Variations in choice sets and identification of mixed logit models: Monte Carlo evidence

Sebastien Lecocq

Unité de recherche 1303
alimentation et sciences sociales

Variations in Choice Sets and Empirical Identification of Mixed Logit Models: Monte Carlo Evidence

LECOCQ SÉBASTIEN

AOUT 2008

Working Paper ALISS 2008-08

INRA UR 1303 ALISS
65, Bd de Brandebourg
94205 Ivry-sur–Seine Cedex
France

http://www.paris.inra.fr/aliss

Variations in Choice Sets and Empirical Identification of Mixed Logit Models: Monte Carlo Evidence*

LECOCQ Sébastien[1]

[1] *INRA, UR1303 ALISS, F-94200 Ivry-sur-Seine, France*
Email: lecocq@ivry.inra.fr

Abstract: Empirical identification of Mixed Logit models is known to depend on the richness of the data in terms of variations in the explanatory variables. In this paper, we wonder whether choice set variations observed in scanner-based consumer surveys are sufficient to enable identification. Using Monte Carlo experiments, we show that when a random parameter is applied to a binary variable, Mixed Logit models are identified only if the number of options varies across a sufficient number of individuals and/or choice situations. Conversely, when a random parameter is applied to a continuous variable, models are identified without any choice set variation.

Keywords: MIXED LOGIT MODEL, CHOICE SET VARIATION, EMPIRICAL IDENTIFICATION, MONTE CARLO EXPERIMENT

JEL CODES: C15, C25, C81

* I would like to thank F. Etilé and T. Magnac for their comments and suggestions.

1 Introduction

Mixed Multi-Nomial Logit (or Mixed Logit for short) models are highly flexible discrete choice models that have been developed to allow for unobserved heterogeneity in taste parameters and to generate unrestricted substitution patterns.[1] Basically, they have the form of a standard Multi-Nomial Logit (MNL) model, where choices are explained by projecting consumer preferences onto a set of attributes describing the available options. However, in contrast to MNL models whose parameters do not vary across individuals, their parameters are allowed to vary randomly in the population according to a given distribution. Though Mixed Logit models have been known for quite a long time, they have been more extensively used in empirical work over the last decade only, thanks to the advent of simulation techniques and to faster computers which have eased the burden of estimation (due to high dimensional integrals appearing in choice probabilities). They have been proved an attractive tool for addressing economic questions such as the measurement of market power, the identification of the nature of competition, the simulation of the impact of a new product introduction or a new pricing strategy (see Dubé et al. 2002).

Applications of Mixed Logit models can be distinguished according to the type of data used for estimation. In economics, most applications rely on data that are either aggregated at the market level (at least to some extent) or collected through stated choice experiments. Market level data can mainly be found in the recent empirical industrial organization literature, including Berry et al. (1995) for the automobile industry, Nevo (2001) for the ready-to-eat cereal industry, Davis (2006) for movie theatres, to name a few. In many respects, the methodology required to estimate Mixed Logit models on aggregate data is quite different from the one that has to be implemented in the case of individual data (for an overview, see Nevo 2000). The latter being the focus of the paper, market level data are not discussed further on in what follows.

Stated choice data are individual data generated through experimental designs. Preferences are recovered by constructing hypothetical choice situations, where each situation consists of more than two options from which a respondent is asked to choose. Examples of papers using stated choice data to estimate a Mixed Logit model are Brownstone and Train (1998) on alternative-fuel vehicles, Revelt and Train (1998) on appliance efficiency level, or Rigby and Burton (2005) on genetically modified food. A major advantage of these data is that they contain sufficient variations in the attributes of the options over experiments and/or respondents to guarantee the identification of the underlying preference parameters. A well-known shortcoming however is that they are based on choice situations that are hypothetical, the respondent facing none of the constraints he would face in a real-life situation.[2]

[1] McFadden and Train (2000) show, under some mild regularity conditions, that any discrete choice model derived from random utility maximisation has choice probabilities that can be approximated arbitrarily closely by a Mixed Logit model.

[2] Constructing stated preference experiments from choices made by respondents in real preference settings can nevertheless enhance their realism (see Hensher and Greene 2003, Train and Wilson 2008, and references therein).

Individual data, or more precisely household data, gathered from the observation of real choice situations have also been considered in applications of Mixed Logit models: many in marketing (Allenby et al. 1998, Allenby and Lenk 1995, and Allenby and Rossi 1999 on ketchup, tuna and detergents), and some in economics (Bonnet and Simioni 2001 on camembert, Bjørner et al. 2004 on toilet paper, paper towels and detergents, or Athey and Imbens 2007 on yogurt). Such revealed choice (or survey) data provide information on the choices made by individuals in actual situations, such as the choice of a brand in a store under a budget constraint for example. However, the richness of the information they provide, in terms of variability, is generally lower than in stated choice data. In particular, revealed choice data raise a number of issues related both to the set of options faced by an individual at the time of his choice and to the variations in the attributes of the options. A good illustration of these issues can be given by considering the case of a scanner-based consumer survey.

Typically, a consumer scanner dataset provides, for each purchased product, information on the unit price and some characteristics of the chosen brand. However, it does not give any information on the brands that were on the shelves at the same time but that were not purchased. In other words, the consumer's choice set is unknown. It may also vary over time, as new brands may emerge while others may be removed from the market, and it may vary across stores, as some brands may be found almost everywhere while others may only be available in one chain of supermarkets (store or retailer brands). To get round that first issue, most papers suppose that the set of brands faced by an individual at the time of his choice is composed of all the brands that were purchased (and registered in the data) in the same week in the same chain of stores. This is for instance the assumption retained in the three applications mentioned above. These papers slightly differ in the estimation strategy though. Bonnet and Simioni (2001) estimate the model for each store separately, hence constraining the choice set to be identical for all individuals in a given week and, therefore, to vary merely over weeks. In contrast, Athey and Imbens (2007) and Bjørner et al. (2004) estimate the model on the purchases pooled over both weeks and stores, in order to obtain a choice set that varies not only across time but also across individuals (through stores).

A second issue is related to the fact that there are many product characteristics that never change, though the price usually does (due to promotions or store pricing strategy for instance) and that it might be the case for some other particular attributes (generally considered in the marketing literature, such as "is on promotion", "is featured in advertising", etc.). For a given option indeed, neither the brand nor the information mentioned on the package or label vary through time and stores. In many consumer data, the only variations in the choice set that we can expect to observe are therefore variations from one week to another and/or from one store to another either in the price of each brand, or in the number of distinct brands constituting the choice set, that is the set of brands purchased in the same week in the same store by the sampled individuals, or both.

While the number of papers involving the estimation of a Mixed Logit model is constantly increasing, identification issues have largely been overlooked in the literature. On the theoretical side, no formal proof or rule of identification has been given until very recently. Walker (2002) and Walker et al. (2007) establish rules that can be applied before estimation, looking at the covariance matrix of utility differences, and using a new equality condition, in addition to the standard order and rank conditions, to determine theoretical identification. Bajari et al. (2007) are the first to provide a theoretical proof showing that the random coefficients in a Mixed Logit model are nonparametrically identified. Their result however requires all characteristics to be continuous: the characteristics only need to vary locally but cannot be discrete. Furthermore, the set of alternatives they consider varies across all observations (as in stated choice data), and they do not show how their result would be affected if that assumption was released.

On the empirical side, several papers by Walker (2002), Hensher and Greene (2003), Walker et al. (2007), Chiou and Walker (2007), and Cherchi and Ortúzar (2008) focus on identification "in practice", arguing that even when a model is shown to be **theoretically** identified (through a mathematical result), it may not be **empirically** identified. All agree that empirical identification of Mixed Logit models relies on the richness of the data in terms of variations in the explanatory variables, and that care must be taken about asking more from the data than they can actually provide. In most revealed choice data yet, variations are by construction much less important than in stated choice data, where choice situations are voluntarily designed to differ one another as much as possible (Sándor and Wedel 2002), considering various sets of options and various combinations of attributes for different situations and different individuals. Whether variations found in typical survey data are sufficient to empirically identify a Mixed Logit model is the point that we want to address in this paper.

From a methodological point of view, our paper is entirely based on Monte Carlo simulations and, in that respect, is closely connected to those by Chiou and Walker (2007), and especially Cherchi and Ortúzar (2008).[3] Chiou and Walker (2007) use both actual and synthetic datasets to show how simulation methods can mask identification problems in the estimation of discrete choice models such as Mixed Logits, which may then appear as identified even though they are theoretically unidentified. Cherchi and Ortúzar (2008) investigate the effect of data information richness on empirical identification of Mixed (binomial) Logit models. Focusing on continuous characteristics and choice sets varying over all observations, they find that the richness of the data does matter and, in particular, that identification problems arise when the characteristic associated to the random parameter has a low variability between alternatives. They also show that observing more than one choice per individual (panel data) makes empirical identification easier and strongly reduces the effect of sample size.

[3] See also Andrews and Currim (2005) who conduct an extensive simulation experiment to investigate how scanner data preparation strategies (data pruning and entity aggregation) can impact Mixed Logit estimates of price and promotion responses.

In the present paper, Monte Carlo experiments are used to answer another set of questions, though some are similar, all related to variations that may be observed (or not) in the choice set from one consumer to another and/or from one period to another: can we identify a Mixed (multinomial) Logit model when the choices recorded in the data are made by individuals all facing a same set of alternatives or do we need this set to be different across periods and/or, to some extent, across individuals? Basically, each experiment considers a set of ten alternatives all differentiated by two characteristics: one is represented by a continuous variable, taking different values for different options (a price for example), and the other by a discrete (binary) variable, equal to one for a subset of options and to zero for the complement (organic or not, sugar-free or not, etc.). For a given option, the binary attribute is designed in all experiments to be the same for all consumers in all choice situations; in contrast, the continuous attribute may or may not, depending on the experiment, differ across individuals, choice situations, or both. Similarly, the number of alternatives in the choice set is also allowed to vary (or not) across consumers and/or choice situations. The impact of these variations, in the continuous variable and in the number of options, on empirical identification is then investigated.

The paper is organised as follows. Mixed Logit models, together with the main identification issues related to their estimation by simulation methods, are presented in section 2. Section 3 describes the Monte Carlo design. Section 4 reports the results and discusses their implications for real consumer data. Section 5 concludes.

2 Mixed Logit Models

Denote J_{it} the choice set faced by consumer $i = 1, ..., N$ in choice situation $t = 1, ..., T$. A choice situation here can refer to a time period, a market, the combination of the two, or a situation in a stated choice experiment. Let x_{ijt} be the vector of characteristics, price included, observable by the econometrician and describing the alternative $j \in J_{it}$. Each choice set contains an outside (or no purchase) option, denoted $j = 0$ by convention and whose observed attributes are all set to zero. Assuming linearity in its arguments, the utility that individual i obtains from the consumption of one unit of j in period t can be written as

$$u_{ijt} = \beta_i' x_{ijt} + \varepsilon_{ijt}, \tag{2.1}$$

where β_i is the vector gathering together the tastes of consumer i for each observed characteristic, and ε_{ijt} an error term representing his unobserved (by the econometrician) idiosyncratic taste for alternative j in choice situation t, which is assumed to be zero mean and independent from the attributes of j. The tastes for the observed characteristics are supposed to vary in the population in the following way

$$\beta_i = \overline{\beta} + \Sigma\nu_i, \tag{2.2}$$

where $\overline{\beta}$ is the vector of average tastes in the population, ν_i the vector of unobserved tastes of consumer i, and Σ a diagonal matrix of parameters, denoted σ and interpreted as deviations from average tastes.[4]

Define y_{ijt} an indicator variable equal to 1 if individual i chooses alternative j at time t and to 0 otherwise. Each consumer being supposed to choose the option that maximises his utility and further assuming that ties occur with zero probability, the choice criterion is

$$\begin{aligned} y_{ijt} &= 1 \quad \text{if} \quad u_{ijt} > u_{ikt} \quad \forall j \neq k, \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

Response probabilities are then obtained by summing the choices implied by the individual utility model on the distribution in the population of interest of the consumers' unobserved tastes, ν_i and ε_{ijt}. Suppose that ν_i and ε_{ijt} are independent, that ε_{ijt} is i.i.d. with a type I extreme value distribution, and that the elements of ν_i are independently and normally distributed with zero mean and unit variance. Under the additional assumption that consumer's heterogeneity enters the utility function only through the additive error term ε_{ijt}, that is $\nu_i = 0$, the model reduces to the standard MNL model. Although the MNL model is very attractive because of its extreme tractability, it restricts substitution patterns in an unreasonable fashion. Indeed, it is well-known that the MNL model satisfies the Independence of Irrelevant Alternatives (IIA) property, which states that the ratio of choice probabilities for any pair of options is independent from the existence and characteristics of all other options. In other words, introducing a new alternative or modifying the attributes of an existing alternative will change the probabilities of all other options proportionately, keeping the ratios of choice probabilities unchanged.

Consider the hypothetical situation where ν_i would be different from zero but observed. In this case, the above model would simply reduce to the MNL formulation with choice probabilities given by

$$\Lambda(y_{ijt} = 1 \mid x_{ijt}, \nu_i; \theta) = \frac{\exp(\overline{\beta}' x_{ijt} + \nu_i' \Sigma x_{ijt})}{1 + \sum_{k \in J_{it}, k \neq 0} \exp(\overline{\beta}' x_{ikt} + \nu_i' \Sigma x_{ikt})}, \tag{2.3}$$

where $\theta = \text{vec}(\overline{\beta}, \text{diag}(\Sigma))$ is the vector of parameters to be estimated, $\Lambda(y_{ijt} = 1 \mid x_{ijt}, \nu_i; \theta)$ is the probability that alternative j is chosen by individual i at time t conditional on x_{ijt}, ν_i and θ, and the utility derived from the consumption of the outside alternative is normalised to zero. Since ν_i is not observed, the probability unconditional on ν_i is

$$P(y_{ijt} = 1 \mid x_{ijt}; \theta) = \int_{\nu} \Lambda(y_{ijt} = 1 \mid x_{ijt}, \nu_i; \theta) f(\nu_i) d\nu_i, \tag{2.4}$$

where $f(\nu_i)$ is the joint density function of ν_i, which by assumption is the product of standard univariate normal densities. Given that each component of ν_i adds a dimension to the integral and that there may be as many components as there are characteristics, it is not possible to compute (2.4) by integrating out over ν_i analytically.

[4] Tastes for characteristics are thus orthogonal one another. This assumption is made to keep the analysis as simple as possible, and correlations could have been considered using a nondiagonal Σ matrix.

Simulation techniques that enable the approximation of high dimensional integrals have to be used. The most common consists in replacing the choice probability by the following unbiased, smooth and tractable simulator

$$\check{P}(y_{ijt} = 1 \mid x_{ijt}; \theta) = \frac{1}{S} \sum_{s=1}^{S} \Lambda(y_{ijt} = 1 \mid x_{ijt}, \nu_{is}; \theta), \quad (2.5)$$

where ν_{is} denotes the s-th draw from the distribution of ν_i. Halton draws are generally preferred to pseudo-random draws as they are known to be more efficient (see Train 1999). Besides, as recommended by McFadden and Train (2000), these draws must be kept fixed during the iterative process that adjust θ. The simulated log-likelihood function can then be written as

$$\check{L}(\theta) = \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{j \in J_{it}} y_{ijt} \ln \check{P}(y_{ijt} = 1 \mid x_{ijt}; \theta). \quad (2.6)$$

Unfortunately, a particular feature of (simulated) log-likelihood functions for Mixed Logit models is that they are not as well-behaved as those for standard MNL models: they are not globally concave and exhibit biases when too low numbers of draws are used in the simulator (Chiou and Walker 2007). The main identification issues pointed out in the literature are often related to that feature.

As pointed out by Walker (2002), most statistical software packages do not report when a model is not identified, because of poor optimisation routine, bad starting values, or too few draws used to simulate probabilities. Results from the estimation of unidentified models are thus unknowingly reported in the literature and used in applications and textbooks (see the couple of examples given in Walker 2002). Yet, there are clues of identification problems: large standard errors indicating a near singular hessian, instability in parameter (and standard error) estimates when different starting values are used in the optimisation process or when the number of draws considered in the simulator is increased. According to Chiou and Walker (2007), the number of draws revealing the identification issue depends on the data, model, and type of draws, and there is no general "rule of thumb" to determine what would constitute a high or low number of draws. These authors distinguish between two types of identification problems: (1) theoretical unidentification, when the model cannot be correctly estimated in principle (regardless the data at hand), usually because it counts too many parameters, and (2) empirical unidentification, when the data cannot support the model even if the model is estimable in principle. In the first situation, they show that low numbers of draws can mask identification issues, which means that a model that is not theoretically identified can appear to result in identified estimates at low numbers of draws.[5] Regarding the second situation, they show that when the number of draws is used to estimate an unidentified model is large enough, the flatness of the likelihood

[5] The fact that identification problems only appear with a sufficient number of draws is confirmed by Cherchi and Ortúzar (2008): for alternatives showing a low variability in their characteristics, estimates are unstable and, depending on the initial seed used to generate the data, the model cannot be estimated unless very frew draws are used.

function is revealed and results either in exploding parameter estimates or in a singular hessian; conversely, when insufficient draws are used, the simulated likelihood function exhibits local concavity, thereby masking identification issues and producing misleading results.

The present paper is part of this literature. Focusing on empirical identification, it builds on Monte Carlo results to provide guidelines about the features the data need to have, in terms of variations in the choice set, to be able to support the estimation of a Mixed Logit model.

3 Monte Carlo Design

A series of 48 experiments is conducted. All experiments enter the following framework: 1000 Monte Carlo replications, 2000 simulated individuals per replication, 5 choice situations per individual, at most 10 options per choice situation, options are differentiated by 2 characteristics and 100 Halton draws are used to approximate choice probabilities;[6] they differ one another by the nature of the variables random coefficients are applied to and the type of variations observed in the choice set.

Specifically, in all experiments, ten alternatives, including an outside option, are constructed from two attributes, represented by a continuous variable (a price in logarithm for instance) and a discrete (binary) variable. These variables are drawn independently, respectively from a normal distribution with zero mean and unit variance,[7] and from a uniform distribution rounded to the nearest integer in order to equal either 0 or 1 (values taken by both characteristics are set to zero for the outside option). As the resulting variables are common to all experiments, they are drawn only once. Below, we first describe the aspects of the design that vary between but not within experiments and then those that vary with each Monte Carlo replication.

Choice sets are defined at the beginning of each experiment and are maintained fixed all along. Two distinct sources of variations in the choice set are investigated: (a) variations in the continuous characteristic of the alternatives, and (b) variations in the number of alternatives.[8] Combined, they lead to four choice set configurations that, for convenience, we denote (a, b), where a and b equal one if the corresponding source of variations is found in the choice set and zero otherwise. As variations can occur both across choice situations (data with repeated choices) and across individuals (stated choice data and, to some extent, survey data where individuals belong to different markets, stores, etc.), the four configurations may be observed in those two dimensions. Let $t(a, b)$ and $i(a, b)$ indicate that the variations implied by configuration (a, b) are observed in the choice situation and individual dimensions, respectively. Then, it is possible to define 16 different choice sets, which can be represented in a table with variations across choice situations in rows and variations across individuals in columns (see table 3 below).

[6] These common values were chosen on the basis of preliminary results detailed in the next section.

[7] Drawing from a normal distribution with non-zero mean does not affect the results.

[8] We do not consider variations in the discrete characteristic, given that the usual information it provides on a product does not vary (brand name, quality label, sugar- or fat-free, etc.).

In the four choice sets appearing in the first row, $t(0,0)$, there is no variation across choice situations: the characteristics and the number of alternatives faced by a given individual are identical from one period to another. Hence, each consumer faces a single and same choice set in all choice situations, and chooses the same option since his preferences are not supposed to change through time; this is equivalent to a setting where consumers are observed only once (cross-section framework). In the first of these first-row choice sets, in column $i(0,0)$, there is no individual variation either: it can be seen as a "universal" configuration where the same ten options are proposed to all consumers in all choice situations. The second choice set, in column $i(1,0)$, still consists of ten alternatives for all consumers but, except the outside option, their continuous attribute now differ across individuals (but not across periods). In other words, all individuals are confronted with ten options, but there are not two of them who face the same ten; think of the same ten yogurts differently priced for each consumer. In the third choice set, in column $i(0,1)$, there is no variation in the characteristics of the alternatives, but the number of alternatives differs from one consumer to another; it does not vary through time. For each individual, we randomly draw a subset of options to be faced (from two to ten elements, the outside and first inside options being always included) from the first (universal) configuration; on average, consumers face six alternatives per choice situation, with a standard deviation of 1.4. Finally, the fourth choice set, in column i $(1,1)$, is identical to the third except that the continuous attribute also varies across individuals (as in the second choice set).

The last three rows of the table introduce variations across choice situations in addition to variations across consumers: the continuous characteristic in row $t(1,0)$, the number of options in row $t(0,1)$, and both in row $t(1,1)$, vary from one period to another; an individual may then choose different alternatives in different choice situations (panel framework). For example, the four choice sets in column $i(0,0)$ do not allow any variation across consumers; in this setting, variations are only observed from one choice situation to another, in the continuous attribute in rows $t(1,0)$ and $t(1,1)$, or in the number of options in rows $t(0,1)$ and $t(1,1)$. At the other extremity, configurations in column $i(1,1)$ allow individual variations in both the continuous characteristic and the number of alternatives. The choice set in row $t(1,0)$ adds variations across choice situations in the continuous attribute; in row $t(0,1)$, the continuous attribute does not vary across periods, but the number of options does; eventually, row $t(1,1)$ denotes a choice set where the two types of variations are present in both individual and choice situation dimensions. As we mentioned above, although all consumers face the same number of choice situations in all experiments, configurations in the first row are equivalent to a cross-section framework whereas those in the last three correspond to a balanced panel framework. These two data structures are distinguished in order to see to what extent observing more than one choice per individual is helpful to recover preference parameters.

Table 1 presents some simple descriptive statistics on the alternatives in the different choice set configurations. In table 1a, the binary variable indicating whether the discrete characteristic is exhibited or not by an option equals one for six options out of ten; this variable is common to all individuals and all choice situations. Similarly, in the first column, the continuous variable of a given alternative does not vary either (it takes only one value), but it does more and more when shifting to the right of the table, varying first with t (5 values), then with i (2000 values), and finally with both t and i (10000 values). In column (2), the mean (standard deviation) of the five values drawn per option ranges from -0.366 (0.564) for option 2 to 0.719 (1.961) for option 5. Of course, the variable being drawn from a standard normal distribution, its mean and standard deviation tend towards zero and one, respectively, as the number of values it takes increases (see columns (3) and (4)). The magnitude of the simulated standard deviations is fairly comparable to what can be found in real consumer data. For instance, on average, price standard deviations are equal to 0.9 and 1.0, depending on the chain of supermarkets, in Bonnet and Simioni (2001), and to 0.7, 1.1 and 0.4 in Bjørner et al. (2004), depending on the product. The probability of belonging to the choice set is shown for each alternative in table 1b. It equals one for the first two options regardless the configuration, and for the eight others in the first column. In the last three columns, these eight other options are included in the choice set with a probability either ranging from 0.2 to 0.8 (second column), or all approaching 0.5 (third and fourth columns). For comparison purposes, the probability reported in Athey and Imbens (2007) ranges from 0.11 to 0.88.

[Table 1 about here]

The continuous/discrete nature of the characteristic whose coefficient is set to be randomly distributed is another source of distinction between experiments (and contrasts with Cherchi and Ortúzar 2008, who only consider continuous characteristics). Three model specifications are considered: taste heterogeneity is placed on the continuous attribute only (β_1), on the discrete attribute only (β_2), and on both. The last specification in particular may help verify a suggestion that sometimes can be found in the literature saying, without any further justification, that it is conventional wisdom to keep one taste nonrandom for identification (for example, see Chiou and Walker 2007).

A sample of 2000 consumers is generated by drawing values of unobservables in the relevant distributions. Draws are specific to individuals and do not vary over time. The unobserved tastes for the observed characteristics, ν, are drawn from a standard univariate normal distribution with zero mean and unit variance, and the unobserved tastes for the alternatives, ε, from a type I extreme value distribution, drawing a vector, η, from the uniform distribution and applying the transformation $\varepsilon = -\ln(-\ln(\eta))$. Given these unobservables, we compute the utility that each consumer gets from each option using (2.1), setting $\overline{\beta} = (-0.2, 0.5)'$ and all diagonal elements of Σ to 0.3,[9] and we assign to each individual the option that gives him the highest satisfaction. The Monte Carlo draw-specific dataset is then complete and can be used to estimate the Mixed Logit model. Operations from the draw of unobservables to the estimation are repeated 1000 times in each experiment.[10]

4 Simulation results

Before turning to simulation results, let us briefly explain how were determined the values that are common to all experiments. During the whole process, we concentrated on the choice set characterized by the largest variations, $t(1,1)$ and $i(1,1)$, where both the number of alternatives and the continuous characteristic of each alternative vary across choice situations and across individuals. We started by performing a benchmark experiment based on 1000 Monte Carlo replications, 500 individuals, 5 choice situations, 5 alternatives (outside option included), and 100 Halton draws. Table 2, column (1), reports the estimates and Root Mean Square Errors (RMSE) obtained from the estimation of three Mixed Logit models, which differ one another only in the attribute taste heterogeneity is applied to.

[Table 2 about here]

As can be seen, the results in this first setting strongly depend on the model specification: when a random parameter is placed on the continuous attribute only (β_1 random and β_2 fixed in table 2a), estimates cannot be distinguished from the true values; in contrast, when a random parameter is placed on the discrete characteristic (β_2 random, regardless whether β_1 is fixed in table 2b or random in table 2c), the estimates of the random coefficient standard deviation, σ_2, are biased by almost 50% upward, with RMSE that are much larger than those found for any other estimate.[11] Some features of the benchmark experiment were then modified in order to see whether it was possible to reduce the biases. In what follows, we focus on the case where β_2 is random and β_1 fixed (results are very similar with β_1 random). Results are presented in the remaining columns of table 2b. In column (2) for instance, the experiment is identical to the benchmark in all respects except that 5000 Monte Carlo replications are now performed.

[9] Cherchi and Ortúzar (2008) show that identification issues do not depend on the values taken by the elements in Σ, but only on the richness of the associated data.

[10] The Gauss 6.0 language subroutine used both to generate Monte Carlo samples and to estimate the Mixed Logit model is available from the author upon request. The simulated log-likelihood function is maximised using the maxlik (version 5.0) routine. In order to reduce optimisation time, we programmed the simulated gradient vector and used the product of simulated scores to approximate the simulated hessian matrix.

[11] Using different seeds in the process generating the characteristics and the Monte Carlo samples, or different starting values in the optimisation routine does not alter the estimates in any significant way. Conversely, enabling variations of the choice set in both the continuous and discrete attributes improves the estimates in the last two specifications; the estimate of σ_2 in the second specification, for example, becomes 0.373 with a RMSE equal to 0.293.

Estimates show that the increase in the number of replications does not change anything (else than computation time). Note that Chiou and Walker (2007) and Andrews and Currim (2005) use synthetic data generated from much less Monte Carlo draws (one and ten, respectively). In our setting, it was impossible using such small numbers of replications to obtain estimates that were robust to different seeds when β_2 was random, even for very large sample sizes (up to 100000 individuals); considering more individuals and/or more than one replication for such sample sizes was computationally untractable.

The next experiments were performed multiplying by two, with respect to the benchmark, the number of individuals on the one hand, and the number of options included in the choice set on the other hand. In both cases, results improve in a similar way, with estimates for σ_2 that are now biased 25% and 33% upward (see columns (3) and (4), respectively). They further improve when the two changes are made simultaneously, as can be seen in column (5). Again multiplying the number of individuals generated in (5) by two, we obtain estimates in column (6) that are reasonably close to true values, with biases for σ_2 that are lower than 10%. The RMSE values for this parameter remain large however, especially when compared to those obtained for σ_1. Unfortunately, we could not get better ones. We could have considered more individuals and/or more alternatives, but it would have been very time expensive.[12] Eventually, we estimated the model with 50 and 200 Halton draws to test for the robustness of the estimates in (6) to the number of draws used in the approximation of the integral in (2.4). Results reported in the last two columns of table 2b are found to be fairly stable, even though RMSE slightly increase with the number of Halton draws. As a consequence, all the estimates that follow are based on the common settings defined in column (6): 1000 Monte Carlo replications, 2000 individuals, 5 choice situations, 10 options, and 100 Halton draws.

Simulation results for our 48 experiments of interest are reported in table 3. Similarly to table 2, estimates and RMSE are shown under three Mixed Logit specifications: β_1 random and β_2 fixed in table 3a, β_1 fixed and β_2 random in table 3b, and β_1 and β_2 random in table 3c. As mentioned in section 3, cells in each part of table 3 represent different types of choice sets, the features of which may vary either across individuals, or across choice situations, or both. Before going into details, it is possible to draw a first overall lesson from those results: variations in the choice set only matter when a random parameter is applied to the discrete characteristic; when β_1 is random and β_2 fixed (table 3a), each of the 16 table cells indeed shows parameters that are all identified. Although variations in the continuous attributes strongly reduce the RMSE for σ_1 (by about two-third), the estimation of this particular specification does not seem to require the choice set to vary over individuals or choice situations. In that respect, table 3a is not further informative and detailed comments below therefore focus on tables 3b and 3c.

[12] As an indication, more than two days were necessary using a Dual Core processor (2.66GHz, 3.93 GB RAM) to obtain the estimates reported in column (6) of table 2b.

[Table 3 about here]

Results in cell $[t(0,0), i(0,0)]$ of tables 3b and 3c show that the mean and variance of the random parameter applied to the discrete characteristic are biased when there is no variation in the choice set. Although both estimates get closer to true values when the continuous attribute varies over choice situations (cell $[t(1,0), i(0,0)]$), individuals (cell $[t(0,0), i(1,0)]$), or both (cell $[t(1,0), i(1,0)]$), they remain severely biased (by almost 50% for σ_2).[13] Conversely, no bias can be found when the number of alternatives faced from one observation to another varies sufficiently: the biases decrease when that number varies across periods (cells $[t(\cdot,1), i(\cdot,0)]$), and they almost fully disappear when it varies across individuals (cells $[t(\cdot,\cdot), i(\cdot,1)]$).[14] In other words, when a random coefficient is applied to a binary variable describing a set of products and when, for each product, the value taken by the variable is the same for all observations (individuals and weeks), which is typically the case for most qualitative informations displayed on packages, its related parameters (mean and variance) are identified only if the number of options in the choice set varies across a sufficient number of consumers and/or choice situations. In the context of table 3c, this result also means that there is no ground to the idea that consistent estimation of a Mixed Logit model would require at least one parameter to be fixed. All coefficients can be specified as random: as far as there is enough variations in the number of options composing the choice set, their mean and variance can be identified. Eventually, another lesson that can be drawn from this table is that estimating a Mixed Logit model with a randomly distributed taste for the discrete attribute is possible on cross-section data (see the first row) if the number of alternatives varies across individuals; if it does not, then panel data (as in the three remaining rows) are needed, with the number of alternatives included in the common choice set varying across as many choice situations as possible.

Actually, it is unlikely that the 16 configurations examined in table 3 all provide a complete and accurate description of any real dataset, but some can be thought of as more relevant than others. For instance, individual variations contained in the configurations listed in the last three columns are common in typical stated choice data, which are collected through procedures that generate as much variation as possible between individuals. Such configurations are not realistic in revealed choice data, especially when thinking of a brand choice in a supermarket. Variations of the choice set from one consumer to another would indeed be an extreme situation as in most cases they may only occur between broad groups of people who buy in different stores. Revealed choice data would thus rather correspond to some combinations between configurations in the

[13] Interestingly, assuming that it is the discrete, and not the continuous, attribute that varies across individuals and choice situations yields unbiased estimates for all parameters: in particular, the estimates for σ_2 are 0.301 (0.189) and 0.313 (0.183) in the second and third specifications, respectively.

[14] In the first cases, the biases do not completely vanish because of the small number of choice situations considered in the simulations, and we expect them to tend toward zero as this number increases. For instance, considering 10 choice situations instead of 5, the estimates for $\overline{\beta}_2$ and σ_2 in cell $[t(0,1), i(0,0)]$ of table 3b are 0.505 (0.021) and 0.304 (0.180), respectively.

first column and those in the three others. Hence, combining the first and second columns would represent data where all consumers in a given choice situation face the same set of products, regardless the store they purchase in, but with prices varying from one store to another (think of stores selling all producer brands, no retailer brand, and having a specific pricing strategy); the first column combined with the third would correspond to data where individuals buying in a given store only face a subset of all existing products, each sold at a single price (stores selling all producer brands, their own retailer brands, and having the same pricing strategy for common brands); and finally, combination of the first and fourth columns would be identical to the previous one but with common brand prices varying across stores. Given that for many products the main producer brands can all be found in most stores, identification then strongly relies on retailer brands, solely sold in the relevant stores, and on their importance in terms of market shares.

If the datasets used in the three applications that served as illustrations in the introduction had to be characterized by one of these combinations, the latter would probably be the best for all. However, the way the models are estimated makes things different. Remember that Bjørner et al. (2004) and Athey and Imbens (2007) estimate their model on the purchases pooled over stores; the choice set therefore varies in size across several broad groups of individuals (through stores), and the combination of configurations depicted in the first and fourth columns remains the most relevant to describe the estimating sample. It suggests that unbiased estimates could have been obtained had they had cross-section data only. In Bonnet and Simioni (2001), the model is estimated for each store separately, which implies that in a given choice situation (a week) the choice set is identical for all individuals. The relevant configurations to qualify their samples are thus clearly those that are reported in the first column and, conversely to the two preceding papers, repeated observations (over 52 weeks) were therefore necessary to have sufficient variations in the number of alternatives belonging to the choice set and identify the model. Since consumers may not always buy a particular product in the same store, and since stores may introduce new products or withdraw some others, such variations in the number of options across choice situations can potentially be found quite often in consumer panel data. In that respect, the choice set configurations defined in the third and fourth rows of table 3 hence appear as the most relevant in all three applications.

5 Conclusion

Over the last decade, the Mixed Logit model has become a popular tool for analyzing discrete response. While the number of applications has rapidly grown, only a few papers have discussed identification issues yet. From the empirical perspective, specifically, all agree that identification strongly relies on the richness of the data in terms of variations in the explanatory variables. Although this statement is true for any model, it is considered as particularly relevant for Mixed

Logit models, which are highly demanding in that respect. As pointed out by Walker (2002), the difficulty with these models, as well as with other discrete choice model formulations, is that most statistical software packages are likely to report apparently sensible estimation results even if they are not identified. Results from the estimation of unidentified models may therefore be unknowingly reported in the literature.

Several types of data have been used in Mixed Logit applications. Apart from aggregated (at the market level) data, individual data collected through experimental designs are those that are usually found in the economic literature. Individual survey data, such as consumer scanner data, have also been used, though in a smaller number of papers. By construction, variations observed in survey (revealed choice) data are much less important than those found in experimental (stated choice) data. For instance, variations that can typically be observed in most consumer survey data are variations from one week to another and/or from one store to another in the price and in the number of products constituting the consumer choice set. In this paper, we wonder whether such limited variations are sufficient to enable the identification of Mixed Logit models.

We report simulation results obtained from a series of Monte Carlo experiments and provide evidence on the variations that must be observed in the data, and more specifically in the choice set, to help identify such models. Each experiment considers ten options all differentiated by two characteristics, one continuous (taking different values for different alternatives) and one binary (equal to one for some options and to zero for the others). For a given alternative, the continuous attribute is further allowed to take different values for different observations (choice situations, individuals), depending on the experiment; in contrast, the binary characteristic is constrained to be the same for all observations. Variations, across consumers and/or choice situations, in the continuous attribute as well as in the number of options recorded in the choice set, and their impact on Mixed Logit estimates are investigated. We show that these variations do not matter when a random coefficient is applied to the continuous characteristic only: all parameters are identified in each configuration, even in the extreme case where the choice set is the same for all individuals in all choice situations, although the estimate of the random coefficient variance is much more precise when the continuous attribute varies. Conversely, when a random coefficient is applied to the discrete attribute, regardless the nature of the other coefficient (fixed or random), the mean and variance of the random parameter are strongly biased if there is no variation in the choice set, and remain so if the number of options in the choice set does not vary either across enough consumers or across enough choice situations; if it does, all parameters are identified and there is no need to have at least one coefficient fixed.

Hence, variations in the number of alternatives across a sufficient number of choice situations or stores are required to identify Mixed Logit models when a random coefficient is applied to a binary variable. In consumer survey data, retailer brands and the entry/exit of brands guarantee the existence of some actual variations in those two dimensions. At this stage however, it seems worthwhile to make a distinction between variations that are actually observed by consumers in stores and variations that are observed by the econometrician in the data. Constructing choice sets from the data may generate artificial variations. Indeed, even without the introduction or withdrawal of any product, a choice set constructed from the data may change from one period to another, because some options chosen at least once in a given period may not be chosen at all in another period. A possible reason for observing such a situation is purchase infrequency: if consumers of a product do not buy it in every period, some available brands (at least those having small market shares) may not be recorded in the data. Similarly, constructed choice sets may differ between stores, even in the absence of store-specific brands, because a brand available in two stores may be chosen in only one. This case may occur if the survey sampling schedule under-represents consumers of a given store. In other words, the recorded choice set (observed by the econometrician) may be a subset of the choice set faced by consumers, and may therefore vary whereas the real choice set does not. The additional difficulty then is that usual maximum likelihood estimates for many parametric discrete choice models, including Mixed Logit models, are inconsistent when obtained using data on a subset of the options available to individuals. But this goes beyond the scope of the present paper, and the interested reader can refer to Fox (2007), for both a discussion of the issue and the presentation of an alternative and consistent semiparametric estimator based on maximum score.

References

[1] Allenby GM, Arora N, Ginter JL. 1998. On the Heterogeneity of Demand. Journal of Marketing Research 35: 384-389.

[2] Allenby GM, Lenk PJ. 1995. Reassessing Brand Loyalty, Price Sensitivity, and Merchandising Effects on Consumer Brand Choice. Journal of Business and Economic Statistics 13: 281-289.

[3] Allenby GM, Rossi PE. 1999. Marketing Models of Consumer Heterogeneity. Journal of Econometrics 89: 57-78.

[4] Andrews RL, Currim IS. 2005. An Experimental Investigation of Scanner Data Preparation Strategies for Consumer Choice Models. International Journal of Research in Marketing 22: 319-331.

[5] Athey S, Imbens GW. 2007. Discrete Choice Models with Multiple Unobserved Choice Characteristics. International Economic Review 48: 1159-1192.

[6] Bajari P, Fox JT, Kim K, Ryan S. 2007. A Simple Nonparametric Estimator for the Distribution of Random Coefficients in Discrete Choice Models. University of Chicago, Working Paper.

[7] Berry S, Levinsohn J, Pakes A. 1995. Automobile Prices in Market Equilibrium. Econometrica 63: 841-890.

[8] Bjørner TB, Hansen LG, Russell CS. 2004. Environmental Labeling and Consumers' Choice — An Empirical Analysis of the Effect of the Nordic Swan. Journal of Environmental Economics and Management 47: 411-434.

[9] Bonnet C, Simioni M. 2001. Assessing Consumer Response to Protected Designation of Origin Labelling: A Mixed Multinomial Logit Approach. European Review of Agricultural Economics 28: 433-449.

[10] Brownstone D, Train K. 1998. Forecasting New Product Penetration with Flexible Substitution Patterns. Journal of Econometrics 89: 109-129.

[11] Cherchi E, Ortúzar JD. 2008. Empirical Identification in the Mixed Logit Model: Analysing the Effect of Data Richness. Network and Spatial Economics 8: 109-124.

[12] Chiou L, Walker JL. 2007. Masking Identification of Discrete Choice Models under Simulation Methods. Journal of Econometrics 141: 683-703.

[13] Davis P. 2006. Spatial Competition in Retail Markets: Movie Theaters. RAND Journal of Economics 37: 964-982.

[14] Dubé JP, Chintagunta P, Petrin A, Bronnenberg B, Goettler R, Seetharaman PB, Sudhir K, Thomadsen R, Zhao Y. 2002. Structural Applications of the Discrete Choice Model. Marketing Letters 13: 207-220.

[15] Fox JT. 2007. Semiparametric Estimation of Multinomial Discrete Choice Models using a Subset of Choices. RAND Journal of Economics 38: 1002-1019.

[16] Hensher DA, Greene WH. 2003. The Mixed Logit Model: The State of Practice. Transportation 30: 133-176.

[17] McFadden D, Train K. 2000. Mixed MNL Models for Discrete Response. Journal of Applied Econometrics 15: 447-470.

[18] Nevo A. 2000. A Practitioner's Guide to Estimation of Random Coefficients Logit Models of Demand. Journal of Economics and Management Strategy 9: 513-548.

[19] Nevo A. 2001. Measuring Market Power in the Ready-to-Eat Cereal Industry. Econometrica 69: 307-342.

[20] Revelt D, Train K. 1998. Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level. Review of Economics and Statistics 80: 1-11.

[21] Rigby D, Burton M. 2005. Preference Heterogeneity and GM Food in the UK. European Review of Agricultural Economics 32: 269-288.

[22] Sándor Z, Wedel M. 2002. Profil Construction in Experimental Choice Designs for Mixed Logit Models. Marketing Science 21: 455-475.

[23] Train K. 1999. Halton Sequences for Mixed Logit. University of California, Berkeley, Working Paper.

[24] Train K, Wilson WW. 2008. Estimation on Stated-Preference Experiments Constructed from Revealed-Preference Choices. Transportation Research, Part B: Methodology 40: 191-203.

[25] Walker JL. 2002. The Mixed Logit (or Logit Kernel) Model: Dispelling Misconceptions of Identification. Transportation Research Record 1805: 86-98.

[26] Walker JL, Ben-Akiva M, Bolduc D. 2007. Identification of Parameters in Normal Error Component Logit-Mixture (NECLM) Models. Journal of Applied Econometrics 22: 1095-1125.

Table 1a: Descriptive statistics on characteristics

Options	Continuous				Discrete
	$[t(0,\cdot), i(0,\cdot)]$	$[t(1,\cdot), i(0,\cdot)]$	$[t(0,\cdot), i(1,\cdot)]$	$[t(1,\cdot), i(1,\cdot)]$	
0	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0
1	−0.091 (0.000)	0.309 (1.185)	−0.028 (0.964)	−0.001 (0.991)	0
2	0.668 (0.000)	−0.366 (0.564)	−0.036 (0.999)	−0.002 (1.008)	1
3	0.599 (0.000)	−0.089 (0.748)	−0.009 (0.980)	−0.017 (0.998)	1
4	−0.099 (0.000)	−0.082 (1.089)	−0.034 (0.993)	0.004 (0.993)	1
5	1.763 (0.000)	0.719 (1.961)	−0.006 (1.008)	−0.010 (1.004)	0
6	0.223 (0.000)	0.304 (0.815)	0.023 (0.967)	0.001 (0.996)	1
7	0.881 (0.000)	0.085 (0.756)	0.023 (0.988)	0.004 (1.000)	1
8	1.146 (0.000)	0.271 (0.890)	−0.039 (0.971)	0.001 (0.997)	1
9	−0.173 (0.000)	−0.236 (0.677)	0.007 (1.003)	0.003 (1.000)	0
All	0.546 (0.000)	0.102 (0.965)	−0.011 (0.996)	−0.002 (0.999)	

Note: Standard deviations in parentheses.

Table 1b: Descriptive statistics on choice sets

Options	Probability of being included			
	$[t(\cdot,0), i(\cdot,0)]$	$[t(\cdot,1), i(\cdot,0)]$	$[t(\cdot,0), i(\cdot,1)]$	$[t(\cdot,1), i(\cdot,1)]$
0	1.000	1.000	1.000	1.000
1	1.000	1.000	1.000	1.000
2	1.000	0.400	0.507	0.502
3	1.000	0.400	0.516	0.507
4	1.000	0.800	0.519	0.500
5	1.000	0.600	0.504	0.494
6	1.000	0.400	0.484	0.493
7	1.000	0.800	0.519	0.500
8	1.000	0.800	0.485	0.486
9	1.000	0.200	0.478	0.494
	Average number of options			
	10 (0.000)	6.400 (0.490)	6.013 (1.378)	5.977 (1.400)

Note: Standard deviations in parentheses.

Table 2a: Monte Carlo θ estimates, β_1 random and β_2 fixed

	θ_0	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\overline{\beta}_1$	-0.2	-0.201 (0.035)	-0.202 (0.035)	-0.200 (0.024)	-0.199 (0.029)	-0.198 (0.020)	-0.199 (0.015)	-0.199 (0.015)	-0.199 (0.015)
$\overline{\beta}_2$	0.5	0.500 (0.046)	0.499 (0.045)	0.500 (0.031)	0.501 (0.042)	0.502 (0.031)	0.501 (0.022)	0.501 (0.022)	0.501 (0.022)
σ_1	0.3	0.300 (0.146)	0.298 (0.146)	0.289 (0.114)	0.286 (0.100)	0.291 (0.069)	0.297 (0.047)	0.295 (0.047)	0.297 (0.046)

Table 2b: Monte Carlo θ estimates, β_1 fixed and β_2 random

	θ_0	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\overline{\beta}_1$	-0.2	-0.203 (0.031)	-0.203 (0.031)	-0.201 (0.021)	-0.201 (0.024)	-0.200 (0.017)	-0.200 (0.013)	-0.200 (0.013)	-0.200 (0.013)
$\overline{\beta}_2$	0.5	0.511 (0.052)	0.510 (0.052)	0.507 (0.036)	0.515 (0.057)	0.509 (0.039)	0.505 (0.027)	0.505 (0.026)	0.505 (0.027)
σ_2	0.3	0.424 (0.362)	0.427 (0.369)	0.370 (0.298)	0.398 (0.327)	0.333 (0.255)	0.306 (0.212)	0.318 (0.203)	0.315 (0.215)

Table 2c: Monte Carlo θ estimates, β_1 and β_2 random

	θ_0	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\overline{\beta}_1$	-0.2	-0.204 (0.036)	-0.205 (0.037)	-0.202 (0.025)	-0.201 (0.029)	-0.199 (0.020)	-0.200 (0.015)	-0.200 (0.015)	-0.200 (0.015)
$\overline{\beta}_2$	0.5	0.512 (0.054)	0.511 (0.055)	0.508 (0.037)	0.517 (0.057)	0.511 (0.040)	0.505 (0.027)	0.505 (0.027)	0.505 (0.027)
σ_1	0.3	0.317 (0.151)	0.312 (0.153)	0.299 (0.114)	0.290 (0.102)	0.294 (0.069)	0.298 (0.048)	0.296 (0.048)	0.299 (0.047)
σ_2	0.3	0.458 (0.389)	0.461 (0.392)	0.389 (0.305)	0.420 (0.337)	0.354 (0.257)	0.325 (0.207)	0.326 (0.201)	0.315 (0.217)

Notes: Root Mean Square Errors in parentheses; (1) = Benchmark, (2) = (1) with 5000 Monte Carlo replications, (3) = (1) with 1000 individuals, (4) = (1) with 10 alternatives, (5) = (1) with 1000 individuals and 10 alternatives, (6) = (1) with 2000 individuals and 10 alternatives, (7) = (6) with 50 Halton draws, (8) = (6) with 200 Halton draws.

Table 3a: Monte Carlo θ estimates, β_1 random and β_2 fixed

	θ_0	$i(0,0)$	$i(1,0)$	$i(0,1)$	$i(1,1)$
$t(0,0)$					
$\overline{\beta}_1$	-0.2	-0.202 (0.027)	-0.200 (0.014)	-0.202 (0.032)	-0.200 (0.015)
$\overline{\beta}_2$	0.5	0.502 (0.029)	0.500 (0.023)	0.501 (0.029)	0.500 (0.022)
σ_1	0.3	0.281 (0.144)	0.295 (0.042)	0.279 (0.151)	0.293 (0.054)
$t(1,0)$					
$\overline{\beta}_1$	-0.2	-0.199 (0.015)	-0.200 (0.014)	-0.199 (0.017)	-0.199 (0.015)
$\overline{\beta}_2$	0.5	0.500 (0.022)	0.501 (0.022)	0.500 (0.022)	0.500 (0.022)
σ_1	0.3	0.297 (0.027)	0.300 (0.038)	0.298 (0.033)	0.297 (0.049)
$t(0,1)$					
$\overline{\beta}_1$	-0.2	-0.201 (0.024)	-0.200 (0.015)	-0.204 (0.032)	-0.200 (0.015)
$\overline{\beta}_2$	0.5	0.502 (0.026)	0.500 (0.022)	0.504 (0.030)	0.501 (0.022)
σ_1	0.3	0.278 (0.147)	0.296 (0.049)	0.289 (0.152)	0.295 (0.052)
$t(1,1)$					
$\overline{\beta}_1$	-0.2	-0.200 (0.017)	-0.200 (0.014)	-0.199 (0.017)	-0.199 (0.015)
$\overline{\beta}_2$	0.5	0.501 (0.026)	0.501 (0.022)	0.500 (0.022)	0.501 (0.022)
σ_1	0.3	0.298 (0.031)	0.297 (0.046)	0.296 (0.034)	0.297 (0.047)

Note: Root Mean Square Errors in parentheses.

Table 3b: Monte Carlo θ estimates, β_1 fixed and β_2 random

	θ_0	$i(0,0)$	$i(1,0)$	$i(0,1)$	$i(1,1)$
$t(0,0)$					
$\overline{\beta}_1$	-0.2	-0.194 (0.079)	-0.202 (0.011)	-0.200 (0.021)	-0.201 (0.012)
$\overline{\beta}_2$	0.5	1.958 (7.305)	0.621 (0.271)	0.504 (0.028)	0.504 (0.027)
σ_2	0.3	3.541 (15.00)	0.707 (0.800)	0.314 (0.217)	0.308 (0.212)
$t(1,0)$					
$\overline{\beta}_1$	-0.2	-0.201 (0.012)	-0.202 (0.012)	-0.200 (0.013)	-0.200 (0.013)
$\overline{\beta}_2$	0.5	0.666 (0.526)	0.603 (0.202)	0.504 (0.028)	0.505 (0.028)
σ_2	0.3	0.820 (1.204)	0.668 (0.677)	0.312 (0.213)	0.315 (0.216)
$t(0,1)$					
$\overline{\beta}_1$	-0.2	-0.201 (0.020)	-0.201 (0.012)	-0.201 (0.021)	-0.200 (0.013)
$\overline{\beta}_2$	0.5	0.528 (0.060)	0.526 (0.057)	0.505 (0.027)	0.504 (0.027)
σ_2	0.3	0.436 (0.347)	0.416 (0.322)	0.311 (0.217)	0.311 (0.209)
$t(1,1)$					
$\overline{\beta}_1$	-0.2	-0.201 (0.013)	-0.201 (0.012)	-0.200 (0.013)	-0.200 (0.013)
$\overline{\beta}_2$	0.5	0.524 (0.054)	0.527 (0.060)	0.505 (0.027)	0.505 (0.027)
σ_2	0.3	0.413 (0.324)	0.415 (0.332)	0.309 (0.213)	0.306 (0.212)

Note: Root Mean Square Errors in parentheses.

Table 3c: Monte Carlo θ estimates, β_1 and β_2 random

	θ_0	$i(0,0)$	$i(1,0)$	$i(0,1)$	$i(1,1)$
$t(0,0)$					
$\overline{\beta}_1$	−0.2	−0.200 (0.052)	−0.202 (0.014)	−0.202 (0.033)	−0.200 (0.015)
$\overline{\beta}_2$	0.5	1.949 (5.469)	0.607 (0.218)	0.507 (0.034)	0.505 (0.029)
σ_1	0.3	0.277 (0.142)	0.300 (0.042)	0.283 (0.151)	0.294 (0.054)
σ_2	0.3	3.573 (11.03)	0.679 (0.699)	0.328 (0.217)	0.318 (0.213)
$t(1,0)$					
$\overline{\beta}_1$	−0.2	−0.201 (0.015)	−0.201 (0.014)	−0.199 (0.017)	−0.200 (0.015)
$\overline{\beta}_2$	0.5	0.648 (0.353)	0.594 (0.181)	0.505 (0.028)	0.505 (0.028)
σ_1	0.3	0.305 (0.030)	0.304 (0.039)	0.299 (0.033)	0.298 (0.050)
σ_2	0.3	0.800 (0.954)	0.640 (0.631)	0.325 (0.207)	0.329 (0.215)
$t(0,1)$					
$\overline{\beta}_1$	−0.2	−0.202 (0.025)	−0.201 (0.015)	−0.204 (0.033)	−0.200 (0.015)
$\overline{\beta}_2$	0.5	0.530 (0.061)	0.527 (0.059)	0.508 (0.033)	0.505 (0.027)
σ_1	0.3	0.287 (0.143)	0.300 (0.051)	0.291 (0.149)	0.297 (0.054)
σ_2	0.3	0.449 (0.343)	0.437 (0.326)	0.326 (0.210)	0.324 (0.205)
$t(1,1)$					
$\overline{\beta}_1$	−0.2	−0.200 (0.017)	−0.201 (0.014)	−0.200 (0.017)	−0.200 (0.015)
$\overline{\beta}_2$	0.5	0.528 (0.058)	0.527 (0.058)	0.505 (0.028)	0.505 (0.027)
σ_1	0.3	0.299 (0.031)	0.301 (0.046)	0.298 (0.035)	0.298 (0.048)
σ_2	0.3	0.461 (0.348)	0.434 (0.325)	0.322 (0.208)	0.325 (0.207)

Note: Root Mean Square Errors in parentheses.

ALISS Working Papers

2008

2008-07 Giraud-Héraud, E. ; Hammoudi, A. ; Hofmann, R. ; Soler, L.G. **Vertical relationships and safety standards in the food marketing chain,** *Aliss Working Paper 2008-07,* juillet 2008, 30 p.

2008-06 Régnier, F. ; Masullo A. **Une affaire de goût ?". Réception et mise en pratique des recommandations nutritionnelles,** *Aliss Working Paper 2008-06,* juillet 2008, 51 p.

2008-05 Giraud-Héraud, E. ; Grazia, C. ;Hammoudi, A. **Strategies for the development of brands in the agrifood chains,** *Aliss Working Paper 2008-05,* juillet 2008, 34 p.

2008-04 Cardon, P ; Gojard, S. **Les personnes âgées face à la dépendance culinaire : entre délégation et remplacement,** *Aliss Working Paper 2008-04*, Juillet 2008, 22 p.

2008-03 Allais, O. ; Bertail, P. ; Nichèle, V. **The effects of a "Fat Tax" on the nutrient intake of French Households,** *Aliss Working Paper 2008-03,* Juin 2008, 36 p.

2008-02 Etilé, F. **Food Price Policies and the Distribution of Body Mass Index: Theory and Empirical Evidence from France,** *Aliss Working Paper 2008-02,* Juin 2008, 52 p.

2008-01 Boizot-Szantaï, C., Etilé, F. **Le prix des aliments et la distribution De l'Indice de Masse Corporelle des Français,** *Aliss Working Paper 2008-01*, Mai 2008, 19 p.