How fare simulation tools in urban public transport can benefit from smart card data analysis?

Catherine Bouteiller, Bruno Faivre d'Arcier

Abstract

In public transport, traditional magnetic cards or tickets have been replaced by smart cards very progressively since the 70s. Smart card systems collect day to day variability of users' behavior at a very detailed spatial and temporal resolution and data mining techniques deliver a good reconstruction for Origin and Destination (OD) matrices. This research assesses what the appropriate pricing to optimize trips and social welfare would be, if a new metro line was built and added to an existing public transport network. Our model leverages a traffic function based on the generalized cost. For all OD in the smartcard database, we calculate the potential shift traffic from existing lines to the new one and we calculate the revenue as a function of the fare in the new system. Big data treatment will be necessary to achieve our work by analyzing all origins and destinations...

C.Bouteiller (Corresponding author), Post doc
Laboratory of transport economics LET, Lyon
Email: Catherine.bouteiller@gmail.com

B. Faivre D'Arcier, Pr
Laboratory of transport economics LET, Lyon
Email: bfdarcier@let.ish-lyon.cnrs.fr

Introduction

In public transport, traditional magnetic cards or tickets have been replaced by smart cards very progressively since the 70s. This is nearly achieved in the largest cities (London, Singapore, Seoul…). Advantages of using smart cards are well known and commonly accepted by public transport authorities' strategy. As mentioned by Pelletier et al. (2011) "the smart card improves the quality of the data, gives transit a more modern look, and provides new opportunities for innovative and flexible fare structuring". Smart card systems collect day to day variability of users' behavior at a very detailed spatial and temporal resolution and data mining techniques deliver a good reconstruction for Origin and Destination (OD) matrices (Munizaga & Palma, 2012; Ma et al. 2013; Agard et al. 2013, Trépanier et al. 2009).

This research assesses what the appropriate pricing to optimize trips and social welfare would be if a new metro line was built and added to an existing public transport network. Its originality lies in using smartcard data reconstruction OD matrices to make a fare simulation tool for policy makers.

Grand Paris line 15 will be launched in 2020. It is an example of introducing a new orbital metro line inducing significant time savings for commuters. Passengers are given a real alternative between their usual route and the new one. The new route can offer several advantages: gain of time spent during the trip, gain in terms of number of transfers, comfort gain, security gain, price advantages. This could be an opportunity for an agency to review its fare policy. Our model, based on existing travel flows derived from the itinerary reconstruction by means of smartcard data, establishes a fare grid based on travelers' preferences and current travel patterns.

1 Is it common to use smartcard data to fix fare simulation tools in public transport?

Very few researchers (Zureiqt, 2008; Lovric et al, 2012) have used smartcard data to fix fare simulation tools. Numerous works on tariff optimization were led to maximize the profit of the operators or the social welfare of the users. The thesis of Feng-Ming Tsai (2009) describes several types of works. The models mentioned are often applied to very simplified networks, most of the time single line with an analytical model which considers that the users are sensitive to the quality of the service, or to the headways, to the waiting times and to the fares. Lam and Zhou (2000) suggest maximizing the income of the operators by using the price and route choice variables. Lee and Tsai (2004), use a function to maximize the profit and minimize the generalized costs, including headways variables. Another literature analysis regarding tariff modelling in public transports was realized by Kraus (2012) who distinguishes two types of studies: those optimizing at the same time the fares and the services, and those concentrating on the pricing for public transports. In our bibliographical study, we retained simulation methods developed from smart card data. We retained more particularly two studies combining the use of ticketing data and the tariff simulations: Zureiqat (2008) worked on London Oyster disaggregated smartcard data and tried to measure how every customer's segment (identified according to its travel habits and social profile) could react to a tariff modification. The elasticities are then calculated. A more global segmentation is then realized to determine market shares by fare. The results of Zureiqat’s study are interesting because it is an update of prices elasticities. There are relatively similar to those known: -0.4 in the short term for subway trips and -0.64 in the long term. Zureiqat shows that an increase of price from 1.5£ to 2£ in peak hour in London subway (8:30 to 9:15) could diminish patronage by 9% and thus congestion. We retained a second study, combining smartcard data analysis and modelling: It is the one from Lovric and al. (2012) who designed a DSS (Decision Support System), model multi-agents to limit the congestion by modulating price of the OV Chipkaart (Netherlands) for pricing peak hour / off-peak hour. Unfortunately, the simulation tool does not take into account the possible effects of transfers on other modes, and uses global elasticity data and no user’s profiles information.

2 Why a simulation method based on Navigo smartcard data?

2.1 Context

Grand Paris line 15 will be launched in 2020. It is an example of introducing a new orbital metro line inducing significant time savings for commuters. Passengers are given a real alternative between their usual route and the new one. The new route can offer several advantages: gain of time spent during the trip, gain in terms of number of transfers, comfort gain, security gain, price advantages. This could be an opportunity for an agency to review its fare policy. Our model, based on existing travel flows derived from the itinerary reconstruction by means of smartcard data, establishes a fare grid based on travelers' preferences and current travel patterns. For this research, we used data from public transport pass users in the Paris' Metropolitan area for one month. These data represent 8 373 000 transportation pass contracts in use; 734 000 000 validations processed over three months. Data are completely anonymous. We also used all information available on the new orbital line of Grand Paris that will be launched in 2020. The details of line 15 are already available as open data: stations localizations and connections with the existing metro and RER network, inter-station distances and commercial speed of the trains.

2.2 The affectation of traffic between two rival route

To determine how many users are going to use a route rather than another, we choose to use a function for affecting itineraries, which is generally used for the roads. The mechanism of affectation of traffic between two routes with different generalized costs is known in the road sector thanks to the law of Abraham (1961). We can determine individuals' proportion named Pc for choosing a route according to the duration time on the routes and the prices offered on both routes. This function is given by the following formula (2.1), where C1 is a generalized cost on the route 1, C2 is the generalized cost on the route 2. ß is a factor that is used in the road context. Abraham formula uses a ß of 10. In the urban context, tests were realized to calibrate the ß empirically.

$$Pc2ij = \frac{\left(\frac{C1ij}{C2ij}\right)^{\beta}}{1+\left(\frac{C1ij}{C2ij}\right)^{\beta}} \tag{2.1}$$

i is an OD; j a customer segment (senior, student,…), the route 2 is the new line, route 1 is the existing network. We can determine the price which maximizes the transfer on new route 2 considering the characteristics of both routes.

C1 is a generalized cost that can be expressed as following:

C1ij = P1 +Vj xT1ij (2.2)

V is the Value of Time for a user segment j, T is duration of the trip on ODi and network 1 (existing network)

Fig 2 shows the modal shift from the new line (Grand Paris Line 15) to the existing network for different levels of price (price change on the new Grand Paris line 15 only). The generalized cost function applied is considered for an adult profile, with a value of time of 12.2€ and for a student profile with a value of time of 8€. A ß of 6 is giving the slope of the curve. It has been determined empirically with existing competing routes on the current network.

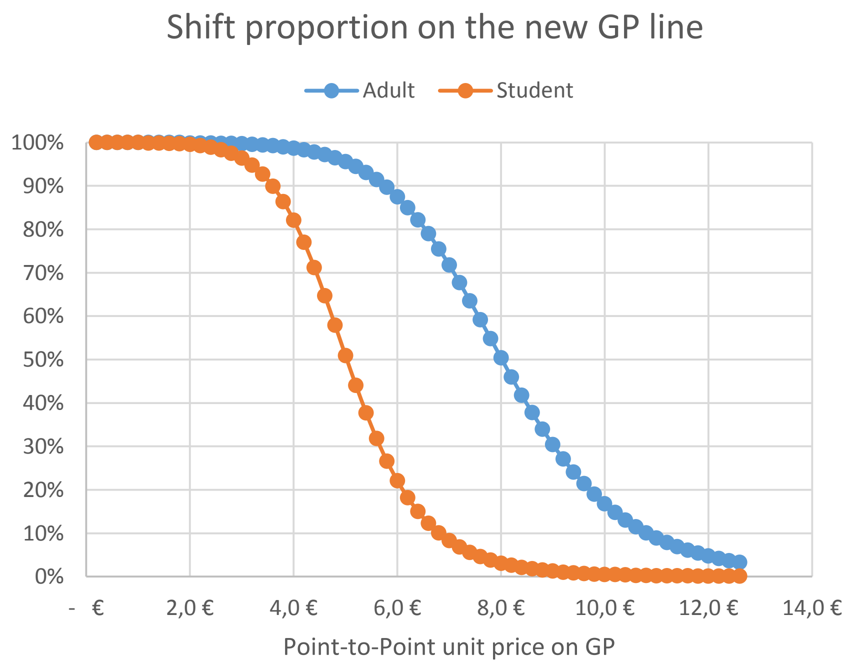

Fig. 1. Modal shift on new Grand Paris Line for two social profiles with two different values of time parameters

If the price on the new line is inferior to the one on the existing network, the users will prefer the new line as the generalized cost is more advantageous. Then, as the unit price increases, the interest is lesser. The inflexion point of 50 % of modal part allocated to every network show the level of price that reorients public transport users to the existing network. The price on the new line can be completely prohibitive and advantage in terms of travel time gained on the new built line could be canceled by an inappropriate princing.

2.3 How princing a new metro line taking into account existing smartcard data ?

If Rt is the total revenue function, it can be rewritten to express the price on the new line 2. The public transport price on the existing line 1 doesn't change.

$$Rtij = R1ij+R2ij$$

The revenue on the new line is the number of customers multiplied by the price on the new Grand Paris line, knowing that the number of customers is calculated from the formula of Abraham's law as a proportion of the initial total number of customers noted N0R2ij = N0ij x Pc2ij x P2ij

The price on the new line P2i for an OD i can express itself as the product of a unit cost per mile (unit price P2u) by a distance D2i.

P2i = P2u x D2i.

Rtij = R1ij+R2ij

R1ij + R2 ij >=R0

We can rewrite these equations and calculate the total revenue on the new network and the existing one (2.3).

$$R1ij = N0ij \times P1 \times \left[1- \frac{\left(\frac{P1 + VjT1ij}{(\mathbf{P2u} \times D2ij) + VjT2ij} \right)^{\beta}}{1+ \left(\frac{P1 + VjT1ij}{(\mathbf{P2u} \times D2ij) + VjT2ij} \right)^{\beta}} \right]$$

$$R2ij = N0ij \times (\mathbf{P2u} \times D2ij) \times \left[\frac{\left(\frac{P1 + VjT1ij}{(\mathbf{P2u} \times D2ij) + VjT2ij} \right)^{\beta}}{1+ \left(\frac{P1 + VjT1ij}{(\mathbf{P2u} \times D2ij) + VjT2ij} \right)^{\beta}} \right]$$

$$Rt = N0ij \;\; \times \left[P1 + \left(((\mathbf{P2u} \times D2ij) \; - P1) \right) \times \frac{\left(\frac{P1 + VjT1ij}{(\mathbf{P2u} \times D2ij) + VjT2ij} \right)^{\beta}}{1+ \left(\frac{P1 + VjT1ij}{(\mathbf{P2u} \times D2ij) + VjT2ij} \right)^{\beta}} \right]$$

(2.3)

The revenue functions above (2.3) are displayed on the graph of the figure 2. There are as many curves of total revenues as OD, because we calculate for every OD the price level from which the revenue goes by a maximum. The price level on the usual network does not change.

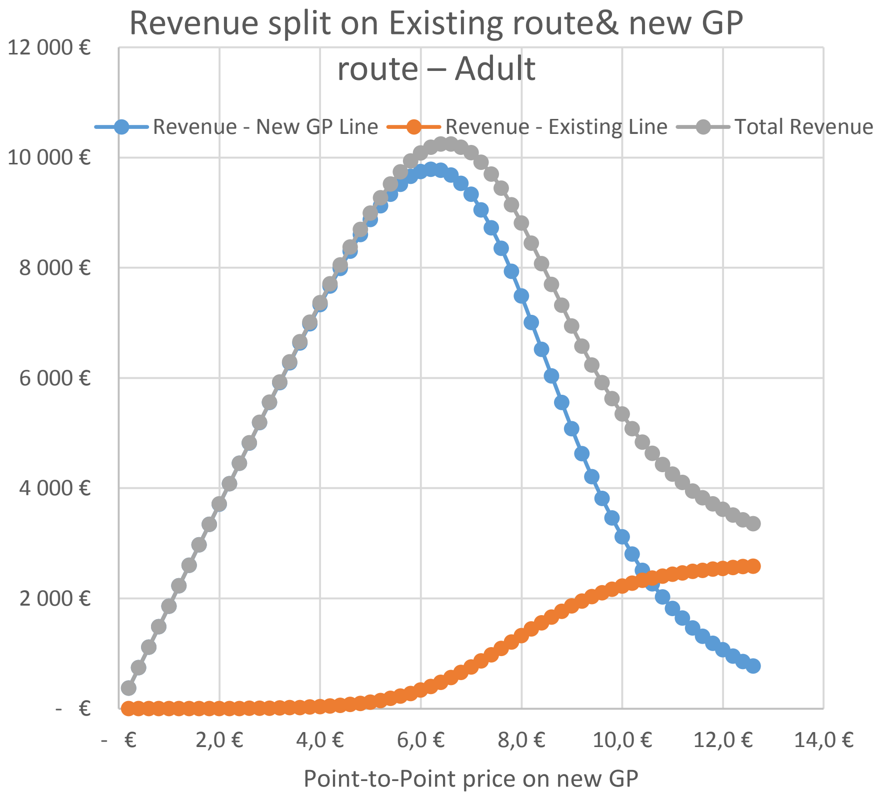

Fig. 2. Visualizing Revenue curves on existing network and new one

2.4 From a unit price per OD to an average kilometric price by distance range

Accessing a pricing by OD would be little readable and very complex. We suggest establishing a zonal kilometric pricing. In the hypothesis where we would test a zonal price with zonal unit costs per mile, Rt must be aggregated by distance traveled noted "d" in the formula (2.4) below.

$$Rt = \sum_{j} \sum_{d} \sum_{i} N0ij \times \left[P1 + \left(((P2u \times D2i) - P1)) \times \frac{\left(\frac{P1 + VjT1ij}{(P2u \times D2ij) + VjT2ij} \right)^{10}}{1 + \left(\frac{P1 + VjT1ij}{(P2u \times D2ij) + VjT2ij} \right)^{10}} \right) \right]$$

(2.4)

The unit price which maximizes the total is given either by graphic resolution, or by looking for the unit price which cancels the by-product of this function. We are going to test this formula on all the rival routes which we identified by the cartographic analysis.

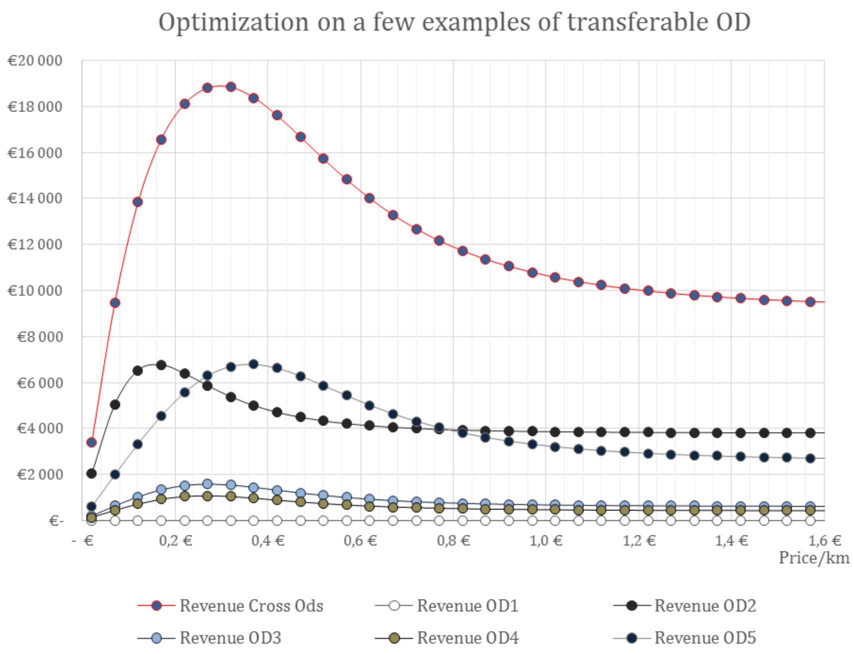

Fig. 3. Curves by OD transferable from existing network to new Grand Paris Line 15

3 Results

As we were not capable of testing our functions on all the origins and the destinations, we worked on five stations origins of the future Grand Paris Line 15 and we studied all the OD which has their origin within a radius of 500 m of the future station of Grand Paris Line 15. About 705 were identified. 40 OD were handled individually because they were representative in terms of users' volume on a time slot, for 1 month and for 1 tariff segment. They represent only 5.6 % of all the OD on this time slot and naturally the results in absolute value are not usable.

The figure 4 below shows all OD recorded by Navigo Pass from Villejuif station and those that can be transferred on the new line.

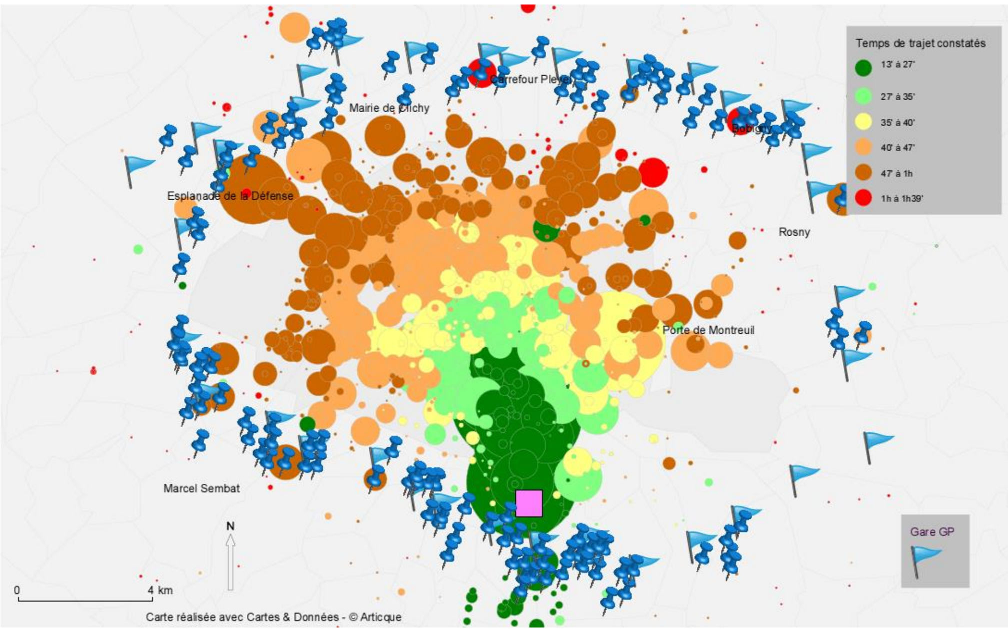

Fig. 4. OD visualization: one origin Villejuif (square)

On the above picture, the size of the bubble give indications on the number of journey during peak hours between Villejuif station (square) and the locations recorded with Navigo pass. The color of the bubble indicates average time recorded for the journeys. Blue spins are the stations located within 500m distance of a future grand Paris station. Journeys that can be transferred to Grand Paris Express line are the ones that are located where a blue spin is.11% of the ODs can be transferred on the new Grand Paris line 15.

Once the ODs are identified, it's possible to calculate the appropriate price and fix the levels of transfer.

To determine the unit price two alternatives are possible; the first one is to determine the price which maximizes the profit. If we do so, only 45% of the population targeted will be interested in the new network, as the level of prices will be too high (table1)
The second possibility is to fix levels of transfer and deduct the price level: for example, in the configuration b of the table 2, a level of global adjournment of 85 % is fixed as acceptable, in particular 90 % of OD transfers between 15 and 20 kilometers. The global revenue will increase by 2%.

Table 1. Results configuration 1

	Price / km	**Shift %**
<3Km	2.17	60%
5-7Km	0.87	69%
7-10 Km	0.42	56%
10-15 Km	0.62	69%
15-20 Km	0.37	50%
>20 Km	0.17	40%

If profit is maximized, we observe a final 163% increase of revenue but 45% overall shift to the Grand Paris network.

Table 2. Results configuration 2

	Price / km	**Shift %**
<3Km	1.92	70%
5-7Km	0.87	70%
7-10 Km	0.22	85%
10-15 Km	0.42	90%
15-20 Km	0.14	90%
>20 Km	0.02	82%

. If we fix limits to obtain as results for modal shift on the new network (between 70% and 90% report), then we observe 82% overall shift to the Grand Paris network and a 2% increase of revenue.

CONCLUSION AND LIMITS

576 public transport stops have been identified within 500 m of the future line Grand Paris Line 15 stops. If we consider all the OD potentially transferable, more than 300 000 OD has to be studied. For this article, 705 OD only have been identified from 5 origins. It was not possible to reconstitute all the routes without big data techniques. Our model leverages a traffic function (Abraham, 1961) based on the generalized cost. For all OD in the smartcard database, we calculate the potential shift traffic from existing lines to the new one (logit model) and we calculate the revenue as a function of the fare in the new system. The cost function used takes into account price, time and value of time by social profiles. With the simulation tool, it is possible to fix a modal shift target, for example 85% overall shift to the new line, and define the fare grid that leads to a total increase of revenue (for instance +2% with a distance based price on the new line). Our results should be refined to take into account all ODs recorded, all profiles and all and peak and off-peak periods as well. Big data treatment is necessary to achieve our work by analyzing all origins and destinations and we still need to calibrate the B variable to adjust the model.

References

Abraham C., Coquand, R. (1961). La répartition du trafic entre itinéraires concurrents, réflexions sur le comportement des usagers, application au calcul des péages, Revue Générale des routes et aérodromes, N°357, pp57-79

Agard, B.; Partovinia, V.; Trépanier, M. (2013). Assessing public transport travel behavior from data with advanced data mining techniques, Communication à la 13th WCTR, 2013, Rio de Janeiro.

Feng-Ming Tsai. (2009). optimizing fare structure and service frequency for an intercity transit system, A Dissertation Submitted to the Faculty of New Jersey Institute of Technology, thesis,168p

Kraus, M.. (2012). Road pricing with optimal mass transit. J. Urban Econ. 72 2,81–86.

Lam, W. H. K., and J. Zhou. (2000). "Optimal Fare Structure for Transit Networks with Elastic Demand." Transportation Research Record 1733, pp. 8-14

Lee, C. K., Tsai, T. H. (2004). "Demand-Responsive Pricing Method for the Product Line of Taiwan High-Speed Rail." Transportation Research Record 1863, pp. 1-8.

Lovric, et al. (2012). Sustainable revenue management: A enabled agent-based modeling approach, Decis.Support Syst, 13oi:10.1016/j.dss.2012.05

Ma, X., Wu, Y-J., Wang.Y., Chen. F., Liu. J. (2013). Mining data for Transit rider's travel patterns, transportation Research, Part C 36 p1-12

Munizaga, M.A., Palma, C. (2012). Estimation of a disaggregate multi-modal public transport Origin-Destination matrix from passive smartcard data from Santiago, Chile, Transportation research Part C, 24, pp. 9-18.

Pelletier, M.-P., Trépanier, M. and Morency, C. (2011). Data use in public transit: A literature review. CIRRELT 2009-46, Centre Interuniversitaire de recherché sur les réseaux d'entreprise, la logistique et le transport, 17pp

Trepanier, M., Morency, C., Agard, B. (2009). Calculation of transit performance measures using smartcard data. Journal of Public Transportation 12 1, 79–96.

Zureiqat, H.M. (2008). Fare policy analysis for public transport: a discrete-continous Modeling Approach Using Panel Data, Master of Science in Transportation, MIT, 117p