Fig. B4: Probability density in the Poisson model (cf. Fig. B2)

B.10 Recapitulation of the simplifying assumptions

As for reminder, let us recapitulate the sequence of simplifying assumptions that we adopted about the technical operations of the system:

S1/ A Markov model, by equating customer handling times with exponential random variables.

S2/ The simplification of access times, and the averaging of transit and access times, in the formulas for transition rates.

S3/ Using the mean number of vacant vehicles to calculate the mean conditional access time.

S4/ To come close to a Markov model with a Poisson form, the approximation of transition rates and the conversion of the negative range.