Nonlinear Pricing in Transportation:
An application to transit system of Santiago de Chile

Marco Batarce*
Toulouse School of Economics

Marc Ivaldi
Toulouse School of Economics

May 2010

Abstract

This paper computes optimal nonlinear prices for public transportation in Santiago of Chile. 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 users’ utility maximization problem to derive the likelihood function and apply Bayesian methods for inference. Using data from Santiago of 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. As result, the nonlinear pricing schedule produces total benefits slightly greater than the linear pricing; however the former requires significantly lower subsidy.

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

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 firm operations and therefore a nonlinear tariff minimizes allocative distortions caused by setting prices equal to marginal cost when the firm is a monopoly.

However, the economic literature has focused on how to price roads with marginal costs and to internalize the congestion since many years (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, 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 objective of this paper is to develop a methodology to compute optimal nonlinear prices for public transportation and to apply it to Santiago de Chile. 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 that 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 demands depend on the price faced by each user.

Our methodology takes into consideration two problems. First, the asymmetry of information 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, modeling methods use demand segmentation according to departure time, origin and destination, income and car ownership. This allows 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

---

1In 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)

2For instance, Ortúzar and Willumsen (2002) do not mention problems of asymmetric information in transportation planning.
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. Then, they play a game of incomplete information. Congestion effects are considered explicitly in transportation analysis by means of network equilibrium models (De Cea and Fernández, 2001), but only through increases in travel time and waiting time in the case of bus network. The equilibrium approach only considers time effects of congestion on the users’ utility and neglects, for instance, innate aversion to congestion or aversion to pollution produced by congestion.

In principle, utility functions should satisfy: i) implementation conditions for nonlinear tariffs, ii) existence of equilibrium in the individual utility maximization problem. The parameterization of the utility function used here satisfies both criteria. In addition, the demand functions derived from this parameterization are consistent with observed behavior, such as zero demand for a number of individuals in one or more alternatives of transportation.

For the estimation, we adapt the methodology developed by Kim et al. (2007). They derive the likelihood function of the data from the first order conditions of the individual’s utility maximization problem. They assume a random component in the marginal utility. Bayesian methods are used for statistical inference. In particular, a Monte Carlo Markov Chain is implemented to simulate the posterior distribution of the parameters. We use data from Santiago de Chile to estimate the utility function.

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. However, the subsidies are lower in the nonlinear case. They are Ch$10 and Ch$40 per trip for the nonlinear schedule and the linear one, respectively. The total demand for trips is the same under both pricing schemes. Nevertheless, in the nonlinear case there is a lower proportion of user traveling by bus. This implies an efficiency-exclusion trade-off in the pricing policy.

The paper is organized as follows. Section 2 presents the model and the implementation conditions according to Guesnerie and Laffont (1984). We derive optimal prices for the general case where user’s utility depends on the expected number of the trips on the network. Section 3 presents the parameterization. It is consistent with the data and satisfies conditions of section 2. Section 4 describes the estimation procedure based on Bayesian inference and Monte Carlo Markov Chain methods. The application to Santiago de Chile is reported in section 5. Finally, possible extensions and limitations are presented in section 6.
2 Optimal prices with congestion

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, $\theta$, 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 $\theta$.

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, $\theta$, with the exception of the distribution. The individual’s problem is

$$\max_x u(x, X, P(x); \theta)$$

s.t. $X = \int_\theta P(x, X, P(\cdot))f(\mu)d\mu$

The solution is a demand function, $x(\theta, X, P(\cdot))$, which depends on $\theta$, $X$, 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 $\theta$, since $(x, t)$ is constructed using the demand schedule.

A schedule $(x, t)$ for which the user reveals her type parameter $\theta$ 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 $\theta$ 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 $C^2$-class;

(CS) *constant sign of marginal rate of substitution*: the sign of the vector $\frac{\partial^2 U}{\partial x \partial t}(x, t, \theta)$ 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 $\left\| \frac{\partial U}{\partial x} \right\| \leq K |t|$, uniformly in $x, \theta$.

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

$$ U(x, t, X, \theta) = u(x, X, \theta) - t $$

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

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

$$ \frac{dx}{d\theta}(\theta) \geq 0 \quad (IC2) $$

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 need to collect $\$(1 + \lambda)$ in taxes to pay $\$1$ to the firm.
The aggregated consumer’s surplus is

\[ CS = \int_\theta S(\theta) f(\theta) d\theta = \int_\theta [u(x(\theta), X, \theta) - t(\theta)] f(\theta) d\theta \]

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

\[ PS = \int_\theta [t(\theta) - cx(\theta)] f(\theta) d\theta - C_0 \]

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

\[
\max_{(x,t)} \{ CS + (1 + \lambda)PS \} \\
\text{s.t. } \frac{\partial u}{\partial x}(x, X, \theta) \frac{dx}{d\theta}(\theta) = \frac{dt}{d\theta}(\theta) \]
\[ \frac{dx}{d\theta}(\theta) \geq 0 \]
\[ x(\theta) \geq 0 \]

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

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

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

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

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

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

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

\[
\frac{\partial v}{\partial x}(x, \theta) + w(X) = \frac{dP}{dx}(X) \equiv \pi(x) 
\]
From (6), there is a direct relationship between $\theta$ and $x$. Let $\theta(z)$ represent $\min\{\theta : x(\theta) = z\}$ (Spulber, 1989). The optimal marginal prices can be expressed as

$$\pi(x) = c + \frac{\lambda}{1 + \lambda} \frac{\partial^2 u}{\partial x \partial \theta}(x, \theta(x)) H(\theta(x)) + \xi(\theta) \frac{d\omega}{dX}(\xi(\theta))$$

(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 consumer’s informational rent. It is due to the incentive compatibility requirement. Since $H(\theta)$ is decreasing, a higher $\theta$ implies a lower the rent extracted by the planner. The intuition behind such reduction is that it is more socially efficient offer a lower marginal price to individuals who obtain a greater benefit for travel (recall we assume $U$ increasing in $\theta$). 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, optimal pricing schedule internalizes congestion effects.

It is worth three remarks. 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 total number of users in the network. Second, 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 contrasts 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 consumer’s utility function and private information, together with implementability of a nonlinear pricing schedule.

Regarding the first one, corner solutions should be admissible. Indeed, some individuals in the sample choose more than one mode along a 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

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

where

\[
\psi_i = \alpha_i + \beta_i E(x_k) + \varepsilon_i
\]

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 \( i \). \( \theta \) represents consumers’ private information and it has support \([\underline{\theta}, \overline{\theta}]\). \( \gamma \) measures the weight of the consumption of a numeraire good \( z \) and equals the marginal utility of income.

The quality index \( \psi_i \) depends on observable and unobservable components. In (10), \( \alpha_i \) represents the observable one. Typically, \( \alpha_i \) captures attributes such as comfort, safety or reliability. The unobservable part is represented by \( \varepsilon_i \) and can be interpreted as taste variation across individuals or as subjective perception of quality. The standard assumptions are that individuals observe \( \varepsilon_i \), 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 \( \beta_i \) measures the impact of such effect in the utility of the alternative \( i \).

\( \theta \) 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 \( \theta \) is greater than \( \gamma p_i - \psi_i \). The last term can be interpreted as a generalized cost for making a trip. Therefore, \( \theta \) is a reservation utility for carrying out the activity that motivates the trip.
Given the interpretation of $\theta$, 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 along one day, there is not enough variability to estimate a demand system for trips by mode and purpose, and obtain an accurate estimation of $\theta$ for each purpose.

The demand functions are as follows

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

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 $\theta$ and $\varepsilon_i$. In particular, we adopt normal distributions for $\theta$ and $\varepsilon_i$, for all $i$, therefore the distribution of $\eta$ 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) $\theta$ is distributed in $[\underline{\theta}, \overline{\theta}]$ with a density function $f$ that is common knowledge; and

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

Then, if $\eta_k = \theta + \varepsilon_k$ and $f_{\eta_k}$ 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

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

where

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

In addition, the solution always exists and is unique.

**Proof.** See Appendix A. \(\blacksquare\)
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 \( \alpha_j \) between individuals with available car \((\alpha_j^c)\) and those without car \((\alpha_j^{nc})\). The probability of being of type \( \kappa \), using the notation introduced in Proposition 1, is as follows

\[
\Pr(\text{being type } \kappa) = \Pr(\eta_j = \kappa - \alpha_j^c) \cdot \Pr(\text{available car}) + \\
\Pr(\eta_j = \kappa - \alpha_j^{nc}) \cdot \Pr(\text{no available car})
\]

\[
= f_{\eta_j}(\kappa - \alpha_j^c) \cdot q + f_{\eta_j}(\kappa - \alpha_j^{nc}) \cdot (1 - q)
\]

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 of section 5, \( \theta \) and \( \varepsilon_j \) 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 two alternatives: Mode 1 is car and mode 2 is public transport (bus and subway). In the first stage, equation (6) is solved on \( x_2(\theta) \). 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

\[
x_2(\theta) = \exp \left\{ \theta + \alpha_2 + \beta_2 \xi_1(\bar{\theta}) - \gamma c_2 + \frac{1 - \lambda}{\lambda} H(\theta) \right\} - 1
\]

where \( \xi_1(\bar{\theta}) \) is given by the solution of equation (12).

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

\[
\pi_2(x_2(\theta)) = c_2 - \frac{1 - \lambda}{\lambda} \frac{H(\theta)}{\gamma}
\]

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

\( x_2(\theta) \) 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 Monte Carlo Markov Chain (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

\[
\frac{\partial U}{\partial x_i}(x) = \psi_i + \theta - \ln(x_i + 1) = \lambda p_i \text{ if } x_i > 0
\]

\[
\frac{\partial U}{\partial x_i}(x) = \psi_i + \theta - \ln(x_i + 1) < \lambda p_i \text{ if } x_i = 0
\]

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

\[
V_i \equiv \alpha_i + \beta_i E(x_k) + \theta - \ln(x_i + 1) - \gamma p_i = -\varepsilon_i \text{ if } x_i > 0
\]

\[
V_i \equiv \alpha_i + \beta_i E(x_k) + \theta - \ln(x_i + 1) - \gamma p_i < -\varepsilon_i \text{ if } x_i = 0
\]

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, \( \varepsilon \). We use equations (19) and (20) to obtain the distribution of \( x \) by applying the change-of-variable theorem.

If \( \varepsilon \) follows a multivariate normal distribution with mean zero and variance \( \Omega \), 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

\[
l(x) = \Pr(x_i > 0, x_j = 0; i = 1, \ldots, m, j = m + 1, \ldots, n)
\]

\[
= -V_n \ldots -V_{m+1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi(-V_1, \ldots, -V_m, \varepsilon_{m+1}, \ldots, \varepsilon_n | 0, \Omega) |J| d\varepsilon_{m+1}, \ldots, d\varepsilon_n
\]

where \( \phi \) 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 \( i, j = 1, \ldots, m \), 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 $\varepsilon_a = (\varepsilon_1, \ldots, \varepsilon_m)'$ and $\varepsilon_b = (\varepsilon_{m+1}, \ldots, \varepsilon_n)'$, such that

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

with $\varepsilon_a$ and $\varepsilon_b|\varepsilon_a$ are 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 $V_a = (V_1, \ldots, V_m)'$. Therefore, the likelihood function (21) can be rewritten as

$$
l(x) = \Pr(x_i > 0, x_j = 0; i = 1, \ldots, m, j = m + 1, \ldots, n)
$$

$$
= \phi_{\varepsilon_a}(-V_1, \ldots, -V_m|0, \Omega_{aa}) |J|
\cdot \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \phi_{\varepsilon_b|\varepsilon_a}(v_{m+1}, \ldots, v_n|\mu, \Sigma) |J| dv_{m+1}, \ldots, dv_n
$$

4.2 Inference

Our interest is on the point estimation, together with the distribution of $\theta$. Remember that $\theta$ and $\varepsilon$ captures individuals’ heterogeneity regarding preferences for transportation, equation (9). However, due to identification restrictions we are only able to estimate the distribution of $\theta$.

We estimate a random parameter model in which $\theta$ 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) as follows

$$
x \sim \text{according to the likelihood func. } l(x|\zeta, \theta, \tilde{p}), \text{ eq. (22)}
$$

$$
\zeta \sim N(\zeta_0, \Sigma_{\zeta_0})
$$

$$
\theta \sim N(\mu_\theta, \sigma^2_\theta)
$$

$$
\mu_\theta \sim N(m, s^2)
$$

$$
\sigma^2_\theta \sim IG(a, b)
$$
where $\zeta = (\bar{\pi}_1, \bar{\beta}_1, \ldots, \bar{\pi}_N, \bar{\beta}_N, \bar{\gamma})$ is a vector of transformed parameters of the structural model in equation (9). 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

$$\beta_i < 0 \Rightarrow \beta_i = -\exp(\bar{\beta}_i) \quad (23)$$

$$\gamma > 0 \Rightarrow \gamma = \exp(\bar{\gamma}) \quad (24)$$

There are no requirements on $\alpha_i$, therefore $\bar{\pi}_i = \alpha_i$. Condition (23) implies that the congestion has a negative effect in the utility. Equation (24) comes from the fact that $\gamma$ represents the marginal utility of income, which is positive. Then, the vector $\zeta$ distributes according to a normal with mean $\zeta_0$ and variance matrix $\Sigma_{\zeta_0}$.

The random parameter $\theta$ is normally distributed with parameters $\mu_\theta$ and $\sigma_\theta$, which are the parameters of interest. Likewise, $\mu_\theta$ distributes normally with mean $m$ and variance $s^2$, and $\sigma_\theta$ distributes according to an inverted gamma with parameters $a$ and $b$. These priors for $\mu_\theta$ and $\sigma_\theta$ allows to obtain closed expressions for the posteriors, given a sample of the individual parameter, $\theta$ (Train, 2003).

The posterior distribution of the parameters is obtained by Monte Carlo Markov Chain method. An advantage of the MCMC method is that it allows us to obtain individual-level estimates for $\theta$. In particular, we use a Gibbs sampling method with a Metropolis step to infer the posterior distribution of the model’s parameters (Geweke, 2005).

### 4.3 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 done. 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 be not totally, theoretically consistent with the underlying utility function. Oppositely, the approach adopted in this paper satisfies theoretical consistency.

5 Application to Santiago

This section presents the estimation results using data from Santiago of Chile and the computation of nonlinear prices. We assume that there are three available transportation modes: Car, public transportation (bus and subway), and non motorized modes (walking and bicycling). 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 corresponds to surveys conducted during the normal season and 3,191 to 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.

To estimate, we consider only information from individuals making at least one trip (regardless of the mode), except students. In addition, we exclude all information coming from combining modes. This results in a sample of 25,642 individuals and 86,773 trips, from which 16,773 are made by car, 26,292 by bus, 2,970 by subway, and 37,688 by non-motorized modes (walking and/or bicycle). The filtered sample represents 96.5% of the total trips within the sample of non-students.

The price of a car trip is calculated as a function of travel distance, gasoline consumption per kilometer 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 imputed³.

Table 1

Characteristics of households and persons in the sample

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Households</td>
<td>14,152</td>
</tr>
<tr>
<td>Mean household size</td>
<td>4.0 pers/hh</td>
</tr>
<tr>
<td>Mean household income</td>
<td>515.9 Ch$ *10³</td>
</tr>
<tr>
<td>Mean income per capita</td>
<td>128.9 Ch$ *10³</td>
</tr>
<tr>
<td>Households with car</td>
<td>33 %</td>
</tr>
<tr>
<td>Mean cars per household</td>
<td>0.4 car/hh</td>
</tr>
<tr>
<td>Persons</td>
<td>43,692</td>
</tr>
<tr>
<td>Workers</td>
<td>44.9 %</td>
</tr>
<tr>
<td>Students</td>
<td>29.6 %</td>
</tr>
<tr>
<td>With driving license</td>
<td>26.0 %</td>
</tr>
</tbody>
</table>

Table 2

Characteristics of the trips in the sample

<table>
<thead>
<tr>
<th>Trips</th>
<th>Trips/person</th>
<th>Mean travel time (min.)</th>
<th>Mean distance (km.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All modes</td>
<td>147,872</td>
<td>3.39</td>
<td>26.5</td>
</tr>
<tr>
<td>Car</td>
<td>20,345</td>
<td>3.67</td>
<td>27.3</td>
</tr>
<tr>
<td>Bus</td>
<td>40,244</td>
<td>2.12</td>
<td>45.7</td>
</tr>
<tr>
<td>Subway</td>
<td>4,805</td>
<td>1.75</td>
<td>43.1</td>
</tr>
<tr>
<td>Walking</td>
<td>56,573</td>
<td>2.94</td>
<td>11.1</td>
</tr>
<tr>
<td>Bicycle</td>
<td>3,343</td>
<td>2.44</td>
<td>18.7</td>
</tr>
</tbody>
</table>

5.2 Parameter estimation

In addition to the heterogeneity represented by $\theta$ and $\varepsilon$, 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

³This price is flat (it does not vary with distance nor with time schedules) and only unitary tickets are charged.
that the social planner (who use 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 only its distribution is estimated.

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

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

Model 1 : $\psi_i = \alpha_i + \varepsilon_i$ \hspace{1cm} (25)

Model 2 : $\psi_i = \delta_{\text{car}}(\alpha^c_i + \beta^c_i E(x_k)) + (1 - \delta_{\text{car}})(\alpha^{nc}_i + \beta^{nc}_i E(x_k)) + \varepsilon_i$ \hspace{1cm} (26)

Model 3 : $\psi_i = \alpha_i + \beta_i E(x_k) + \varepsilon_i$ \hspace{1cm} (27)

Thus, the model is estimated with different quality indexes for public transportation and non-motorized modes 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 variable is given by the fraction of individuals with car in the population, the same fraction in the sample is a consistent estimator.

The results for the three 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 and fourth columns, Model 2 considers the discrete heterogeneity and congestion effects. The last two columns present the results for the model without discrete heterogeneity, Model 3. Table 3 also includes the likelihood ratio tests under the null hypotheses: (H$_{10}$) no congestion effects (Model 1 vs. Model 2), and (H$_{20}$) no discrete heterogeneity (Model 3 vs. Model 2).

Regarding the final value of the likelihood function, Model 2 is better than the other two models. The likelihood ratio tests reject both null hypotheses with high confidence level. Thus, model 2 is statistically better than the other models.

The standard errors of the parameters $\beta$ 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 $H_{10}$.

Model 2 has a marginal utility of income (parameter $\gamma$) consistent with the literature. Indeed, according to this specification, the demand price elasticities (in absolute value) are 0.66 and 0.52 for car and public transportation, respectively. They are calculated using $\gamma$ 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 elasticities demand for trips by automobile in the range of (in absolute value) 0.00-0.52 and for trips by transit system in the range of 0.01-0.96. Oppositely, when considering models 1 and 3 the values of the price elasticity turn to be inconsistent with those in Oum and Waters (2000). This constitutes another reason in favor of model 2.

The estimation results are in line with what is expected ex ante. For example, the values of the parameters \( \alpha \), which represent the perceived quality of each alternative, imply that car is the best quality mode and non-motorized is the worst mode. This feature is present in the three models. Regarding the value of \( \beta \) in Model 2, the congestion matters more for individuals without available car.

### Table 3

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th></th>
<th>Model 2</th>
<th></th>
<th>Model 3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu )</td>
<td>-0.3384</td>
<td>0.0046</td>
<td>0.2059</td>
<td>0.0076</td>
<td>0.2984</td>
<td>0.0051</td>
</tr>
<tr>
<td>( \sigma )</td>
<td>0.0061</td>
<td>0.0004</td>
<td>0.0079</td>
<td>0.0007</td>
<td>0.0068</td>
<td>0.0006</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>0.0018</td>
<td>0.0001</td>
<td>0.0179</td>
<td>0.0001</td>
<td>0.0002</td>
<td>0.0001</td>
</tr>
<tr>
<td>( \alpha_1 )</td>
<td>1.3946</td>
<td>0.0063</td>
<td>1.4928</td>
<td>0.0079</td>
<td>0.7106</td>
<td>0.0051</td>
</tr>
<tr>
<td>( \alpha_2 )</td>
<td>0</td>
<td>—</td>
<td>0</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>( \alpha_2^{car} )</td>
<td>0.8752</td>
<td>0.0042</td>
<td>2.4488</td>
<td>0.0080</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>( \alpha_2^{nocar} )</td>
<td>-0.0769</td>
<td>0.0065</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>( \alpha_3 )</td>
<td>-0.3439</td>
<td>0.0052</td>
<td>-0.8900</td>
<td>0.0056</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>( \alpha_3^{car} )</td>
<td>0.7624</td>
<td>0.0014</td>
<td>0.2373</td>
<td>0.0056</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>( \alpha_3^{nocar} )</td>
<td>-1.6104</td>
<td>0.0053</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>-0.1345</td>
<td>0.0021</td>
<td>-0.0064</td>
<td>0.0001</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>-0.0061</td>
<td>0.0001</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>( \beta_2^{car} )</td>
<td>-0.1079</td>
<td>0.0003</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>( \beta_2^{nocar} )</td>
<td>-1.6104</td>
<td>0.0053</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Sim. log-lik.      -96687 | -93468 | -99934

LRT(\( H_0^{1} \)) (Model 1/Model 2) 6437
LRT(\( H_0^{2} \)) (Model 3/Model 2) 12931

Heterogeneity is statistically significant (see the value of standard error of \( \sigma_\theta \)). However, the value of the coefficient estimated is very small. This may be due to the fact that we have fixed the variance of \( \varepsilon \) equal to one. The consequence of such a parameterization is that we impose a total variance equal
to \((1 + \sigma_\theta^2)^4\). The estimated \(\sigma_\theta\) 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 car trips the correlation is 0.64. For bus trips the correlation is 0.43.

### 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, \(\kappa\), in the population is given by the mixture distribution in equation (14).

Since the parameters \(\theta\) and \(\varepsilon_2\) are assumed normally distributed with support on \( \mathbb{R} \), \(\kappa\) representing private information is also distributed on \( \mathbb{R} \). The numerical solution of equations (15) and (16) are badly behaved in the tail of the distribution, which results in the need to define a closed interval for the support of \(\kappa\). Based on some robustness tests, we define the support as \([\mu_\theta - 3.5(\sigma_\theta + \varepsilon), \mu_\theta + 3.5(\sigma_\theta + \varepsilon)]\). This interval leaves out only 0.26% of individuals. Thus, we neglect the impact of changing the support of \(\kappa\).

The planner’s problem requires knowledge of 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 (Sectra, 2002). If we consider that capacity of the system 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 a 20% of the average cost.

We also need to assume a cost of public funds, \(\lambda\), 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 \(\kappa^*\) such that \(x(\kappa^*) = 0\). We define the solution \(\hat{x}(\kappa) = \max(0, x(\kappa^*))\). In this case \(\kappa^* = 0.82\). Figure 1 in the upper panel shows the optimal number of trips as a function of individual’s type, together with optimal prices as a function of number of trips (lower panel).

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 4 presents the final optimal menu. As reference point, the price ticket in the bus system of Santiago is approximately 290 (Ch$).

We compare total benefits of both optimal nonlinear and linear pricing schedule. Table 5 shows the mean of expected benefits. Nonlinear prices are considerably better with respect to linear ones. The benefits are 29% higher in the first case. For the menu of tickets, presented in Table 4, the benefits are 25% greater than those obtained with a linear pricing schedule.

---

4 Recall that the heterogeneity is represented by the parameters \(\theta\) and \(\varepsilon\).

5 In the general case, this requires solving an optimal control problem with bounded control (Kamien and Schwartz, 1991)
Therefore, in terms of benefits, the results reflect that a nonlinear pricing policy is preferable when the provision of public transportation is financed with costly public funds.

Table 4
Optimal menu of bus tickets

<table>
<thead>
<tr>
<th>Ticket</th>
<th>Price (Ch$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 trip</td>
<td>280</td>
</tr>
<tr>
<td>2 trips</td>
<td>500</td>
</tr>
<tr>
<td>3 trips</td>
<td>690</td>
</tr>
<tr>
<td>4 trips</td>
<td>860</td>
</tr>
<tr>
<td>Daily</td>
<td>1030</td>
</tr>
</tbody>
</table>

Table 5
Benefits under different pricing schedules

<table>
<thead>
<tr>
<th>Pricing schedule</th>
<th>Mean benefits (Ch$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marginal-cost price</td>
<td>497</td>
</tr>
<tr>
<td>Implemented linear price</td>
<td>541</td>
</tr>
<tr>
<td>Optimal linear price</td>
<td>576</td>
</tr>
<tr>
<td>Optimal nonlinear prices</td>
<td>590</td>
</tr>
</tbody>
</table>

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 6 shows the demands under linear and nonlinear pricing.

There exists a trade-off between economic efficiency and exclusion in the market. Indeed, the implemented sub-optimal linear price exhibits a higher demand and a lower proportion of excluded users. Despite equal demands for both linear and nonlinear prices, the exclusion is higher in the nonlinear case. This means there are more trips per user under the nonlinear pricing.

Table 6
Demand for public transportation under different pricing schedules

<table>
<thead>
<tr>
<th>Pricing schedule</th>
<th>Mean of trips</th>
<th>Individuals choosing PT (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marginal-cost price</td>
<td>2.51</td>
<td>73.3</td>
</tr>
<tr>
<td>Implemented linear price</td>
<td>1.33</td>
<td>58.9</td>
</tr>
<tr>
<td>Optimal linear price</td>
<td>1.81</td>
<td>65.8</td>
</tr>
<tr>
<td>Optimal nonlinear prices</td>
<td>1.85</td>
<td>56.1</td>
</tr>
</tbody>
</table>
Finally, table 7 presents net revenues. Only the implemented pricing schedule lead to positive net revenues. To eliminate the subsidy, the nonlinear pricing schedule may consider a fixed fee. However, doing so, the number of people choosing public transportation decreases. Thus, there is a trade-off between exclusion and public financing of the service.

<table>
<thead>
<tr>
<th>Pricing schedule</th>
<th>Mean revenues (Ch$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marginal-cost price</td>
<td>-292</td>
</tr>
<tr>
<td>Implemented linear price</td>
<td>49</td>
</tr>
<tr>
<td>Optimal linear price</td>
<td>-41</td>
</tr>
<tr>
<td>Optimal nonlinear prices</td>
<td>-10</td>
</tr>
</tbody>
</table>

6 Final comments

To our knowledge, no study has used a similar approach to compute optimal nonlinear prices for bus services. There exists 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.

Further 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.

Our model takes 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.

Finally, linearity in the utility of the numeraire good implies a nil income effect. Such 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.
References


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


Kim, J., G.M. Allenby and P.E. Rossi (2007) Product attributes and models of multiple discrete-


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


World Bank.

Paper*, University of California, Berkeley.


Appendix A

Proof of Proposition 1.

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

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

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

\[
g'(z) = \gamma e^{\alpha + \gamma z} \int_{-(a+\gamma z)}^{\overline{\theta}} e^{\theta} f(\theta) d\theta < 0 \quad \forall z
\]

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,

\[
g(0) = e^{\alpha} \int_{-\overline{\theta}}^{\overline{\theta}} e^{\theta} f(\theta) d\theta - \int_{-\overline{\theta}}^{\overline{\theta}} f(\theta) d\theta
\]

\[
= \int_{-\overline{\theta}}^{\overline{\theta}} [e^{\alpha + \theta} - 1] f(\theta) d\theta
\]

\[
> 0
\]

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

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

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

\[
= e^{\overline{\theta}} \int_{\overline{\theta}}^{\overline{\theta}} e^{\theta} f(\theta) d\theta - \int_{\overline{\theta}}^{\overline{\theta}} f(\theta) d\theta
\]

\[
= 0
\]

The uniqueness of the solution results from monotonicity of \( g \).
Figure 1: Trips as a function of type $\kappa$ and price as a function of trips