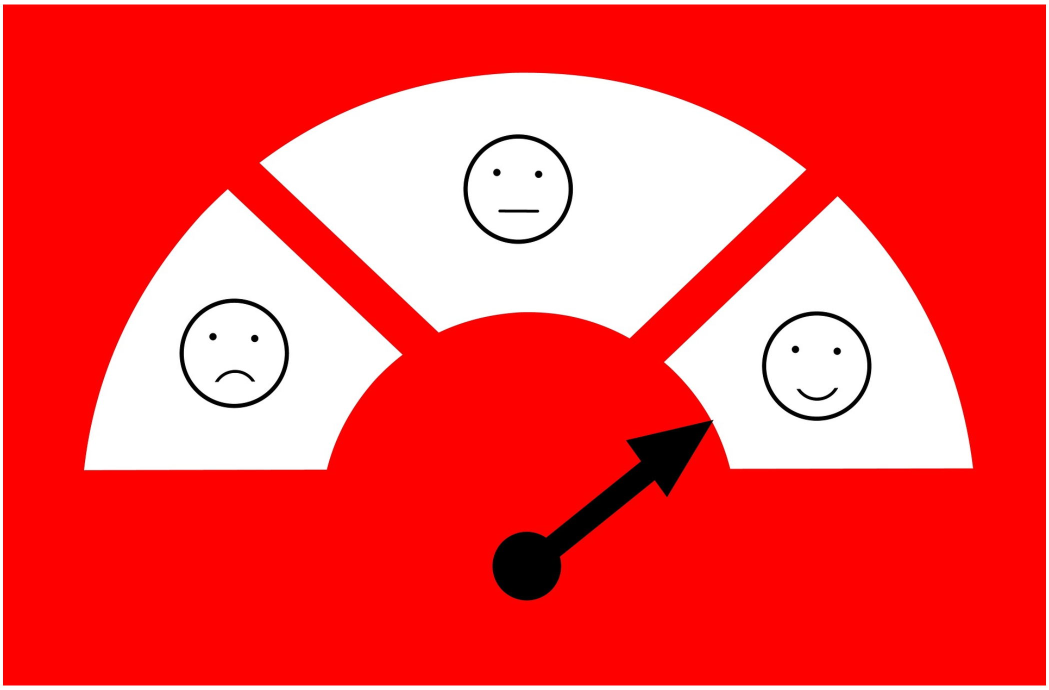

Passenger satisfaction maximization within a demand-based optimization framework

Meritxell Pacheco Paneque

Shadi Sharif Azadeh

Michel Bierlaire

Transport and Mobility Laboratory (EPFL)

May 2019

STRC | 19th Swiss Transport Research Conference
Monte Verità / Ascona, May 15 – 17, 2019

Transport and Mobility Laboratory (EPFL)

Passenger satisfaction maximization within a demand-based optimization framework

Meritxell Pacheco Paneque, Michel Bierlaire
Transport and Mobility Laboratory
École Polytechnique Fédérale de Lausanne
TRANSP-OR, Station 18, CH-1015 Lausanne
phone: +41-21-693-81-00
fax: +41-21-693-80-60
{meritxell.pacheco, michel.bierlaire}@epfl.ch

Shadi Sharif Azadeh
Operations Research & Logistics group
Erasmus University Rotterdam, Netherlands
H11-27 Tinbergen Building, Rotterdam
phone: +31-10-402-2378
fax: +31-10-408-9162
sharifazadeh@ese.eur.nl

May 2019

Abstract

In this research, in order to maximize passenger satisfaction, we consider an approach that integrates discrete choice models, which are the state-of-the-art of the disaggregate modeling of the demand, in Mixed Integer Linear Programming (MILP) models, which are usually considered to address supply decisions such as the price of a service. Typically, passenger satisfaction is modeled in terms of the consumer surplus, but thanks to the linear characterization of the choice model in the demand-based optimization framework, we can express passenger satisfaction directly in terms of the expected maximum utility (one of the variables of the linear formulation), which notably simplifies the methodology. Our goal is the maximization of passenger satisfaction in a short-distance commuting context while accounting for different settings with respect to road tolling and investment in public transportation, also known as revenue recycling The idea behind is that the income collected from the tolls on highways is used to improve the public transportation network. Since supply decisions performed by transportation authorities are associated with investments, we need to include in the model constraints on the investment associated with both the road tolling and the improvement of the public transportation network in order not to exceed the available budget. For the sake of illustration, the resulting formulation is tested on a syntethic case study in the Lausanne-Morges region.

Keywords

discrete choice models, passenger satisfaction, consumer surplus, revenue recycling, road tolling

1 Introduction

Green-house gas emissions, air pollutants and congestion are the main negative externalities associated with urban road transportation (Anas and Lindsey, 2011). According to the theory of welfare economics and externalities (Pigou, 1920), a tax or toll is needed to correct these externalities. In practice, policy instruments associated with road pricing include mostly congestion pricing in the form of a toll to confront drivers with the cost of the congestion delays they impose on other drivers.

In order to address the above-mentioned externalities, several strategies consider a revenue recycling mechanism in which the revenue collected by road tolls is used for transport investments or subsidies. This revenue is typically used for the implementation of the system, and the remaining funds might be used, for instance, for additional road and/or public transportation investments to encourage modal shift (Levinson, 2010).

A key factor for a successful implementation of a road pricing policy is its acceptability (Kidokoro, 2010). Lack of public acceptance has been the most important barrier to road pricing schemes. Acceptability of road pricing is notably affected by the use of the generated toll revenue, and as discussed in Lyons *et al.* (2004), it improves significantly when the revenue is dedicated to the development of transport. Furthermore, in the absence of revenue recycling, some real examples show that the welfare is distributed more unequally.

The objective of this research is to define a Mixed Integer Linear Programming (MILP) model for the maximization of consumer surplus, or welfare, which is considered here as an aggregated measure of passenger satisfaction, while accounting for revenue recycling. This paper represents a first attempt in this context by applying the demand-based optimization framework introduced in Pacheco *et al.* (2017), which describes a choice model linearization that can be embedded in any MILP model, and provides an application in the context of profit maximization. To avoid unboundedness, we need to include in the formulation budget constraints limiting the investments that transportation authorities can carry out. Moreover, the model can accommodate any other assumptions related to the revenue recycling strategy that wants to be implemented.

The remainder of the paper is organized as follows. Section 2 reviews the literature on consumer surplus for passenger maximization and revenue recycling in public transportation. Section 3 details the MILP formulation that uses the choice model linearization to maximize passenger satisfaction with a revenue recycling strategy. Section 4 presents a proof of concept of this methodology by means of a synthetic case study, and Section 5 discusses the main conclusions and future avenues of research.

2 Literature review

This section reviews the concepts of passenger satisfaction and revenue recycling. Section 2.1 overviews the definition and measurement of the former, whereas Section 2.2 includes various examples of applications with different demand representations and pricing schemes for the latter. Section 2.3 summarizes the main conclusions of the two parts of the literature review.

2.1 Passenger satisfaction

The concept of passenger satisfaction originates from customer satisfaction research, which is a popular field of study in marketing. In the context of transportation, it often refers to the service offered and the passenger reaction to the service (St-Louis *et al.,* 2014). It depends not only on the attributes associated with the trip and the transportation mode, but also on the socio-economic characteristics and attitudes of passengers. Thus, a disaggregate representation of the demand including this information is beneficial for the measurement of passenger satisfaction.

As mentioned in Section 1, we formulate the demand by means of a discrete choice model. One of the indicators that can be derived from such models, and that is useful for policy analysis, is the consumer surplus (sometimes denominated welfare or benefit). In the microeconomic consumer theory, in the context of continuous goods, consumer surplus is defined as the difference between what a consumer is willing to pay for one good and what she actually pays for it, which corresponds to the experienced satisfaction. In the case of discrete choice models, the role of price is taken by the utility of the good (the transportation mode in this case). At the same time, the expected maximum utility represents a scalar summary of the expected *worth* of a set of alternatives (Ben-Akiva and Lerman*,* 1985), whose difference in two situations corresponds to the difference in consumer surplus between the two situations in the case of the Multivariate Extreme Value (MEV) models, such as the logit model (see Section 3.2).

We highlight two works in transportation where a discrete choice model represents the demand, and passenger satisfaction is considered during the supply decision making. In Robenek *et al.* (2016), a MILP model is defined to maximize the train operating company's profit while maintaining an ε level of passenger satisfaction (negative of the generalized cost). The results show that large improvements on the passenger satisfaction can be achieved while maintaining a low profit loss for the operator. In Atasoy *et al.* (2015), the Flexible Mobility on Demand (FMOD) system, an innovative concept offering a personalized menu of services to passengers, is built on an assortment optimization framework that optimizes such a menu while accounting for the trade-off between the consumer surplus (passenger satisfaction) and operator's profit.

2.2 Revenue recycling in public transportation

As reviewed in Tsekeris and Voss (2010), the relation between road pricing and public transportation is specially examined in the literature as a stimulus-answer relation rather than an aspect of a global strategy of the transport system. The use of disaggregate demand models provides valuable insight into the feasibility of road pricing and public transportation management schemes, as well as the trade-offs between efficiency and equity impacts of such schemes. However, the discussed approaches usually rely on a priori restrictions on the demand elasticities and substitution patterns, which are not supported by the economic theory of demand.

In the road pricing design process, public transportation can be taken into account through different considerations, being the revenue recycling, i.e., the reinvestment of revenues from road tolling to public transportation or to the road network one of them. In the absence of revenue recycling, several examples analyzed in Levinson (2010) conclude that the welfare is distributed more unequally after introducing the road pricing than before. Indeed, revenue recycling offers a way of ameliorating adverse equity impacts, being the High Occupancy Toll (HOT) lanes the pricing strategy least likely to raise public concerns given that the revenue can be recycled to benefit public transportation users in the corresponding corridor.

Huang (2000) and Huang (2002) compare different pricing schemes for a simplified corridor network in a bi-modal context (auto and transit). In the former, a deterministic demand model is considered for two groups of commuters with given size and for which only the internal elasticity is modeled. Two of the strategies include revenue recycling: a marginal cost-based fare with time-invariant toll for subsidizing public transportation, and an optimization problem where the transit fare and the road toll are unknown quantities to be decided to minimize the sum of the total social costs of the highway and the mass transit subject to revenue recycling. In the latter, the demand is assumed to be elastic, all commuters are supposed to be identical and the modal split is governed by a logit model. A first-best pricing model (optimal road toll and transit fare through maximizing the net social benefit of the system), and a second-best pricing model (first-best pricing model without road toll) are considered. The numerical results show that the first-best strategy generates higher social welfare improvement than the second-best.

Also in the bi-modal context, Basso and Jara-Díaz (2012) define an optimal (welfare maximizing) pricing and design of transport services while accounting for car congestion. The optimization variables are the congestion toll, the transit fare and the transit frequency, and the consumers choose the mode based on the perceived price. The demand is formulated with a general choice model, specified later with deterministic utilities and with a logit model. From the several findings, we highlight the feasibility of the self financing of the transport sector and the different optimal modal split when accounting for behavior as opposed to minimizing the total cost.

2.3 Conclusions

The measure of passenger satisfaction considered in the following is the expected maximum utility. In Section 3.2, we provide a detailed explanation on the mathematical relation between this quantity and the consumer surplus. From the reviewed references, we notice that the use of the consumer surplus to quantify passenger satisfaction is a common practice when a discrete choice model is used to model the demand (Atasoy *et al.,* 2015, Basso and Jara-Díaz*,* 2012).

In order to illustrate the use of a revenue recycling strategy within the passenger satisfaction maximization framework, we consider two supply decisions: the highway toll and the public transportation fare. As explained in Section 3.3, we assume that a single transportation authority manages both the highway (car alternative) and the public transportation alternative. The budget to invest in both alternatives is composed of the already available investment budget (which can be assumed to be 0), and the revenues collected from the tolls and fares.

3 Methodology

In this section, we present the MILP formulation designed to maximize passenger satisfaction while accounting for a revenue recycling strategy. In Section 3.1 we summarize the linearization of the choice model introduced in Pacheco *et al.* (2017). Section 3.2 characterizes the objective function (passenger satisfaction maximization) and Section 3.3 depicts the constraints that define a general revenue recycling scheme. Section 3.4 sums up the complete formulation.

3.1 Choice model linearization through simulation

Discrete choice models are the state-of-the-art of demand modeling at the disaggregate level. The associated probability formulas are typically nonlinear and non convex, which makes it difficult to include them in a MILP model requiring a disaggregate demand representation. The probabilistic nature of these models is overcome in Pacheco *et al.* (2017) by relying on simulation to specify the model directly in terms of the utility functions (instead of the choice probabilities). This allows us to define a set of linear constraints that characterizes the preference structure and the behavioral assumption of choice models, which can be inserted in any MILP formulation.

We assume a population of N individuals (indexed by n) and we denote the choice set by C

(indexed by i). The utility function associated by individual n with alternative $i \in C_n$ (the set of alternatives considered by n) is

$$U_{in} = V_{in} + \varepsilon_{in}, \qquad \forall i \in C_n, n, \tag{1}$$

where V_{in} denotes the systematic part of the utility function and ε_{in} the random term. The probabilistic nature of the choice model is addressed with simulation by generating R draws from the distribution of ε_{in}. We denote by U_{inr} the utility associated with alternative $i \in C_n$ by individual n in draw r and by ξ_{inr} the r-th draw from the distribution of ε_{in}:

$$U_{inr} = V_{in} + \xi_{inr}, \qquad \forall i \in C_n, n, r, \tag{2}$$

We notice that U_{inr} is not a random variable.

We define the variables U_{nr} to model the highest utility for customer n in scenario r:

$$U_{nr} = \max_{i \in C_n} U_{inr}, \qquad \forall n, r, \tag{3}$$

We can then define the binary variables w_{inr} to capture the choice, which are 1 if U_{nr} is achieved at alternative i, and 0 otherwise (each individual is choosing exactly one alternative). The demand of alternative i can be obtained by averaging the sum of the choice variables over R:

$$D_i = \frac{1}{R} \sum_{r=1}^{R} \sum_{n=1}^{N} w_{inr}, \qquad \forall i \in C. \tag{4}$$

The utility function (11), the constraints linearizing definition (3) and the ones that ensure that only one alternative is chosen per individual and draw can be included in any MILP formulation as a disaggregate demand representation. A concrete application on passenger satisfaction maximization accounting for a revenue recycling scheme is developed in Sections 3.2 and 3.3. We note that the only condition for the integration of the linearized choice model in an MILP model is that its decision variables appear linearly in V_{in}. Such variables are called endogenous variables, as they are present in both the optimization model and the choice model.

This representation of a choice model enables to capture the interactions between the demand and the supply (modeled by the MILP model) under the same formulation, whose objective function can be defined either from the point of view of the operator (supply) or the individuals (demand). In Pacheco *et al.,* 2017, we have defined a MILP formulation that aims at maximizing the profit of an operator that sells services to a market, whereas in Section 3.2 we account for the individuals' perspective by optimizing the passenger satisfaction.

3.2 Passenger satisfaction maximization

As discussed in Section 2.3, we define passenger satisfaction directly in terms of the expected maximum utility $\mathbf{E}[\max_{i \in C_n} U_{in}]$ (Ben-Akiva and Lerman, 1985). The largest utility within the choice set represents the benefit obtained by individual n from the chosen alternative. As utility is a random variable, we are interested in the average of this benefit, that is, the average of the maximum utility. This quantity is called the expected maximum utility, and allows us to compare different choice situations and to identify the benefit associated with them.

In the case of the logit model, the expected maximum utility is calculated as

$$\mathbf{E}[\max_{i \in C_n} U_{in}] = \frac{1}{\mu}(\ln \sum_{i \in C_n} e^{\mu V_{in}} + \gamma), \qquad \forall n, \tag{5}$$

where μ is the scale parameter and γ is the Euler's constant. This quantity is equivalent to the consumer surplus up to a constant. In general, the expected maximum utility is used to compute differences (comparison of two scenarios), which enables to ignore the γ terms as they cancel out. Hence, the difference of individual n consumer surplus between two situations is

$$\frac{1}{\mu}(\ln \sum_{i \in C_n} e^{\mu V_{in}^2}) - \frac{1}{\mu}(\ln \sum_{i \in C_n} e^{\mu V_{in}^1}), \qquad \forall n, \tag{6}$$

which is the difference between individual n expected maximum utility in the two situations.

More generally, the expected maximum utility of Multivariate Extreme Value (MEV) models is computed by replacing the sum in (5) by the corresponding choice probability generating function (CPGF). These models represent a family of choice models that allow for the correlation among the error terms of the utility functions (e.g., logit model, nested logit model).

In this context, the expected maximum utility of individual n is approximated by

$$\frac{1}{R}\sum_{r=1}^{R} \mathbf{E}[\max_{i \in C_n} U_{inr}] \overset{(3)}{=} \frac{1}{R}\sum_{r=1}^{R} \mathbf{E}[U_{nr}] = \frac{1}{R}\sum_{r=1}^{R} U_{nr}, \qquad \forall n. \tag{7}$$

The first equality comes from the definition (3), and the second one from the fact that U_{nr} is not a random variable (U_{inr} are not random variables), and therefore its expected value is directly U_{nr}. We define the measure of passenger satisfaction by aggregating the expected maximum utilities for all individuals, i.e.,

$$\frac{1}{R}\sum_{n=1}^{N}\sum_{r=1}^{R} U_{nr}, \qquad \forall n. \tag{8}$$

3.3 Revenue recycling strategy

We consider a population consisting of N individuals that need to perform a trip within a certain time horizon, and a choice set C composed by, at least, two modes of transportation: private (car for short) and public transportation (PT). We define a revenue recycling strategy where a single transportation authority needs to perform two supply decisions: the toll associated with the use of the highway and the fare associated with the trip performed by public transportation (see Section 2.3). These decisions are endogenous variables of the formulation, as we assume that they have an impact on the behavior of the individuals (i.e., they should be present in the utility functions). We notice that the purpose of this supply configuration is to illustrate the use of the framework, and that other features could be included and/or replaced.

As mentioned in Section 1, in the absence of constraints limiting the investments to be made by the transportation authorities, the optimization problem defined by the maximization of (8) becomes unbounded, since the aggregated expected maximum utility increases as the investment on the supply decisions increases. Thus, we need to include a constraint ensuring that the investment does not exceed the available budget.

In such a context, the investment comprises the costs associated with the implementation and operation of the tolling infrastructure, and the operating costs associated with public transportation. On the other hand, the budget is defined as the sum of the already available budget by the transportation authority and the revenues collected from the tolls and the fares. The former is known beforehand, and can be assumed to be 0, whereas the second is endogenous to the formulation, as it depends on the expected demand for both modes.

We denote by B and I the budget and the investment, respectively. We limit the investment by adding the constraint $I \leq B$ to the model. The budget B is composed of an initial available budget B^0 and the collected revenues. We denote by $p_{\text{car},n}$ and $p_{\text{PT},n}$ the toll and the fare to be paid by individual n, respectively (it can be the same for everyone). Thus, B is calculated as

$$B = B^0 + \frac{1}{R} \sum_{n=1}^{N} \sum_{r=1}^{R} [p_{\text{car},n} w_{\text{car},n,r} + p_{\text{PT},n} w_{\text{PT},n,r}]. \tag{9}$$

Notice that constraint (9) is not linear on the decision variables of the optimization problem, as we find the product of the toll (fare) with the choice variables. To linearize it, we need to define a continuous variable η_{inr} to capture the product $p_{in} w_{inr}$. Then, lower and upper bounds need to be assumed on the continuous variable of the product (i.e., p_{in}). The corresponding linearizing constraints are discussed in Section 3.4 (constraints (16)–(19)).

On the other hand, the investment I is composed of the investments I^{car} and I^{PT}. We assume that the former involves a fixed cost F^{car} (costs associated with the toll facility for the considered time horizon) and a variable cost that is defined by a cost per transaction c^{car}, which is assumed to be the same for everyone. For the sake of simplicity, we only take into account a fixed cost F^{PT} for public transportation, since most of the expenses associated with this mode (labor, fuel, electricity, etc.) are independent of the number of passengers carried. The total investment I is calculated as

$$I = I^{\text{car}} + I^{\text{PT}} = F^{\text{car}} + \frac{1}{R} c^{\text{car}} \sum_{n=1}^{N} \sum_{r=1}^{R} w_{\text{car},n,r} + F^{\text{PT}}. \tag{10}$$

We note that additional assumptions on the different elements defining the budget and the investment can be added to the formulation.

3.4 Demand-based passenger satisfaction maximization with revenue recycling

The formulation (8),(11)–(19) represents the passenger satisfaction maximization problem with revenue recycling. Constraint (11) corresponds to the utility function for alternative i, where p_{in} denotes the cost associated with mode i and β_{in} the associated parameter. Furthermore, the utility function also includes an exogenous term $g_{in}(x_{in})$, which contains other variables that are not decision variables of the MILP formulation, but contribute to explain the choice, and the random draw ξ_{inr}. We notice that we could also allow for other transportation modes, in which case the corresponding utility functions should be added to the model.

Constraints (12)–(14) model the behavioral assumption stating that the alternative with the highest utility (and only that alternative) is chosen by individual n at draw r. Given that $M_{inr} = m_{nr} - \ell_{inr}$, where $\ell_{inr} \leq U_{inr} \leq m_{inr}$ and $m_{nr} = \max_{j \in C} m_{jnr}$, it is clear that $w_{inr} = 0$ when U_{inr} is not the highest utility, and thanks to constraint (14), which also ensures that one and only one alternative is chosen, $w_{inr} = 1$ when $U_{nr} = U_{inr}$. The budget-investment relationship is formulated in constraint (15). The linearizing constraints. Finally, constraints (16)–(19) are the linearizing constraints associated with the variable η_{inr}, where a_{in} and b_{in} are lower and upper bounds on p_{in}, respectively. Constraints (16)–(17) are binding when $w_{inr} = 0$, and impose $\eta_{inr} = 0$, whereas constraints (18)–(19) are binding when w_{inr} and impose $\eta_{inr} = p_{inr}$. Notice that these are the two possible values of the product $p_{in} w_{inr}$.

$$
\begin{aligned}
Z = \max \quad & \frac{1}{R}\sum_{n=1}^{N}\sum_{r=1}^{R} U_{nr} && && (8)\\
\text{s.t.} \quad & U_{inr} = \beta_{in}p_{in} + g_{in}(x_{in}) + \xi_{inr} && \forall i,n,r && (11)\\
& U_{inr} \leq U_{nr} && \forall i,n,r && (12)\\
& U_{nr} \leq U_{inr} + M_{inr}(1 - w_{inr}) && \forall i,n,r && (13)\\
& \sum_{i\in C} w_{inr} = 1 && \forall n,r && (14)\\
& F^{\text{car}} + \frac{1}{R}c^{\text{car}}\sum_{n=1}^{N}\sum_{r=1}^{R} w_{\text{car},n,r} + F^{\text{PT}} \leq B^0 + \frac{1}{R}\sum_{i\in C}\sum_{n=1}^{N}\sum_{r=1}^{R}\eta_{inr} && && (15)\\
& a_{in}w_{inr} \leq \eta_{inr} && \forall i>0,n,r && (16)\\
& \eta_{inr} \leq b_{in}w_{inr} && \forall i>0,n,r && (17)\\
& p_{in} - (1 - w_{inr})b_{in} \leq \eta_{inr} && \forall i>0,n,r && (18)\\
& \eta_{inr} \leq p_{in} - (1 - w_{inr})a_{in} && \forall i>0,n,r && (19)
\end{aligned}
$$

4 Proof of concept

We consider a simple but meaningful case study for the proof of concept of the framework described in Section 3. In Section 4.1 we present the scenario that we are interested in analyzing. The logit model explaining the behavior of the travelers is estimated in Section 4.2. In Section 4.3 we show the creation of a synthetic sample (representing the population of interest) to run the optimization problem, whose results are compared with an initial situation in Section 4.4.

4.1 Scenario

In this section, we describe the scenario used to test the formulation by defining the population of interest and the set of relevant transportation modes. To this end, we restrict ourselves to a certain time horizon and two geographical zones among which the trips take place.

More precisely, we get inspiration from the geographical definition of subregions (called cordons) of the Lausanne-Morges region for data collection of public and private transportation by *Lausanne Région* (Guillaume-Gentil *et al.,* 2015). Figure 1 shows the cordons the region is divided into, for which the incoming, outgoing and inner trips are measured. We consider here the trips with origin in the city center of Morges (cordon 4) and destination in the city center of Lausanne (cordon 1). This selection is motivated by the fact that 66.6% of the trips made by private transportation during the morning peak hour in 2014 in the Lausanne-Morges region (cordon 6) used the highway, and although this share is supposed to be smaller between both city centers, it allows for a more tractable problem in terms of the characterization of the trips.

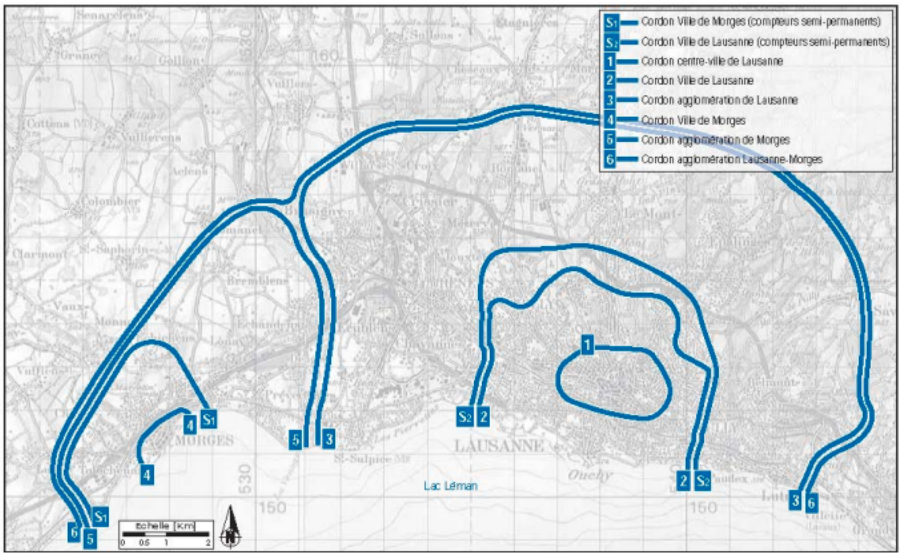

Figure 1: Geographical division of the Lausanne-Morges region into cordons. Source: Guillaume-Gentil *et al.,* 2015

We consider three modes connecting the two areas described above: car, public transportation and slow modes (SM). We assume that the car alternative is related to highway A1 and E23, whose usage is subject to a toll, and that public transportation comprises only the train, as it provides the fastest connection and is the only mode linking the two zones without the need for transfers. Given the distance between the two cities (between 11 and 15 km depending on the origin and destination of the trip), we account only for bicycle as a slow mode.

Furthermore, to better specify the trips, we select the morning peak hour as the time horizon, which is defined from 7 a.m. to 8 a.m.. In this way, we can provide more stable values for travel time (specially for car), and we can simplify some of the characteristics of the target population, such as the trip purpose. We assume that the population consists of individuals that need to perform one trip between the two cities for business purposes.

We notice that we do not have data for the scenario described in this section, neither at the disaggregate level nor at the aggregate one. As mentioned before, the trips measured by *Lausanne Région* are the incoming, outgoing and internal trips within each of the cordons in Figure 1, but not in between cordons. Thus, in order to specify and calibrate the choice model, we consider a dataset from the *Optima* case study, which deals with the estimation of a mode choice behavior model for inhabitants in Switzerland using revealed preference data (Section 4.2). We then create a synthetic sample by generating disaggregate data for the variables that are in the choice model and are not decision variables of the optimization model (Section 4.3).

4.2 Choice model

The data related to the Optima case study was obtained within a project conducted by Professors Vincent Kaufmann (LASUR, EPFL), Michel Bierlaire (TRANSP-OR, EPFL) and Martin Schuler (CEAT, EPFL) with the objective to show the market potential for combined mobility, especially within agglomerations. The survey was conducted between 2009 and 2010 by CarPostal, the public transport branch of the Swiss Postal Service. The main purpose of this survey was to collect data for analyzing the travel behavior of people in low-density areas, where CarPostal typically serves. The collected information (1124 completed surveys) contains origin, destination, cost, travel time, chosen mode and activity at the destination. Moreover, socio-economic information about the respondents and their households was collected, such as gender, number of children and age.

We specify a choice model that includes the most relevant attributes of the three alternatives: the travel time (in minutes) and the travel cost (in CHF) for car and public transportation, and the distance (in km) for slow modes. Regarding the socioeconomic characteristics of the individuals, we only incorporate the interaction between the cost and the net income level of the household per month, which we define as a categorical variable taking three levels: low income (less than 4000 CHF), medium (between 4001 and 8000 CHF), and high (more than 8001 CHF).

The systematic part of the utility functions for the three alternatives are specified in equations (20)–(22). The alternative specific constant (ASC) is normalized to 0 for the slow modes. The specifications of car and PT contain a generic coefficient (β_{Time}) for the variables TimeCar_n and TimePT_n, which are defined as the total travel time in minutes between the origin and the destination of the trip. The cost variables CostCar_n and CostPT_n, which expressed the gas cost and the cost associated with the trip in CHF, respectively, are interacted with the net monthly income of the household, which is expressed as a categorical variable taking three levels, represented by the binary variables LowIncome_n (< 4000 CHF), MedIncome_n (between 4001 CHF and 8000 CHF), and HighIncome_n (> 8000 CHF). Three generic coefficients are defined for each income category: $\beta_{\text{CostLow}}, \beta_{\text{CostMed}}, \beta_{\text{CostHigh}}$. The slow modes only take into consideration the distance of the trip in km (Distance_n).

$$\begin{aligned}
\text{V}_{\text{car},n} &= \text{ASC}_{\text{car}} + \beta_{\text{Time}} \cdot \text{TimeCar}_n + \beta_{\text{CostLow}} \cdot \text{CostCar}_n \cdot \text{LowIncome}_n \\
&\quad + \beta_{\text{CostMed}} \cdot \text{CostCar}_n \cdot \text{MedIncome}_n + \beta_{\text{CostHigh}} \cdot \text{CostCar}_n \cdot \text{HighIncome}_n \qquad (20)\\
\text{V}_{\text{PT},n} &= \text{ASC}_{\text{PT}} + \beta_{\text{Time}} \cdot \text{TimePT}_n + \beta_{\text{CostLow}} \cdot \text{CostPT}_n \cdot \text{LowIncome}_n \\
&\quad + \beta_{\text{CostMed}} \cdot \text{CostPT}_n \cdot \text{MedIncome}_n + \beta_{\text{CostHigh}} \cdot \text{CostPT}_n \cdot \text{HighIncome}_n \qquad (21)\\
\text{V}_{\text{SM},n} &= \beta_{\text{Distance}} \cdot \text{Distance}_n \qquad (22)
\end{aligned}$$

The estimates of the parameters can be found in Table 1. The model has been estimated in a sample of 446 individuals, after excluding some of the observations reporting missing values for the variables of interest. Moreover, we have also excluded the observations with a rural residence location and with a leisure purpose of the trip. All else being equal, the public transportation is preferred over the car (ASC_{PT} = 1.57), and both are preferred over the bicycle (ASC_{car} = 0.96). The sensitivity towards cost and time have negative sign, as expected. The former is higher for the medium category, followed by the low and the high income categories. This is not expected, but might be due to the low representation of this income category in the considered dataset (only 34 individuals). With respect to the distance, the parameter has also negative sign, as expected, as the higher the distance to the destination, the lower the utility to consider a slow mode. Even though β_{Distance} is not significantly different from 0 (applying a 95% confidence level), we decide to keep it in the specification as it is a relevant variable characterizing the slow modes (moreover, we notice that only 27 individuals in the sample have chosen slow modes).

Table 1: Estimates of the logit model defined by the systematic utilities (20), (21) and (22)

Estimation results

	Parameter	Param. estimate	Rob. std err	Rob. t-stat	Rob. p-value
1	ASC_{car}	0.958	0.731	1.31	0.19
2	ASC_{PT}	1.57	0.638	2.45	0.0142
3	β_{CostHigh}	-0.105	0.0482	-2.18	0.0292
4	β_{CostLow}	-0.143	0.0619	-2.3	0.0214
5	β_{CostMed}	-0.198	0.059	-3.36	0.00079
6	β_{Distance}	-0.125	0.0659	-1.89	0.0586
7	β_{Time}	-0.016	0.00426	-3.75	0.000177

Summary statistics

Number of observations=446

$\mathcal{L}(0) = -489.981$

$\mathcal{L}(\hat{\beta}) = -290.890$

$\bar{\rho}^2 = 0.392$

The purpose of the choice model estimated here is to illustrate the logic of the formulation described in Section 3.4. Hence, we have not performed any statistical tests to confront this specification with other specifications. This model includes the most relevant attributes of the transportation modes under consideration, and a socioeconomic characteristic of the population to capture the heterogeneity of the individuals. These variables are considered in Section 4.3 for the generation of the synthetic sample that will be used to test the approach.

4.3 Synthetic sample

As with the choice model, this sample is created with the objective of testing the introduced methodology. To this end, we perform appropriate assumptions for the variables present in the choice model and randomly generate their values for each individual in order to create a sample of 50 individuals that represents the population of the scenario described in Section 4.1.

In order to characterize the trips, we assume that the distance between the origin and the destination consists of a distance to be traveled within the origin zone (city center of Morges), a distance to be traveled within the destination zone (city center of Lausanne), and a distance between the two zones. We assume that the first distance ranges between 100 m and 1.5 km, whereas the second one varies between 200 m and 3 km, as the second cordon is larger than the first. Then, for each individual we randomly draw from a uniform distribution with minimum and maximum the specified values. The distance connecting the two zones is fixed to 12 km.

The variable Distance_n is the sum of the three above-mentioned distances. To calculate the travel time by PT and by car we need to perform additional assumptions on the speed and, in the case of the former, on the waiting time and in-vehicle time.

In the case of TimeCar_n, we assume that the speed of cars within the origin and destination zones is 15 km/h during peak hours (Christidis and Ibáñez-Rivas, 2012), whereas the speed associated with the use of the highway varies between 45 km/h and 70 km/h (according to Google Maps) for the morning peak hour. Once we know the average speeds, we just need to convert the distances into travel times and sum them to obtain TimeCar_n.

TimePT_n is calculated as the sum of the access time, waiting time, in-vehicle time and egress time. We define the access and egress time as the times spent within each zone. For the former, we assume a walking speed of 5 km/h, while for the later we assume the same speed if the distance is shorter than 1.5 km, and a larger speed (15 km/h, the same as for cars in urban environments during peak hours) if the distance is higher, representing the fact that the individual is not walking from the train station but taking a faster mode (e.g., bus) instead. As the purpose of the trips is business-related, we assume that the individuals are familiar with the schedule, and do not plan to arrive to the station way in advance. We therefore assume a uniform distribution between 0 and 8 minutes (which corresponds to the expected waiting time between the trains in the morning peak hour with the highest headway). Since we do not take into account the departure time, we simplify the definition of the in-vehicle time by using the same value for everyone: the weighted average of the in-vehicle times (2 connections with a duration of 12 minutes, 2 with a duration of 13 minutes, 1 with a duration of 15 minutes and 1 with a duration of 18 minutes), which is equal to 13.8 minutes.

The last variable we need to generate is the income. To do so, we use the data on net monthly income at the Swiss level for 2016 (Federal Statistical Office), which state that the percentage of low, medium and high income (as defined in Section 4.2) are 34.4%, 50.6% and 15%, respectively. We simply assign a level of income to each individual in the sample that ensures that the mentioned shares are preserved.

4.4 Benchmark

In this section, we compare the results obtained from the optimization problem (Section 3.4) with an initial situation that we define next, as the value of the expected maximum utility is not well defined unless some benchmark level is established. In both situations, we assume that both the fare and the toll are the same for everyone, i.e., we consider p_i instead of p_{in}.

For the initial situation, we define the PT fare by dividing the cost of the monthly ticket (137 CHF for 2019) by the number of working trips in a month (252 working days in 2019 in canton Vaud, which makes an average of 21 working days per month, i.e., 42 trips per month): 3.27 CHF, which is assumed to be the same for everyone. With respect to the cost of car, we take into account the associated variable costs, which include the maintenance and repairs, tires, gas and depreciation. This cost corresponds to 37.9% of the total cost associated with the car, and since the cost per kilometer is assumed to be 0.71 CHF, the variable cost is 0.27 CHF/km (Touring Club Suisse). We note that the cost by kilometer is then multiplied by the trip distance to obtain the total cost of the car alternative.

Now we need to make assumptions on the initial budget and the required investment. With respect to the former, we assume that there is no initial budget ($B^0 = 0$), which means that the revenues collected from both paid alternatives need to cover the involved costs. For the fixed cost of car and public transportation, we consider the cost per person and kilometer associated with car (0.076 CHF), and the same cost associated with the railway transportation (0.032 CHF, Federal Office for Spatial Development), and multiply them by the size of the sample and the average distance to obtain fixed costs whose order of magnitude is appropriate to the size of the problem ($F^{car} = 54.53$ CHF and $F^{PT} = 22.96$ CHF). Finally, for the car variable cost c^{car}, we consider the cost associated with each transaction (payment) that takes place in the highway, which is assumed to be 0.44 CHF (KPMG). We notice that a more accurate calculation for these costs is necessary to obtain results that can be directly applied in reality, and use these values with the sole objective of illustrating the logic of the formulation.

Table 2 contains the obtained results for $R = 100$ draws. We observe a decrease of the fare with the presence of the toll, as expected. We notice that this value is rather high, which might be due to the assumptions on the fixed costs. The demand for public transportation increases, since the demand for car has decreased, whereas the demand for slow modes remains quite similar. Finally, we observe an increase of the passenger satisfaction with the new price configuration.

Table 2: Fare and toll (CHF), expected demand for the three alternatives (%) and expected maximum utility for the discussed situations for a sample of $N = 50$ individuals and $R = 100$

Situation	Fare	Toll	Demand PT	Demand car	Demand SM	Expected Max. Utility
Initial	3.27	0	57.50	36.22	6.28	156.27
Optimized	1.56	2.30	71.8	23.02	5.18	163.77

5 Conclusions

This paper presents a first attempt by the authors to adapt the demand-based optimization framework developed in Pacheco *et al.* (2017) to the context of passenger satisfaction with revenue recycling. We aim at maximizing the passenger satisfaction, which is defined as the expected maximum utility. The optimization problem is completed with the constraints that linearly characterize the choice model, and the ones defining the revenue recycling strategy. For the sake of illustration, the resulting formulation is tested on a synthetic case study in the Lausanne-Morges region.

The framework enables to accommodate any supply decision variable of interest, which can also be included in the choice model specification to capture the supply-demand interaction. In this case, we have only considered the toll and the fare to illustrate the approach, but we plan to account for other variables such as the frequency of the public transportation mode. The revenue recycling strategy is also general in the sense that both the toll and the fare are decision variables, and the only requirement is that the necessary investment does not exceed the available budget, in such a way that the revenues collected by both modes are not allocated beforehand. Thus, this formulation provides a flexible approach and allows to integrate different policies and to evaluate diverse goals (e.g., modal shift).

The next step consists in incorporating the frequency of the public transportation mode, which might require the tracking of the available capacity to provide meaningful results. This will extend the model by including capacity constraints, which will make the problem computationally more expensive. In terms of testing the formulation, we will consider an advanced choice model from the literature to better represent the behavior of individuals, and we will define different and more complex scenarios that will allow a better assessment of the changes in passenger satisfaction for the revenue recycling strategy under consideration. For the construction of these new scenarios, an extensive research on the assumptions and state-of-the-art mechanisms within a revenue recycling context will be performed.

6 References

Anas, A. and R. Lindsey (2011) Reducing urban road transportation externalities: Road pricing in theory and in practice, *Review of Environmental Economics and Policy*, **5**, 66–88, 06 2011.

Atasoy, B., T. Ikeda, X. Song and M. E. Ben-Akiva (2015) The concept and impact analysis of a flexible mobility on demand system, *Transportation Research Part C: Emerging Technologies*, **56**, 373 – 392, ISSN 0968-090X.

Basso, L. J. and S. R. Jara-Díaz (2012) Integrating congestion pricing, transit subsidies and mode choice, *Transportation Research Part A: Policy and Practice*, **46** (6) 890 – 900, ISSN 0965-8564.

Ben-Akiva, ., Moshe E. and S. R. Lerman (1985) *Discrete choice analysis: theory and application to travel demand / Moshe Ben-Akiva, Steven R. Lerman*, MIT Press Cambridge, Mass, ISBN 0262022176.

Christidis, P. and J. N. Ibáñez-Rivas (2012) Measuring road congestion, *Technical Report*, European Union. Institute for Prospective and Technological Studies, Joint Research Centre.

Guillaume-Gentil, S., M. Cazorla and A. Mousset (2015) Comptage periodiques de trafic ti - tc 2014.

Huang, H.-J. (2000) Fares and tolls in a competitive system with transit and highway: the case with two groups of commuters, *Transportation Research Part E: Logistics and Transportation Review*, **36** (4) 267 – 284, ISSN 1366-5545.

Huang, H.-J. (2002) Pricing and logit-based mode choice models of a transit and highway system with elastic demand, *European Journal of Operational Research*, **140** (3) 562 – 570, ISSN 0377-2217.

Kidokoro, Y. (2010) Revenue recycling within transport networks, *Journal of Urban Economics*, **68** (1) 46–55.

Levinson, D. (2010) Equity effects of road pricing: A review, *Transport Reviews*, **30** (1) 33–57.

Lyons, G., G. Dudley, E. Slater and G. Parkhurst (2004) Evidence based review: Attitudes to road pricing, *Project Report. Department for Transport*.

Pacheco, M., S. S. Azadeh, M. Bierlaire and B. Gendron (2017) Integrating advanced discrete choice models in mixed integer linear optimization, *Technical Report*, **TRANSP-OR 170714**, Transport and Mobility Laboratory, Ecole Polytechnique Fédérale de Lausanne.

Pigou, A. (1920) *The Economics of Welfare*, no. v. 1 in The Economics of Welfare, Macmillan and Company, Limited.

Robenek, T., Y. Maknoon, S. S. Azadeh, J. Chen and M. Bierlaire (2016) Passenger centric train timetabling problem, *Transportation Research Part B: Methodological*, **89**, 107 – 126, ISSN 0191-2615.

St-Louis, E., K. Manaugh, D. van Lierop and A. El-Geneidy (2014) The happy commuter: A comparison of commuter satisfaction across modes, *Transportation Research Part F: Traffic Psychology and Behaviour*, **26**, 160 – 170, ISSN 1369-8478.

Tsekeris, T. and S. Voss (2010) Public transport and road pricing: A survey and simulation experiments, *Public Transport*, **2**, 87–109, 05 2010.