Estimating non-parametric random utility models, with an application to the value of time in heterogeneous populations

by F. Bastin[1], C. Cirillo[2] and Ph. L. Toint[3]

Report 07/15 December 15, 2007

[1] Department of Computing Science and Operational Research,
University of Montreal, Canada
email: bastin@iro.umontreal.ca

[2] Department of Civil and Environmental Engineering,
University of Maryland, USA
email: ccirillo@umd.edu

[3] Department of Mathematics,
University of Namur,
61, rue de Bruxelles, B-5000 Namur, Belgium,
Email: philippe.toint@fundp.ac.be

Estimating non-parametric random utility models, with an application to the value of time in heterogeneous populations

Fabian Bastin[*], Cinzia Cirillo[†], Philippe L. Toint[‡]

December 15, 2007

Abstract

The estimation of random parameters by means of mixed logit models is becoming current practice amongst discrete choice analysts, one of the most straightforward applications being the derivation of willingness to pay distribution over a heterogeneous population. In numerous practical cases, parametric distributions are a priori specified and the parameters for these distributions are estimated. This approach can however lead to many practical problems. Firstly, it is difficult to assess which is the more appropriate analytical distribution. Secondly, unbounded distributions often produce values ranges with difficult behavioural interpretation. Thirdly, little is known about the tails and their effects on the mean of the estimates. (Hess, Bierlaire, Polak [18], Cirillo and Axhausen [6]).

In this paper, we propose to capture the randomness present in the model by using a new nonparametric estimation method, based on the approximation of inverse cumulative distribution functions. This technique is applied to simulated data and the ability to recover both parametric and non-parametric random vectors is tested. The non-parametric mixed logit model is also used on real data derived from a Stated Preference survey conducted in the region of Brussels (Belgium) in 2002. The model presents multiple choices and is estimated on repeated observations.

1 Introduction

In mixed logit models estimation, investigators traditionally use parametric models involving specific functional forms and a finite number of unknown parameters. Among them, some are considered as random variables in order to reflect population heterogeneity.

The early applications of mixed logit have mainly used normal distributions. The use of unbounded distributions however appears inappropriate in many cases: certain

[*]Department of Computing Science and Operational Research, University of Montreal, Canada.
[†]Department of Civil and Environmental Engineering, University of Maryland, USA.
[‡]Department of Mathematics, University of Namur, Belgium.

attributes are assumed to be positively (or negatively) valued by all individuals; moreover, a zero cost coefficient causes problems for the evaluation of the willingness to pay. In order to circumvent these difficulties, more recent models use bounded distributions, often obtained as simple transformations of normals. Train and Sonnier [27] specify mixed logit models with lognormal, censored normal and Johnson Sb distributions bounded on both sides. They also suggest to adopt Bayesian procedure in order to avoid estimation problems encountered with log-normal distributions parameters.

Some investigators have also questioned whether the underlying theory is capable of conveying sufficient information to enable a correct and successful specification of parametric models and have instead proposed the less restrictive nonparametric or semiparametric approaches to the problem. In that context, Dong and Koppelman [9] assume that distributions are represented by a finite number of points and use the Bayesian method to recover their mass and the associated probabilities. They indicate that Maximum Likelihood Mixed Logit failed to recover the true mass points from simulated data, although no reasons are given to explain that fact. The empirical analysis reported by these authors show that Mass Point Mixed Logit is superior to Parametric Mixed logit. Those results are nevertheless limited by the use of only two points along each of the parameter dimensions. Hess, Bierlaire and Polak [17] propose discrete mixture of GEV models over a finite set of distinctive support points. The major advantage of this approach is the lack of need for simulation processes. The authors however report several issues in the estimation of such models mainly due to the non-linearity non-concavity of the log-likelihood function.

Hensher [15] resolves the problem of behaviourally incoherent sign changes by imposing a global sign condition on the marginal disutility expression and gives an application on the valuation of travel-time savings for car commuters. He adopts a globally constrained Rayleigh distribution for total travel time parameterization, although his focus is not on the specific analytical distribution, but on the behavioural appeal of the imposition of a global sign condition. Train and Weeks [28] place distributional assumptions on the willingness to pay and derive the distribution of the coefficients. Their major finding is that models using normal and lognormal distributions for coefficients (models in preference space) fit the data better than those in willingness to pay space but provide less reasonable distribution for the willingness to pay. They also conclude that it is not possible to identify the distribution to use in all situations and that the best distribution-fit is highly situation-dependent. Fosgerau [13] employs various non-parametric techniques to investigate the distribution of the travel-time savings from a stated choice experiment. The methodology adopted by this author relies on a two-step estimation procedure: a Klein and Spady [20] estimator is first used to estimate parameters in a linear index binary choice model with no assumptions on the error term distribution; the distribution of the error term is then estimated. This method does not account for repeated observations and applies only to binomial choices. Recently, Fosgerau and Bierlaire [12] have proposed a semi nonparametric (SNP) specification, based on Legendre polynomials, to test if a random parameter of a discrete choice model follows a given distribution. The SNP technique has been successfully applied to simulated data for testing the null hypothesis that the true (and known) distribution is a normal or lognormal and to a simple real case study; the test is adapted for just one random parameter at a time.

The purpose of this paper is to introduce a new non-parametric approach to resolve the difficulties associated with the identification of underlying unknown base random distributions. Our proposal is characterized by the explicit estimation of the shape of the unknown distributions, expressed via their cumulative distributions, as a part of the complete calibration procedure. Due to multiple modelling assumptions, such as linear utilities or the use of a Gumbel distribution to characterize the unobserved part of the utility function, the recovering of the true distributions inside the population is a tricky, and perhaps vain, task. It is nevertheless meaningful to capture the randomness nature of some parameters to correctly apprehend to heterogeneity inside the population. We will exhibit in this paper that our approach, even based on rough approximations, is able to fulfil this objective.

In order to perform this estimation efficiently, we use a B-spline parametrization of the inverse cumulative distribution functions and the associated monotonicity constraints are then included in the log-likelihood maximization. B-splines are known to provide a concise formulation for curves that are composed of many polynomial pieces, thereby automatically controlling the overall curve smoothness (Farin [11]). This technique is often used for nonparametric regression (Fox [14]). Recent applications in such areas as meteorology, medicine and price modeling can be found in Singh, McNamara, and Lozanoff [25], Jarvis and Stuart [19], and Bao and Wan [2]. To date there have only been a handful of applications of this class of functions in econometrics (Engle, Granger, Rice, and Weiss [10], Koenker, Ng and Portnoy [21]), partly due to the difficulty to impose some conditions, as monotonicity. To our best knowledge, it is new for mixed logit models estimation. The random variables of the objective functions are here assumed to be continuous, bounded, and independent, and we are interested by the inverse cumulating distribution functions. These functions are modeled by means of cubic B-splines with strictly increasing base coefficients, a sufficient condition to construct monotonic (increasing) functions. As a result, the number of parameters that have to be estimated increases; the information on the shape of the random variables however should help the analyst to find the right parametric distribution for the random parameters (if this exists).

The paper is organized as follows. Section 2 briefly recalls the Mixed logit model formulation and the estimation techniques adopted to solve the related maximum log-likelihood problem. Non-parametric estimation of continuous variables is developed on Section 3, together with a short review of the constrained optimization techniques used to ensure a correct function shape. Section 4 presents results obtained on simulated data and discusses the ability to recover both parametric and non-parametric random vectors. Results on a real case study are given in Section 5. Conclusions and perspectives for research are finally presented.

2 Mixed Multinomial Logit Model (MMNL) formulation

The mixed logit formulation is nowadays extensively used in transport modeling for its flexibility. In particular, MMNL models estimate taste variation, avoid the problem of restricted substitution pattern in standard logit model and account for state dependency across observations.

We consider a set of I individuals, each one having to choose one alternative within a finite set $\mathcal{A}_j$. We associate an utility U_{ij} to each alternative A_j in $\mathcal{A}_i$, as perceived by individual i. Relying on the econometric theory, we also assume that individuals aim to maximize their utility, but we do not observe all the components. Instead, we decompose the utility U_{ij} as the sum of a deterministic part $V_{ij}(\beta)$, where β is a vector to estimate, and a random, unobserved part ϵ_{ij}. The probability choice is then

$$L_{ij}(\beta) = P[V_{ij}(\beta) + \epsilon_{ij} \geq V_{in}(\beta) + \epsilon_{in},\ \forall A_n \in \mathcal{A}(i)].$$

The probability expression is of course dependent of the distribution choice for ϵ_{ij}. When the ϵ_{ij}'s are assumed to be i.i.d. Gumbel's amongst the individuals and alternatives, we obtain the traditional logit probabiilty.

In the mixed logit framework, we relax the assumption that β is a constant vector, but instead a random vector with cumulative distribution function (CDF) $F_B(\beta)$ so that the probability choice L_{ij} is now conditional on the realization β, and the unconditional probability is

$$P_{ij} = E_B[L_{ij}(\beta)] = \int L_{ij}(\beta) dP_B(\beta). \tag{1}$$

We therefore cannot directly estimate B, so we will assume that it can be described as $B = B(\Gamma, \theta)$, where Γ is some random vector, and θ some constant parameters vector to estimate. In other terms, we assume some distribution family for B, parametrized by θ. If, moreover, the vector β is continuous, we can rewrite (1) as

$$P_{ij}(\theta) = \int L_{ij}(\gamma, \theta)\phi(\gamma, \theta) d\gamma,$$

where $\phi(\gamma, \theta)$ is the density of B, with parameters vector θ.

In the case when the same individual can express several choices, we observe for each individual the sequence of choices $y_i = (j_{i1}, \ldots, j_{iT_i})$, that can be assumed to be correlated, and we will consider the data as panel data. A simple way to accomodate this situation is to assume the heterogeneity is present on the population level only, but not on the individual level. The probability to observe the individuals choices is then given by the product of logit probabilities $L_{ij_{it}}$ (Train [26]):

$$P_{iy_i}(\theta) = \int \left(\prod_{t=1}^{T_i} L_{ij_{it}}(\gamma, \theta) \right) \phi(\gamma, \theta) d\gamma.$$

2.1 MMNL Model estimation

The vector of unknown parameters is estimated by maximizing the log-likelihood function, i.e. by solving the problem:

$$\max_{\theta} LL(\theta) = \max_{\theta} \frac{1}{I} \sum_{i=1}^{I} \ln P_{iy_i}(\theta), \tag{2}$$

where y_i is the vector of alternative choices made by the individual i. This involves the computation of $P_{iy_i}(\theta)$ for each individual i ($i = 1, \ldots, I$), which is impractical since it requires the evaluation of one multidimensional integral per individual. To approximate this integral, a popular approach is to choose for each individual a point set $S_R = \{u_1, \ldots, u_R\}$ in $(0, 1]^s$, where s is the problem dimension, i.e. the number of random coefficients, convert the vectors u_{r_i} to the (multivariate) distribution of Γ, and then take the average value of the function over S_R. This leads to the simulated probability

$$SP_{iy_i}^R = \frac{1}{R} \sum_{r_i=1}^{R} \prod_{t=1}^{T_i} L_{ij_{it}}(\gamma_{r_i}, \theta),$$

where R is the number of random draws γ_{r_i}. As a result, θ is now computed as one solution of the simulated log-likelihood problem

$$\max_{\theta} SLL^R(\theta) = \max_{\theta} \frac{1}{I} \sum_{i=1}^{I} \ln SP_{iy_i}^R(\theta).$$

We will denote by θ_R^* one solution of this last approximation (often called Sample Average Approximation, or SAA), while θ^* denotes the solution of the true problem (2).

3 Non-parametric estimation of continuous variables

Most of the studies devoted to estimation of parameters without assumptions on the underlying distributions are concerned with discrete distributions. Such a discrete treatment could lead to an arbitrary population segmentation, which can be avoided if we turn to continuous distributions.

Each component of the random vector inherent to the mixed logit function is itself random, and if we assume independence between these components, we can consider each one separately. Then all what we have to do is to draw from univariate random variables. If X is an univariate random distribution, a well-known technique to generate draws from its distribution consist to sample an uniform on $[0, 1]$, hereafter denoted by $U[0, 1]$, and to apply the inverse cumulative distribution function F_X^{-1} to these draws:

$$S_X = \{F_X^{-1}(U),\ U \sim U[0, 1]\},$$

where S_X represented the sample set drawn from the random variable X. It is usually assumed that F_X^{-1} is available (or at least some numerically good approximation of it), the distribution X being known. This method is known as the inversion technique in the random numbers generation litterature (Devroye [8], Law [22]), and is also popular in the context of variance reduction methods (see for instance L'Ecuyer [23]).

We will capitalize on this approach by assuming that the distribution of the random variable X is not known, but that F_X, or more precisely F_X^{-1}, can still be approximated in some way. If X is a random continuous variable, the only properties that F_X^{-1} has to satisfy are:

- $F_X^{-1} : [0, 1] \to \mathcal{R}$,

- F_X^{-1} is monotonically increasing,
- F_X^{-1} is continuous.

In other terms, we have to estimate an arbitrary continuous real function whose domain is $[0,1]$, and which is monotonically increasing.

Functions approximation is a large field of mathematics, and various techniques are possible. We however seek an adequate balance between estimation capabilities and satisfaction of the conditions ensuring that we can interpret the function as an inverse cumulative distribution function. Moreover, the density exists only for continuous distributions, and is the derivative of the cumulative distribution function. It is also usually easier to estimate a function rather than its derivative. All these considerations lead us to propose to estimate the inverse cumulative distribution function and to express it as some element of a functional space:

$$F_X^{-1}(\cdot) = \sum_{k=1}^{\infty} p_k h_k(\cdot),$$

where $\{h_k,\ k = 0, \ldots, \infty\}$ constitute a basis of this space and p_k are the coordinates of F_X (if the basis cardinal is finite, and equal to n, we just set h_k and p_k to 0, for $k > n$).

If we furthermore assume that the random variable X has a bounded support, an elegant way to achieve such a balance is the use of B-spline functions. The bounded support assumption is not too much restrictive, since extreme behaviours, corresponding to values of X tending to plus or minus infinity, are usually not welcome since that are difficult to interpret. We therefore consider the bounded support assumption as an advantage rather than a drawback of the proposition.

A B-spline function of degree p is a polynomial function of degree p, defined on the interval $[a, b]$, that can be expressed as a linear combination of $n+1$ basis functions $N_{i,p}(u)$, as follows:

$$C(u) = \sum_{i=0}^{n} P_i N_{i,p}(u).$$

The coefficients $P_0, P_1, \ldots, P_n$ are called the control points, and u is the knot vector $(u_0 = a, u_1, \ldots, u_m = b)$. The basis functions can be constructed by recurrence on the degree p:

$$N_{i,0} = \begin{cases} 1 & \text{if } u \in [u_i, u_{i+1}), \\ 0 & \text{otherwise.} \end{cases}$$

and

$$N_{i,p} = \frac{u - u_i}{u_{i+p} - u_i} N_{i,p-1}(u) + \frac{u_{i+p+1} - u}{u_{i+p+1} - u_{i+1}} N_{i+1,p-1}(u),$$

so that n is equal to $m - p - 1$.

There are several types of knot vectors, but one especially convenient for our purposes is the nonperiodic (or clamped or open) knot vector, which has the form

$$U = \{\underbrace{a, \ldots, a}_{p+1}, u_{p+1}, \ldots, u_{m-p-1}, \underbrace{b, \ldots, b}_{p+1}\},$$

that is the first and last knots have multiplicity $p + 1$.

It is possible to show that the function $C(u)$ is $p-1$ continuously differentiable. In this paper, we will consider cubic B-splines, i.e. we will set p to 3. Another particularly nice property with respect to our needs is that with these basis and knots choices, $C(u)$ is monotonically increasing if $P_0 \le P_1 \le \ldots \le P_n$. As we will describe in the next Section, this property can be algorithmically guaranteed. The resulting spline construction is illustrated in Figure 1.

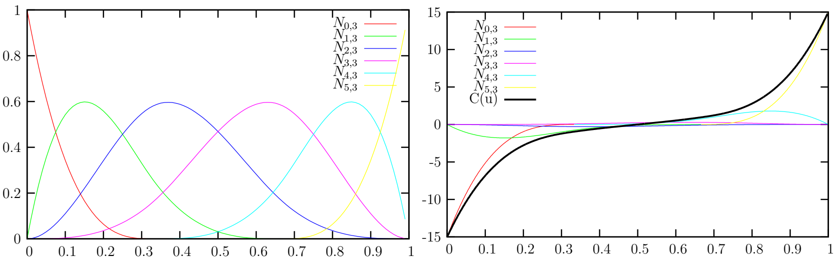

Figure 1: Basis B-splines and monotonically increasing spline.

For a more complete review of B-splines properties, we refer the reader to Piegl and Tiller [24].

3.1 Constrained Optimization

When estimating the log-likelihood function, we have to solve a problem of the form

$$\max_{x \in C} f(x),$$

where $C \subset \mathcal{R}^n$ is the feasible region. In our case, C will have the form

$$C = \{x_{di} \le x_{di+1} \le \ldots \le x_{di+n}\},$$

for some i's in $\mathcal{N}_0$. That is C is the set of real numbers satisfying the monotonicity constraints. Such constraints can be easily dealt with projections. For simplicity, we temporarily assume that we only have one non-parametric coefficient, so that C defines k ordered variables. C is then call the order-simplex, as illustrated in Figure 2 for the three-dimensional case.

The projection onto the order-simplex can be performed easily and efficiently, since several algorithms of complexity $O(n)$ have been designed (Ayer, Brunk, Ewing, Reid and Silverman [1], Best and Chakravarti [4]). Moreover, it is possible to adapt a well-known approach for nonlinear optimization, the trust-region algorithm, to benefit from such projections (see Chapter 12 of Conn, Gould, and Toint [7]). The main idea of a trust-region algorithm involves the calculation, at iteration k (with current estimate θ_k), of a trial point $\theta_k + s_k$ by approximately maximizing a model m_k of the objective function inside a trust region defined as

$$\mathcal{B}_k = \{\theta \text{ such that } \|\theta - \theta_k\|_k \le \Delta_k\},$$

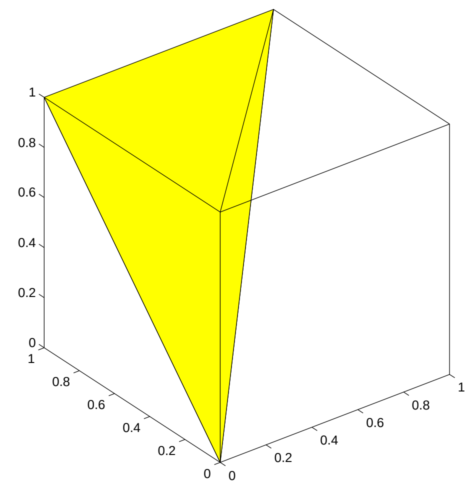

Figure 2: Order-simplex

where Δ_k is called the trust-region radius. We will here use a quadratic model:

$$m_k(s) = SLL^R(\theta_k) + s^T \nabla_\theta SLL^R(\theta_k) + \frac{1}{2} s^T H_k s, \tag{3}$$

where H_k is a symmetric approximation of the Hessian $\nabla^2_{\theta\theta} SLL^R(\theta_k)$. The step s_k is computed by first attempting to identifying the active constraints by maximizing the model (3) along the projected gradient path (illustrated in Figure 3). Once these active constraints have been guessed, the model is further maximized in the corresponding face using an approximate projected truncated conjugate-gradient technique. The predicted and actual increases in the value of the objective function are then compared by computing the ratio

$$\rho_k = \frac{SLL^R(\theta_k + s_k) - SLL^R(\theta_k)}{m_k(\theta_k + s_k) - m_k(\theta_k)},$$

If this ratio is larger than a certain threshold, set to 0.01 in our tests, the trial point becomes the new iterate, and the trust-region radius is (possibly) enlarged. More precisely, if ρ_k is larger than 0.75, we set the trust-region Δ_k to be the maximum between θ_k and $2\|s_k\|$, otherwise we set $\Delta_k = 0.5\Delta_k$. If the ratio is below the bound, the trial point is rejected and the trust region is shrunk by a factor of 2, in order to improve the correspondence of the model with the true objective function, thereby concluding the iteration. Details of the convergence theory and constraint identification for the proposed algorithm may be found in Chapter 12 of [7].

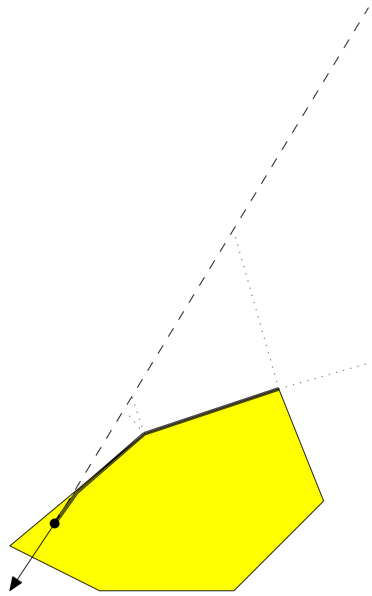

Figure 3: Projected conjugated gradient path

4 Simulations

Our first experiment aims at verifying that known distributions can be approximately recovered by our estimation technique. In our simulated experiments we create two synthetic populations; the first data set is cross sectional and simulates 2000 individuals giving just one response, the second data set is a panel of 1000 individuals contributing with two observations each. The design contains four alternatives, normally distributed with parameters $N(0.5, 1)$, and one independent variable. We run two simulations on each of the data sets described. The first supposes that the coefficient to be estimated is normally distributed with parameters $N(0, 4)$. The second assumes that the coefficient is lognormally distributed with mean 1.133 and standard deviation 0.604. Results are illustrated in the following figures, where the cumulative distribution function (CDF) is shown on the left and the inverse of the cumulative distribution function (ICDF) on the right, and the spline is constructed using the knot vector $\{0, 0, 0, 0, 1/3, 2/3, 1, 1, 1, 1\}$. As far as normal distribution is concerned, we note that B-splines approximate quite well the true distribution of the coefficient except on the tails. This should be expected since we approximate an unbounded distribution with a bounded distribution. The approximation is less accurate when trying to reproduce a coefficient with lognormal distribution, but the general behaviour is captured. In both cases results are better with panel data.

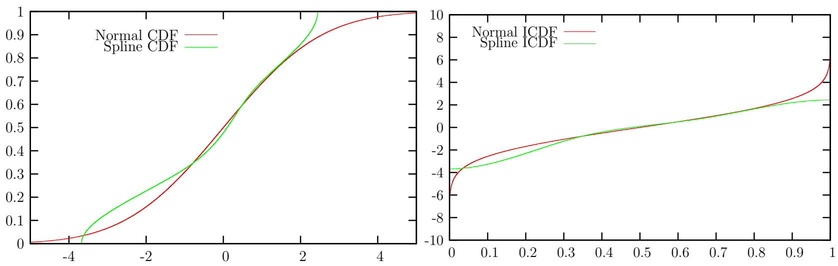

Figure 4: Spline reproducing normal distribution on cross-sectional data

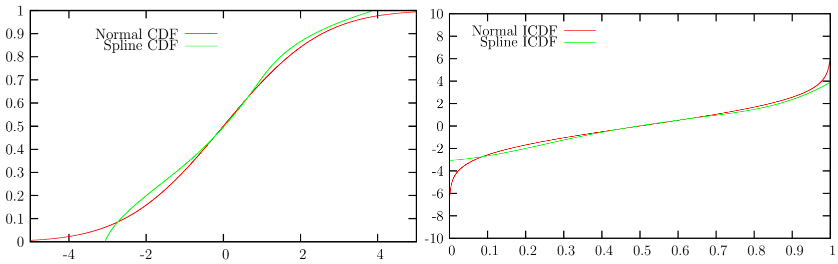

Figure 5: Spline reproducing normal distribution on panel data

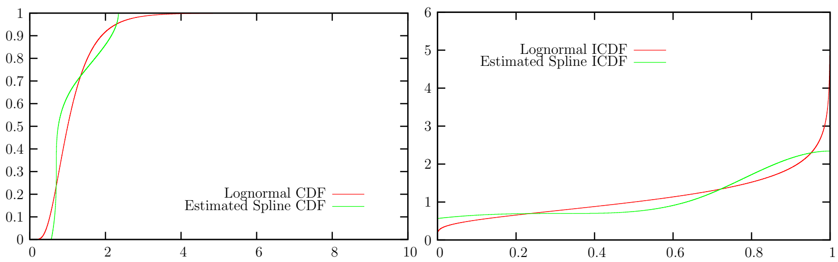

Figure 6: Spline reproducing lognormal distribution on cross-sectional data

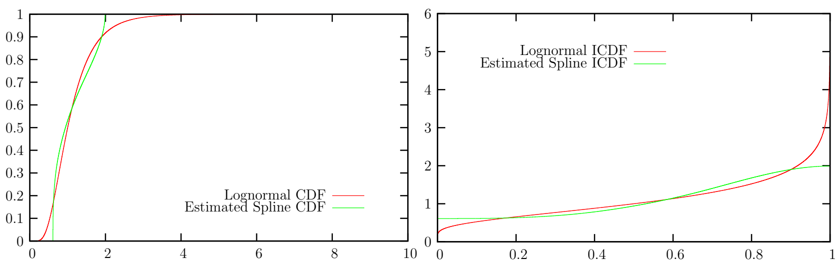

Figure 7: Spline reproducing lognormal distribution on panel data

5 Real case study: IRIS survey

Stated preference data are used to validate our methodology on a real case. The IRIS dataset is derived from a survey conducted in the region of Brussels (Belgium) in the autumn of 2002. The respondents are car users, intercepted during morning peak hours on the ring that gives access to the city from the suburban areas. They were presented with up to three scenarios, each containing four choice options: car, car with delayed departure time, public transport and car on a High Occupancy Vehicles lane (this latest alternative being only prospective). For a more detailed description about the scope of the survey, the way it was administrated, the design and the principal findings we refer the reader to Bastin, Cirillo and Toint [3]. The final mode choice model is estimated on 2602 observations belonging to 871 individuals. The original model specification contained 18 exogenous variables, of which seven were assumed to be randomly distributed; in particular the two components of time (congested travel time and free-flow travel time) were assumed to be lognormal and cost was kept constant. Here, we extend the original analysis to normal and non-parametric distribution (based on B-Spline) for both time and cost coefficients. In total six model specifications have been estimated:

(1) times normally distributed, cost constant (T N);

(2) times and cost normally distributed (T-C N);

(3) times log-normally distributed, cost constant (T L);

(4) times and cost log-normally distributed (T-C L);

(5) times B-Spline distributed, cost constant (T BS);

(6) times and cost B-Spline distributed (T-C BS).

For the B-Spline coefficients, seven coefficients (P1,P2,...,P7) have been estimated, where P_1 and P_7 give the bounds of the distribution, and the knot vector is defined on the percentiles 0, 0.25, 0.5, 0.75, and 1. Monte Carlo simulations based on 2000 random draws per individual have been adopted to calculate the maximum likelihood. The results provided ihave been averaged over ten simulations. All the runs have been executed with a modified version of AMLET, available in open source from the site http://www.grt.be/amlet. The coefficient estimates are reported in Table 1. Due to space limitations we do not report t-statistics, but all the coefficients except the alternative specific constant of public transport are significant at 5% level. These results confirm the existence of both time and cost taste heterogeneity across the population. The fit of the model improves when we replace lognormal distributions with normal distributions; the use of non-parametric distributions reinforces that trend (Table 2). When we apply normal distribution to both congested (Cong T Time) and free-flow travel time (FF T Time) parameters we obtain that for the first component about 16% and for the second about 20% of the population have positive values. These percentages do not change much when we estimate non-parametric distributions for both congested and free-flow travel time. However, B-Spline indicates a reduction of positive values for cost coefficient from the 31.6% obtained with normal distribution to 11.9%, a large proportion of the population having a cost coefficient close to 0. In order to compute the values of travel time savings (VTTS), we simply draw on time and cost, and compute the corresponding ratio. The effects of the different distributions adopted on values of travel time savings (VTTS) are shown on Table 3 where the 25th, 50th and 75th percentiles are calculated; the resulting distributions and the inverse of the cumulative distribution function are presented on Figure 10 and Figure 11, where we used 750.000 draws; for the non-parametric distributions we also present the values corrected by truncating the distributions at zero Table 4. For completeness, we also present the mean VTTS, defined as the ratio $60E[\mathrm{time}]/E[\mathrm{cost}]$. Note that in order to be able to compare all distributions, we transform the lognormal CDF (only defined on $\mathcal{R}^+$, and estimated using the opposite of associated observations) to its symmetric distribution with respect to the vertical axis on 0.

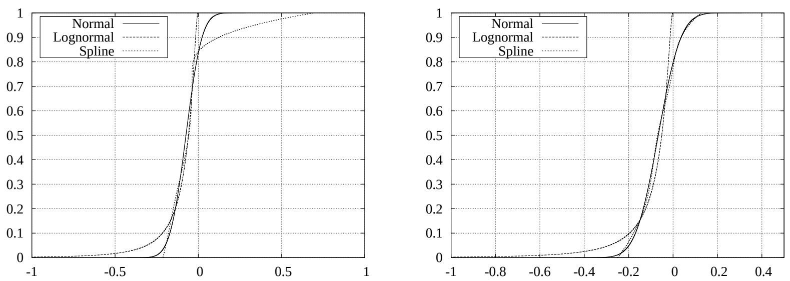

Figure 8: Congested Travel Time and Free-Flow Travel Time CDF

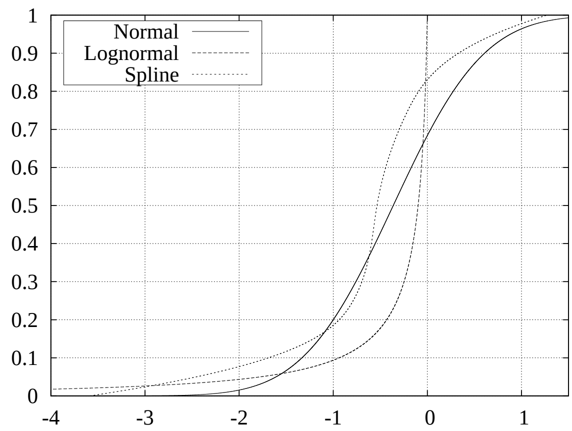

Figure 9: Cost CDF

We observe quite a lot of variability in VTTS values, especially when cost is assumed to be random. Lognormal distribution applied to both time and cost parameters

Parameter	T N	T-C N	T L	T-C L	T BS	T-C BS
Car Passenger(CP) μ	-1.252	-1.305	-1.153	-1.151	-1.226	-1.289
HOV (HOV) μ	-4.773	-4.903	-5.253	-5.358	-4.606	-4.888
HOV (HOV) σ	4.668	4.759	4.919	4.954	4.583	4.897
Shared car on HOV (HOVs) μ	-6.704	-7.138	-7.259	-7.741	-6.561	-7.321
Shared car on HOV (HOVs) σ	6.009	6.149	6.372	6.554	5.904	6.413
Public Transport (PT) μ	-0.869	-0.958	-0.839	-0.841	-0.824	-0.817
Cong T Time P1	-0.074	-0.078	-2.822	-2.786	-0.212	-0.250
Cong T Time P2	0.076	0.077	1.004	1.003	-0.182	-0.204
Cong T Time P3					-0.141	-0.152
Cong T Time P4					-0.041	-0.045
Cong T Time P5					-0.036	-0.039
Cong T Time P6					-0.036	-0.029
Cong T Time P7					0.696	0.656
FF T Time P1	-0.066	-0.070	-2.976	-2.929	-0.249	-0.270
FF T Time P2	0.080	0.083	1.051	1.078	-0.178	-0.225
FF T Time P3					-0.095	-0.093
FF T Time P4					-0.080	-0.083
FF T Time P5					0.010	0.002
FF T Time P6					0.013	0.025
FF T Time P7					0.126	0.078
Cost P1	-0.291	-0.363	-0.282	-2.318	-0.309	-3.587
Cost P2		0.757		1.758		-1.345
Cost P3						-0.6086
Cost P4						-0.584
Cost P5						-0.267
Cost P6						0.187
Cost P7						1.269
Toll (HOV) μ	-0.494	-0.499	-0.508	-0.513	-0.493	-0.359
Dist.(CD,CP,CDs,CPs,HOV,HOVs) μ	0.192	0.212	0.214	0.238	0.186	0.218
Dist.(CD,CP,CDs,CPs,HOV,HOVs) σ	0.236	0.253	0.251	0.274	0.239	0.257
Trip Freq. -once a week (PT) μ	3.423	3.982	3.235	3.0746	3.482	3.701
Comfort no-seats (PT,PTs) μ	-1.211	-1.409	-1.169	-1.298	-1.272	-1.411
Comfort crowded (PT,PTs) μ	-1.872	-2.0409	-1.877	-2.024	-1.906	-2.059
Earlier Dep. Time (CP,CPs) μ	-2.836	-2.939	-3.233	-3.302	-2.703	-3.009
Earlier Dep. Time (CP,CPs) σ	2.559	2.594	2.878	2.876	2.506	2.723
Later Dep. Time (CP,CPs) μ	-1.908	-2.054	-2.442	-2.586	-1.887	-2.129
Later Dep. Time (CP,CPs) σ	2.026	2.145	2.575	2.728	2.0885	2.277
Much Later Dep. Time (CP,CPs) μ	-2.464	-2.615	-2.685	-2.808	-2.390	-2.647
Self-employed (CD-HOV) μ	1.763	1.740	1.878	1.799	1.669	1.699
Manager (HOV) μ	1.206	1.172	1.272	1.306	1.203	1.187
Num. cars -3 per HHLD (CD) μ	2.297	2.437	2.027	2.112	2.237	2.375

Table 1: Real data: IRIS parameters estimation

Dist.	T N	T-C N	T L	T-C L	T BS	T-C BS
Loglikelihood β	-3.1460	-3.1399	-3.1604	-3.1511	-3.1453	-3.1339

Table 2: Real data: Final log-likelihood values for the six models estimated

	Quant.	T N	T-C N	T L	T-C L	T B-S	T-C BS
Cong T Time	25%	4.75	3.36	6.41	16.10	7.20	2.42
	50%	15.31	4.52	12.64	37.61	11.05	4.36
	75%	25.89	10.49	24.88	87.77	26.12	16.68
	Mean	15.323	12.96	20.90	13.17	7.90	4.74
FF T Time	25%	2.56	2.10	5.33	14.03	0.89	0.70
	50%	13.72	6.24	10.84	32.65	13.53	3.07
	75%	24.91	11.35	22.02	75.74	22.18	14.35
	Mean	13.74	11.61	18.798	12.35	12.64	7.47

Table 3: Real data: value of travel time savings

	Quant.	T N (corr.)	T-C N (corr.)	T BS (corr.)	T-C BS (corr.)
Cong T Time	25%	10.25	4.65	7.84	4.56
	50%	18.54	5.59	15.03	6.95
	75%	27.98	12.54	28.47	17.01
	Mean	19.93	7.94	18.35	7.83
FF T Time	25%	9.64	4.40	9.55	3.96
	50%	17.97	4.87	17.11	5.60
	75%	27.68	12.62	25.67	15.23
	Mean	19.58	7.92	18.59	7.65

Table 4: Value of travel time savings from B-Spline truncated at zero

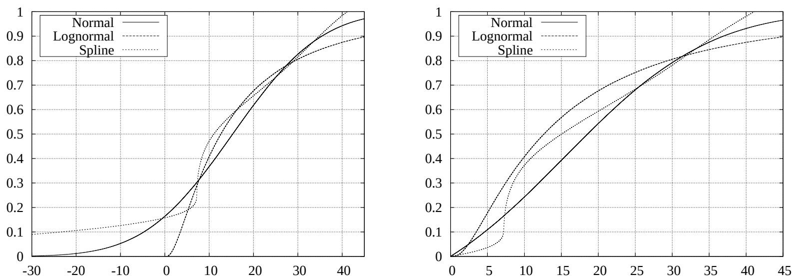

Figure 10: Value of Congested Travel Time

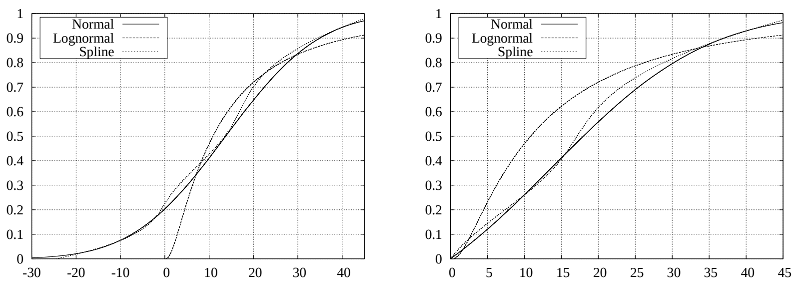

Figure 11: Value of Free-Flow Travel Time

produce high VTTS values for the known effect of fat tails [16]. Normal distributions result into low VTTS values given the relatively high percentages of "wrong" positive values for time and cost. Non-parametric distributions produce VVTS values that are close to those obtained with normal distributions; in particular non-parametric distributions on time and cost result into a congested VTTS higher than free-flow VTTS (as expected). Computational time does increase with model flexibility; when three parameters are specified as non-parametric optimization time is about 74 minutes on a MacBook Pro, which is 6 times higher than the time required to estimate the same model with three normal distributions instead. Still, these values are very reasonable especially if we consider that improvements are possible by adopting quasi-Monte Carlo draws [5] or an adaptive algorithm as in Bastin et al. [3], as well as various refinements of the current code. To summarize, we found that non-parametric B-Splines (a) on both time and cost coefficients provide the best fit, (b) reduce significantly the percentages of the population showing positive values for cost but leave unchanged the proportion of positive time values, (c) give VTTS ranges that do not suffer from fat tail effects. This suggests that the lognormal assumption, even if it more coherent with the econometric theory, is not reasonable here, and the non-parametric approach has the advantage over the normal to bound the distributions.

6 Conclusion

The estimation of heterogeneity in travel time saving has become one of the most discussed subjects in travel behaviour. This problem is often approached with advanced demand models that allow the estimation of random coefficients with parametric distribution. This approach is not without drawbacks, especially the choice of the "best" distribution is not a trivial task. In this paper we have proposed to turn to non-parametric methods by adopting B-spline curves as polynomial approximations of arbitrary distributions, and we have implemented them into a classical mixed logit formulation. Constrained optimization methods are used to deal with the monotonicity of the inverse of the cumulative distribution functions. The method has been applied to both simulated and real data. The results from simulated data show that B-Spline have been able to recover the initial and known parametric distributions. We have also used real data to validate our methodology and found that for the IRIS data set not only the goodness of fit of the non-parametric model is better, but that it also gives VTTS distributions that would be very difficult to recover with classic parametric distributions. Thus parametric approach can fail to detect the real distribution and non-parametric random variables could guide the analysts in search for the real shape of the coefficients distributions. The proposed approach is highly flexible and can handle more than one random parameter at a time. However, improved model flexibility comes at the expense of a (moderate) increase in computational costs, that are higher to those needed to run classical parametric mixed logit models. Finally, most methods to generate multivariate distributions rely on the marginal functions and some dependence treatment between these marginals. These methods usually exploit the inversion technique to generate these marginals, so a natural extension of the proposed method will be to examine how to apply it to such multivariate random vector generation techniques. This issue, together with recent applications proposed in econometrics to study state dependency, will be part of our future investigations.

Acknowledgments

The authors would like to express their gratitude to Thierry Duquenne from Région de Bruxelles Capitale for granting access to the IRIS data set.

References

[1] Miriam Ayer, H.D. Brunk, G.M. Ewing amd W.T. Reid, and E. Silverman. An empirical distribution function for sampling with incomplete information. *Annals of Mathematical Statistics*, 26:641–648, 1955.

[2] H. X. H. Bao and A.T.K. Wan. On the use of spline smoothing in estimating hedonic housing price models: Empirical evidence using hong kong data. *Real estate economics*, 32(3):487–507, 2004.

[3] Fabian Bastin, Cinzia Cirillo, and Philippe L. Toint. Application of an adaptive Monte Carlo algorithm to mixed logit estimation. *Transportation Research Part B*, 40(7):577–593, 2006.

[4] M. J. Best and N. Chakravarti. Active set algorithms for isotonic regression: a unifying framework. *Mathematical Programming*, 47, 1990.

[5] Chandra R. Bhat. Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research*, 35B(7):677–693, August 2001.

[6] Cinzia Cirillo and Kay W. Axhausen. Evidence on the distribution of values of travel time savings from a six-week diary. *Transportation Research Part A*, 40:444–457, 2006.

[7] Andrew R. Conn, Nicholas I. M. Gould, and Philippe L. Toint. *Trust-Region Methods*. SIAM, Philadelphia, USA, 2000.

[8] Luc Devroye. *Non-Uniform Random Variate Generation*. Springer-Verlag, 1986.

[9] X. Dong and Frank S. Koppelman. Mass point mixed logit model: Development and application. paper presented at the 10th International Conference on Travel Behaviour Research, Lucerne, Switzerland, 2003.

[10] R. F. Engle, C. W. J. Granger, J. Rice, and A. Weiss. Semiparametric estimates of the relation between weather and electricity. *Journal of the American Statistical Association*, 81:310–320, 1986.

[11] G. Farin. *NURBS for curve and surface design*. SIAM, 1991.

[12] M. Fosgerau and Michel Bierlaire. A practical test for the choice of mixing distribution in discrete choice models. *Transportation Research Part B*, 41:784–794, 2007.

[13] Mogens Fosgerau. Investigating the distribution of the value of travel time savings. *Transportation Research Part B*, 40:688–707, 2006.

[14] John Fox. *Multiple and generalised non-parametric regression*. Sage Publications, Inc, 2000.

[15] David A. Hensher. The signs of the times: Imposing a globally signed condition on willingness to pay distributions. *Transportation*, 33(3):205–222, 2006.

[16] David A. Hensher and William H. Greene. The mixed logit model: The state of practice. *Transportation*, 30(2):133–176, 2003.

[17] Stéphane Hess, Michel Bierlaire, and John W. Polak. Discrete mixture of gev models. Paper presented at 5th Swisss Transport Research Conference, 2005.

[18] Stéphane Hess, Michel Bierlaire, and John W. Polak. Estimation of value of travel-time savings using mixed logit models. *Tranportation Research Part A*, 39(3):221–236, 2005.

[19] C. H. Jarvis and N. Stuart. A comparison among strategies for interpolating maximum and minimum daily air temperatures, part ii: the interaction between the number of guiding variables and the type of interpolation method. *Journal of Applied Meteorology*, 40:1075–1084, 2001.

[20] R. Klein and R. Spady. An efficient semi-parametric estimator for binary response models. *Econometrica*, 61(2):387–422, 1993.

[21] Roger Koenker, Pin Ng, and Stephen Portnoy. Quantile smoothing splines. *Biometrika*, 81:673–680, 1994.

[22] Averill M. Law. *Simulation Modeling & Analysis*. McGraw-Hill, 4th edition edition, 2007.

[23] Pierre L'Ecuyer. Efficiency improvement and variance reduction. In *Proceedings of 1994 Winter Simulation Conference*, pages 122–132, Orlanda, USA, 1994.

[24] L. A. Piegl and W. Tiller. *The NURBS Book*. Springer-Verlag, New York, NY, USA, second edition, 1996.

[25] G. D. Singh, J. A. McNamara, and S. Lozanoff. Morphometry of the cranial base in subjects with class iii malocclusion. *Journal of Dental Research*, 76:694–703, 1997.

[26] Kenneth Train. *Discrete Choice Methods with Simulation*. Cambridge University Press, New York, USA, 2003.

[27] Kenneth Train and G. Sonnier. Mixed logit with bounded distributions of correlated partworths. In A. Alberini and R. Scarpa, editors, *Applications of Simulations Methods in Environmental Resource Economics*, chapter 7. Kluwer Academics Publisher, Dordrecht, The Netherlands, 2005.

[28] Kenneth Train and M. Weeks. Discrete choice models in preference space and willingness-to-pay space. In A. Alberini and R. Scarpa, editors, *Applications of Simulations Methods in Environmental Resource Economics*, chapter 7. Kluwer Academics Publisher, Dordrecht, The Netherlands, 2005.