ON THE EFFICIENCY OF MIXED LOGIT PARAMETER ESTIMATES: ANALYSING THE EFFECT OF DATA RICHNESS

Elisabetta Cherchi
CRiMM - Dipartimento di Ingegneria del Territorio
Facoltà di Ingegneria - Università di Cagliari, Piazza d'Armi, 16 - 09123 Cagliari, Italia
Tel: 39 70 675 5274, Fax: 39 70 675 5261, E-mail: echerchi@unica.it

Juan de Dios Ortúzar
Departamento de Ingeniería de Transporte y Logística
Pontificia Universidad Católica de Chile, Casilla 306, Cod. 105, Santiago 22, Chile
Tel: (56 2) 354 48 22, Fax: (56 2) 553 02 81, E-mail: jos@ing.puc.cl

ABSTRACT

The objective of this paper is to analyse the effect that the richness of a dataset has on the efficiency of parameters estimated with multinomial and mixed logit models, trying to isolate the effect due to the attribute ranges from that due to the number of attribute levels. Efficiency is inversely related to the value of the attributes associated to the parameters and to the number of observations. Moreover, it depends also on a second order function of the probabilities, such that as the variability of the data and the value of the scale factor increase, the efficiency of the parameters decreases. We found that the variability of the data does not always increase the efficiency of the estimates; rather, in some cases it might not be beneficial to increase data variability or it might even be better to have a smaller range of variation. This is a crucial problem if we need to gather revealed preference data or to design a stated preference experiment.

Keywords: effciency of parameter estimates, richness of data, mixed logit

RESUMEN

El objetivo de este trabajo es analizar el efecto que tiene la riqueza de los dato en la eficiencia de los parámetros estimados en modelos logit, multinomial y mixto, aislando el efecto debido al rango de variación de los atributos del efecto debido al número de niveles de cada atributos. La eficiencia de los parámetros está inversamente relacionada con el valor de los atributos y el número de observaciones. Sin embargo, es también una función de segundo orden de las probabilidades de elección de las alternativas, de manera que si la variabilidad de los datos y el valor del factor de escala del modelo crecen, la eficiencia de los parámetros estimados decrece. En este trabajo encontramos que no siempre la eficiencia de los parámetros estimados aumenta con la variabilidad de los datos. Al contrario, hay casos en que no existen ventajas en aumentar la variabilidad de los datos, o puede ser mejor utilizar un rango de variación más pequeño. Este es un problema crucial si se debe recolectar datos de preferencias reveladas o diseñar experimentos de preferencias declaradas.

Palabras clave: eficiencia de los parámetros estimados, riqueza de los datos, modelo Logit mixto

1. INTRODUCTION

The Maximum Likelihood (ML) method is the most widely used technique to estimate discrete choice models. The ML method is straightforward to apply and, under relatively general conditions, guarantees that the estimated parameters are asymptotically consistent and efficient. That is, as the sample gets larger the distribution of the estimators collapses on the true parameter values and their variance is the smallest. However, the ML method does not always guarantee that in the case of real samples (i.e. samples of relatively small size) the estimated parameters are efficient or unbiased (i.e. their expected values equate the true values). Moreover, while the log-likelihood function is always asymptotically unbiased in the multinomial logit (MNL) model, it is biased in the mixed multinomial logit (MMNL) model, due to the log transformation of the simulated probability used in estimation (Borsch-Supan and Hajivassiliou, 1993).

Efficiency is measured by the inverse of the covariance matrix and this is inversely related to the square of the attribute values associated to the parameters. It is well known that a higher variability in the data allows estimating more robust models, and variability is often associated to the amount of explanation that can be extracted from the data. However, even in the simple MNL the variance depends not only on the data variability but also on the number of observations and on a second order function of the probabilities. We show that this function is such, that as the variability of the data and the value of the scale factor increases, the efficiency of the parameter decreases because the logit probability depends both on the data variability and the variance of the error term (through the scale factor).

So, the variability of the data does not always increase the efficiency of the estimates; rather, in some cases it might not be beneficial to increase data variability or it might even be better to have a small range of variation. This is a crucial problem if we need to collect revealed preference (RP) data or to design a stated preference (SP) experiment. While in the RP case the degree of variability is often related to the number of observations and cannot always be controlled by the modeller, in the case of SP data it is the analyst who decides what type of data (trade-off) should be presented to respondents. In both cases it is crucial first to be clear about how the efficiency of the estimates vary with the data and what portion of the phenomenon are we able to explain with the estimated models; then, we can consider what we should/could do to improve efficiency.

Bliemer and Rose (2004; 2005) analysed the effect of the number of alternatives, attributes, and attribute levels on the “optimal sample size” for a SP experiment; this was defined as the number of observations needed such that adding more would not significantly improve the asymptotic efficiency of the estimated parameters. In particular, they found that the number of attribute levels did not play a role in determining the optimal sample size for a MNL model, while the attribute level range could be the possible explanation for some convergence problems encountered in their experiments; however, they did not explicitly analyse the effect of the attribute range. They also discussed the relation between the number of attribute levels and attribute ranges, as the range usually increases with the levels, but they did not attempt to disentangle these two effects.

Our work complements that of Bliemer and Rose (2004; 2005) as we specifically analyse the effect of attribute variability, trying to isolate the effects of the different elements that determine this variability, i.e. the range and number of levels of the attributes. We also analyse how the efficiency of the estimated parameters varies if a customised or fixed SP design is used and if it depends on whether the attribute levels are defined as absolute values or as percentage variations.

In section 2 we discuss theoretically the effect of data variability on the efficiency of the estimated parameters and its relation with the dimension of the variance of the error terms and size of the sample used, for both RP and SP data. In section 3 we describe a Monte Carlo experiment designed to simulate a collection of datasets as well as the results of the estimation of several models. Section 4 discusses the efficiency of the estimated parameters and section 5 summarizes our conclusions.

2. ON THE EFFICIENCY OF THE ESTIMATED PARAMETERS

A good estimator should be unbiased, consistent and efficient. In the case of discrete choice models only asymptotic properties have been demonstrated, thus an estimator is asymptotically unbiased and efficient if, as the sample gets larger, the expected value of the estimated parameters ($\hat{\beta}$) equals the true parameters (β) and its variance is the smallest[1]:

$$\lim_{n\to\infty} E(\hat{\beta}) = \beta \tag{1}$$

$$\mathrm{var}(\hat{\beta}) = \min_k \mathrm{var}(\hat{\beta}_k) \qquad \forall k \tag{2}$$

On the other hand, the estimator is consistent if and only if, its error and variance approach zero in the limit:

$$\lim_{n\to\infty} E(\hat{\beta}-\beta)^2 = \lim_{n\to\infty}\left[(error\,\hat{\beta})^2 + \mathrm{var}(\hat{\beta})\right] = 0 \tag{3}$$

It is also well known that the maximum-likelihood (ML) method guarantees, under relatively general conditions, that the estimated parameters are asymptotically consistent and efficient. In particular, using the Cramér-Rao theorem, the minimum variance estimator is that which meets the following bound:

$$\mathrm{var}(\hat{\boldsymbol{\beta}}) \geq \mathbf{B}^{-1} \tag{4}$$

where **B** is the asymptotic variance-covariance matrix of the estimates i.e. the variance of the asymptotic distribution (called also information matrix or Hessian). Now, in ML estimation $\mathbf{B} = -E\left[\nabla^2 \ell\right]$, where $\ell = \ln \prod_{n=1}^{N} p_{jn}^{y_{jn}}$, is the log-likelihood function with respect to the parameters evaluated at their true values. Thus the variance of the k^{th} estimated parameter (which is the k^{th} element of the diagonal of **B**) is equal to:

[1] The definition of efficiency often includes also the condition that the estimator is unbiased (i.e. it fulfils both equations (1) and (2)) to facilitate the comparison among estimators.

$$E\left[\frac{\partial^2 \ell}{\partial \beta_k^2}\right] \cong \sum_{n=1}^{N} \sum_{j \in I_n} \left[\frac{\partial^2 \left(y_{jn} \ln p_{jn}(\mathbf{x}_{jn};\boldsymbol{\beta})\right)}{\partial \beta_k^2}\right]_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \tag{5}$$

where p_{jn} is the probability that individual n will choose alternative j among the alternatives belonging to her choice set (I_n) and y_{jn} equals one if j is the alternative actually chosen by the individual n and zero otherwise. Note that as the likelihood function depends on the estimated parameters and on the attributes ($\mathbf{x}$) associated to each parameter ($\boldsymbol{\beta}$), to compute the expectation of the information matrix we should know the actual value of the parameter and the true distributions of the attributes to take their expected values. As this is not the case, the best approximation of the expectation of the information matrix is the second derivative of the log-likelihood function calculated at the point of the estimated parameters.

Equation (5) shows that the efficiency of the estimated parameter depends on the sample size, on the values of the attributes associated to the estimated parameters and on the probabilities associated to the chosen alternative. In particular, the logit probability depends, among other things, on the data variability and on the variance of the error term (through the scale factor); this makes understanding the sensitivity of the efficiency of the estimated parameters a complex task. We will analyse these effects in some detail in the following paragraphs, with the aim of finding out what makes the efficiency of the estimated parameters to increase.

2.1. Efficiency in the MNL

Let us assume first a binary logit model (two alternatives j=1,2) with only one parameter (k=1). This allows simplifying the notation and the discussion, but it does not change the general results; at the end of this section we will provide the generalisation to the MNL. Assuming one or more parameters is equivalent, because the expression of the variance in equation (5) does not change with the number of parameters, but it allows to simplify the notation as we can omit the index for the k-th attribute and the summation over alternatives available to person n (n =1, ..., N). Under these two assumptions, equation (5) simplifies to:

$$E\left[\frac{\partial^2 \ell}{\partial \beta^2}\right] \cong \sum_{n=1}^{N} x_{jn} \left[\frac{\partial\left(y_{jn} - p_{jn}(\mathbf{x}_{jn};\boldsymbol{\beta})\right)}{\partial \beta}\right]_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \tag{6}$$

If the parameter is specified generic between both alternatives, such that $\Delta x_n = (x_{jn} - x_{in})$, then the variance of the estimated parameter is:

$$\text{var}(\hat{\beta}) = -\frac{1}{\sum_n \Delta x_n^2 \hat{p}_{jn}(1-\hat{p}_{jn})} \tag{7}$$

where $\hat{p}_{jn}$ is the probability that individual n chooses the alternative which is actually chosen (this is why y_{jn} does not appear), calculated at the point of the estimated parameter ($\beta = \hat{\beta}$).

Equation (7) clearly shows why the efficiency of the ML estimator increases as the sample (*N*) and/or the data variability (Δx_n) get bigger. In the RP data case the degree of variability is often related to the number of observations because it cannot be controlled by the modeller; in the SP data, instead, the number of choice tasks faced by each individual and the attribute ranges matter. In fact, if the attribute levels are defined as a percentage variation of the actual values ($\Delta x_{nt} = \delta_t x_{jn}$), the variance can be written as:

$$\mathrm{var}(\hat{\beta}) = -\frac{1}{\sum_n x_{jn}^2 \sum_t \delta_t^2 \hat{p}_{jnt}(1-\hat{p}_{jnt})} \tag{8}$$

while for absolute attribute levels ($\Delta x_{nt} = \Delta x_t \quad \forall n$), the variance can be written as:

$$\mathrm{var}(\hat{\beta}) = -\frac{1}{N\sum_t \Delta x_t^2 \hat{p}_{jt}(1-\hat{p}_{jt})} \tag{9}$$

However, the previous equations show that even for the simple binary logit the variance depends not only on the data variability and number of observations, but also on a second order function of the probabilities. In particular, this function is such that as the variability of the data and the value of the scale factor increases, the efficiency of the parameter decreases. This is because the logit probability depends, among other things, also on the data variability and on the variance of the error term (through the scale factor).

Figure 1 was drawn using equation (7) and shows how the efficiency of the parameter varies with the variability of the attribute and with sample size. In particular, the graph shows that the data variability does not always increase the efficiency of the estimate; rather, in some cases it might not be beneficial to increase data variability or it might even be better to have a smaller range of variation. On the contrary, efficiency clearly improves with sample size. Figure 2 is also based on equation (7) but considers the effect of the scale factor explicitly on the estimated parameter in the estimated probability ($\hat{\beta} = \lambda\beta$).

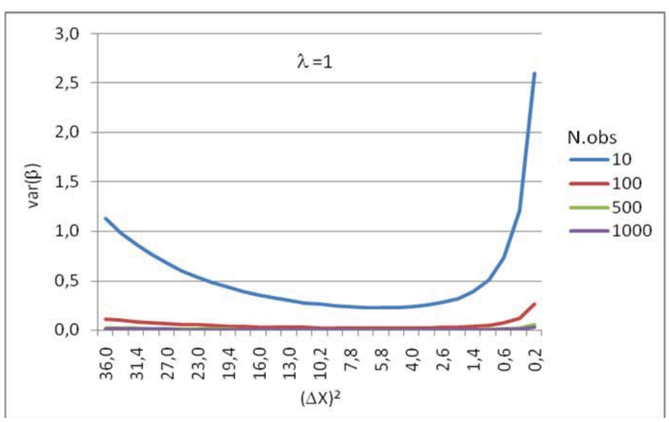

Figure 1: Efficiency of the estimate (for λ =1)

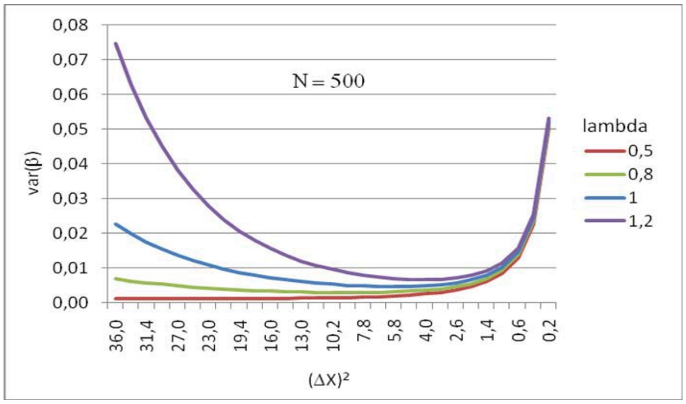

Figure 2: Efficiency of the estimates (for N = 500)

Figure 2 shows that the efficiency of the parameter increases with the variability of the attribute but only for scale factors over 0.5. This effect, that might seem counterintuitive, is due to the effect that the scale factor has on the variability of the data, because efficiency reduces as data variability diminishes; and is also due the second order function of the probability that tends to zero as the probability of the chosen alternative approximates one.

The extension of these analyses to the multinomial case is not difficult. The variance of the parameters estimated in the MNL can be derived from equation (5), computing the second derivatives and rearranging terms:

$$E\left[\frac{\partial^2 \ell}{\partial \beta_k^2}\right] \cong -\sum_{n=1}^{N}\left[\sum_{j\in I_n}\sum_{i>j} p_{jn} p_{in} (x_{jnk} - x_{ink})^2\right]_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \tag{10}$$

Note that the terms inside the square brackets are all positive; thus the efficiency of the estimates increases with N and with the variability in the data. The effect of the probability is the same as discussed previously.

2.2. Efficiency in the Mixed MNL

The mixed logit model (Train, 2003) is characterized by a choice probability that can be expressed as the integral of standard logit probabilities, evaluated at parameters α, over a density of parameters:

$$p_{jn} = \int \frac{e^{\beta_{jn} L_{jn}}}{\sum_{i\in I_n} e^{\beta_{in} L_{in}}} f(\beta_{jn} | \Omega) d\beta \tag{11}$$

where Ω are the population parameters of the distribution, L_{jn} is any transformation of the relevant attributes for individual n and alternative j, and p_{jn} is the probability of individual n choosing alternative j among the alternatives in her choice set (I_n). When a mixed logit is estimated, the covariance matrix is still calculated as the inverse of the matrix of second derivatives of the log-likelihood function, but the probabilities are now simulated probabilities (sp):

$$\operatorname{var}(\hat{\beta}) = -1 / \left.\frac{\partial^2 sp}{\partial \beta^2}\right|_{\beta=\hat{\beta}} = -1 / \frac{\partial^2}{\partial \beta^2} \left. \left(\sum_n \frac{1}{R} \sum_{r=1}^{R} p_{jn}^{(r)} \right) \right|_{\beta=\hat{\beta}^{(r)}} \tag{12}$$

where R is the number of draws used to calculate the simulated probability. Equation (12) shows the expression of the covariance matrix in the mixed logit model with generic parameters. As can be seen, the variance of the mean of the random parameter is more complex than those reported in equations (7)-(9) but the structure is analogous:

$$\operatorname{var}(\hat{\beta}) = \frac{-1}{\sum_n \Delta x_n^2 \left(2 \sum_{r=1,R} (\hat{p}_{jn}^{(r)})^2 (1-\hat{p}_{jn}^{(r)}) \sum_{r=1,R} \hat{p}_{jn}^{(r)} - \sum_{r=1,R} (\hat{p}_{jn}^{(r)})^2 \sum_{r=1,R} \hat{p}_{jn}^{(r)} (1-\hat{p}_{jn}^{(r)}) \right) / \left(\sum_{r=1,R} \hat{p}_{jn}^{(r)} \right)^2} \tag{13}$$

Note that it is also inversely related to the square value of the attributes associated to each parameter (as in the case of fixed parameters) and is also a function of the probabilities.

A major problem in the estimation of the mixed logit model relates to the log transformation of the simulated probabilities: in fact, while the log-likelihood function is unbiased in the simple (binary or multinomial) logit model, in the mixed logit model the estimates are biased due to the log transformation of the simulated probability (Börsch-Supan and Hajivassiliou, 1993). As illustrated by Walker (2001), the error in the simulated log-likelihood can be calculated making a second order expansion of the simulated probability (sp) around the true probability (p):

$$\ln(sp) \cong \ln(p) + \frac{1}{p}(sp - p) - \frac{1}{2(sp)^2}(sp - p)^2 \tag{14}$$

Calculating the expected value of equation (14), we can get the error in the simulated log-likelihood ($s\ell$) as:

$$s\ell - \ell \cong -\frac{\operatorname{var}(sp)}{2p^2} \tag{15}$$

This is the main reason why a sufficiently large number of draws is required to have efficient estimates in a mixed logit model. As reported by Walker (2001), to minimize the bias in the simulated log-likelihood the choice probabilities must be simulated with precision and this increases with the number of draws. At the same time, as shown in equations (7)-(10), the data (sample size and attribute variation) are determinant in increasing the efficiency of the estimated parameters.

3. SIMULATED EXPERIMENTS

Following Williams and Ortúzar (1982), a collection of datasets were simulated in which pseudo-observed individuals behaved according to a choice process determined by the analyst. Simple samples, with only two alternatives, two generic attributes (travel time and cost) and a Gumbel error (ε_q), were generated, but the marginal utility of travel time (β_q) was varied such that the generated sample showed random heterogeneity in tastes. In all experiments the attributes and travel time parameters were generated according to a censored Normal distribution to avoid mass points on the truncations that can induce estimation problems (Cherchi and Polak, 2005).

The datasets were generated according to the following utility functions:

$$U_{nt1} - U_{nt2} = -\beta_n(Time_{nt1} - Time_{nt2}) - 1.0(Cost_{n1} - Cost_{n2}) + \varepsilon_{nt1} - \varepsilon_{nt2} \tag{16}$$

In the first experiment we assumed that individuals n evaluate only one choice task (t =1), as in a RP data set. We first generated the vectors of Gumbel errors and of cost differences between both alternatives (with mean equal to -4.5, standard deviation equal to 2.45, and lower and upper limits equal to [–10; –0.1]). Then, keeping these vectors fixed, several samples of 2,000 observations were generated varying the standard deviation of the distribution of travel time differences between both alternatives, as illustrated in Table 1. In this way we were able to control for the richness of information in the data (in this case the travel time attribute) and its effect on the efficiency of the estimated parameters. To avoid results being dependent on a particular case, each sample was generated several times (following Train, 2003, up to 50 repetitions were generated[2]) with different seeds.

Table 1: Samples generated varying travel time attributes

Samples	*$Time_1 - Time_2$*			
	mean	**standard deviation**	**coefficient of variation**	**limits**
RP1	4.0	0.31	0.08	[2.9; 5.1]
RP2	4.0	0.44	0.11	[2.5; 5.6]
RP3	4.0	0.62	0.16	[1.9; 6.2]
RP4	4.0	0.76	0.19	[1.4; 6.7]
RP5	4.0	0.98	0.25	[0.6; 7.5]
RP6	4.0	1.37	0.34	[0.2; 9.0]
RP7	4.1	1.86	0.45	[0.1; 9.5]
RP8	4.3	2.27	0.52	[0.1; 9.9]

In the second experiment, instead, we assumed that each individual evaluated more than one choice situation, as in SP experiments. A Monte Carlo analysis was set out to generate a collection of samples that reproduced SP choices. Samples were generated varying - once at a time - the number of choice tasks, the number of individuals and the attributes' ranges. The choice experiments were generated varying the travel time between both options around a value previously generated for each individual, and leaving the rest of the utility functions unchanged. In particular, both the cost and travel time parameters were kept fixed among the choice tasks and equal to the value generated for sample A1, while a new Gumbel error vector was generated for this experiment that varied among choice tasks; sample A1 was used as the base sample.

Thus, in the first dataset (SP1), 25 individuals were considered, each evaluating only two choice

[2] In some cases model results were extremely stable, thus less than 50 repetitions were used.

tasks (the final sample had 50 pseudo-individuals), one equal to the base sample (i.e. the value generated in sample A1) and another where the travel time differences between both alternatives for each individual were increased by 5%. In the second set (SP2), the choice situations were again only two, the base sample and a second one where the travel time differences were increased by 10%, and so on. The characteristics of the samples generated are illustrated in Table 2.

Table 2: Samples generated varying sample size and number of choice tasks

Samples	**No. of choice tasks**	**Sample size** (Pseudo-individuals)	**No. of individuals**	**% variation**
SP1	2	50	25	(=, 5%)
SP2	2	50	25	(=, 10%)
SP3	2	100	50	(=, 5%)
SP4	2	100	50	(=, 20%)
SP5	2	100	50	(=, 50%)
SP6	2	100	50	(=, 10%)
SP7	2	200	100	(=, 5%)
SP8	2	200	100	(=, 10%)
SP9	2	200	100	(=, 50%)
SP10	2	1,000	500	(=, 50%)
SP11	2	2,000	1,000	(=, 5%)
SP12	2	2,000	1,000	(=, 50%)
SP13	4	200	50	(=, 5%, 10%, 15%)
SP14	4	400	100	(=, 5%, 10%, 15%)
SP15	4	1,000	250	(=, 5%, 10%, 15%)
SP16	4	2,000	500	(=, 5%, 10%, 15%)

For each dataset discussed above, and for each repetition of the Monte Carlo experiment, several MMNL models allowing for random heterogeneity in the travel time parameters were estimated. In particular, Table 3 reports the results of the MMNL estimation using the datasets described in Table 1, while Table 4 illustrates the results of models estimated using some of the datasets described in Table 2. All the results are the average estimates over the number of times the data were generated (Monte Carlo repetitions).

The first result worth noting is that, as extensively discussed in Cherchi and Ortúzar (2007), in the MMNL model there is a clear problem in reproducing the true parameter when there is low variability in the data (i.e. in the difference between alternatives). However this effect mainly occurs when each individual provides only one piece of information (as in RP data), while it almost disappears in presence of repeated observations (as in SP data). In fact, results in Table 3 show that only with the samples generated with CV = 0.19 or greater (models RP4-RP8), the MMNL estimates were always empirically identified whatever seed was used. On the contrary, models illustrated in Table 4 show that the occurrence of poor results was extremely low and only happened for the dataset generated with two choice tasks and 50 individuals or less.

Table 3: Model estimation results: travel time parameter β_q ~ N(-0.9, 0.39)

	RP1	RP2	RP3	RP4	RP5	RP6	RP7	RP8
Sample CV(Tv_1-Tv_2)	0.08	0.11	0.16	0.19	0.25	0.34	0.45	0.52
Travel time (mean)	-0.4006	-0.5473	-0.5861	-0.609	-0.623	-1.6905	-0.9108	-0.7255
t-test vs 0	*(-2.34)*	*(-1.24)*	*(-4.53)*	*(-4.73)*	*(-5.39)*	*(-3.80)*	*(-5.40)*	*(-6.74)*
t-test vs "true" value	*-2.92*	*-0.80*	*-2.43*	*-2.26*	*-2.40*	*1.78*	*0.06*	*-1.62*
lambda	**0.45**	**0.61**	**0.65**	**0.68**	**0.69**	**1.88**	**1.01**	**0.81**
Travel time (st.dev.)	-0.0162	0.7778	0.2073	0.2645	0.3098	1.3176	0.7426	0.4752
t-test vs 0	*(-0.05)*	*(-0.99)*	*(-1.26)*	*(-1.55)*	*(-2.15)*	*(-3.14)*	*(3.56)*	*(3.54)*
t-test vs "true" value	*-1.15*	*-0.49*	*1.11*	*0.74*	*0.56*	*-2.21*	*-1.69*	*-0.63*
lambda	**0.04**	**1.99**	**0.53**	**0.68**	**0.79**	**3.38**	**1.90**	**1.22**
Travel cost	-0.5499	-1.1878	-0.6292	-0.6822	-0.7041	-1.8471	-1.1669	-0.8926
t-test vs 0	*(-26.09)*	*(-1.31)*	*(-5.79)*	*(-5.13)*	*(-5.80)*	*(-3.60)*	*(-5.10)*	*(-6.22)*
t-test vs "true" value	*-21.35*	*0.21*	*-3.41*	*-2.39*	*-2.46*	*1.65*	*-0.73*	*0.75*
lambda	**0.55**	**1.19**	**0.63**	**0.68**	**0.70**	**1.85**	**1.17**	**0.89**
Constant (alt. 1)	-0.2277	-1.7146	0.3035	0.2296	0.2366	0.3705	-0.1532	0.2251
t-test vs 0	*(-0.33)*	*(-0.91)*	*(-0.76)*	*(-0.65)*	*(-0.82)*	*(-0.67)*	*(-0.51)*	*(1.01)*
t-test vs "true" value	*-0.33*	*-0.91*	*-0.76*	*-0.65*	*-0.82*	*-0.67*	*0.51*	*-1.01*
$l(\theta)$	-1049.96	-1034.22	-1031.54	-1018.41	-1017.12	-976.93	961.43	-935.66
No. of draws	125-2000	125-2000	1000-2000	2000	2000	2000	2000	2000
Sample size	2000	2000	2000	2000	2000	2000	2000	2000

Table 4: Model estimation results: travel time parameter: CV(Tv_1-Tv_2) = 0.08

	SP7	SP9	SP11	SP13	SP14	SP16
No. of choice tasks	2	2	2	4	4	4
No. of individuals	100	100	1000	50	100	500
% variation	5%	50%	5%	5-10-15%	5-10-15%	5-10-15%
Travel time (mean)	-1.3324	-1.2282	-1.1626	-1.6235	-1.1378	-1.2399
t-test vs 0	*(-3.05)*	*(-3.42)*	*(-9.99)*	*(-2.66)*	*(-3.86)*	*(-9.72)*
t-test vs "true" value	*-0.052*	*0.004*	*0.507*	*0.034*	*-0.292*	*0.217*
lambda	**1.51**	**1.42**	**1.30**	**1.77**	**1.38**	**1.37**
Travel time (st.dev.)	0.9676	0.8042	0.8759	1.0692	0.8587	0.8756
t-test vs 0	*(3.32)*	*(3.28)*	*(10.87)*	*(3.52)*	*(5.46)*	*(11.86)*
t-test vs "true" value	*0.426*	*-0.003*	*0.633*	*0.019*	*0.896*	*0.397*
lambda	**1.68**	**1.42**	**1.49**	**1.80**	**1.61**	**1.50**
Travel cost	-1.2677	-1.4262	-1.3561	-1.8072	-0.8107	-1.3821
t-test vs 0	*(-2.85)*	*(-3.64)*	*(-11.28)*	*(-3.05)*	*(-2.75)*	*(-10.91)*
t-test vs "true" value	*0.525*	*-0.008*	*0.193*	*-0.022*	*1.426*	*0.170*
lambda	**1.26**	**1.42**	**1.35**	**1.80**	**0.81**	**1.38**
$l(\theta)$	-88.9871	-92.7012	-897.4769	-77.8909	-159.9737	-778.2788
No. of draws	1000	1000	1000	1000	1000	1000
Sample size	200	200	2000	200	400	2000

If we look at the t-tests against zero it is clear, and not surprising, that the significance of the standard deviation (and thus the efficiency) of the random travel time parameter increases with information richness (i.e. the difference between the attributes in both alternatives). However, efficiency does not guarantee unbiasedness; in fact, all models reported in Table 3 show significant differences in the scale factors (*lambda*) among estimated attributes, which implies that some parameter is not estimated correctly.

As discussed in Section 2, a major problem with the MMNL model is that the simulated probability is biased and a sufficiently large number of draws is required to have efficient estimates. All models were estimated varying the number of draws from 1 up to 10,000 and we found that for over approximately 100 draws, results were highly stable. Moreover, all models reproduced correctly a highly significant random heterogeneity in travel time, even for only two choice tasks per individual and also for very few individuals (only 100). At the same time, as shown in equations (8)-(10), the data (sample size and attributes variation) are determinant in increasing the efficiency of the estimated parameters. Comparing the models in Table 4 with model RP1 in Table 3, it is clear that the MMNL estimates improve (low variance of the estimated parameters) with the number of choice tasks performed by each individual, even when the number of individuals is low (see model SP13 in Table 4).

4. EFFICIENCY OF THE ESTIMATES

To test the efficiency of the MMNL estimates we computed the measure of asymptotic efficiency (MAE) proposed by Bliemer and Rose (2004):

$$MAE = \max \left| \mu(\hat{\beta}_{jn}) - \beta_j^{true} \pm \sigma(\hat{\beta}_{jn}) \right| / \beta_j^{true} \tag{17}$$

where:

$$\mu(\hat{\beta}_{jn}) = \frac{1}{R} \sum_{r=1,R} \hat{\beta}_{jn}^{(r)}$$

$$\sigma(\hat{\beta}_{jn}) = \sqrt{\frac{\sum_{r=1,R} (\hat{\beta}_{jn}^{(r)} - \mu(\hat{\beta}_{jn}))^2}{R-2}} \tag{18}$$

are, respectively, the mean and standard deviation of the estimated parameters ($\hat{\beta}_{qj}$). The estimator is asymptotically unbiased if, in the limit, its mean is equal to the true mean and its standard deviation equals zero. This test allows verifying whether the estimated means and standard errors are close enough to the true parameters.

As discussed in Section 2 all parameters estimated in a discrete choice model are scaled by an unknown factor (λ) proportional to the inverse of the standard deviation. It does not matter the absolute value of the scale but that the estimated parameters are deflected by the same scale value. Thus, when comparing the estimated values with the true values in equations (17) and (18), the MAE measure can be influenced by the scale factor of the estimated parameters. Therefore, we also computed a scaled MAE measure, where at each iteration the ratio $\hat{\beta}_{jn}^{(r)} / \lambda^{(r)}$ was used instead of the exact estimated parameter $\hat{\beta}_{jn}^{(r)}$.

Table 5 reports the results on asymptotic efficiency for some of the models estimated with single and repeated observations. First of all it is interesting to note that the MAE (not scaled) measure is always quite high, due to the effect of the scale factors implicit in the estimated parameters. Note also that model RP8, which is the best one under the usual statistics applied in practice and the one with the less biased estimates also, has correctly a much smaller MAE (also scaled) than model RP1, although the MAE measure for the standard deviation estimated with model RP8 is not too small. At the same time, as expected, the scaled MAE measure for the repeated observation (SP) models is good and better than that computed for the RP data (note that only model RP1 can be compared with the SP results, as the SP data were generated as variations of dataset RP1).

Table 5: Model asymptotic efficiency

	MAE					*MAE scaled*				
	RP1	**RP8**	**SP7**	**SP11**	**SP16**	**RP1**	**RP8**	**SP7**	**SP11**	**SP16**
Travel time (mean)	0.5469	0.2470	0.7939	0.3815	0.4627	0.4237	0.1658	0.0849	0.0815	0.0749
Travel time (st.dev.)	1.1081	0.4855	1.0295	0.5670	0.6025	1.2147	0.2081	0.0857	0.0992	0.1043
Travel cost	0.4721	0.2907	0.5885	0.4130	0.4800	0.5872	0.1228	0.0217	0.0315	0.0457

5. CONCLUSIONS

We analysed the effect of data information richness on the efficiency of estimated discrete choice parameters trying to disentangle the several effects that influence their efficiency in the MNL and MMNL models. Efficiency is inversely related to the attributes associated to the parameters and to the number of observations; but, it also depends on a second order function of the probabilities, in a way that as the variability of the data and the scale factor increases the efficiency of the parameter decreases.

Using simulated data, we found that the efficiency of the parameters increases with the variability of the attributes but only for scale factors bigger than 0.5. This effect, that might seem counterintuitive, is actually due to the effect that the scale factor has on the variability of the data, because efficiency reduces as data variability diminishes; it is also due the second order function of the probability that tends to zero as the probability of the chosen alternative approximates one. We also found that efficiency clearly improves with sample size. However, the variability of the data does not always increase the efficiency of the estimates, rather, in some cases it might not be beneficial to increase data variability or it might even be better to have a small range of variation. The problem of efficiency with size is crucial if we need to gather RP data or to build a SP experiment. In particular, we found that the efficiency of the estimated parameter improves significantly when models are estimated using data sets with repeated observations, even for data sets of small size.

ACKNOWLEDGEMENTS

The second author is grateful to the Chilean Fund for Scientific and Technological Research (FONDECYT), through Project 1050672, and to the Ministry of Planning's Millennium Scientific Initiative, through the Institute on Complex Engineering Systems, for having partially funded this research

REFERENCES

Bliemer, M.C.J. and J.M. Rose (2004) Sample size requirements for estimating multinomial logit models using stated choice experiments. **Working Paper**, Institute of Transport and Logistics Studies, University of Sydney.

Bliemer, M.C.J. and J.M. Rose (2005) Efficiency and sample size requirements for stated choice studies. **Transportation Research B** (submitted).

Börsch-Supan, A. and V. Hajivassiliou (1993). Smooth unbiased multivariate probability simulators for maximum likelihood estimation of limited dependent variable models. **Journal of Econometrics** 58, 347-368.

Cherchi, E. and J.de D. Ortúzar (2007) Empirical identification in the mixed logit model: analysing the effect of data richness. **Network and Spatial Economics** (in press).

Cherchi, E. and J.W. Polak (2005) The assessment of user benefits using discrete choice models: implications of specification errors under random taste heterogeneity. **Transportation Research Record 1926**, 61-69.

Chiou, L. and J. Walker (2007) Masking identification in the discrete choice model under simulation methods. **Journal of Econometrics** (in press)

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

Walker, J. (2001) **Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures, and Latent Variables.** PhD Thesis, Department of Civil and Environmental Engineering, MIT.

Williams, H.C.W.L. and J. de D. Ortúzar (1982) Behavioural theories of dispersion and the mis-specification of travel demand models. **Transportation Research** 16B, 167-219.