Transportation Research Part B 43 (2009) 365–377

Contents lists available at ScienceDirect

Transportation Research Part B

journal homepage: www.elsevier.com/locate/trb

The constrained multinomial logit: A semi-compensatory choice model

Francisco Martínez *, Felipe Aguila, Ricardo Hurtubia

University of Chile, P.O. Box 228-2 Santiago, Chile

ARTICLE INFO

Article history:
Received 21 November 2005
Received in revised form 13 June 2008
Accepted 17 June 2008

Keywords:
Discrete choice
Logit models
Constrained behavior
Consumer's utilities

ABSTRACT

The traditional formulation of logit models applied to transport demand assumes a compensatory (indirect) utility function in which the consumers' strategy assumes a trade-off between attributes. Several authors have criticized this approach because it fails to recognize attribute thresholds in consumer behavior, or a more generic domain where such a compensatory strategy is contained. In this paper, a mixed strategy is proposed, which combines the compensatory strategy valid in the interior of the choice domain with cutoff factors that restrain choices to the domain edge. The proposed model combines the multinomial logit model with a binomial logit factor that represents soft cutoffs. This approach extends previous contributions in several ways and allows multiple dimensions for cutoff factors. In addition to considering individual behavior, it introduces system constraints such as capacity and inter-agent interactions (choice externalities). This extension yields a non-linear problem, which is solved by analyzing the fixed point problem. Additionally, a set of evaluation tools, a social utility of the constrained problem, and a measure of the shadow price of each constraint, are proposed.

1. Introduction

Following Domencic and McFadden's book (1975), the random utility model assuming a Gumbel distribution for utilities has been widely applied in urban studies, producing an extensive literature of logit models based on different covariance matrix structures, such as multinomial, nested, and mixed logit models, among others. The main microeconomic foundation assumptions of these models is the compensatory strategy followed by individuals, i.e., their decision strategy assumes a trade-off between attributes. This assumption has been criticized by several scientists who claim that non-compensatory behavior is potentially more realistic, as, for example, the elimination-by-aspect (EBA) process (Tversky, 1972). A natural approach to relaxing the compensatory assumption, proposed by Manski (1977) and followed by Swait and Ben-Akiva (1987), Ben-Akiva and Boccara (1995) and Cantillo and Ortúzar (2005), among others, is to explicitly model the choice set generation process using a two-stage approach: first, the feasible choice set is generated for each individual and, second, a compensatory model calculates the choice probability conditional on the choice set. The appeal of this approach is that it permits different models to simulate the phenomena associated with each stage (Cascetta and Papola, 2001), but it is computationally complex because the number of possible choice sets explodes with the number of alternatives, with a maximum of $2^m - 1$ choice sets for m alternative options. Heuristic approaches have been proposed to reduce this difficulty, such as the pairwise comparisons of alternatives suggested by Morikawa (1995). Nevertheless, the choice set formation process is not sufficiently efficient if the number of alternatives is large, as in the case of spatial choices (e.g., trip destination and location choices), and is not applicable in more complex processes involving intensive choice making calculations, such as equilibrium and optimization processes.

* Corresponding author. Tel.: +56 2 9784380.
E-mail address: fmartine@ing.uchile.cl (F. Martínez).

0191-2615/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.trb.2008.06.006

The model proposed by Cascetta and Papola (2001) follows a different approach. It extends the compensatory utility function in order to implicitly simulate (rather than generate) the perception/availability of an alternative, leading to a one-step approach called the implicit availability/perception model (IAP). In this model, the choice-set of alternatives is a fuzzy set, where each element has a degree of membership to the choice set; thus, the choice-set is "soft"[1] rather than "crisp".

Swait (2001) models choice behavior incorporating a wide range of decision strategies using a third approach. He extends the standard deterministic utility maximization problem by including constraints on the values that the attributes and prices can attain for a choice to be feasible, which define a set of lower and upper bounds or cutoffs for each alternative. These constraints represent a feasible domain where the individual is willing to or can make choices, with attribute bounds reproducing ideological cutoffs, (for example the EBA process), economic constraints (e.g., income or time budgets), and physical limits. Thus, the author proposes a flexible version of the deterministic utility optimization problem by relaxing constraints which are introduced as linear penalties in the utility function which are activated if cutoffs are violated. This is also an implicit approach, a one-step method based on a "reduced" form model of behavior. The underpinning rationale is given by Swait: "it is behaviorally equivalent whether the decision-maker simply chooses the best good that satisfies the constraints, or alternatively, first screens based on constraints, and then chooses the best alternative". While we agree with Swait's rationale, his implementation using a linear relaxation of cutoffs introduces a "kink" in the utility (changes the slope) at the cutoff. The problem is that kinks in the utility function make it non-differentiable at the cutoff, which introduces a problem in certain complex calculation processes, such as equilibrium and optimization processes, or in systems with a large number of agents making choices on choice sets which change their attributes in the process (like price adjustment to equilibrium).

The constrained logit model (CLM) proposed in this paper combines aspects of Swait's model and the IAP model. It follows a one-stage approach using a reduced utility function that implicitly imposes cutoffs to choice makers. Our constrained utility function is similar to the IAP model in that it applies the binomial logit to simulate soft cutoffs by a continuous and differentiable extended utility function. However, we extend the approach for a full set of constraints on attributes and prices, so the CLM constrains choices to a multi-dimensional domain. We also advance the analysis for the case of a multinomial version of the CLM, denoted CMNL. For this model, we innovate by studying the more complex case of system constraints, where the set of alternative attributes depend on the choices potentially made by the whole population of decision-makers. In this case, these constraints introduce endogenous variables in the forecasting process in order to represent the complex issue of externalities in consumption.

The extension that enables the model to cope with externalities opens up the set of applications to a variety of complex cases where endogenous variables are relevant. In fact, it makes the demand model able to simulate emerging behavior observed at a macro level while being generated based on micro level (individual) decisions. However, it also introduces the need to study the solution(s) that emerge from such a model. In this paper, we prove that the proposed model has a unique solution and that the fixed point iteration converges to that solution. The proving theorem, however, applies *only* for the binomial logit used in CMNL model. The theorem makes the model applicable to a wide range of real problems where attributes are endogenous, where the demand model inevitably embeds an equilibrium problem. A good example of these attributes is the residential location model, where zone quality is defined as a function of the location choice of all agents; thus, one's demand for a location option depends on others' location choice behavior. This case describes Shelling's (1978) model, where individuals' micro motives, such as the will to live among peers, leads to a macro land use distribution reflecting spatial socioeconomic segregation.

The applications of the CMNL cover a wide range, as we discuss below. The model enhances the logit model for current applications since consumers always face a variety of constraints, for example, income, time, and choice attributes that violate individuals' limits. Additionally, on the supply side, producers may also be modeled as choice makers facing constraints that can be explicitly modeled using the CMNL, for example, in the real estate market where zoning constraints are numerous for developers.

The CLM's theoretical framework is presented and discussed in the following sections. Next, we define choice probabilities for the special case of the multinomial logit, which specifies the constrained multinomial logit model (CMNL), and we analyze the model calibration with empirical examples. Then, we study the use of the CMNL to forecast demand, with a focus on the non-linear effect introduced by endogenous constraints. This model is then further studied to produce two evaluation tools: a social benefit measure and the shadow price for each constraint. The paper ends with a brief presentation of the range of potential applications of the CLM model in spatial studies.

2. The constrained discrete choice problem

Consider the following class of optimization problems widely used in microeconomic theory to describe agent behavior of discrete goods. Each agent n behaves according to the indirect utility function U_n when deciding the best choice among a set of I discrete alternatives contained in the set C of available alternatives. Assume that the utility function depends on a set of $K-1$ attributes, denoted by vector $X \in R^{(K-1)\times I}$, and on the price vector $p \in R^I$.

From microeconomics, the rational consumer's choice behavior among the set of discrete options C is described by the following optimization problem:

[1] Soft constraint means that the constraint can be violated to some limited extent.

$$\begin{aligned} \max_{\delta_{ni}} \quad & \sum_{i \in C} \delta_{ni} U_n(X_i, p_i) \\ \text{s.a} \quad & \sum_{i \in C} \delta_{ni} = 1, \quad \delta_{ni} \in \{0,1\} \quad \forall i = \{1,2,\ldots,I\} \in C \end{aligned} \tag{1}$$

where δ_{ni} represents the individual's choice, X_i is the vector of exogenous attributes that describes alternative *i*, p_i is the exogenous price of the alternative, and $U(X,p)$ is the indirect utility function.

A key point is that the utility function embeds, by definition, all budget constraints faced by the individual in making a choice, such as the individual's income and time, and by assumption, all other constraints as well. The violation of this assumption implies potentially strong errors in forecasting demand. Therefore, in modeling the consumer behavior described in (1), the modeler faces the difficulty of specifying an indirect utility function able to reproduce the complexity of real behavior under constraints. For example, the behavior of choosing under the elimination by aspects process, where some attributes surpass levels that saturate utility, or the behavior in the vicinity of the constraints imposed by income and time expenditures. Although the specification of such utility functions is theoretically feasible, functions that can comply with that requirement have a complex non-linear form and are difficult to implement in real studies, the majority of which use the simple linear form for utilities or other forms limited by the data available to calibrate them.

Thus, instead of searching such a complex function, one can modify the consumers' problem into:

$$\begin{aligned} \max_{\delta_{ni}} \quad & \sum_{i \in C} \delta_{ni} \overline{U}_n(X_i, p_i) \\ \text{s.a} \quad & \sum_{i \in C} \delta_{ni} = 1, \quad \delta_{ni} \in \{0,1\} \ \forall i \in C \\ & a_{nk} \leqslant X_{ik} \leqslant b_{nk} \quad \forall i \in C, \ k = \{1,\ldots,K-1\} \\ & a_{nK} \leqslant p_i \leqslant b_{nK} \quad \forall i \in C \end{aligned} \tag{2}$$

where the consumer's feasible/acceptable domain D_n is explicit. For simplicity, hereafter, we define vector $Z_i = (X_i, p_i) \in R^K$, which contains attributes and the price of the alternative *i*.

Note that in making constraints on the attributes and prices explicit, the modeler has to specify a more simple function $\overline{U}$ able to correctly represent behavior only in the interior of the domain; the behavior at the border of the domain is controlled in (2) by explicit constraints. This motivates the decomposition of the consumer's problem into a compensatory utility function in the interior of the consumers' domain ($\overline{U}$), and an explicit model of the behavior in the vicinity of the domain. Such decomposition makes use of a simple functional form for $\overline{U}$ more plausible. This motivated Manski's (1977) two-stage approach, where the first stage is a model that identifies the subset C_n where all constraints are complied, followed by a second stage that models the maximum utility choice within this subset using a compensatory utility.

We denote bounds for attributes and prices by the following vectors for lower and upper bounds respectively: $\theta_n^{\mathrm{L}} = [a_{n1}, a_{n2}, \ldots, a_{nK}], \theta_n^{\mathrm{U}} = [b_{n1}, b_{n2}, \ldots, b_{nK}]$. Note that these bounds are assumed to be exogenous and independent of the specific alternative, but the approach may be easily extended to consider the case of alternatives with specific bounds. Note also that bounds are assumed as specific to the choice maker; the case of constraints equal to all individuals is a special and simpler case of problem (2).

Problem (2) can be written by the following unconstrained Lagrange function:

$$\max_{\delta_{ni}} L = \sum_{i \in C} \left(\delta_{ni} \overline{U}_n(X_i, p_i) + \sum_k \lambda_{nik}(b_{nk} - Z_{ik}) - \sum_k \kappa_{nik}(a_{nk} - Z_{ik}) \right) + \gamma_n \left(\sum_{i \in C} \delta_{ni} - 1 \right) \tag{3}$$

where γ, λ and κ are Lagrange multipliers associated to the constraints. The fact that problem (3) is unconstrained motivates the implicit approach followed by Swait (2001) and Cascetta and Papola (2001), and is used in the following section to develop a model of the consumer behavior problem (2) based on the random utility theory.

3. The constrained random utility

Consider now the classical model where the utility function is a random variable, that is $\overline{U}_n = \overline{V}_n + \varepsilon_n$, with $\overline{V}_n$ as a systematic compensatory utility, and ε_n a random term. The widely used logit model is derived upon assuming that random terms are distributed Gumbel, which implies that $\varepsilon \in [-\infty, \infty]$, and utilities are unconstrained. Thus, by definition, logit models assume unconstrained utilities, thus choice behavior is also unconstrained unless the choice set is specified, which eliminates all alternatives whose attribute vector lies out of the consumer's choice domain, which is the two-stage approach mentioned above. Naturally, one can choose a distribution of utilities that, in contrast to the Gumbel distribution, is constrained; this would yield a family of models which are not studied in this paper.

Alternatively, we use the implicit method. We define a "constrained utility" function that induces the individual to make choices belonging to her feasible domain D_n with a certain probability. As will be evident later, this probability may be as high as desired, but not certain, because we allow cutoffs to be violated with a given probability $\eta = \{\eta_k, k = 1, \ldots, K\}$, where each element in this vector is the violation probability associated with the respective attribute k's cutoff. Additionally, the constrained utility function is defined as compensatory in the interior of the individual's domain, but non-compensatory in a vicinity of the domain.

To obtain a utility function constrained to a domain, as in the IAP model, we also augment the usual compensatory utility function by a new cutoff term, called the utility penalty. Thus, a compensatory term (V^c) and an additive cutoff term define the constrained utility as follows:

$$V_n(Z_i) = V_n^c(Z_i) + \frac{1}{\mu}\ln\phi_{ni}(Z_i) + \varepsilon_{ni} \tag{3}$$

with ε assumed as Gumbel distributed $(0,\mu)$. Notice that a difference from Cascetta and Papola's (2001) model is that here the cutoff term is amplified by the inverse of the Gumbel scale parameter $(1/\mu)$, which increases the penalty as the utility dispersion increases. This difference means that the cutoff term in our model becomes more relevant as variability of utility increases (μ decreases).

Since the set of constrains affecting the choice of a given alternative may be more than only one, then the penalty term is now defined as the following multi-dimensional composite cutoff factor $\phi_{ni} = \prod_{k=1}^{K}\phi_{nki}^{L}\cdot\phi_{nki}^{U}$, which is composed of a set of K lower and upper cutoffs. Now, each elementary cutoff factor is defined as a binomial logit function. This function has been used previously for its simplicity by Swait and Ben-Akiva (1987), Ben-Akiva and Boccara (1995) and Cascetta and Papola (2001). We used it not only for its simplicity, but also because it provides some relevant properties required to prove the theorem for the case of endogenous constraints, as shown below.

Then, for each n_{ki}, we define the elementary lower and upper cutoff factors by

$$\phi_{nki}^{L} = \frac{1}{1+\exp(\omega_k(a_{nk} - Z_{ki} + \rho_k))} = \begin{cases} 1 & \text{if } (a_{nk} - Z_{ki}) \to -\infty \\ \eta_k & \text{if } a_{nk} = Z_{ki} \end{cases} \tag{4a}$$

$$\phi_{nki}^{U} = \frac{1}{1+\exp(\omega_k(Z_{ki} - b_{nk} + \rho_k))} = \begin{cases} 1 & \text{if } (b_{nk} - Z_{ki}) \to \infty \\ \eta_k & \text{if } b_{nk} = Z_{ki} \end{cases} \tag{4b}$$

We define $\omega_j > 0, \forall j$, because negative values simply convert (4a) into (4b) and vice versa. Additionally, we define

$$\rho_k = \frac{1}{\omega_k}\cdot\ln\left(\frac{1-\eta_k}{\eta_k}\right) \tag{4c}$$

The parameter η is the cutoff tolerance, which defines the choice probability at the boundary, and ρ_k is defined in the same units as the kth variable. This tolerance can be as small as desired, but not zero, implying that the model cannot be applied for deterministic compliance of constraints; some degree of tolerance is structurally imposed. For simplicity, η is specified as constant for all agents, but an individual specific constant is also possible. The performance of other functions for the cutoff may be explored; for example, Cascetta and Papola (2001) analyze the Gamma distribution for the single (not composite) cutoff factor.

Observe that the generalized cutoff factor is (quasi)innocuous for any feasible alternative, i.e., those with vector $Z_i \in D_n$, because $\phi_{ni} \to 1$; conversely, if any element $Z_{ki} \notin D_n$ then $\phi_{ni} \to 0$, the alternative's utility tends to minus infinity, and the choice probability tends to zero, performing a soft compliance of the constraint. Fig. 1 depicts the binomial (lower and upper) cutoff functions, and Fig. 2 shows that the parameter ω controls the softness of the cutoff by defining the slope of the cutoff function.

Note that in the case of a deterministic behavior, theoretically only the most binding constraint on each variable will affect choices, because the rest are feasible by definition. In contrast, in our stochastic approach, even the least binding cutoff on one attribute will have some effect on choices. Additionally, the more the constraint is violated, the larger the effects of the cutoff. For example, consider the case of a choice alternative with a price close to a self-imposed maximum expenditure, which defines the first constraint; the second less binding, but stricter constraint is the individual's income. In the model, the utility will tend to be reduced primarily by the first cutoff, but some extra reduction is produced by the second cutoff. This combination of effects and the presence of a violation tolerance seem to be plausible in a real context because they can represent the stochastic nature of constraints associated to unpredictable events affecting the consumer's disposable income, such as accidents or health problems. The method may also be applied to consider constraints defined as a mix of attributes.

Our approach can be compared with Swaits' (2001) model because both models penalize utilities of choices out of the domain, but while his model assumes a linear penalty function, ours is non-linear. Indeed, our utility penalty factor is

$$\ln\phi_{ni} = \ln\left[\prod_{k}^{K}\phi_{nki}^{L}\phi_{nki}^{U}\right] = \sum_{k}^{K}(\ln[\phi_{nki}^{L}] + \ln[\phi_{nki}^{U}])$$

then

$$\ln\phi_{ni} = -\sum_{k}^{K}(\ln[1+\exp\omega(a_{nk} - Z_{ki} + \rho_k)] + \ln[1+\exp\omega(Z_{ki} - b_{nk} + \rho_k)]) \tag{5}$$

This penalty is negative (disutility) for all Z_i out of the individual's domain D_n, and increases exponentially as one (or more) attributes fall further out of the domain. Another relevant difference is that Swaits' linear penalty yields a continuous utility

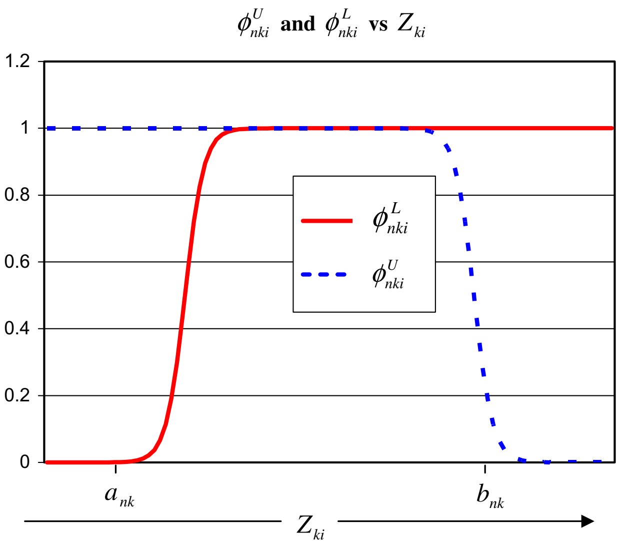

Fig. 1. The lower and upper binomial cutoff functions ϕ^U_{nki} and ϕ^L_{nki} vs. Z_{ki}.

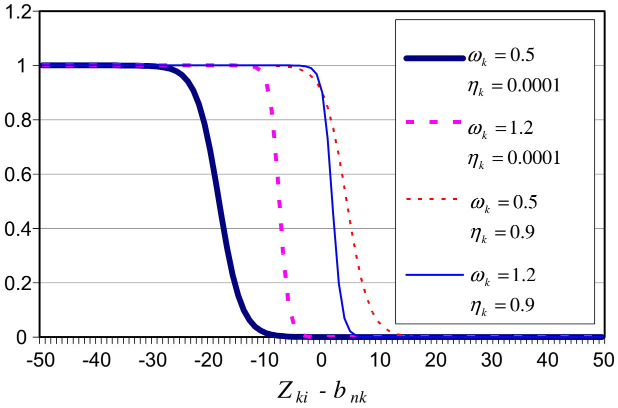

Fig. 2. The effect of the scale parameter on upper cutoff factors.

function but is not differentiable at the cutoff values of attributes, while the advantage of the non-linear approach is that utilities are continuous and differentiable for all $Z \in R$.

At this point we argue, along with other authors previously mentioned, that the optimization problem with soft constrained utilities is the natural representation of the individual's choice problem. This argument arises from the observation that in social sciences, cutoff limits are naturally soft because individual choices are subject to the individuals' perceptions, even in the case of physical constraints such as infrastructure capacity. Then, in models where a choice maker represents a set of individuals with differentiated reactions to constraints, the emerging demand for a given alternative is naturally represented by the soft cutoff rather than a deterministic cutoff.

The cutoff tolerance parameter ρ may be conceived in the context of a process with memory, because the tolerance for accepting penalties may be specified as dependent on previous experiences, in a way such that those individuals that had chosen alternatives in the vicinity of the domain limit have a better knowledge of the penalty and the benefits/costs of the choices made. Thus, we can postulate that the tolerance differs across similar individuals in all aspects, including their experience in choosing alternatives in the vicinity of specific constraints.

4. The constrained multinomial logit model (CMNL)

The individual choice problem under the constrained utility function defined in Eq. (3), is

$$\underset{i \in C}{\text{Max}}\, V_n(Z_i) = V_n^C(Z_i) + \frac{1}{\mu} \ln \phi_{ni} + \varepsilon_{ni} \tag{6}$$

with C the universal set of alternatives and $\phi_{ni} = \phi_n(Z_i)$. Expression (6) is the reduced stochastic objective function that represents a stochastic version of the choice problem (2), with V as the indirect utility function which complies with the feasible domain. The solution of this problem yields the following constrained choice probabilities:

$$P_{ni} = \text{Prob}\left[V_{ni}^C + \frac{1}{\mu}\ln\phi_{ni} + \varepsilon_{ni} \geqslant \max_{j\in C}\left(V_{nj}^C + \frac{1}{\mu}\ln\phi_{nj} + \varepsilon_{nj}\right)\right] \tag{7}$$

Moreover, the assumption that the constrained utility is distributed as an identical and independent Gumbel yields the following multinomial probability function:

$$P_{ni} = \frac{\phi_{ni}\cdot\exp(\mu V_{ni}^C)}{\sum_{j\in C}\phi_{nj}\cdot\exp(\mu V_{nj}^C)} \tag{8}$$

This expression represents a choice probability which complies with the feasible domain D_n, which tends asymptotically to zero if any of the alternative attributes violate any cutoff. At the boundary of the domain, the usual compensatory probability is multiplied by tolerance probability factors ηs. This model, named the constrained multinomial logit (CMNL), preserves the closed form of the equivalent classical compensatory logit models and the technical properties of the multinomial logit with a non-linear utility function given by (6). Mathematically, the model reproduces the function of compensatory utility models in the interior of the domain, but in the vicinity of the domain boundaries, diverts so that the probability drastically falls to (near) zero. The transition from the compensatory model to a constrained model is smooth, because the utility functions are continuous and differentiable in the whole real space.

4.1. Calibration issues and applications

It is now important to make some comments regarding calibration of the CMNL model. The population is assumed classified into socioeconomic groups, whose members are then assumed to behave identically, except for idiosyncratic variability represented by ε_n. Thus, we calibrate parameters for a representative individual identified by index n in Eq. (6) using observations of consumers belonging to this category and their choices. The set of parameters include those of the compensatory utility and the set $(\rho_{nk},\omega_{nk};\forall n,k)$ of the cutoff functions. The data required for the model calibration includes the set of exogenous attributes (vector Z) of the choice set, the observation of the chosen alternative in each observation, and socioeconomic indices of the choice makers, which is the standard data for logit models. Additionally, all exogenous constraints should be identified, and the values of the constraints are inputs for the calibration. As in the unconstrained multinomial logit model, the parameter μ is not identifiable; it is embedded in the parameters calibrated for the compensatory utility V^C and does not affect cutoffs when expressed as in Eq. (8).

The methodology for the calibration of parameters distinguishes two cases. The first case considers exogenous constraints, where (a_{nk} and b_{nk}) are inputs for each attribute k (when they apply). In this case, the standard maximum likelihood method with non-linear utility functions is applicable. The second case considers endogenous or self-imposed constraints, for example, when a minimum quality is required for some attributes as a requisite for the alternative to be considered in the option set; here, the a and b parameters are unknown for the modeler, and have to be identified endogenously, for example, using the calibration method proposed by Cascetta and Papola (2001), which takes into account interdependencies on the random terms.

In both cases, the calibration procedure faces the difficulty that cutoff parameters are associated to behavior in the very specific region of attributes near the domain border, while within the domain, they may affect the parameters of the compensatory utility. This fact calls attention to a potential identification problem of parameters ρ and ω using revealed preferences, unless the calibration sample contains sufficient information of the choices near the domain border. In any case, the calibration data may be divided into two subsets, one defined by choices in the interior of the domain, which can only identify the compensatory utility function, and a second one defined by choices in the vicinity of the domain border, which is appropriate to identify cutoff parameters. In the case where the calibration data came from a stated preferences (SP) experiment, it is possible to design such experiment to explore and report on choice behavior in both sub-domains, where compensatory behavior applies, and at the cutoff vicinity. However, this is not always possible, or at least less directly applicable, with revealed preference data.

We report on two experiments. The first one refers to the calibration process and results obtained for a land use model. We calibrate bid functions (denoted as B) in a CMNL model describing a location auction, given by

$$P_{n/i} = \frac{\phi_{ni}\exp(B_{ni})}{\sum_{g\in C}\phi_{gi}\exp(B_{gi})}$$

with $\phi_{ni}^k = \frac{1}{1+\exp(\omega(\delta_n - I_i))}$, and $B_{ni} = \alpha_n + \sum_k \beta_{nk} x_{ik}$. We modeled the probability that a household type n is the highest bidder in location i, $P_{n/i}$, in an auction where all households in set C are allowed to bid for i. In this example, we used Eq. (4a) to model a lower cutoff, reflecting that some consumers will not be willing to bid in zones having an average income (I_i) below certain minimum, which are defined endogenously by ρ_n and different for each group n. We have assumed $\omega_n = \omega\,\forall n$ and $\delta_n = a_n + \rho_n$, which simplifies Eq. (4a).

Two MNL and CMNL models are calibrated in this experiment and the results are shown in Table 1 in Appendix 1. Both include the attribute "resident average income in the zone", the MNL applied in a compensatory way, and the CMNL as a cutoff function for this variable.

We used a cross-section revealed preferences data set with 600 observations, obtained from a household survey in Santiago city in 2002, which contains information on the socioeconomic characteristics of the household, the residential location, and the dwelling attributes. Land use data was used to define zone attributes. We considered five income groups (indexed by n, which increases with income); the variables included in the bid function included are floor space of the building, average income of residents in the zone (both as logarithms), and accessibility. Only one cutoff was considered to model the segregation behavior of households in their residential location choice, such as high income groups allocated among peer income groups.

The results show that including the zonal income variable improves the log-likelihood indicator in both models, but the way in which it is defined, as a compensatory term or as a cutoff term, makes no relevant difference in reproducing the observed location distribution (e.g., they have a similar log-likelihood indicator). Nevertheless, we obtain some important conclusions from this exercise. First, that the cutoff parameters (ω, ρ) are identified, some with high level of significance. Second, notice that in the CMNL model the estimates of the constant parameters α_n are lower than in the MNL, which means that part of the behavioral information explained by the constants in the latter model are transferred to the cutoff in the CMNL model. This is a positive effect because the behavior described by the CMNL is more responsive to changes in the zone income variable than a simple constant. Finally, although these models are similar in representing the observed location distribution, they produce very different forecasts of location patterns when they are applied to scenarios where the constrained attribute changes significantly.

The second experiment uses a synthetic data set for mode choice with 679 cases, based on data collected on a corridor (Las Condes – Centro) of Santiago city. Two models were calibrated: one is a standard multinomial logit model (MNL) with a linear utility function, and the other a constrained multinomial logit model (CMNL) with a cutoff parameter on the walking time, which represents the travelers' maximum time in which they are willing to walk. The walking time cutoff was assumed to be 20 min, which was obtained as the value that yields the maximum log-likelihood value. The parameter η was assumed as $\eta = 0.0086$, which was calculated as the percentage of travelers which violates the cutoff in the data set. Additionally, we partitioned the sample assuming that 20% of travelers behave as having a compensatory behavior ($\phi = 1$), while the rest are assumed to be affected by the cutoff.

The results show (see Table 2 in Appendix 1) that the original parameters assumed to generate the data set are recovered by the calibration process; that is, the models are significantly different (based on the χ^2 test on the likelihood index) and the CMNL model has a better fit than the MNL model.

Additional experiments show that the likelihood index is sensitive to the partition of the data set into those observations, affected or not by the cutoff. Moreover, they show that if data is not partitioned at all, some cutoff parameters cannot be identified because of their correlation with parameters of the compensatory utility. These results call for further research on the calibration methodology and on data collection.

5. Forecasting issues

The model structure also has relevant implications in the context of forecasting demand when (some) constraints are endogenously defined by all consumers, which we call system constraints. In the presence of these constraints, forecasting demand involves solving an equilibrium problem where the demand model's structure is crucial for studying the equilibrium.

System constraints can belong to two categories. One includes constraints exogenous to the consumer, for example, if the alternative has capacity (e.g., road and public transport capacity, land space, and numerous policy regulations). Another category contains endogenous constraints, such as individual thresholds, where the associated attributes are defined by the outcome of all other consumers' choices; for example, neighborhood quality in residential location choice when quality is defined, and for instance, by socioeconomic, racial, or religious condition of neighbors and in-vehicle congestion in transport choice. In economic terminology, these endogenous thresholds are called consumption externalities, characterized by producing fundamental, real, and complex market effects.

A large number of exogenous system constraints may be expressed by the following (linear) expression:

$$\bar{a}_{ij}^{\mathrm{L}} \leqslant \sum_n y_{ij} P_{ni} \leqslant \bar{b}_{ij}^{\mathrm{U}} \tag{9}$$

where y_{ij} are exogenous parameters that define the amount of the scarce resource j used if alternative i is chosen, P_{ni} is the probability that consumer n chooses alternative i, and $\bar{a}_{ij}$ and $\bar{b}_{ij}$ are the lower and upper constraints for the jth system constraint affecting alternative i. We define the aggregated demand for resources j generated by alternative i, given by $\overline{Y}_{ij}(P) = \sum_n y_{ij} P_{ni}$.

To introduce exogenous system constraints in the demand model, we apply the reduced (or constrained) utility approach that internalizes all system constraints on each individual choice process. Again, we define the vector of system constraints for each of the I alternatives and J constraints for each alternative:

$$\bar{\theta}_i^{\mathrm{L}} = [\bar{a}_{i1}, \bar{a}_{i2}, \ldots, \bar{a}_{ij}], \quad \bar{\theta}_i^{\mathrm{U}} = [\bar{b}_{i1}, \bar{b}_{i2}, \ldots, \bar{b}_{ij}] \tag{10}$$

which define the alternative's sub-domain $\overline{D}_i$.

As above, the constrained utility function Eq. (3) is further augmented by penalties of violating the system constraints, yielding:

$$\overline{V}_n(Z_i) = V_n^C(Z_i) + \frac{1}{\mu}\ln\phi_{ni}(Z_i) + \frac{1}{\mu}\ln\Phi_i(P_i) + \varepsilon_{ni} \tag{11}$$

where the system cutoff factor is defined as a function of the choice probabilities on alternative i, given by matrix P_i, for all individuals. Additionally, $\Phi_i = \prod_{j=1}^{J}\Phi_{ij}^{\mathrm{L}} \cdot \Phi_{ij}^{\mathrm{U}}$ with each elemental term defined by

$$\Phi_{ij}^{\mathrm{L}}(P_i) = \frac{1}{1+\exp(\bar{\omega}_j(a_{ij} - \overline{Y}_{ij}(P_i) + \bar{\rho}_j))} = \begin{cases} 1 & \text{if } (\bar{a}_{ij} - \overline{Y}_i) \to -\infty \\ \bar{\eta}_j & \text{if } \bar{a}_{ij} = \overline{Y}_{ij} \end{cases} \tag{12a}$$

$$\Phi_{ij}^{\mathrm{U}}(P_i) = \frac{1}{1+\exp(\bar{\omega}_j(\overline{Y}_{ij}(P_i) - \bar{b}_{ij} + \bar{\rho}_j))} = \begin{cases} 1 & \text{if } (\bar{b}_{ij} - \overline{Y}_{ij}) \to \infty \\ \bar{\eta}_j & \text{if } \bar{b}_{ij} = \overline{Y}_{ij} \end{cases} \tag{12b}$$

$$\bar{\rho}_j = \frac{1}{\bar{\omega}_j} \cdot \ln\left(\frac{1-\bar{\eta}_j}{\bar{\eta}_j}\right) \tag{12c}$$

with $\bar{\omega}_j > 0\,\forall j$, and

Endogenous system constraints naturally represent consumption externalities because they introduce interdependencies in consumption between consumer agents. These externalities may affect utilities through changes in prices (pecuniary externalities) or by directly changing attributes (technological externalities). The CMNL model can represent these externalities by making $Z = Z(P)$ in Eq. (4), and the system constraint is represented by endogenous cutoffs on these attributes. Therefore, the model described above does provide the flexibility to accommodate this case.

The combination of all constraints restricts individual choice probability to the domain $\widetilde{D}_n = D_n \bigcap_j^J D_j$, which is defined by the augmented constraints vector $\tilde{\theta}_n = \theta_n^{\mathrm{L}} \cup \theta_n^{\mathrm{U}} \cup \bar{\theta}^L \cup \bar{\theta}^U;\ \tilde{\theta}_n \in R^{2(K+I)}$. Then, the CMNL model (Eq. (8)) can be extended to recognize system externalities as follows:

$$P_{ni} = \frac{\tilde{\phi}_{ni}(P_i) \cdot \exp(\mu V_{ni}^C(P_i))}{\sum_{j\in C}\tilde{\phi}_{nj}(P_j) \cdot \exp(\mu V_{nj}^C(P_j))} \tag{13}$$

where P is the constrained choice probability, and $\tilde{\phi}_{ni} = \phi_{ni} \cdot \Phi_i$ is the composite cutoff factor including individual and system constraints.

Notice that a system constraint effectively makes the individual utility dependent on other consumers' choices, then dependent on others' utility levels, by means of the joint consumption of capacity and by consumption externalities. Because the constraints are defined at a system demand level, they do not induce further identification issues in the calibration process, as they can be assumed exogenous for the consumer. However, it does raise the issue of the complexity induced by interactions between consumers in the forecasting process, which is a matter discussed in the rest of this section.

Observe that Eq. (13) represents an equilibrium equation for the choice probabilities, which makes each individual's choice dependent on the other individuals' choice. This describes a fixed point problem $P = f(P)$, with P representing the probabilities matrix of individual choices across travel alternatives, composed by a system of $I \cdot N$ non-linear equations and the same number of variables. The algorithm studied to solve the equilibrium condition is the fixed point iterative process, $P^t = f(P^{t-1})$, where t is the iteration number, which starts with any initial matrix P^0. In the Appendix, we prove the following theorem:

Theorem (Existence, uniqueness, and convergence). *The CMNL model has a unique fixed point solution, and the fixed point iteration converges to the solution if:*

1. $\frac{1}{\lambda} > 2 \cdot \max_{mz}\left\{\sum_{ni}\left|\frac{\partial V_{ni}^c}{\partial P_{mz}}\right| + |n| \cdot \left(\sum_{l=1}^{J} y_{zl}| + \sum_i \sum_{l=1}^{K}\left|\frac{\partial Z_{li}}{\partial P_{mz}}\right|\right)\right\}$
2. $\frac{1}{\lambda} > \left(|n| \cdot \sum_z \sum_{l=1}^{J}|y_{zl}| + \sum_{mzs}\sum_{l=1}^{K}\left|\frac{\partial Z_{ls}}{\partial P_{mz}}\right| + \max_{ni}\left[\sum_{mz}\left\{\left|\frac{\partial V_{ni}^c}{\partial P_{mz}}\right| + \sum_{s\in C}\left|\frac{\partial V_{ns}^c}{\partial P_{mz}}\right|\right\} + \sum_{mz}\sum_{l=1}^{K}\left|\frac{\partial Z_{li}}{\partial P_{mz}}\right| + |n| \cdot \sum_{l=1}^{J}|y_{il}|\right]\right)$

where $\lambda = \max\{\omega; \mu\}$ *is maximum between the scale parameters of the binomial and multinomial functions.*

Proof. See Appendix 2. □

The conditions on λ are obtained by imposing on $f(P)$ the satisfaction of Banach theorem's conditions by imposing that its Jacobean norm be less than 1. This is a sufficient, although not necessary, condition for contractiveness of the fixed point function, which implies that the convergence to a unique solution is assured. Observe that the convergence condition is very strong, because they are sufficient conditions over the whole domain. In fact, violation of these bounds does not necessarily imply lack of convergence; it means only lack of guaranteed convergence.

The theorem has the benefit that it assures that the CMNL fixed point algorithm converges to the unique solution under certain conditions on the model parameters. The convergence condition requires that a minimum dispersion be present on individuals' choice behavior, i.e.: if the choice process is close to deterministic, the convergence conditions are not guaranteed. In the theorem, such a condition imposes maximum values for the dispersion parameters ω and μ of the binomial and multinomial functions, respectively.

Observe that as the number of alternatives increases, each one having nonzero probability (a condition of logit models), then all probabilities tend to differ from one or zero, which means that dispersion in the choice process increases and that the convergence condition in the forecasting demand procedure is more likely to be fulfilled. Thus, in real large scale problems, the minimum dispersion condition is more often likely to be satisfied. We have performed a large number of simulation exercises solving the fixed point problem (13), with small and large problems, obtaining a high convergence performance considering the complexity of the non-linear system of Eq. (13). See, for example, Martínez and Hurtubia (2006) for the application of the CMNL model in the land use case where the number of zoning constraints for suppliers is very large; in this case, suppliers chose maximum profit locations.

The theorem constitutes a fundamental advantageous property of the CMNL model on applications used to forecast transport and land use markets. Indeed, under the presence of externalities and cutoffs, the market equilibrium problem involves solving complex non-linear problems. Most applications simply ignore these effects, but this shortcoming wrongly assumes that endogenous attributes are exogenous variables; thus, forecasts of demand are likely to violate constraints and miscalculate utilities, demand, and equilibrium prices. The theorem may be extended to other logit structures, for example, to the Nested and Mixed Logit, which remains for further research.

6. Evaluation tools

The above defined CMNL model is used in this section to derive two evaluation tools. The first one is a measure of the social benefit associated to choices made under a constrained context, defined as the expected maximum of individuals' utilities aggregated across the population. The second one measures the social cost of policies that constrain consumption (e.g., capacities and regulations), measured as the shadow price of each elemental constraint.

Consider the CMNL utility function of Eq. (11), evaluated at the demand solution and calibrated parameters, that is, at the forecast of the utility level and demand for alternatives. It is possible to examine the expected maximum utility level that the consumer can obtain from a set of alternative choices restrained to the domain $\widetilde{D}_n$, which is given by the following logsum formula:

$$\widetilde{U}_{n/C} = \frac{1}{\mu} \ln \left[\sum_{i \in C} \tilde{\phi}_{ni} \cdot \exp(\mu V_{ni}^C) \right] \tag{14}$$

This equation measures the individual's maximum expected benefit obtained from the choice-set C, which we can use to analyze the impact of urban policies on individuals' satisfaction.

The aggregate utility across N consumers is

$$\widetilde{U}_C = \frac{1}{\mu} \sum_{n}^{N} \ln \left[\sum_{i \in C} \tilde{\phi}_{ni} \cdot \exp(\mu V_{ni}^C) \right] \tag{15}$$

which represents the utilitarian social measure of the consumers' benefits; this measure ignores issues of wealth distribution.[2] Notice that the domain of this social utility function is $\widetilde{D}_C = \bigcup_n^N \widetilde{D}_n$ defined by the augmented vector $\tilde{\theta}_C = \bigcup_n \tilde{\theta}_n; \tilde{\theta}_C \in R^{(2K+I)\cdot N}$. Notice also that the parameter μ is normally unknown in applied MNL models, because it is theoretically embedded in the parameters calibrated for compensatory utility $\widehat{V}^C = \mu\widetilde{V}$; then, in this case, the parameter μ can be correctly assumed to be equal to one.

Eq. (15) provides a measure of the social benefit yield by the urban system, which can be used for evaluating different policy scenarios. One novel and logical application is to compare the benefits associated to different regulation sets: for example, in the residential location choice process or transport regulations that affect demand of specific transport modes. The benefit of changing the regulations from a scenario a to a scenario b, is given by: $\Delta\widetilde{U} = \widetilde{U}(\tilde{\theta}_b) - \widetilde{U}(\tilde{\theta}_a)$.

From this social benefit measure, one can derive the marginal social utility of violating a given constraint, or the value of marginally loosening the constraint, which is known as the shadow price of the constraint. The shadow price (S_j) associated to the jth constraint, denoted $\tilde{\theta}_j \in \tilde{\theta}_C$ with $j = 1,\ldots,L$ and $L = (2K + I)$, is calculated as the marginal utility of relaxing the constraint. Then

$$S_j = \frac{\partial U_C}{\partial \tilde{\theta}_j} = \frac{1}{\mu} \sum_{n}^{N} \sum_{i \in C} \widetilde{P}_{ni} \left[\sum_{l \in L} \omega_l \frac{(1 - \tilde{\phi}_{nl})}{\tilde{\phi}_{nl}} \frac{\partial \tilde{\phi}_{nl}}{\partial \tilde{\theta}_j} \right] \tag{16}$$

Again, in applied studies, the scale parameter μ can be assumed equal to one.

[2] Wealth distributions with different equity criteria can be introduced by adding differentiated social values for consumers' benefits.

Observe that the shadow price is strictly non-negative because all terms are positive. It increases as demand for alternatives close to the edge of the domain also increases, because cutoff factors tend to zero and $\omega_l > 0, \forall l$, so each term in parentheses has the positive sign of $\frac{\partial \tilde{\phi}}{\partial \theta}$; thus, S_j is strictly positive in that case. Conversely, if the choice pattern is sufficiently far from the cutoff in the interior of the domain, then $\tilde{\phi}$ tends to one, and shadow prices tend to zero. These two cases are consistent with the theoretical value of shadow prices under constrained behavior.

The terms in brackets recognize that our model includes multiple constraints, individual thresholds and system capacities that are potentially interdependent; if they were independent, then the cross-derivatives are equal to zero and the shadow price is only dependent on the corresponding cutoff. This is a relevant point because cross-dependency between cutoffs is likely to occur. Think, for example, of the effects of increasing the level of the individual's acceptance of travel time by car (an individual constraint); then, more users are expected to show up on congested roads, thus increasing the level of congestion (due to a system constraint) and, therefore, increasing the shadow price of road capacity constraints. Another example is in land use, where a stronger zone regulation (a system constraint), like the minimum density induces several effects on land values and location patterns, which may activate residents' thresholds (an individual constraint) on neighbor environment.

7. Applications

The potential application of constrained logit models covers the whole range of discrete choice processes in economic systems, both on the demand and supply sides, where endogenous and exogenous individual and system constraints are numerous.

In modeling the transport system, the model can be applied both for demand and supply choices. In travel demand, the usual cutoffs are budget and time resources, which are assumed exogenous in the context of transport decisions. Examples of endogenous cutoffs are associated with thresholds on several attributes: minimum activity level at destination for attracting trips, maximum spent on travel, maximum waiting, and access times to public transport. Another cutoff is the maximum walking limit, which may be taken as exogenous for the handicapped and elderly, or as endogenous for other travelers. In vehicle route-assignment models, road and vehicle capacities are exogenous cutoffs, while accepted maximum time at traffic jams is endogenous.

In location and land use modeling, cutoffs are particularly relevant. Households spend a significant proportion of their income budget on housing cost, either mortgage or rental cost, and real estate prices vary across space also significantly, which implies that housing choices are crucially determined by the income budget for the vast majority of the population. This makes prices also directly dependent on this constraint, which is reflected in the urban dynamic model developed by Martínez and Hurtubia (2006). Moreover, if real estate attributes are usually numerous, then attribute thresholds may also be numerous and diverse. Relevant location options is another interesting case, because agents are likely to have cognitive constraints to evaluating all alternative zones in a city; hence, cutoffs help to model this issue more realistically by restraining the scope of the spatial search. A similar argument applies for the destination choice in the travel demand model and for all spatial choice processes. The non-negative profit constraint in a real estate production model is also a reasonable economic assumption for the behavior of suppliers, in addition to planning regulations which represent the most numerous and diverse set of constraints for real estate supply.

8. Conclusions

Advances in discrete choice modeling have not slowed in the last three decades, but challenges in replication of the actual behavior of agents are still open for further research. Better techniques are clearly needed to deal with the high complexity of this problem, and more specific models are required for the large variety of applications. Thus, models that explicitly incorporate specific and complete sets of constraints to the choice process are clearly relevant. For instance, in random utility models, the explicit specification of the constraints in the deterministic component of the indirect utility function increases the predictive power of the deterministic part of the indirect utility function and improves the quality of calibrated parameters.

This paper proposes a method which builds upon previous techniques to make discrete choice random utility models more realistic, by adding to the theoretically sound compensatory utility functions the additional flexibility to cope with constraints to individuals' behavior. One advantage of this method is that it does not impose any limitation on the compensatory utility function; on the contrary, it enhances the performance of any function in the domain's border.

The basic method of implementing cutoffs as a binomial logit function was embedded in multinomial logit models, yielding enhanced discrete choice models with the following characteristics. Physical and economic constraints (called exogenous) and attribute thresholds (endogenous constraints) are modeled as soft cutoffs controlled by a stochastic tolerance factor. Appropriate cutoff factors reproduce the wide range of individual and system constraints. A new reduced utility function is maximized yielding a multinomial logit probability function, where usual compensatory utilities are replaced by the new constrained utility. The result is the constrained multinomial logit model (CMNL), which preserves the close form of the MNL model, allowing the choice domain to be constrained by as many cutoffs as required, limiting both upper and lower levels of variables.

Although some comments regarding the calibration of the CMNL are included in the paper, we recognize that the standard maximum likelihood method yields a nonlinear optimization problem substantially different than the known MNL model. Therefore, the calibration of the CMNL model parameters needs further research on efficient algorithms.

It is worth emphasizing that demand models are often used in forecasting future demand in an equilibrium context, where several variables become endogenous. With this in mind, the paper also analyses the use of the CMNL model in this context, because several cutoffs become endogenous, introducing extra complexity in solving the model to find the demand, because it represents an equilibrium (fixed point) problem. A theorem proves that the solution of this problem exists and is unique under minimum dispersion on choices, and that the standard fixed point algorithm converges to the solution. Additionally, empirical tests show that convergence is highly efficient regarding the complexity of the non-linear equations involved.

The paper also proposes two evaluation results. One is a social benefit measure for a constrained setting and an evaluation method for regulatory scenarios. The other one is the shadow price for each cutoff. These are useful tools for the economic evaluation of policies affecting perceptions of attribute cutoffs or system capacities.

Acknowledgements

The authors are thankful for the collaboration and fruitful comments from R. Henríquez, M. Castro and M. Munizaga. This research was partially funded by FONDECYT 1010422 and 1060788 and the Millennium Institute of Complex Engineering Systems.

Appendix 1. See Tables 1 and 2.

Appendix 2. In this Appendix, we prove the theorem of existence, uniqueness, and convergence of the fixed point problem associated with choice externalities.

Theorem A1 (Existence of endogenous cutoffs solutions). *The CMNL model has a fixed point solution.*

Proof. A direct application of the Brower's fixed point theorem yields this result for function (13) on the probability simplex $\Delta = \left\{x \in \Re^{|N||I|} / \sum_{ni} x_{ni} = 1, x_{ni} \geqslant 0\right\}$. □

Table 1
Calibration of cutoff parameters for a land use model

Parameter	Income group	MNL $\phi = 1$	CMNL $\phi \neq 1$
α_n	2	−3.329	−0.238**
	3	−8.130	−2.272**
	4	−14.228	−4.781
	5	−24.808	−10.257
ln(floor space)	2	−1.840	−0.012**
	3	−0.078*	0.493**
	4	−0.857	1.022
	5	0.346**	2.502
ln(zone income)	2	0.723*	
	3	1.075	
	4	2.013	
	5	2.442	
Accessibility	1	0.283**	0.492**
	2	1.636	1.692
	3	2.262	2.295
	4	3.926	3.966
	5	3.125*	3.377
δ_n	1		−63.769**
	2		−23.953
	3		−18.290
	4		−11.601**
	5		−0.069**
ω			0.242
Log-likelihood		−3.313	−3.316
Nr observations		600	600

Note: Estimates without asterix are significant (t-test>1.96), except when indicated by * (1.7<test-t<1.96) and by ** (t-test<1.7).

Table 2
A mode choice model with cutoffs (synthetic data set)

Parameter	MNL	CMNL
Car driver	−**2.0074**	−**2.3400**
	(−4.4)	(−5.3)
Car passenger	−**2.1220**	−**2.3857**
	(−6.2)	(−7.2)
Share taxi	−**1.4364**	−**1.5315**
	(−4.7)	(−5.1)
Subway	**2.4430**	**2.3027**
	(7.2)	(7,1)
Bus	**0.0000**	**0.0000**
Car driver – subway	−**1.9062**	−**1.9687**
	(−4.6)	(−5.0)
Car pass. – subway	−**1.2733**	−**1.2474**
	(−5.1)	(−5.1)
Shared taxi – subway	−**1.6940**	−**1.8113**
	(−5.1)	(−5.6)
Bus – subway	−**0.4267**	−**0.5179**
	(−1.6)	(−2.1)
Nr of license	**2.2374**	**2.2838**
	(5.3)	(5.6)
Sex	−**0.3108**	−**0.3179**
	(−1.4)	(−1.5)
Travel time	−**0.0846**	−**0.0848**
	(−4.8)	(−5.0)
Waiting time	−**0.162**	−**1.207**
	(−8.4)	(−8.3)
Walking time	−**0.2435**	−**0.2199**
	(−2.1)	(−1.9)
Total out-of-pocket/leisure time	−**0.0061**	−**0.0045**
	(−4.1)	(−3.0)
ω_k (walking time)	–	**0.1321**
		(−6.8)
Likelihood (q)	−943.86	−901.30
Likelihood (cte)	−1027.94	−1001.39
ρ^2	0.082	0.100
LR (constant)	168.16	200.18
Nr of observations	697	679

Theorem A2 (*Convergence of endogenous cutoffs fixed points*). *$\Delta = \left\{x \in \Re^{|N||I|} / \sum_{ni} x_{ni} = 1, x_{ni} \geqslant 0\right\}$ be the probability simplex and λ the maximum scale factor of the CMNL model. If the functions $V_{ni}(P), n \in N, i \in C$ are of class C^1, then there exists $\tilde{\lambda} > 0$ such that $\forall \lambda \in (0, \tilde{\lambda})$ the CMNL model has a unique fixed point solution over Δ, and the fixed point iteration converge to the solution.*

Proof. We will find $\tilde{\lambda} > 0$ such that, under theorem conditions on λ, the Jacobian of the logit function presented in Eq. (13) has norm less than one. This means that the function is contractive, so the application of the Banach fixed point theorem yields the results of existence, uniqueness, and convergence.

Let $M_{rl}^{ni} = \max_{P \in \Delta}\left\{\left|\frac{\partial V_{ni}^C}{\partial P_{rl}}\right|, \left|\frac{\partial Z_{ni}}{\partial P_{rl}}\right|\right\}$; we have that $M_{rl}^{ni} < \infty$, because functions $V_{ni}(P)$ are of class C^1, and so are functions $Z_{ni}(P)$.

Now we calculate the Jacobian ∞-norm and bound it strictly by 1. This is

$$\|J\|_\infty = \max_{mz} \sum_{ni} \left|\frac{\partial f_{ni}}{\partial P_{mz}}\right|$$

where f is the CMNL function presented in Eq. (13) and P is the CMNL probability. The f function is such that

$$\frac{\partial f_{ni}}{\partial P_{mz}} = P_{ni}\left\{\mu \frac{\partial V_{ni}^c}{\partial P_{mz}} + \sum_{l=1}^{K} w_k \frac{\partial Z_{li}}{\partial P_{mz}}(\phi_{nli}^{U} - \phi_{nli}^{L}) + \delta_z^i \sum_{l=1}^{J} \bar{w}_l y_{il}(\phi_{il}^{U} - \phi_{il}^{L}) \right.$$
$$\left. - \sum_{s \in C} P_{ns}\left(\mu \frac{\partial V_{ns}^c}{\partial P_{mz}} + \sum_{l=1}^{K} w_k \frac{\partial Z_{ls}}{\partial P_{mz}}(\phi_{nls}^{U} - \phi_{nls}^{L}) + \delta_z^s \sum_{l=1}^{J} \bar{w}_l y_{sl}(\phi_{sl}^{U} - \phi_{sl}^{L})\right)\right\}$$

where δ_z^i equals 1 if, and only if, $i = z$ and 0 otherwise; the rest of the notation follows the text.

Successive applications of the triangular inequality, the fact that $|\phi_{\cdot}^{U} - \phi_{\cdot}^{L}| \leqslant 1$, the strict positivity condition on the scale factors μ, w_k, w_j and the probabilities, and the M bounds, yield the following:

$$\left|\frac{\partial f_{ni}}{\partial P_{mz}}\right| \leqslant P_{ni}\left\{ \mu M_{mz}^{ni} + \sum_{l=1}^{K} w_k M_{mz}^{li} + \delta_z^i \sum_{l=1}^{J} \bar{w}_l |y_{il}| + \sum_{s\in C} P_{ns}\left(\mu M_{mz}^{ns} + \sum_{l=1}^{K} w_k M_{mz}^{ls} + \delta_z^s \sum_{l=1}^{J} \bar{w}_l |y_{sl}| \right) \right\}$$

Let $\lambda = max\{\mu; max_k w_k; max_j \bar{w}_j\}$ be the maximum dispersion parameters over the binomial and multinomial functions of the CMNL model. We have

$$\left|\frac{\partial f_{ni}}{\partial P_{mz}}\right| \leqslant \lambda P_{ni}\left\{ M_{mz}^{ni} + \sum_{l=1}^{K} M_{mz}^{li} + \delta_z^i \sum_{l=1}^{J} |y_{il}| + \sum_{s\in C} P_{ns}\left(M_{mz}^{ns} + \sum_{l=1}^{K} M_{mz}^{ls} + \delta_z^s \sum_{l=1}^{J} |y_{sl}| \right) \right\}$$

Then for the ∞-norm we can write:

$$\begin{aligned} \|J\|_\infty &\leqslant \lambda \max_{mz} \sum_{ni} P_{ni}\left\{ M_{mz}^{ni} + \sum_{l=1}^{K} M_{mz}^{li} + \delta_z^i \sum_{l=1}^{J} |y_{il}| + \sum_{s\in C} P_{ns}\left(M_{mz}^{ns} + \sum_{l=1}^{K} M_{mz}^{ls} + \delta_z^s \sum_{l=1}^{J} |y_{sl}| \right) \right\} \\ &= \lambda \max_{mz} \sum_{ni} P_{ni}\{\widetilde{M}_{mz}^{ni} + \max_{s\in C} \widetilde{M}_{mz}^{ns}\} \leqslant \lambda \max_{mz} \sum_{ni} \{\widetilde{M}_{mz}^{ni} + \max_{s\in C} \widetilde{M}_{mz}^{ns}\} \end{aligned}$$

where $\widetilde{M}_{rl}^{ni} = M_{mz}^{ni} + \sum_{l=1}^{K} M_{mz}^{li} + \delta_z^i \sum_{l=1}^{J} |y_{il}|$

Strictly bounding the above norm bounds by one and noting that all matrix norms are equivalent, we obtain that there exist $M > 0$ such that $\|J\| \leqslant \lambda M < 1$. Thus, defining $\tilde{\lambda} = 1/M$ and provided that $\lambda < \tilde{\lambda}$, f becomes a contractive function. Finally, an application of the Banach Theorem to f over Δ for any $\lambda \in (0, \tilde{\lambda})$ yields the proof. □

References

Ben-Akiva, M.E., Boccara, B., 1995. Discrete choice models with latent choice sets. International Journal of Research in Marketing 12 (1), 9–24.

Cantillo, V., Ortúzar, J.deD., 2005. A semi-compensatory discrete choice model with explicit attribute thresholds of perception. Transportation Research Part B 39 (7), 641–657.

Cascetta, E., Papola, A., 2001. Random utility models with implicit availability/perception of choice alternatives for the simulation of travel demand. Transportation Research Part C 9 (4), 249–263.

Domencic, T.A., McFadden, D., 1975. Urban Travel Demand: A Behavioral Approach. North-Holland.

Manski, C.F., 1977. The structure of random utility models. Theory and Decisions 8, 229–254.

Martínez, F., Hurtubia, R., 2006. Dynamic model for the simulation of equilibrium states in the land use market. Networks and Spatial Economics 6 (1), 55–72.

Morikawa, T., 1995. A hybrid probabilistic choice set model with compensatory and non-compensatory choice rules. In: Proceedings 7th World Conference on Transport Research. Travel Behaviour, vol. 1. Pergamon, Oxford.

Shelling, T., 1978. Micromotives and Macrobehavior. W.W. Norton and Co.

Swait, J.D., 2001. A non-compensatory choice model incorporating attribute cut-offs. Transportation Research Part B 35 (7), 903–928.

Swait, J.D., Ben-Akiva, M.E., 1987. Incorporating random constraints in discrete models of choice set generation. Transportation Research Part B 21 (2), 91–102.

Tversky, A., 1972. Elimination by aspects: a theory of choice. Psychological Review 79, 281–299.