

Article

## A Maximum Entropy Estimator for the Aggregate Hierarchical Logit Model

Pedro Donoso <sup>1</sup>, Louis de Grange <sup>2,\*</sup> and Felipe González <sup>2</sup>

<sup>1</sup> Laboratorio de Modelamiento del Transporte y Uso del Suelo (LABTUS), Departamento de Ingeniería Civil, Universidad de Chile, Santiago PO Box 10-D, Chile; E-Mail: pdonoso@labtus.cl

<sup>2</sup> Escuela de Ingeniería Civil Industrial, Universidad Diego Portales, Santiago, 8370179, Chile; E-Mail: felipe.gonzalezr@udp.cl

\* Author to whom correspondence should be addressed; E-Mail: louis.degrange@udp.cl; Tel.: +56-2-6762400; Fax: +56-2-6768130.

Received: 20 June 2011; in revised form: 15 July 2011 / Accepted: 20 July 2011 /

Published: 2 August 2011

---

**Abstract:** A new approach for estimating the aggregate hierarchical logit model is presented. Though usually derived from random utility theory assuming correlated stochastic errors, the model can also be derived as a solution to a maximum entropy problem. Under the latter approach, the Lagrange multipliers of the optimization problem can be understood as parameter estimators of the model. Based on theoretical analysis and Monte Carlo simulations of a transportation demand model, it is demonstrated that the maximum entropy estimators have statistical properties that are superior to classical maximum likelihood estimators, particularly for small or medium-size samples. The simulations also generated reduced bias in the estimates of the subjective value of time and consumer surplus.

**Keywords:** hierarchical logit model; lagrange multipliers; maximum entropy; maximum likelihood

**PACS Codes:** 02.50.Cw

**MSC Codes:** 62P20; 97M40

---

## 1. Introduction

In urban transportation planning, travel demand is frequently represented by multinomial or hierarchical logit discrete choice models, particularly for the selection of destinations, routes, and transportation modes. These models are also used in land use planning for modeling the real estate supply and activity location. Traditionally, both classes of choice models are deduced from the paradigm of the rational user who uses transportation services or real state units so as to maximize his or her utility as given by an assigned probability distribution. Estimation of the model parameters relies on classical statistical criteria, most commonly the method of maximum likelihood.

These models can also be deduced as the solution to certain constrained entropy maximization problems. The Lagrange multipliers of the constraints constitute alternative estimators of the population parameters to those generated by the known maximum likelihood approach, and in this work will be called maximum entropy estimators.

In [1] it was shown that maximum likelihood estimators are identical to maximum entropy estimators in the case of multinomial logit models. Here, this equivalence will be investigated for the hierarchical logit model [2].

The maximum entropy approach has been used primarily to formulate aggregate trip demand models, especially users' single decisions. Particularly prominent are the spatial trip distribution formulations such as the doubly constrained gravitational model proposed by [3] and its later modifications and extensions (see [4–10]). Other important applications of maximum entropy are the combined models which integrate the different transportation decisions including trip generation, destination choice, mode choice and route choice. These models first appeared in the early 1980s and were followed over the next two decades by further important developments (e.g., [1,11–21]). In every maximum entropy application reported in the specialized literature, the demand model is the solution of an entropy maximization problem (or an equivalent formulation) with exogenous parameters in its objective function and a set of linear constraints. Applying the optimality conditions to these problems generates combined multinomial or hierarchical logit demand models, depending on the form of the objective function and its constraints. A microeconomic interpretation of the maximum entropy estimator of multinomial logit models and its equivalence to the maximum likelihood estimator is presented in [22].

The endogenous and exogenous parameters of these models are estimated by applying certain statistical techniques used for calibrating econometric models, most notably the maximum likelihood method. This is a sensible strategy for a multinomial logit model given the equivalence of the two estimators noted above, but may not be the best option if the model is hierarchical logit.

To analyze the equivalence of the maximum likelihood and maximum entropy estimators in the context of a hierarchical logit model, we first carry out a theoretical analysis and then calculate both estimators based on Monte Carlo simulations. The estimators are then evaluated in terms of their bias with respect to known population parameters, their efficiency and consistency properties as well as certain general goodness-of-fit criteria such as mean square error. From the results obtained we determine the differences between the two approaches for various data scenarios.

Our conclusion is that the maximum entropy estimators provide a viable alternative for estimating hierarchical logit models; indeed, in the light of the simulations they appear to be superior to maximum likelihood estimates, especially with small sample sizes.

## 2. Formulation and Estimation of Hierarchical Logit Model

### 2.1. Formulation of Hierarchical Logit (HL) Model

In the HL model (see [2]) the utility of alternative  $a$  in group  $g$  for individual  $q$  of type  $i$  is given by:

$$U_{agqi} = V_{agi} + V_{gi} + \varepsilon_{agqi} + \varepsilon_{gqi} \quad \text{for all } a, g, q = 1, \dots, N_i^0, i \tag{1}$$

where  $N_i^0$  is the number of individuals of type  $i$ . If  $N_i^0 = 1$  for all  $i$ , the model is disaggregate by individuals; if not, the sub index  $q$  is omitted for simplicity. The terms  $V_{agi}$  and  $V_{gi}$  are the deterministic components of the utility perceived by a type  $i$  individual from alternative  $a$  and group  $g$ , respectively. We assume that both terms are linear functions of attributes ( $x$  and  $w$ ) and parameters ( $\beta$  and  $\gamma$ ) that are either generic or specific to an alternative or group and type of individual. Thus:

$$V_{agi} = \sum_k \beta_k x_{agik} \quad \text{for all } a, g, i \tag{2}$$

$$V_{gi} = \sum_m \gamma_m w_{gim} \quad \text{for all } g, i \tag{3}$$

The  $\varepsilon_{agqi}$  are i.i.d. random variables with a Gumbel distribution  $(0, \mu_g)$  where  $\mu_g > 0$ . The  $\varepsilon_{gqi}$  variables are such that  $\varepsilon'_{gqi} = \varepsilon_{gqi}^* + \varepsilon_{gqi} \quad \forall g, q, i$ , where the  $\varepsilon'_{gqi}$  are i.i.d. r.v.'s with a Gumbel distribution  $(0, \lambda)$ ,  $\lambda > 0$  and the  $\varepsilon_{gqi}^*$  are i.i.d. r.v.'s with a Gumbel distribution  $(0, \mu_g)$ . The HL model of the probability that a type  $i$  individual chooses alternative  $a$  in group  $g$  ( $p_{ag/i}$ ) is:

$$p_{ag/i} = p_{a/gi} p_{g/i} \quad \text{for all } a, g, i \tag{4}$$

where  $p_{a/gi}$  is the probability that a type  $i$  individual chooses alternative  $a$  given that he or she chose group  $g$ , and  $p_{g/i}$  is the probability that a type  $i$  individual chooses group  $g$ . If we denote

$V_{gi}^* = \frac{1}{\mu_g} \ln \left( \sum_a \exp(\mu_g V_{agi}) \right)$  for all  $g, i$ , the latter two probabilities are then given by:

$$p_{a/gi} = \frac{\exp(\mu_g V_{agi})}{\sum_{a'} \exp(\mu_g V_{a'gi})} \quad \text{for all } a, g, i \tag{5}$$

$$p_{g/i} = \frac{\exp(V_{gi}^* + V_{gi})}{\sum_{g'} \exp(V_{g'i}^* + V_{g'i})} \quad \text{for all } g, i \tag{6}$$

For (6) we assume that  $\lambda = 1$  so that the model parameters can be identified. With this assumption,  $\mu_g \geq 1$  for all  $g$ ; if  $\mu_g = 1$  for all  $g$ , the HL formulation reduces to the multinomial logit (MNL) model.

2.2. Estimation of HL Model by Maximum Likelihood (ML)

The maximum likelihood estimators for the parameters of the HL model defined by (4)–(6) are obtained as the solution of the following optimization problem (see microeconomic interpretation in Appendix A):

$$\max_{\{\mu, \beta, \gamma\}} \ln L = \sum_{i,g} N_{gi}^0 \ln p_{g/i}(\mu, \beta, \gamma) + \sum_{i,g,a} N_{agi}^0 \ln p_{a/gi}(\mu_g, \beta) \tag{7}$$

where  $N_{agi}^0$  is the observed number of type  $i$  individuals who chose alternative  $a$  in group  $g$ ,  $N_i^0 = \sum_{g,a} N_{agi}^0$  for all  $i$  and  $N_{gi}^0 = \sum_a N_{agi}^0$  for all  $g, i$ :

The first-order conditions of (7) are:

$$\frac{\partial \ln L}{\partial \beta_k} = \sum_{i,g} \frac{N_{gi}^0}{p_{g/i}} \frac{\partial p_{g/i}}{\partial \beta_k} + \sum_{i,g,a} \frac{N_{agi}^0}{p_{a/gi}} \frac{\partial p_{a/gi}}{\partial \beta_k} = 0 \quad \text{for all } k \tag{8}$$

$$\frac{\partial \ln L}{\partial \mu_g} = \sum_{i,g} \frac{N_{gi}^0}{p_{g/i}} \frac{\partial p_{g/i}}{\partial \mu_g} + \sum_{i,a} \frac{N_{agi}^0}{p_{a/gi}} \frac{\partial p_{a/gi}}{\partial \mu_g} = 0 \quad \text{for all } g \tag{9}$$

$$\frac{\partial \ln L}{\partial \gamma_m} = \sum_{i,g} \frac{N_{gi}^0}{p_{g/i}} \frac{\partial p_{g/i}}{\partial \gamma_m} = 0 \quad \text{for all } m \tag{10}$$

It is readily demonstrated that:

$$\frac{\partial p_{g/i}}{\partial \beta_k} = p_{g/i} \left( \sum_a p_{a/gi} x_{agik} - \sum_{g'} p_{g'/i} \sum_a p_{a/g'i} x_{ag'ik} \right) \quad \text{for all } g, i, k \tag{11}$$

$$\frac{\partial p_{a/gi}}{\partial \beta_k} = p_{a/gi} \mu_g \left( x_{agik} - \sum_{a'} p_{a'/gi} x_{a'gik} \right) \quad \text{for all } a, g, i, k \tag{12}$$

Substituting (11) and (12) in (8) we obtain the following equation associated with  $\beta_k$ :

$$\sum_{i,g,a} \left( N_i^0 p_{g/i} p_{a/gi} + (\mu_g - 1) (N_{gi}^0 p_{a/gi} - N_{agi}^0) \right) x_{agik} = \sum_{i,g,a} N_{agi}^0 x_{agik} \quad \text{for all } k \tag{13}$$

As we know, if  $\mu_g = 1$  for all  $g$ , we just have the MNL model and (13) reduces to the well-known expression which states that the sum of the values of attribute  $x_k$  for the alternatives chosen by the various individuals is equal to the sum predicted by the estimated choice probabilities. Then, the maximum likelihood estimators of a MNL model reproduce the average values of its explanatory variables (travel time, cost, etc.) and, if specific constants for each alternative are specified, the market (*i.e.*, observed) modal shares.

However, it is evident from (13) that this condition does not hold for the HL model since this would require that the following additional condition be satisfied:

$$\sum_{i,g,a} (\mu_g - 1) (N_{gi}^0 p_{a/gi} - N_{agi}^0) x_{agik} = 0 \quad \text{for all } k \tag{14}$$

Also, if we define a specific constant  $\beta_{0ag}$  for each alternative  $a$  in each group  $g$ , the Equation (13) associated with this parameter is:

$$\sum_i N_i^0 p_{g/i} p_{a/gi} + (\mu_g - 1) \sum_i (N_{gi}^0 p_{a/gi} - N_{agi}^0) = \sum_i N_{agi}^0 \quad \text{for all } a, g \quad (15)$$

Thus, the HL model does not reproduce the observed modal shares for each alternative in each group as this would require that  $\mu_g = 1 \forall g$  (the MNL model) or that  $\sum_i N_{gi}^0 p_{a/gi} = \sum_i N_{agi}^0$  for all  $a, g$ , that is, that the observed modal shares be reproduced conditional upon the choice from each group. Furthermore, if we sum over  $a'$  on both sides of (15) we obtain:

$$\sum_i N_i^0 p_{g/i} = \sum_i N_{gi}^0 \quad \text{for all } g \quad (16)$$

Thus, the observed modal shares of each group are reproduced when specific constants for each alternative are specified. By similar reasoning we obtain the same conclusion if we define specific constants for each group.

### 2.3. Estimation of HL Model by Maximum Entropy (ME)

Consider the following optimization problem, denoted Entropy Maximization with Hierarchical Probabilities (EMHP):

$$\max_{\{p_{a/gi}, p_{g/i}\}} - \sum_{a,g,i} N_i^0 p_{ag/i} \ln p_{ag/i} = - \sum_{a,g,i} N_i^0 p_{g/i} p_{a/gi} \ln(p_{g/i} p_{a/gi}) \quad (17)$$

subject to:

$$- \sum_{i,a} N_i^0 p_{g/i} p_{a/gi} \ln p_{a/gi} = - \sum_{i,a} N_{agi}^0 \ln \left( \frac{N_{agi}^0}{N_{gi}^0} \right) \quad \text{for all } g \left( \frac{1}{\mu_g} \right) \quad (18)$$

$$\sum_{i,g,a} N_i^0 p_{g/i} p_{a/gi} x_{agik} = \sum_{i,g,a} N_{agi}^0 x_{agik} \quad \text{for all } k \quad (\beta_k) \quad (19)$$

$$\sum_{i,g} N_i^0 p_{g/i} w_{gim} = \sum_{i,g} N_{gi}^0 w_{gim} \quad \text{for all } m \quad (\gamma_m) \quad (20)$$

$$\sum_a p_{a/gi} = 1 \quad \text{for all } g, i \quad (\alpha_{gi}) \quad (21)$$

$$\sum_g p_{g/i} = 1 \quad \text{for all } i \quad (\rho_i) \quad (22)$$

In this formulation,  $\left(\frac{1}{\mu_g}\right)$ ,  $(\beta_k)$ ,  $(\gamma_m)$ ,  $(\alpha_{gi})$  and  $(\rho_i)$  are the Lagrange multipliers of constraints

(18) to (22), respectively. The structure of the EMHP problem is similar to the entropy maximization problem generated by a multinomial logit model (see [1]) with the exception of (18), which restricts entropy in each group. Total entropy can be decomposed as follows:

$$- \sum_{a,g,i} N_i^0 p_{ag/i} \ln p_{ag/i} = - \sum_{i,g} N_i^0 p_{g/i} \ln p_{g/i} - \sum_{a,g,i} N_i^0 p_{g/i} p_{a/gi} \ln p_{a/gi} \quad (23)$$

The constraint (18) imposes that the second term on the right side of (23) be constant, so the objective function (17) can be reduced to  $-\sum_{i,g} N_i^0 p_{g/i} \ln p_{g/i}$ . The Lagrangian of the reduced EMHP

problem is:

$$\begin{aligned}
 L = & -\sum_{i,g} N_i^0 p_{g/i} \ln p_{g/i} + \sum_g \frac{1}{\mu_g} \left( -\sum_{i,a} N_i^0 p_{g/i} p_{a/gi} \ln p_{a/gi} + \sum_{i,a} N_{agi}^0 \ln \left( \frac{N_{agi}^0}{N_{gi}^0} \right) \right) + \\
 & + \sum_k \beta_k \left( \sum_{i,g,a} N_i^0 p_{g/i} p_{a/gi} x_{agik} - \sum_{i,g,a} N_{agi}^0 x_{agik} \right) + \\
 & + \sum_m \gamma_m \left( \sum_{i,g} N_i^0 p_{g/i} w_{gim} - \sum_{i,g} N_{gi}^0 w_{gim} \right) + \\
 & + \sum_{i,g} \alpha_{gi} \left( \sum_a p_{a/gi} - 1 \right) + \sum_i \rho_i \left( \sum_g p_{g/i} - 1 \right)
 \end{aligned}
 \tag{24}$$

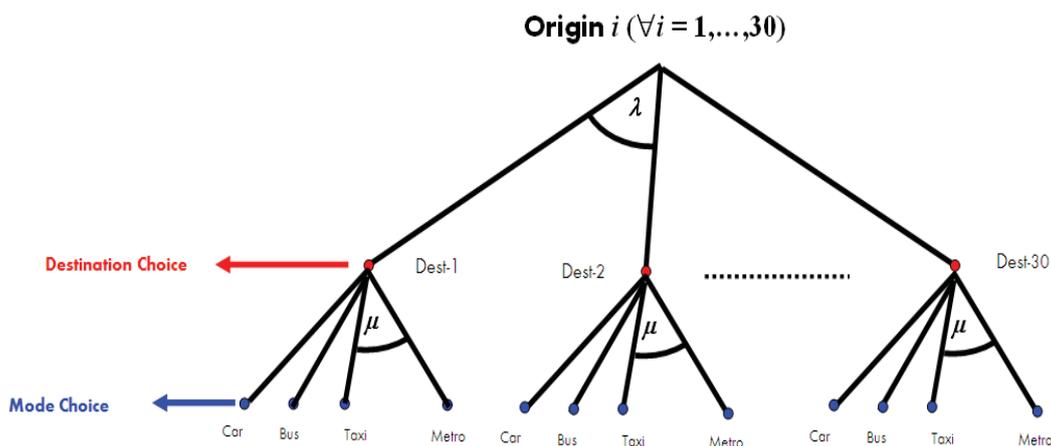
The optimality conditions, applied to this function, lead to the solutions  $p_{a/gi}$  and  $p_{g/i}$  which satisfy (5), (6), (2) and (3) and thus constitute an HL model.

The ME estimators of the model are the Lagrange multipliers  $\frac{1}{\mu_g}$ ,  $\beta_k$  and  $\gamma_m$ , which are obtained by substituting (5) and (6) into constraints (18), (19) and (20) and solving the equations numerically using, for instance, the Newton’s method. Since, unlike the ML estimators, the ME estimators satisfy constraint (19), the average values of the explanatory variables can be reproduced, as can the market modal shares if constraints are specified for each alternative. We may thus conclude that the ML estimators are different than the ME ones. This result will be analyzed empirically in the next section.

### 3. Simulation Analysis of ML and ME Estimators

To compare the two parameter estimation approaches (ML and ME) we conducted a series of Monte Carlo simulations of a hierarchical model of combined destination choice and modal share using various sample sizes and values for the parameter  $\mu$ . Although this HL model is particular, it illustrates clearly the differences between both approaches. The simulated tree structure is depicted in Figure 1.

Figure 1. Simulated hierarchical tree structure.



We defined 30 origin and destination zones and four transportation modes (private car, bus, taxi and metro) available for trips between any origin and destination pair in either direction. For simplicity, we assumed that the parameter  $\mu$  was the same for all destinations. The parameter  $\lambda$  was set at unity ( $\lambda = 1$ ). No specific attributes were assumed for the different groups.

The utility functions for the four modes are given below. In each case, the explanatory variables are travel time and travel cost and we include mode constants:

$$V_{Car} = \beta_{Car}^0 + \beta^{Time} T_{Car} + \beta^{Cost} C_{Car} \tag{25}$$

$$V_{Bus} = \beta_{Bus}^0 + \beta^{Time} T_{Bus} + \beta^{Cost} C_{Bus} \tag{26}$$

$$V_{Taxi} = \beta_{Taxi}^0 + \beta^{Time} T_{Taxi} + \beta^{Cost} C_{Taxi} \tag{27}$$

$$V_{Metro} = \beta_{Metro}^0 + \beta^{Time} T_{Metro} + \beta^{Cost} C_{Metro} \tag{28}$$

The explanatory variable parameters are generic, that is, the same for each mode. The values of these population parameters are set out in Table 1.

**Table 1.** Population parameter values in simulations.

PARAMETER	VALUE (*)
$\beta_{Car}^0$	0.9
$\beta_{Bus}^0$	0
$\beta_{Taxi}^0$	0.5
$\beta_{Metro}^0$	0.4
$\beta^{Time}$	-0.25
$\beta^{Cost}$	-0.006
SVT (**)	41.47
$\lambda$	1
$\phi = \lambda/\mu = 1/\mu$	0.5 (***)

(\*) Values are defined by the authors, and are similar to the estimates reported in [21];  
 (\*\*) Subjective value of time, estimated at  $(-0.25/-0.006) = 41.47$ ; (\*\*\*) Values of 0.2 and 0.9 were considered in complementary analysis, which is reported in the appendices.

The values for the explanatory variables (time and cost) were extracted from a 2001 transportation survey for Greater Santiago of Chile [23]. Their means and standard deviations for each transportation mode are given in Table 2.

A total of 1,000 Monte Carlo simulations were conducted with each of five different sample sizes containing 500, 1,000, 5,000, 10,000 and 20,000 observations, respectively. We thus obtained 1,000 ML and ME estimators of the parameters for each sample size, from which the estimates of bias, variance and mean squared error were derived.

The likelihood and entropy maximization problems were solved numerically using Newton’s method. The convergence criterion in all simulations was 0.1%, meaning that the percentage difference between the estimates of each parameter obtained from two consecutive iterations of the method did not exceed 0.001.

**Table 2.** Mean and standard deviation of explanatory variables by mode.

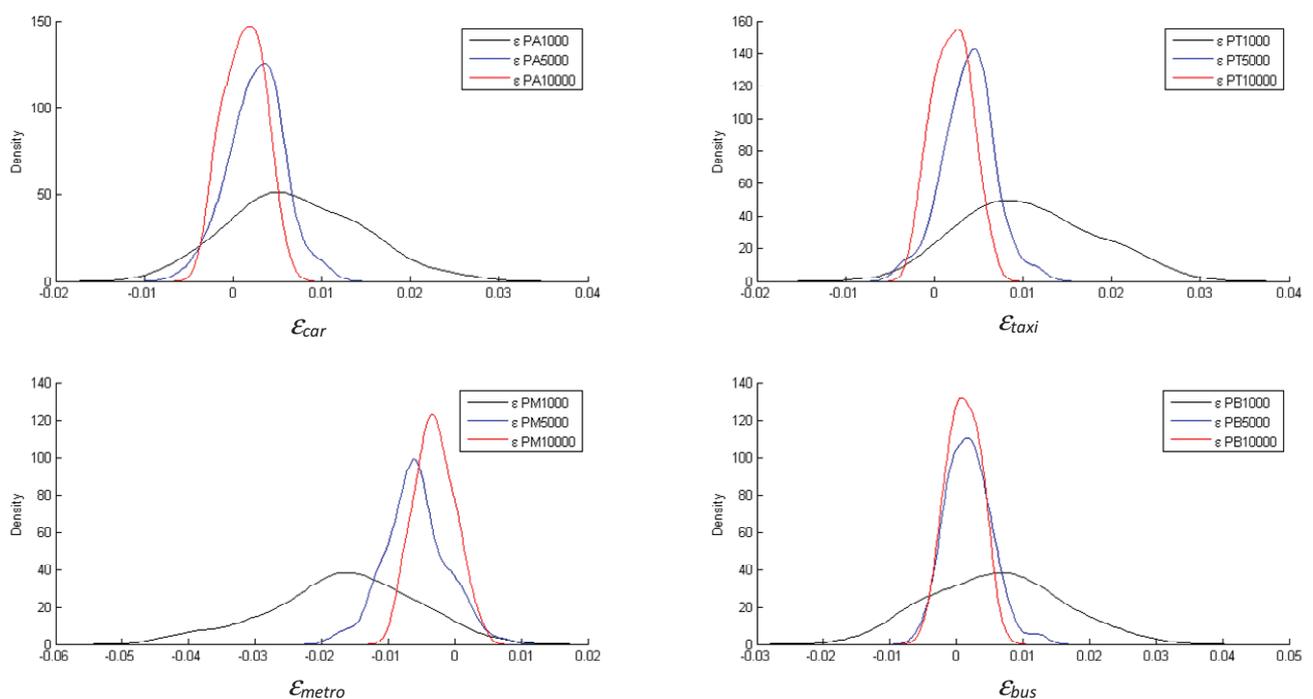
MODE	VARIABLE	MEAN (*)	STD DEV (*)
Car	Travel time	16	11
	Cost	2,031	138
Taxi	Travel time	17	11
	Cost	2,279	148
Bus	Travel time	54	12
	Cost	409	25
Metro	Travel time	45	7
	Cost	833	73

Source: [23]. (\*) Travel time in minutes and cost in Chilean pesos as of 2009.

3.1. Bias, Variance and Mean Squared Error (MSE)

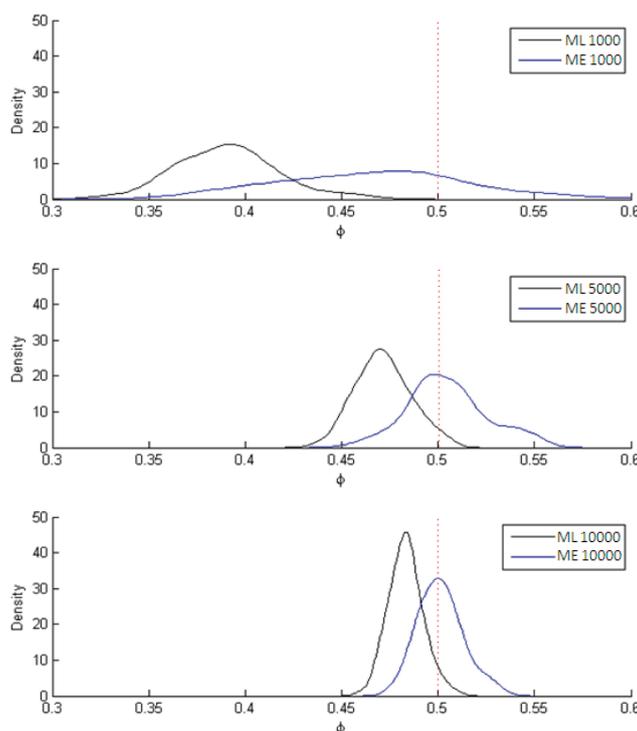
The 1,000 ML and ME estimators were used to construct histograms for simultaneously analyzing bias, variance and mean squared error. In Figure 2 we compare histograms of the observed modal split and modeled modal split for the maximum likelihood estimator, for three sample size (1,000, 5,000 and 10,000); in maximum entropy estimator, both observed and modeled modal split are identical by construction [see Equation (19)]. Each variable  $\epsilon_a$  was defined, for each simulations and mode  $a$  (car, taxi, bus and metro) by the following expression:  $\epsilon_a = \sum_{gi} (p_{agi} - N_{agi}^0) / N_{agi}^0$ . We observe that, just asymptotically, modeled modal split converges to observed modal split [see Equation (15)].

**Figure 2.** Observed vs. modeled modal split using maximum likelihood estimation, for parameter value  $\phi = 1/\mu = 0.5$ .

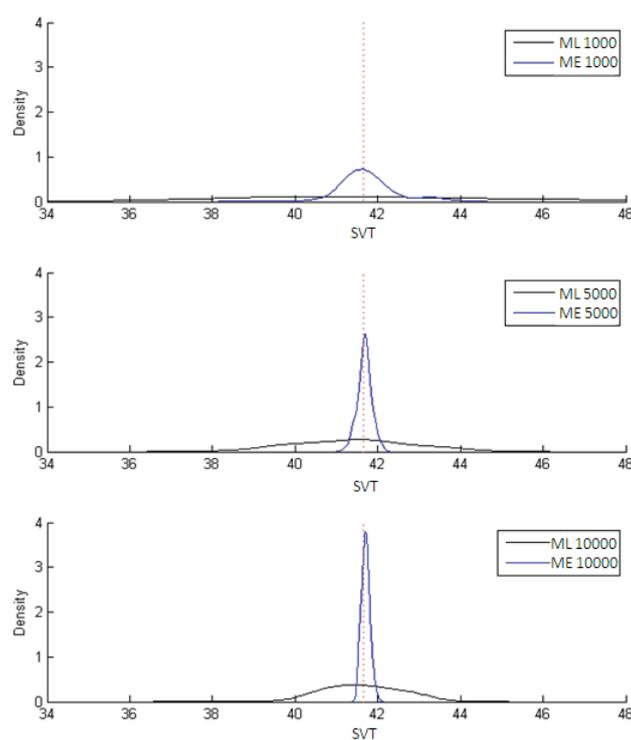


The estimates of the parameter  $\phi = 1/\mu$  are shown in Figure 3 while the estimates of the subjective value of time (SVT) are displayed in Figure 4. In both figures the dotted line marks the population parameter (used in the simulation) while the black curve traces out the ML estimates and the blue curve the ME estimates. The results in all cases are for 1,000; 5,000 and 10,000 observations, the three instances in which the differences between the two estimators most clearly stand out.

**Figure 3.** Results of ML and ME estimators for parameter value  $\phi = 1/\mu = 0.5$ .



**Figure 4.** Results of ML and ME estimators for  $SVT = \beta^{Time} / \beta^{Cost}$ .



As can be seen in Figure 3, the ML estimate of  $\mu$  is relatively biased but consistent while the ME estimate is significantly less biased but also less efficient (*i.e.*, greater variance).

Figure 4 shows that both estimators of SVT are unbiased, although the ME estimator is clearly more efficient. This is confirmed in Table 3, which summarizes the results on bias, variance and MSE for both ML and ME. Also clear from the table is that the ME estimators are less biased than the ML ones, though the variances of the latter are smaller. The MSE, however, is always lower for the ME estimators.

**Table 3.** Summary of results for ML and ME estimators.

METHOD	PARAMETER	SAMPLE SIZE	BIAS	VARIANCE	MSE (*)
ML	1/ $\mu$	500	0.16844	0.00096	0.02933
	1/ $\mu$	1,000	0.10912	0.00074	0.01265
	1/ $\mu$	5,000	0.03067	0.00016	0.00110
	1/ $\mu$	10,000	0.01655	0.00009	0.00036
	1/ $\mu$	20,000	0.00821	0.00004	0.00011
	VST	500	2.57679	47.61970	54.25957
	VST	1,000	0.68741	13.87630	14.34884
	VST	5,000	0.16172	3.01595	3.04210
	VST	10,000	0.12386	1.23927	1.25461
	VST	20,000	0.11537	0.71021	0.72352
ME	1/ $\mu$	500	0.12867	0.00384	0.02039
	1/ $\mu$	1,000	0.03050	0.00250	0.00343
	1/ $\mu$	5,000	0.00494	0.00047	0.00050
	1/ $\mu$	10,000	0.00182	0.00017	0.00017
	1/ $\mu$	20,000	0.00023	0.00006	0.00006
	VST	500	1.70225	12.01048	14.90812
	VST	1,000	0.35554	1.95599	2.08240
	VST	5,000	0.19228	0.22072	0.25769
	VST	10,000	0.08879	0.07469	0.08257
	VST	20,000	0.03281	0.01565	0.01672

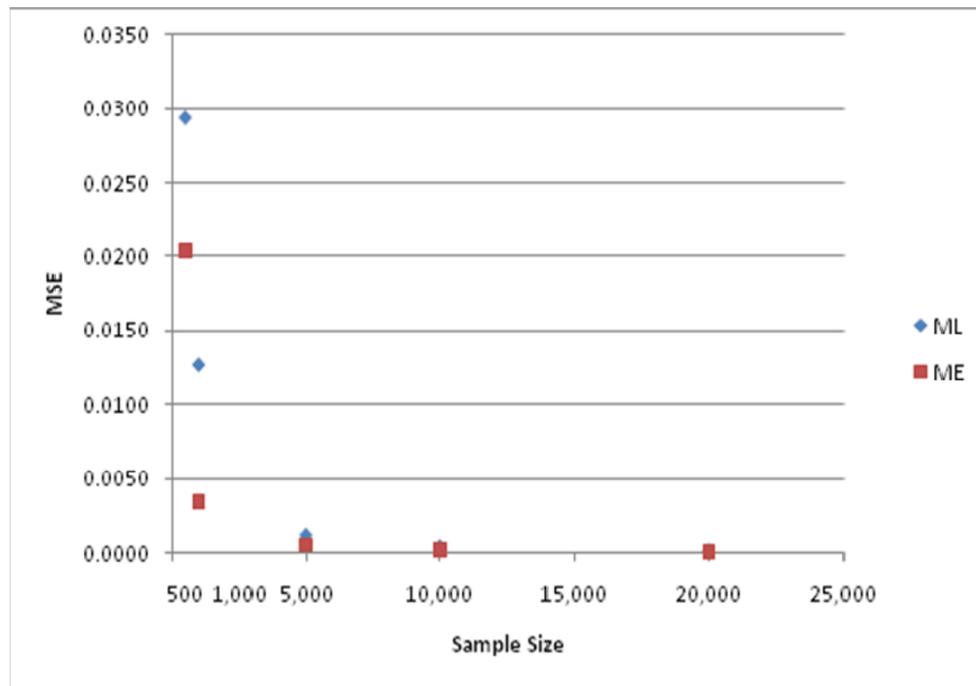
(\*) Defined as the sum of the variance and the square of the bias.

For parameter values  $\phi = 1/\mu = 0.2$  and  $\phi = 1/\mu = 0.9$ , the bias, variance and MSE estimates with the same sample sizes are found in Appendix B. In every case both bias and MSE are smaller for the ME estimator.

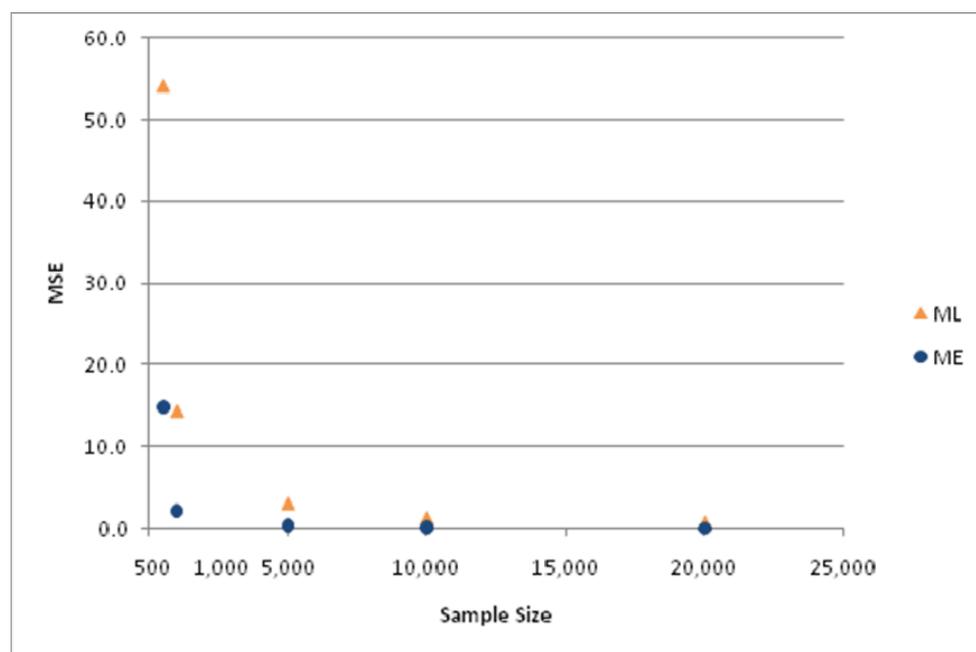
The differences in MSE for the two estimators are depicted in Figures 5 and 6. The various results just presented lead to the conclusion that the ME estimator is systematically superior to the ML estimator in this hierarchical tree structure, particularly for small sample sizes.

The differences between both estimation approaches may respond to their abilities for reproducing the market modal shares of the calibration samples. It is well established that when the utility functions of a multinomial model contain a constant term by mode the modeled and observed modal shares are always the same [24]. This is not the case for the hierarchical logit model when its parameters are estimated by ML. Under the ME approach, however, the problem constraints force the market modal shares to be reproduced.

**Figure 5.** MSE of ML and ME estimators for parameter value  $\phi = 1/\mu = 0.5$ .



**Figure 6.** MSE of ML and ME estimators for SVT.

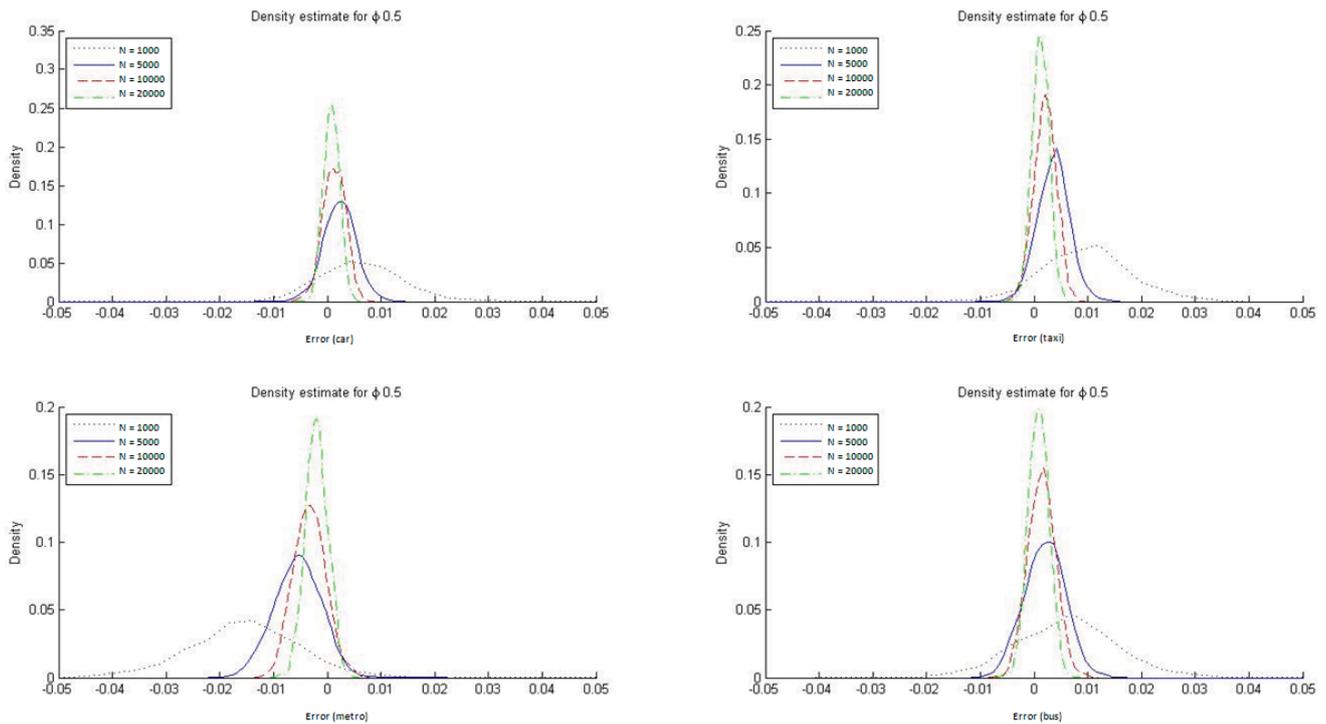


To determine the ability of the ML estimators to reproduce the market modal shares in HL models, we constructed histograms of DMS, defined as the difference between the modeled and observed modal shares (see Figure 7). These differences were estimated for each of the four transportation modes. As is apparent from the figure, the ML estimators of the hierarchical logit model do not, on average, reproduce the sample modal shares, particularly in small samples.

It is also evident from Figure 7 that with HL models the ML estimator reproduces the market modal shares only asymptotically whereas by construction the ME estimators always reproduce them. In

Appendix C we report the DMS distributions for the four transportation modes when  $\phi = 1/\mu = 0.2$  and  $\phi = 1/\mu = 0.9$ . The conclusions are similar for the two cases, but with  $\phi = 1/\mu = 0.9$  it is especially clear that the ML estimate for the HL model reproduces the modal shares must more accurately. This result is to be expected given that a value for the parameter  $\phi$  closer to 1 implies that the HL model is more similar to an MNL one, which always reproduces the observed modal shares when estimated with ML.

**Figure 7.** Distribution of DMS with ML estimation ( $\phi = 1/\mu = 0.5$ ).



### 3.2. Estimate of Consumer Surplus

Also of interest is the comparison of the ML and ME estimates of consumer surplus generated by our combined destination and mode choice model. This measure of the welfare perceived by the transportation system user is expressed by the expected maximum utility (EMU), which is written as follows:

$$-\sum_{a,g,i} N_i^0 p_{ag/i} \ln p_{ag/i} = -\sum_{i,g} N_i^0 p_{g/i} \ln p_{g/i} - \sum_{a,g,i} N_i^0 p_{g/i} p_{a/gi} \ln p_{a/gi} \tag{29}$$

This expression gives the consumer surplus of individual  $i$  in group  $g$ . The group refers to the origin-destination pair representing the trip taken, and since our model is an aggregate one, all individuals in a given group have the same utility function  $V_{agi}$  for each mode alternative  $a$ . This being the case, we have  $EMU_{gi} = EMU_g$ , and can therefore estimate the average EMU as:

$$EMU_{average} = \frac{\sum_{a,g} t_{ag} EMU_g}{\sum_{a,g} t_{ag}} \tag{30}$$

where  $t_{ag}$  is the number of individuals in group  $g$  that takes alternative  $a$ . We will use average EMU to make comparisons with the results generated by the simulations for various sample sizes (the sum of the  $t_{ag}$  will therefore equal the size of the sample from which the parameters that give the EMU are estimated).

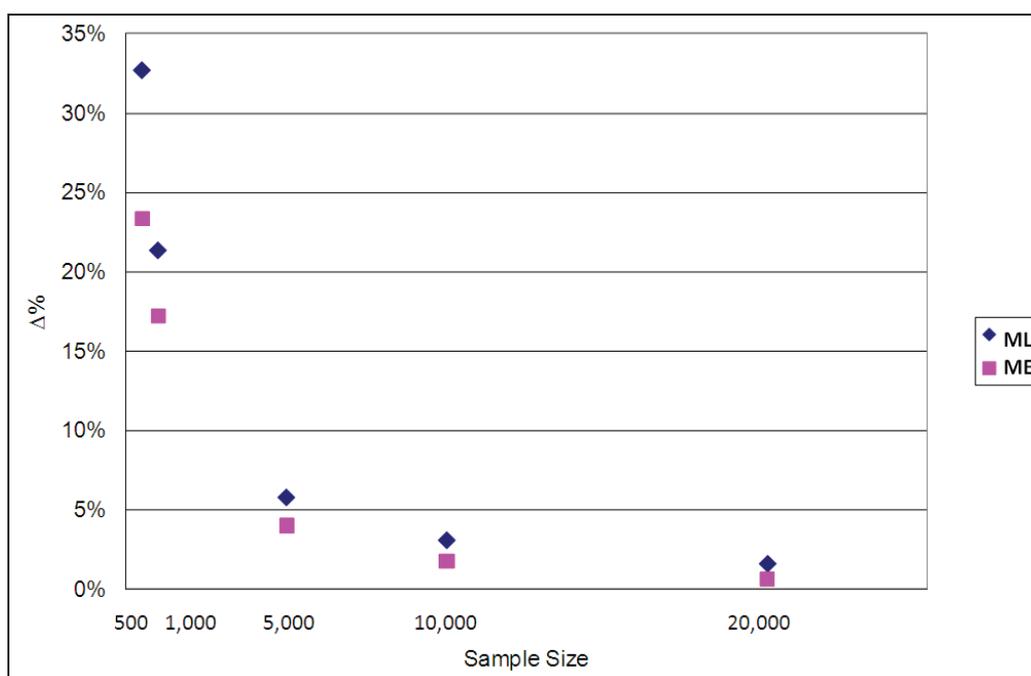
The average EMU values estimated for each case using the ML and ME estimators are compared with the population parameters used for the simulations in Table 4. The percentage differences between the estimated and population parameter values are graphed in Figure 8. It is clear from both the table and the figure that the EMU estimates produced by the ML parameters are more biased than those of the ME parameters, especially when sample sizes are small, though they both converge asymptotically to the true values when the sample size increases. The corresponding results for  $\phi = 1/\mu = 0.2$  and  $\phi = 1/\mu = 0.9$  are reported in Appendix D, confirming that for all three cases ( $\phi = 0.2$ ,  $\phi = 0.5$  and  $\phi = 0.9$ ) the ME estimator of average EMU is less biased than the ML one.

**Table 4.** Average EMU.

SAMPLE SIZE	SIMULATION	ML	$\Delta\%$ ML (*)	ME	$\Delta\%$ ME (*)
500	7.0574	4.7481	32.7%	5.4107	23.3%
1,000	7.0601	5.5514	21.4%	5.8450	17.2%
5,000	7.0918	6.6807	5.8%	6.8073	4.0%
10,000	7.0755	6.8564	3.1%	6.9500	1.8%
20,000	7.0679	6.9539	1.6%	7.0223	0.6%

(\*) Calculated as the difference between population (simulation) and estimated EMU divided by population EMU.

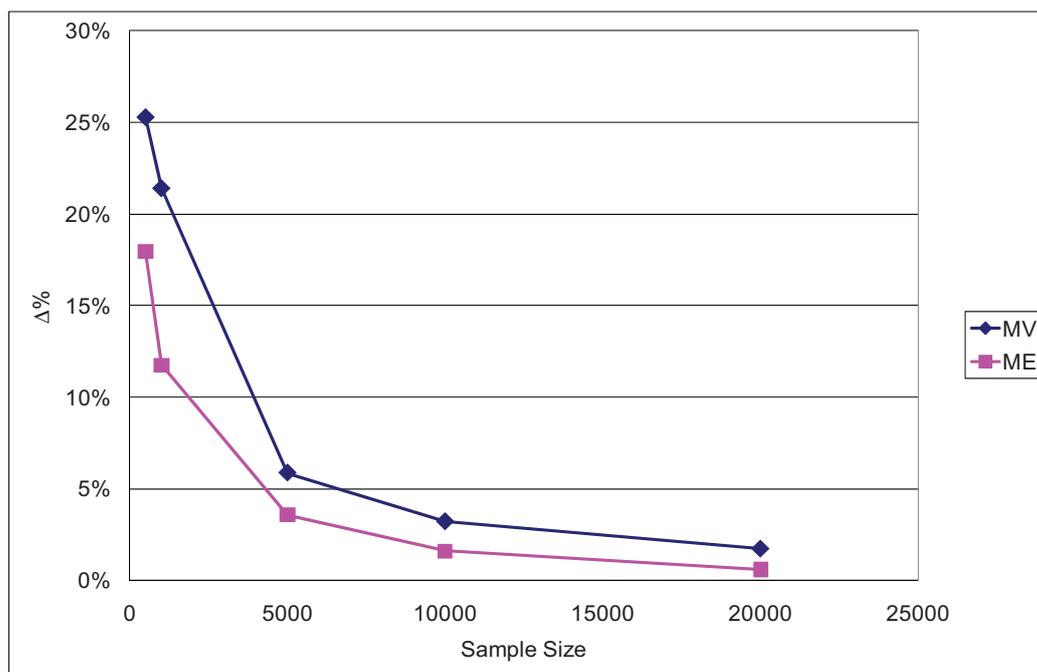
**Figure 8.** Percentage differences between population and estimated average EMU ( $\phi = 1/\mu = 0.5$ ).



### 3.3. Out-of-Sample Prediction

To study the behavior of the ML and ME estimators of consumer surplus we varied the travel times of the four transportation modes and estimated the resulting changes in average EMU as given by (13). The travel time variations consisted in reducing this factor by 10% for all four modes. The results obtained are shown for various sample sizes in Figure 9.

**Figure 9.** Percentage differences in  $\Delta\text{EMU}$  for a 10% reduction in travel time ( $\phi = 1/\mu = 0.5$ ).



In every case it can be observed that the ML overestimates EMU to a greater extent than does ME. The overestimation effect is particularly significant for the small samples (500 and 1,000 data items), which is consistent with Figure 9. The corresponding results for  $\phi = 1/\mu = 0.2$  and  $\phi = 1/\mu = 0.9$  are reported in Appendix E, confirming that for all three cases ( $\phi = 0.2$ ,  $\phi = 0.5$  and  $\phi = 0.9$ ) the ME overestimate of  $\Delta\text{EMU}$  is smaller than the ML one.

## 4. Conclusions

In the context of aggregate transportation demand forecasting and land use planning, entropy maximization problems are often formulated, mainly because their solutions are the well-known multinomial logit and hierarchical logit models. The parameters of these models are normally estimated using the maximum likelihood method, but they can also be estimated by solving the entropy maximization problems directly. These latter estimators are referred to here as maximum entropy estimators. It has long been known (see [1,22]) that both estimation methods lead to the same results if the model is multinomial logit.

This work extended the analysis to the case of aggregate hierarchical logit models. We began by formulating a general problem of maximizing the hierarchical entropy and deducing that its solution is a hierarchical logit model. We then observed that the maximum entropy estimators were different from

the maximum likelihood ones, especially in that the latter do not reproduce either the average values of the explanatory variables or the observed market modal shares (of the calibration sample) whereas the maximum entropy estimators do reproduce them by construction.

The two estimators were then subjected to various empirical analyses. The population parameters of a relatively general travel demand hierarchical model were estimated using Monte Carlo simulations with samples of various sizes. The results obtained showed that the maximum entropy estimator is superior to the maximum likelihood estimator, especially for smaller sample sizes. More specifically, the maximum entropy estimator exhibited less bias and a smaller mean square error. The reduced bias in turn results in underestimations of consumer surplus with maximum likelihood.

Though similar analyses of other hierarchical structures are required, we may conclude on the basis of the results presented here that the maximum entropy approach is a better alternative for estimating hierarchical logit aggregate models than the maximum likelihood approach, particularly with small or medium-size samples such as those typically used in actual transportation planning processes.

## References

1. Anas, A. Discrete choice theory, information theory and the multinomial logit and gravity models. *Transp. Res.* **1983**, *17*, 13–23.
2. Williams, H.C.W.L. On the formation of travel demand models and economic evaluation measures of user benefit. *Environ. Plan.* **1977**, *9*, 285–344.
3. Wilson, A.G. *Entropy in Urban and Regional Modeling*; Pion: London, UK, 1970.
4. Morrison, W.; Thumann, R. Lagrangian multiplier approach to the solution of a special constrained matrix problem. *J. Reg. Sci.* **1980**, *20*, 279–292.
5. Fotheringham, A. A new set of spatial interaction models: The theory of competing destinations. *Environ. Plan.* **1983**, *15A*, 15–36.
6. Fotheringham, A. Modeling hierarchical destination choice. *Environ. Plan.* **1986**, *18*, 401–418.
7. Fang, S.; Tsao, J. Linearly-constrained entropy maximization problem with quadratic cost and its applications to transportation planning problems. *Transp. Sci.* **1995**, *29*, 353–365.
8. Thorsen, I.; Gitlesen, J.P. Empirical evaluation of alternative model specifications to predict commuting flows. *J. Reg. Sci.* **1998**, *38*, 273–292.
9. De Grange, L.; Ibeas, A.; Gonzalez, F. A hierarchical gravity model with spatial correlation: Mathematical formulation and parameter estimation. *Netw. Spat. Econ.* **2009**, doi:10.1007/s11067-008-9097-0.
10. De Grange, L.; Fernandez, J.E.; De Cea, J. A consolidated model of trip distribution. *Transp. Res.* **2010**, *46*, 61–75.
11. Boyce, D.; LeBlanc, L.; Chon, K.; Lee, Y.; Lin, K. Implementation and computational issues for combined models of location, destination, mode and route choice. *Environ. Plan.* **1983**, *15*, 1219–1230.
12. Boyce, D.; LeBlanc, L.; Chon, K. Network equilibrium models of urban location and travel choices: A retrospective survey. *J. Reg. Sci.* **1988**, *28*, 159–183.
13. Safwat, K.; Magnanti, T. A combined trip generation, trip distribution, modal split and traffic assignment model. *Transp. Sci.* **1988**, *22*, 14–30.

14. Brice, S. Derivation of nested transport models within a mathematical programming framework. *Transp. Res.* **1989**, *23*, 19–28.
15. Fernandez, J.E.; De Cea, J.; Florian, M.; Cabrera, E. Network equilibrium models with combined modes. *Transp. Sci.* **1994**, *28*, 182–192.
16. Oppenheim, N. *Urban Travel Demand Modeling*; John Wiley & Sons: New York, NY, USA, 1995.
17. Abrahamsson, T.; Lundqvist, L. Formulation and estimation of combined network equilibrium models with applications to stockholm. *Transp. Sci.* **1999**, *33*, 80–100.
18. Boyce, D.; Bar-Gera, H. Validation of multiclass urban travel forecasting models combining origin-destination, mode, and route choices. *J. Reg. Sci.* **2003**, *43*, 517–540.
19. Ham, H.; Tschangho, J.; Boyce, D. Implementation and estimation of a combined model of interregional, multimodal commodity shipments and transportation network flows. *Transp. Res.* **2005**, *39*, 65–79.
20. Garcia, R.; Marin, A. Network equilibrium with combined modes: Models and solution algorithms. *Transp. Res.* **2005**, *39*, 223–254.
21. De Cea, J.; Fernandez, J.E.; De Grange, L. Combined models with hierarchical demand choices: A multi-objective entropy optimization approach. *Transp. Rev.* **2008**, *28*, 415–438.
22. Donoso, P.; De Grange, L. A Microeconomic interpretation of the maximum entropy estimator of multinomial logit models and its equivalence to the maximum likelihood estimator. *Entropy* **2010**, *12*, 2077–2084.
23. SECTRA. Encuesta Origen Destino de Viajes 2001 para el Gran Santiago. Secretaría Interministerial de Planificación de Transporte, 2002, Santiago de Chile. Available online: [http://www.sectra.cl/transporte/transporte\\_urbano\\_eod\\_frm.html](http://www.sectra.cl/transporte/transporte_urbano_eod_frm.html) (accessed on 21 March 2010).
24. Ortuzar, J.d.D.; Willumsen, L.G. *Modeling Transport*; John Wiley & Sons: Chichester, UK, 2001.

## Appendices

### Appendix A: Microeconomic Interpretation of the Entropy Maximization Dual Problem

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

$$EMU_i = \ln \sum_{g/i} \exp(V_{gi}^* + V_{gi}), \quad V_{gi}^* = \frac{1}{\mu_g} \ln \left( \sum_{a/g,i} \exp(\mu_g V_{agi}) \right) \quad \text{for all } g, i \quad (31)$$

For the hierarchical logit model, it is known that

$$P_{ag/i} = P_{a/gi} P_{g/i} = \frac{\exp(\mu_g V_{agi})}{\sum_{a'} \exp(\mu_g V_{a'gi})} \frac{\exp(V_{gi}^* + V_{gi})}{\sum_{g'} \exp(V_{g'i}^* + V_{g'i})} \quad (32)$$

Since  $P_{g/i} = \frac{\exp(V_{gi}^* + V_{gi})}{\sum_{g'} \exp(V_{g'i}^* + V_{g'i})}$ , we have

$$\sum_{g'} \exp(V_{g'i}^* + V_{g'i}) = \frac{\exp(V_{gi}^* + V_{gi})}{p_{g/i}} \tag{33}$$

$$\ln \left( \sum_{g'} \exp(V_{g'i}^* + V_{g'i}) \right) = V_{gi}^* + V_{gi} - \ln p_{g/i} = EMU_i \tag{34}$$

Multiplying (33) by  $p_{g/i}$  and then summing over  $g$ , we obtain

$$p_{g/i} EMU_i = p_{g/i} (V_{gi}^* + V_{gi}) - p_{g/i} \ln p_{g/i} / \sum_g \tag{35}$$

$$EMU_i = \sum_{g/i} p_{g/i} (V_{gi}^* + V_{gi}) - \sum_{g/i} p_{g/i} \ln p_{g/i} \tag{36}$$

Since by (31) and (32),  $V_{gi}^* = \frac{1}{\mu_g} \ln \left( \sum_{a/g,i} \exp(\mu_g V_{agi}) \right)$  and  $p_{a/gi} = \frac{\exp(\mu_g V_{agi})}{\sum_{a'} \exp(\mu_g V_{a'gi})}$ , we easily get

$$V_{gi}^* = V_{agi} - \frac{1}{\mu_g} \ln p_{a/gi} \rightarrow p_{a/gi} V_{gi}^* = p_{a/gi} V_{agi} - p_{a/gi} \frac{1}{\mu_g} \ln p_{a/gi} / \sum_{a/g,i} \tag{38}$$

$$V_{gi}^* = \sum_{a/g,i} p_{a/gi} V_{agi} - \frac{1}{\mu_g} \sum_{a/g,i} p_{a/gi} \ln p_{a/gi} \tag{39}$$

Finally, substituting (38) into (36) we obtain

$$EMU_i = \sum_{g/i} p_{g/i} \left( \sum_{a/g} p_{a/gi} V_{agi} - \frac{1}{\mu_g} \sum_{a/g} p_{a/gi} \ln p_{a/gi} + V_{gi} \right) - \sum_{g/i} p_{g/i} \ln p_{g/i} \tag{39}$$

$$EMU_i = \sum_{g/i} p_{g/i} \left( \sum_{a/g} p_{a/gi} V_{agi} \right) + \sum_{g/i} p_{g/i} V_{gi} - \sum_{g/i} \frac{1}{\mu_g} p_{g/i} \left( \sum_{a/g,i} p_{a/gi} \ln p_{a/gi} \right) - \sum_{g/i} p_{g/i} \ln p_{g/i} \tag{40}$$

Based on (40) we can now formulate the following optimization problem that each individual must solve:

$$\begin{aligned} \max_{\{p_{a/gi}, p_{g/i}\}} & \sum_{g/i} p_{g/i} \left( \sum_{a/g} p_{a/gi} V_{agi} \right) + \sum_{g/i} p_{g/i} V_{gi} - \sum_{g/i} \frac{1}{\mu_g} p_{g/i} \left( \sum_{a/g,i} p_{a/gi} \ln p_{a/gi} \right) - \sum_{g/i} p_{g/i} \ln p_{g/i} \\ \text{s.t.} : & \sum_{a/g,i} p_{a/gi} = 1 \quad \text{for all } g, i \quad (\Phi_{gi}) \\ & \sum_{g/i} p_{g/i} = 1 \quad \text{for all } i \quad (\gamma_i) \end{aligned} \tag{41}$$

The optimality conditions for (41) are then

$$p_{a/gi} = \frac{\exp(\mu_g V_{agi})}{\sum_{a'} \exp(\mu_g V_{a'gi})}, \quad \Phi_{gi} = p_{g/i} \left[ \frac{1}{\mu_g} \ln \sum_{a/g,i} \exp(\mu_g V_{agi}) + 1 \right] \tag{42}$$

$$P_{g/i} = \frac{\exp(V_{gi}^* + V_{gi})}{\sum_{g'} \exp(V_{g'i}^* + V_{g'i})}, \quad \gamma_i = \ln \sum_{g/i} \exp(V_{gi}^*) \tag{43}$$

Assuming a linear and additive utility function  $V_{agi} = \sum_k \beta_k x_{agik}$  for each individual, and given that multiple individuals make their optimal decision (based on mixes strategies) simultaneously, we demonstrate that the optimization problem (41) is the equivalent problem of the optimization problem formulated in Section 2.3. Therefore, the value of the objective function of the latter problem at the optimum is the sum over all individuals of their maximum expected utilities from the available alternatives.

**Appendix B: Estimates of Bias, Variance and Mean Square Error for Parameter Values**

$\phi = 1/\mu = 0.2$  and  $\phi = 1/\mu = 0.9$

**Table 5.** Summary of results for ML and ME estimators ( $\phi = 1/\mu = 0.2$ ).

METHOD	PARAMETER	SAMPLE SIZE	BIAS	VARIANCE	MSE (*)
ML	1/μ	500	0.09615	0.00198	0.01122
	1/μ	1,000	0.06589	0.00030	0.00464
	1/μ	5,000	0.00966	0.00011	0.00020
	1/μ	10,000	0.00716	0.00005	0.00010
	1/μ	20,000	0.00274	0.00001	0.00002
	VST	500	1.10301	46.22001	47.43665
	VST	1,000	0.45788	28.97585	29.18551
	VST	5,000	0.23895	3.70613	3.76323
	VST	10,000	0.07104	2.73371	2.73876
	VST	20,000	0.04468	0.91350	0.91550
ME	1/μ	500	0.12461	0.00017	0.01570
	1/μ	1,000	0.05836	0.00048	0.00389
	1/μ	5,000	0.00517	0.00007	0.00010
	1/μ	10,000	0.00410	0.00004	0.00006
	1/μ	20,000	0.00300	0.00003	0.00004
	VST	500	0.81525	5.40041	6.06504
	VST	1,000	0.64647	3.43284	3.85076
	VST	5,000	0.20451	0.28043	0.32226
	VST	10,000	0.09542	0.12176	0.13087
	VST	20,000	0.04107	0.07484	0.07653

(\*) Defined as the sum of the variance and the square of the bias.

**Table 6.** Summary of results for ML and ME estimators ( $\phi = 1/\mu = 0.9$ ).

METHOD	PARAMETER	SAMPLE SIZE	BIAS	VARIANCE	MSE (*)
ML	$1/\mu$	500	0.22624	0.00432	0.05550
	$1/\mu$	1,000	0.12514	0.00285	0.01851
	$1/\mu$	5,000	0.03456	0.00049	0.00168
	$1/\mu$	10,000	0.01835	0.00033	0.00067
	$1/\mu$	20,000	0.01331	0.00016	0.00034
	VST	500	1.27926	28.32571	29.96221
	VST	1,000	1.16837	11.69802	13.06312
	VST	5,000	0.12918	1.92596	1.94265
	VST	10,000	0.08153	0.96521	0.97186
	VST	20,000	0.05641	0.38886	0.39204
ME	$1/\mu$	500	0.23444	0.00784	0.06280
	$1/\mu$	1,000	0.09184	0.00452	0.01296
	$1/\mu$	5,000	0.01017	0.00074	0.00084
	$1/\mu$	10,000	0.00754	0.00027	0.00032
	$1/\mu$	20,000	0.00450	0.00016	0.00018
	VST	500	0.70652	5.28093	5.78010
	VST	1,000	0.65276	2.30125	2.72735
	VST	5,000	0.16269	0.18473	0.21119
	VST	10,000	0.09341	0.09402	0.10275
	VST	20,000	0.00434	0.00270	0.00272

(\*) Defined as the sum of the variance and the square of the bias.

**Appendix C: Distribution of DMS with ML Estimation for Parameter Values  $\phi = 1/\mu = 0.2$  and  $\phi = 1/\mu = 0.9$**

**Figure 10.** Distribution of DMS with ML estimation ( $\phi = 1/\mu = 0.2$ ).

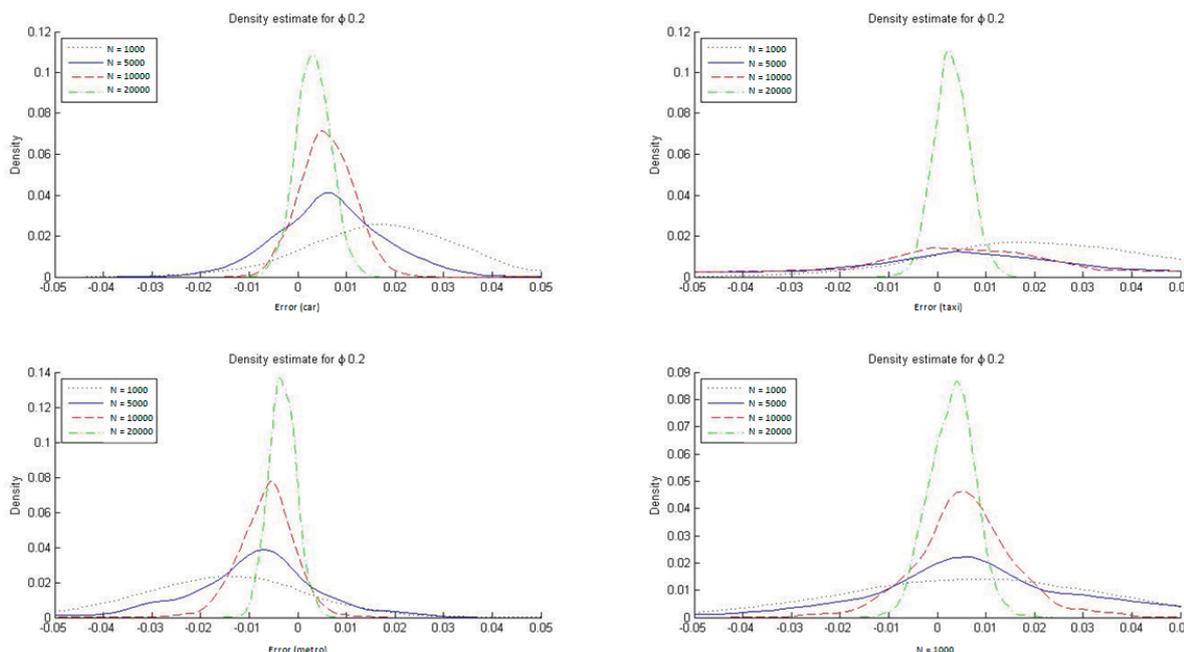
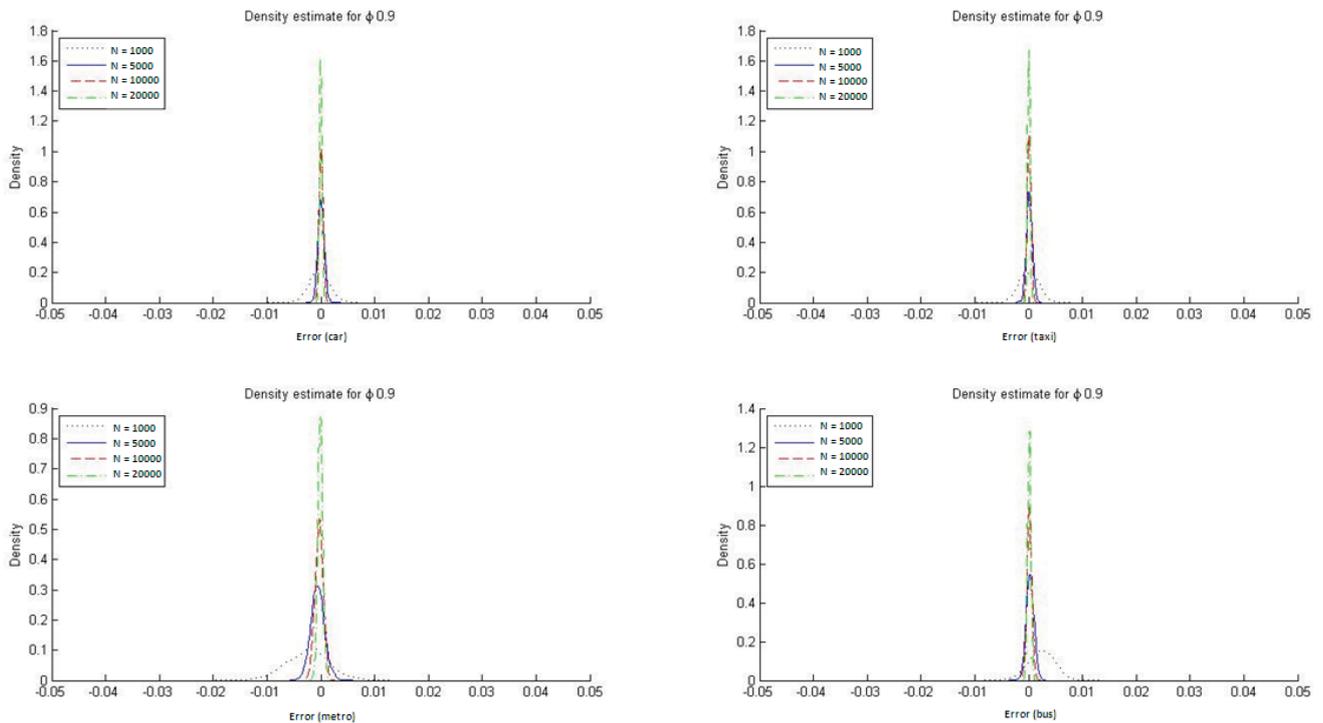


Figure 11. Distribution of DMS with ML estimation ( $\phi = 1/\mu = 0.9$ ).



Appendix D: Estimates of EMU for Parameter Values  $\phi = 1/\mu = 0.2$  and  $\phi = 1/\mu = 0.9$

Table 7. Average EMU ( $\phi = 1/\mu = 0.2$ ).

SAMPLE SIZE	SIMULATION	ML	$\Delta\%$ ML (*)	ME	$\Delta\%$ ME (*)	$\Delta\%$ ML/ME (**)
500	3.4436	1.7934	47.9%	2.1346	38.0%	26.1%
1,000	3.4765	2.3030	33.8%	2.5821	25.7%	31.2%
5,000	3.4075	3.2077	5.9%	3.2767	3.8%	52.8%
10,000	3.4228	3.2882	3.9%	3.3239	2.9%	36.1%
20,000	3.4205	3.3780	1.2%	3.3936	0.8%	58.1%

(\*) Calculated as the difference between population (simulation) and estimated EMU divided by population EMU; (\*\*) Calculated as the ratio of  $\Delta\%$  MV to  $\Delta\%$  ME minus 1.

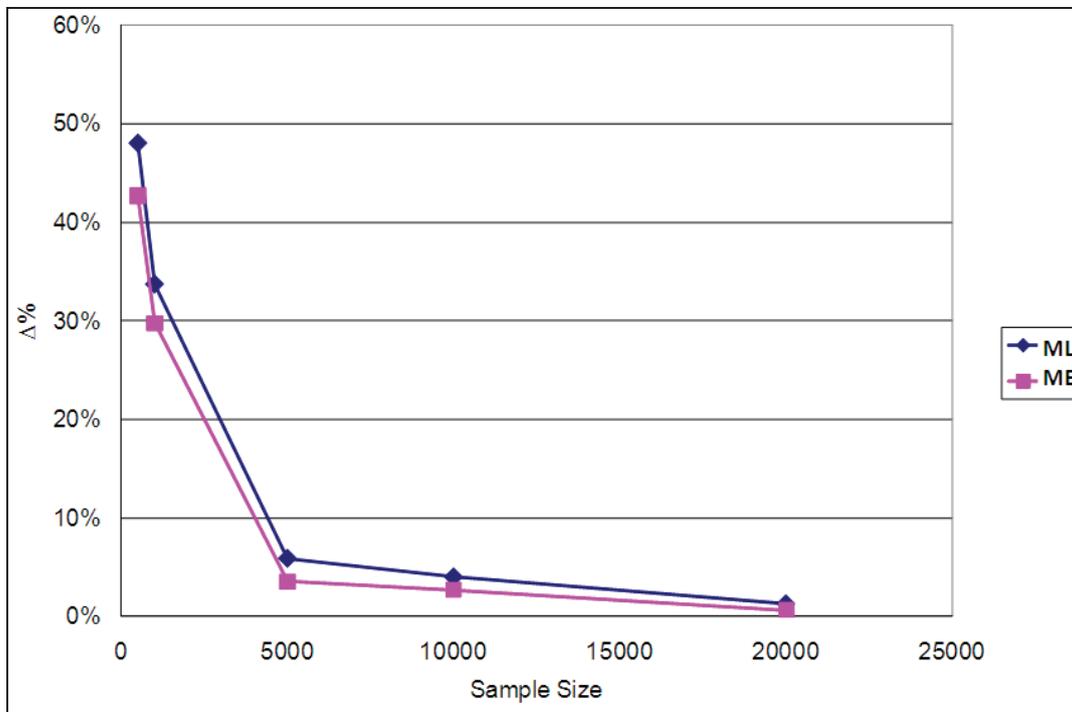
Table 8. Average EMU ( $\phi = 1/\mu = 0.9$ ).

SAMPLE SIZE	SIMULATION	ML	$\Delta\%$ ML (*)	ME	$\Delta\%$ ME (*)	$\Delta\%$ ML/ME (**)
500	11.7592	8.8470	24.8%	9.6120	18.3%	35.6%
1,000	11.7305	10.0668	14.2%	10.6354	9.3%	51.9%
5,000	11.7876	11.3636	3.6%	11.5302	2.2%	64.8%
10,000	11.7835	11.5360	2.1%	11.6765	0.9%	131.2%
20,000	11.7728	11.6164	1.3%	11.6959	0.7%	103.3%

(\*) Calculated as the difference between population (simulation) and estimated EMU divided by population EMU; (\*\*) Calculated as the ratio of  $\Delta\%$  MV to  $\Delta\%$  ME minus 1.

**Appendix E: Estimates of  $\Delta EMU$  for a 10% Reduction in Travel Time for the Parameters  $\phi = 1/\mu = 0.2$  and  $\phi = 1/\mu = 0.9$**

**Figure 12.** Percentage differences in  $\Delta EMU$  for a 10% reduction in travel time ( $\phi = 1/\mu = 0.2$ ).



**Figure 13.** Percentage differences in  $\Delta EMU$  for a 10% reduction in travel time ( $\phi = 1/\mu = 0.9$ ).

