Assessment of User Benefits in Presence of Random Taste Heterogeneity

Comparison of Parametric and Nonparametric Models

Elisabetta Cherchi, Cinzia Cirillo, and John W. Polak

The ultimate objective of most practical transport modeling is the evaluation of the economic impacts of alternative policy measures. During the past decade, significant advances have been made in the ability of operational models to accommodate flexible patterns of taste heterogeneity. The majority of applications use parametric distributions in which the parameters of the distribution are specified a priori. However, in reality individual tastes rarely follow a standard distribution, especially in the extremes of the distribution in which behavior is likely to differ substantially from that of the rest of the population. The major problem in these cases is the suitability of estimated models for determining economic welfare because they may yield erroneous or counterintuitive results. Recent research proposes using nonparametric models to capture randomness in individuals' tastes, in which the shape of the unknown distribution is defined as part of the estimation process. But there is still no evidence regarding the implication of using these more advanced model structures for determining economic welfare. The aim of this paper is to investigate to what extent imposing a well-shaped distribution of individual tastes affects the appraisal of transport policy measures. In particular the focus is on random taste heterogeneity. With the use of simulated data, an investigation is done on how parametric and nonparametric models perform under different assumptions for true taste distributions. Finally, implications of the accuracy of alternative methods of computing benefit measures are critically analyzed.

The use of transport demand models for assessing user benefits is a basic element of the transport planning process, and considerable effort has been devoted to improving the capabilities of travel demand modeling techniques. Following the rapid adoption of the mixed multinomial logit (MMNL) model, one issue that has received a great deal of research attention is the characterization of taste distributions (*1*). Most applications use parametric distributions (either bounded or unbounded), in which the shape of the distribution is specified a priori by the analyst. Unbounded distributions by construction may have long tails and zero values and may allow parameters to be either positive or negative (*2, 3*). This causes several problems: opposite signs or extreme values (too large or too small) are often obtained, which can jeopardize the user benefit analysis. Bounded distributions overcome these problems (*4, 5*), but still assume a well-shaped distribution and can overestimate the true mean toward the bounds, distorting the welfare measures (*6*).

The most recent and promising advances adopt nonparametric models to capture the randomness in individuals' tastes. Here, the shape of the distribution is unknown and defined as part of the estimation process. Several approaches have been proposed in the literature to estimate semiparametric or nonparametric random coefficients in discrete choice models. Dong and Koppelman use two points along each of the parameter dimensions to represent the distribution and a Bayesian approach to recover their mass and the associated probabilities (*7*). Using a real data set they found that maximum likelihood mixed logit failed to recover the true mass points from simulated data and that the mass point mixed logit is superior to the parametric mixed logit. Hess et al. propose a discrete mixture of generalized extreme value (GEV) models over a finite set of distinctive support points (*8*). Two or three support points distribution are estimated on the Danish value of time data, and up to six points discrete mixture of GEV are compared with mixture of two normals in a simulated study (*9*). Fosgerau employs various nonparametric techniques to investigate the distribution of travel-time savings from a stated choice experiment (*10*). The methodology adopted does not account for repeated observations and applies only to binomial choices. Recently, Fosgerau and Bierlaire proposed a seminonparametric specification to test whether a random parameter of a discrete choice model follows a given (normal or lognormal) distribution (*11*). The test is adapted for just one random parameter at a time. Finally Bastin et al. propose a new nonparametric approach for explicitly estimating the shape of the unknown distributions, expressed via their cumulative distributions, as part of the complete calibration procedure (F. Bastin, C. Cirillo, and F. L. Toint, Estimating Non-Parametric Random Utility Models with an Application to the Value of Time in Heterogeneous Populations, under review, *Transportation Science*).

All these applications focus simply on random parameter estimation and on the derivation of the point estimates of willingness to pay (WTP). WTP is perhaps the most popular benefit measure in transport studies, but it represents a point estimate and its validity as a benefit measure is limited to small changes in the variables that do not make individuals change alternatives. Moreover, the extensive work devoted to exploring the implications of random taste heterogeneity in WTP is questionable because WTP per se (i.e., the ratio between the marginal utility of the attribute one wants to evaluate and the marginal utility of cost) is used only in the rule-of-half, the most aggregated of the welfare measures (*12, 13*). All other welfare measures, including the

E. Cherchi, Centro Ricerche Modelli di Mobilità, Facoltà di Ingegneria–Università di Cagliari, Piazza d'Armi 16, 09123 Cagliari, Italy. Alternate address for E. Cherchi: Centre for Transport Studies, Imperial College London, Exhibition Road, London SW7 2AZ, United Kingdom. C. Cirillo, Department of Civil and Environmental Engineering, University of Maryland, College Park, MD 20742. J. W. Polak, Centre for Transport Studies, Imperial College London, Exhibition Road, London SW7 2AZ, United Kingdom. Corresponding author: E. Cherchi, echerchi@unica.it.

Transportation Research Record: Journal of the Transportation Research Board, No. 2132, Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 78–86.
DOI: 10.3141/2132-09

logsum, require that the marginal utility of income be estimated separately from any other parameter.

Few studies in the literature analyze the welfare implications of the presence of taste heterogeneity for nonmarginal changes. Train's analysis of anglers' choice of fishing sites found that a random parameter mixed logit model and a multinomial logit model gave practically the same welfare measure but only when the policy implies a change of the same absolute value in all alternatives (*14*). This result is interesting because, in this particular case (i.e., when an alternative generic attribute is increased by the same absolute quantity in all alternatives), the benefit is exactly equal to the WTP multiplied by the absolute quantity by which the attribute is increased. Therefore this result falls in the WTP category. On the contrary, when the policy implies different changes among alternatives, Train found that accounting or not for random taste heterogeneity yields significantly different estimated benefits. Persson analyzes, in regard to health economics, the welfare implications of a change in the price of different inputs but using only the mean value of the random coefficients (*15*). Disregarding the extra information provided by the distribution of the parameters equates to treating the MMNL as a multinomial logit (MNL) model; therefore the measure adopted by Persson cannot be considered entirely correct.

In the transportation context, Amador et al. computed the benefit measures in the travel time parameter when accounting for random taste heterogeneity (*16*). Their main finding was that estimates from a simple logit model underestimate the benefit by as much as 32% as compared with the benefit obtained using a random parameter logit model. Cherchi and Polak highlighted a new problem related to the consistency in the treatment of heterogeneity between estimation and benefit calculation (*6*). They found that errors in modeling random heterogeneity can have unpredictable effects on benefit estimates. Although the estimation of user benefit generally improves with the quality of the specification of heterogeneity, cases arise in which good results can be obtained with a very poorly specified model.

To the best of the authors' knowledge, there is still no evidence on the implications of using nonparametric model structures for determining economic welfare for nonmarginal changes.

The overall aim of this paper is twofold. The first objective is to investigate to what extent nonparametric models are able to recover situations in which individual tastes are characterized by nonsymmetric distributions, and in particular by different behavior in the tails. The paper provides further evidence on the capability of more flexible techniques recently proposed in the literature to reproduce ill-shaped distributions. This is an interesting issue that, notwithstanding recent advances, has not been fully explored. In addition, by focusing attention on the compensating variation and on the logsum measure, the implication of different taste patterns on the appraisal of transport policy measures is studied. A controlled experiment was set up to concentrate on the "pure" effect of taste heterogeneity. This procedure has the advantage of enabling the analyst to explore the effects of different specification errors, without incurring the practical complications and uncertainties due to real survey data treatment. Moreover, the use of pseudo-observed individuals allows calculation of the exact compensating variation and the ability to compute, for each individual, the difference between the expectation and the "real" benefit.

The rest of the paper is organized as follows. The next section reviews the parametric and the nonparametric approaches for modeling taste heterogeneity in the mixed logit model. Next, there is a brief outline of the welfare economic concepts in discrete choice situations and a discussion on how the assumptions behind aggregate benefit measures (conventionally used in practice) reduce the explanatory capability of the mixed logit models. A description of the design of the numerical experiments undertaken is given next, followed by the results of the models estimated. The final section draws conclusions from the analysis and highlights key issues, for practitioners and researchers.

ACCOUNTING FOR RANDOM HETEROGENEITY IN MMNL MODEL

The MMNL is any model whose choice probabilities can be expressed as the integral of standard logit probabilities, evaluated at parameters β, over a density f of parameters (*17*):

$$P_{qj} = \int L_{qj}(\beta)\, f(\beta|\Omega)\, d\beta \tag{1}$$

where

- P_{qj} = unconditional probability of individual q choosing alternative j among any ith alternative belonging to her or his set (I_q) of available alternatives,
- Ω = population parameters of the distribution, and
- L_{qj} = probability of individual q choosing alternative j conditional on β.

The only requirement for an MMNL model is that the random component of individual utility have an additive GEV Type 1 error, which generates the standard logit probability, and the vector of parameters β can assume any desired distribution (or can be fixed). In fact, as demonstrated by McFadden and Train, any well-behaved random utility model can be approximated to any degree of accuracy by an MMNL model (*1*).

The typical MMNL model with parametric distribution of the random parameters is obtained assuming that L_{qj} is the typical simple logit formula, such that the unconditional probability of individual q choosing alternative j takes the following form:

$$P_{qj} = \int \frac{\exp\left(\beta_{qj} \cdot X_{qj}\right)}{\sum_i \exp\left(\beta_{qi} \cdot X_{qi}\right)} f\left(\beta|b_j, \textstyle\sum\right) d\beta \tag{2}$$

where X_{qj} is a vector of known (by the modeler) characteristics of alternative j for individual q; $\beta_{qj} \approx f(\beta|b_j, \Sigma)$ is a vector of taste parameters distributed over the population, according to a predefined density f, with mean b_j, and variance–covariance matrix Σ, to be estimated.

The continuous mixed distribution in Equation 2 is typically estimated through simulation, over R repetitions:

$$P_{qj} = \frac{1}{R} \sum_{r=1,\ldots,R} \frac{\exp\left(\beta^r_{qj} \cdot X_{qj}\right)}{\sum_i \exp\left(\beta^r_{qi} \cdot X_{qi}\right)} \tag{3}$$

Following Train, in regard to the mixed logit model, a nonparametric model can be viewed as a discrete mixing distribution with a sufficient number of support points (*18*). In fact, let $s_c = f(\beta_c|\Omega)$ be the share of the population that has coefficient β_c for each class $c = 1, \ldots, C$. In this case the unconditional probability of individual q choosing alternative j will be the summation over the C classes of the continuous mixed distribution:

$$P_{qj} = \sum_{c=1,\ldots,C} s_c \int L_{qj}(\beta)\,\phi\left(\beta|b_c, \textstyle\sum_c\right) d\beta \tag{4}$$

where $\phi(\beta|b_c, \Sigma_c)$ is the unconditional density of the parameters in each class.

Most studies on parameter estimation without assumptions on the underlying distributions are concerned with discrete distributions. In this paper continuous nonparametric distributions are adopted based on B-spline, recently introduced in mixed logit estimation by Bastin et al. (see documentation above). Here, each component of the random vector inherent in the mixed logit function is itself random; if independence between these components is assumed, each component can be considered separately. If X is a univariate random distribution, a well-known technique to generate draws from its distribution consists in sampling from a uniform on [0,1] and in applying the inverse cumulative distribution function F_X^{-1} to these draws. It is furthermore assumed that the random variable X has a bounded support based on B-spline functions. The bounded support assumption is not unduly restrictive because extreme behavior, corresponding to values of X tending to plus or minus infinity, is usually undesirable. In many practical cases the bounded support assumption can be an advantage rather than a drawback.

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:

$$C(u) = \sum_{i=0}^{n} P_i N_{i,p}(u) \tag{5}$$

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:

$$N_{1,0} = \begin{cases} 1 & \text{if } u \in [u_i, u_{i+1}] \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

and

$$N_{i,p} = \frac{u - u_i}{u_{i+p} - u_i} N_{i,p-1}(u) + \frac{u_{i+p+1} - u}{u_{i+p+1} - u_{i+1}} N_{i+1,p-1}(u) \tag{7}$$

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

The knot vector chosen for the purpose of this paper is the nonperiodic (clamped or open) knot vector that takes the form

$$U = \Big\{\underbrace{a, \ldots, a}_{p+1}, u_{p+1}, \ldots, u_{m-p-1}, \underbrace{b, \ldots, b}_{p+1}\Big\} \tag{8}$$

In which the first and last knots have the multiplicity of $p + 1$.

In this application cubic B-spline is considered, and therefore $p = 3$ is set.

ACCOUNTING FOR RANDOM HETEROGENEITY IN WELFARE MEASURES

The welfare definition, "the compensating variation in income, whose loss would just offset the fall in price, and leave the consumer no better off than before," was proposed by Hicks in 1956 (*19*). The compensating variation (CV) thus represents the minimum amount by which a consumer would have to be compensated after a price change to stay at the same level of utility as before. [This paper refers simply to the CV because it is the most commonly used welfare measure. In fact the CV is evaluated at the new prices but at the initial utility, while the analogous measure (the equivalent variation) is evaluated at the old prices but at the final utility level, which is often impossible to compute in practice.]

Following the general notation proposed by McFadden (*20*), the compensating variation for a change in the transport attributes from $(c'_{qj}, t'_{qj}, X'_{qj})$ to $(c''_{qj}, t''_{qj}, X''_{qj})$ is the quantity CV such that

$$\max_{j \in A_q} U(I_q - c'_{qj}, t'_{qj}, X'_{qj}; s_q, \eta_{qj}) = \max_{j \in A_q} U\left(I_q - \mathrm{CV}_q - c''_{qj}, t''_{qj}, X''_{qj}; s_q, \eta_{qj}\right) \tag{9}$$

where

$U(\cdot)$ = utility that individual q obtains from a particular travel choice j;
c_j and t_j = travel cost and time for alternative j, respectively;
X_{qj} = vector of other observed attributes of travel;
s_q = vector of individual's observed characteristics;
η_{qj} = vector of unobserved attributes of travel and individual characteristics; and
I_q = set of alternatives available to each individual.

In the most general case, computing the CV is a challenging task, because it depends on η, whose distribution induces a distribution of income compensation levels. Because in real cases η_q is unknown, the best approximation of the "real" CV one can estimate is its expected value $(E[\mathrm{CV}(I - c_j, t_j, X_{qj}; s_q, \eta_q)])$. Moreover, since the expectation is not preserved by nonlinearities, if the utility is nonlinear in income and thus in CV (i.e., in the presence of income effect) the compensating variation is not independent on the maximization over j and thus on the expectation. In this case, McFadden proposes a simulation procedure that consists in drawing a vector of random terms (η), searching iteratively for the value of CV that satisfies Equation 9 and calculating the mean of the CV(r) over the R iterations of these two steps (*20*).

The logsum measure is derived from Equation 9 under the assumptions that there is no income effect and that at least one of the unknown terms (say, $\eta_{qj} = \{\beta_{qj}, \epsilon_{qj}\}$) is additive with a GEV Type 1 distribution. The first assumption, as said before, makes the CV independent of the expectation and allows (i.e., it produces correct results) calculation of CV as the quantity that equates the expected maximum utility before and after policy application. The second assumption allows one to derive the typical logsum formula because the expected maximum utility of a GEV Type 1 distribution is equal to the logsum expression plus a constant (*21, 22*). A detailed derivation of the benefit measures can be found in Cherchi and Polak (*23*). Hence, if the utility in Equation 9 reduces to $U(\bullet) = -\lambda_q c''_{jq} + f_j(t_{qj}, X_{qj}, s_q; \beta_{qj}) + \epsilon_{qj}$, with λ_q the marginal utility of income and ϵ_{qj} GEV Type 1, the expected compensating variation (E[CV]) is given by

$$\mathrm{CV}_q = \frac{1}{\lambda_q}\left\{\begin{aligned} &\log\sum_{i=1}^{J}\left[\exp\left(-\lambda_q c''_{qj} + f_j(\cdot)\right)\right] \\ &-\log\sum_{j=1}^{J}\left[\exp\left(-\lambda_q c'_{qj} + f_j(\cdot)\right)\right]\end{aligned}\right\} \tag{10}$$

Equation 10 does not require that individual tastes (or the marginal utility of income or any other parameters) be fixed over the population. In fact, the logsum formula also holds in the presence of random heterogeneity. In the case of the unconditional welfare measure, the individual compensating variation can be computed through simulation (SCV), that is, drawing β^r from the joint distribution of the random parameters (*10*), calculating CV_q with Equation 9 or 10 and averaging over R repetitions:

$$\mathrm{SCV}_q = \frac{1}{R}\sum_{r=1}^{R}\mathrm{CV}_q\left(\beta^r\right) \tag{11}$$

The method was defined for the parametric distributions, in which each β^r is drawn from the underlying distribution, its mean and standard deviation being known. In the case of nonparametric models, the method is analogous, but each β^r is drawn from the following polynomial function defined previously:

$$C(u) = \sum_{i=0}^{n} P_i N_{i,p}(u) \tag{12}$$

Finally, the estimate of the aggregate logsum for the population is

$$\mathrm{CV} = \sum_{q=1}^{Q}\delta_q \mathrm{SCV}_q \tag{13}$$

where each individual in the sample represents a certain number of individuals in the population, so it is the weight, δ_q, of each individual q on the population.

SIMULATION EXPERIMENTS

To explore the effect of taste heterogeneity on benefit measures a collection of data sets was simulated in which pseudo-observed individuals behave according to a choice process determined by the analyst. These individuals were generated from a kernel of data drawn from a sample of real individuals collected in 1998 for a mode choice study among car, bus, and train (*24*). The advantage of using this kernel is that it provides both detailed level of service (LOS) information for alternative modes and extensive socioeconomic information on the individuals and on the distributional properties of these attributes across the sample. By so doing it was possible to approximate the generated data set, be it simulated, to a real situation.

The variables (type, average values, and standard deviations) used to build the simulated utility function are shown in Table 1. For the LOS variables, a truncated normal distribution has been adopted to guarantee that the minimum and maximum values of each attribute do not exceed the limits measured in the real sample.

In the experiment the utility functions assigned to each individual q are linear in the parameters and in the attributes. The model specification includes a full set of alternative specific constants and four level of service variables [travel time (tt), travel cost (c), waiting time (wt), and frequency of public transportation (f)].

$$\begin{aligned}
U_{q,\text{car}} &= \beta_{\text{car}} + \beta_c c_{q,\text{car}} + \beta_{\text{tt}}\text{tt}_{q,\text{car}} + \beta_{\text{wt}}\text{wt}_{q,\text{car}} + \epsilon_{q,\text{car}}\\
U_{q,\text{bus}} &= \quad\quad + \beta_c c_{q,\text{bus}} + \beta_{\text{tt}}\text{tt}_{q,\text{bus}} + \beta_{\text{wt}}\text{wt}_{q,\text{bus}} + \beta_f f_{q,\text{bus}} + \epsilon_{q,\text{bus}}\\
U_{q,\text{train}} &= \beta_{\text{train}} + \beta_c c_{q,\text{train}} + \beta_{\text{tt}}\text{tt}_{q,\text{train}} + \beta_{\text{wt}}\text{wt}_{q,\text{train}} + \beta_f f_{q,\text{train}} + \epsilon_{q,\text{train}}
\end{aligned} \tag{14}$$

TABLE 1 Variables Used to Build Simulated Systematic Utility Functions

	Average Value	Standard Deviation	Truncation Upper Limit	Truncation Lower Limit
Car				
tt	16.4	14	2	28
c	1.99	1.3	1	4
wt	2.84	4	1	4
Bus				
tt	27.5	13	5	60
c	1.54	0.5	1	3
wt	14.4	7	1	20
f	6	2	1	12
Train				
tt	16.9	9.8	4	25
c	1.74	0.5	1	3.5
wt	20.1	9	1	30
f	3	2	1	6

NOTE: Travel and walking time are in minutes, travel cost in euros, frequency in number of means per hour, and income in euros per trip.

In the case of fixed parameters the values of the parameters were taken from an MNL estimated with the above real data and have the following point values: travel time (–0.1), cost (–0.6), walking time (–0.22), frequency (0.6), car specific constant (2.0), and train specific constant (–0.5).

The additive random term (ϵ_{qj}) in Equation 14 is GEV Type 1. This is a useful assumption because it is consistent with the basic structure of the MMNL and with the logsum welfare measure used in the application. Taste heterogeneity was then introduced by allowing car travel time to vary randomly over the population. Four samples were generated by drawing from truncated normal distributions with different means, standard deviations; and mass points at zero (Table 2). Table 2 also shows the Persson's skewness coefficient, which measures the asymmetry of the distribution, defined as $S_k = 3(\mu - M_d)/\sigma$ (where μ is the mean, σ the standard deviation, and M_d the median). Each sample contained 6,000 pseudoindividuals. Truncated distributions create samples in which two classes of behavior can be identified: one characterized by random taste (i.e., normal) and the other with deterministic preferences. This produces a very interesting case study; situations in which the real taste distribution in the sample does not have exactly the same shape assumed in the estimation are very likely to arise. This problem will be explored in the section on results, and its effect on the calculation of user benefits will be discussed.

To evaluate the implications of the assumption about random heterogeneity on the benefit measures, the true compensating variation (CV_true) was computed using Equation 9, and the true parameters were generated as described above. The method comprises the following three steps:

1. Compute the corresponding random utility for each individual and each alternative,
2. Search for the maximum over the alternatives, and
3. Calculate the CVs that equate the two maxima for each individual.

This is therefore an exact measure of the compensating variation and will be used as a reference measure. The logsum is then calculated by applying Equation 10, and the error is computed with respect to the reference measure.

TABLE 2 Simulated Values for Truncated or Censored Normal Distribution

Sample	Mean	Standard Deviation	Mass on Inferior Limit (%)	Mass on Superior Limit (%)	Persson's Skewness Coefficient
1	–0.23	0.29	0.0	41.3	–1.360
2	–0.29	0.32	0.0	33.3	–0.883
3	–0.44	0.38	0.0	18.6	–0.351
4	–0.80	0.43	0.3	3.7	–0.007

Calculated by Munizaga et al., the market share predicted under different model specifications ($\hat{N}_i$) was also compared with the real market share (N_i) (i.e., the simulated number of individuals choosing option *i* under the policy); the following chi-square index was used for that purpose (*25*):

$$\chi^2 = \sum_i \frac{\left(\hat{N}_i - N_i\right)^2}{N_i} \tag{15}$$

MODELING RESULTS

By using the samples described above, several MMNL models were estimated under the assumptions commonly made in practice in cases in which modelers have no information about the actual underlying preferences. In particular, parametric and nonparametric models were estimated and estimation results were compared on the basis of (*a*) goodness-of-fit statistics, (*b*) their ability to reproduce the true taste distribution, and (*c*) their performance in assessing benefits for nonmarginal changes. Among the parametric models, unbounded normal and bounded normal were tested (bounds were defined a priori, not estimated). For comparative purposes simple MNL were also estimated.

As discussed in the previous section, all simulated individuals have negative preferences for travel time. However, in practice, underlying preferences are not known. Therefore modelers typically estimate several models using different unbounded and bounded distributions to characterize these preferences and judge which is the most appropriate distribution on the basis of the estimation results.

Results of the parametric models, along with the comparison with the MNL model, are shown in Table 3. The unbounded model is estimated with the typical normal distribution, which allows for long positive and negative tails. Long tails of both signs are usually a drawback in the estimation; however unbounded distributions have the great advantage of being symmetric, which can be beneficial in regard to accuracy in the estimation of the true distribution's mean. The bounded model uses a normal truncated distribution, with the same truncations used to generate the true parameter distribution ($-2 < \beta_q < -0.0001$). Bounded distributions are more realistic (these models were estimated using Kenneth Train's Gauss code, elsa.berkeley.edu/~train/software.html), but are also usually skewed, which may shift the estimated mean significantly with respect to the true mean (*6*). For reasons of space the *t*-test is not shown here, but it was always far greater than the critical value at 95% significance level.

First note that all models perform fairly well for all samples. As expected, all models that allow for heterogeneity in tastes perform much better than the MNL models (the likelihood ratio test is always rejected at the 95% significance level). In line with the theory, the values of all parameters in the MNL are lower than the corresponding parameters estimated in the MMNL models because the latter have a smaller variance (and hence a higher scale parameter, as clearly illustrated in Table 3) (*3*). The scale, reported for each estimated parameter, is computed as the ratio between the estimated and the true parameter and represents indeed the scale of the model. The value of the scale is not useful per se, but the heterogeneity of the scale parameters among the estimated parameters for each model reveals how well the estimated parameters recover the true ones. The more homogeneous the scale, the more the estimated parameters recover the true ones. For that purpose, the coefficient of variation of the scales is reported for each model. It is clear that the ability of the MMNL to recover the true parameters increases as the distribution of the tastes becomes better shaped (smaller mass points in the tails). As expected the bounded normal recovers the true parameters better than the unbounded one, and the MNL is the worst performer. Interestingly however, the normal distribution performs quite well, and results are fairly similar to those obtained with truncated normal distribution for all samples analyzed. Moreover, the marginal superiority of the bounded over the unbounded distribution does not vary across the samples (i.e., with skewness of the true taste distribution), confirming the good behavior of the unbounded normal distributions.

Table 4 shows the results obtained with the nonparametric models. The spline is constructed using the knot vector {0, 0, 0, 0, 1/10, 2/5, 3/5, 9/10, 1, 1, 1, 1}; eight control points (P1, P2, . . . , P8) have been estimated, where P1 and P8 give the bounds of the distribution. Monte Carlo simulations based on 2,000 random draws per individual have been adopted to simulate the maximum likelihood (these models are estimated with a modified version of AMLET, available in open source at www.grt.be/amlet). The adjusted rho-squared (that properly accounts for the number of estimated parameters, which is higher in the nonparametric models) is always higher than in the parametric models, and this superiority is more pronounced for distributions with large mass points (Sample 1) than for better shaped distributions (as in Sample 4). In Figure 1 the cumulative distribution function of car travel time obtained with the B-spline on eight support points (called spline_10-40-60-90) is compared with the estimated truncated normals and with a spline based on a different knot vector having seven control points (spline_25-50-75). The nonparametric models are able to recover the mass at zero in all samples. The nonparametric distributions fail to recover the other extreme of the distribution in the first three samples, and in the fourth sample the negative tail is correctly recovered. In this latter case, on increasing the knot vector from seven points to eight points a sharp improvement was observed in negative tail estimation. Increasing the number of support points in the nonparametric distribution improves precision, but there is a limit to the information that can be drawn from extreme behaviors.

How different model specifications affect the computation of the welfare measures is also explored in Table 5. This is a crucial analysis, not only because the economic evaluation of alternative

TABLE 3 Estimation Results: MNL and Parametric Models

	Sample 1		Sample 2		Sample 3		Sample 4	
Attribute	Estimate	Scale	Estimate	Scale	Estimate	Scale	Estimate	Scale
MNL								
tt_Car (mean)	–0.0426	0.13	–0.0697	0.24	–0.0871	0.20	–0.1499	0.19
tt_Car (SD)	—		—		—			—
tt-pt_bus&train	–0.0573	0.57	–0.0636	0.64	–0.0662	0.66	–0.0827	0.83
Cost	–0.2196	0.37	–0.2613	0.44	–0.2978	0.50	–0.3900	0.65
wt	–0.1308	0.59	–0.1393	0.63	–0.1424	0.65	–0.1885	0.86
Freq.	0.3298	0.55	0.3648	0.61	0.3818	0.64	0.4989	0.83
ASC_car	0.3311	0.17	0.5244	0.26	0.0705	0.04	–0.9839	–0.49
ASC_train	–0.7583	1.52	–0.7150	1.43	–0.7098	1.42	–0.6297	1.26
L(max)	–4,603.69		–4,546.29		–4,566.42		–3,405.01	
Rho^2 adjusted	0.3004		0.3091		0.3060		0.4822	
Coeff.var		0.84		0.66		0.76		1.11
Normal Unbounded								
tt_Car (mean)	–0.1450	0.44	–0.2516	0.86	–0.4874	1.10	–0.9771	1.22
tt_Car (SD)	0.4973	1.13	0.4142	1.28	0.4851	1.29	0.5317	1.24
tt-pt_bus&train	–0.1420	1.42	–0.1411	1.41	–0.1401	1.40	–0.1333	1.33
Cost	–0.6888	1.15	–0.7259	1.21	–0.7933	1.32	–0.7789	1.30
wt	–0.2955	1.34	–0.2930	1.33	–0.2911	1.32	–0.2950	1.34
Freq.	0.8009	1.33	0.8066	1.34	0.8141	1.36	0.8024	1.34
ASC_car	1.1513	0.58	1.8997	0.95	2.2797	1.14	2.2696	1.13
ASC_train	–0.8802	1.76	–0.8089	1.62	–0.7932	1.59	–0.6846	1.37
L(max)	–4,356.16		–4,286.84		–4,225.75		–3,111.86	
Rho^2 adjusted	0.3379		0.3484		0.3577		0.5267	
Coeff.var		0.38		0.20		0.12		0.06
Inferior limit	4.8284		3.8902		4.3632		4.3399	
Superior limit	–5.1184		–4.3935		–5.3379		–6.2940	
% of tt_Car > 0	38.53		27.17		15.75		3.31	
Normal Bounded								
tt_Car (mean)	–0.1569	0.47	–0.2619	0.89	–0.5029	1.14	–1.0197	1.27
tt_Car (SD)	0.4909	1.12	0.4258	1.32	0.5023	1.34	0.5660	1.32
tt-pt_bus&train	–0.1391	1.39	–0.1399	1.40	–0.1399	1.40	–0.1336	1.34
Cost	–0.6808	1.13	–0.7220	1.20	–0.7985	1.33	–0.7803	1.30
wt	–0.2904	1.32	–0.2913	1.32	–0.2908	1.32	–0.2957	1.34
Freq.	0.7828	1.30	0.7981	1.33	0.8098	1.35	0.8022	1.34
ASC_car	1.3393	0.67	2.0602	1.03	2.4128	1.21	2.4028	1.20
ASC_train	–0.8735	1.75	–0.8085	1.62	–0.7971	1.59	–0.6885	1.38
L(max)	–4,348.89		–4,278.48		–4,218.38		–3,110.31	
Rho^2 adjusted	0.3390		0.3497		0.3588		0.5269	
Coeff.var		0.35		0.18		0.10		0.04
Inferior limit	0.2134		0.1252		0.0885		0.0331	
Superior limit	–2.0192		–2.3047		–2.5820		–2.5991	
% of tt_Car > 0	37.46		26.93		15.83		3.58	

NOTE: L = log likelihood.

policy measures is the ultimate objective of most practical transport modeling, but also because the welfare measures can provide further information to evaluate model results other than statistical tests. Indeed, in line with the findings of other researchers (*6, 13*), the results seem to confirm that poorly specified models can sometimes yield the best results in regard to computation of user benefits for nonmarginal changes. Table 5 refers to a simple increase of 25% in car travel time. The nonparametric models, although slightly superior in regard to statistical tests, do not reproduce the true welfare values better than do the parametric models. The welfare measures were computed in both cases (i.e., for both the parametric and nonparametric distributions), truncating the taste distributions to assume only negative values. This

TABLE 4 Estimation Results: Nonparametric Models

	Sample 1		Sample 2		Sample 3		Sample 4	
Attribute	Estimate	Scale	Estimate	Scale	Estimate	Scale	Estimate	Scale
tt_Car (P1)	−1.4736		−1.4600		−1.3143		−2.7382	
tt_Car (P2)	−1.4736		−1.4600		−1.2777		−1.3617	
tt_Car (P3)	−1.0870		−0.6268		−1.2777		−1.3617	
tt_Car (P4)	−0.2745		−0.5263		−0.6693		−1.3617	
tt_Car (P5)	0.0497		−0.0222		−0.3726		−0.6856	
tt_Car (P6)	0.0497		0.0350		0.0441		−0.6856	
tt_Car (P7)	0.0497		0.0350		0.0441		0.0213	
tt_Car (P8)	0.0497		0.0368		0.0441		0.0942	
tt-P_bus&train	−0.1426	1.43	−0.1417	1.42	−0.1406	1.41	−0.1331	1.33
Cost	−0.7130	1.19	−0.7257	1.21	−0.8033	1.34	−0.7777	1.30
wt	−0.2998	1.36	−0.2963	1.35	−0.2918	1.33	−0.2943	1.34
Freq.	0.8037	1.34	0.8109	1.35	0.8102	1.35	0.7988	1.32
ASC_car	2.1368	1.07	2.5103	1.26	2.6032	1.30	2.3051	1.15
ASC_train	−0.8754	1.75	−0.8078	1.62	−0.8062	1.61	−0.6885	1.38
L(max)	−4324.29		−4270.21		−4212.97		−3105.47	
Rho^2 adjusted	0.3418		0.3522		0.3587		0.5289	
Coeff.var		—		2.22		0.07		0.05
tt_Car (25%)	−0.7590	1.33	−0.6284	1.26	−1.0109	1.45	−1.3427	1.22
tt_Car (50%)	−0.1260	1.27	−0.2750	1.38	−0.5237	1.31	−1.0237	1.28
tt_Car (75%)	0.0406		−0.0036	36.0	−0.1412	1.48	−0.6550	1.32
% of tt_Car > 0	36.4		25.0		14.0		3.2	

is what Cherchi and Polak called the problem of lack of consistency between estimation and evaluation (*6*). No matter what distribution is taken in the estimation process, in calculating the benefit it is necessary to constrain the distribution to just the negative part. This problem typically occurs with the MMNL, and results show that nonparametric distributions also suffer from this problem.

CONCLUSIONS

This paper has explored the use of advanced demand models for assessing user benefits and determining welfare measures for nonmarginal changes in transportation planning. The application proposed is based not only on parametric distributions, in which the shape of the distribution is specified a priori (either bounded or unbounded), but also on more flexible nonparametric techniques. Different methods have recently been proposed in the literature to capture the randomness in individuals' tastes without imposing a defined distribution, but no results are reported for the consequent calculations of the compensating variation.

The analysis is based on four simulated experiments in which two classes of behavior are present in the sample: one that follows the normal distribution and a second that has deterministic preferences (characterized by two masses at the truncation points). The use of synthetic data is necessary to study the ability of the proposed techniques, not only to recover the real shape of the individual taste heterogeneity, but also to correctly reproduce the true compensating variation in benefit measures.

Results show that the nonparametric models are in general superior to the parametric models for all the samples examined in regard to goodness of fit; this superiority however is not as pronounced for large mass points at zero. The percentage of the population with zero value for car travel time is correctly recovered by B-spline in cases in which the mass at zero is consistent and in which this mass is fairly small.

From a benefits estimation perspective, the results seem to confirm that theoretically "superior" models (i.e., with a superior specification) do not always yield better results in regard to computation of user benefits for nonmarginal changes. In this case, in fact, it was found that the nonparametric models always performed worse than their parametric counterparts. These results suggest that improving modeling instruments without considering the measures for which these modeling outcomes will be used might produce undesirable erroneous results.

Overall, the work highlights that notwithstanding the major advances in discrete choice modeling, and particularly in nonparametric models, significant problems remain in the application of these techniques to the characterization of taste heterogeneity and related issues of user benefit assessment. These results are still preliminary but suggest that the potentiality (benefits and limits) of the nonparametric models still needs to be fully explored. At the same time, the results also suggest the need to look more carefully at the implications of using these more advanced models for computing economic welfare. This certainly has a strong impact, not only from the research point of view but, more important, in regard to recommendations that can be drawn for practitioners.

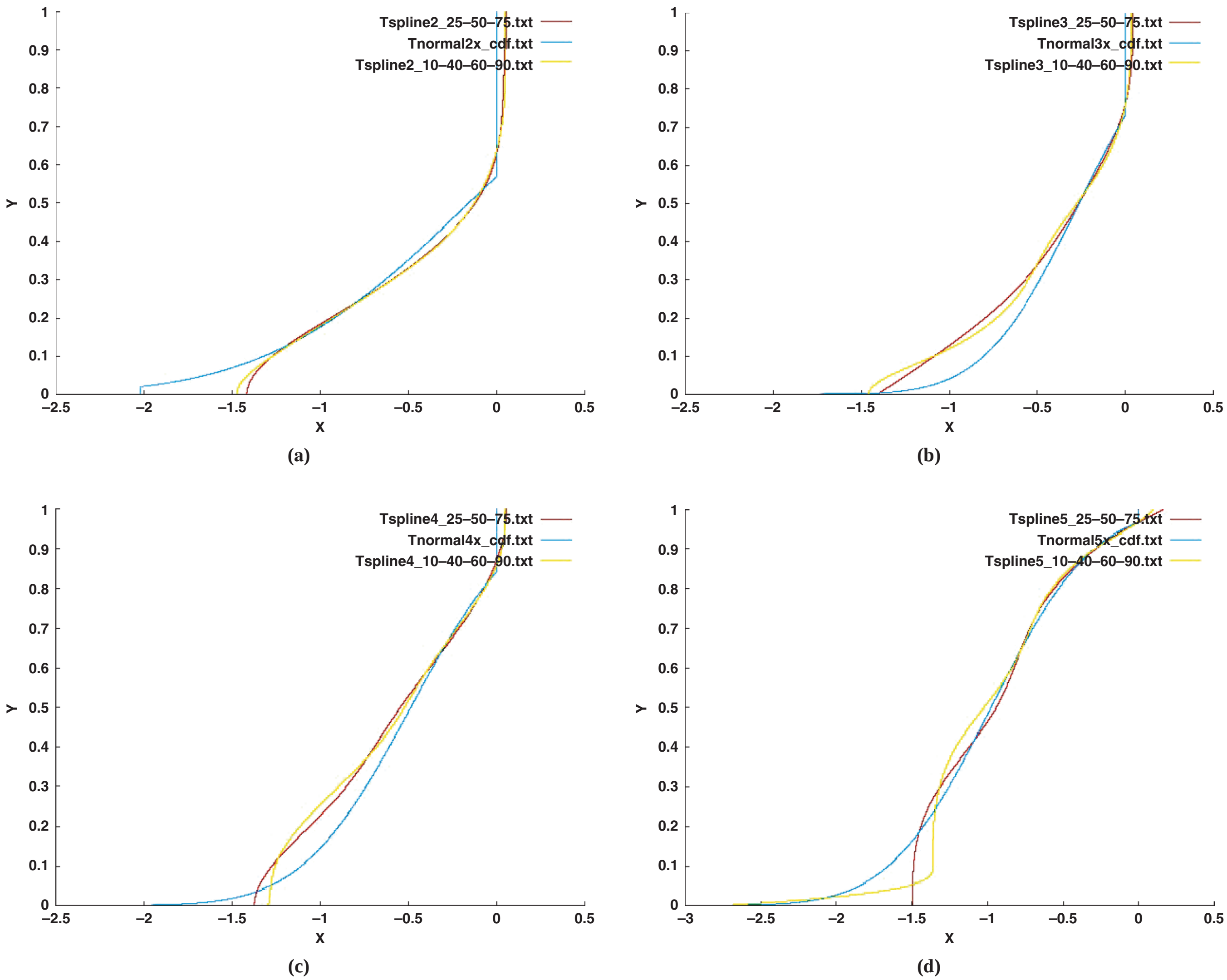

FIGURE 1 Performance of nonparametric model versus true taste distribution.

TABLE 5 Estimates of User Benefit Under Different Model Specifications

Type of Distribution	Sample 1 %	Sample 2 %	Sample 3 %	Sample 4 %
CV (estimated–simulated) versus simulated				
Normal bounded	18.52	19.14	4.20	6.78
Normal unbounded	8.53	13.78	2.20	–7.95
Nonparametric	20.32	22.91	7.59	–9.29
Logsum (estimated–simulated) versus simulated				
Normal bounded	21.34	16.75	3.55	–7.72
Normal unbounded	11.48	11.72	1.76	–8.60
Nonparametric	20.83	19.48	6.22	–10.1
χ^2 (market share simulated versus estimated)				
Normal bounded	17.86	13.94	31.50	37.80
Normal unbounded	27.38	11.54	26.90	31.16
Nonparametric	22.08	33.38	47.84	35.09

NOTE: 10,000 repetitions were used.

REFERENCES

1. McFadden, D., and K. Train. Mixed MNL Models for Discrete Response. *Journal of Applied Econometrics,* Vol. 15, No. 5, 2000, pp. 447–470.
2. Meijer, E., and J. Rouwendal. *Measuring Welfare Effects in Models with Random Coefficients.* Research Report No. 00F25. SOM Research School, University of Gröningen, Netherlands, 2000.
3. Sillano, M., and J. de D. Ortúzar. WTP Estimation with Mixed Logit Models: Some New Evidence. *Environment and Planning Part A,* Vol. 37, No. 3, 2005, pp. 525–550.
4. Train, K., and G. Sonnier. Mixed Logit with Bounded Distributions of Correlated Partworths. In *Applications of Simulation Methods in Environmental Resource Economics* (A. Alberini and R. Scarpa, eds.), Kluwer Academics Publisher, Dordrecht, Netherlands, 2005.
5. Hess, S., M. Bierlaire, and J. W. Polak. Estimation of Value of Travel-Time Savings Using Mixed Logit Models. *Transportation Research Part A,* Vol. 39, No. 3, 2005, pp. 221–236.
6. Cherchi, E., and J. W. Polak. Assessing User Benefits with Discrete Choice Models: Implications of Specification Errors Under Random Taste Heterogeneity. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1926,* Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 61–69.

7. Dong, X., and F. S. Koppelman. Mass Point Mixed Logit Model: Development and Application. Presented at 10th International Conference on Travel Behaviour Research (IATBR), Lucerne, Switzerland, 2003.
8. Hess, S., M. Bierlaire, and J. W. Polak. Discrete Mixture of GEV Models. Presented at 5th Swiss Transport Research Conference, Monte Verità/Ascona, Switzerland, 2005.
9. Burge, P., and C. Rohr. DATIV: SP Design: Proposed Approach for Pilot Survey. Tetra-Plan in cooperation with RAND Europe and Gallup A/S, 2004.
10. Fosgerau, M. Investigating the Distribution of the Value of Travel Time Savings. *Transportation Research Part B,* Vol. 40, No. 8, 2006, pp. 688–707.
11. Fosgerau, M., and M. Bierlaire. A Practical Test for the Choice of Mixing Distribution in Discrete Choice Models. *Transportation Research Part B,* Vol. 41, No. 7, 2007, pp. 784–794.
12. Jara-Díaz, S. R., and M. Farah. Valuation of User's Benefits in Transport Systems. *Transport Reviews*, Vol. 8, No. 3, 1988, pp. 197–218.
13. Cherchi, E., J. W. Polak, and G. Hyman. The Implications of Income, Taste and Substitution Effects for the Assessment of User Benefits Using Discrete Choice Models. Presented at European Transport Conference, Strasbourg, France, 2004.
14. Train, K. Recreation Demand Models with Taste Differences over People. *Land Economics,* Vol. 74, No. 2, 1998, pp. 230–239.
15. Persson, T. H. Welfare Calculations in Models of the Demand for Sanitation. *Applied Economics,* Vol. 34, No. 12, 2002, pp. 1509–1518.
16. Amador, F. J., R. M. González, and J. de D. Ortúzar. Preference Heterogeneity and Willingness to Pay for Travel Time Savings. *Transportation*, Vol. 32, No. 6, 2005, pp. 627–647.
17. Train, K. *Discrete Choice Methods with Simulation.* Cambridge University Press, Cambridge, United Kingdom, 2003.
18. Train, K. EM Algorithms for Nonparametric Estimation of Mixing Distributions. *Journal of Choice Modeling,* Vol. 1, No. 1, 2008, pp. 40–69.
19. Hicks, J. R. *A Revision of Demand Theory.* Oxford University Press, Oxford, United Kingdom, 1956.
20. McFadden, D. Disaggregate Behavioral Travel Demand's RUM Side. A 30-Year Retrospective. Prepared for 9th International Conference on Travel Behaviour Research (IATBR), Gold Coast, Australia, 2000.
21. Williams, H. C. W. L. On the Formation of Travel Demand Models and Economic Evaluation Measures of User Benefit. *Environment and Planning Part A,* Vol. 9, No. 3, 1977, pp. 285–344.
22. Daly, A. J., and S. Zachary. Improved Multiple Choice Models. In *Determinants of Travel Choice* (D. A. Hensher and M. Q. Dalvi, eds.), Saxon House, Westmead, United Kingdom, 1978.
23. Cherchi, E., and J. W. Polak. *The Implications of Income, Taste and Substitution Effects for the Assessment of User Benefits Using Discrete Choice Models.* Working paper. Centre for Transport Studies, Imperial College London, United Kingdom, 2004.
24. Cherchi, E., and J. de D. Ortúzar. Mixed RP/SP Models Incorporating Interaction Effects: Modelling New Suburban Train Services in Cagliari. *Transportation,* Vol. 29, 2002, pp. 371–395.
25. Munizaga, M. A., B. G. Heydecker, and J. de D. Ortúzar. Representation of Heteroscedasticity in Discrete Choice Models. *Transportation Research Part B,* Vol. 34, No. 3, 2000, pp. 219–240.

The Transportation Demand Forecasting Committee sponsored publication of this paper.