A Logit Model With Endogenous Explanatory Variables and Network Externalities

Louis de Grange · Felipe González · Ignacio Vargas · Rodrigo Troncoso

Published online: 13 January 2015
© Springer Science+Business Media New York 2015

Abstract A novel logit model is presented that explicitly includes endogeneity in explanatory variables whose values depend on individual choice decisions that involve network externalities or social interactions such as those impacting road congestion or public transport comfort and convenience. The proposed specification corrects for this particular type of endogeneity. The model is derived from a linearly constrained maximum entropy optimization problem that incorporates the network externalities or social interactions causing the endogeneity. It is validated through simulations and an application to a case of transport mode choice in a Chilean city using real data.

Keywords Logit model · Endogeneity · Bias · Network externalities · Social interactions · Fixed point · Maximum entropy

JEL Codes C13 · C5 · C6 · R4

1 Introduction

Logit multinominal discrete choice models are frequently used in marketing to model consumer preferences, and are also employed in transport system planning to represent...
trip demand and ground use. They are traditionally built from random utility models and estimated using maximum likelihood (McFadden 1974; Ortúzar 1982; 1983; Train 2003; Ortúzar and Willumsen 2011). Alternatively, however, they have been derived as the solution to certain constrained entropy maximization problems in which the Lagrange multipliers of the constraints are the model’s parameters (Anas 1983; Donoso and De Grange 2010; Donoso et al. 2011).

In discrete choice models, interaction between supply and demand (Berry et al. 1995; Petrin and Train 2010) and the omission of variables unobservable to the researcher (Villas-Boas and Winer 1999) may cause endogeneity. In both cases, the problem can be handled by introducing instrumental variables into the estimation. The existence of endogeneity has also been traced to the definition of the choice set facing the individual (Haab and Hicks 1997; Louviere et al. 2005).

In this study, endogeneity is considered to be the result of network externalities or social interactions of the type peculiar to transport systems. More specifically, it emerges from the fact that the attractiveness of a particular alternative depends on the number of persons choosing it (Yamins et al. 2003; Blumenfeld-Lieberthal 2009; Bogart 2009). But whereas the attractiveness of a technology product, for example, often varies positively with user numbers, the choice of a transport mode can be negatively impacted by user numbers because it is influenced negatively by congestion. To deal with this phenomenon we propose a novel specification for logit models that corrects the endogeneity bias introduced by network externalities and social interactions.

The approach that will be taken is based on entropy maximization with explicit incorporation of the endogeneity caused by network externalities or social interactions into logit multinomial discrete choice models. An equivalent maximum entropy optimization problem is formulated in which the explanatory variables facing individuals depend on the decisions they make (e.g., with private transport, trip time depends on congestion; with public transport it depends on wait time or crowding at bus stops or metro stations). This interdependence produces endogeneity in the model’s explanatory variables, whose values will therefore depend on the individuals’ decisions.

The result of this approach is an alternative version of the logit multinomial model that has the functional form of a fixed-point equation in which the choice probabilities depend on themselves. Estimation is by maximum likelihood, and no additional variables (e.g., instruments) other than the original ones in the model are required, a major advantage of the proposed method.

The remainder of this article is organized into four sections. Section 2 introduces the theoretical framework for the proposed methodology and briefly surveys the literature on endogeneity in discrete choice models. Section 3formulates the Logit model with endogenous explanatory variables. Section 4 presents a numerical example using simulations and real data for a city in northern Chile. Finally, Section 5 sums up the main conclusions and contributions of this study.

2 Theoretical Framework and Literature Survey

Logit-type discrete choice models are developed using either of two approaches. One of them is based on random utility models (McFadden 1974; Williams 1977; Train 2003;
Ortúzar and Willumsen (2011) while the other involves formulating a maximum entropy optimization problem (Anas 1983; Boyce 2007; Hasan and Dashti 2007; De Cea et al. 2008; Donoso and De Grange 2010; Donoso et al. 2011; De Grange et al. 2010, 2011, 2013; Kitthamkesorn et al. 2014). Our proposed model takes the latter approach, specifying a maximum entropy optimization problem with constraints that explicitly incorporates the endogeneity phenomenon in the model’s explanatory variables.

In the random utility approach, an individual $i$ faced with a set of alternatives chooses the one that produces the greatest utility. Thus, the individual will choose alternative $m$ when $U_i^m > U_i^{m'} \forall m' \neq m$. The utility function $U_i^m$ is typically decomposed additively as $U_i^m = V_i^m + \varepsilon_i^m$, where $V_i^m$ is a deterministic component depending on observable variables and $\varepsilon_i^m$ a random component that is not observable. The observer does not directly observe the individuals’ actual utility functions $U_i^m$ but rather the choices they make and the attributes of each alternative, these latter constituting the definition of $V_i^m$. The deterministic component is typically expressed as a linear function in the attributes. Thus, if the $k$th attribute or explanatory variable faced by individual $i$ in alternative $m$ is defined as $x_{ki}^m \forall i, m, k$, then $V_i^m = \sum_k \beta_k^m x_{ki}^m$, where $\beta_k^m$ are the parameters to be estimated and represent the relative weights of each attribute.

The multinomial logit (MNL) model is obtained by assuming the random component of each utility function is independent and identically Gumbel-distributed (McFadden 1974; Ben-Akiva and Lerman 1985; Train 1986, 2003; Ortúzar and Willumsen 2011). The probability that individual $i$ chooses alternative $m$ is then given in general terms by

$$P_i^m = \frac{e^{V_i^m}}{\sum_{m'} e^{V_i^{m'}}}$$  \hspace{1cm} (1)

Under the second approach, on the other hand, (1) is derived by solving the following maximum entropy optimization problem (the full derivation is given in Appendix 1):

$$\min_{\{p_i^m\}} \sum_i \sum_m p_i^m (\ln p_i^m - 1)$$

s.t.:

\begin{align*}
\sum_m p_i^m &= 1 & (\Phi_i) \\
\sum_i p_i^m x_{ki}^m &= c_k^m & (\beta_k^m)
\end{align*}  \hspace{1cm} (2)

where variables $x_{ki}^m$ are the measurable exogenous attributes individual $i$ perceives in alternative $m$; and $c_k^m$ are the observed values for each attribute $k$ in each alternative $m$. Also, $e_k^m = \sum t \delta_i^m x_{ti}^m$, where $\delta_i^m$ is 1 if individual $i$ chooses alternative $m$ and 0 otherwise. The values of $x_{ki}^m$ and $\delta_i^m$ are traditionally obtained from surveys, measurements and calibration samples.

Anas (1983) gives a formal statement of the equivalence between the multinomial model based on random utility theory (1) and the maximum entropy problem (2). Donoso and De Grange (2010) show that an analysis of (2) yields two useful
interpretations of this equivalence: first, the maximum entropy problem is consistent with the rational decisions of welfare-maximizing individuals, and second, the likelihood function of the MNL model is equal to the problem’s dual. However, the equivalence is valid only if the constraints in (2) are linear, which is the case only when the attributes or variables $x^m_{ki}$ are exogenous.

In general terms, endogeneity in discrete choice models is present when part of the deterministic utility function $V^m_i$ is correlated with the error term $\varepsilon^m_i$ (Berry et al. 1995; Louviere et al. 2005; Guevara and Ben-Akiva 2009; Walker et al. 2011). When this occurs, the estimates of the parameters $\beta^m_{ki}$ may be inconsistent.

Endogeneity may appear for a variety of reasons, such as a model specification error due to the omission of important variables or the ability of individuals to influence the formation of the choice sets (Louviere et al. 2005). In this study we consider it to be the product of social interactions, which may be positive or negative. More specifically, the attributes of the alternatives are assumed to depend on the level of choice aggregation. For example, as more individuals choose to travel by private car, trip times may increase due to congestion. Likewise, overcrowding in buses or the metro, or wait times at stops or stations, may increase if there are many travellers at peak hours.

In addition, individuals may make decisions based on the actions of others due to incomplete information, such as occurs in herd behaviour (Banerjee 1992) or informational cascades (Bikhchandani et al. 1992). For example, diners may avoid a restaurant if there are few customers inside, taking it as a sign of high prices or poor quality. Similarly, a bus stop with no-one or many people waiting may indicate a disruption in service and thus discourage potential riders.

In all of these cases the explanatory or attribute variables $x^m_{ki}$ become endogenous variables of the type $x^m_{ki} = x^m_{ki}(t^m)$, where $t^m = \sum_i p^m_i$ is the total demand for alternative $m$ so that $x^m_{ki} = x^m_{ki}\left(\sum_i p^m_i\right)$.

Various ways of handling endogeneity in discrete choice models have been proposed in the literature. Blundell and Powell (2004) suggest a semi-parametric approach to testing for the exogeneity of continuous explanatory variables in binary choice models. Maximum likelihood is commonly employed with these models, but it requires an explicit parametric specification of the way in which each endogenous variable depends on a set of instruments and the errors. Furthermore, a joint distribution of the random component of the utility functions and the error component in the relationship between an endogenous variable and the instruments must be specified (Lewbel 2007). These requirements can limit maximum likelihood’s usefulness due to the difficulties often involved in specifying these relationships correctly. De Grange et al. (2009) present an alternative method for estimating aggregate logit models by multiple linear regression that is based on proxy variables, thus circumventing the endogeneity problems arising when least-squares estimators are used. Batarce and Ivaldi (2014) formulate a structural model for travel demand in which the congestion level is endogenously determined in the equilibrium of a game with a continuum of players. The estimation uses the first-order conditions of the users’ utility maximization problem to derive the likelihood function and then a two-step semi-parametric method.
Of the various approaches that have been suggested to correct for endogeneity, five in particular are worthy of note:

i. control function approach
ii. the Berry, Levinsohn and Pakes (BLP) approach
iii. dual approach
iv. latent variables approach
v. special regressor approach

Each of these are briefly described in what follows.

i. Control function approach. This method is described in Heckman (1976), Hausman (1978), Heckman and Robb (1985), Villas-Boas and Winer (1999), Blundell and Powell (2004), Guevara and Ben-Akiva (2006), and Petrin and Train (2010). It involves two stages. First, the endogenous variable is regressed on exogenous instruments; then, the residual (or a function of it) is incorporated into the utility function as an additional explanatory variable called the control function (Louviere et al. 2005; Guevara and Ben-Akiva 2009). As with maximum likelihood, this approach requires that the relationship between the endogenous regressors and the instrument be correctly specified.

ii. BLP approach. As described by Berry et al. (1995), this method is applicable when endogeneity arises with groups of individual decision-makers, such as when the endogenous variable is separable for individuals in geographically distinct locations. In other words, the endogenous variable varies among different groups of individuals but is the same for individuals within a single group. The method consists in linearising the discrete choice model (normally multinomial logit) and then applying classical instrumental variable techniques (Louviere et al. 2005).

iii. Dual approach. This method has been suggested by Matzkin (2004). It is applicable to both linear and non-linear models with endogeneity and can be interpreted as a dual of the instrumental variable approach. It consists in finding new explanatory variables \( \tilde{X} \) that are correlated with the original endogenous variable \( X \) and may also be endogenous, but the difference \( \varepsilon \) between them must be completely random (i.e., \( \varepsilon = (\rho \tilde{X} - \gamma X) \) where \( (\rho, \gamma) > 0 \)) so the method does not require exogenous instruments \( Z \).

iv. In discrete choice models, the new variable \( \tilde{X} \) is included in the utility function together with the endogenous variable \( X \). Train and Winston (2004) model the choice of vehicle make as a function of the resale price \( \tilde{X} \), arguing that the latter variable is correlated with the new vehicle price, the endogenous variable \( X \).

v. Latent variable approach. This method is set out in Ben-Akiva and Boccara 1995; Walker and Ben-Akiva 2002; and Guevara and Ben-Akiva 2009. It involves the explicit incorporation into the model of unobservable or latent variables (e.g., psychological factors, attitudes or perceptions). The missing variable causing the endogeneity is thus considered to be a latent variable and is modelled as part of a system of structural equations that are a function of observable variables. It is then incorporated into the utility function.
vi. **Special regressor approach.** This method has been widely used with discrete and ordered choice models and models with censored data and selection bias (Lewbel 1998, 2000, 2007; Magnac and Maurin 2007, 2008) as well as models with truncated data (Khan and Lewbel 2007) and panel data (Ai and Gan 2010). The special regressor estimator is a control function but the identification of the endogenous regressor’s coefficient considers the remaining terms of the latent variable as a special regressor.

The proposed method set out in the following section is similar to the control function approach but has the advantage of not requiring that a functional form be specified, nor does it require exogenous instruments.

### 3 Formulation and Estimation of Logit Model With Endogenous Explanatory Variables (MNLE)

#### 3.1 Mathematical Formulation

As was explained in Section 2, we represent the network externalities or social interactions of the explanatory variables by functions of the type $x_{ki}^m = x_{ki}^m(t^m)$, where $t^m = \sum_i p_i^m$ is the total demand for alternative $m$ so that $x_{ki}^m = x_{ki}^m \left( \sum_i p_i^m \right)$.

We can now solve the following equivalent optimization problem, incorporating explicitly the variables $x_{ki}^m = x_{ki}^m(t^m)$ with $\lambda = 1$ (for an extension to the hierarchical logit model with endogeneity, see Appendices 2 and 3):

$$\min_{\{p_i^m\}} \sum_i \sum_m p_i^m (\ln p_i^m - 1)$$

s.t.: $\sum_i p_i^m x_{ki}^m = \sum_i \delta_i x_{ki}^m \quad \forall m \ (\beta_k^m)$

$$\sum_m p_i^m = 1 \quad \forall i \ (\Phi_i)$$

The optimality conditions of problem (3) are

$$\frac{dL}{dp_i^m} = \ln p_i^m - \Phi_i - \sum_k \beta_k^m \left( x_{ki}^m + p_i^m \frac{dx_{ki}^m}{dp_i^m} - \delta_i \frac{dx_{ki}^m}{dp_i^m} \right) = 0$$

$$p_i^m = \exp \left( \Phi_i + \sum_k \beta_k^m \left( x_{ki}^m + (p_i^m - \delta_i) \frac{dx_{ki}^m}{dp_i^m} \right) \right)$$

Summing over $m$ and taking the natural logarithm, we have

$$\sum_m p_i^m = 1 \rightarrow \sum_m \exp \left( \Phi_i + \sum_k \beta_k^m \left( x_{ki}^m + (p_i^m - \delta_i) \frac{dx_{ki}^m}{dp_i^m} \right) \right) = 1$$
\[ \Phi_i = -\ln \sum_m \exp \left( \sum_k \beta_k \left( x_{ki}^m + \left( p_i^m - \delta_i^m \right) \frac{dx_{ki}}{dp_i^m} \right) \right) \]  

(7)

Substituting (7) into (5), we get the probability that individual \( i \) chooses alternative \( m \):

\[ p_i^m = \frac{\exp \left( \sum_k \beta_k \left( x_{ki}^m + \left( p_i^m - \delta_i^m \right) \frac{dx_{ki}}{dp_i^m} \right) \right)}{\sum_{m'} \exp \left( \sum_k \beta_k \left( x_{ki}^{m'} + \left( p_i^{m'} - \delta_i^{m'} \right) \frac{dx_{ki}}{dp_i^{m'}} \right) \right)} \]  

(8)

This expression is similar to the traditional logit multinomial specification except that it explicitly incorporates the phenomenon of endogeneity of the variables \( x_{ki}^m \) due to network externalities or social interactions. We thus describe it as a logit model with endogenous explanatory variables (MNLE). It is a fixed-point function in \( p_i^m \), whose estimation runs up against two difficulties: first, solving the fixed point; and second, estimating \( \frac{dx_{ki}}{dp_i^m} \).

MNLE can be considered a member of the family of discrete choice models with social interactions and heterogeneity (Brock and Durlauf 2001, 2006; Soetevent and Kooreman 2007; Dugundji and Gulyás 2008; Amador et al. 2008). In these formulations, an individual’s decision may be affected by the decisions of other groups of individuals in society.

When \( \frac{dx_{ki}}{dp_i^m} = 0 \), (8) reduces to the traditional MNL model but when \( \frac{dx_{ki}}{dp_i^m} \neq 0 \), a traditional MNL such as (1) would be incorrectly specified because the term \( \left( p_i^m - \delta_i^m \right) \frac{dx_{ki}}{dp_i^m} \) present in (8) is missing. This omission results in biased estimations of the parameters \( \beta_k^m \). The missing term corrects the explanatory variables \( x_{ki}^m \) by incorporating the effect of the network externality or social interaction type of endogeneity.

3.2 Parameter Estimation

Multinomial logit and other analogous models are typically estimated by maximum likelihood. For the fixed-point model (8) just described, however, a practical complication arises due to the presence of the \( p_i^m \) term on the right-hand side of the equation. A simple way of getting around this difficulty is to estimate the model in two steps. In the first step, the term \( \frac{dx_{ki}}{dp_i^m} \) is estimated exogenously. For example, if the variable \( x_{ki}^m \) represents trip time by private transport, the parameter \( \frac{dx_{ki}}{dp_i^m} \) can be estimated from the road network flow-delay functions, as will be done here when the model is applied to real data (Section 4.2). If the model is being used for empirical work, either additional data on the parameter value or valid instruments will be needed to obtain a consistent estimate. Yet another way is to make a conjecture and then conduct a sensitivity analysis on it to get an idea of the order of magnitude of the possible bias.

A starting point must also be specified for an estimate of \( p_i^m \). This can be obtained using the multinomial logit model (Raveau et al. 2011; De Grange et al. 2013), which is
the equivalent of setting \( \frac{dx^m_i}{dp^m_i} = 0 \) in (8) and provides a warm start. Thus,

\[
p^m_{i,0} = \frac{\exp \left( \sum_k \beta_{k,0}^m x^m_{ki} \right)}{\sum_{m'} \exp \left( \sum_k \beta_{k,0}^{m'} x^{m'}_{ki} \right)}
\]

(9)

where the values of \( \beta_{k,0}^m \) are the maximum likelihood estimators of the parameters.

Once \( p^m_{i,0} \) and \( \frac{dx^m_i}{dp^m_i} \) have been estimated we proceed to the second step, which is to estimate directly via maximum likelihood the parameters \( \beta_{k}^m \) of the following model:

\[
p^m_{i,1} = \frac{\exp \left( \sum_k \beta_{k,1}^m \left( x^m_{ki} + \left( p^m_{i,0} - \delta^m_i \right) \frac{dx^m_{ki}}{dp^m_i} \right) \right)}{\sum_{m'} \exp \left( \sum_k \beta_{k,1}^{m'} \left( x^{m'}_{ki} + \left( p^{m'}_{i,0} - \delta^{m'}_i \right) \frac{dx^{m'}_{ki}}{dp^{m'}_i} \right) \right)}
\]

(10)

This process is iterated until \( \beta_{k,n}^m \approx \beta_{k,n-1}^m \) \( \forall k, m \), which implies \( p^m_{i,n} \approx p^m_{i,n-1} \) at the fixed point defined in (10). For the case where \( \frac{dx^m_i}{dp^m_i} \) is constant \( \forall i \), the sufficient conditions of existence and uniqueness for the equilibrium solution of this iterative process are set out in what follows.

**Theorem** Let \( \eta^m = \sum_k \left( \beta_k^m \frac{dx^m_{ki}}{dp^m_i} \right) \), where \( \frac{dx^m_i}{dp^m_i} = \text{const} \), \( \forall i \). If either (i) \( \max_m |\eta^m| < 2 \), or (ii) \( \sum_m |\eta^m| < 4 \), then both the existence and the uniqueness of a fixed point are assured and the linear convergence of the iterative method just described is guaranteed.

**Proof** Let \( f(p)=(p^m_i)\in \mathbb{R}^i \times n \rightarrow f(p)\in \mathbb{R}^i \times n \) be

\[
f(p) = \left( \frac{\exp \left( \eta^m \left( p^m_i - \delta^m_i \right) + \sum_k \beta_k^m x_{ik}^m \right)}{\sum_{m'} \exp \left( \eta^{m'} \left( p^{m'}_i - \delta^{m'}_i \right) + \sum_k \beta_k^{m'} x_{ik}^{m'} \right)} \right)^m_i
\]

(11)

where parameters \( \beta_k^m \) and \( \eta^m \) are the points that optimize the log-likelihood function of the MNL. This is a continuous function (De Grange et al. 2013) and the recursive estimation is such that \( f(p^{(n)})=p^{(n+1)} \).

Let \( p, q \) be any two points. For any norm,

\[
\|f(p)-f(q)\| \leq \|J(r)(p-q)\| \leq \|J(r)\|\|p-q\|
\]

(12)
where \( J \) is the Jacobian of \( f(\cdot) \) and \( r \) is a point such that \( r = \theta p + (1-\theta)q, \ \theta \in [0,1] \).

\[
J^m_{ij} = \frac{\partial f^m_i}{\partial p^j} = \begin{cases} 
\eta^m f^m_i (1-f^m_i), & i = j; k = m \\
-\eta^m f^m_i f^m_k, & i = j; k \neq m \\
0, & \# \neq j 
\end{cases}
\quad (13)
\]

Matrix norm \( \|\cdot\|_1 \) and matrix norm \( \|\cdot\|_\infty \) are applied to Jacobian (13) to get (14) and (15), respectively. Thus,

\[
\|J(r)\|_1 = \max_i \max_m |\eta^m| f^m_i (1-f^m_i) + |\eta^m| f^m_i \sum_{k \neq m} f^k_i \leq 2 \max_i |\eta^m| \max_i f^m_i (1-f^m_i) \leq \frac{1}{2} \max_m |\eta^m|, \forall r
\]

\[
\|J(r)\|_\infty = \max_i \max_m |\eta^m| f^m_i (1-f^m_i) + \sum_{k \neq m} |\eta^k| f^k_i \leq \max_i \max_m f^m_i (1-f^m_i) \sum_k |\eta^k| \leq \frac{1}{4} \sum_k |\eta^k|, \forall r
\]

Since \( f^m_i f^k_i \leq f^m_i (1-f^m_i) \leq \frac{1}{4} \), \( \forall i, k, m \), we obtain

\[
\max_m |\eta^m| < 2 \Rightarrow \| f(p) - f(q) \|_1 < L |p-q|_1, \quad L \in [0, 1) \quad (14)
\]

\[
\sum_{m'} |\eta^{m'}| < 4 \Rightarrow \| f(p) - f(q) \|_\infty < L |p-q|_\infty, \quad L \in [0, 1) \quad (15)
\]

If condition (i) is satisfied, \( f(\cdot) \) is proven to be a contraction for norm \( \|\cdot\|_1 \), and if condition (ii) is satisfied, \( f(\cdot) \) is a contraction for norm \( \|\cdot\|_\infty \). In both cases, by the Banach fixed point theorem, a unique fixed point exists. Furthermore, since all norms are equivalent in a finite-dimensional space, the linear convergence of the iterative method is demonstrated for any norm.

### 4 Numerical Examples

In this section we review the performance of the model set out above in the presence of endogeneity using two sets of data, one generated by a Monte Carlo simulation and the other taken from an urban transport survey in the Chilean city of Antofagasta.

#### 4.1 Monte Carlo Simulation

Consider a simple case in which there are only two transport alternatives (private car and bus) and one explanatory variable (trip time) with a common parameter \( \beta_{\text{time}} \). The utility functions for the two modes are

\[
V_{\text{car}}^{\text{it}} = \beta_0 + \beta_{\text{time}} T_{\text{car}}^{\text{it}} \quad (16)
\]

\[
V_{\text{bus}}^{\text{it}} = \beta_{\text{time}} T_{\text{bus}}^{\text{it}} \quad (17)
\]

where the private car trip time in minutes observed by individual \( i \) is a linear function of the total number of car trips in a given period \( t (F_{\text{car}}^{\text{it}}) \) plus a random term \( \varepsilon_{\text{car}}^{\text{it}} \), and is thus expressed as

\[
T_{\text{car}}^{\text{it}} = \alpha + \gamma F_{\text{car}}^{\text{it}} + u_{\text{car}}^{\text{it}} \quad (18)
\]
where \( \alpha = 3 \), \( \gamma = 1/15 \), and \( F_{t}^{\text{car}} \sim U(100;500) \), so that the total flow of cars during period \( t \) is 100 to 500 vehicles uniformly distributed. Different trip time intervals can be chosen as a function of aggregate demand. Since the number of trips will be different in each simulation, the number of observations (i.e., the sample size) in each case will also differ (each traveller or user is an observation). A total of 130 simulations were conducted with sample sizes ranging from 150 to 800 users or observations.

The variable \( u_{it}^{\text{car}} \) distributes uniformly between 0 and \( (\alpha + \gamma F_{t}^{\text{car}})/2 \). The bus trip time is independent of demand and is defined as \( T_{it}^{\text{bus}} \sim U(8;60) \). Finally, the population parameters are set at \( \beta_{\text{time}} = -0.25 \) and \( \beta_{0} = -0.15 \).

Since by construction \( \sum_{l} p_{i}^{m} = \sum_{l} \delta_{i}^{m} \), it is easily demonstrated for this simulation exercise that \( \sum_{l} p_{it}^{\text{car}} = \sum_{l} \delta_{it}^{\text{car}} = F_{t}^{\text{car}} \), and therefore \( \frac{dF_{t}^{\text{car}}}{dp_{it}^{\text{car}}} = 1 \) and \( \frac{dT_{it}^{\text{car}}}{dp_{it}^{\text{car}}} = \frac{dF_{t}^{\text{car}}}{dT_{it}^{\text{car}}} = 1 \) and \( \frac{dF_{t}^{\text{car}}}{dT_{it}^{\text{car}}} = \frac{1}{\gamma} = 15 \).

The values estimated for \( \beta_{\text{time}} \) using the traditional MNL model (1) above and MNLE model (8), the latter explicitly reflecting the endogeneity in the private car trip time variable, are compared by the histograms in Fig. 1. The histograms for the parameter \( \beta_{0} \) are shown in Fig. 2. In Appendix 4 we show simulation results using \( \gamma = 1 \) and \( \gamma = 2 \).

The mean of the \( \beta_{\text{time}} \) estimator for both models in the presence of trip time congestion \( (\gamma = 1/15) \) is shown in Table 1 along with the result of a simulation using the same population parameters \( (\beta_{\text{time}} = -0.25 \) and \( \beta_{0} = -0.15 \) but no congestion (i.e.,

Fig. 1 Distribution of the \( \beta_{\text{time}} \) parameter estimator for the MNL and MNLE models \( (\gamma = 1/15) \)
\(\gamma=0\) and thus no endogeneity. The bias in the simulation estimate of the mean of \(\beta_{\text{time}}\) using MNL with endogeneity in trip time (e.g., congestion) is evident in the figure of \(-0.125\). At this value the null hypothesis \(H_0: \beta_{\text{time}}=-0.25\), the known population parameter, is rejected with a high level of significance. The mean estimate produced by MNLE, on the other hand, is not significantly different from the known parameter value and the null hypothesis cannot be rejected. In the bottom row of the table, it can be seen that when there is no endogeneity \((\gamma=0)\) the classic MNL’s estimate is consistent.

The results for parameter \(\beta_0\) are given in Table 2. As with \(\beta_{\text{time}}\), the MNL estimate in the presence of endogeneity is biased whereas the MNLE estimate is statistically unbiased (as is MNL without endogeneity in the bottom row of the table).

A dispersion graph of the bias estimated in each simulation for the \(\beta_{\text{time}}\) parameter and the level of demand or flow \(F_t^{\text{car}}\) impacting the level of congestion is shown in

![Graph showing distribution of beta_0 parameter estimator for MNL and MNLE models (gamma=1/15)](image)

**Table 1** \(\beta_{\text{time}}\) parameter estimators

| MODEL         | \(\hat{\beta}_{\text{time}}\) | Standard Error | \(t\) Test (\(\beta_{\text{time}}=-0.25\)) |
|---------------|-------------------------------|----------------|----------------------------------|
| MNL \((\gamma=1/15)\) | -0.125                        | 0.046          | 2.704                            |
| MNLE \((\gamma=1/15)\)   | -0.262                        | 0.058          | -0.211                           |
| MNL \((\gamma=0)\)       | -0.255                        | 0.056          | -0.084                           |
As is apparent, increasing congestion induces a downward bias in the MNL model estimates. Thus, the greater is the congestion the greater will be the MNL estimate bias unless the endogeneity is corrected.

As regards the models’ goodness-of-fit, four indicators are reported in Table 3: log-likelihood evaluated at the parameter estimate values ($L^*$), log-likelihood evaluated at zero ($L^0$), and rho-square ($\rho^2$) and adjusted rho-square ($\bar{\rho}^2$) where

$$\rho^2 = 1 - \frac{L^*}{L^0}, \quad \bar{\rho}^2 = 1 - \frac{L^*-K}{L^0}$$

and $K$ is the number of estimated parameters. The values shown are the averages of the simulation results for each statistic. They clearly show that MNLE has better goodness-of-fit on various indicators.

**Table 2** $\beta_0$ parameter estimation

| MODEL          | $\hat{\beta}_0$ | Standard Error | $t$ Test ($\beta_0=-0.15$) |
|---------------|-----------------|----------------|-----------------------------|
| MNL ($\gamma=1/15$) | -1.102          | 0.431          | -2.209                      |
| MNLE ($\gamma=1/15$) | -0.194          | 0.342          | -0.128                      |
| MNL ($\gamma=0$)    | -0.131          | 0.185          | 0.105                       |

Fig. 3. Relation between $\beta_{time}$ estimator bias and flow in MNL models ($\gamma=1/15$)
The Horowitz test (1983) for comparing non-nested discrete choice models can also be used to compare the two models. The null hypothesis is the model with more parameters does not have a better fit. The value of this test statistic is given by

$$
\Phi \left\{ -2 \left( \hat{\rho}_h^2 - \hat{\rho}_l^2 \right) \cdot L^0 + (K_h - K_l) \right\}^{\frac{1}{2}} \sim N(0, 1)
$$

where:

- $\hat{\rho}_l^2$ is the adjusted likelihood ratio index for the model with the lowest ($l$) value;
- $\hat{\rho}_h^2$ is the adjusted likelihood ratio index for the model with the highest ($h$) value (in our case, it was the MNLE);
- $K_h, K_l$ are the numbers of parameters in models $h$ and $l$, respectively;
- $\Phi$ is the standard normal cumulative distribution function.

Using a 95 % level of confidence the criterion for rejecting the null hypothesis if $|\Phi^{-1}| > 1.96$. The average value of $\Phi^{-1}$ for various simulations was $-3.934$ indicating that the MNLE was had a better fit than MNL.

These findings can be complemented with the Hausman specification test statistic (Hausman 1978; Marquez-Ramos et al. 2011), which in the present case is expressed as

$$
H = \left[ \hat{\beta}_{\text{MNL}} - \hat{\beta}_{\text{MNLE}} \right]^T \left[ \text{var} \left( \hat{\beta}_{\text{MNL}} \right) - \text{var} \left( \hat{\beta}_{\text{MNLE}} \right) \right]^{-1} \left[ \hat{\beta}_{\text{MNL}} - \hat{\beta}_{\text{MNLE}} \right] \sim \chi^2_q
$$

where $\left( \hat{\beta}_{\text{MNL}} \right)$ is the parameter vector estimated by the MNL model and $\left( \hat{\beta}_{\text{MNLE}} \right)$ the corresponding vector estimated by MNLE. Analogously, $\text{var} \left( \hat{\beta}_{\text{MNL}} \right)$ is the variance and covariance matrix of the MNL parameter estimates and $\text{var} \left( \hat{\beta}_{\text{MNLE}} \right)$ the corresponding matrix for the MNLE. The null hypothesis of the test is that the selected parameters in both models are statistically equal. $H$ is chi-squared distributed with $q$ degrees of freedom (in this case 2, the number of parameters, and the critical value at the 5 % level of significance is 5.99). The value of the statistic is $H = 18.58 > 5.99$ supporting that the MNL estimator is more biased than that of the MNLE.

Thus, for the simulations that were conducted, the endogeneity in the trip time explanatory variable caused by congestion causes in the MNL model estimate whose order of magnitude is significant compared to the parameter values. By contrast, MNLE produces consistent estimators that fit the data better.

A sensitivity analysis was carried out on the $\frac{dT_{\text{car}}}{dp_{it}^a}$ term. As noted above, the relationship between trip time and demand defined for the simulations was such that $\frac{dT_{\text{car}}}{dp_{it}^a} = \frac{1}{\gamma} \approx 0.07$. By modifying this value slightly in the MNLE model estimation,
variations are produced in the $\beta_{time}$ parameter bias and the log-likelihood value. The results are graphed for the former in Fig. 4 and for the latter in Fig. 5. Figure 4 shows the

![Fig. 4](image1.png)  
**Fig. 4** Sensitivity analysis: $dX/dP$ and log-likelihood

![Fig. 5](image2.png)  
**Fig. 5** Sensitivity analysis: $dX/dP$ and $\beta_{time}$Bias
values taken by the likelihood function for different values of \( \frac{dx_n^m}{dp_i^m} \) (by trial and error). Figure 5 show the bias for different values of \( \frac{dx_n^m}{dp_i^m} \) (by trial and error). As may be observed, the highest log-likelihood values are obtained when \( \frac{dx_n^m}{dp_i^m} \) is relatively close to 1/15≈0.07 while the lowest bias levels in \( \beta_{time} \) are also found close to that point. These relationships will be useful for estimating the MNLE model in real-world situations where the functional form of \( \frac{dx_n^m}{dp_i^m} \) is not known.

4.2 Application of MNLE Model to Real Data

The multinomial model with endogeneity (MNLE) as expressed by (8) was tested on real data from a trip survey conducted in the northern Chilean city of Antofagasta (SECTRA 2011). Antofagasta is the city with the highest per capita income of Chile, and the fifth most populous city with about 400,000 inhabitants in an area of 125,000 square kilometers. Figure 6 shows the hourly distribution of trips in the city.

The peak period (7 am to 9 am) represents 20 % of daily trips. Observed modal split is 48 % for Bus, 33 % for Private Car and 19 % for Share Taxi. The average travel time for Private Car is approximately 10.6 minutes. For the off-peak period the average travel time is close to 7 minutes, and the observed modal split is 52 % for Bus, 26 % for Private Car and 23 % for Share Taxi.

A total of 903 individual commuter trips during the morning peak hour (7 am to 9 am) were surveyed in three transport modes:

i. Private Car
ii. Bus
iii. Share Taxi

For comparison purposes a conventional MNL model was also tested. The explanatory variables in the two models were trip time for mode \( m \ (T_m^m) \) and cost (\( cosinc_i \)), the

---

**Fig. 6** Trips distribution, week day
latter divided by traveller income. The utility functions for the three models are presented in Table 4. The estimation process was performed using BIOGEME (Bierlaire 2003; Bhat and Guo 2004).

The modal constant of the bus mode ($\beta_{bus} = 0$) was set as the reference (zero) value and the other two specific modal constants were then estimated with respect to it. Generic parameters ($\beta_k^m = \beta_k, \forall m$) were used for trip time and cost.

To estimate the models, a functional form had to be chosen for $\frac{dx^m}{dp_i}$, $\forall k, m$. It was assumed that only the trip time variable (i.e., not cost) could be endogenous. Thus, $\frac{dx^m}{dp_i}$ was interpreted as the ratio of the variation in total trip time between peak and off-peak periods for private car travel to the variation between peak and off-peak periods in the total number of trips. Analytically, this ratio can be written as

$$\frac{dT}{dp_i} \approx \frac{T_{\text{total peak}} - T_{\text{total off-peak}}}{F_{\text{total peak}} - F_{\text{total off-peak}}}$$  \hspace{1cm} (22)$$

where $T_{\text{total peak}}$ and $T_{\text{total off-peak}}$ are the total private car trip times in peak and off-peak periods, respectively, and $F_{\text{total peak}}$ and $F_{\text{total off-peak}}$ are the corresponding total private car

| Mode       | Function                                      |
|------------|-----------------------------------------------|
| Private Car| $V_i^{\text{car}} = \beta_0^{\text{car}} + \beta_{\text{time}} \cdot T_i^{\text{car}} + \beta_c \cdot \text{cosinc}_i$ |
| Bus        | $V_i^{\text{bus}} = \beta_{\text{time}} \cdot T_i^{\text{bus}} + \beta_c \cdot \text{cosinc}_i$ |
| Share Taxi | $V_i^{\text{share}} = \beta_0^{\text{share}} + \beta_{\text{time}} \cdot T_i^{\text{share}} + \beta_c \cdot \text{cosinc}_i$ |

Fig. 7 Sensitivity analysis: dX/dP and log-likelihood considering real data (MNLE-2)
trips. The conjectured value of $\frac{dT_{\text{car}}}{dp_{\text{car}}}i$ was thus estimated using the Antofagasta survey data (SECTRA 2011) to be 1.82. The model in this case is called MNLE-1.

Another way to estimate $\frac{dT_{\text{car}}}{dp_{\text{car}}}i$ is by a “trial and error” process (see Fig. 7). Using many different values of $\frac{dT_{\text{car}}}{dp_{\text{car}}}i$, we choose one that produces the highest value of the log-likelihood function. In this second case, we obtain $\frac{dT_{\text{car}}}{dp_{\text{car}}}i = 1.6$ and a better goodness of fit. The model in this second case is called MNLE-2.

The results of the estimation for MNL, MNLE-1 and MNLE-2 are set out in Table 5. All of the parameter estimates have the correct sign and a high level of statistical significance.

For the MNLE-1, the Horowitz test was applied to these estimates was $\Phi^{-1} = -5.304$ rejecting the null hypothesis of the test and indicating a better fit of the MNLE model compared to the MNL.

These outcomes are reinforced by the results of the Hausman test applied to the two groups of estimators in Table 5. The chi-square critical value for $H$ statistic at a 95 % confidence level with four degrees of freedom is 9.49. In this case we have $H = 13.97 > 9.49$, rejecting the null hypothesis of the test. This confirms that MNLE achieves a closer fit and supports the suitability of our conjecture setting $\frac{dT_{\text{car}}}{dp_{\text{car}}}i$ at 1.82.

For the MNLE-2, the Horowitz test was applied to these estimates was $\Phi^{-1} = -5.402$ rejecting the null hypothesis. In the Hausman test we have obtained that $H = 14.18 > 9.49$, rejecting again the null hypothesis of the test.

Finally, using the Horowitz test to compare MNLE-1 with MNLE-2, we found no statistical difference between the two models.

### Table 5 Parameter estimates for MNL and MNLE models (*)

| Parameter       | MNL          | MNLE-1 ($\frac{dx_{\text{car}}^m}{dp_{\text{car}}^m} = 1.82$) | MNLE-2 ($\frac{dx_{\text{car}}^m}{dp_{\text{car}}^m} = 1.6$) |
|-----------------|--------------|-------------------------------------------------|-------------------------------------------------|
| $\beta_0^{\text{car}}$ | -0.082 (-0.51) | -0.518 (-2.51) | -0.569 (-2.70) |
| $\beta_0^{\text{share}}$ | -0.473 (-4.42) | -0.521 (-4.74) | -0.528 (-4.80) |
| $\beta_c$ | -0.122 (-6.54) | -0.135 (-6.75) | -0.136 (-6.77) |
| $\beta_{\text{time}}$ | -0.023 (-4.72) | -0.039 (-5.75) | -0.041 (-5.77) |
| $L^*$ | -679.54 | -665.48 | -663.81 |
| $L^0$ | -864.74 | -864.74 | -864.74 |
| $\rho^2$ (**) | 0.214 | 0.230 | 0.232 |
| $\tau^2$ (****) | 0.210 | 0.226 | 0.227 |
| $PCP$ (*****) | 68.4 % | 68.9 % | 68.9 % |
| $RSS$ (***** | 242.64 | 235.86 | 235.03 |

(*): Figures in parentheses are statistical significance levels.

(**): $\rho^2 = 1 - L^*/L^0$

(***): $\tau^2 = 1 - (L^* - 4) / L^0$

(****): $PCP$ is percent of correct prediction.

(*****): $RSS = \sum \frac{(\delta^m - \hat{y}_i^m)^2}{2}$
5 Conclusions

A new approach was presented for dealing with endogeneity in choice models where the attributes of the alternatives or the explanatory variables and their values are endogenous because they depend on individual choice decisions. This type of endogeneity typically arises in the context of social interactions or networks subject to network externalities. The article addressed a transport network setting in which public transport trip times depend on wait times or crowding at bus stops or train stations and private car trip times are subject to road congestion.

The proposed methodology explicitly incorporates this type of endogeneity into logit multinomial discrete choice models via the formulation of an equivalent maximum entropy optimization problem. The solution of the problem is a logit model with a fixed-point functional form that is calibrated through maximum likelihood estimation. This model is extendible to hierarchical multinomial formulations.

The approach was tested by comparing the new model’s performance with that of a traditional logit model in two different applications, one using simulated data and the other employing real data. The simulated data were generated with endogeneity in the explanatory variable. The new model corrected the bias in the parameter estimates produced by the traditional formulation. The greater was the degree of endogeneity (i.e., level of congestion), the greater was the estimated bias. The proposed model also achieved a tighter fit to the data according to several goodness-of-fit indicators and statistical tests.

In the case of the real data, drawn from a trip time survey conducted in a Chilean city, the degree of endogeneity was not known with certainty and a conjecture had to be estimated using the data themselves. The estimation of the model demonstrated that if the endogeneity was of the order of magnitude conjectured, the traditional logit model’s estimators were considerably biased whereas the proposed formulation fit the data relatively well.

Acknowledgments This research was supported by FONDECYT Grant No. 11121199.

Appendix 1: Microeconomic Deduction of the Maximum Entropy Problem for the Multinomial Logit Model

Assuming a Gumbel-distributed function, the maximum expected utility \( EMU_i \) for an individual is

\[
EMU_i = \frac{1}{\lambda} \ln \sum m \exp(\lambda V_i^m) \tag{23}
\]

It is known that for the Gumbel distribution,

\[
p_i^m = \frac{\exp(\lambda V_i^m)}{\sum m' \exp(\lambda V_i^{m'})} \rightarrow \sum m' \exp(\lambda V_i^{m'}) = \frac{\exp(\lambda V_i^m)}{p_i^m} \tag{24}
\]
Substituting (24) into (23) we obtain

\[ EMU_i = \frac{1}{\lambda} \ln \left( \frac{\exp(\lambda V^m_i)}{p^m_i} \right) = V^m_i - \frac{1}{\lambda} \ln p^m_i \]  

(25)

Multiplying (25) by \( p^m_i \) and summing over \( m \), we have

\[ p^m_i \cdot EMU_i = p^m_i V^m_i - \frac{1}{\lambda} p^m_i \ln p^m_i \]  

(26)

\[ EMU_i \sum_m p^m_i = \sum_m p^m_i V^m_i - \frac{1}{\lambda} \sum_m p^m_i \ln p^m_i \]  

(27)

\[ EMU_i = \sum_m p^m_i V^m_i - \frac{1}{\lambda} \sum_m p^m_i \ln p^m_i \]  

(28)

\[ EMU_i = \sum_m p^m_i V^m_i - \frac{1}{\lambda} \sum_m p^m_i \ln p^m_i = 0 \]  

(29)

It is evident from (29) that the entropy can be interpreted as the difference between the maximum expected utility of the individual and the average utility the individual faces. From this result we can also formulate the following optimization problem that each individual solves:

\[
\max \{ p^m_i \} \quad EMU_i = \sum_m V^m_i p^m_i - \frac{1}{\lambda} \sum_m p^m_i (\ln p^m_i - 1) \\
\text{s.t.} : \sum_m p^m_i = 1 \quad (\Phi_i)
\]  

(30)

Problem (30) represents the optimal decisions the individual will make in a mixed strategy. The optimality conditions for (30) are

\[ p^m_i = \frac{\exp(\lambda V^m_i)}{\sum_{m'} \exp(\lambda V^{m'}_i)} , \quad \Phi_i = \frac{1}{\lambda} \ln \left( \sum_m \exp(\lambda V^m_i) \right) \]  

(31)

Assuming the linear additive utility function \( V^m_i = \sum_k \beta^m_{ik} x^m_k \) for each individual \( i \) and that multiple individuals make their optimal decision simultaneously (based on mixed strategies), we formulate the following optimization problem:

\[
\min \{ p^m_i \} \quad Z_1 = \sum_i \sum_m \frac{1}{\lambda} p^m_i \ln p^m_i \\
\text{s.t.} : \sum_i p^m_i x^m_i = X^m_k \quad \forall k, m \quad (\beta^m_k) \\
\sum_m p^m_i = 1 \quad \forall i \quad (\Phi_i)
\]  

(32 to 34)

where the variable \( p^m_i \) is the probability (in a mixed strategy) that individual \( i \) chooses alternative \( m \) and constraints (34) impose that each individual’s probabilities
must add up to 1. The parameter $\lambda$ is a scalar that must be positive. Constraints (33), whose Lagrange multipliers give the specific constants ($\beta_{0m}$) for each alternative, ensure that the observed proportion of individuals choosing alternative $m$ is the same as the proportion predicted by the model. This result is also obtained when the constant terms are incorporated in maximum likelihood estimations of MNL models based on random utility theory.

Note also that one of the constraints in (33) is linearly dependent on the others, implying that one of the $\beta_{0m}$ parameters must be 0. This is equivalent to what is done in random utility theory, where one of the constant terms is also set to 0 for identifiability of parameters.

Finally, the $x_{ik}^m$ variables are the explanatory variables of alternative $m$ perceived by individual $i$ and $X_k^m$ is the generalized total cost of variable $k$ for alternative $m$.

Appendix 2: Microeconomic Deduction of the Maximum Entropy Problem for the Hierarchical Logit Model

The maximum expected utility ($EMU_i$) of an individual in a hierarchical choice structure is given by

$$EMU_i = \frac{1}{\alpha} \ln \sum_n \exp(\alpha L_i^n), \quad L_i^n = \frac{1}{\lambda} \ln \sum_{m \in n} \exp(\lambda V_{im}^m)$$

(35)

In this case we know that

$$p_{im}^{nm} = \frac{\exp(\alpha L_i^n)}{\sum_{n'} \exp(\alpha L_i^{n'})} \frac{\exp(\lambda V_{im}^m)}{\sum_{m' \in n} \exp(\lambda V_{im}^m)} = p_i^n \cdot p_i^m$$

(36)

Since $p_i^n = \frac{\exp(\alpha L_i^n)}{\sum_{n'} \exp(\alpha L_i^{n'})}$,

$$EMU_i = \frac{1}{\alpha} \ln \left( \frac{\exp(\alpha L_i^n)}{p_i^n} \right) = L_i^n - \frac{1}{\alpha} \ln p_i^n$$

(37)

Multiplying both sides of (37) by $p_i^n$ and summing over $n$, we have

$$p_i^n EMU_i = p_i^n L_i^n - \frac{1}{\alpha} p_i^n \ln p_i^n$$

(38)

$$EMU_i = \sum_n p_i^n L_i^n - \frac{1}{\alpha} \sum_n p_i^n \ln p_i^n$$

(39)

Also, we know that $L_i^n = \frac{1}{\lambda} \ln \sum_{m' \in n} \exp(\lambda V_{im}^m)$.
Therefore, since
\[ p_i^m = \frac{\exp(\lambda^m V_i^m)}{\sum_{m' \in n} \exp(\lambda^m V_i^m)} \]
we readily obtain
\[ L_i^m = V_i^m - \frac{1}{\lambda^m} \ln p_i^m = \sum_{m \in n} \frac{1}{\lambda^m} \ln p_i^m - \frac{1}{\lambda^m} \sum_{m' \in n} p_i^{m'} \ln p_i^m \]  \hspace{1cm} (40)

Finally, substituting (40) into (39) we get
\[ EMU_i = \sum_n \frac{1}{\alpha} \sum_{m} p_i^n \ln p_i^n - \sum_n \frac{1}{\lambda^n} \sum_{m} p_i^m \ln p_i^m - \frac{1}{\lambda^n} \sum_{m'} \sum_n p_i^{m'} \ln p_i^{m'} \]  \hspace{1cm} (41)

From this expression we can formulate the following optimization problem that each individual solves:
\[ \max_{\{p_i^n, p_i^m\}} \sum_n p_i^n \sum_{m} p_i^m V_i^m - \sum_n \frac{1}{\lambda^n} \sum_{m} p_i^m \ln p_i^m - \frac{1}{\alpha} \sum_n p_i^n \ln p_i^n \]  
\[ \text{s.t. : } \sum_{m' \in n} p_i^{m'} = 1 \quad \forall i, n \quad (\Phi_i^n) \]  
\[ \sum_n p_i^n = 1 \quad \forall n \quad (\gamma_i) \]  \hspace{1cm} (42)

The optimality conditions for the problem are
\[ p_i^m = \frac{\exp(\lambda^m V_i^m)}{\sum_{m' \in n} \exp(\lambda^m V_i^m)} \quad , \quad \Phi_i^n = p_i^n \left[ \frac{1}{\lambda^n} \ln \sum_{m \in n} \exp(\lambda^m V_i^m) + 1 \right] \]  \hspace{1cm} (43)
\[ p_i^n = \frac{\exp(\alpha L_i^n)}{\sum_n \exp(\alpha L_i^n)} \quad , \quad \gamma_i = \frac{1}{\alpha} \ln \sum_n \exp(\alpha L_i^n) \]  \hspace{1cm} (44)

Assuming a linear additive utility function \( V_i^m = \sum_k \beta_k^m x_{ki} \) for each individual \( i \) and that multiple individuals make their optimal decision simultaneously (based on mixed strategies), we formulate the following optimization problem:
\[ \min_{\{p_i^n, p_i^m\}} \sum_i \sum_n \frac{1}{\lambda^n} \sum_{m \in n} p_i^m \ln p_i^m + \frac{1}{\alpha} \sum_i \sum_n p_i^n \ln p_i^n \]  
\[ \text{s.t. : } \sum_{m \in n} p_i^m = 1 \quad \forall i, n \quad (\Phi_i^n) \]  \hspace{1cm} (45)
\[ \sum_n p_i^n = 1 \quad \forall i \quad (\gamma_i) \]  \hspace{1cm} (46)
\[ \sum_i p_i^n p_i^m x_{ki} = X_k^m \quad \forall k, m \in n \quad (\beta_k^m) \]  \hspace{1cm} (47)

The uniqueness of solution (45) - (48) imposes that \( \lambda^n > 0 \) and \( \lambda^n > \alpha \) (\( \forall n \)), or in other words, \( 0 < \phi^n = \frac{\alpha}{\lambda^n} < 1 \). This condition is the equivalent of the condition in
hierarchical logit models based on random utility in which parameter \( \phi^n = \frac{\alpha}{\lambda^n} < 1 \) so that the variances will not be negative. Recall that for reasons of parameter identifiability, in classic econometric estimation one of the parameters (\( \alpha \) or \( \lambda^n \)) must be fixed, leading to the use of upper and lower normalizations. If we assume an upper normalization of (\( \alpha = 1 \)), then \( \phi^n = \frac{1}{\lambda^n} < 1 \).

**Appendix 3: Extension of Model with Endogenous Explanatory Variables to the Hierarchical Logit Case**

The equivalent optimization problem whose optimality conditions give the hierarchical logit model is the following (Donoso et al. 2011):

\[
\begin{align*}
\min_{\{p^m_i\}} & \quad \sum_{i} \left( \sum_{n} p^m_i \frac{1}{\lambda^n} \sum_{m} p^m_i \ln p^m_i \right) + \frac{1}{\alpha} \sum_{i} \left( \sum_{n} p^m_i \ln p^m_i \right) - \sum_{i} \left( \sum_{n} p^m_i \sum_{m} p^m_i V^m_i \right) \\
\text{s.t.} : & \quad \sum_{m \in n} p^m_i = 1 \quad \forall i, n \quad (\Phi_i) \\
& \quad \sum_{n} p^m_i = 1 \quad \forall n \quad (\gamma_i)
\end{align*}
\]

where \( n \) represents the nest and \( V^m_i = \sum_{k} \beta^m_k x^m_{ki} \).

Assuming that \( x^m_{ki} = x^m_{ki} (t^m) \), and in this case \( t^m = \sum_{i} \left( p^m_i \cdot p^n_i \right) \), \( \forall m \in n \), then

\[
\begin{align*}
\frac{\partial x^m_{ki}}{\partial p^n_i} & = \frac{\partial x^m_{ki}}{\partial t^m} \frac{\partial t^m}{\partial p^n_i} = \frac{\partial x^m_{ki}}{\partial t^m} p^n_i \\
\frac{\partial x^m_{ki}}{\partial p^n_i} & = \frac{\partial x^m_{ki}}{\partial t^m} \frac{\partial t^m}{\partial p^n_i} = \frac{\partial x^m_{ki}}{\partial t^m} p^n_i \\
\end{align*}
\]

(50)

Dividing (50) by (51), we have

\[
\begin{align*}
\frac{\partial x^m_{ki}}{\partial p^n_i} / \frac{\partial p^n_i}{\partial p^n_i} & = p^n_i / p^n_i \rightarrow p^n_i \frac{\partial x^m_{ki}}{\partial p^n_i} = p^n_i \frac{\partial x^m_{ki}}{\partial p^n_i} \\
\end{align*}
\]

(52)

The optimality conditions for (49) are

\[
\begin{align*}
\frac{dL}{dp^n_i} \bigg|_{\forall m \in n} & = \frac{1}{\lambda^n} \left( \ln p^n_i + 1 \right) + \phi^n_i + \sum_{k, m \in n} \beta^m_k \left( x^m_{ki} + p^n_i \frac{dx^m_{ki}}{dp^n_i} \right) = 0 \\
\frac{dL}{dp^n_i} & = \frac{1}{\lambda^n} \sum_{m \in n} p^n_i \ln p^n_i + \frac{1}{\alpha} \left( \ln p^n_i + 1 \right) + \gamma_i + \sum_{m \in n} \sum_{k, m \in n} \beta^m_k \left( x^m_{ki} + p^n_i \frac{dx^m_{ki}}{dp^n_i} \right) = 0 \\
\frac{dL}{dp^n_i} & = \frac{1}{\lambda^n} \sum_{m \in n} p^n_i \ln p^n_i + \frac{1}{\alpha} \left( \ln p^n_i + 1 \right) + \gamma_i + \sum_{k, m \in n} \beta^m_k \left( x^m_{ki} + p^n_i \frac{dx^m_{ki}}{dp^n_i} \right) = 0
\end{align*}
\]

(53) (54) (55)
From (53) we obtain

$$\ln p_i^m = -\lambda^* \Phi_i^m - \lambda^n \sum_{k, m \in n} \beta_k^m \left( x_{ki}^m + p_i^m \frac{d x_{ki}^m}{d p_i^m} \right) - 1$$  \hspace{1cm} (56)$$

$$p_i^m = \exp \left( -\lambda^* \Phi_i^m - \lambda^n \sum_{k, m \in n} \beta_k^m \left( x_{ki}^m + p_i^m \frac{d x_{ki}^m}{d p_i^m} \right) - 1 \right)$$  \hspace{1cm} (57)$$

Summing over \( m \) in (57), we get

$$\sum_{m \in n} p_i^m = 1 \rightarrow \exp(-\lambda^* \Phi_i^m - 1) \sum_{m \in n} \exp \left( -\lambda^n \sum_{k, m \in n} \beta_k^m \left( x_{ki}^m + p_i^m \frac{d x_{ki}^m}{d p_i^m} \right) \right) = 1$$  \hspace{1cm} (58)$$

Dividing (57) by (58), we have

$$p_i^m = \frac{\exp \left( -\lambda^n \sum_{k, m \in n} \beta_k^m \left( x_{ki}^m + p_i^m \frac{d x_{ki}^m}{d p_i^m} \right) \right)}{\sum_{m \in n} \exp \left( -\lambda^n \sum_{k, m \in n} \beta_k^m \left( x_{ki}^m + p_i^m \frac{d x_{ki}^m}{d p_i^m} \right) \right)}, \forall m \in n$$  \hspace{1cm} (59)$$

This expression represents the probability that individual \( i \) chooses alternative \( m \) given that he or she is in nest \( n \). In other words, it is the conditional probability of choosing the nest.

Taking the natural logarithm of (59),

$$\ln p_i^m = -\lambda^n \sum_{k, m \in n} \beta_k^m \left( x_{ki}^m + p_i^m \frac{d x_{ki}^m}{d p_i^m} \right) - \ln \sum_{m \in n} \exp \left( -\lambda^n \sum_{k, m \in n} \beta_k^m \left( x_{ki}^m + p_i^m \frac{d x_{ki}^m}{d p_i^m} \right) \right)$$  \hspace{1cm} (60)$$

Multiplying both sides by \( p_i^m \) and summing over \( m \),

$$\sum_{m \in n} p_i^m \ln p_i^m = -\lambda^n \sum_{m \in n} p_i^m \sum_{k, m \in n} \beta_k^m \left( x_{ki}^m + p_i^m \frac{d x_{ki}^m}{d p_i^m} \right)$$

$$- \sum_{m \in n} p_i^m \ln \sum_{m \in n} \exp \left( -\lambda^n \sum_{k, m \in n} \beta_k^m \left( x_{ki}^m + p_i^m \frac{d x_{ki}^m}{d p_i^m} \right) \right)$$  \hspace{1cm} (61)$$

$$\sum_{m \in n} p_i^m \ln p_i^m = -\lambda^n \sum_{k, m \in n} \beta_k^m \left( x_{ki}^m + p_i^m \frac{d x_{ki}^m}{d p_i^m} \right) - \ln \sum_{m \in n} \exp \left( -\lambda^n \sum_{k, m \in n} \beta_k^m \left( x_{ki}^m + p_i^m \frac{d x_{ki}^m}{d p_i^m} \right) \right)$$  \hspace{1cm} (62)$$
Dividing both sides by $\lambda^n$,

$$\frac{1}{\lambda^n} \sum_{m \in n} p_i^m \ln p_i^m = - \sum_{k, m \in n} \beta_k^m \left( x_{ki} + p_i^m \frac{dx_{ki}^m}{dp_i^m} \right) - \frac{1}{\lambda^n} \ln \sum_{m \in n} \exp \left( - \lambda^n \sum_{k, m \in n} \beta_k^m \left( x_{ki} + p_i^m \frac{dx_{ki}^m}{dp_i^m} \right) \right)$$

(63)

Substituting the expression $\frac{1}{\lambda^n} \sum_{m \in n} p_i^m \ln p_i^m$ from (63) into (55), and recalling equality (52), we obtain

$$\frac{1}{\alpha} \left( \ln p_i^n + 1 \right) - \frac{1}{\lambda^n} \ln \sum_{m \in n} \exp \left( - \lambda^n \sum_{k, m \in n} \beta_k^m \left( x_{ki} + p_i^m \frac{dx_{ki}^m}{dp_i^m} \right) \right) + \gamma_i = 0$$

(64)

This can be rewritten as

$$\ln p_i^n = \frac{\alpha}{\lambda^n} \ln \sum_{m \in n} \exp \left( - \lambda^n \sum_{k, m \in n} \beta_k^m \left( x_{ki} + p_i^m \frac{dx_{ki}^m}{dp_i^m} \right) \right) - \alpha \gamma_i - 1$$

(65)

$$p_i^n = \exp(-\alpha \gamma_i - 1) \exp \left( \frac{\alpha}{\lambda^n} \ln \sum_{m \in n} \exp \left( - \lambda^n \sum_{k, m \in n} \beta_k^m \left( x_{ki} + p_i^m \frac{dx_{ki}^m}{dp_i^m} \right) \right) \right)$$

(66)

Summing over $n$, we have

$$\sum_n p_i^n = 1 \rightarrow \exp(-\alpha \gamma_i - 1) \sum_n \exp \left( \frac{\alpha}{\lambda^n} \ln \sum_{m \in n} \exp \left( - \lambda^n \sum_{k, m \in n} \beta_k^m \left( x_{ki} + p_i^m \frac{dx_{ki}^m}{dp_i^m} \right) \right) \right) = 1$$

(67)

Dividing (66) by (67),

$$p_i^n = \frac{\exp \left( \frac{\alpha}{\lambda^n} \ln \sum_{m \in n} \exp \left( - \lambda^n \sum_{k, m \in n} \beta_k^m \left( x_{ki} + p_i^m \frac{dx_{ki}^m}{dp_i^m} \right) \right) \right)}{\sum_n \exp \left( \frac{\alpha}{\lambda^n} \ln \sum_{m \in n} \exp \left( - \lambda^n \sum_{k, m \in n} \beta_k^m \left( x_{ki} + p_i^m \frac{dx_{ki}^m}{dp_i^m} \right) \right) \right)}$$

(68)

This equation gives the probability that individual $i$ chooses nest $n$.

Finally, from (59) and (68) we obtain the hierarchical logit model for estimating the probability $p_i^{n,m} = p_i^n \cdot p_i^m$ that individual $i$ chooses alternative $m$ in nest $n$. 
If we now define $\eta^m = \sum_k \beta_k^m \frac{dx_{kli}^m}{dp_i^m}$, we finally arrive at

$$
\begin{align*}
\hat{p}_i^{n,m} &= \frac{\exp\left(\frac{\alpha}{\lambda^n} \ln \sum_n \exp\left(-\lambda^n \sum_{k, \text{men}} \beta_k^m \left(x_{kli}^m + p_i^m \frac{dx_{kli}^m}{dp_i^m}\right)\right)\right)}{
\sum_n \exp\left(\frac{\alpha}{\lambda^n} \ln \sum_n \exp\left(-\lambda^n \sum_{k, \text{men}} \beta_k^m \left(x_{kli}^m + p_i^m \frac{dx_{kli}^m}{dp_i^m}\right)\right)\right)} \\
&= \frac{\exp\left(-\lambda^n \sum_{k, \text{men}} \beta_k^m \left(x_{kli}^m + p_i^m \frac{dx_{kli}^m}{dp_i^m}\right)\right)}{
\sum_{k, \text{men}} \exp\left(-\lambda^n \sum_{k, \text{men}} \beta_k^m \left(x_{kli}^m + p_i^m \frac{dx_{kli}^m}{dp_i^m}\right)\right)}
\end{align*}
$$

This expression is also a fixed-point function that can be solved by defining an instrument for $p_i^m$.

Appendix 4: Simulation Results Using $\gamma=1$ y $\gamma=2$

![Fig. 8 Distribution of the $\beta_{\text{time}}$ parameter estimator for the MNL and MNLE models ($\gamma=1$)](image-url)
References

Ai C, Gan L (2010) An alternative root-n consistent estimator for panel data binary choice models. J Econ 157:93–100
Amador FJ, González RM, de Ortúzar JD (2008) On confounding preference heterogeneity and income effect in discrete choice models. Netw Spat Econ 8:97–108
Anas A (1983) Discrete choice theory, information theory and the multinomial logit and gravity models. Transp Res 17B:13–23
Banerjee A (1992) A simple model of herd behavior. Q J Econ 107:797–817
Batarce M, Ivaldi M (2014) Urban travel demand model with endogenous congestion. Transp Res 59A:331–345
Bhat CR, Guo J (2004) A mixed spatially correlated logit model: formulation and application to residential choice modeling. Transp Res 38B:147–168

Fig. 9 Distribution of the $\beta_{time}$ parameter estimator for the MNL and MNLE models ($\gamma=2$)

Table 6 $\beta_{time}$ parameter estimators

| MODEL          | $\hat{\beta}_{time}$ | Standard Error | $t$ Test ($\beta_{time}=-0.25$) |
|----------------|-----------------------|----------------|---------------------------------|
| MNL ($\gamma=1$) | $-0.108$             | $0.045$        | $3.167$                         |
| MNLE ($\gamma=1$) | $-0.294$             | $0.075$        | $-0.578$                        |
| MNL ($\gamma=2$) | $-0.119$             | $0.057$        | $2.318$                         |
| MNLE ($\gamma=2$) | $-0.305$             | $0.079$        | $-0.695$                        |
Lewbel A (2007) Endogenous selection or treatment model estimation. J Econ 141:777–806
Louviere J, Train K, Ben-Akiva M, Bhat C, Brownstone D, Cameron T, Carson C, Deshazo J, Fiebig D, Greene W, Hensher D, Waldman D (2005) Recent progress on endogeneity in choice modeling. Mark Lett 16:255–265
Magnac T, Maurin E (2007) Identification and information in monotone binary models. J Econ 139:76–104
Magnac T, Maurin E (2008) Partial identification in monotone binary models: discrete regressors and interval data. Rev Econ Stud 75:835–864
Marquez-Ramos L, Martinez-Zarzoso I, Perez-Garcia E, Wilmsmeier G (2011) “Special issue on Latin-american research” maritime networks, services structure and maritime trade. Netw Spat Econ 11:555–576
Matzkin R (2004) Unobservable Instruments. working paper, Economics, Northwestern University
McFadden D (1974) Conditional logit analysis of qualitative choice behavior. In: Zarembka P (ed) Frontiers in Econometrics. Academic, New York
de Ortúzar JD (1982) Fundamentals of discrete multimodal choice modelling. Transp Rev 2:47–78
de Ortúzar JD (1983) On the equivalence of modified logit models: some comments. Transportation 11:383–385
Willemsen LG, de Ortúzar JD (2011) Modeling Transport. John Wiley & Sons, Chichester
Petrin A, Train K (2010) A control function approach to endogeneity in consumer choice models. J Mark Res 47:370–379
Raveau S, Muñoz JC, De Grange L (2011) A topological route choice model for metro. Transp Res 45A:138–147
SECTRA (2011) Actualización y Diagnóstico del Sistema de Transporte Urbano de la Ciudad de Antofagasta, Etapa I. www.sectra.cl
Soetevent A, Kooreman P (2007) A discrete-choice model with social interaction: with an application to high school teen behaviour. J Appl Econ 22:599–624
Train K (1986) Qualitative Choice Analysis: Theory Econometrics, and an Application to Automobile Demand. The MIT Press, Cambridge
Train K (2003) Discrete Choice Methods with Simulation. Cambridge University Press, Cambridge
Train K and Winston C (2004) Vehicle Choice Behavior and the Declining Market Share of U.S. Automakers. Working paper, Economics, University of California, Berkeley
Villas-Boas J, Winer R (1999) Endogeneity in brand choice models. Manag Sci 45:1324–1338
Walker J, Ben-Akiva M (2002) Generalized random utility model. Math Soc Sci 43:303–343
Walker J, Ehlers E, Banerjee I, Dugundji R (2011) Correcting for endogeneity in behavioral choice models with social influence variables. Transp Res 45A:362–374
Williams HCMWL (1977) On the formation of travel demand models and economic evaluation measures of user benefit. Environ Plan 9A:285–344
Yamins D, Rasmussen S, Fogel D (2003) Growing Urban Roads. Netw Spat Econ 3:69–85