TRANSPORTATION SCIENCE
Vol. 44, No. 4, November 2010, pp. 537–549
ISSN 0041-1655 | EISSN 1526-5447 | 10 | 4404 | 0537

DOI 10.1287/trsc.1100.0321

Estimating Nonparametric Random Utility Models with an Application to the Value of Time in Heterogeneous Populations

Fabian Bastin
Department of Computing Science and Operational Research, University of Montréal, Montréal, Québec H3C 3J7, Canada, and CIRRELT, Department of Computing Science and Operational Research, University of Montréal, Montréal, Québec H3C 3J7, Canada, bastin@iro.umontreal.ca

Cinzia Cirillo
Department of Civil and Environmental Engineering, University of Maryland, College Park, Maryland 20742, ccirillo@umd.edu

Philippe L. Toint
Department of Mathematics, University of Namur, B5000 Namur, Belgium, philippe.toint@fundp.ac.be

The estimation of random parameters by means of mixed logit models is now current practice for the analysis of transportation behaviour. One of the most straightforward applications is the derivation of willingness-to-pay distribution over a heterogeneous population, an important element for dynamic tolling strategies on congested networks. In numerous practical cases, the underlying discrete choice models involve parametric distributions that are a priori specified and whose parameters are estimated. This approach can however lead to many problems for realistic interpretation, such as negative value of time, etc.

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 nonparametric random vectors is tested. The nonparametric mixed logit model is also used on real data derived from a stated preference survey conducted in the region of Brussels (Belgium). The model presents multiple choices and is estimated on repeated observations. The obtained results provide a more realistic interpretation of the observed behaviours.

Key words: mixed logit; nonparametric estimation; B-spline; constrained optimization
History: Received: December 2007; revision received: April 2009; accepted: January 2010. Published online in *Articles in Advance* May 24, 2010.

1. Introduction

In strategic transportation planning, one is very often led to analyzing the behaviour of a population faced with new transport alternatives or constraints. Such an analysis is crucial both for estimating economic viability of the projects under study and for measuring public acceptance of new policies. A crucial tool for this approach is supplied today by discrete choice modelling, a technique whose purpose is to estimate the probablity of choice between a set of (transportation) alternatives for members of a possibly heterogeneous population. This technique has been used in a large number of transportation projects, such as travel mode choice (Bhat 1998), destination choice (Daly 1982), route choice (Cascetta et al. 2002; Bierlaire and Frejinger 2008), air travel choice (Proussaloglou and Koppelman 1999), activity analysis (Srinivasan and Bhat 2005), car ownership, and brand and model choice (Hensher et al. 1992) among many others. The method used is to associate to each individual and to each alternative a so-called utility function, which describes the relative advantages and drawbacks of the alternative as perceived by the individual. Discrete choice models come in a variety of types, whose suitability for a specific purpose generally depends on the nature of the problem at hand. In particular, heterogeneity of the population under study calls for methods that are capable of taking this very heterogeneity into account, typically by introducing random elements in the model. Mixed-logit approaches are among the most elaborate and popular tools of this type. In such models, investigators traditionally use parametric distributions involving specific functional forms and a finite number of unknown parameters. Among them, some are considered as random variables to reflect population heterogeneity. These techniques have been pioneered by McFadden and Train (2000) and have gained wide acceptance among practitioners, especially in the transportation field (see Hensher and Greene 2003 for a relevant survey).

The early applications of mixed logit have mainly focussed on using normal and other existing distributions, which led to a number of practical difficulties. The first is the lack of criteria for assessing which is the contextually more appropriate among a wide range of analytical distributions. A second problem is that unbounded distributions, although popular, often produce value ranges with difficult behavioural interpretation.[1] And, finally, little is known about the tails and their effects on the mean of the estimates (Hess, Bierlaire, and Polak 2005b; Cirillo and Axhausen 2006). Unbounded distributions therefore appear to be inappropriate in many cases, in particular because certain attributes are assumed to be positively (or negatively) valued by all individuals. For instance, estimating the willingness to pay for tolled usage of transportation infrastructure can be problematic due to zero or positive time and cost coefficients in the user's derived utility functions. To circumvent these difficulties, more recent models use bounded distributions, often obtained as simple transformations of normals. Train and Sonnier (2005) specify mixed logit models with lognormal, censored normal, and Johnson Sb distributions bounded on both sides. They also suggest to adopt Bayesian procedures in order to avoid estimation problems encountered with lognormal distributions parameters. A similar approach is followed by Sillano and Ortúzar (1999), where Bayesian techniques are used to estimate individual parameters; however, the choice between classical and Bayesian methods is not always clear (see Huber and Train 2001). Moreover, as indicated by Jackman (2000), the use of Bayesian techniques and associated software is not without difficulties, in particular in detecting convergence of the sampling process. At variance, the mixed logit methods considered in this paper enjoy a very strong convergence analysis (see Bastin, Cirillo, and Toint 2006b).

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 (2003) 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 shows 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 (2005a) propose discrete mixture of generalized extreme-value 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 nonlinearity and nonconcavity of the log-likelihood function.

Hensher (2006) 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 (2005) 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 (2006) employs various nonparametric 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 (1993) estimator is first used to estimate parameters in a linear index binary choice model with no assumptions on the error term distribution, then the distribution of the error term is estimated; however, this method does not account for repeated observations and applies only to binomial choices. Recently, Fosgerau and Bierlaire (2007) have proposed a seminonparametric (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 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 nonparametric 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

[1] Barring the undesirable effect of deficiencies in the data.

of the unknown distributions, expressed via their cumulative distributions, as a part of the complete calibration procedure. Because of 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 difficult, and perhaps vain, task. It is nevertheless meaningful to capture the randomness nature of some parameters in order to correctly apprehend heterogeneity inside the population. We will show in this paper that our approach, even based on simple approximations, is suitable for this objective, and also that it can be applied to realistic transportation problems related to congestion pricing and management.

To estimate the shape of the unknown distributions efficiently, we suggest using a B-spline parametrization of the inverse cumulative distribution functions, the associated monotonicity constraints being then explicitly included in the log-likelihood maximization. B-splines are known to provide a concise formulation for curves that are composed of polynomial pieces, thereby automatically controlling the overall curve smoothness (Farin 1991). This technique is often used for nonparametric regression (Fox 2000). Recent applications in such areas as meteorology, medicine, and price modelling can be found in Singh, McNamara, and Lozanoff (1997); Jarvis and Stuart (2001); and Bao and Wan (2004). To date there have only been a handful of applications of this class of functions in econometrics (Engle et al. 1986; Koenker, Ng, and Portnoy 1994), 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 here are assumed to be continuous, bounded, and independent, and we are interested by the inverse cumulative 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 resulting 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. Nonparametric estimation of continuous variables is developed on §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 nonparametric random vectors. Results on a real-case study are given in §5. Conclusions and perspectives for research are finally presented.

2. Mixed Multinomial Logit Model Formulation

The mixed multinomial logit (MMNL) formulation is at the present time extensively used in behavioural analysis for transport modelling, mostly 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 now briefly review their essential concepts.

We consider a set of I individuals, each one having to choose one alternative within a finite set $\mathcal{A}_j$. We associate a 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 at maximizing their utility, but we do not observe all its components. Instead, we decompose the utility U_{ij} as the sum of a deterministic part $V_{ij}(\beta)$, where β is a vector to be estimated, and a random, unobserved part ϵ_{ij}. The probability $L_{ij}(\beta)$ that individual i chooses alternative j is then the probability that $U_{ij} \geq U_{in}$ for all A_n in $\mathcal{A}(i)$, that is

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

This 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 among the individuals and alternatives, we obtain the traditional logit probability.

In the mixed logit framework, we relax the assumption that β is a constant vector, but instead is a random vector B. As a result, the probability choice L_{ij} is now conditional on the realization β of this random vector, and the unconditional probability is then given by

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

where $E_B(\cdot)$ is the expectation with respect to the distribution of B, and P_B the probability associated with this distribution. We therefore cannot directly estimate B, so we will additionally assume that it can be parametrized as $B = B(\Gamma, \theta)$, where Γ is some random vector with a known distribution (a normal for instance), and θ is some constant parameter vector to be estimated. In other terms, we assume some distribution family for B, parameterized by θ. If, moreover, the vector B is continuous, we can rewrite (1) as

$$\begin{aligned} P_{ij}(\theta) &= \int L_{ij}(\gamma, \theta)\phi_B(\gamma, \theta)\, d\gamma \\ &= \int \frac{e^{V_{ij}[\beta(\gamma,\theta)]}}{\sum_l e^{V_{il}[\beta(\gamma,\theta)]}} \phi_B(\gamma, \theta)\, d\gamma, \end{aligned}$$

where γ is a realization of the random vector Γ, $\phi_B(\gamma, \theta)$ is the density of $B(\Gamma, \theta)$, with parameters vector θ and the index l covers all alternatives.

In the case where 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 that the heterogeneity is present on the population level only, but not on the individual, level. The probability to observe the individual's choices is then given by the product of logit probabilities $L_{ij_{it}}$ (see Train 2003; see also Sillano and Ortúzar 1999):

$$P_{iy_i}(\theta) = \int \left(\prod_{t=1}^{T_i} L_{ij_{it}}(\gamma, \theta)\right) \phi_B(\gamma, \theta)\, d\gamma$$
$$= \int \left(\prod_{t=1}^{T_i} \frac{e^{V_{ij_{it}}[\beta(\gamma,\theta)]}}{\sum_l e^{V_{il}[\beta(\gamma,\theta)]}}\right) \phi_B(\gamma, \theta)\, d\gamma.$$

2.1. MMNL Model Estimation

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

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

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 because 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

$$\mathrm{SP}^R_{iy_i} = \frac{1}{R} \sum_{r_i=1}^{R} \prod_{t=1}^{T_i} L_{ij_{it}}(\gamma_{r_i}, \theta), \tag{3}$$

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

$$\max_\theta \mathrm{SLL}^R(\theta) = \max_\theta \frac{1}{I} \sum_{i=1}^{I} \ln \mathrm{SP}^R_{iy_i}(\theta). \tag{4}$$

We will denote by θ^*_R one solution of this last approximation (often called sample average approximation, or SAA), whereas θ^* denotes the solution of the true problem (2).

3. Nonparametric Estimation of Continuous Variables

The procedure outlined in the previous paragraph is consistent and applicable, provided one finds a reasonable way to choose the distribution $\phi_B(\gamma, \theta)$ for the unknown random vector B. A first approach is by using a parametric distribution for ϕ_B whose parameters are adjusted by trial and error, the quality of which is then measured by the final likelihood value and by the interpretability of the results (such as value of time). However this technique is time consuming and prone to misinterpretation, and a more formal approach remains desirable. One possible such approach is to use discrete distributions for ϕ_B. Such a discrete treatment however leads to an arbitrary population segmentation, which can be avoided if we turn to continuous distributions, which we find preferable.

In our proposal, each component of the vector B inherent to the mixed logit function is itself random, and if we assume independence between these components, can therefore be considered each one separately. Usefully, this allows us to draw from univariate random variables. More specifically, if X is an univariate random distribution, a well-known technique to generate draws from its distribution consists in sampling a uniform distribution on $[0,1]$, hereafter denoted by $U[0,1]$, and in applying the inverse cumulative distribution function F_X^{-1} to these draws, yielding

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

where S_X represents 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 literature (Devroye 1986; Law 2007), and is also popular in the context of variance reduction methods (see for instance L'Ecuyer 1994).

We propose to 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}: $[0,1] \to \mathcal{R} \cup \{-\infty\} \cup \{+\infty\}$; and
- F_X^{-1} is nonincreasing.

For simplicity, we restrict ourselves to continuous variates, so that F_X^{-1} is continuous and strictly increasing. In other terms, we have to estimate an arbitrary continuous real function whose domain is $[0,1]$, and which is monotonically increasing, while seeking an adequate balance between estimation capabilities and satisfaction of the conditions ensuring that we can interpret this 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.

If we furthermore assume that the random variable X has a bounded support, an elegant way to achieve the sought balance is the use of B-spline functions. The bounded support assumption is not too restrictive in practice, because extreme behaviours, corresponding to values of X close to plus or minus infinity, are usually not welcome because they are difficult to interpret. We therefore consider the bounded support assumption as an advantage rather than a drawback of our proposition. We thus propose to use the B-spline approximation of degree three given by

$$F_X^{-1}(u) \approx C(u) = \sum_{i=0}^{n} \pi_i N_{i,3}(u),$$

where the piecewise cubic polynomial functions $N_{i,3}(u)$ form an easily computable basis for B-spline functions in $[0, 1]$ (see Farin 1991). The coefficients $|\pi_0, \pi_1, \ldots, \pi_n$ are called the control points, and the functions $N_{i,3}(u)$ depend on the choice of special points, called knots, in $[0, 1]$ (we use here a classical knot type, known as nonperiodic, or clamped or open). It is then possible to show that the approximating function $C(u)$ is twice continuously differentiable and that, with these basis and knots choices, $C(u)$ is monotonically increasing if the control points have the same property, that is if $\pi_0 \leq \pi_1 \leq \cdots \leq \pi_n$. As we will describe in the next section, this property can be algorithmically guaranteed. The resulting spline construction is illustrated in Figure 1.

For a more complete review of B-splines properties, we refer the reader to Piegl and Tiller (1996).

3.1. Constrained Optimization

If we now use our B-splines approximations for the inverse cumulative distributions associated with the random vector B, the random distribution Γ according to which the γ_{r_i} are drawn in (3) now also depends on the parameters used to define the B-spline approximation itself, namely the control points and the positions of the nonparametric knots in $[0, 1]$. Because it is notoriously difficult to estimate knot positions from data, we assume in what follows that these are fixed. The variables of the maximum likelihood problem (4) may now be viewed as consisting of two subsets corresponding to the behavioural parameters in the model and to the $n \times s$ dimensional vector $\pi = \{\{\pi_{di}\}_{i=0}^{n}\}_{d=1}^{s}$ (the control points of the B-spline approximation), where the index i in this last expression ranges over the control points and the index $d = 1, \ldots, s$ ranges over the random coefficients in our problem. In other words, we may define $\theta = (\tau, \pi)$, where τ is the vector of behavioural parameters. The maximization problem now takes the form

$$\max_{\tau,\, \pi \in C} f(\tau, \pi),$$

where $f(\cdot, \cdot)$ is some complicated function given by (4), and where

$$C = \prod_{d=1}^{s} \{\pi_{d0} \leq \pi_{d1} \leq \cdots \leq \pi_{dn}\}$$

is the feasible region defined by the monotonicity constraints on the B-splines control points.

For the benefit of the algorithmically minded reader, we now outline the carefully taylored numerical algorithm which we propose for solving this constrained maximization problem. To this aim, we temporarily assume, for simplicity, that we only have one nonparametric coefficient ($s = 1$), so that C defines n ordered variables. The vector C is then called the order simplex, which is illustrated in Figure 2 for the three-dimensional case. The projection onto the order-simplex can be performed easily and efficiently, because several algorithms of complexity $O(n)$ have been designed (Ayer et al. 1955; Best and Chakravarti 1990). Moreover, it is possible to adapt a well-known approach for nonlinear optimization, the trust-region

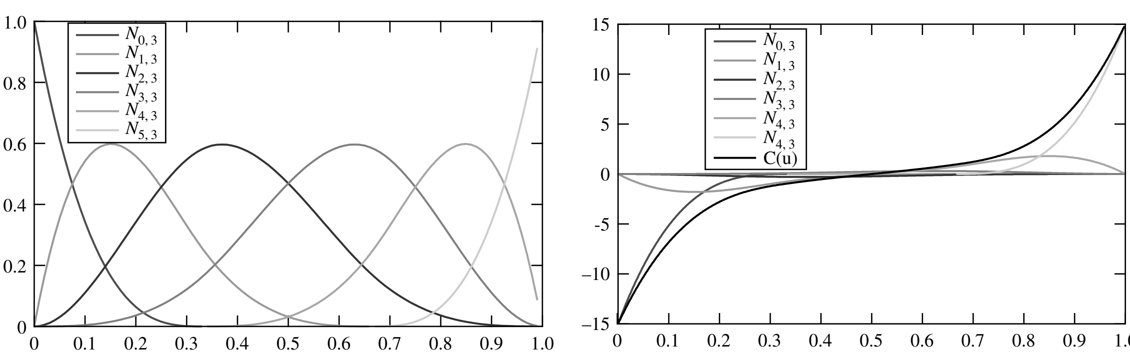

Figure 1 Basis B-Splines and Monotonically Increasing Spline

Transportation Science 44(4), pp. 537–549, © 2010 INFORMS

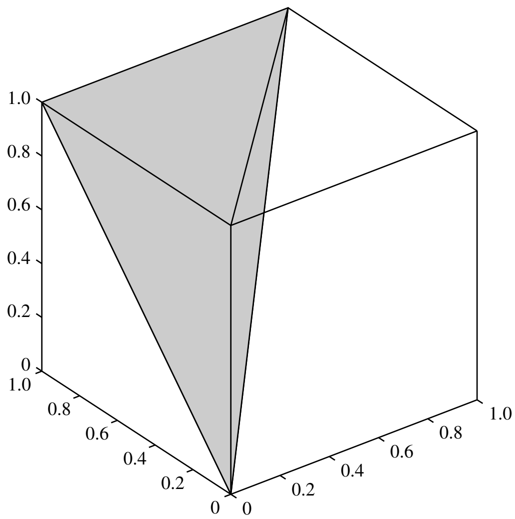

Figure 2 **The Order Simplex in $\mathscr{R}^3$**

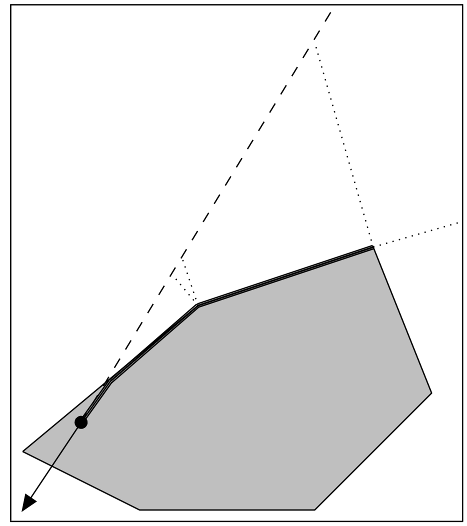

Figure 3 **Projected Gradient Path**

algorithm, to benefit from such projections (see Conn, Gould, and Toint 2000). The main idea of a trust-region algorithm involves the calculation, at iteration k (with current estimate (τ_k, π_k)), of a trial point $(\tau_k + \Delta\tau_k,\ \pi_k + \Delta\pi_k)$ by approximately maximizing a model m_k of the objective function inside a trust region defined as

$$\mathscr{B}_k = \big\{(\tau, \pi) \in \mathscr{R}^{|\beta|+sn},\\ \text{such that } \|(\tau, \pi) - (\tau_k, \pi_k)\|_2 \le \Delta_k\big\},$$

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

$$m_k(p) = p^T \nabla_\theta \,\mathrm{SLL}^R(\tau_k, \pi_k) + \tfrac{1}{2} p^T H_k p, \qquad (5)$$

where H_k is a symmetric approximation of the Hessian $\nabla^2_{\theta\theta}\,\mathrm{SLL}^R(\tau_k, \pi_k)$. The step $p_k = (\Delta\tau_k, \Delta\pi_k)^T$ is computed by first attempting to identify the active constraints by maximizing the model (5) along the "projected gradient path." Once these active constraints have been guessed, the model $m(p)$ is further maximized in the intersection of corresponding face of C with $\mathscr{B}_k$, using an approximate projected truncated conjugate-gradient technique (see Toint 1981; Steihaug 1983), as illustrated in Figure 3. The predicted and actual increases in the value of the objective function are then compared by computing the ratio

$$\rho_k = \frac{\mathrm{SLL}^R(\tau_k + \Delta\tau_k, \pi_k + \Delta\pi_k) - \mathrm{SLL}^R(\tau_k, \pi_k)}{m_k(s_k) - m_k(0)}.$$

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.[2]

[2] More precisely, if ρ_k is larger than 0.75, we set the trust region Δ_{k+1} to be the maximum between Δ_k and $2\|p_k\|$; otherwise, we set $\Delta_{k+1} = 0.5\Delta_k$.

If the ratio is below the bound, the trial point is rejected and the trust region is shrunk by a factor of two, in order to improve the correspondence of the model with the true objective function, thereby concluding the iteration. Details of the strong convergence theory and constraint identification properties for the proposed algorithm may be found in (Conn, Gould, and Toint 2000), which support our experimental observation that the method is robust in practice.

4. Simulations

Our first experiment aims at verifying by simulation that known distributions can be approximately recovered by our estimation technique. For this verification we have created two synthetic populations; the first data set is cross-sectional and simulates 2,000 individuals giving just one response, the second data set is a panel of 1,000 individuals contributing two observations each. The design contains four alternatives and one independent variable normally distributed with parameters $N(0.5, 1)$. We ran two simulations on each of the data sets described and built the associated mixed logit models. The first simulation supposes that the coefficient to be estimated is normally distributed with parameters $N(0, 4)$, and the second assumes that the coefficient is lognormally distributed with mean 1.133 and standard deviation 0.604. The spline estimated is constructed using a knot vector based on the percentiles 0, 1/3, 2/3, and 1. Once the mixed logit models were estimated, we retrieved the estimated distribution of B and compared it with its true description. Results are illustrated in Figures 4–7, where, for each of the four simulated cases, we report on the left the two cumulative distribution functions

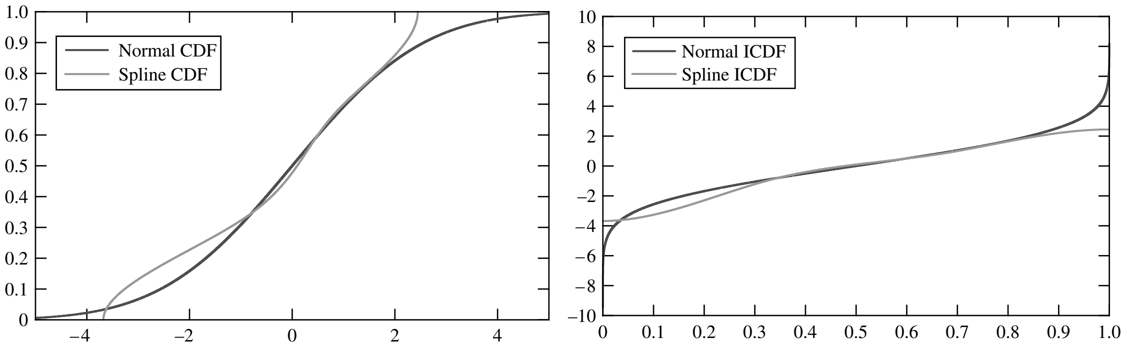

Figure 4 Spline Reproducing Normal Distribution on Cross-Sectional Data

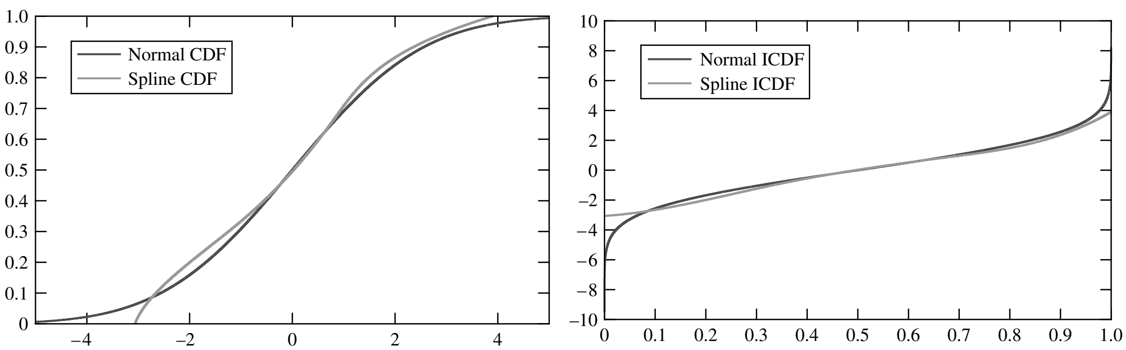

Figure 5 Spline Reproducing Normal Distribution on Panel Data

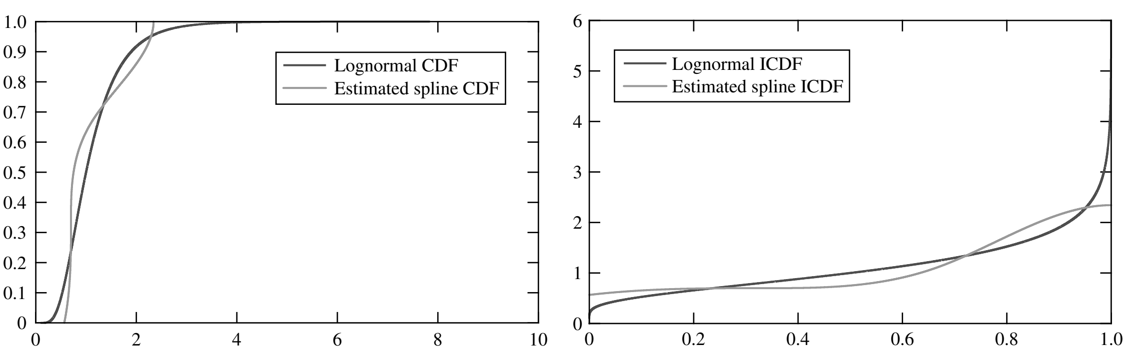

Figure 6 Spline Reproducing Lognormal Distribution on Cross-Sectional Data

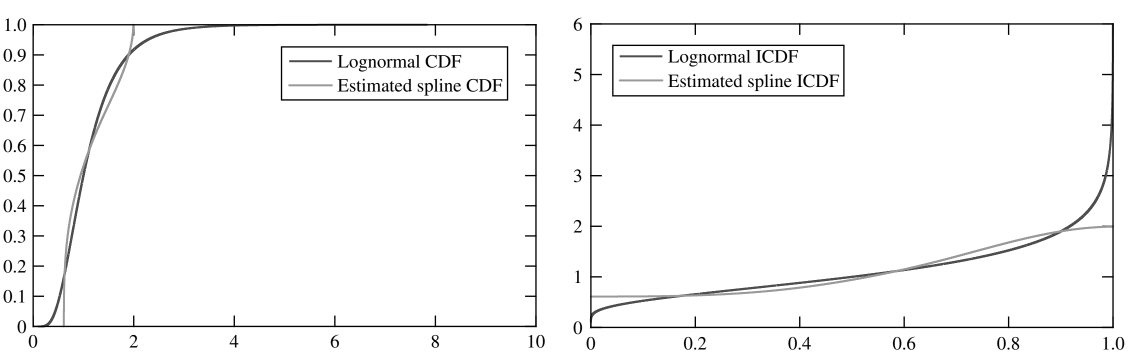

Figure 7 Spline Reproducing Lognormal Distribution on Panel Data

(CDF) and on the right the respective inverse cumulative distribution functions (ICDF). Examining these figures, one may conclude that B-splines approximate the normal distribution well, except for the tails. This should be expected because we approximate an unbounded distribution with a bounded one. The approximation is less accurate when trying to reproduce a coefficient with lognormal distribution, but the general behaviour is captured nevertheless. In both cases results are better with panel data.

5. Real-Case Study: Induction, Research, and Information Skills Survey

We now turn to testing our new methodology in the more practical context, and use stated preference data to validate our methodology on a real case. The IRIS project covers daily mobility in the Brussels (Belgium) region at large and its associated data set is derived from a survey conducted in this region during the autumn 2002. The respondents are car users, intercepted during morning peak hours on the highway ring that gives access to the city from the suburban areas. Special care was exercized in collecting the data in view of the problems arising from behaviourally inconsistent values of estimated parameters. They were presented with up to three scenarios, each containing four transport mode 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, its method of administration and design, and its principal findings we refer the reader to Bastin, Cirillo, and Toint (2006a). The mode choice model presented in this reference is estimated on 2,602 observations belonging to 871 individuals, and its original specification contains 18 exogenous variables, of which seven are assumed to be randomly distributed. In particular, the two components of time (congested travel time and free-flow travel time) were assumed to be lognormally distributed and cost was kept constant. In the analysis presented here, we consider exactly the same data set and started by attempting to use a purely deterministic approach (a logit model), knowing a priori that heterogeneity could not be represented in such a model. The results were nevertheless globally interpretable and the final value of the log likelihood was −4.1765 (the value at zero being −4.6292). This provides a sound basis for comparison, but our purpose in this section is to show that this can be considerably improved by taking population heterogenity into account, using our new modelling technique. We therefore extend the original analysis based on lognormal distributions to normal and nonparametric distribution (estimated by means of B-spline) for both time and cost coefficients. We thus consider six model specifications in total (in addition to the deterministic logit):

(1) Times normally distributed, cost constant (TN);
(2) Times and cost normally distributed (T-CN);
(3) Times lognormally distributed, cost constant (TL);
(4) Times and cost lognormally distributed (T-CL);
(5) Times B-spline distributed, cost constant (TBS);
(6) Times and cost B-spline distributed (T-CBS).

Seven control points ($\pi_1, \pi_2, \ldots, \pi_7$) have been estimated for each B-spline, where π_1 and π_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 have been averaged over ten simulations. All runs have been conducted with a modified version of AMLET, available in open source from the site http://www.grt.be/amlet, and the obtained coefficient estimates are reported in Table 1. The results confirm the existence of both time and cost taste heterogeneity across the population. First, we note from Table 2 that the model fit considerably improves when we introduce random elements in our specification. To verify this conclusion, we computed the log-likelihood ratio test for comparing log-likelihood values (see Ben-Akiva and Lerman 1985, §3.8.2, p. 74), leading to an extremely clear rejection[3] of the deterministic model compared to T-CN. Furthermore, the deterministic model produced very similar values for free flow and congested travel times, thereby essentially ignoring variability. By contrast, the variability present in the data duly appears in the new models in the form of significant standard deviations. It can be seen in Table 2 that the fit also improves when lognormal distributions are replaced with normal distributions, and the use of nonparametric distributions reinforces that trend (Table 2). When we apply normal distribution to both congested (CongT Time) and free-flow travel time (FFT 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 nonparametric distributions for both congested and free-flow travel time. However, B-spline indicates a desirable 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 zero. We also confirmed the advantage of the B-spline model by

[3] The value of the statistics is 1805.58 for a $\chi^2(5)$ whose threshold value at the 99% level is 15.09.

Table 1 Real Data: IRIS Parameters Estimation

Parameter	TN	T-CN	TL	T-CL	TBS	T-CBS
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
CongT Time π_1	−0.074	−0.078	−2.822	−2.786	−0.212	−0.250
CongT Time π_2	0.076	0.077	1.004	1.003	−0.182	−0.204
CongT Time π_3					−0.141	−0.152
CongT Time π_4					−0.041	−0.045
CongT Time π_5					−0.036	−0.039
CongT Time π_6					−0.036	−0.029
CongT Time π_7					0.696	0.656
FFT Time π_1	−0.066	−0.070	−2.976	−2.929	−0.249	−0.270
FFT Time π_2	0.080	0.083	1.051	1.078	−0.178	−0.225
FFT Time π_3					−0.095	−0.093
FFT Time π_4					−0.080	−0.083
FFT Time π_5					0.010	0.002
FFT Time π_6					0.013	0.025
FFT Time π_7					0.126	0.078
Cost π_1	−0.291	−0.363	−0.282	−2.318	−0.309	−3.587
Cost π_2		0.757		1.758		−1.345
Cost π_3						−0.6086
Cost π_4						−0.584
Cost π_5						−0.267
Cost π_6						0.187
Cost π_7						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 frequency—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 Departure Time (CP, CPs) μ	−2.836	−2.939	−3.233	−3.302	−2.703	−3.009
Earlier Departure Time (CP, CPs) σ	2.559	2.594	2.878	2.876	2.506	2.723
Later Departure Time (CP, CPs) μ	−1.908	−2.054	−2.442	−2.586	−1.887	−2.129
Later Departure Time (CP, CPs) σ	2.026	2.145	2.575	2.728	2.0885	2.277
Much later Departure 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
Number of cars—3 per household (CD) μ	2.297	2.437	2.027	2.112	2.237	2.375

Note. The μ and σ are the mean and standard deviations of the random variable X for the normal distribution, and of $\log(X)$ for the lognormal.

computing the log-likelihood ratio test for comparing log likelihood of models T-CN and TC-BS, leading to a clear rejection[4] of T-CN. 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 Figures 8 and 9, where we used 750,000 draws;

[4] The value of the statistics is 104.52 for a $\chi^2(15)$ whose threshold value at the 99% level is 30.58.

Table 2 **Real Data: Final Log-Likelihood Values for the Six Models Estimated**

Dist.	TN	T-CN	TL	T-CL	TBS	T-CBS
Log likelihood	−3.1460	−3.1399	−3.1604	−3.1511	−3.1453	−3.1339

Table 3 **Real Data: Value of Travel-Time Savings**

	Quant.	TN	T-CN	TL	T-CL	TB-S	T-CBS
CongT 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
FFT 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

for the nonparametric 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 $60E[\text{time}]/E[\text{cost}]$. Note that in order to be able to compare all distributions, we transform the lognormal CDF (only defined on $\mathscr{R}^+$, and estimated using the opposite of associated observations) to its symmetric distribution with respect to the vertical axis on zero.

We observe substantial variability in VTTS values, especially when cost is assumed to be random. Lognormal distribution applied to both time and cost parameters produces high VTTS values, as a known effect of fat tails (Hensher and Greene 2003). Normal distributions result into low VTTS values given the relatively high percentages of "wrong" positive values for time and cost. Nonparametric distributions produce VTTS values that are close to those obtained with normal distributions, as shown in Figure 10; in particular nonparametric distributions on time and cost result into a congested VTTS higher than free-flow VTTS (as expected). Computational time does however increase with model flexibility; when three parameters are specified as nonparametric, optimization time is about 74 minutes on a MacBook Pro 2 GHz, which is six times larger than the time required to estimate the same model with three normal distributions instead. We nevertheless consider that this additional computing cost remains very reasonable, especially if one considers that improvements are possible by adopting quasi-Monte Carlo draws (Bhat

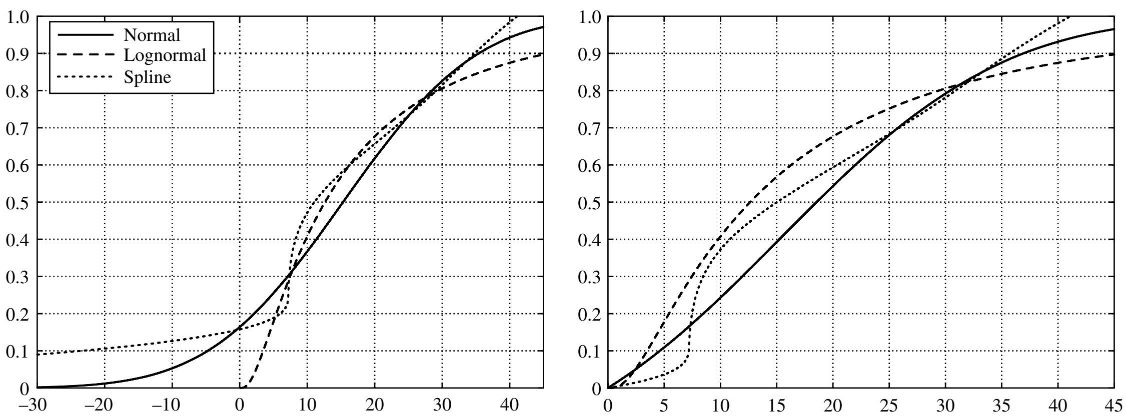

Figure 8 **Value of Congested Travel Time**

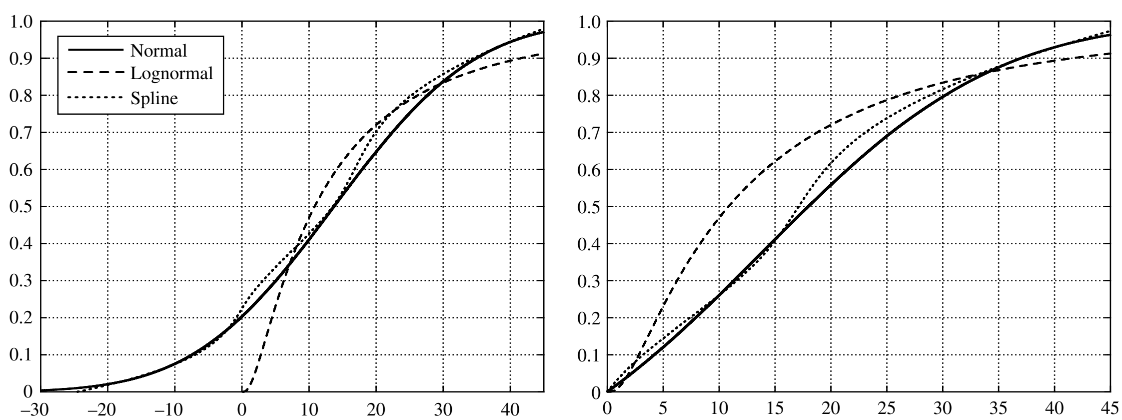

Figure 9 **Value of Free-Flow Travel Time**

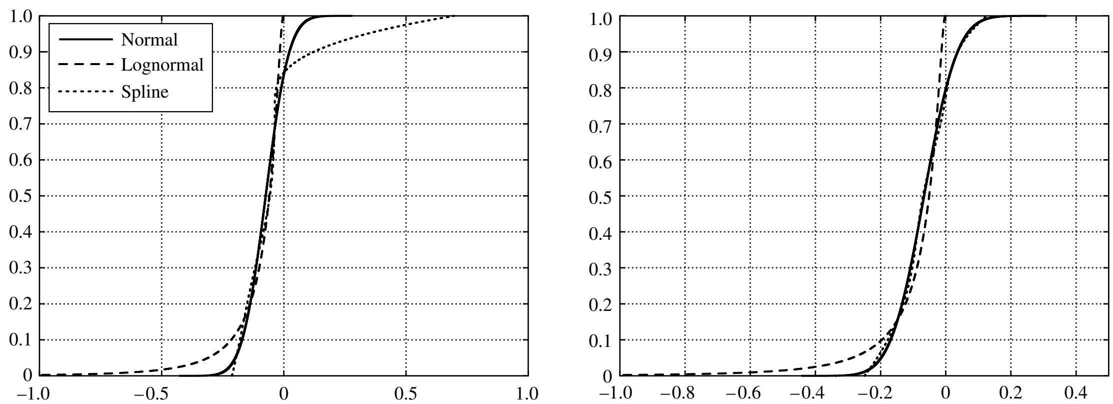

Figure 10 Congested Travel Time and Free-Flow Travel Time CDF

Table 4 Value of Travel-Time Savings from B-Spline Truncated at Zero

	Quant.	TN (corr.)	T-CN (corr.)	TBS (corr.)	T-CBS (corr.)
CongT 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
FFT 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

2001) or an adaptive algorithm (as in Bastin, Cirillo, and Toint 2006b) as well as various refinements of the current code. To summarize, we found that nonparametric B-splines on both time and cost coefficients (a) provide the best model fit to the data, (b) significantly reduce the percentages of the population showing positive values for cost while leaving the proportion of positive time values unchanged (see Figure 11), and (c) give VTTS ranges that do not suffer from fat-tail effects. This suggests that the lognormal assumption, even if more coherent with the econometric theory, may not be reasonable in the studied context, and that the nonparametric approach outperforms the specifications based on the normal to bound the distributions.

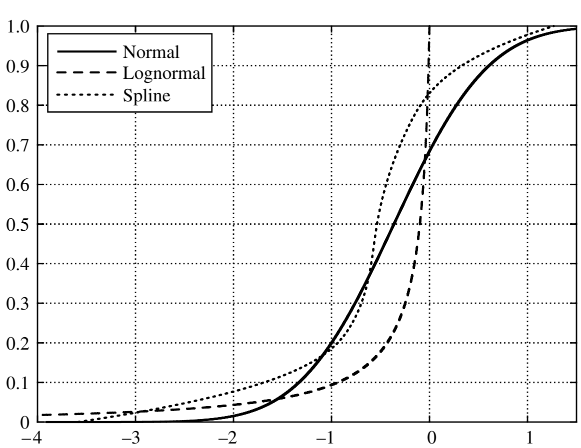

Figure 11 Cost CDF

6. Conclusion

The estimation of heterogeneity in travel-time saving has been the subject of much discussion in the travel behaviour arena, mostly because of its implications in terms of tolling strategy and congestion valuation. 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 because the choice of the “best” distribution is not a trivial task. In this paper, we have proposed to turn to nonparametric 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 for simulated data show that B-spline are able to recover the initial known parametric distributions. We also used real (IRIS) data to validate our methodology and found that not only the goodness of fit of the nonparametric model is better than with other MMNL techniques, but also that it gives VTTS distributions that would be very difficult to recover with classical parametric distributions. In particular, the fraction of the population with implausible cost parameter sharply decreases with the new approach, without leading to excessive VTTS, as would result from using the lognormal distribution. If this last model were to be believed, the financial burden for the users would be higher and the scheme compromised.

These arguments indicate that a purely parametric approach can fail to detect the real distribution, whereas nonparametric random variables may guide the analysts in their search for the real shape of the coefficient 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 than for using classic parametric mixed logit models. Finally, we note that most methods for generating 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, which makes natural an extension of the proposed method where one would apply such multivariate random vector generation techniques. This issue, together with recent applications proposed in econometrics to study state dependency, is part of ongoing research.

Acknowledgments

The authors express their gratitude to Thierry Duquenne from Région de Bruxelles Capitale for granting access to the IRIS data set. Thanks also to the anonymous referees for their hindsightful comments. The first author also gratefully acknowledges the support of Natural Sciences and Engineering Research Council Grant 342368-07.

References

Ayer, M., H. D. Brunk, G. M. Ewing, W. T. Reid, E. Silverman. 1955. An empirical distribution function for sampling with incomplete information. *Ann. Math. Statist.* **26**(4) 641–648.

Bao, H. X. H., A. T. K. Wan. 2004. On the use of spline smoothing in estimating hedonic housing price models: Empirical evidence using Hong Kong data. *Real Estate Econom.* **32**(3) 487–507.

Bastin, F., C. Cirillo, Ph. L. Toint. 2006a. Application of an adaptive Monte Carlo algorithm to mixed logit estimation. *Transportation Res. Part B* **40**(7) 577–593.

Bastin, F., C. Cirillo, Ph. L. Toint. 2006b. Convergence theory for nonconvex stochastic programming with an application to mixed logit. *Math. Programming Ser. B* **108**(2–3) 207–234.

Ben-Akiva, M., S. R. Lerman. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand*. The MIT Press, Cambridge, MA.

Best, M. J., N. Chakravarti. 1990. Active set algorithms for isotonic regression: A unifying framework. *Math. Programming* **47**(3) 425–439.

Bhat, C. R. 1998. Accommodating variations in responsiveness to level-of-service measures in travel mode choice modeling. *Transportation Res. A* **32**(7) 495–507.

Bhat, C. R. 2001. Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Res. B* **35**(7) 677–693.

Bierlaire, M., E. Frejinger. 2008. Route choice modeling based on network-free data. *Transportation Res. C* **16**(2) 187–198.

Cascetta, E., F. Russo, F. A. Viola, A. Vitetta. 2002. A model of route perception in urban road networks. *Transportation Res. A* **36**(7) 577–592.

Cirillo, C., K. W. Axhausen. 2006. Evidence on the distribution of values of travel time savings from a six-week diary. *Transportation Res. Part A* **40**(5) 444–457.

Conn, A. R., N. I. M. Gould, Ph. L. Toint. 2000. *Trust-Region Methods*. SIAM, Philadelphia.

Daly, A. 1982. Estimating choice models containing attraction variables. *Transportation Res. B* **16**(1) 5–15.

Devroye, L. 1986. *Non-Uniform Random Variate Generation*. Springer-Verlag, New York.

Dong, X., F. S. Koppelman. 2003. Mass point mixed logit model: Development and application. Presentation, *10th Internat. Conf. Travel Behav. Res., Lucerne, Switzerland*.

Engle, R. F., C. W. J. Granger, J. Rice, A. Weiss. 1986. Semiparametric estimates of the relation between weather and electricity. *J. Amer. Statist. Assoc.* **81**(394) 310–320.

Farin, G. E. 1991. *NURBS for Curve and Surface Design*. SIAM, Philadelphia.

Fosgerau, M. 2006. Investigating the distribution of the value of travel time savings. *Transportation Res. Part B* **40**(8) 688–707.

Fosgerau, M., M. Bierlaire. 2007. A practical test for the choice of mixing distribution in discrete choice models. *Transportation Res. Part B* **41**(7) 784–794.

Fox, J. 2000. *Multiple and Generalised Non-Parametric Regression*. Sage Publications, Newbury Park, CA.

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

Hensher, D. A., W. H. Greene. 2003. The mixed logit model: The state of practice. *Transportation* **30**(2) 133–176.

Hensher, D. A., N. C. Smith, F. W. Milthorpe, P. O. Barnard. 1992. *Dimensions of Automobile Demand: A Longitudinal Study of Household Automobile Ownership and Use*. Vol. 22, *Studies in Regional Science and Urban Economics*. Elsevier, Amsterdam.

Hess, S., M. Bierlaire, J. W. Polak. 2005a. Discrete mixture of GEV models. *Proc. 5th Swiss Transport Res. Conf., Monte Verita, Ascona, Switzerland*.

Hess, S., M. Bierlaire, J. W. Polak. 2005b. Estimation of value of travel-time savings using mixed logit models. *Transportation Res. Part A* **39**(3) 221–236.

Huber, J., K. Train. 2001. On the similarity of classical and Bayesian estimates of individual mean partworths. *Marketing Lett.* **12**(3) 257–267.

Jackman, S. 2000. Estimation and inference via Bayesian simulation: An introduction to Markov chain Monte Carlo. *Amer. J. Political Sci.* **44**(2) 375–404.

Jarvis, C. H., N. Stuart. 2001. 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. *J. Appl. Meteorology* **40**(6) 1075–1084.

Klein, R. W., R. H. Spady. 1993. An efficient semi-parametric estimator for binary response models. *Econometrica* **61**(2) 387–422.

Koenker, R., P. Ng, S. Portnoy. 1994. Quantile smoothing splines. *Biometrika* **81**(4) 673–680.

Law, A. M. 2007. *Simulation Modeling and Analysis*, 4th ed. McGraw-Hill, New York.

L'Ecuyer, P. 1994. Efficiency improvement and variance reduction. *Proc. 1994 Winter Simulation Conf.*, IEEE Press, Piscataway, NJ, 122–132.

McFadden, D. L., K. Train. 2000. Mixed MNL models for discrete response. *J. Appl. Econometrics* **15**(5) 447–470.

Piegl, L., W. Tiller. 1996. *The NURBS Book*, 2nd ed. Springer-Verlag, New York.

Proussaloglou, K. E., F. S. Koppelman. 1999. The choice of carrier, flight, and fare class. *J. Air Transport Management* **5**(4) 193–201.

Sillano, M., J. D. Ortúzar. 1999. Willingness-to-pay estimation with mixed logit models: Some new evidence. *Environment and Planning A* **37**(3) 525–550.

Singh, G. D., J. A. McNamara, S. Lozanoff. 1997. Morphometry of the cranial base in subjects with class III malocclusion. *J. Dental Res.* **76** 694–703.

Srinivasan, S., C. R. Bhat. 2005. Modeling household interactions in daily in-home and out-of-home maintenance activity participation. *Transportation* **32**(5) 523–544.

Steihaug, T. 1983. The conjugate gradient method and trust regions in large scale optimization. *SIAM J. Numer. Anal.* **20**(3) 626–637.

Toint, Ph. L. 1981. Towards an efficient sparsity exploiting Newton method for minimization. I. S. Duff, ed. *Sparse Matrices and Their Uses*. Academic Press, London, 57–88.

Train, K. 2003. *Discrete Choice Methods with Simulation*. Cambridge University Press, New York.

Train, K., G. Sonnier. 2005. Mixed logit with bounded distributions of correlated partworths. A. Alberini, R. Scarpa, eds. *Applications of Simulation Methods in Environmental Resource Economics*, Chapter 7. Kluwer Academics Publisher, Dordrecht, The Netherlands.

Train, K., M. Weeks. 2005. Discrete choice models in preference space and willingness-to-pay space. A. Alberini, R. Scarpa, eds. *Applications of Simulation Methods in Environmental Resource Economics*, Chapter 7. Kluwer Academics Publisher, Dordrecht, The Netherlands.