Travel Demand Model with Heterogeneous Users and Endogenous Congestion: An application to optimal pricing of bus services*

Marco Batarce† **Marc Ivaldi**

Toulouse School of Economics

April 2011

Abstract

We formulate and estimate a structural model for travel demand, in which users have heterogeneous preferences and make their transport decisions considering the network congestion. A key component in the model is that users have incomplete information about the preferences of other users in the network and they behave strategically when they make transportation decisions (mode and number of trips). Therefore, the congestion level is endogenously determinate in the equilibrium of the game played by users. For the estimation, we use the first order conditions of the users' utility maximization problem to derive the likelihood function and apply Bayesian methods for inference. Using data from Santiago, Chile, the estimated demand elasticities are consistent with results reported in the literature and the parameters confirm the effect of the congestion on the individuals' preferences. Finally, we compute optimal nonlinear prices for buses in Santiago, Chile. As a result, the nonlinear pricing schedule produces total benefits slightly greater than the linear pricing. Also, nonlinear pricing implies fewer individuals making trips by bus, but a higher number of trips per individual.

*This paper was presented in the Kuhmo-Nectar Conference on Transport Economics, Valencia, 2010, as "Nonlinear Pricing in Transportation: An application to transit system of Santiago de Chile".

† Corresponding author. Email address: Marco-antonio.Batarce-moreno@sip.univ-tlse1.fr

1 Introduction

Traditionally, congestion effects have been considered explicitly in transportation analysis by means of network equilibrium models (e.g. Beckmann et al., 1956; De Cea and Fernández, 2001). In these models, the only effect of congestion is an increase in travel time and waiting time in the case of a bus network (e.g., Spiess, 1984; Spiess and Florian, 1989; De Cea and Fernández, 1993). Although the congestion modeling with network assignment models has been successful, its complexity prevents its use for the design of some transport policies. For example, to implement optimal pricing, generally it is used simplified approaches (De Borger, 2001; Van Dender and Proost, 2003; Ahn, 2009; Wang et al., 2008). Moreover, the network equilibrium approach do not consider other effects than time on the users' utility and neglects, for instance, any innate aversion to congestion or aversion to pollution produced by the congestion.

In this paper, we develop a methodology to estimate a travel demand model with endogenous congestion. In doing so, we recognize that users are heterogeneous in their preferences about transportation. This heterogeneity is modeled by adding an idiosyncratic parameter in the utility function, which is private information for each agent, but its distribution is common knowledge. In addition, we recognize that users make travel decisions considering the congestion level or the expected total number of trips in the network. Thus, individuals behave strategically and maximize their utility. They play an incomplete information game, in which the strategy is to decide the number of trips in each available transportation mode. In equilibrium, individual demand depends on the price faced by each user.

Our methodology takes into consideration two problems. First, the information asymmetry faced by the social planner designing a pricing policy. Indeed, the planner does not know the users' characteristics, which influence their transport preferences (e.g., income, subjective value of time, traveling distances, intrinsic aversion to congestion, etc.). This source of asymmetric information significantly affects transportation planning. Even though it has not been explicitly recognized in the practice of transportation planning[1], modeling methods use demand segmentation according to departure time, origin and destination, income and car ownership. This allows planners to partially control users' preferences for these attributes.

Second, users might take into account the congestion when making transportation decisions. Indeed, in order to decide how many trips to make, where to go, and which mode to use, users take into account the level of congestion in the city or in the route they use. However, they do not know the travel decisions made by other users. In that case, they play a game of incomplete information.

[1] For instance, Ortúzar and Willumsen (2002) do not mention problems of asymmetric information in transportation planning.

We apply our model to compute optimal nonlinear prices for buses in Santiago. The primary economic motivation for introducing optimal pricing is that it enhances economic efficiency. Urban transportation is not an exception to this rule. In particular, the provision of public transportation services also should be subject to optimal prices. Moreover, as Wilson (1993) remarks, nonlinear pricing can be used to maximize the consumer's net benefits from the firms' operations and therefore a nonlinear tariff minimizes allocative distortions caused by setting prices equal to marginal cost when the firm is a monopoly.

However, for many years, the economic literature has focused on how to price roads with marginal costs and to internalize the congestion (Dupuit, 1844; Pigou, 1920; Walters, 1961; De Borger, 2001; Lindsey and Verhoef, 2001, etc.). By contrast, since public transportation services do not exhibit this distortion[2], the discussion on optimal fares has been centered on the size of the subsidies and policy measures to reduce pollution and congestion (e.g., Timilsina and Dulal, 2008).

The paper is organized as follows. Section 2 presents the basic theoretical model and the required conditions for the implementation of a nonlinear pricing scheme (Guesnerie and Laffont, 1984). We derive optimal prices for the general case where the user's utility depends on the expected number of the trips on the network.

Section 3 presents the parameterization of the utility function. We adopt a discrete/continuous choice approach for the parameterization (Hanemann, 1984). This way, the demand functions derived from the parameterized utility are consistent with observed behavior, such as zero demand for a number of individuals in one or more modes of transportation. The individual heterogeneity is represented by an idiosyncratic parameter in the utility function. In addition, the congestion level is endogenously determinate as the equilibrium of an incomplete information game played by users. For the adopted parameterization, such equilibrium is the solution of a fixed point equation. We prove existence and uniqueness of the solution. The utility function also satisfies the implementability conditions for nonlinear tariffs (Guesnerie and Laffont, 1984).

Section 4 describes the statistical model that can accommodate the observed discrete/continuous choices. In order to derive the likelihood function, we use the first order condition of the individual's utility maximization problem (Kim et al., 2002). We assume random components in utility, which are associated to quality perception of the transport modes. Then, we specify a distribution function for these components and the corresponding likelihood function. Thus, we obtain an econometric model which is consistent with the microeconomic model. The heterogeneity is modeled as a random parameter and its distribution is specified parametrically. Doing so, we obtain an error component model or mixed model. The estimation procedure is based on Bayesian inference and the Markov

[2]In transportation science it is recognized that in the public transportation network there is congestion due to the limited capacity of the services (De Cea and Fernández, 1993).

Chain Monte Carlo (MCMC) method. In particular, the Markov Chain Monte Carlo is implemented to simulate the parameter posterior distribution. This allows us to avoid the integration of the likelihood function. In Section 4, we also study the model identification. In this respect, we conclude the model needs the normalization of one parameter to be identified.

Section 5 reports the results of the application to Santiago de Chile. We use data from Santiago to estimate the utility function and compute optimal prices, both linear and nonlinear. Concerning the model estimation, in line with the literature, we find that transportation demand exhibits low price elasticity (see Oum and Walter, 2000) and a significant congestion effect (see Bousquet and Ivaldi, 2001). Regarding the nonlinear pricing schedule, we obtain a significant quantity discount. The net benefits derived from a nonlinear schedule are slightly higher than those obtained with a linear price. Also, the total demand is higher under the nonlinear pricing schedule, but it exhibits a lower proportion of user traveling by bus.

Finally, in Section 6 we conclude and discuss limitations and possible extensions.

2 Theoretical model and optimal prices

2.1 Model and implementability conditions

Consider a pricing policy that recognizes the users are heterogeneous in their preferences. It is assumed that the regulator (or the planner) knows only the distribution of preferences in the relevant population. Therefore, the regulator cannot identify the characteristics of the user being served for the purpose of optimal price discrimination. The regulator must design a price mechanism in which users self-select according to their individual characteristics by the size of their purchase. Self-selection is induced by means of a quantity-dependent pricing schedule offered by the firm. The user faces a nonlinear outlay schedule $P(x)$, where x is the quantity consumed.

Assume that users have preferences depending on a vector of trips by mode of transportation, x; the expected total number of trips in the network (or the congestion level), X; and nonlinear outlay schedule, $P(\cdot)$. Individuals have unobservable characteristics, which are private information. They are represented by an idiosyncratic parameter, θ, continuously distributed with density function f and support $[\underline{\theta}, \overline{\theta}] \subseteq \mathbb{R}$. The preferences are summarized by a utility function $U = u(x, X, P; \theta)$. By assumption, the utility function is strictly increasing and concave on x, strictly decreasing on X and increasing in θ.

Consider a static framework where the user maximizes her utility choosing x during a fixed period of time. All users behave strategically and take into account the number of trips chosen by others in the network at the same period. They do not know the others' preferences, θ, with the exception of the distribution. The individual's problem is

$$\max_x u(x, X, P(x); \theta)$$
$$\text{s.t. } X = \int_{\underline{\theta}}^{\overline{\theta}} x(\mu, X, P(\cdot)) f(\mu) d\mu$$

The solution is a demand function, $x(\theta, X, P(\cdot))$, which depends on θ, X, and the payment schedule $P(\cdot)$. Notice that the expression for the total number of trips implies a fixed point equation. We discuss this feature in section 3.

The choice problem for the user may be recast in the following form. Rather than offering a quantity-dependent price schedule $P(x)$, the regulator can offer a quantity and payment schedule that depends on each user's declaration of her type, $(x(\theta), t(\theta))$. For any quantity-dependent schedule $P(\cdot)$, it is possible to construct the equivalent revelation schedule. Define (x, t) by $x(\theta) = x(\theta, X, P(\cdot))$ and $t(\theta) = P(x(\theta, X, P(\cdot)))$. Thus, the user has the incentive to reveal correctly her type θ, since (x, t) is constructed using the demand schedule.

A schedule (x, t) for which the user reveals her type parameter θ is referred to as a direct revelation mechanism. The requirement that truth telling is optimal for users is called incentive compatibility.

Formally, consider the utility defined as $U(x, X, t; \theta)$, where x is a vector of chosen trips, t is a monetary transfer, and θ is private information representing heterogeneity. A direct revelation mechanism is implementable if U satisfies the following conditions (Guesnerie and Laffont, 1984):

(M) *monotonicity*: U is strictly decreasing in the transfers;

(D) *differentiability*: U is continuously differentiable of C^2class;

(CS) *constant sign of marginal rate of substitution*: the sign of the vector $\frac{\partial}{\partial\theta}\left(\frac{\partial U/\partial x}{\partial U/\partial t}\right)$ remains constant; and

(B) *boundary behavior of the utility*: for any $(x, t, \theta) \in X \times [\underline{\theta}, \overline{\theta}], \exists K > 0$ such that for t large enough$\left\|\left(\frac{\partial U/\partial x}{\partial U/\partial t}\right)_{(x,t,\theta)}\right\| \leqslant K\,|t|$, uniformly in x, θ.

In addition, if the utility is quasi-linear in the monetary transfers in the form

$$U(x, t, X, \theta) = u(x, X, \theta) - t \tag{1}$$

then the direct mechanism $(x(\theta), t(\theta))$ is incentive compatible if it satisfies the following constraints (Guesnerie and Laffont, 1984)

$$\frac{\partial u}{\partial x}(x, X, \theta)\frac{dx}{d\theta}(\theta) = \frac{dt}{d\theta}(\theta) \;\; (IC1) \tag{2}$$

$$\frac{dx}{d\theta}(\theta) \geqslant 0 \;\; (IC2) \tag{3}$$

Users participating in the market must obtain positive net surplus. If the opportunity cost of nonparticipation is normalized to zero, then the individual rationality requires (x, t) to be such that

$$S(\theta) \equiv \max_{(x,t)} U(x, t, X, \theta) \geq 0$$

2.2 Optimal pricing schedule

In what follows, we consider that the utility function is separable in trips and expected congestion level. Therefore

$$u(x, X, \theta) = v(x, \theta) + xw(X)$$

where w measures the impact of congestion per trip. It satisfies $w(0) = 0,\ dw/dX \leq 0$.

To compute optimal nonlinear pricing schedule, the criterion is the maximization of social net benefits. We consider a Ramsey-type pricing schedule. The social planner obtains revenues from the operation of the firm and pays the production cost with costly public funds. This means government needs to collect $\$(1 + \lambda)$ in taxes to pay \$1 to the firm.

The aggregated consumer's surplus is

$$CS = \int_{\underline{\theta}}^{\overline{\theta}} S(\theta) f(\theta) d\theta = \int_{\underline{\theta}}^{\overline{\theta}} [u(x(\theta), X, \theta) - t(\theta)] f(\theta) d\theta$$

We consider that the production cost is composed by a fixed cost, C_0, and constant marginal cost, c. Therefore the producer's surplus is the following

$$PS = \int_{\underline{\theta}}^{\overline{\theta}} [t(\theta) - cx(\theta)] f(\theta) d\theta - C_0$$

With these ingredients, the social planner's problem is given by

$$\begin{aligned} &\max_{(x,t)} \{CS + (1 + \lambda) PS\} \\ &\text{s.t.} \quad \frac{\partial u}{\partial x}(x, X, \theta) \frac{dx}{d\theta}(\theta) = \frac{dt}{d\theta}(\theta) \\ &\qquad \frac{dx}{d\theta}(\theta) \geqslant 0 \\ &\qquad x(\theta) \geqslant 0 \end{aligned} \tag{4}$$

where the constraint $x(\theta) \geq 0$ is a feasibility condition.

To solve the problem, we define the variable $\xi(\theta) = \int_{\underline{\theta}}^{\theta} x(\mu) f(\mu) d\mu$. Using the incentive compatibility condition $(IC1)$, we eliminate the variable t in the problem (4). Then, the problem is transformed into

$$\max_x \int_{\underline{\theta}}^{\overline{\theta}} \left[v(x(\theta),\theta) + \frac{\lambda}{1+\lambda}\frac{\partial u}{\partial \theta}(x(\theta),\theta)H(\theta) - cx(\theta) \right] f(\theta) d\theta \tag{5}$$
$$+\xi(\overline{\theta})w(\xi(\overline{\theta})) - C_0$$
$$\text{s.t.} \quad \frac{d\xi}{d\theta}(\theta) = x(\theta)f(\theta)$$
$$\frac{dx}{d\theta}(\theta) \geqslant 0$$
$$x(\theta) \geqslant 0$$

where $H(\theta) \equiv (1 - F(\theta))/f(\theta)$ is the inverse of the hazard rate of θ.

Ignoring the constraint $\frac{dx}{d\theta}(\theta) \geqslant 0$ and considering the interior solution for x, the solution of the problem is given by the following equation

$$\frac{\partial v}{\partial x}(x,\theta) - \frac{\lambda}{1+\lambda}\frac{\partial^2 v}{\partial x \partial \theta}(x,\theta)H(\theta) - c + w(\xi(\overline{\theta})) + \xi(\overline{\theta})\frac{dw}{d\xi}(\xi(\overline{\theta})) = 0 \tag{6}$$

If the consumer does not take into account the impact of her trips on the congestion level, her optimal choice is such that

$$\frac{\partial v}{\partial x}(x,\theta) + w(X) = \frac{dP}{dx}(x) \equiv \pi(x) \tag{7}$$

From (6), there is a direct relationship between θ and x. Let $\theta(z)$ represent $\min\{\theta : x(\theta) = z\}$ (Spulber, 1989). The optimal marginal prices can be expressed as

$$\pi(x) = c + \frac{\lambda}{1+\lambda}\frac{\partial^2 u}{\partial x \partial \theta}(x,\theta(x))H(\theta(x)) + \xi(\overline{\theta})\frac{dw}{dX}(\xi(\overline{\theta})) \tag{8}$$

The marginal price departs from first best (marginal cost pricing) because of two sources of distortion: asymmetric information and congestion. The second term in the RHS is the source of consumers' informational rent. It is due to the incentive compatibility requirement. Indeed, the agent of type $\theta_2 > \theta_1$ can always pretend his type is θ_1, make $x(\theta_1)$ trips, pay the price $p(\theta_1)$ and thus get a positive utility. However, the agent of type θ_1 cannot gain anything by pretending to be type θ_2 because in doing so he gets a negative utility. The ability of higher types to behave as lower types is responsible for their informational rent. This rent is the price that the planner has to pay for higher types to reveal their information (for details, see Salanie, 1997; Laffont and Martimort, 2002). In particular, since $H(\theta)$ is decreasing, a higher θ implies a lower the rent extracted by the planner. The intuition behind such reduction is that it is more socially efficient to offer a lower marginal price to individuals who obtain a greater benefit for travel (recall we assume U increasing in θ). The last term in equation (8) is the marginal effect of the congestion in the utility. In other words, it is the marginal cost of congestion. Thus, an optimal pricing schedule internalizes congestion effects.

Here, three remarks are worth making. First, in contrast to standard models of network effects, trips are consumed in variable quantities by heterogeneous users. Therefore, the magnitude of the network effect depends on the total quantity consumed, rather than the total number of users in the network. Second, the user's utility due to network effects depends only on the number of trips and not on the user's type, such as in Sundararajan (2004). Third, in contrast with telecommunication models, the externality negatively impacts on the utility (such as Hahn, 2003).

3 Empirical model

3.1 Parameterization

In order to calculate nonlinear tariffs with congestion effects, the model needs to satisfy some conditions: consistency and tractability of the consumer's utility function and private information, together with the implementability of a nonlinear pricing schedule.

Regarding the first condition, corner solutions should be admissible. Indeed, some individuals in the sample choose more than one mode at day when they travel, but they do not use all available modes. A way of capturing this feature is by specifying a nonlinear separable utility function.

Concerning the second condition, the consumer must be able to maximize her utility and determine the (perceived) congestion level, under the assumption that the distribution of the private information is common knowledge. In other words, there must be an equilibrium.

We build on the parametric model of Bousquet and Ivaldi's (2001) paper. They model urban transportation demand, where users' preferences include private information. Furthermore, the level of utility depends on the total number of trips by car. This variable enters in the utility function through a quality index, which depends not only on the network congestion, but also on each alternative's attributes. Users choose the optimal number of trips for each alternative, given the monetary costs and quality (the congestion level). The authors obtain a closed form for demand functions and congestion.

A limitation of Bousquet and Ivaldi's (2001) paper is that the utility function is not well defined in nil consumption levels, given that it does not admit corner solutions (demand functions are positive for all price levels). This is not consistent with the users' observed behavior.

For a set of n available modes with alternative k being the only mode producing the externality, we define the utility function as follows

$$U(x,z) = \sum_{i=1}^{n} x_i(1 + \psi_i + \theta) - (x_i + 1)\ln(x_i + 1) + \gamma z \tag{9}$$

where

$$\psi_i = \alpha_i + \beta_i E(x_k) + \varepsilon_i \tag{10}$$

The variable x is a vector with non-negative components and represents the number of trips by mode. Thus, x_i corresponds to trips made by mode i. θ represents consumers' private information and it has support $[\underline{\theta}, \overline{\theta}]$. γ measures the weight of the consumption of a numeraire good z and equals the marginal utility of income.

The quality index ψ_i depends on observable and unobservable components. In (10), α_i represents the observable one. Typically, α_i captures attributes such as comfort, safety or reliability. The unobservable part is represented by ε_i and can be interpreted as taste variation across individuals or as subjective perception of quality. The standard assumptions are that individuals observe ε_i, for all i, but it is private information.

In addition, the quality index includes the effect of the congestion produced by the total demand of the alternative k, $E(x_k)$. More specifically, the k-th alternative is car and, therefore, the quality index depends on the expected value of trips by car. This statistics captures externalities induced by users choosing trips by car and endured by users of all of the other modes. Thus, it is assumed that only cars produce congestion. The parameter β_i measures the impact of such effect in the utility of the alternative i.

θ can be interpreted as a measure of the intrinsic utility obtained from making a trip. Transportation demand is derived from demand for activities (Ben-Akiva and Bowman, 1998), thus individuals choose to make a trip because they obtain a positive net utility by doing an activity at the end of the trip. Indeed, given the price per trip, p_i, the first order conditions of the utility maximization imply that x_i is positive only if θ is greater than $\gamma p_i - \psi_i$. The last term can be interpreted as a generalized cost for making a trip. Therefore, θ is a reservation utility for carrying out the activity that motivates the trip.

Given the interpretation of θ, the demand can be segmented according not only to travel modes but also to trip purpose. That segmentation is part of standard modeling techniques in transportation planning (Ortúzar and Willumsen, 2003) However, since we have trips data only for one day, there is not enough variability to estimate a demand system for trips by mode and purpose, and obtain an accurate estimation of θ for each purpose.

The demand functions are as follows

$$x_i = \max\{\exp(\theta + \psi_i - \gamma p_i) - 1, 0\} \tag{11}$$

From equation (11), $E(x_k)$ is the expectation of a truncated random variable and does not have a closed form. The existence of equilibrium in the individual's utility maximization problem and the congestion term are stated in a proposition below.

In the model, private information has dimension $N+1$ which is represented by the vector $(\theta, \varepsilon_1, ..., \varepsilon_N)$. However, it can be reduced to N. To do so, the utility can be written as follows $\eta_i = \theta + \varepsilon_i$. This way, it resembles the multiproduct nonlinear pricing model of Armstrong (1996). In general, the distribution of private information is given by the convolution of the distributions of θ and ε_i. In particular, we adopt normal distributions for θ and ε_i, for all i, therefore the distribution of η is also normal. The most important implication of such dimension reduction is that $E(x_k)$ is the expectation of a one-dimension random variable. Next, the proposition states the existence of an equilibrium in the game played by users.

Proposition 1 *Suppose that*

a) The utility function is given by expressions (9) and (10),

b) θ is distributed in $[\underline{\theta}, \overline{\theta}]$ with a density function f that is common knowledge; and

c) ε_k distributed in $[\underline{\varepsilon}_k, \overline{\varepsilon}_k]$ with a density function g_k that is common knowledge, and is independent of θ

Then, if $\eta_k \equiv \theta + \varepsilon_k$ and f_{η_k} denote its density with support $[\underline{\eta}_k, \overline{\eta}_k]$, the externality term, $E(x_k)$, is given by the solution of the equation

$$E(x_k) = e^{\alpha_k + \beta_k E(x_k) - \gamma p_k} \int_{\eta_k^*}^{\overline{\eta}_k} e^{\eta} f_{\eta_k}(\eta) d\eta - \int_{\eta_k^*}^{\overline{\eta}_k} f_{\eta_k}(\eta) d\eta \tag{12}$$

where

$$\eta_k^* = -(\alpha_k + \beta_k E(x_k) - \gamma p_k) \tag{13}$$

In addition, the solution always exists and is unique.

Proof. See Appendix A. ■

The parameterization also allows for individual heterogeneity in a discrete form. In Section 5, the model is estimated taking into account car availability as a source of heterogeneity. In that case, the parameters associated to the modes' quality are differentiated according to car availability. Thus, the joint distribution of all sources of private information is a mixture distribution. Given equation (10), we distinguish the j-th alternative α_j between individuals with available car (α_j^c) and those without car (α_j^{nc}). The probability of being of type κ, using the notation introduced in Proposition 1, is as follows

$$\begin{aligned} \Pr(\textit{being type } \kappa) &= \Pr(\eta_j = \kappa - \alpha_j^c) * \Pr(\textit{available car}) + \\ &\quad \Pr(\eta_j = \kappa - \alpha_j^{nc}) * \Pr(\textit{no available car}) \\ &= f_{\eta_j}(\kappa - \alpha_j^c) * q + f_{\eta_j}(\kappa - \alpha_j^{nc}) * (1 - q) \end{aligned} \tag{14}$$

where q is the probability of being an individual with car. The RHS in the last line of equation (14) is a mixture distribution. In the application presented later on, θ and ε_j are assumed normal, therefore the distribution of the individual type is a mixture of normal distributions.

3.2 Optimal nonlinear price

Next, we apply the results of Section 2 to the model with utility function given by equation (9) and considering only three alternatives: Mode 1 is car and mode 2 is bus and mode 3 is subway. In the first stage, equation (6) is solved on $x_2(\eta_2)$. Notice that we are only interested in the optimal prices for public transportation. The optimal number of trips in that mode is given by the following equation

$$x_2(\eta_2)= \exp\left\{\eta_2+\alpha_2+\beta_2 E(x_1)-\gamma c_2-\frac{\lambda}{1+\lambda}H(\eta_2)\right\}-1 \tag{15}$$

where $E(x_1)$ is given by the solution of the equation (12). Notice the demand function in equation (15) is the interior solution of the problem (5). Similarly to the demand in equation (11), the demand function (15) allows for corner solutions. Indeed, users do not make trips by mode 2 if his type $\eta_2^* = (\varepsilon_2+\theta)^*$ is such that $(\eta_2^*+\alpha_2+\beta_2 E(x_1)-\gamma c_2-\frac{\lambda}{1+\lambda}H(\eta_2^*)) = 0$. The complete solution for the optimal demand function is $x_2^*(\eta_2) = \max\{x_2(\eta_2), 0\}$.

The result for x_2 is used to obtain the marginal price, $\pi_2(\eta_2)$ (given 8)

$$\pi_2(x_2(\eta_2)) = c_2 + \frac{\lambda}{1+\lambda}\frac{H(\eta)}{\gamma} \tag{16}$$

The congestion does not have any impact on the optimal marginal price, since it is the automobile that produces the externality. This is also a result of the linearity and separability assumed in the utility function.

$x_2^*(\eta_2)$ satisfies the monotonicity constraints (condition $(IC2)$), given the functional form used here.

Finally, the quasi-linear form of the utility simplifies the computation of the nonlinear prices. However, it rules out income effects. The separability of the utility implies that the demand cross elasticities are zero. We retain an additive utility structure under the assumption that the utility derived from one mode is independent of others. The utility function needs to have decreasing marginal returns to capture satiation.

4 Estimation

In this section, we specify a statistical model that can accommodate both interior and corner solutions. To estimate the parameters, we use Bayesian inference and the Markov Chain Monte Carlo (MCMC) method.

4.1 Likelihood specification

Given the utility function in equation (9), the first order conditions for the maximization problem subject to the budget constraint, respect to the alternative i, are

$$\frac{\partial U}{\partial x_i}(x) = \psi_i + \theta - \ln(x_i + 1) = \lambda p_i \;\; if \; x_i > 0 \tag{17}$$
$$\frac{\partial U}{\partial x_i}(x) = \psi_i + \theta - \ln(x_i + 1) < \lambda p_i \;\; if \; x_i = 0 \tag{18}$$

where λ is a Lagrange multiplier and p_i is the observed price of mode i. If the numeraire good is always consumed, λ is equal to γ (the marginal utility of income). Using the quality index given by equation (10) and rearranging terms, we obtain

$$V_i \equiv \alpha_i + \beta_i E(x_k) + \theta - \ln(x_i + 1) - \gamma p_i = -\varepsilon_i \;\; if \; x_i > 0 \tag{19}$$
$$V_i \equiv \alpha_i + \beta_i E(x_k) + \theta - \ln(x_i + 1) - \gamma p_i < -\varepsilon_i \;\; if \; x_i = 0 \tag{20}$$

Equations (19) and (20) are similar to those appearing in standard choice models, where marginal utility is constant and does not depend on the quantity demanded. However, here marginal utility depends on quantities because of the non-linearity in the utility function. The goal is to derive the distribution of observed demand, x. In order to do that, we specify a distribution for the vector of taste variations, ε. We use equations (19) and (20) to obtain the distribution of x by applying the change-of-variable theorem.

If ε follows a multivariate normal distribution with mean zero and variance Ω, the likelihood function of the data is a mixture of density ordinates (equation 19) and point masses (equation 20) corresponding to non zero and zero demand, respectively. Suppose that there are n transportation modes and the first m alternatives have non zero demand, the likelihood for one individual is given by

$$\begin{aligned} l(x) &= \Pr(x_i > 0, x_j = 0; i = 1, .., m, j = m+1, .., n) \\ &= \int_{-\infty}^{-V_n} ... \int_{-\infty}^{-V_{m+1}} \phi(-V_1, .., -V_m, \varepsilon_{m+1}, .., \varepsilon_n | 0, \Omega) \, |J| \, d\varepsilon_{m+1}, .., d\varepsilon_n \end{aligned} \tag{21}$$

where ϕ is multivariate normal density, $V_i = V_i(x, p)$ and J is the Jacobian, that is $J_{ij} = \partial V_i(x,p)/\partial x_j$ with $i, j = 1, ..., m$, which appears because of the change of variable.

Since the multidimensional integral of the likelihood function (21) cannot be evaluated directly, it is transformed to the product of two factors -following Kim et al. (2002). The vector of taste variations is decomposed in $\varepsilon_a = (\varepsilon_1, ..., \varepsilon_m)'$ and $\varepsilon_b = (\varepsilon_{m+1}, ..., \varepsilon_n)'$, such that

$$\begin{bmatrix} \varepsilon_a \\ \varepsilon_b \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Omega_{aa} & \Omega_{ab} \\ \Omega_{ba} & \Omega_{bb} \end{bmatrix} \right)$$

with ε_a and $\varepsilon_b|\varepsilon_a$ being normally distributed. $\varepsilon_a \sim N(0, \Omega_{aa})$ and $\varepsilon_b|\varepsilon_a = V_a \sim N(\mu, \Sigma)$ where $\mu = \Omega_{ba}\Omega_{aa}^{-1}V_a$, $\Sigma = \Omega_{bb} - \Omega_{ba}\Omega_{aa}^{-1}\Omega_{ab}$ and $V_a = (V_1, ..., V_m)'$. Therefore, the likelihood function (21) can be rewritten as

$$\begin{aligned} l(x) &= \Pr(x_i > 0, x_j = 0; i = 1, .., m, j = m+1, .., n) \qquad (22) \\ &= \phi_{\varepsilon_a}(-V_1, .., -V_m | 0, \Omega_{aa}) \, |J| \\ &\quad \cdot \int_{-\infty}^{-V_n} ... \int_{-\infty}^{-V_{m+1}} \phi_{\varepsilon_b|\varepsilon_a}(v_{m+1}, .., v_n | \mu, \Sigma) \, |J| \, dv_{m+1}, .., dv_n \end{aligned}$$

4.2 Inference

Our interest is on the point estimation, together with the distribution of θ. Remember that θ and ε capture individuals' heterogeneity regarding preferences for transportation, equation (9). However, due to identification restrictions we are only able to estimate the distribution of θ.

We estimate a random parameter model in which θ distributes according to an unknown distribution function, which is specified parametrically.

The Bayesian approach is a natural way to estimate a random parameter model, since it considers the parameters of interest as random variables and it investigates their posterior distributions given prior information. Moreover, the Bayesian method avoids integrating the likelihood function. In contrast, maximum likelihood estimation requires integrating out the likelihood function over the distribution of the random parameters, while simulated maximum likelihood requires that the number of simulations increases with the number of observations in order to reduce the simulation bias (Train 2003, Gourieroux and Monfort, 2003).

To this end, we formulate a Bayesian hierarchical model (Carlin and Louis, 1996; Gill, 2002) for an individual h ($h = 1, ..., H$) as follows

$$\begin{aligned} x_h &\sim l(x_h | p_h; \zeta, \Gamma, \theta), \text{ likelihood function eq. (22)} \\ (\zeta, \Gamma) &\sim N(\nu, \Sigma) \\ \theta_h &\sim N(\mu_\theta, \sigma_\theta) \\ \mu_\theta &\sim N(m, s) \\ \sigma_\theta^2 &\sim IG(a, b) \end{aligned}$$

where $\zeta = (\alpha_1, \beta_1, ..., \alpha_n, \beta_n, \gamma)$ is a vector of parameters of the structural model in equation (9) and Γ is a vector with the elements of the Cholesky decomposition of Ω. We also assume the random parameter θ_h is normally distributed with parameters μ_θ and σ_θ, which are the parameters of interest. Likewise, μ_θ distributes normally with mean m and variance s^2, and σ_θ distributes according to an inverted gamma with parameters a and b. These priors for μ_θ and σ_θ allows to obtain closed expressions for the posteriors, given a sample of the individual parameter, θ (Train, 2003).

The posterior distribution of the parameters is obtained by the Markov Chain Monte Carlo method. In particular, we use a Gibbs sampling method with a Metropolis step to infer the posterior distribution of the model's parameters (Geweke, 2005). An advantage of the MCMC method is that it allows us to obtain individual-level estimates for θ_h. Indeed, with the MCMC method we can draw a sample from the posterior distribution of θ_h conditional on $(x_1, p_1; \zeta, \Gamma, \mu_\theta, \sigma_\theta)$. Then, according to Von-Mises's theorem, the mean of this sample is a consistent estimator of θ_h, in the sense of classical estimation (see Train 2003).

We summarize the simulation procedure, for details see Kim et al. (2002). First, choose initial parameters ζ^0, μ_θ^0, σ_θ^0, Γ^0 and hyper-parameters ζ_0, Σ_{ζ_0}, m, s, a, and b. Compute de congestion effect, $E(x_k)$, for all h, given the initial parameters. Then, repeat for $t = 1, ..., (T_0 + T_1)$ the following

1. Draw θ_h^t, $h = 1, ..., H$, according to random walk Metropolis-Hasting sampling, and using the individual likelihood, $l(x_h|p_h; \zeta^{t-1}, \Gamma^{t-1}, \theta_h^{t-1})$, given θ_h^{t-1}, ζ^{t-1} and Γ^{t-1}

2. Compute μ_θ^t accorging to the posterior distribution given the sample θ_h^t, $h = 1, ..., H$, σ_θ^{t-1}, and the prior $N(m, s)$

3. Compute σ_θ^t accorging to the posterior distribution given the sample θ_h^t, $h = 1, ..., H$, μ_θ^t, and the prior $IG(a, b)$

4. Draw (ζ^t, Γ^t) according to random walk Metropolis-Hasting sampling, and using the total likelihood, $\prod_h l(x_h|p_h; \zeta^{t-1}, \Gamma^{t-1}, \theta_h^t)$, given θ_h^{t-1}, ζ^{t-1} and Γ^{t-1}

5. Compute the congestion efect, $E(x_{hk}|p_h; \zeta^t, \Gamma^t, \theta_h^t, \mu_\theta^t, \sigma_\theta^t)$, solving the equation (12) for $h = 1, ..., H$

6. Update $t = t + 1$ and go to step 1.

Finally, the estimators are computed as the means and variances of the last T_1 drawn parameters.

4.3 Identification

To verify our model is identified we consider only two available alternatives: mode 1 is car and mode 2 is bus. Trips by car are the only ones producing congestion. Individuals travelling by car or bus undergo the effects of congestion. In addition, we consider that the vector components of taste variations are independent, although they have different variance, denoted by σ_{ε_i}. We do not loss generality using two alternatives because the in case with more modes, which do not produce congestion, we can identify the especific parameters similarly to the bus parameters.

The identification is not clear from the frequentist point of view. The problem arise because the congestion measure, $E(x_1)^3$, is not only a nonlinear function of some structural parameters (α_1, β_1) and the price p_1, but also it is a function of the parameters of the random component distributions $(\sigma_{\varepsilon_1}, \mu_\theta, \sigma_\theta)$. By contrast, according to the Bayesian approach, the model is identified. In fact, Kass et al. (1998) asserts there is no identification problem for MCMC methods, provide the posterior is proper. In turn, Geweke (2005) states if the parameter is identified, its mean and variance are identified. However, these assertions are not free of controversy (for example, see San Martin and González, 2010). We argue the model is identified because the MCMC method focuses in the data generating process not marginalized with respect to θ_h (San Martin and González, 2010). Therefore, we can introduce additional conditions on the parameters which allow us to identify the model.

Considering the model as a system of equations where the depended variable is $y_i \equiv \ln(x_i + 1)$, $i = 1, 2$, and the independent variables are p_i and $E(x_1)$ in equation (19). Therefore, under the distributional assumptions of the previous sections, the statistical model conditional on θ_h for an individual h choosing both modes is given by

$$\begin{aligned} P(y_h|p_h;\zeta,\theta_h,\sigma_{\varepsilon_1},\sigma_{\varepsilon_2},\mu_\theta,\sigma_\theta) \quad = \quad & \frac{1}{\sigma_{\varepsilon_1}}\phi\left(\frac{y_{h1}-(\alpha_1+\theta_h+\beta_1 E(x_1)-\gamma p_{h1})}{\sigma_{\varepsilon_1}}\right)\cdot \\ & \frac{1}{\sigma_{\varepsilon_2}}\phi\left(\frac{y_{h2}-(\alpha_2+\theta_h+\beta_2 E(x_1)-\gamma p_{h2})}{\sigma_{\varepsilon_2}}\right) \end{aligned} \tag{23}$$

$$E(\theta_h|\mu_\theta,\sigma_\theta) = \mu_\theta \tag{24}$$

$$Var(\theta_h|\mu_\theta,\sigma_\theta) = \sigma_\theta^2 \tag{25}$$

where ϕ is the standard normal density function.

We say the model is identified if for two vectors $(\zeta,\theta_h,\sigma_{\varepsilon_1},\sigma_{\varepsilon_2},\mu_\theta,\sigma_\theta)$ and $(\widetilde{\zeta},\widetilde{\theta}_h,\widetilde{\sigma}_{\varepsilon_1},\widetilde{\sigma}_{\varepsilon_2},\widetilde{\mu}_\theta,\widetilde{\sigma}_\theta)$ such that

$$P(y_h|p_h;\zeta,\theta_h,\sigma_{\varepsilon_1},\sigma_{\varepsilon_2},\mu_\theta,\sigma_\theta) = P(y_h|p_h;\widetilde{\zeta},\widetilde{\theta}_h,\widetilde{\sigma}_{\varepsilon_1},\widetilde{\sigma}_{\varepsilon_2},\widetilde{\mu}_\theta,\widetilde{\sigma}_\theta) \tag{26}$$

for all (y_h, p_h), then

$$(\zeta,\theta_h,\sigma_{\varepsilon_1},\sigma_{\varepsilon_2},\mu_\theta,\sigma_\theta) = (\widetilde{\zeta},\widetilde{\theta}_h,\widetilde{\sigma}_{\varepsilon_1},\widetilde{\sigma}_{\varepsilon_2},\widetilde{\mu}_\theta,\widetilde{\sigma}_\theta)$$

Since the model is parametric and similar to a linear regression, it is straightforward to verify the identified parameters are γ, α_2, β_2, σ_{ε_1}, and σ_{ε_2}. We cannot identify α_1 and θ_h separately, thus we normalize $\alpha_1 = 1$. This normalization is also useful to identify the remaining parameters. Identification of μ_θ and σ_θ is due to the indetification of θ_h for all h. In fact, using the conditions (24) and (25) we

[3]For the sake of exposition we do not write the conditioning variables of the expectation of x. Hereafter, it should bear in mind this notation.

have

$$\begin{aligned}\mu_\theta &= E(\theta_h|\mu_\theta,\sigma_\theta) = E(\theta_h|\widetilde{\mu}_\theta,\widetilde{\sigma}_\theta) = \widetilde{\mu}_\theta \\ \sigma_\theta^2 &= Var(\theta_h|\mu_\theta,\sigma_\theta) = Var(\theta_h|\widetilde{\mu}_\theta,\widetilde{\sigma}_\theta) = \widetilde{\sigma}_\theta^2\end{aligned}$$

Therefore, we identify β_1 from the following equation

$$\beta_1 E(x_1|p_{h1};\beta_1,\mu_\theta,\sigma_\theta,\sigma_{\varepsilon_1}) = \widetilde{\beta}_1 E(x_1|p_{h1};\widetilde{\beta}_1,\mu_\theta,\sigma_\theta,\sigma_{\varepsilon_1})$$

As $E(x_1)$ is a monotone function of β_1 (we prove this in appendix), the last equation implies $\beta_1 = \widetilde{\beta}_1$.

These results impose a requirement on the data. Indeed, to identify the parameters, it is necessary that all travelling alternatives are chosen by a subset of individuals in the sample. This does not mean that individuals have to make trips in all available modes. However, it implies that the likelihood function should not be only composed by point masses for any alternative. This enables us to apply eq. (19). Finally, for identification, the individuals' choice set should include an alternative unaffected by the congestion and with a price ticket varying across the sample.

Finally, we need to impose some restrictions on the original parameters in order to obtain theoretically consistent estimates (Train and Sonnier, 2003). These restrictions may be seen as prior information coming from the theory. The transformations are

$$\beta_i < 0 \Rightarrow \beta_i = -\exp(\overline{\beta}_i) \tag{27}$$

$$\gamma > 0 \Rightarrow \gamma = \exp(\overline{\gamma}) \tag{28}$$

$$\sigma_{\varepsilon_i} > 0 \Rightarrow \sigma_{\varepsilon_i} = \exp(\overline{\sigma}_{\varepsilon_i}) \tag{29}$$

There are no requirements on α_i, therefore $\overline{\alpha}_i = \alpha_i$. Condition (27) implies that the congestion has a negative effect in the utility. Equation (28) comes from the fact that γ represents the marginal utility of income, which is positive. Finally, condition (29) is required because σ_{ε_i} is the standard deviation of ε_i. Then, the vector ζ distributes according to a normal with mean ζ_0 and variance matrix Σ_{ζ_0}.

4.4 Alternative estimation approaches

Bousquet and Ivaldi (2001) assume that there are two types of simultaneous individual decisions. Given the congestion level and the travel cost, the user chooses the number of trips in each mode of transportation. The optimal number of trips is used to obtain the conditional indirect utility function. The choice of transportation modes, x, is determined by this conditional maximum level of utility. The authors also assume that the number of trips distributes double Poisson and the indirect utility functions are observed with a certain degree of error, which distributes according to an Extreme Value distribution Type I, producing choice probabilities with a multinomial logit form. This approach corresponds to the discrete/continuous demand modeling (Hanemann, 1984). However, it does not take into account the relationship between the distribution of the error term in the utility function (in the discrete choice level) and the distribution of the quantity consumed (in the continuous decision level), in contrast to Hanemann (1984), and Dubin and McFadden (1984).

Nevertheless, a correctly specified discrete/continuous model according to Hanemann (1984) is not useful because the modes are not perfect substitutes. Indeed, there are individuals in the sample who choose simultaneously two or more transportation modes, but the Hanemann's models consider situations where only one alternative is chosen.

Finally, another equivalent approach is the estimation of a system of demands with truncated regression. However, two remarks should be made. On the one hand, if the utility function specification does not allow for an analytic expression of the demand function, the approach is not useful. On the other hand, a flexible demand function may not be totally, theoretically consistent with the underlying utility function. By contrast, the approach adopted in this paper satisfies theoretical consistency.

5 Application to Santiago

This section presents the estimation results using data from Santiago de Chile and the computation of nonlinear prices. We assume that there are three available transportation modes: car, bus, and subway. Cars are supposed to produce congestion in the road network, which is undergone by both car and bus users. The optimal nonlinear prices are calculated for public transportation applying the results of Section 4.

5.1 Data

We use data provided by the travel survey "Encuesta Origen-Destino de Viajes de Santiago" ("Origin and Destination Travel Survey of Santiago"), carried out by the National Agency of Transportation Planning (Sectra, 2002). A total of 15,537 households were surveyed, out of which 12,346 correspond to surveys conducted during the normal season and 3,191 to the summer season. The survey gathers information about households (e.g. number of individuals in the household, total income, number of vehicles, etc.), household's members (e.g. age, type of job or study, driving license, etc.), and trips made during a day by each member of the household (departure time, arrival time, origin, destination, mode, price ticket, parking cost, walking time and distance, etc.). Some descriptive statistics for the sample are provided in Tables 1 and 2.

Table 1

Characteristics of households and persons in the survey

Households	14,152	
Mean household size	4.0	pers/hh
Mean household income	515.9	Ch$ $\cdot 10^3$
Mean income per capita	128.9	Ch$ $\cdot 10^3$
Households with car	33	%
Mean cars per household	0.4	car/hh
Persons	43,692	
Workers	44.9	%
Students	29.6	%
With driving license	26.0	%

Table 2

Characteristics of the trips in the sample

	Trips	Trips/person	Mean travel time (min.)	Mean distance (km.)
All modes	147,872	3.39	26.5	4.7
Car	20,345	3.67	27.3	6.4
Bus	40,244	2.12	45.7	8.4
Subway	4,805	1.75	43.1	9.5
Walking	56,573	2.94	11.1	0.8
Bicycle	3,343	2.44	18.7	2.3

To estimate, we consider only information gathered in the normal season from individuals making at least one trip by car, bus, or subway, except students. In addition, we exclude all information coming from combining modes. This results in a sample of 10,866 individuals and 36,613 trips, from which 10,245 are made by car, 16,907 by bus, 2,034 by subway, and 6,579 by non-motorized modes (walking and/or bicycle). For estimation, we also use expansion factors in order to represent the population. The expanded sample represents 1,693,591 individuals.

The price of a car trip is calculated as a function of travel distance, gasoline consumption per kilometre and fuel price. For individuals in the sample making trips by car, the price of a trip was calculated as the average cost. For those who do not report trips by car, the price is estimated as a function of the average travel distance by bus.

The price of the bus ticket is included in the data for those individuals that report travelling by bus. For individuals that declare not to travel by bus, the price ticket was imputed.[4]

5.2 Parameter estimation

In addition to the heterogeneity represented by θ and ε, the model includes a discrete source of heterogeneity to capture differences between individuals with availability of car or not. Individuals with different choice sets value the alternatives in a different scale. Typically, the transportation demand analysis has segmented the demand according to car ownership or availability (Ortúzar and Willumsen, 2003; Ben-Akiva and Lerman, 1985), which implies estimating a model for each segment and assuming that the social planner (who uses the model for policy design) is able to segment the population. The approach adopted here does not need to segment the population because car availability is private information and its distribution is estimated only.

The set of parameters of the utility function differs if the individual has an available car or not. Consider the dichotomous variable $\delta_{car} \in \{0, 1\}$, which takes value one for individuals with available car and zero otherwise. The models to estimate are reflected on the following equations

$$
\begin{aligned}
U(x,z) \quad &= \quad \sum_{i=1}^{n} x_i(1+\psi_i+\theta) - (x_i+1)\ln(x_i+1) + \gamma z \\
\text{Model 1} \quad &: \quad \psi_i = \alpha_i + \varepsilon_i \qquad (30) \\
\text{Model 2} \quad &: \quad \psi_i = \delta_{car}(\alpha_i^c + \beta_i^c E(x_k)) + (1-\delta_{car})(\alpha_i^{nc} + \beta_i^{nc} E(x_k)) + \varepsilon_i \qquad (31)
\end{aligned}
$$

Thus, the model is estimated with different quality indexes for bus and subway according to car availability.

The distribution of car availability in the population is not estimated simultaneously with the parameters of the model, but it is estimated from a sample of individuals. Since the distribution of such a variable is given by the fraction of individuals with car in the population, the same fraction in the sample is a consistent estimator.

As a first analysis, we especify a general variance matrix for ε, but the Markov chain did not converge. This is due to the data do not allow us to indentify all the parameters in the matrix, mainly because there is no individual choosing the three availables modes. Similar fails in identification occur with error component logit-mixture models (Walker et al., 2007). Thus, we normalize the variance matrix of ε with the restriction $\sigma_{\varepsilon_1} = 1$.

The results for the two models are presented in Table 3. Model 1, in the first two columns, does not include the congestion effect, but it includes discrete heterogeneity due to car availability. Therefore, the parameters of the quality index are differentiated with respect to that characteristic. In the third

[4] This price is flat (it does not vary with distance nor with time schedules) and only unitary tickets are charged.

and fourth columns, Model 2 considers the discrete heterogeneity and congestion effects. Table 3 also includes the likelihood ratio tests under the null hypotheses: (H_0^1) no congestion effects (Model 1 vs. Model 2). Regarding the final value of the likelihood function, Model 2 is better than Model 1. The likelihood ratio tests reject both null hypotheses with high confidence level.

The standard errors of the parameters β confirm the hypothesis that individuals take into account congestion when making their decisions. The same is true with the likelihood ratio test for the hypothesis H_0^1.

Table 3

Estimated parameters of utility function

	Model 1		Model 2	
	Param.	Std. err.	Param.	Std. err.
μ_θ	1.2747	0.00041	1.8546	0.00062
σ_θ	0.0304	0.00004	0.0286	0.00004
γ	0.0027	0.00000	0.0057	0.00001
α_1	0	-	0	-
α_2^{car}	-0.6753	0.00063	-0.4935	0.00079
α_2^{nocar}	-0.1331	0.00051	-0.4002	0.00047
α_3^{car}	-3.3873	0.00187	-3.8893	0.00076
α_3^{nocar}	-2.7897	0.00155	-3.3368	0.00152
β_1			-0.3390	0.00029
β_2^{car}			-0.4778	0.00064
β_2^{nocar}			-0.1548	0.00016
Γ_{11}	1	-	1	-
Γ_{21}	-0.1564	0.00017	-0.1518	0.00015
Γ_{22}	0.4217	0.00014	0.4228	0.00014
Γ_{31}	-0.9378	0.00081	-0.9902	0.00073
Γ_{32}	-1.1795	0.00065	-1.2101	0.00047
Γ_{33}	0.1981	0.00036	0.2292	0.00028
Sim. log-lik.	-132,013		-131,866	
LRT(H_0^1) (Model 1/Model 2)				293

Table 4 shows the variance matrix of ε in the case of Model 2. This matrix is computed with the estimates of the Cholesky decomposition, Γ.

Model 2 has a marginal utility of income (parameter γ) consistent with the literature. Indeed, according to this specification, the demand price elasticities (in absolute value) are 0.211, 0.165, and

0.171 for car, bus and subway, respectively. They are calculated using γ and the sample mean of the prices, since the elasticities are linear in such variables. Oum and Waters (2000) summarize demand elasticities for transport of passengers estimated in several studies. They report demand elasticities for trips by automobile in the range of (in absolute value) 0.00-0.52 and for trips by bus in the range of 0.01-0.96.

The estimation results are in line with what is expected ex ante. For example, the values of the parameters α, which represent the perceived quality of each alternative, imply that car is the best quality mode and subway is the worst mode. However, note that these parameters take into account access time as part of the unobservable quality. Since the subway network is small with respect to the bus network, the subway access time negatively affects the quality of the service. This feature is present in the two models. Regarding the value of β in Model 2, the congestion matters more for individuals with available car.

Heterogeneity is statistically significant (see the value of standard error of σ_θ). However, the value of the coefficient estimated is very small. This may be due to the fact that we have fixed the variance of ε_1 equal to one. The consequence of such a parameterization is that we impose a total variance equal to $(1+\sigma_\theta^2)$[5]. The estimated σ_θ may imply that the heterogeneity in the population is less or equal to one.

Finally, we compute the correlation between the observed trips and the modeled ones. In the case of trips by car, the correlation is 0.64. For bus trips the correlation is 0.43. Figures 1 and 2 show the observed and modeled trips for car and bus, respectively.

Table 4

Variance matrix of ε

	ε_1	ε_2	ε_3
ε_1	1.000		
ε_2	-0.152	0.202	
ε_3	-0.990	-0.361	2.497

5.3 Optimal prices

Nonlinear prices for public transportation are calculated based on the former Model 2. Given that the selected model includes three sources of private information, the distribution of an arbitrary type, κ, in the population is given by the mixture distribution in equation (14).

Since the parameters θ and ε_2 are assumed normally distributed with support on $\mathbb{R}$, κ -which represents private information- is also distributed on $\mathbb{R}$. The numerical solution of equations (15) and

[5] Recall that the heterogeneity is represented by the parameters θ and ε.

(16) are badly behaved in the tail of the distribution, which results in the need to define a closed interval for the support of κ. Based on some robustness tests, we define the support as $[\mu_\theta - 3.5(\sigma_\theta + \sigma_{\varepsilon_2}),$ $\mu_\theta + 3.5(\sigma_\theta + \sigma_{\varepsilon_2})]$. This interval leaves out only 0.1% of individuals. Thus, we neglect the impact of changing the support of κ.

The planner's problem requires knowing the marginal cost of a trip by public transportation in Santiago. According to the National Agency of Transportation Planning, the average cost per trip is Ch$215[6] (Sectra, 2002). If we consider that the capacity of the system is constant, the marginal cost per passenger is small compared to fixed costs (capital cost of the vehicle, fuel cost, driver's wage). We assume a marginal cost equal to 20% of the average cost. We also need to assume a cost of public funds, $(1 + \lambda)$, which is set in 1.5.

In order for (15) to satisfy the non negativity constraint in the planner's problem (5), we need to find κ^* such that $x(\kappa^*) = 0$.[7] We define the solution $\tilde{x}(\kappa) = \max(0, x(\kappa^*))$. In this case, $\kappa^* = 0.69$.

Implementing optimal nonlinear tariffs requires taking into account that trips are discrete. Therefore, the optimal price is estimated for a menu of bus tickets. Table 5 presents the final optimal menu. The actual price ticket in the bus system of Santiago is approximately 290 (Ch$). Therefore, the optimal nonlinear prices are far higher than the fare actually charged. The optimal linear prices is 433 (Ch$). These pricing schedules seem to be politically infeasible.

Table 5

Optimal menu of bus tickets ($\lambda = 0.5$)

Ticket	Price (Ch$)
1 trip	820
2 trips	1,250
3 trips	1,550
4 trips	1,780
Daily	1,970

We compare the benefits of actual and optimal pricing schedules, both linear and nonlinear. Table 6 shows the mean of the expected benefits, consumer surplus and net revenues. Regarding total benefits, nonlinear prices are better with respect to actual ones. However, regarding consumer surplus, the results are reversed. The actual linear price is better. This reveals that the high total benefits obtained with nonlinear prices are mainly due to the net revenues from the operation of the firm.

Under a public financing scheme, having positive net revenues is equivalent to collecting public funds using the transit system. Therefore, the pricing schedule becomes a tax, which is not designed

[6] Ch$ denotes the Chilean currency, Peso. In 2001, 1 US$ = 630 Ch$

[7] In the genal case, this require solving an optimal control problem with bounded control (Kamien and Schwartz, 1991)

for that purpose. From a mathematical point of view, it is inconsistent that the planner's budget constraint does not bind and the Lagrange's multiplier, λ, is different from zero. Therefore, in the planner's problem, the net revenues have a weight higher than the optimal one. In order to overcome this inconsistency, we solve the planner's problem (4) considering λ as a variable. In the optimum, λ should be non zero and the budget constraint should be binding.

Table 6

Benefits under different pricing schedules ($\lambda = 0.5$)

Pricing schedule	Total benefits (Ch$)	Consumer surplus (Ch$)	Net revenues (Ch$)
Marginal-cost price	2,197	2,635	-292
Actual linear price	2,413	2,028	257
Optimal linear price	2,443	1,727	477
Optimal nonlinear prices	2,563	1,497	711

In this case, the non negativity constraint implies that the $\kappa^* = 0.32$, such that $x(\kappa^*) = 0$. The Lagrange multiplier is $\lambda = 0.036$. Figure 3 in the upper panel shows the optimal number of trips as a function of the individual's type, together with different pricing schedules as a function of number of trips (lower panel). Table 7 shows the menu of optimal prices.

Table 7

Optimal menu of bus tickets ($\lambda = 0.036$)

Ticket	Price (Ch$)
1 trip	280
2 trips	390
3 trips	460
4 trips	530
Daily	690

We solve the planner's problem for a linear price. The optimal price in this case is 162 (Ch$) and the Lagrange's multiplier is $\lambda = 0.105$.

Again, we compare total benefits, consumer surplus, and net revenues for both the optimal nonlinear and linear pricing schedule (see Table 8). Nonlinear prices are slightly better with respect to linear ones. Even though both nonlinear and linear pricing produce total benefits similar to marginal cost pricing, they do not require subsidies. However, if we consider the value of public funds is 1.5, in the case of marginal cost pricing, the total benefits reduce to 2,194 (Ch$). This implies that with optimal pricing schedules we obtain total benefits 6% higher. Moreover, with optimal pricing the consumers' surplus is 15% higher than the actual pricing.

In addition, we compare the total demand under different pricing schedules, taking into account the demand for public transportation and the number of individuals choosing public transportation. Table 9 shows the demands under linear and nonlinear pricing.

There is a trade-off between economic efficiency and exclusion of the market. Indeed, the optimal nonlinear price exhibits the highest benefits and the lowest participation. Despite the fact that demand in the linear pricing schedule is lower than in the nonlinear one, the exclusion is higher in the latter case. This means there are more trips per user under the nonlinear pricing.

Table 8

Benefits under different pricing schedules and λ is optimally determined

Pricing schedule	Total benefits (Ch$)	Consumer surplus (Ch$)	Net revenues (Ch$)
Marginal-cost price	2,330	2,632	-292
Actual linear price	2,291	2,023	256
Optimal linear price	2,326	2,326	0
Optimal nonlinear prices	2,336	2,336	0

Table 9

Demand under different pricing schedules and λ is optimally determined

Pricing schedule	Mean of trips	Individuals choosing bus (%)
Marginal-cost price	2.70	100
Actual linear price	2.22	99
Optimal linear price	2.46	99
Optimal nonlinear prices	2.62	98

6 Final comments

This methodology to model transportation demand with endogenous congestion should be considered as a first step in this type of models. It may be extended to consider more complex behavior. For instance, it is possible to model mode and destination choice under a discrete choice framework with endogenous congestion. This way, the model would be in line with traditional transportation demand modeling. Also, departure time choice is an interesting problem to consider endogenous congestion effects.

Even though our results depend on the parametric assumtions adopted, it is possible to extend some of them to more general settings. In particular, we have preliminary the results concerning to equlibrium existence with more general utility functions. However, model identification is strongly dependent on the parameterization and it requires specific analysis.

To our knowledge, no study has used a similar approach to compute optimal nonlinear prices for bus services. There exist two main approaches to compute optimal prices in transportation services: the demand-based and the network equilibrium-based approach. Our methodology belongs to the first group. However, it has the advantage that it takes into account congestion effect, which makes it more general (as in the second group). Although, such network models can take into account the congestion effects, the computation of optimal prices requires solving a mathematical program with equilibrium constraints. Thus, such an approach seems to be unfeasible in a city as big as Santiago.

Our model makes some simplifying assumptions to compute optimal prices. One of them is utility separability, which implies a negligible substitutability between modes. Nevertheless, transportation demand analysis recognizes that different modes are indeed substitutes and that cross price elasticities are significant. If this is the case, the demand for public transportation depends on the demand for trips by car. Then, optimal bus fares have an indirect effect through the congestion level. This is an issue for research in a next stage.

Linearity in the utility of the numeraire good implies a nil income effect. Such an effect may be relevant for individuals with low income, for whom transportation expenditure represents a high proportion of their budget. However, this assumption simplifies the computation of nonlinear prices.

Finally, frurther analysis has to be done on the estimation of marginal costs. Here, we use an existing estimation of the average cost. Indeed, a deeper analysis of such a variable exceeds the scope of this work.

References

Ahn, K. (2009) Road Pricing and Bus Service Policies. *Journal of Transport Economics and Policy*, Vol. 43, 25–53.

Armstrong, M. (1996) Multiproduct Nonlinear Pricing. *Econometrica*, Vol. 64, No. 1, 51–75.

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

Ben-akiva, M. and J.L. Bowman (1998) Activity based travel demand model systems. In P. Marcotte and S. Nguyen, eds. *Equilibrium and Advanced Transportation Modeling.* Kluwer Academic Publishers, Dordrecht, 1998.

Bousquet, A. and M. Ivaldi (2001) An Econometric Measure of Externalities Individual Valuation in Urban Transport. *Working Paper IDEI*, Toulouse.

Carlin, B. and T. Louis (1996) *Bayesian Methods for Data Analysis.* Chapman & Hall/CRC.

De Borger, B. (2001) Discrete choice models and optimal two-part tariffs in the presence of externalities: optimal taxation of cars. *Regional Science and Urban Economics*, Vol. 31, 471–504.

De Cea, J. and E. Fernández (1993) Transit Assignment for Congested Public Transport Systems: An Equilibrium Model. *Transportation Science*, Vol. 27, No. 2, pp. 133–147.

De Cea, J. and E. Fernández (2001) ESTRAUS: a simultaneous equilibrium model to analyze and evaluate multimodal urban transportation systems with multiple user classes. *Proceedings of the Ninth World Conference on Transport Research.*

Dubin, J.A., D.L. McFadden (1984) An Econometric Analysis of Residential Electric Appliance Holdings and Consumption. *Econometrica*, Vol. 52, No. 2 345–362.

Dupuit, J. (1844) De la mesure de l'utilité des travaux publics. *Annales des ponts et chaussées*, Second series, 8.

Geweke (2005) *Contemporary Bayesian Econometrics and Statistics.* Wiley-Interscience.

Gill (2002) *Bayesian Methods: A Social and Behavioral Sciences Approach.* CRC Press.

Gourieroux and Monfort (2003) *Simulation-Based Econometric Methods.* Oxford University Press, USA

Guesnerie, R. and J.J. Laffont (1984) A Complete Solution to a Class of Principal-Agent Problems with an Application to the Control of a Self-Managed Firm. *Journal of Public Economics*, Vol. 25, 329–369.

Hahn, J.H. (2003) Nonlinear pricing of telecommunications with call and network externalities. *International Journal of Industrial Organization*, Vol. 21, 949–967

Hanemann, M. (1984) Discrete/Continuous Models of Consumer Demand. *Econometrica*, Vol. 52, No. 3, 541–562.

Kamien, M. and N. Schwartz, (1991) *Dynamic Optimization.* Elsevier Science.

Kass, R., B. Carlin, A. Gelman and R. Neal (1998) Markov chain Monte Carlo in practice: A roundtable discussion. *The American Statistician*, Vol. 52, 93-100.

Kim, J., G.M. Allenby and P.E. Rossi (2002) Modeling Consumer Demand for Variety. *Marketing Science*, Vol. 21, No. 3, 229–250.

Kim, J., G.M. Allenby and P.E. Rossi (2007) Product attributes and models of multiple discreteness. *Journal of Econometrics*, Vol. 138, 208–230.

Laffont, J.J. and D. Martimort (2002) *The Theory of Incentives: The Principal-Agent Model.* Princeton University Press, Princeton, New Jersey.

Lindsey, C.R. and E.T. Verhoef (2001) Traffic congestion and congestion pricing. In D.A. Hensher and K.J. Button, eds. *Handbook of Transport Systems and Traffic Control, Handbooks in Transport 3.* Elsevier, Amsterdam, 77–105.

Ortúzar, J.D. and L. Willumsen (2002) *Modelling Transport*, Wiley and Son, Chechister.

Oum, T.H. and W.G. Waters II (2000) Transport Demand Elasticities. In D.A. Hensher and K.J. Button, eds. *Handbook of Transport Modelling* . Oxford: Elsevier Science Ltd.

Pigou, A.C. (1920) *Wealth and Welfare.* MacMillan, London.

Salanie, B. (1997) *The Economics of Contracts: A Primer.* MIT Press, Cambridge, MA.

San Martín, E. and J. González (2010) Bayesian identifiability: Contributions to an inconclusive debate. *Chilean Journal of Statistics*, Vol. 1, No. 2, 69-91.

SECTRA (2002) *Análisis Modernización del Transporte Público, VI Etapa.* Ministerio de Planificación, Gobierno de Chile. Santiago, Chile.

Spiess, H. (1984) Contribution 'a la théorie et aux outils de planification des réseaux de transport urbains. *Publication 382, Centre de Recherche sur les Transports*, Univ. de Montréal.

Spiess, H., and M. Florian (1989) Optimal strategies: A new assignment model for transit networks. *Transportation Research*, Vol. 23B, 83–102.

Spulber, D.F. (1989) *Regulation and Markets.* MIT Press, Cambridge, Massachusetts.

Sundararajan, A. (2004) Nonlinear pricing and type-dependent network effects. *Economics Letters,* Vol. 83, 107–113.

Timilsina, G.R. and H.B. Dulal (2008) Fiscal Policy Instruments for Reducing Congestion and Atmospheric Emissions in the Transport Sector: A Review. *Policy Research Working Paper* 4652, World Bank.

Train, K. and G. Sonnier (2003) Mixed Logit with Bounded Distributions of Partworths, *Working Paper*, University of California, Berkeley.

Train, K. (2003) *Discrete Choice Methods with Simulation*, Cambridge University Press, New York.

Van Dender, K. and S. Proost (2003) Marginal Social Cost Pricing for all Transport modes and the effects of modal budget constraints. *Working Paper*, Center for Economic Studies, Katholieke Universiteit Leuven.

Walker, J.L., M. Ben-akiva, and d. Bolduc (2007) Identification of Parameters in Normal Error Component Logit-Mixture (NECLM) Models. *Journal of Applied Econometrics*, Vol. 22, 1095–1125.

Walters, A.A. (1961) The theory and measurement of private and social cost of highway congestion. *Econometrica*, Vol. 29, No. 4, 676–697.

Wang, J., R. Lindsey and H. Yang (2008) Nonlinear Pricing on Private Roads with Congestion and Toll Collection Costs. *Working Paper*, Department of Economics, University of Alberta.

Wilson, R.A. (1993) *Nonlinear Pricing*, New York, Oxford University Press.

Appendix A

Proof of Proposition 1.

Consider the function $g(z)$, $z \in \mathbb{R}_+$, defined by

$$g(z) = e^{a+\gamma z} \int_{-(a+\gamma z)}^{\overline{\theta}} e^{\theta} f(\theta) d\theta - \int_{-(a+\gamma z)}^{\overline{\theta}} f(\theta) d\theta$$

where f is a density function with support $[\underline{\theta}, \overline{\theta}]$, $\underline{\theta} < -a < \overline{\theta}$ and $\gamma < 0$.

From the definition follows that g is a continuous function and is also monotonically decreasing. Indeed

$$g'(z) = \gamma e^{a+\gamma z} \int_{-(a+\gamma z)}^{\overline{\theta}} e^{\theta} f(\theta) d\theta < 0 \; \forall z$$

To prove the existence of a solution of the fixed point equation given by (12) it suffice show that $g(0) > 0$ and there exists $\overline{z} > 0$ such that $g(\overline{z}) = 0$. Indeed,

$$\begin{aligned} g(0) &= e^{a} \int_{-a}^{\overline{\theta}} e^{\theta} f(\theta) d\theta - \int_{-a}^{\overline{\theta}} f(\theta) d\theta \\ &= \int_{-a}^{\overline{\theta}} \left[e^{a+\theta} - 1 \right] f(\theta) d\theta \\ &> 0 \end{aligned}$$

since $e^{a+\theta} - 1$ is positive for all $\theta > -a$.

Now, consider $\overline{z} = -(\overline{\theta} + a)/\gamma > 0$ then

$$\begin{aligned} g(\overline{z}) &= e^{a+\gamma \overline{z}} \int_{-(a+\gamma \overline{z})}^{\overline{\theta}} e^{\theta} f(\theta) d\theta - \int_{-(a+\gamma \overline{z})}^{\overline{\theta}} f(\theta) d\theta \\ &= e^{-\overline{\theta}} \int_{\overline{\theta}}^{\overline{\theta}} e^{\theta} f(\theta) d\theta - \int_{\overline{\theta}}^{\overline{\theta}} f(\theta) d\theta \\ &= 0 \end{aligned}$$

The uniqueness of the solution results from monotonicity of g.

Appendix B

Derivatives of congestion with respect to the parameters of the model

Consider the function $z : \mathbb{R}^d \times \mathbb{R}_+ \to \mathbb{R}_+$, such that $(x, \sigma) \to z(x)$, defined implicitly by the equation

$$z = e^{g(x,z)} \int_{-g(x,z)}^{\overline{\theta}} e^{\theta} f(\theta|\sigma) d\theta - \int_{-g(x,z)}^{\overline{\theta}} f(\theta|\sigma) d\theta \tag{A.1}$$

where f is a density function conditional in σ with support $[\underline{\theta}, \overline{\theta}]$ and g is a continuous real-valued function defined over $X \times \mathbb{R}_+$, with $X \subset \mathbb{R}^d$

Consider that the function g is such that eq. (A.1) there exist a solution for all $x \subset X$. We can apply implicit derivation and the Leibniz's rule for differentiation under the integral sign. Then, the derivative of z with respecto to x_i is given by

$$\frac{\partial z}{\partial x_i}(x) = \frac{\partial g}{\partial x_i}(x,z) F(g(x,z),\sigma) \left[1 - \frac{\partial g}{\partial z}(x,z) F(g(x,z),\sigma)\right]^{-1} \tag{A.2}$$

where

$$F(y,\sigma) = e^{y} \int_{-y}^{\overline{\theta}} e^{\theta} f(\theta|\sigma) d\theta$$

Consider $x = (\alpha_1, \beta_1, \gamma, p_1)$, $z = \mathcal{X}_1$ and $g(x,z) = \alpha_1 + \beta_1 \mathcal{X}_1 - \gamma p_1$, with $\beta < 0$. Then, applying eq. (A.2), it follows

$$\begin{aligned}
\frac{\partial \mathcal{X}_1}{\partial \alpha_1} &= F(\alpha_1 + \beta_1 \mathcal{X}_1 - \gamma p_1, \sigma) \left[1 - \beta F(\alpha_1 + \beta_1 \mathcal{X}_1 - \gamma p_1, \sigma)\right]^{-1} > 0 \\
\frac{\partial \mathcal{X}_1}{\partial \beta_1} &= \mathcal{X}_1 F(\alpha_1 + \beta_1 \mathcal{X}_1 - \gamma p_1, \sigma) \left[1 - \beta F(\alpha_1 + \beta_1 \mathcal{X}_1 - \gamma p_1, \sigma)\right]^{-1} > 0
\end{aligned}$$

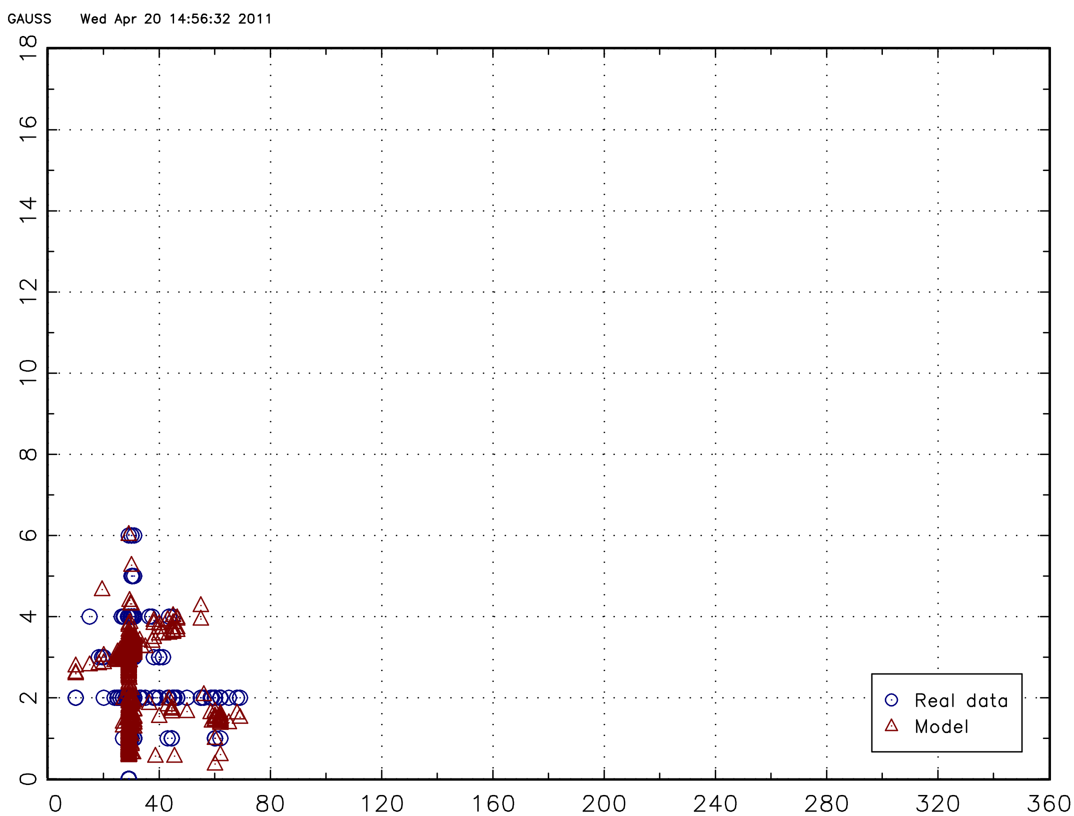

Figure 1: Comparison of real and modeled trips by car

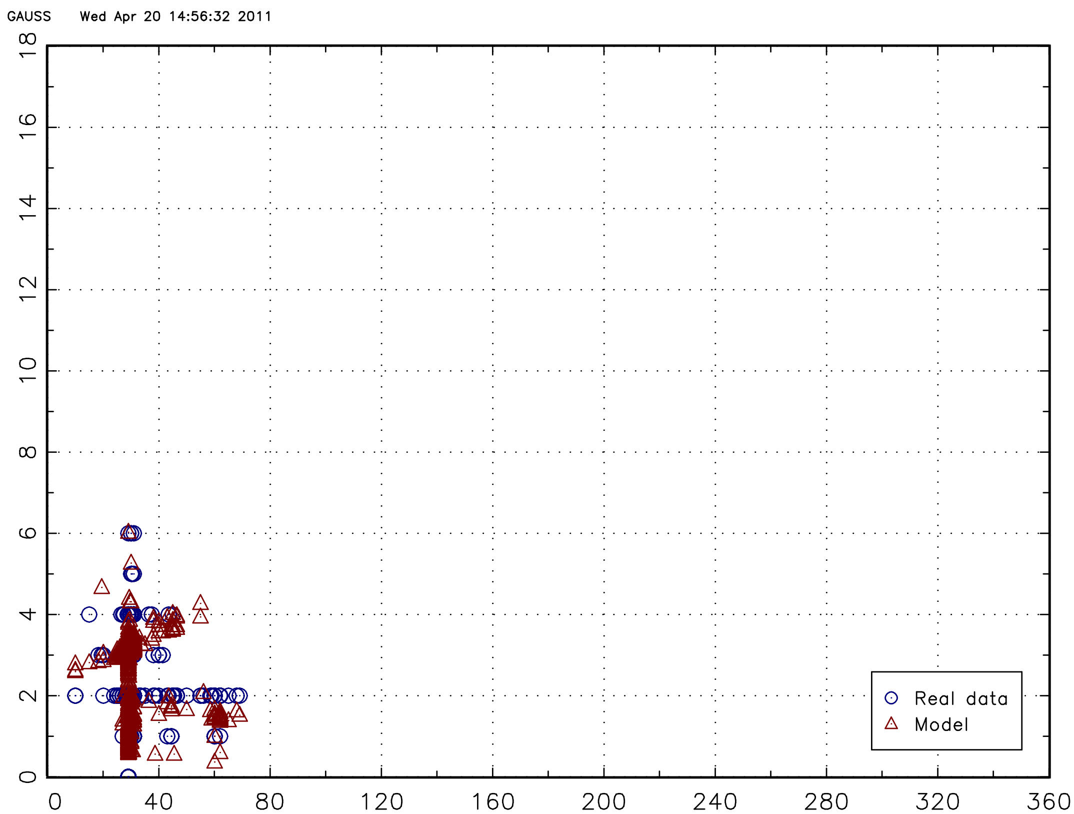

Figure 2: Comparison of real and modeled trips by bus

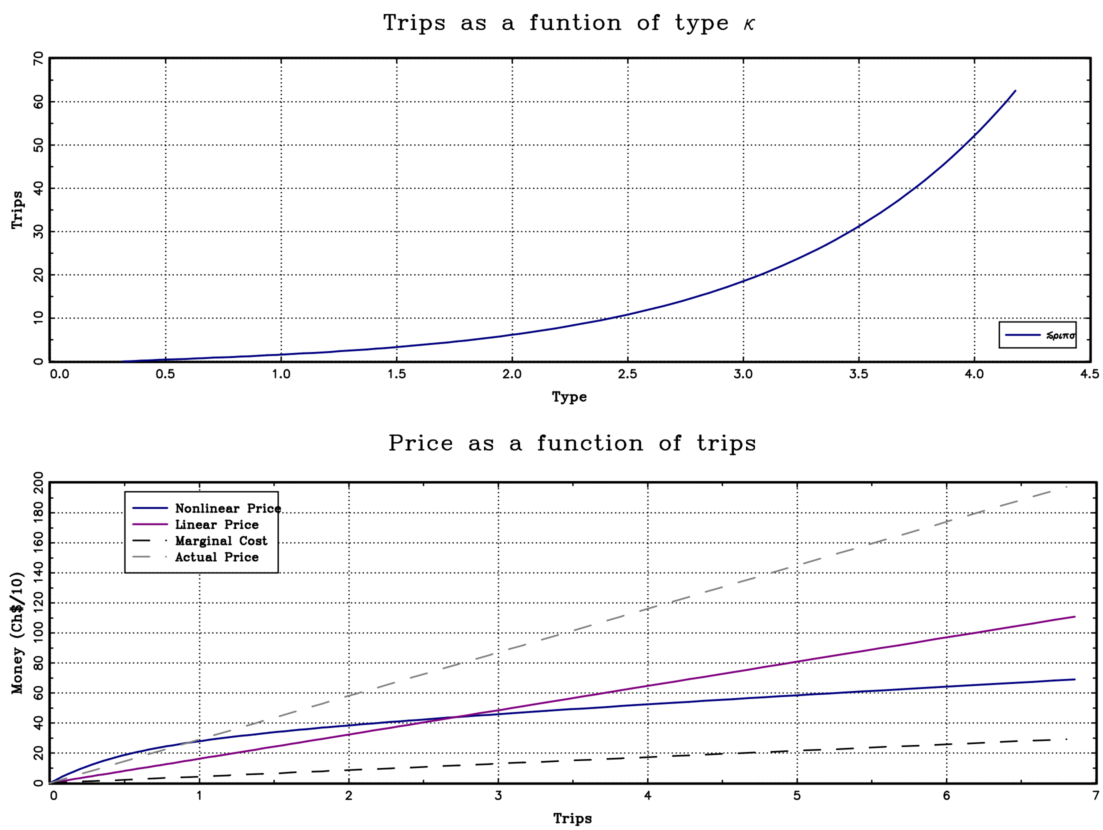

Figure 3: Trips as a function of individual's type (upper panel) and pricing schedules (lower panel)