

Constrained nested logit model: formulation and estimation

José Luis Espinosa-Aranda¹ · Ricardo García-Ródenas¹ ·
María Luz López-García¹ · Eusebio Angulo¹

© The Author(s) 2017. This article is an open access publication

Abstract A model of traveller behaviour should recognise the exogenous and endogenous factors that limit the choice set of users. These factors impose constraints on the decision maker, which constraints may be considered implicitly, as *soft* constraints imposing thresholds on the perception of changes in attribute values, or explicitly as *hard* constraints. The purpose of this paper is twofold: (1) To present a constrained nested logit-type choice model to cope with hard constraints. This model is derived from the entropy-maximizing framework. (2) To describe a general framework to deal with (dynamic) non-linear utilities. This approach is based on Reproducing Kernel Hilbert Spaces. The resulting model allows the dynamic aspect and the constraints on the choice process to be represented simultaneously. A novel estimation procedure is introduced in which the utilities are viewed as the parameters of the proposed model instead of attribute weights as in the classical linear models. A discussion on over-specification of the proposed model is presented. This model is applied to a synthetic test problem and to a railway service choice problem in which users choose a service depending on the timetable, ticket price, travel time and seat availability (which imposes capacity constraints). Results show (1) the relevance of incorporating constraints into the choice models, (2) that the constrained models appear to be a better fit than the counterpart unconstrained choice models; and (3) the viability of the approach, in a real case study of railway services on the Madrid–Seville corridor (Spain).

✉ José Luis Espinosa-Aranda
JoseL.Espinosa@uclm.es

Ricardo García-Ródenas
Ricardo.Garcia@uclm.es

María Luz López-García
marialuz.lopez@uclm.es

Eusebio Angulo
Eusebio.Angulo@uclm.es

¹ Escuela Superior de Informática, Universidad de Castilla La Mancha, Paseo de la Universidad, 4, 13071 Ciudad Real, Spain

Keywords Constrained nested logit model · Non-compensatory behaviour · Cut-offs · Dynamic choices · Reproducing Kernel Hilbert Spaces

Introduction

Discrete choice models have long been recognized for their ability to capture a broad range of transport-related choice phenomena. For quite some time, there has been a growing interest in research into traveller behaviour, to explore the choice set formation and its representation. A proper modelling of user behaviour requires including the *endogenous* and *exogenous* factors that affect the decision-making process, as they induce constraints in the choice set formation. The *endogenous factors* are inherent to the user, limiting the *universal* choice set. The *exogenous factors* are instead originated by the decisions of other users and by the existing supply of goods or services.

Constraints imposed by endogenous or exogenous factor are called *soft* if they reduce the probability to choose a given alternative but does not completely exclude it from the choice set. *Hard* constraints instead cannot be violated and impose the exclusion of the alternative. For example in the choice of residential location, a user can eliminate from the choice set alternatives whose price is above a threshold (hard constraint due to endogenous factors), or it can be that the utility of an alternative drops if an attribute takes a value above or below a given threshold, but the alternative is still available (soft constraint due to endogenous factors). Analogously the allocation of seats between railway services and the choices of other passengers (exogenous constraint) may define the latent choice set for a specific user (hard constraint). Note that exogenous factors may induce soft constraints, e.g., the vehicle capacity in transit services (subway or bus) does not impose an upper bound in the number of passengers but represents a factor of discomfort.

As will be clarified in the next section, the literature review, most work concentrates on the use of soft constraints applied to model endogenous factors. Hard constraints have received less attention, despite their importance in modelling such endogenous factors as the interaction of supply and demand. If we are analysing a supply and demand equilibrium problem and we wish to estimate a demand model, the data available about demand is the result of choices made by that demand and the limitations imposed by the current supply. Most applications simply ignore these effects; thus, forecasts of demand for a different supply scenario from that used in the estimation are likely to violate constraints and miscalculate demand.

A notable example of the interaction of supply and demand can be seen in the problem of rolling out a refuelling infrastructure for Alternative Fuel Vehicles (AFV). This effect is known as the *Chicken-or-Egg* dilemma. The refuelling infrastructure imposes hard constraints on the user choice set. Users with no effective access to a refuelling station will not contemplate buying an AFV. Most discrete choice models designed to analyse the introduction of AF vehicles up to now have considered infrastructure as just another attribute (soft constraint). Fúnez-Guerra et al. (2016) show the problem for the Spanish case, and underlines that a discrete choice model that does not consider the refuelling infrastructure as a hard constraint significantly overestimates sales of AF vehicles as against a model which does. The hard constraints separate the effects of supply from those associated with demand allow changes in supply to be properly accounted for when making estimations. The main aim of this study is to explore an approach that can handle hard constraints in the

decision-making process to find a mathematical formulation for the specific case of (nested) logit models.

A second problem addressed in this study is how to introduce general nonlinear utilities into this type of constrained model. This is especially important when considering dynamic phenomena. The machine learning community has made extensive and successful use of Reproducing Kernel Hilbert Spaces. This paper adapts these techniques to the second problem. The challenge lies in its estimation, since there may be a parameter for each observation, rather than one for each attribute considered. For this reason, and because constrained nested logit model do not allow a closed formula, we have created a novel estimation method for this problem.

Literature review

As mentioned before most of the literature concentrates on the use of soft constraints applied to model endogenous factors. Several theoretical frameworks have been proposed to account for these constraints. A rough taxonomy classifies the approaches followed in the literature into:

1. *Implicit* choice-set approaches: These methods incorporate thresholds in the perception/availability of an alternative in the random utility choice model; this has typically been carried out by introducing thresholds (cut-offs) as penalties in the utility function (see Swait 2001; Cascetta and Papola 2001; Elrod et al. 2004; Martínez et al. 2009; Bierlaire et al. 2010 among others).
2. *Explicit* choice-set approaches: These methods consist in a two-stage representation of decision making. In the first stage, the choice-set generation is simulated. The decision-makers screen alternatives and eliminate those that do not reach the relevant attribute cut-off levels from their choice sets. In the second stage, the decision makers choose, applying compensatory decision rules, only from the alternatives remaining in the reduced choice set (see Manski 1977; Swait and Ben-Akiva 1987; Ben-Akiva and Boccara 1995; Cantillo and Ortúzar 2005; Cantillo et al. 2006 among others).

The implicit approaches propose an extension to the linear compensatory utility model, which accommodates both the use of attribute cut-offs and cut-off violations in choice modelling. The major advantage of these methods is related to the computational time required. In particular Swait (2001) incorporates attribute cut-offs into the utility maximization problem formulation. It makes it possible for the consumer to treat the constraints as soft by violating them at some cost. This approach assumes a linear penalty function.

Cascetta and Papola (2001) propose the *implicit availability/perception model (IAP)*. The choice-set of alternatives is a fuzzy set where each element has a degree of membership of the choice set.

Elrod et al. (2004) propose and test a model of decision making that integrates variations of a compensatory and two non-compensatory (i.e., conjunctive and disjunctive) decision strategies which is capable of providing probabilistic predictions for objects anywhere on a closed interval.

Martínez et al. (2009) formulate a *constrained multinomial logit (CMNL)*. This model implements cut-offs as a binomial logit function embedded in multinomial logit models. CMNL model is a heuristic that is based on convenient assumptions about the functional form of the utility function. CMNL model allows the choice domain to be constrained by as many cut-offs as required, limiting both an upper and a lower level of variables.

Therefore, Castro et al. (2013) study the estimation of the CMNL model using the maximum likelihood approach. The CMNL model appears to be suitable for general applications. Using real data, these authors found significant differences in the elasticities between compensatory MNL and semi-compensatory CMNL models.

In the explicit approach, the endogenous factors are brought together in the latent choice set models which consider a set of alternatives for each decision maker. The Manski model (Manski 1977) has served as the standard workhorse model for discrete choice modelling with latent choice sets. The problem with this approach is that it leads to the need to enumerate (exponentially) the set of all alternatives. The estimation of the two-stage models is computationally intensive and their severely restrictive assumptions impede their practical application. Several studies (see Kaplan et al. 2009, 2012) are relaxing the assumptions embedded in these models with respect to the number of alternatives and choice sets, the representation of threshold selection and independently and identically distributed error terms across alternatives at the choice stage.

Bierlaire et al. (2010) show on simple examples that CMNL model is not adequate for modelling the choice set generation process consistently with Manski's framework. Although Li et al. (2015) results are consistent with Bierlaire et al. (2010) findings, they differ in the fact that the CMNL model can successfully recover cut-off and scale parameters in the choice set probability function, while Bierlaire et al. (2010) finds that at the most only one can be recovered. Moreover, Paleti (2015) proposes higher-order approximations of the Manski approach in which the CNML model constitutes a first order approximation. That work also carries out a simulation study and shows additional order of approximation offering incremental improvement in the quality of the parameter estimates.

Less attention is paid to the *exogenous* factors. This problem is studied in the analysis of endogeneity in choice modelling (Louviere et al. 2005). The concept of endogeneity refers to the fact that the individual choice decisions may depend on themselves. Ding et al. (2012) show in a behavioural experiment that some respondents are willing to take a utility penalty (*soft* cut-off) rather than eliminate an alternative when a cut-off violation occurs (*hard* cut-off). However, with exogenous constraints, such as social, temporal, spatial and resource constraints, their fulfilment is mandatory.

Endogeneity may appear for a variety of reasons such as the omission of unobservable variables or the ability of individuals to influence the formation of the choice sets. In this paper, the second reason is analysed. Recently De Grange et al. (2015) have proposed a logit model to explicitly include endogeneity in attributes (explanatory variables) due to network externalities or social interactions. This approach tackles endogeneity with a fixed-point equation.

The previous models consider linear utility functions. In a dynamic choice modelling context it is essential to consider non-linear utility functions when changes in demand trend are to be captured. Other reasons to support the use of non-linear specifications of the utility function are found in the work of Cherchi and Ortúzar (2002) where they test different specifications of the utility function for a new train service design. These authors found that the non-linear specifications appear to be more suitable as not only are better model results obtained, but also the real distribution of the error terms is revealed. In the context of health economics, Van Der Pol et al. (2014) show that welfare estimates are sensitive to different utility function specifications. The results showed that the willingness to wait for hip and knee replacement (WTW) for the different patient profiles varied considerably. Assuming a linear utility function led to much higher estimates of marginal rates of substitution (WTWs) than with non-linear specifications.

Despite the importance of specifying the utility function, this matter has received limited attention in the literature on dynamic choice modelling. Popuri et al. (2008) introduced continuous (dynamic) systematic utility functions for departure time modelling, via sinusoidal functions which interact with covariates.

Anas (1983) demonstrated that *information minimizing* (or *entropy-maximizing modelling*) and *utility maximizing* (*behavioural demand modelling*) should be seen as two equivalent views of the same problem. This author proves that the doubly-constrained gravity model is identical to a multinomial logit model of joint origin-destination choice. Donoso and de Grange (2010) give an interpretation of the entropy maximization problem in the context of microeconomic modelling, attempting to explain the origin of the two problems' equivalence.

In this paper we adopt the entropy-maximizing approach, a recent literature review of this area can be found in Swait and Marley (2013). The entropy-maximizing approach makes possible the addition of non-linear constraints in its formulation. These constraints introduce new information in the forecasting process in order to represent the complex issue of interrelation modelling between the decisions of individuals. The key difference between our entropy maximization problem and those presented in Anas (1983), Donoso and de Grange (2010), Grange et al. (2013) and De Grange et al. (2015) is that general utilities are considered, and they do not necessarily have linear attributes.

Summary and contributions of this paper

This paper makes two major contributions.

- The paper proposes a novel approach for modelling people's behaviour through a discrete choice process considering the existence of hard constraints. Specifically, we propose a mathematical formulation of the nested logit model to cope with hard constraints. We provide a Monte Carlo simulation study on an application with hard constraints to show that the demand estimations performed by the logit-type models are in error when they predict new scenarios in which the current hard constraints in the base scenario of the estimation are modified. This disadvantage does not show up in the constrained models.
- Secondly, we address the building of a general framework to specify dynamic non-linear utilities. The main feature of this approach is that the specification of the utility function is not centered on a specific functional form (linear, polynomial, sinusoidal, etc) but on belonging to a specific space of functions, the so-called Reproducing Kernel Hilbert Spaces (RKHS). The essential advantage of this design is the search for the most suitable shape for the utility within the set of functions for the problem at hand. Furthermore, we propose a method for estimating the constrained nested model with general utilities based on the novel point of view of considering a subset of utilities as parameters for the estimation instead of the classical weightings of the attributes. This approach has been illustrated with the modelling of the selection of railway services.

The paper is organised as follows. “[The constrained nested logit model](#)” section formulates the constrained nested logit model. In this section, the RKHS are defined to represent generic utility functions. Furthermore the Tikhonov regularization method is explained for the estimation of these functions. “[Estimation of the CNL model](#)” section discusses the procedure followed to estimate the constrained nested logit model with this type of non-linear utility. Fifth section analyses numerically the logit-type models with the constrained counterparts and it illustrates the methodology by numerically solving a railway service

selection problem. Finally, last section concludes with a discussion of our findings and future work.

The constrained nested logit model

This section formulates the constrained nested logit model. The proposed scheme is utilized to introduce constraints in the users' individual decision-making processes which may influence their behaviour.

Formulation of the constrained nested logit

Logit-type discrete choice models are developed using random utility models (see McFadden 1974; Ortúzar and Willumsen 2011) or a maximum entropy optimization problem (see Anas 1983; Donoso and de Grange 2010; Grange et al. 2013; De Grange et al. 2015). The first approach obtains the multinomial logit model from the assumption that the random component of each utility function is independent and identically Gumbel-distributed. Anas (1983) gives a proof of the equivalence between both frameworks.

This section presents a constrained nested logit-type choice model derived from the entropy-maximizing framework. A selection process similar to the one shown in Fig. 1 is considered. This type of model represents a decision tree for the discrete choice problem, in which the root of the tree represents the first choice users of type $\ell \in \mathcal{L}$ may make between alternatives (denoted by the index $m \in \mathcal{S}^\ell$), and in each branch of the tree lies another selection process in which users may select among the available sub-alternatives (denoted by the index $s \in \mathcal{S}_m^\ell$). This type of model has been widely used in transport modelling (see Oppenheim et al. 1995; Ortúzar and Willumsen 2011; Fernández et al. 1994; García and Marín 2005).

It is assumed that there exist various types of individual $\ell \in \mathcal{L}$. The parameter \hat{g}^ℓ represents all the individuals of type ℓ who interact with the system, the variable $g^{m\ell}$ is the number of individuals of type ℓ who choose alternative m in the first level, the variable $g_s^{m\ell}$ is the number of individuals of type ℓ who choose alternative s in the second level knowing

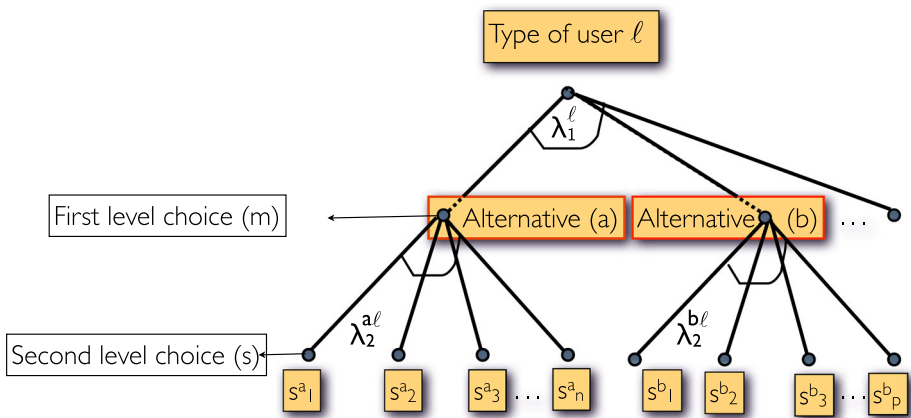


Fig. 1 Nested logit model

that alternative m has been selected previously and $V_s^{m\ell}$ represents the deterministic part of the indirect utility perceived by individual type ℓ conditioned upon choosing alternative m in the first level and alternative s in the second level.

The entropy maximization problem allows the inclusion of constraints, leading to the Constrained Nested Logit model (CNL). These constraints may introduce new factors in the forecasting process which could influence the behaviour of the users. Therefore, the CNL model is formulated as follows:

$$\begin{aligned}
 &\text{minimize} && \sum_{\ell \in \mathcal{L}} \sum_{m \in \mathcal{S}^\ell} \left[\eta_1^{m\ell} g^{m\ell} (\ln g^{m\ell} - 1) + \eta_2^{m\ell} \sum_{s \in \mathcal{S}_m^\ell} g_s^{m\ell} (\ln g_s^{m\ell} - 1) \right. \\
 & && \left. - \sum_{s \in \mathcal{S}_m^\ell} V_s^{m\ell} g_s^{m\ell} \right], \\
 &\text{subject to:} && \sum_{m \in \mathcal{S}^\ell} g^{m\ell} = \widehat{g}^\ell; \quad \forall \ell \in \mathcal{L} \quad (\Phi_\ell) \\
 & && \sum_{s \in \mathcal{S}_m^\ell} g_s^{m\ell} = g^{m\ell}; \quad \forall \ell \in \mathcal{L}, \forall m \in \mathcal{S}^\ell \quad (\Theta_{m\ell}) \\
 & && \widehat{h}_r(\mathbf{g}) \leq b_r; \quad \forall r \in \mathcal{R} \quad (\mu_r)
 \end{aligned} \tag{CNL}$$

where $\mathbf{g} = (\dots, g_s^{m\ell}, \dots)$ is the disaggregated demand by sub-alternatives, $\eta_1^{m\ell} = \frac{1}{\lambda_1} - \frac{1}{\lambda_2^{m\ell}}$, $\eta_2^{m\ell} = \frac{1}{\lambda_2^{m\ell}}$ and λ_1^ℓ and λ_2^ℓ are scalars associated with the variance of the error term of the utilities. Moreover, Φ_ℓ , $\Theta_{m\ell}$ and μ_r are dual variables associated with the constraints.

The first and second constraints of the model express the logical requirements that the sum of the users across each branch of the first level must be the total number of users, and that in each branch, the sum of the users in this particular branch must be the total number of users who previously selected the sub-alternatives. The later constraints represent the hard constraints which are imposed upon the choice process. These constraints are formulated mathematically via the functions \widehat{h}_r and the parameters b_r , which are known and depend on the problem to be solved.

We now see some examples of the possibilities made available by the inclusion of these constraints. Suppose each class ℓ of users consists of only one individual i , $\ell = \{i\}$. In this case $\widehat{g}^\ell = 1$ and the variables $g^{m\ell}$, $g_s^{m\ell}$ represent respectively the probability that individual ℓ chooses alternative m and that, having chosen alternative m , chooses sub-alternative s . The first observation is that the model takes all user decisions collectively and so can capture the interactions between the decisions of users.

Example 1 (Exogenous factors: capacity constraints) Suppose we are modelling flight choices of airline users. At the higher level we decide between standard and low-cost companies. At the lower level users can choose between the different flights available. Each flight s has a capacity limit depending on the number of seats K_s . There exists a set of users L_s , who are likely to choose flight s because of the origin-destination of their trip. In this problem capacity constraints will be active on many flights, especially low-cost ones, and this will significantly affect user choice. This leads to

$$\sum_{\ell \in L_s} g_s^{m\ell} \leq K_s \tag{1}$$

being imposed on the demand estimation, which shows that the expected number of passengers (the left-hand side of the expression) on flight s cannot be greater than the number of seats available on the flight. For some users the choice set will be affected by capacity and they must choose from among available options. In this case, the demand model embeds an equilibrium problem in which all users compete against one another to get tickets.

Example 2 (Exogenous factors: limited availability products) Suppose a set of consumers who make purchases over a period of time. We consider that these individuals are ordered according to the instant at which they make the purchase $\ell_1 < \dots < \ell_n$. If the available resources of each product (alternative s) are limited, these products could run out after a certain time, and that alternative is no longer available to other consumers. Let K_s^ℓ be the number of available units of product s when purchaser ℓ makes the purchase. The following must be satisfied:

$$g_s^{m\ell} \leq K_s^\ell \tag{2}$$

In this example, the optimization problem CNL is separable for each individual ℓ , leading to n independent problems in which there is no interaction between individuals.

The parameters K_s^ℓ are unknown in many practical situations and in their place the initial quantity K_s of product s is known. In this case we replace constraint (2) with the estimation of available capacity:

$$g_s^{m\ell_{j+1}} \leq K_s - \sum_{k=1}^j g_s^{m\ell_k} \tag{3}$$

The decisions of consumers affect the decisions of those preceding them. This leads to an iterative solution process for the CNL. Suppose the variable $g_s^{m\ell_k}$ with $k = 1, \dots, j$, is known, and we calculate the right side of Eq. (3) and solve the CNL problem to calculate the probabilities of choice $g_s^{m\ell_{j+1}}$ of the consumer ℓ_{j+1} , subsequently iterating the process again.

Example 3 (The polarized logit model) Grange et al. (2013) propose the so-called *polarized logit model* which consists of introducing one instrumental constraint in the MNL. The motivation to introduce this constraint is to force the prediction of choice probabilities towards values of 0 or 1. The polarized logit model may be extended to nested logit models considering in the CNL the constraints

$$\hat{h}_m(\mathbf{g}) = \sum_{s \in S_m^\ell} g_s^{m\ell} (1 - g_s^{m\ell}) \leq \varepsilon, \quad \forall \ell, m. \tag{4}$$

Example 4 (Endogenous factors) Martínez et al. (2009) discuss potential application of constrained logit models. In modelling the transport system, the constraints taking into account the endogenous factors associated with thresholds in the attributes, such as minimum activity level at destination for attracting trips, maximum waiting, maximum travel expenditure, and access times to public transport. Examples in location and land use modelling are housing choices which are constrained by the income budget or in relevant location options where the cut-offs help to model the scope of the spatial search of an individual.

Assume that the utility function depends on a set of K attributes, denoted by vector X . Each alternative s is characterised by the vector of attributes $(\dots, X_{s,k}, \dots)$. A user of type ℓ endogenously screens the universal choice set and eliminates all alternatives whose attribute vector lies out of the consumer's choice domain. For example, the user may eliminate the alternatives with a price higher than a self-imposed maximum expenditure b_k^ℓ . The set of constraints

$$\delta_s^\ell X_{s,k} \leq b_k^\ell \tag{5}$$

$$g_s^{m\ell} \leq \delta_s^\ell \tag{6}$$

$$\delta_s^\ell \in \{0, 1\} \tag{7}$$

defines the individual's feasible domain.

The binary variable δ_s^ℓ indicates the validity of the alternative s for the the user type ℓ . The lower cut-offs can be analogously introduced in the CNL model.

Equilibrium issues

The CNL model represents an equilibrium between users. They compete for the existent resources (alternatives) considering their preferences and the imposed constraints which may affect their choices, like the capacity of the system. This fact leads to an equilibrium situation in which each user cannot improve their utility by selecting a different alternative.

Therefore, "Appendix 1" proves that the solution of CNL satisfies the classic Nested Logit probability equations, but show that the constraints introduced in the CNL model penalize the utilities of the lower level, producing a modification to the forecasting of demand depending on the active constraints. A function of the multipliers, the named $W_s^{m\ell'}$, can be interpreted as the shadow price that the type of user ℓ' must pay for choosing the alternative s . Alternative s consumes scarce resources that a set of users must compete for, and this is the price, expressed in terms of utility, that users are prepared to pay to choose that alternative.

Non-linear utility specifications using Reproducing Kernel Hilbert Spaces

Linear utility among attributes is the most commonly-used approach in literature. In this paper the temporal nature of attributes is considered, so non-linear utilities appear to be better suited to the problem. We propose a framework to specify non-linear utility functions based on Reproducing Kernel Hilbert Spaces (RKHS). A quick introduction to RKHS may be found at Daumé (2004). We shall begin by giving the following definition:

Definition 1 (*Reproducing Kernel*) Let \mathcal{H} be a real Hilbert space of functions defined in a compact set $X \subset \mathbb{R}^p$ with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$. A function $K : X \times X \rightarrow \mathbb{R}$ is called a Reproducing Kernel of \mathcal{H} if:

1. $K(\cdot, x) \in \mathcal{H}$ for all $x \in X$.
2. $f(x) = \langle f, K(\cdot, x) \rangle_{\mathcal{H}}$ for all $f \in \mathcal{H}$ and for all $x \in X$.

A Hilbert space of functions that admits a reproducing Kernel is called a RKHS. The reproducing Kernel of a RKHS is uniquely determined. Conversely, if a function $K :$

$X \times X \rightarrow \mathbb{R}$ is positive definite and symmetric (Mercer kernel), then it generates a unique RKHS in which the given kernel acts like a Reproducing Kernel.

We assume that the systematic utility function $V^\ell : X \subset \mathbb{R}^p \rightarrow \mathbb{R}$, where X is the feasible set for the attributes, belongs to a given RKHS \mathcal{H}_K . For simplicity only one type of user ℓ is considered, thus this index is eliminated in this subsection. This assumption leads to the relationship

$$V(x) \in \mathcal{H}_K \tag{8}$$

The utility function can be expressed as a linear combination of the basis of the space \mathcal{H}_K . The kernel function defines a basis $\{K(x, y)\}_{y \in X}$ of the vectorial space \mathcal{H}_K , and thus

$$V(x) = \sum_{y \in Y \subset X} \alpha_y K(x, y) \tag{9}$$

Note that the choice of the kernel function $K(x, y)$ plays the same role as the selection of the specification of the functional form of a non-linear utility function. It is convenient to consider reproducing kernels that lead to spaces \mathcal{H}_K which have a large range of functions in order to represent the utility function $V(x)$ appropriately.

The Eq. (9) gives the functional form of the utilities. Traditionally in the literature, the weightings α_y are the parameters of the model and the utilities are computed from the estimated parameters. In this paper we interchange these roles. To illustrate this, consider the linear utility function:

$$V(x) = \alpha_0 + \alpha^T \cdot x \tag{10}$$

The most common steps followed in the estimation of the utilities are: firstly know the attributes of the alternatives x_s with $s \in \cup_m \mathcal{S}_m$, secondly estimate the parameters (α_0, α^T) of the utility function (10) and thirdly calculate the utilities (see Ben-Akiva and Boccara 1995) by evaluating the utility function over the attributes, i.e

$$V_s^m = V(x_s) = \alpha_0 + \alpha^T \cdot x_s, \quad s \in \mathcal{S}_m \tag{11}$$

In our approach the parameters to be calibrated are a subset of utilities V_s^m and the weightings α_y are computed from these estimated parameters. This view leads to a different order in the estimation process of the utility function:

- Step (i) know the attributes values of the alternatives x_s with $s \in \cup_m \mathcal{S}_m$,
- Step (ii) estimate the utilities V_s^m on a subset of alternatives $\mathcal{D}_1 \subset \cup_m \mathcal{S}_m$, and
- Step (iii) calculate the parameters (α_0, α^T) from the estimated utilities.

The fact that the CNL requires the parameters V_s^m in order to be used focuses the estimation process on these parameters, and so for the alternatives \mathcal{D}_1 it is not necessary to know the functional form and the attributes which are relevant for the decision maker. Moreover, the parameters V_s^m have a clear interpretation but the interpretation of the weightings α_y is unclear.

The Tikhonov Regularization Theory allows the realization of Step (iii) for specifications of utilities based on RKHS. Now we briefly describe Tikhonov discrete regularization theory by RKHS for the problem at hand. The General Theory of Tikhonov Regularization is explained in the book of Tikhonov and Arsenin (1997) and the General Theory of RKHS is defined in Aroszajn (1950).

Assume that the utilities of a subset of alternatives $\mathcal{D}_1 \subseteq \cup_m \mathcal{S}_m$ are known and let n be the cardinality of the set \mathcal{D}_1 . We denote by $\{U_s\}_{s \in \mathcal{D}_1}$ these estimates. Note that they have not been denoted by V as they may contain certain random errors.

Let

$$Y_n := \{x_s\}_{s \in \mathcal{D}_1} \subset X$$

be the set of attributes of the utilities to be estimated and let \mathcal{W}_n be a random sample. That is:

$$\mathcal{W}_n := \{(x_s, U_s) \in X \times \mathbb{R}\}_{s \in \mathcal{D}_1}.$$

Tikhonov Regularization Theory considers the function space

$$\mathcal{V}_n := \text{span}\{K(\cdot, x_s) : s \in \mathcal{D}_1\} \tag{12}$$

where span is the linear hull and projects $V(x)$ onto this space by using the sample \mathcal{W}_n . Tikhonov Regularization Theory makes a stable reconstruction of $V(x)$ by solving the following optimization problem:

$$V^* := \arg \min_{V \in \mathcal{V}_n} \frac{1}{n} \sum_{s \in \mathcal{D}_1} (V(x_s) - U_s)^2 + \gamma \|V\|_{\mathcal{H}_K}^2 \tag{13}$$

where $\gamma > 0$, and $\|V\|_{\mathcal{H}_K} = \langle V, V \rangle_{\mathcal{H}_K}^{1/2}$ represents the norm of V in \mathcal{H}_K . The solution V^* of (13) is called the Regularized γ -Projection of $V(x)$ onto \mathcal{H}_K associated with the sample \mathcal{W}_n .

The representation theorem gives a closed form solution of $V^*(x)$ for the optimization problem (13). This theorem was introduced by Kimeldorf and Wahba (1970) in a spline smoothing context and has been extended and generalized to the problem of minimizing risk of functions in RKHS, see Schölkopf et al. (2001) and Cox and O’Sullivan (1990).

Theorem 1 (Representation) *Let \mathcal{W}_n be a sample of $V(x)$, let K be a (Mercer) kernel and let $\gamma > 0$. Then there is a unique solution V^* of (13) that admits a representation by*

$$V^*(x) = \sum_{s \in \mathcal{D}_1} \alpha_s K(x, x_s), \quad \text{for all } x \in X, \tag{14}$$

where $\alpha = (\alpha_1, \dots, \alpha_n)^T$ is a solution to the system of linear equations:

$$(\gamma n \mathbf{I}_n + K_x) \alpha = U, \tag{15}$$

where \mathbf{I}_n is the identity matrix $n \times n$, $U = (U_{s_1}, \dots, U_{s_n})^T$ and the matrix K_x is given by $(K_x)_{s's} = K(x_{s'}, x_s)$. The expression (14) leads to the estimate of $V(x)$ in Y_n

$$\widehat{V}^* = K_x \alpha \tag{16}$$

The following example illustrates numerically how to carry out step (iii) using Tikhonov Regularization Theory.

Example 1 (A numerical example of the use of Tikhonov Regularization Theory) Suppose that the utilities have been estimated for a given set of attributes x_s (in this case, the time of departure t of a transit service), with the data shown in Table 1.

Table 1 Initial data

	Set of alternatives $s \in \mathcal{D}_1$			
	s_1	s_2	s_3	s_4
t_s	9	10	12	16
U_s	3	5	2	4

In this approach a given functional expression is not specified for the utility function, but it is assumed that it belongs to a RKHS. For example, we consider the RKHS \mathcal{H}_{K_1} with a Gaussian reproducing kernel K_1 and \mathcal{H}_{K_2} with a multi-quadratic kernel. Specifically for this example the following kernels are defined where the values of the parameters are set:

$$K_1(x, y) = \exp((x - y)^2) \tag{17}$$

$$K_2(x, y) = \sqrt{|x - y|} \tag{18}$$

These selections allow the possibility of estimating two utility functions as follows:

$$V^1(t) = \sum_{s=1}^4 \alpha_s^1 K_1(t, t_s) = \alpha_1^1 K_1(t, 9) + \alpha_2^1 K_1(t, 10) + \alpha_3^1 K_1(t, 12) + \alpha_4^1 K_1(t, 16)$$

$$V^2(t) = \sum_{s=1}^4 \alpha_s^2 K_2(t, t_s) = \alpha_1^2 K_2(t, 9) + \alpha_2^2 K_2(t, 10) + \alpha_3^2 K_2(t, 12) + \alpha_4^2 K_2(t, 16)$$

Moreover, to calculate the parameters α , the linear system (15) must be solved. The value of the parameter $\gamma = 0.001$ has been used. The values of α obtained by this method can be seen in Table 2.

Figure 2 represents both utility functions. In the first case the space \mathcal{H}_{K_1} is considered, while the second case considers the space \mathcal{H}_{K_2} , obtaining different utility functions starting from the same set of data. In conclusion, this process allows a non-linear utility function to be derived from a set of known utilities.

We end the section with the following remarks

- *Remark 1. Functional expression* It is worth noting that calculating the vector of parameters α by solving the system of equations (15) allows the analytical expression of $V(x)$, Eq. (14), to be known, and it is possible to calculate the marginal utilities $\frac{\partial V^*}{\partial x_s}$. This allows the subjective values of travel time (SVT) to be calculated (see Jara-Díaz

Table 2 Calculated parameters α

	Gaussian coefficients ($i = 1$)	Multi-quadratic coefficients ($i = 2$)
α_1^i	1.3582	1.4401
α_2^i	4.4476	-1.1676
α_3^i	1.9107	1.5238
α_4^i	3.9841	0.5754

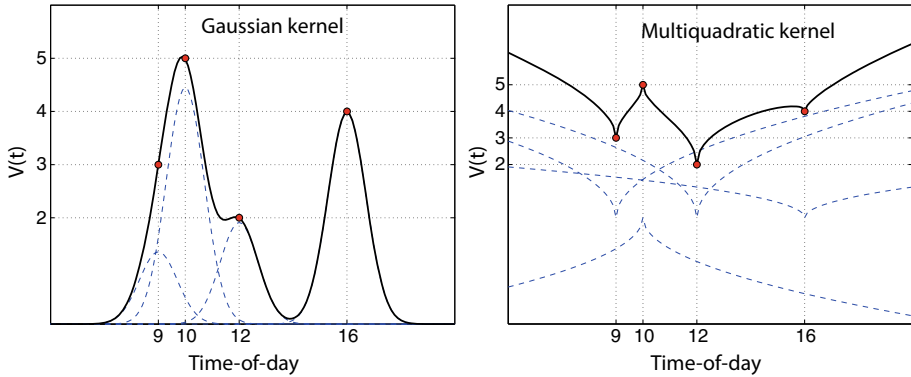


Fig. 2 Calculated utility functions

2000; Amador et al. 2005). Moreover, it allows any utility $V(x_s)$ to be estimated if the vector of attributes of the alternative s , x_s , is known.

- *Remark 2. Various user types* If there were several users ℓ it could happen that one alternative s were common to some of them. This may introduce ambiguities. We modify the definition of set \mathcal{D}_1 to avoid this, considering that its elements are pairs of the form (s, ℓ) , thus leaving nothing undefined. In the case where there are various user types we would estimate a utility function for each of them in this way:

$$V^{*\ell}(x) = \sum_{(s,\ell) \in \mathcal{D}_1} \alpha_s^\ell K(x, x_s^\ell), \quad \text{for all } x \in X, \tag{19}$$

Estimation of the CNL model

CNL parameters are $V, \lambda = (\dots, \lambda_1^\ell, \dots, \lambda_2^{m\ell}, \dots), b_r, \hat{g}^\ell$. It is assumed that the upper bounds b_r of the constraints and \hat{g}^ℓ are known. The remaining parameters (V, λ) must be estimated. In this section a generic estimation methodology is described. As CNL is a strictly convex program (assuming that the functions \hat{h}_r are convex) it poses a single optimum and CNL implicitly defines a function, which obtains the disaggregation of the demand by alternatives for each pair (V, λ) . This is depicted thus:

$$g = \text{CNL}(V, \lambda) \tag{20}$$

The main idea of the approach presented in this paper for estimating the CNL model is to select and estimate a subset of utilities \hat{V}_1 which will be used for calculating the other utilities \hat{V}_2 , repeating this process iteratively, changing the values of \hat{V}_1 with the objective of approximating the demand estimated by the CNL to the real known values.

Assume a sample of N decision-makers, N^ℓ is the number of individuals of type ℓ . Also suppose that the number of individuals of type ℓ who select alternative $s \in \mathcal{S}_m^\ell$ is denoted by $N_s^{m\ell}$ and it is known for a set of combinations $(s, \ell) \in \mathcal{D}_0$. Denote $\mathbf{N} = (\dots, N_s^{m\ell}, \dots)$ with $(s, \ell) \in \mathcal{D}_0$.

Assume that the vector of attributes for each alternative s is known and is denoted by x_s . Let \mathcal{D}_1 be the alternative subset (s, ℓ) in which the utility will be estimated, and let \mathcal{D}_2 be the alternatives in which the utility will be calculated from the estimated function $V^{*\ell}(x)$ by using the Tikhonov regularization theory described in “Non-linear utility specifications using Reproducing Kernel Hilbert Spaces” section. The set of alternatives is decomposed in $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2$ with $\mathcal{D}_1 \cap \mathcal{D}_2 = \{\emptyset\}$.

As a first step, the utility vector is estimated

$$\widehat{V}_1 = (\dots, U_s^{m\ell} \dots); \quad (s, \ell) \in \mathcal{D}_1 \tag{21}$$

and in the second stage the utility function $V^{*\ell}(\cdot)$ is calculated using Eq. (19) and the Eq. (15) for each ℓ and all non-estimated utilities $U_s^{m\ell} = V^{*\ell}(x_s^\ell)$ with $(s, \ell) \in \mathcal{D}_2$ are calculated. Denote

$$\widehat{V}_2 = (\dots, V^{*\ell}(x_s^\ell), \dots); \quad (s, \ell) \in \mathcal{D}_2 \tag{22}$$

The above two stages are schematically represented by

$$\widehat{V}_2 = \mathcal{H}(\widehat{V}_1) \tag{23}$$

Using Eq. (23), the Eq. (20) can be rewritten as:

$$g = \text{CNL}\left(\left(\widehat{V}_1, \mathcal{H}(\widehat{V}_1)\right), \lambda\right) \tag{24}$$

Finally the estimation problem can be stated as:

$$\begin{aligned} &\text{minimize}_{(\widehat{V}_1, \lambda)} F(g, \mathbf{N}) \\ &\text{subject to: } g = \text{CNL}\left(\left(\widehat{V}_1, \mathcal{H}(\widehat{V}_1)\right), \lambda\right) \end{aligned} \tag{25}$$

where F is a *similarity function* between predicted demand by CNL model, g , and the observed values, \mathbf{N} . It is worth noting that the parameters to be estimated are the utility vector \widehat{V}_1 and the vector of scale parameters λ .

An approach widely used for $F(g, \mathbf{N})$ is the minus log-likelihood (LL) function, and it leads to the maximum likelihood (ML) estimation problem:

$$\begin{aligned} &\max_{(\widehat{V}_1, \lambda)} \sum_{(s, \ell) \in \mathcal{D}_0} N_s^{m\ell} \ln(g_s^{m\ell} / N^\ell) \\ &\text{subject to: } g = \text{CNL}\left(\left(\widehat{V}_1, \mathcal{H}(\widehat{V}_1)\right), \lambda\right). \end{aligned} \tag{26}$$

In some cases, as in the numerical Experiment 2 of this paper, disaggregated values by alternatives $N_s^{m\ell}$ are not known. In these cases the least squares method can be adapted to the data.

The likelihood maximization or generalized least squares technique is achieved by embedding the computation of F within a non-linear optimization framework as shown in Fig. 3. The estimation problem of the CNL is formulated as a bi-level optimization model and Fig. 3 shows the application of free-derivative optimization methods to this problem. This calculation scheme is conceptual and susceptible of many implementations. The convergence to the optimal parameters is not guaranteed and it will depend on the free-derivative optimization method applied.

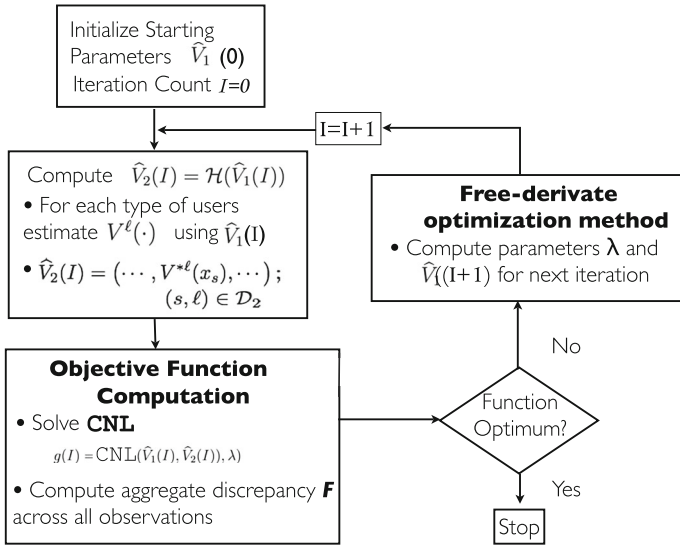


Fig. 3 Schematic representation of CNL estimation procedure

It is worth noting that there exist infinite solutions to the estimation problem (25). Ben-Akiva and Lerman (1985) indicates that discrete choice models have two sources of over-specification. The first source is due to the scale indeterminacy because the units of the underlying latent utilities are not observed, and origin indeterminacy because the zero of the latent utility scale is not observable, so one must be set. The sources of over-specification of the parameters for the CNL proposed are explained in “Appendix 2”.

Numerical analysis

This article describes a methodology for introducing constraints on the Nested Logit (NL) model and a further methodology for estimating it. The first key question to be analysed is how omitting the constraints where they are relevant to the problem affecting the accuracy of predictions of the NL model. The second step is to assess the estimation methodology set out in Sect. 4. The estimation problem to be solved has a bi-level nature in which, to evaluate the objective function, the CNL model must be solved. The proposed method is conceptual and does not specify any given optimization algorithm. The only assumption about the algorithm is that it be derivative-free, as the problem is bi-level and the so there is no guarantee of convergence. Moreover, the bi-level nature presents the challenge of addressing the computational burden in real applications.

These questions have been analysed in the following two experiments.

- *Experiment 1* This experiment was carried out on synthetic data. The objective is to compare the logit model and the nested logit model with its constrained counterparts. As well as checking numerically that the constrained models display better fitting to the data, we show that the predictions of the unconstrained models may not be reliable in scenarios where the constraints are changed.
- *Experiment 2* This experiment is a real application consisting of fitting the CNL model to the rail services choice problem. In this experiment we describe a metaheuristic

methodology based on hybridization of the *Particle Swarm Optimization* with the Nelder–Mead method. The main issue analysed in this experiment is the applicability of the proposed method and its convergence.

A simulation study (*Experiment 1*)

An important class of problems in which constraints appear in the modelling of demand is dynamic pricing with limited inventories, Boer (2015). This type of model enables certain firms, such as in the airline industry, to increase revenue by better matching supply with demand. Chen and Chen (2015) consider that most dynamic pricing problems share the following three main characteristics: (1) products are typically time-sensitive with a fixed selling season, (2) a given and finite amount of inventory of a product available at the beginning of the selling season and (3) multiple prices in the selling season. Feature (2) imposes (exogenous) capacity constraints for the consumers.

In this section we consider the fitting of a demand model to a dynamic pricing problem. The aim is to assess how the constraints affect user choices. For this reason we have greatly simplified the problem and we have considered fixed prices and thus constant utilities during the selling period.

We consider that a given railway service has $K_1 = 25$ seats for first class and $K_2 = 40$ for second class. The users arrive according to a Poisson process with arrival rate $\lambda = 0.8$ min (mean time between arrivals). We assume that the sales process begins 60 min before the departure time. The probability of a user desiring to travel in first class is 0.2 and to travel in second class is 0.8 (i.e the utilities are not time-sensitive). If on buying the ticket there were no seats available in the desired class, the user would choose, with probability 0.5, to change the class, and with the other 0.5 not to make the trip. By a Monte Carlo simulation we have generated data for 25 different days which can be downloaded from <http://bit.ly/simData>.

We have adjusted the MNL model and the NL shown in Fig. 4 to this problem.

These two models have been adjusted with and without capacity constraints (i.e number of available seats). We have considered that the index ℓ represents one individual and these are ordered according to the instant t_ℓ at which they buy the ticket. The decisions of the individuals who arrive before instant t_ℓ affect the choice-set of individual ℓ , since they may have taken all the seats in one or other class. If we let K_i^ℓ be the residual capacity (number of available seats) of the i – th class when user ℓ buys his ticket, the constrained MNL and NL models are formulated as:

$$\begin{aligned}
 &\text{minimize} && \sum_{\ell} \sum_{s \in \{1,2,Ref\}} [\eta_1^{\ell} g_s^{\ell} (\ln g_s^{\ell} - 1) - V_s^{\ell} g_s^{\ell}], \\
 &\text{subject to:} && \sum_{s \in \{1,2,Ref\}} g_s^{\ell} = 1; \quad \forall \ell \\
 &&& g_s^{\ell} \leq K_s^{\ell}; \quad s \in \{1, 2\}, \forall \ell
 \end{aligned}$$

[Constrained MNL]

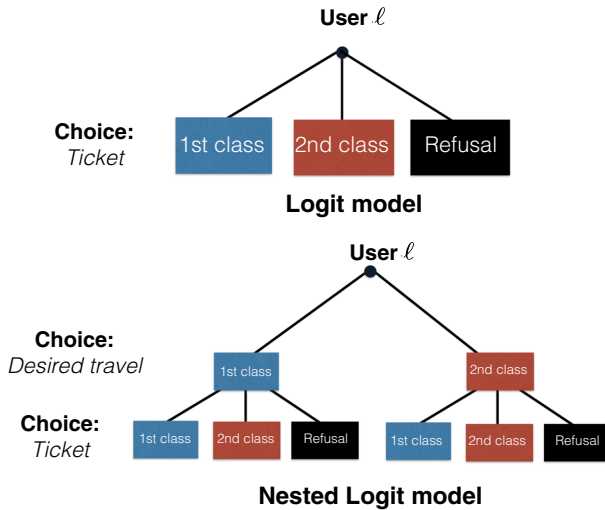


Fig. 4 Adjusted models

$$\begin{aligned}
 &\text{minimize} && \sum_{\ell} \sum_{m \in \{1,2\}} \left[\eta_1^{m\ell} g^{m\ell} (\ln g^{m\ell} - 1) + \eta_2^{m\ell} \sum_{s \in \{1,2,Ref\}} g_s^{m\ell} (\ln g_s^{m\ell} - 1) \right. \\
 & && \left. - \sum_{s \in \mathcal{S}_m} V_s^{m\ell} g_s^{m\ell} \right], \\
 &\text{subject to:} && \sum_{m \in \{1,2\}} g^{m\ell} = 1; \forall \ell \\
 & && \sum_{s \in \{1,2,Ref\}} g_s^{m\ell} = g^{m\ell}; m \in \{1,2\}, \forall \ell \\
 & && g_s^{m\ell} \leq K_s^{\ell}; m, s \in \{1,2\}, \forall \ell
 \end{aligned}$$

[ConstrainedNL]

We used the parameters $\eta_1^{\ell} = 1$ in the the MNL and CMNL models and the values $\lambda_1^{\ell} = 1, \lambda_2^{1,\ell} = 2, \lambda_2^{2,\ell} = 3$ and thus $\eta_1^{1\ell} = \frac{1}{2}, \eta_1^{2\ell} = \frac{2}{3}, \eta_2^{1\ell} = \frac{1}{2}, \eta_2^{2\ell} = \frac{1}{3}$ for the NL and CNL models.

The first numerical trial has the aim of assessing the capacity of the models to describe the data. To this end we estimate the models by maximum likelihood using the following three classical functional specifications for the utilities:

$$\begin{aligned}
 &\text{Constant } V_s^{\ell} = \alpha_s \\
 &\text{Linear } V_s^{\ell} = \alpha_s + \beta_s t_{\ell} \\
 &\text{Quadratic } V_s^{\ell} = \alpha_s + \beta_s t_{\ell} + \gamma_s (t_{\ell})^2
 \end{aligned}$$

and considering α_s, β_s and γ_s the parameters to be estimated. The estimation model is solved by using the Nelder–Mead algorithm implemented in MATLAB (fminsearch function in MATLAB) and limiting the maximum number of iterations 2000. The constrained MNL and NL models were solved using the Sequential Quadratic Programming algorithm implemented in MATLAB (fmincon function in MATLAB).

Table 3 Goodness-of-fit and CPU time for MNL, NL, CMNL and CNL models

Utility	MNL				Constrained MNL			
	L^*	ρ^2	$\bar{\rho}^2$	$CPU(s.)$	L^*	ρ^2	$\bar{\rho}^2$	$CPU(s.)$
Constant	-2024.12	0.0676	0.0662	2.76	-1147.01	0.4716	0.4703	531.67
Linear	-1438.96	0.3371	0.3344	10.17	-1074.19	0.5052	0.5024	1593.20
Quadratic	-1334.08	0.3855	0.3813	36.17	-1060.68	0.5114	0.5073	1935.74
Utility	NL				Constrained NL			
	L^*	ρ^2	$\bar{\rho}^2$	$CPU(s.)$	L^*	ρ^2	$\bar{\rho}^2$	$CPU(s.)$
Constant	-2516.61	0.3188	0.3172	61.41	-1419.18	0.6159	0.6143	13,448.42
Linear	-1705.44	0.5384	0.5352	133.75	-1362.27	0.6313	0.6280	64,126.07
Quadratic	-1694.24	0.5414	0.5366	244.18	-1518.24	0.6323	0.6274	87,688.47

Number of observations = 1976

As regards the models’ goodness-of-fit, three indexes are reported in Table 3: log-likelihood evaluated at the parameter estimate values (L^*), rho-square (ρ^2) and adjusted rho-square $\bar{\rho}^2$ where

$$\rho^2 = 1 - \frac{L^*}{L^0}; \quad \bar{\rho}^2 = 1 - \frac{L^* - k}{L^0} \tag{27}$$

where k is the number of estimated parameters and L^0 is the log-likelihood evaluated at zero. The values of L^0 for the MNL and NL models are $L^0 = -2170.85$ and $L^0 = -3694.94$ respectively.

Table 3 shows two important facts. The first is that the constrained models also improve on their unconstrained counterparts. It can even be seen that the constrained model with constant utility has a better fit than the unconstrained with quadratic utilities. The second conclusion is that the computational cost of the estimation of the constrained models is high. This reveals the need to solve the estimation problem efficiently in order to apply the methodology to real problems.

We now proceed to discuss how to test whether the improvement introduced by a constrained model with respect to an unconstrained one is or is not statistically significant. This was done using the likelihood ratio test. Let $\mu_s^{*\ell}$ be the optimal Lagrangian multipliers of the constraints of the CMNL model; thus, these constraints can be penalized in the objective function and the CMNL can be reformulated as:

$$\begin{aligned} &\text{minimize} && \sum_{\ell} \sum_{s \in \{1,2,Ref\}} [\eta_s^{\ell} g_s^{\ell} (\ln g_s^{\ell} - 1) - (V_s^{\ell} - \mu_s^{*\ell}) g_s^{\ell}] + K, \\ &\text{subject to:} && \sum_{s \in \{1,2,Ref\}} g_s^{\ell} = 1; \quad \forall \ell \end{aligned}$$

[Penalised constrained MNL]

where $K = -\sum_{\ell} \sum_{s \in \{1,2\}} \mu_s^{*\ell} K_s^{\ell}$ is a constant and we take $\mu_{Ref}^{*\ell} = 0$ to unify the notation. This formulation allows the CMNL to be interpreted as an MNL in which the utilities are given by the expression

$$\tilde{V}_s^\ell = V_s^\ell - \mu_s^{*\ell} \tag{28}$$

The statistical hypothesis testing is:

$$H_0 : \mu_s^{*\ell} = 0; (\ell, s) \in \tilde{S} \quad (\text{MNL model})$$

$$H_1 : \mu_s^{*\ell} \neq 0; (\ell, s) \in \tilde{S} \quad (\text{CMNL model})$$

In this experiment we have contrasted the hypothesis that the constraints 20 min before the train leaves are active and we should, from that moment, consider the model constrained. That is

$$\tilde{S} = \{(\ell, s) : t_s^\ell > -20\}$$

We perform a likelihood ratio test. The test statistic is twice the difference in L^* and these values are shown in Table 4. It can be seen that the inclusion of the constraints is only significant in models with constant utilities. When we estimate the linear and quadratic utilities V_s^ℓ in an unconstrained model, the real purpose of the adjustment is to determine the best parameters to reproduce the penalized utilities \tilde{V}_s^ℓ . The unconstrained models consider the constraints implicitly via the fitting of the penalized utilities. The conclusion of the hypothesis test is that in the case of linear and quadratic utilities the explicit inclusion of the constraints does not significantly improve implicit knowledge of them.

Figures 5 and 6 show the estimated probabilities that a user will choose a given alternative depending on the instant of ticket purchase. Figure 5 shows the estimates obtained using the logit models and Fig. 6 using the nested logit models. These models use quadratic utilities. The true probabilities estimated by Monte Carlo simulation have been overlaid using a 10,000 day sample. All the models try to adjust the true probabilities.

The central question is to determine in what situations it is necessary to use the constrained models. The answer is when estimations are to be carried out for scenarios different from the adjusted situation. That is, when the initial constraints are going to vary. Assume we wish to estimate the number of tickets that will be sold by the transport operator if the rolling stock carrying out the service varies. For example we will assume 3 new types of rolling stock with different numbers of seats. Table 5 shows these estimations

Table 4 Hypothesis test for the comparison of MNL, NL, CMNL and CNL models

Utility	MNL	Constrained MNL		
	L^*	L^*	χ^2	p -value
Constant	-2024.12	-1281.1	1486.0	0.0124 ^(*)
Linear	-1438.96	-1196.6	484.7	0.9999
Quadratic	-1334.08	-1176.2	315.8	0.9999
Utility	NL	Constrained NL		
	L^*	L^*		
Constant	-2516.61	-1634.51	1764.2	0.0000 ^(*)
Linear	-1705.44	-1534.94	341.0	0.9999
Quadratic	-1694.24	-1516.92	354.6	0.9999

(*): Figures in parentheses are statistical significance levels. Degrees of freedom = 1366

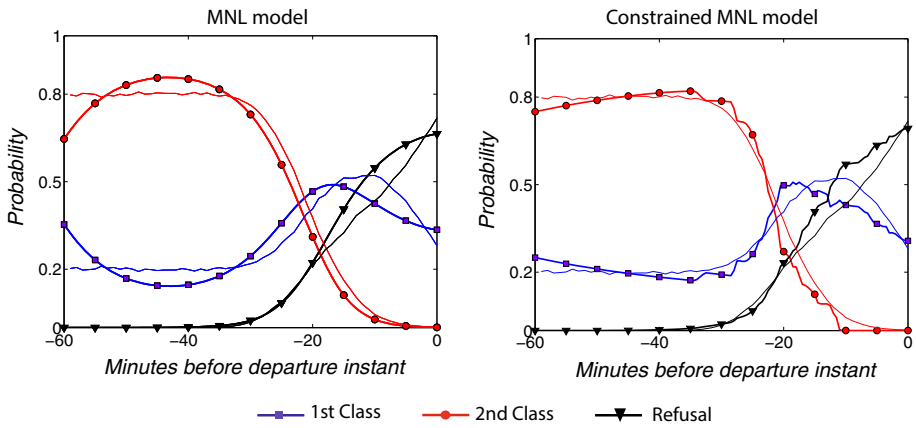


Fig. 5 Predictions of dynamic probabilities using the NML and CMNL models (quadratic utilities) and Monte Carlo simulation

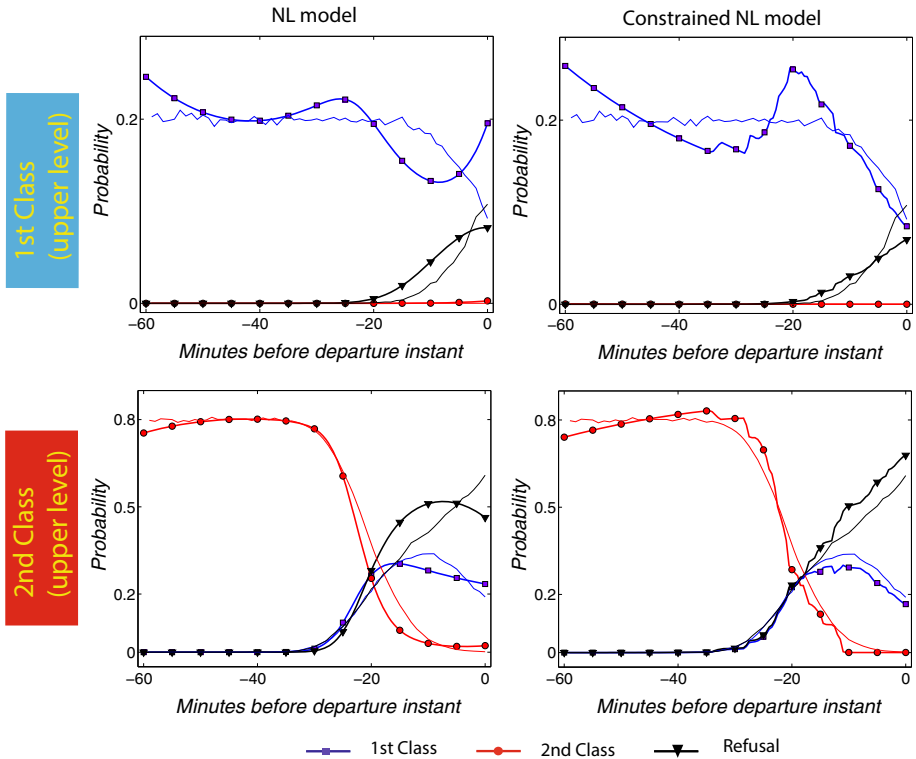


Fig. 6 Predictions of dynamic probabilities using the NL and CNL models (quadratic utilities) and Monte Carlo simulation

for each scenario using the constrained models and the Monte Carlo simulation method. The sample size was 10,000 days, and as this is a large value, we take this estimation as the true value with which to make the comparison.

Table 5 Prediction of ticket demand for several scenarios using the constrained models and the Monte Carlo simulation technique

Scenario	Simulation		Utility	Constrained MNL		Constrained NL	
	Average	σ		Average	σ	Average	σ
Original scenario			Constant	$g_1 = 22.97$ $g_2 = 40.01$	3.79 2.88	$g_1 = 21.92$ $g_2 = 41.99$	4.07 3.02
$K_1 = 25,$ $K_2 = 40$	$g_1 = 22.71$ $g_2 = 39.99$	3.24 0.12	Linear	$g_1 = 23.63$ $g_2 = 40.00$	4.31 2.90	$g_1 = 23.08$ $g_2 = 40.62$	3.27 2.98
			Quadratic	$g_1 = 23.14$ $g_2 = 40.00$	4.48 2.85	$g_1 = 23.28$ $g_2 = 40.45$	3.81 2.94
$K_1 = 25,$ $K_2 = 60$	$g_1 = 16.42$ $g_2 = 56.88$	4.30 4.45	Constant	$g_1 = 14.16$ $g_2 = 58.11$	3.66 3.91	$g_1 = 14.42$ $g_2 = 60.99$	3.44 4.11
			Linear	$g_1 = 14.66$ $g_2 = 57.99$	2.95 3.96	$g_1 = 15.33$ $g_2 = 59.35$	2.89 4.05
			Quadratic	$g_1 = 15.76$ $g_2 = 57.37$	3.21 3.96	$g_1 = 15.97$ $g_2 = 58.40$	2.92 4.06
$K_1 = 10,$ $K_2 = 70$	$g_1 = 9.86$ $g_2 = 61.83$	0.58 6.76	Constant	$g_1 = 7.01$ $g_2 = 65.75$	1.99 6.17	$g_1 = 7.33$ $g_2 = 68.62$	2.08 6.33
			Linear	$g_1 = 8.99$ $g_2 = 65.23$	1.91 6.57	$g_1 = 9.17$ $g_2 = 66.52$	2.01 6.44
			Quadratic	$g_1 = 9.29$ $g_2 = 64.49$	2.03 5.99	$g_1 = 9.43$ $g_2 = 65.81$	2.20 6.69
$K_1 = +\infty,$ $K_2 = +\infty$	$g_1 = 15.05$ $g_2 = 59.99$	3.91 7.70	Constant	$g_1 = 11.16$ $g_2 = 63.09$	1.12 6.37	$g_1 = 11.67$ $g_2 = 66.21$	1.18 6.69
			Linear	$g_1 = 12.77$ $g_2 = 62.75$	1.39 3.37	$g_1 = 13.33$ $g_2 = 64.22$	1.42 6.49
			Quadratic	$g_1 = 13.47$ $g_2 = 62.07$	1.44 6.34	$g_1 = 14.14$ $g_2 = 62.98$	1.47 6.42

Table 6 Prediction of ticket demand using the unconstrained models

Scenario	Utility	MNL		NL	
		Average	σ	Average	σ
ALL	Constant	$g_1 = 23.64$	2.39	$g_1 = 23.63$	2.39
		$g_2 = 40.00$	4.04	$g_2 = 40.00$	4.04
	Linear	$g_1 = 23.63$	2.49	$g_1 = 23.90$	2.40
		$g_2 = 40.00$	5.27	$g_2 = 39.60$	5.41
	Quadratic	$g_1 = 23.64$	2.35	$g_1 = 24.22$	2.38
		$g_2 = 40.00$	5.51	$g_2 = 39.56$	5.44

The unconstrained models predict the same value for all scenarios and this is shown in Table 6. If we observe the scenario which produces the worst estimation using the models with quadratic utility, we see that the relative errors (expressed as percentages) are, for the

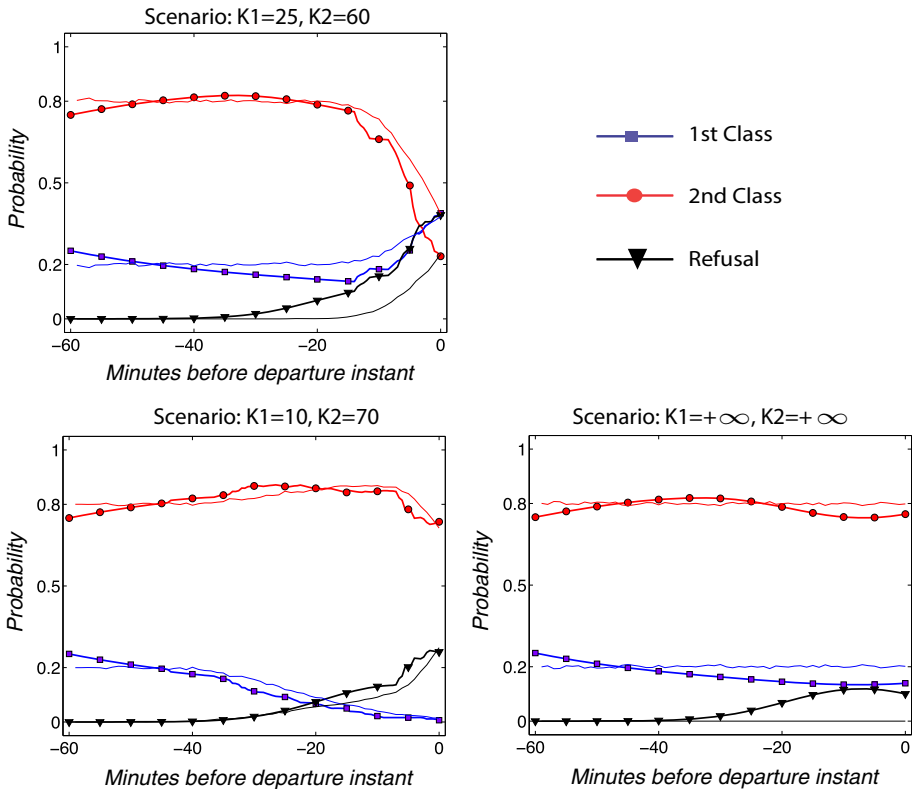


Fig. 7 Predictions of the dynamic probabilities using the CMNL model (quadratic utilities) and Monte Carlo simulation for the three new scenarios

constrained models, 10.49 and 6.40% for the first and second class respectively, while for the unconstrained models these errors are 145.63 and 36.01% respectively.

This shows the need to use the constrained models for this type of estimation. Finally Fig. 7 shows the estimation of the ticket choice probabilities for the three new scenarios using a constrained MNL model. A good fit is seen for these new scenarios, even though they are different from the scenario used in the estimation model.

Application of the CNL model for railway service choice modelling (Experiment 2)

Suppose there are various types of users depending on his/her origin-destination. Index $\ell = (i, j) \in \mathcal{L}$ refers to a trip from station i to station j . Assume an origin-destination matrix $\{\tilde{g}^\ell\}_{\ell \in \mathcal{L}}$ which defines the potential demand. Assume the total demand disaggregated in two alternatives:

- (a) (High-speed) train trips.
- (b) Another means of transport.

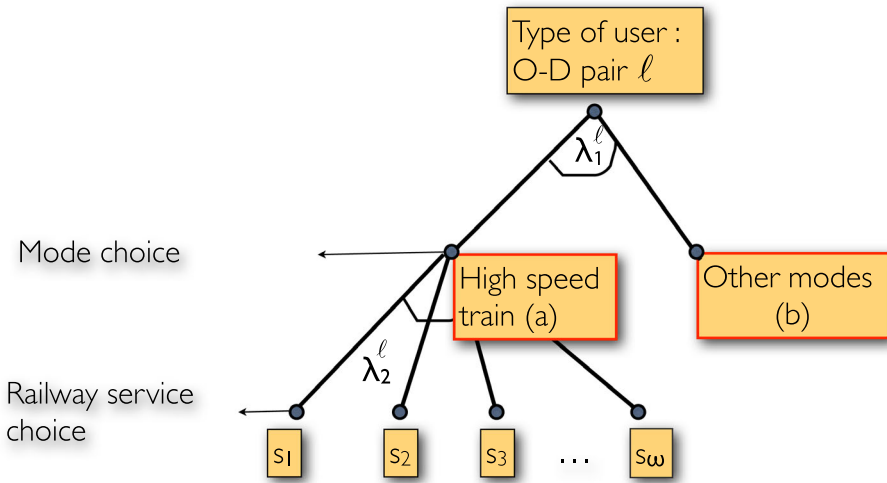


Fig. 8 Hierarchical MNL demand for railway service choice

Assume a logit model which divides the potential demand between alternatives (a) and (b):

$$g^{m\ell} = \frac{\exp(\lambda_1 V^{m\ell})}{\sum_{m \in \{a,b\}} \exp(\lambda_1 V^{m\ell})} \cdot \hat{g}^\ell \quad m \in \{a,b\}, \quad \ell \in \mathcal{L} \tag{29}$$

where $V^{m\ell}$ is the utility of alternative m for the users of the origin-destination pair ℓ .

Note that the index ℓ is deleted in the parameter λ_1 , which means that this value is the same for all origin-destination pairs ℓ .

The model considers a nested logit model to disaggregate the demand considering the feasible timetable for a trip type ℓ . Denote as \mathcal{S}_a^ℓ the feasible set of railway services for making a trip type ℓ . The second level of the nested logit model disaggregates the demand between the different railway services:

$$g_s^{a\ell} = \frac{\exp(\lambda_2 V_s^{a\ell})}{\sum_{s' \in \mathcal{S}_a^\ell} \exp(\lambda_2 V_{s'}^{a\ell})} \cdot g^{a\ell} \quad s \in \mathcal{S}_a^\ell, \quad \ell \in \mathcal{L} \tag{30}$$

Similarly to the upper decision level, the parameter λ_2 is assumed independent of the origin-destination pair ℓ .

Figure 8 shows the nested logit model combined with the capacity constraints of the trains. When a train reaches station j the vehicle has picked up passengers from preceding stations. The number of passengers that can take the train is then restricted by the capacity of the vehicle. Denote by \mathcal{L}_{sj}^+ the set of origin-destination pairs whose users take the service s before station j and leave the vehicle after station j . Also denote by \mathcal{L}_{sj} the set of origin-destination pairs ℓ whose origin is station j and which use s . The capacity constraints of service s in station j is formulated as:

$$\sum_{\ell \in \mathcal{L}_{sj}} g_s^{a\ell} + \sum_{\ell \in \mathcal{L}_{sj}^+} g_s^{a\ell} \leq K_s \quad \text{for all } s \in \mathcal{S}, \quad j \in J_s; \tag{31}$$

where K_s is the capacity of train s , \mathcal{S} is the set of services and J_s represents the set of stations in which service s will stop. The demand model can be stated as:

$$\begin{aligned}
 &\text{minimize} && \sum_{\ell \in \mathcal{L}} \left[\sum_{m \in \{a,b\}} \eta^m g^{m\ell} (\ln g^{m\ell} - 1) + \eta^c \sum_{s \in \mathcal{S}_a^\ell} g_s^{a\ell} (\ln g_s^{a\ell} - 1) \right. \\
 & && \left. - V^{b\ell} g^{b\ell} - \sum_{s \in \mathcal{S}_a^\ell} V_s^{a\ell} g_s^{a\ell} \right], \\
 &\text{subject to:} && g^{a\ell} + g^{b\ell} = \widehat{g}^\ell, \ell \in \mathcal{L} \\
 & && g^{a\ell} = \sum_{s \in \mathcal{S}_a^\ell} g_s^{a\ell}, \ell \in \mathcal{L} \\
 & && \sum_{\ell \in \mathcal{L}_j} g_s^{a\ell} + \sum_{\ell \in \mathcal{L}_j^+} g_s^{a\ell} \leq K_s \text{ for all } s \in \mathcal{S}, \quad j \in J_s;
 \end{aligned}$$

[CNL]

where $\eta^a = \frac{1}{\lambda_1} - \frac{1}{\lambda_2}$ and $\eta^b = \frac{1}{\lambda_1}, \eta^c = \frac{1}{\lambda_2}$.

Espinosa-Aranda et al. (2015) apply this model to the high-speed train timetabling problem.

Case study

To test the above model, a case study has been generated. The main objective is to study the possibility of estimating its parameters. This numerical example looks at the Madrid–Seville corridor of the Spanish High Speed Railway network. This corridor consists of 5 stations: Madrid (MAD), Ciudad Real (CR), Puertollano (PU), Córdoba (COR) and Sevilla (SEV) which produces 20 origin-destination demand pairs (10 per direction of travel) formed by 15,115 passengers/day. Currently this demand is completely covered by 100 services. Figure 9 shows the corridor used by these services, Table 7 indicates the route of each type of service and Table 8 shows the maximum capacity of each type of train.

In this example, there will be 20 types of users (i.e. 20 origin-destination pairs ℓ). Each user type ℓ can travel using a set of services s . Considering the planned schedule, the set of alternatives $(\ell, s) \in \mathcal{D}$ consists of 298 possibilities. In this case the proposed model could estimate 298 parameters.

To estimate the model 25 services have been selected randomly, generating a set \mathcal{D}_1 with 66 possibilities (ℓ, s) and, consequently, 66 parameters $V_s^{a,\ell}$ that should be estimated to calculate the utility of each alternative as explained in “Non-linear utility specifications using Reproducing Kernel Hilbert Spaces” section. Note that the solution of the linear system (15) produces the values α_s^ℓ .

The attributes considered for each possible alternative (ℓ, s) are: (1) the price $x_{s,1}^\ell$, (2) the travel time $x_{s,2}^\ell$ and (3) the timetable $x_{s,3}^\ell$. The vector of attributes is denoted as $x_s^\ell = (x_{s,1}^\ell, x_{s,2}^\ell, x_{s,3}^\ell)$. The data used for the experiment can be downloaded from <http://bit.ly/1gCFw5e>.

To simplify we have set $V^{*\ell}(x) = V^*(x)$ for all origin-destination pairs ℓ in this experiment, and the utility function $V^*(x)$ is defined as:

$$V^*(x) = \sum_{(\ell,s) \in \mathcal{D}_1} \alpha_s^\ell K(x, x_s^\ell), \text{ for all } x \in X, \tag{32}$$

where a Gaussian kernel $K(x, y) = e^{-a\|x-y\|^2}$ is used in which $\|\cdot\|$ is the Euclidean norm $a \in \mathbb{R}^+$. In this case $a = 5$ has been considered. The regularization parameter γ has been



Fig. 9 Madrid–Seville corridor

Table 7 Types of railway services on Madrid–Seville corridor

Type of service	Route	Amount of services	Type of train
1	MAD → CR → PU	11	AVANT
2	MAD → CR → PU → COR	3	AVE
3	MAD → CR → PU → COR → SEV	5	AVE
4	MAD → COR	4	AVE
5	MAD → COR → SEV	9	AVE
6	MAD → SEV	3	AVE
7	COR → SEV	6	MD
8	COR → SEV	9	AVANT
9	SEV → COR	6	MD
10	SEV → COR	9	AVANT
11	SEV → COR → PU → CR → MAD	5	AVE
12	SEV → COR → MAD	8	AVE
13	SEV → MAD	3	AVE
14	COR → PU → CR → MAD	3	AVE
15	COR → MAD	5	AVE
16	PU → MAD → CR	11	AVANT

set as 0.00001. The parameter γ has the function of preventing the system of linear equations (15) being singular. The parameter is chosen thus $\gamma \rightarrow 0^+$. First a small value is tested, as in this example, and if there are no numerical problems (an ill-conditioned

Table 8 Capacity of each type of train on Madrid–Seville corridor

Type	Train capacity K_s (passengers)
AVE	308
AVANT	237
MD	190

problem) the value is acceptable. Otherwise the value of γ is increased until it avoids ill-conditioning.

Estimation methods

The data used as \mathbf{N} are in the public domain and therefore show aggregated information such as the total demand in a determined origin-destination pair or for a type of service. An estimation procedure based on an ML approach is not available because the disaggregated observations for each pair (ℓ, s) are unknown. In this test we have adapted the generalized least squares technique for comparing the known demand behaviour versus the demand predicted by the CHL model. In this example the values of λ_1 and λ_2 have also been estimated.

The optimization method selected for solving the estimation problem (25) is a hybridization of the *Standard Particle Swarm Optimization (SPSO)* (Zambrano-Bigiarini et al. 2013) and the *Nelder–Mead (NM)* (Nelder and Mead 1965) algorithm based on the framework presented in Espinosa-Aranda et al. (2013). Hybrid algorithms try to make full use of the merits of various optimization techniques in order to obtain an efficient method in the search for global optima.

The SPSO has been used successfully in global optimization problems particularly in transportation research (see Angulo et al. 2011, 2013). The main advantages of PSO algorithms could be summarized as follows: they are capable of avoiding local optima,

Table 9 Solution found versus amplitude of feasible region

Case	−B	B	SPSO	SPSO + NM
1	−0.5	0.5	4.7276E+05	2.5866E+05
2	−1	1	3.2272E+05	2.5248E+05
3	−2	2	3.2828E+05	2.5988E+05
4	−5	5	4.8158E+05	2.7113E+05
5	−10	10	5.6018E+05	2.6979E+05
6	−30	30	1.1273E+06	1.1273E+06
7	−50	50	1.2238E+06	1.2238E+06
8	−100	100	1.4023E+06	1.4023E+06
9	−500	500	1.1742E+06	1.1742E+06
10	−1000	1000	1.0889E+06	1.0628E+06
11	−5E+03	5E+03	1.2428E+06	1.2428E+06
12	−1E+04	−1E+04	1.0640E+06	1.0640E+06
13	−5E+04	5E+04	1.3045E+06	1.3045E+06
14	−5E+05	5E+05	1.2318E+06	1.2318E+06
15	−1E+07	1E+07	2.2174E+06	2.2113E+06

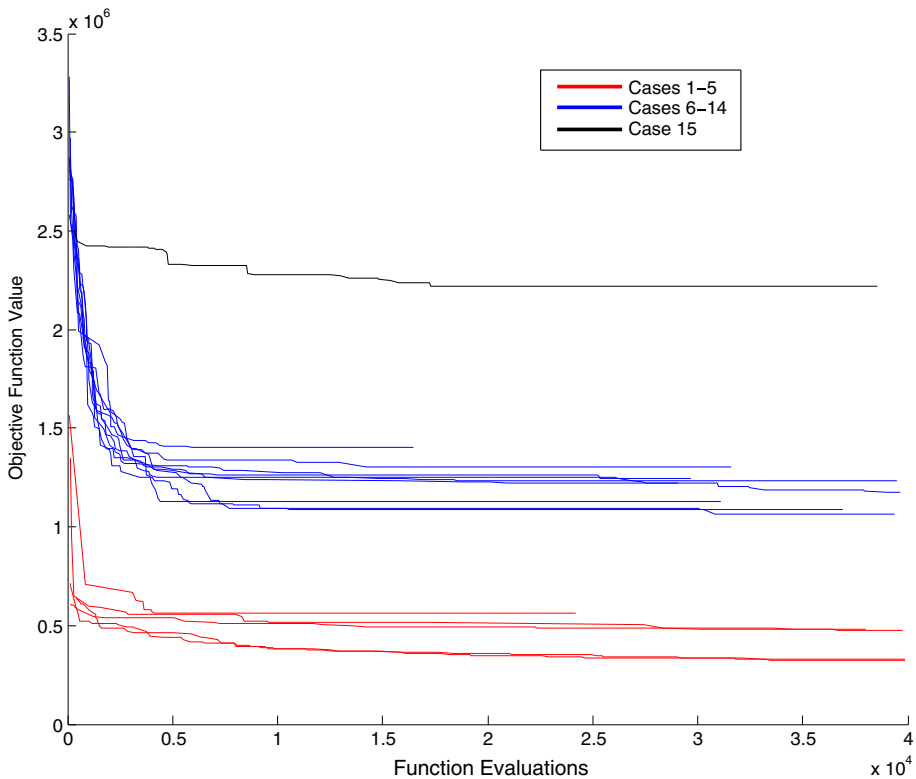


Fig. 10 Evolution of the SPSO + NM algorithm for all cases

doing a search in the entire solution space, are robust against initialization parameters, viable, efficient with a smaller computational burden and have a simple selection of the right parameter values. The NM method is a direct search method that does not use numerical or analytic gradients and has local convergence with a high exploitation capacity.

The resolution procedure for the CNL model has been GAMS 24 with solver CONOPT, showing that in an Intel I7 4 Cores 3.2 GHZ with 16 GB RAM computer the CPU time for each problem is around 0.12 seconds.

The SPSO + NM has been implemented in MATLAB, which calls GAMS to solve each individual CNL problem. The stopping criterion is based on the total number of solved CNL models. The SPSO algorithm was run for 50,000 objective function evaluations. A random start on an interval defined by Eq. (33) was used. The size of the swarm was 40 particles. The PSO-parameters w , c_1 and c_2 for updating velocity are defined as $w = 1/(2 \ln(2))$, $c_1 = 0.5 + \ln(2)$ and $c_2 = c_1$ (Zambrano-Bigiarini et al. 2013). NM is run for 50,000 function evaluations starting from the best solution found by the SPSO algorithm to improve the solution.

“Appendix 2” deals with over-specification based on two considerations. Firstly, the non-observed utilities are set to $V^{bl} = 0$ (the second source of over-specification). The second issue considers that each interval

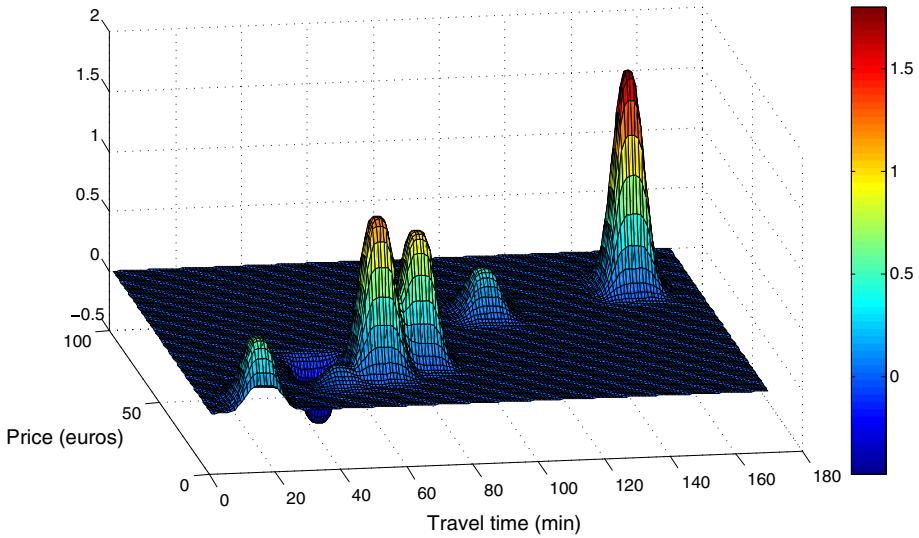


Fig. 11 Utility estimation with fixed departure time (8:30)

$$-B \leq V_s^{a\ell} \leq B \tag{33}$$

with $B > 0$ contains optimal solutions of the estimation problem (the first source of over-specification), thus the imposition of this constraint limits the search space without reducing the quality of the fit. Selecting limits with very small B value (equivalent to $\delta^\ell \rightarrow 0$ in the first source of over-specification) could lead to large values of λ_i , causing a more complex estimation. A trade-off between the range of the interval of the utilities and the order of magnitude of the parameters λ_i must be achieved.

Table 9 shows the computational results depending on the feasible region considered (33). The mean computational cost of each run of the SPSO algorithm was 1.7 h, and with NM, 4.4 h. Therefore the estimation of the CNL model can be computed in an affordable time.

As can be seen, the results show that with small intervals the algorithms are capable of finding a better solution than when searching in a bigger feasible region. This can also be seen in Fig. 10 which depicts the evolution of the objective function during the running of the SPSO algorithm per case study. The red graphs represent case studies 1–5, the blue 6–14 and the black line 15.

Study of the best solution obtained

The best solution is obtained by using the interval $(-B, B) = (-1, 1)$ as the space of parameters. This section shows this solution. Figures 11 and 12 depict the utility function fixing respectively the departure time at 8:30 and the travel time to 63 min.

It can be seen in Fig. 11 that the specification of $V^\ell(x) = V^*(x)$ for each pair ℓ produces the travel time attribute with which to consider the demand effect in each pair ℓ . For example, the largest travel time represents the largest trip $\ell = (MAD, SEV)$ while the smallest time occur in pair $\ell = (CR, PU)$. The utility function captures the demand in each pair.

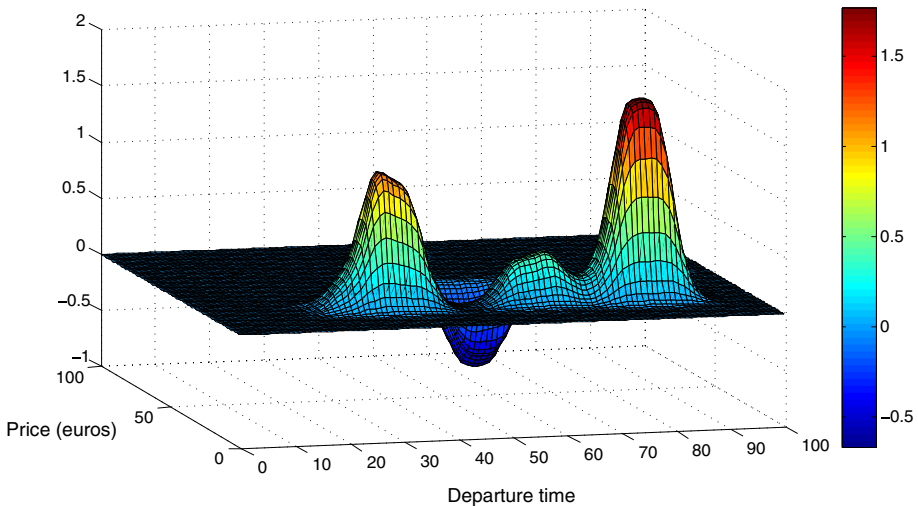


Fig. 12 Utility estimation with fixed travel time (63 min)

Figure 12 depicts the results obtained for a fixed travel time of 63 min (i.e. fixed a origin-destination pair ℓ). This case shows how the utility change depending on the departure time coinciding with the demand peaks at specific times (8:30, 15:00 and 19:30).

Note that Figs. 11 and 12 show negative utility values. To avoid the problem of over-specification we have set the utility of not making the trip by train at $V^{bl} = 0$. The estimates of the other utilities V_s^{al} are calculated with respect to this choice. The fact that these estimations give negative values means only that these utilities are less than the utility of alternative b .

Conclusions

This paper describes the CNL to model both the dynamic and constrained decision spaces in discrete choice contexts. This type of approach is suited to modelling problems in which exogenous and endogenous factors limit the universal choice set of the decision-makers. Applying the model requires additional data to specify the constraints and derivative-free optimization methods for solving the estimation problem.

A key contribution of this paper is the use of Kernel Hilbert Spaces for the specifications of non-linear utility functions. This paper presents a novel point of view for the model estimation based on the consideration of utilities of a set of alternatives as parameters instead of classical attribute weights. The over-specification issues associated with the CNL formulation are also discussed. The introduction of the estimation method requires testing its capacity to infer unbiased estimates of the true parameters. Further research into this subject, and how to specify the type of kernels and its parametrization is necessary to assess whether this methodology can give a perfect reconstruction of the original utility.

Experiment 1 is a simulation study on a constrained problem. The results obtained show that the constrained models have a better fit than their unconstrained counterparts. The unconstrained models take the constraints into account via non-linear utilities and thus for the baseline case there are differences between constrained and unconstrained models only when the utility function is linear. The key conclusion of Experiment 1 is that the

constrained models are robust with respect to the modifications in the baseline constraints. If the unconstrained models are used in new scenarios, therefore, forecasts of demand are likely to violate the new constraints and miscalculate the true demand.

In Experiment 2 a novel railway demand model is used to test the suitability of the proposed approach. Experiment 2 is based on real data for the Madrid–Seville high-speed corridor and proposes a metaheuristic methodology based on the hybridization of the *Particle Swarm Optimization* and Nelder–Mead method to estimate the CNL. The computational cost of solving the bi-level model is 4.4 h. The importance of eliminating over-specification of the model can be seen in the quality of the results obtained. The results show how the CNL model could represent the behaviour of users of the railway network.

Acknowledgements The authors wish to thank the referees for making a range of interesting remarks on an earlier version of this paper. This research was supported by the *Ministerio de Economía y Competitividad* of Spain-FEDER EU Grant with number TRA2014-52530-C3-2-P.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix 1: Equilibrium issues of the CNL model

We shall assume that the functions $\{\widehat{h}_r(\mathbf{g})\}_{r \in \mathcal{R}}$ are convex. The Karush–Kuhn–Tucker conditions are necessary and sufficient for the optimal solution of CNL as it is a (strictly) convex program. Moreover, in this case the model has a unique solution (assuming the feasible region of CNL is not empty) and it is this which will now be characterized.

The Lagrangian function of the problem CNL is of the form

$$L = Z + \sum_{\ell \in \mathcal{L}} \Phi_\ell \left(\sum_{m \in \mathcal{S}^\ell} g^{m\ell} - \widehat{g}^\ell \right) + \sum_{\ell \in \mathcal{L}} \sum_{m \in \mathcal{S}^\ell} \Theta_{m\ell} \left(\sum_{s \in \mathcal{S}_m^\ell} g_s^{m\ell} - g^{m\ell} \right) + \sum_{r \in \mathcal{R}} \mu_r \left(\widehat{h}_r(\mathbf{g}) - b_r \right)$$

where Z represents the objective function of CNL.

The Karush–Kuhn–Tucker conditions for CNL are stated as

$$\frac{\partial L}{\partial g_s^{m\ell}} = \frac{1}{\lambda_2^{m\ell}} \ln g_s^{m\ell} - V_s^{m\ell} + \Theta_{m\ell} + W_s^{m\ell} = 0; \quad \ell \in \mathcal{L}, m \in \mathcal{S}^\ell, s \in \mathcal{S}_m^\ell \tag{34}$$

$$\frac{\partial L}{\partial g^{m\ell}} = \eta_1^{m\ell} \ln g^{m\ell} + \Phi_\ell - \Theta_{m\ell} = 0; \quad \ell \in \mathcal{L}, m \in \mathcal{S}^\ell \tag{35}$$

$$\mu_r \geq 0 \text{ and } \mu_r \left(\widehat{h}_r(\mathbf{g}) - b_r \right) = 0; \quad r \in \mathcal{R} \tag{36}$$

where

$$W_s^{m\ell} = \sum_{r \in \mathcal{R}} \mu_r \frac{\partial \widehat{h}_r}{\partial g_s^{m\ell}}. \tag{37}$$

Next, this paper will focus on the value of the probabilities at the lower level of the CNL. Therefore, solving for $g_s^{m'\ell'}$ in (34),

$$\begin{aligned}
 g_s^{m'\ell'} &= \exp\left\{\lambda_2^{m'\ell'} (V_s^{m'\ell'} - W_s^{m'\ell'} - \Theta_{m'\ell'})\right\} \\
 &= \exp\left(-\lambda_2^{m'\ell'} \Theta_{m'\ell'}\right) \exp\left\{\lambda_2^{m'\ell'} (V_s^{m'\ell'} - W_s^{m'\ell'})\right\}
 \end{aligned}
 \tag{38}$$

and summing over $s \in \mathcal{S}_{m'}^{\ell'}$, which is the set of sub-alternatives for the alternative m' and type of user ℓ' , we get

$$g^{m'\ell'} = \sum_{s \in \mathcal{S}_{m'}^{\ell'}} g_s^{m'\ell'} = \exp\left(-\lambda_2^{m'\ell'} \Theta_{m'\ell'}\right) \sum_{s \in \mathcal{S}_{m'}^{\ell'}} \exp\left\{\lambda_2^{m'\ell'} (V_s^{m'\ell'} - W_s^{m'\ell'})\right\}
 \tag{39}$$

Finally, dividing (38) by (39) we get the probabilities at the lower level of the CNL

$$\frac{g_{s'}^{m'\ell'}}{g^{m'\ell'}} = \frac{\exp\left\{\lambda_2^{m'\ell'} (V_{s'}^{m'\ell'} - W_{s'}^{m'\ell'})\right\}}{\sum_{s \in \mathcal{S}_{m'}^{\ell'}} \exp\left\{\lambda_2^{m'\ell'} (V_s^{m'\ell'} - W_s^{m'\ell'})\right\}}; \quad s' \in \mathcal{S}_{m'}^{\ell'}
 \tag{40}$$

At this point, the probabilities of the CNL in the upper level will be also calculated. Finding $\Theta_{m'\ell'}$ from (39)

$$\Theta_{m'\ell'} = \frac{-1}{\lambda_2^{m'\ell'}} \ln g^{m'\ell'} + \frac{1}{\lambda_2^{m'\ell'}} \ln \left[\sum_{s \in \mathcal{S}_{m'}^{\ell'}} \exp\left\{\lambda_2^{m'\ell'} (V_s^{m'\ell'} - W_s^{m'\ell'})\right\} \right]
 \tag{41}$$

and by replacing it in (35), one obtains

$$\eta_1^{m'\ell'} \ln g^{m'\ell'} + \frac{1}{\lambda_2^{m'\ell'}} \ln g^{m'\ell'} + \Phi_{\ell'} - \frac{1}{\lambda_2^{m'\ell'}} \ln \left[\sum_{s \in \mathcal{S}_{m'}^{\ell'}} \exp\left\{\lambda_2^{m'\ell'} (V_s^{m'\ell'} - W_s^{m'\ell'})\right\} \right] = 0.$$

Now using the definition of $\eta_1^{m'\ell'}$ given in CNL

$$\frac{1}{\lambda_1^{\ell'}} \ln g^{m'\ell'} + \Phi_{\ell'} - L^{m'\ell'} = 0
 \tag{42}$$

where $L^{m'\ell'}$ is the classical *log-sum* given by

$$L^{m'\ell'} = \frac{1}{\lambda_2^{m'\ell'}} \ln \left[\sum_{s \in \mathcal{S}_{m'}^{\ell'}} \exp\left\{\lambda_2^{m'\ell'} (V_s^{m'\ell'} - W_s^{m'\ell'})\right\} \right].
 \tag{43}$$

Finding $g^{m'\ell'}$ from (42)

$$g^{m'\ell'} = \exp\left\{\lambda_1^{\ell'} (L^{m'\ell'} - \Phi_{\ell'})\right\}
 \tag{44}$$

and adding with respect to m

$$\widehat{g}^{\ell'} = \sum_{m \in \mathcal{S}^{\ell'}} \exp\left\{\lambda_1^{\ell'} (L^{m\ell'} - \Phi_{\ell'})\right\}
 \tag{45}$$

Finally, the probability of selecting alternative m' for the type of user ℓ' in the upper level is:

$$\frac{g^{m'\ell'}}{\widehat{g}^{\ell'}} = \frac{\exp\{\lambda_1^{\ell'}(L^{m'\ell'} - \Phi_{\ell'})\}}{\sum_{m \in S^{\ell'}} \exp\{\lambda_1^{\ell'}(L^{m\ell'} - \Phi_{\ell'})\}} = \frac{\exp(\lambda_1^{\ell'} L^{m'\ell'})}{\sum_{m \in S^{\ell'}} \exp(\lambda_1^{\ell'} L^{m\ell'})} \tag{46}$$

Appendix 2: Over-specification of the CNL model

In this Appendix we show that there exist infinite solutions for the estimation problem (25). This is due to over-specification of the parameters (García-Ródenas and Marín 2009; Bierlaire et al. 1997; Daganzo and Kusnic 1993; Ben-Akiva and Lerman 1985). Parameter over-specification must be avoided because although some of the more robust methods succeed in solving the problem, their speed of convergence may be very slow. This problem is due to the singularity of the second derivative matrix of the log-likelihood function.

First source of over-specification

The first source of over-specification arises in the interaction between the structure of the utilities and the parameters λ_j^ℓ with $j \in \{1, 2\}$ and it becomes:

$$\left. \begin{aligned} \widetilde{V}_s^{m\ell} &= \delta^\ell \widehat{V}_s^{m\ell} \\ \widetilde{\lambda}_j^\ell &= \frac{\lambda_j^\ell}{\delta^\ell}; \quad j \in \{1, 2\} \end{aligned} \right\} \ell \in \mathcal{L}$$

The above relationships are schematically denoted as $\widetilde{V} = \delta \widehat{V}$ and $\widetilde{\lambda} = \lambda/\delta$.

Let (\widehat{V}_1, λ) be a vector of parameters for the CNL model and let

$$g = \text{CNL}((\widehat{V}_1, \mathcal{H}(\widehat{V}_1)), \lambda) \tag{47}$$

be the estimated demand.

The objective function of the CNL model is separable in ℓ . If each term in ℓ is multiplied by the constant $\delta^\ell > 0$ then the optimal solution associated with ℓ is not changed. Moreover, the system constraints hold. This leads to:

$$g = \text{CNL}((\delta V_1, \delta \mathcal{H}(\widehat{V}_1)), \lambda/\delta) = \text{CNL}((\widetilde{V}_1, \delta \mathcal{H}(\widehat{V}_1)), \widetilde{\lambda}) \tag{48}$$

It is worth noting that the utilities \widehat{V}_1 are multiplied by δ , the solution of system (15) is multiplied by δ and thus the utilities \widehat{V}_2 are also multiplied by δ because they are linear in their parameters α . Mathematically

$$\mathcal{H}(\delta \widehat{V}_1) = \delta \mathcal{H}(\widehat{V}_1) \tag{49}$$

Using (47), (48) and (49), we obtain

$$g = \text{CNL}((\widehat{V}_1, \mathcal{H}(\widehat{V}_1)), \lambda) = \text{CNL}((\widetilde{V}_1, \mathcal{H}(\widetilde{V}_1)), \widetilde{\lambda}) \tag{50}$$

As the objective function of the estimation model (25), $F(g, \mathbf{N})$, depends only on g , both solutions (\widehat{V}_1, λ) and $(\widetilde{V}_1, \widetilde{\lambda})$ have the same objective value. This shows that there exist infinite optimal solutions of the estimation model.

Thus the scale parameters of the Gumbel error terms are undetermined. In practice, setting one Gumbel term for each ℓ is sufficient for the identification.

Second source of over-specification

The second source of over-specification in the CNL models is adding the same value to the utilities of all the alternatives, which does not affect the log-likelihood of the sample. In this case, we assume that $\mathcal{D}_2 = \{\emptyset\}$. The set of utilities

$$\widetilde{V}_s^{m\ell} = \widehat{V}_s^{m\ell} + \gamma^\ell; \ell \in \mathcal{L} \tag{51}$$

produces the same solution as the optimization model. If utilities \widehat{V} of the objective function CNML are replaced by utilities \widetilde{V} the same objective function value plus the constant is obtained

$$-\sum_{s \in \mathcal{S}_m} \widetilde{V}_s^{m\ell} g_s^{m\ell} = -\sum_{s \in \mathcal{S}_m} (\widehat{V}_s^{m\ell} + \gamma^\ell) g_s^{m\ell} = -\gamma^\ell \widehat{g}^\ell - \sum_{s \in \mathcal{S}_m} \widehat{V}_s^{m\ell} g_s^{m\ell} \tag{52}$$

Bierlaire et al. (1997) have analysed over-specification in nested logit models to the log-likelihood function. These authors have analysed the relationship between any two arbitrary strategies to avoid over-specification, and shown that the two strategies are equivalent under a linear transformation of the variables. Some algorithms are independent of such transformations: Newton’s method and the quasi-Newton methods of the Broyden family are combined with line searches. If these are used, then the way in which the over-specification is eliminated is not important. Daganzo and Kusnic (1993) suggested equating one parameter to zero for each set of parameters mixed up in every source of over-specification, and estimating the rest in order to avoid over-specification in the nested logit model.

References

Amador, F.J., González, R.M., De Dios, Ortúzar J.: Preference heterogeneity and willingness to pay for travel time savings. *Transportation* **32**(6), 627–647 (2005)

Anas, A.: Discrete choice theory, information theory and the multinomial logit and gravity models. *Transp. Res. Part B* **17**(1), 13–23 (1983)

Angulo, E., Castillo, E., García-Ródenas, R., Sánchez-Vizcaíno, J.: Determining highway corridors. *J. Transp. Eng.* **138**(5), 557–570 (2011)

Angulo, E., Castillo, E., García-Ródenas, R., Sánchez-Vizcaíno, J.: A continuous bi-level model for the expansion of highway networks. *Comput. Oper. Res.* **41**(2014), 262–276 (2013)

Aroszajn, N.: Theory of reproducing kernels. *Trans. Am. Math. Soc.* **68**(3), 337–404 (1950)

Ben-Akiva, M., Boccara, B.: Discrete choice models with latent choice sets. *Int. J. Res. Mark.* **12**(1), 9–24 (1995)

Ben-Akiva, M.E., Lerman, S.R.: *Discrete Choice Analysis: Theory and Application to Travel Demand*, vol. 9. MIT press, Cambridge (1985)

Bierlaire, M., Lotan, T., Toint, P.: On the overspecification of multinomial and nested logit models due to alternative specific constants. *Transp. Sci.* **31**(4), 363–371 (1997)

Bierlaire, M., Hurtubia, R., Flötteröd, G.: Analysis of implicit choice set generation using a constrained multinomial logit model. *Transp. Res. Rec.* **2175**, 92–97 (2010)

- Boer, A.: Dynamic pricing and learning: historical origins, current research, and new directions. *Surv. Oper. Res. Manag. Sci.* **20**(1), 1–18 (2015)
- Cantillo, V., Ortúzar, J.: A semi-compensatory discrete choice model with explicit attribute thresholds of perception. *Transp. Res. Part B Methodol.* **39**(7), 641–657 (2005)
- Cantillo, V., Heydecker, B., de Dios, Ortúzar J.: A discrete choice model incorporating thresholds for perception in attribute values. *Transp. Res. Part B Methodol.* **40**(9), 807–825 (2006)
- Cascetta, E., Papola, A.: Random utility models with implicit availability/perception of choice alternatives for the simulation of travel demand. *Transp. Res. Part C Emerg. Technol.* **9**(4), 249–263 (2001)
- Castro, M., Martínez, F., Munizaga, M.: Estimation of a constrained multinomial logit model. *Transportation* **40**(3), 563–581 (2013)
- Chen, M., Chen, Z.L.: Recent developments in dynamic pricing research: multiple products, competition, and limited demand information. *Prod. Oper. Manag.* **24**(5), 704–731 (2015)
- Cherchi, E., Ortúzar, J.D.D.: Mixed RP/SP models incorporating interaction effects: modelling new suburban train services in Cagliari. *Transportation* **29**(4), 371–395 (2002)
- Cox, D., O'Sullivan, F.: Asymptotic analysis of penalized likelihood and related estimators. *Ann. Stat.* **18**(4), 1676–1695 (1990)
- Daganzo, C., Kusnic, M.: Two properties of the nested logit model. *Transp. Sci.* **27**(4), 395–400 (1993)
- Daumé, H.: From zero to reproducing kernel hilbert spaces in twelve pages or less. <http://www.umiacs.umd.edu/~hal/docs/daume04rkhs.pdf> (2004)
- De Grange, L., González, F., Vargas, I., Muñoz, J.: A polarized logit model. *Transp. Res. Part A Policy Pract.* **53**, 1–9 (2013)
- De Grange, L., González, F., Vargas, I., Troncoso, R.: A logit model with endogenous explanatory variables and network externalities. *Netw. Spat. Econ.* **15**(1), 89–116 (2015)
- Ding, Y., Veeman, M., Adamowicz, W.: The influence of attribute cutoffs on consumers' choices of a functional food. *Eur. Rev. Agric. Econ.* **39**(5), 745–769 (2012)
- Donoso, P., de Grange, L.: A microeconomic interpretation of the maximum entropy estimator of multinomial logit models and its equivalence to the maximum likelihood estimator. *Entropy* **12**(10), 2077–2084 (2010)
- Elrod, T., Johnson, R.D., White, J.: A new integrated model of noncompensatory and compensatory decision strategies. *Organ. Behav. Hum. Decis. Process.* **95**(1), 1–19 (2004)
- Espinosa-Aranda, J., García-Ródenas, R., Ramírez-Flores, M., López-García, M., Angulo, E.: High-speed railway scheduling based on user preferences. *Eur. J. Oper. Res.* **246**(3), 772–786 (2015)
- Espinosa-Aranda, J.L., García-Ródenas, R., Angulo, E.: A framework for derivative free algorithm hybridization. *Adapt. Nat. Comput. Algorithms* **7824**, 80–89 (2013)
- Fernández, E., De Cea, J., Florian, M., Cabrera, E.: Network equilibrium models with combined modes. *Transp. Sci.* **28**(3), 182–192 (1994)
- Fúnez-Guerra, C., García-Ródenas, R., Sánchez-Herrera, E.A., Verastegui-Rayó, D., Clemente-Jul, C.: Modeling of the behavior of alternative fuel vehicle buyers: a model for the location of alternative refueling stations. *Int. J. Hydrogen Energy* **41**(42), 19,312–19,319 (2016)
- García, R., Marín, A.: Network equilibrium with combined modes: models and solution algorithms. *Transp. Res. Part B Methodol.* **39**(3), 223–254 (2005)
- García-Ródenas, R., Marín, A.: Simultaneous estimation of the origin-destination matrices and the parameters of a nested logit model in a combined network equilibrium model. *Eur. J. Oper. Res.* **197**(1), 320–331 (2009)
- Jara-Díaz, S.R.: Allocation and valuation of travel time savings. *Handb. Transp.* **1**, 303–319 (2000)
- Kaplan, S., Bekhor, S., Shifan, Y.: Two-stage model for jointly revealing determinants of noncompensatory conjunctive choice set formation and compensatory choice. *Transp. Res. Rec.* **2134**, 153–163 (2009)
- Kaplan, S., Shifan, Y., Bekhor, S.: Development and estimation of a semi-compensatory model with a flexible error structure. *Transp. Res. Part B Methodol.* **46**(2), 291–304 (2012)
- Kimeldorf, G., Wahba, G.: A correspondence between bayesian estimation on stochastic processes and smoothing splines. *Ann. Math. Stat.* **41**(2), 495–502 (1970)
- Li, L., Adamowicz, W., Swait, J.: The effect of choice set misspecification on welfare measures in random utility models. *Resour. Energy Econ.* **42**, 71–92 (2015)
- Louvière, J., Train, K., Ben-Akiva, M., Bhat, C., Brownstone, D., Cameron, T., Carson, R., Deshazo, J., Fiebig, D., Greene, W., Hensher, D., Waldman, D.: Recent progress on endogeneity in choice modeling. *Mark. Lett.* **16**(3–4), 255–265 (2005)
- Manski, C.: The structure of random utility models. *Theor. Decis.* **8**(3), 229–254 (1977)
- Martínez, F., Aguila, F., Hurtubia, R.: The constrained multinomial logit: a semi-compensatory choice model. *Transp. Res. Part B Methodol.* **43**(3), 365–377 (2009)

- McFadden, D.: Conditional logit analysis of qualitative choice behavior. In: Zarembka, P. (ed.) *Frontiers in Econometrics*, pp. 105–142. Academic Press, New York (1974)
- Nelder, J., Mead, R.: A simplex method for function minimization. *Comput. J.* **7**, 308–313 (1965)
- Oppenheim, N., et al.: *Urban Travel Demand Modeling: From Individual Choices to General Equilibrium*. Wiley, New York (1995)
- Ortúzar, J.D., Willumsen, L.G.: *Modelling Transport*, 4th edn. Wiley, New York (2011)
- Paleti, R.: Implicit choice set generation in discrete choice models: application to household auto ownership decisions. *Transp. Res. Part B Methodol.* **80**, 132–149 (2015)
- Popuri, Y., Ben-Akiva, M., Proussaloglou, K.: Time-of-day modeling in a tour-based context: Tel Aviv experience. *Transp. Res. Rec.* **2076**, 88–96 (2008)
- Schölkopf, B., Herbrich, R., Smola, A.J.: A generalized representer theorem. In: Helmbold, D., Williamson, B. (eds.) *Computational Learning Theory. COLT 2001. Lecture Notes in Computer Science*, vol. 2111. Springer, Berlin (2001)
- Swait, J.: A non-compensatory choice model incorporating attribute cutoffs. *Transp. Res. Part B Methodol.* **35**(10), 903–928 (2001)
- Swait, J., Ben-Akiva, M.: Incorporating random constraints in discrete models of choice set generation. *Transp. Res. Part B* **21**(2), 91–102 (1987)
- Swait, J., Marley, A.: Probabilistic choice (models) as a result of balancing multiple goals. *J. Math. Psychol.* **57**(1–2), 1–14 (2013)
- Tikhonov, A., Arsenin, V.Y.: *Solutions of Ill-posed Problems*. Wiley, New York (1997)
- Van Der Pol, M., Currie, G., Kromm, S., Ryan, M.: Specification of the utility function in discrete choice experiments. *Value Health* **17**(2), 297–301 (2014)
- Zambrano-Bigiarini, M., Clerc, M., Rojas, R.: Standard particle swarm optimisation 2011 at cec-2013: a baseline for future PSO improvements. In: *2013 IEEE Congress on Evolutionary Computation (CEC)*, IEEE, pp. 2337–2344 (2013)

José Luis Espinosa-Aranda holds a Ph.D. in Computer Science from University of Castilla-La Mancha. He is currently working as a research assistant of the VISILAB group, University of Castilla-La Mancha, Spain. His research interests are focused on intelligent systems and artificial intelligence.

Ricardo García-Ródenas is an assistant professor of applied mathematics at the University of Castilla-La Mancha. He received a Ph.D. from the Universidad Politécnica de Madrid, Spain, in 2001. Since 2002, his research has focussed on railway system planning, travel demand modeling, and optimization.

María Luz López-García is an assistant professor of applied mathematics at the University of Castilla-La Mancha. She received her Ph.D. degree in computer science from the UCLM, Spain, in 2013. Her current research interests include machine learning applied to transportation systems and artificial intelligence.

Eusebio Angulo holds a Ph.D. in Computer Science from University of Castilla-La Mancha. He is an assistant professor of applied mathematics at the University of Castilla-La Mancha, Spain. His research interests are focused on heuristic algorithm and operations research.