The Constrained Multinomial Logit: a semi-compensatory choice model

Francisco Martínez[1], Felipe Aguila and Ricardo Hurtubia
University of Chile

ABSTRACT

The traditional formulation of logit models applied to transport demand assumes a compensatory (indirect) utility function, that is, the consumers strategy assumes trade-off between attributes. Several authors have criticized this approach because it fails to recognize attributes thresholds in consumers' behavior, or a more generic domain where such 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 CMNL model combines the multinomial logit model with bi-nomial logit factor that represent soft cutoffs. This approach extends previous contribution by allowing multiple dimensions for cutoff factors, but also introduce system constraints such as capacity and inter agents interactions (choice externalities). The analysis of this model includes a method to solve the non-linear fixed point problem that arises when system constraints are considered and a set of evaluation tools: a social utility of the constrained problem and a measure of the shadow price of each constraint.

Key words: discrete choice, logit models, constrained behavior, consumer's utilities.

[1] Corresponding author. POBox 228-2 Santiago, Chile. Phone: 562 9784380. Email: fmartine@ing.uchile.cl

INTRODUCTION

Following Domencich 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 underpinning assumption of these models is the compensatory strategy followed by individuals, i.e. their decision strategy assumes trade-off between attributes. This assumption has been criticized by several scientists who claim that a non-compensatory behavior is potentially more realistic, as for example the elimination-by-aspect (EBA) process (Tversky, 1972). A natural approach to relax the compensatory assumption, proposed by Manski (1977) and followed by Swait and Ben-Akiva (1987), Ben-Akiva and Boccara (1995), Cantillo and Ortúzar (2004), 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 appealing of this approach is that it permits different models to simulate the phenomena associated to 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 has been proposed to reduce this difficulty, as the pair wise comparisons of alternatives suggested by Morikawa (1995). However, the choice set formation process is not sufficiently efficient if the number of alternatives is large, like in the case of spatial choices (e.g. trip destination and location choices), and not applicable in more complex processes involving intensive choice making calculations, like equilibrium and optimization processes.

The model proposed by Cascetta and Papola (2001) extends the compensatory utility function in order to simulate (rather than generate) the perception/availability of an alternative implicitly, leading to a one-step approach named 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"[2] rather than "crisp".

Swait (2001) models choice behavior incorporating a wide range of decision strategies using an alternative 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 represents a feasible domain where the individual is willing -or can- make choices, with attributes 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 that are activated if cutoffs are violated.

As other implicit approaches, the IAP and Swait's models are a one-step method based on a "reduced" form model of behavior. The underpinning rationale is given by Swaits: "it is behaviorally equivalent whether the decision-maker simply chooses the best good that satisfies the constraints, or alternatively, first screens based on constraints, then chooses the best alternative". While we share this statement, Swait's model can be criticized because it introduces a linear relaxation to cutoffs, which means that at the cutoff the utility functions "kinks" (changes the slope) because the penalty is activated. Such kinks makes the utility function non-differentiable at the cutoff, introducing a difficulty in certain complex calculation processes, like equilibrium and optimization processes, or in systems with large number of agents making choices on choice-sets that

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

change their attributes in the process (like price adjustment to equilibrium); conversely, the IAP model may be specified to avoid such difficulty.

The constrained logit model (CLM) proposed in this paper combines aspects of Swait's 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 applies the bi-nomial logit to simulate soft cutoffs by a continuous and differentiable extended utility function. However, we simulate 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 study the more complex case of system constraints, where alternatives' 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 to represent the complex issue of externalities in consumption.

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

THE CONSTRAINED CHOICE PROBLEM

Consider the following class of optimization problems widely used in microeconomic theory to describe agent's 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 alternatives contained in the set C. Assume that the utility function depends on K-1 dimensional attributes set, denoted by vector $X \in R^{(K-1)\times I}$, and on the alternative price $p_m \in p$, with vector $p \in R^I$. We can define a set of attributes/prices cutoffs for the n^{th} agent, including a lower and an upper cutoff for each attribute/price k, denoted by a_{nk} and b_{nk} respectively, which dictates acceptable attribute/price values. Thus, consider the following vectors:

$$q_n^L = [a_{n1}, a_{n2}, \dots, a_{nK}], \quad q_n^U = [b_{n1}, b_{n2}, \dots, b_{nK}] \tag{1}$$

This defines the domain D_n for the individual's feasible choices, with the convention that parameters a_{nK} and b_{nK} are price bounds. Note that bounded parameters are assumed independent of the specific alternative, which is a usual case, but this can be extended to consider the case of alternatives' specific bounds.

Then, the rational choice behavior is modeled by the following optimization problem:

$$\begin{aligned} &\max_{d_{ni}} \sum_{i\in C} d_{ni} U_n(X_i, p_i) \\ &s.a \quad \sum_{i\in C} d_{ni} = 1, \quad d_{ni} \in \{0,1\} \quad \forall i \in C \\ &\qquad a_{nk} \le X_{ik} \le b_{nk} \quad \forall i \in C, k = \{1,\dots,K-1\} \\ &\qquad a_{nK} \le p_i \le b_{nK} \quad \forall i \in C \end{aligned} \tag{2}$$

where d_{ni} represents the individual's choice, X_i is the vector of attributes that describes alternative i, p_i is the price of the alternative and $U(X,p)$ is the indirect utility function. The problem maximizes the aggregated utility across the set of chosen alternatives, subject to the condition that constraints can not be violated in any chosen alternative. In the following we define vector $Z_i=(X_i, p_i) \in R^K$, which contains all attributes (including the price) of an alternative. Note that constraints are assumed specific to the choice maker; the case of constraints equal to all individual is a special and more simple case of problem (2).

It is noteworthy that problem (2) assumes that attributes are exogenous to the choice process; below we extend this problem assuming $X = X(d)$, named as endogenous constraints, which represent choice externalities that are relevant in forecasting demand.

THE MULTIDIMENSIONALLY CONSTRAINED UTILITY

Consider now the classical model where the behavior function is a random variable, that is $U_n = \overline{V}_n + e_n$, with $\overline{V}_n$ a systematic compensatory utility and e_n the random term. Then, individual's[3] choices are represented by the probabilities associated with the distribution of e_n across alternatives in C. The widely used logit model is derived upon assuming that random terms are distributed Gumbel, which implies that $e \in [-\infty, \infty]$, then utilities are unconstrained.

In order to restrain behavior to the individual's feasible set, our method defines a "constrained utility" function that induce the individual to make choices that belong to her feasible domain D_n with 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 to some extent, such that the probability of choosing an alternative out of D_n is limited to a maximum $h = \{h_k, k = 1,..., K\}$. Additionally, the constrained utility function is assumed 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 multi dimensions domain, our approach is similar to the IAP model, because we also augment the usual compensatory utility function by a new cutoff term, called utility penalty. Thus, a compensatory term (Vc) and an additive cutoff term define the constrained utility, as follows:

$$V_n(Z_i) = V_n^C(Z_i) + \frac{1}{m}\ln f_{ni}(Z_i) + e_{ni} \qquad (3)$$

with e assumed Gumbel distributed $(0, m)$. Notice that the cutoff term is amplified by the Gumbel scale parameter, which increases the penalty as the utility dispersion increases. Thus, the cutoff term may be understood as a displacement of the systematic utility term, or the utility penalty, such that the resulting choice probability complies with the cutoff constraint with some given probability h.[4]

The penalty term contains the generalized cutoff factor f_{ni}, which is defined as a composite factor of the set of elementary attributes/price, lower and upper, cutoffs by $f_{ni} = \prod_{k=1}^{K} f_{nki}^{L} \cdot f_{nki}^{U}$. Each elementary cutoff factor is defined as a binomial logit function because it is an interesting and useful example: it is simple for the presentation of the model and it has been similarly used by Swait and Ben-Akiva (1987), Ben_Akiva and Boccará (1995) and Cascetta and Papola (2001), but more importantly, it provides some relevant properties when the model is used in studies with endogenous constraints, as shown below. Then:

$$f_{nki}^{L} = \frac{1}{1 + \exp\left(w_k\left(a_{nk} - Z_{ki} + r_k\right)\right)} = \begin{cases} 1 & if \quad \left(a_{nk} - Z_{ki}\right) \to -\infty \\ h_k & if \quad a_{nk} = Z_{ki} \end{cases} \qquad (4a)$$

---

[3] We use individual as a general expression for the decision maker, which more generally may be defined as an economic agent because we include institutions and companies.

[4] Cascetta and Papola (2001) propose a similar utility penalty but without the $1/m$ factor.

$$f_{nki}^{U} = \frac{1}{1+\exp\left(w_k\left(Z_{ki} - b_{nk} + r_k\right)\right)} = \begin{cases} 1 & if \ \left(b_{nk} - Z_{ki}\right) \to \infty \\ h_k & if \ \ b_{nk} = Z_{ki} \end{cases} \tag{4b}$$

that represents the elementary lower and upper cutoff factors. Additionally, we define:

$$r_k = \frac{1}{w_k} \cdot \ln\left(\frac{1-h_k}{h_k}\right) \tag{4c}$$

The performance of other functions may be explored, for example Cascetta and Papola (2001) also analyze the Gamma distribution for the single (not composite) cutoff factor.

Observe that the generalized cutoff factor is (quasi) innocuous for any feasible alternatives, i.e. those with $Z_i \in D_n$, because $f_{ni} \to 1$; conversely, if any element $Z_{ki} \notin D_n$ then $f_{ni} \to 0$ and the choice probability also tends to zero for this alternative performing a soft compliance of the constraint. Also note that each elementary cutoff factor in equations (4) may be interpreted as binomial choice with two alternatives: respect or violate the specific cutoff. The parameter w represents the scale factor of the binomial logit function that measures the behavior dispersion regarding violation of cutoffs. Figure 1 depicts the binomial – lower and upper – cutoff functions, and Figure 2 shows that the parameter w controls the softness of the cutoff. The other parameters are the cutoff tolerance, with r_k defined in the same units as the k^{th} variable and h defined as a choice probability tolerance. This tolerance can be as small as desired but not zero, implying that the model can not be applied for deterministic compliance of cutoffs; some degree of tolerance is structurally imposed. For simplicity in the presentation h is specified constant for all agents, but an individual specific constant is also possible.

Figure 1: The lower and upper binomial cutoff functions

f^{U}_{nki} **and** f^{L}_{nki} **vs** Z_{ki}

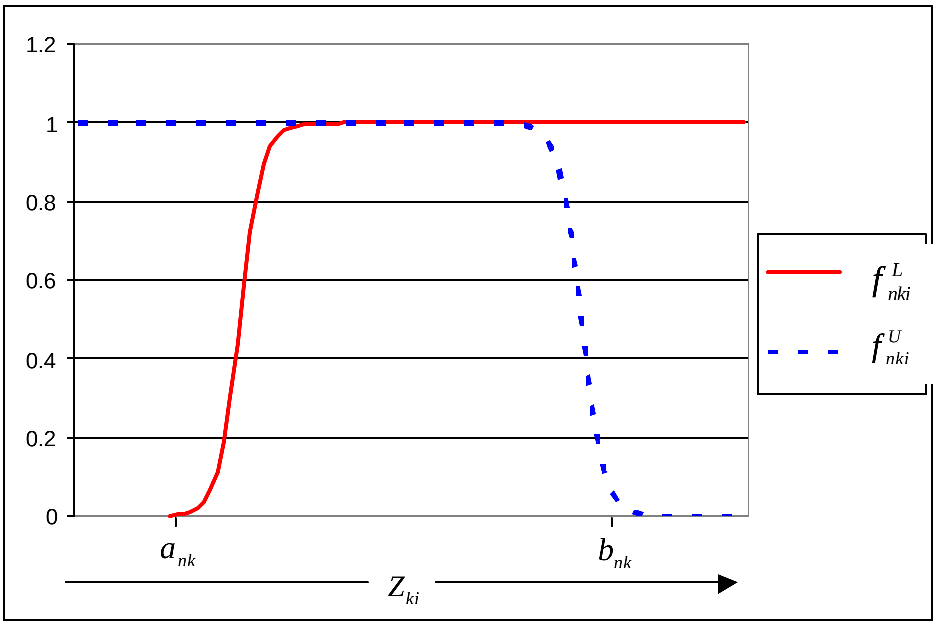
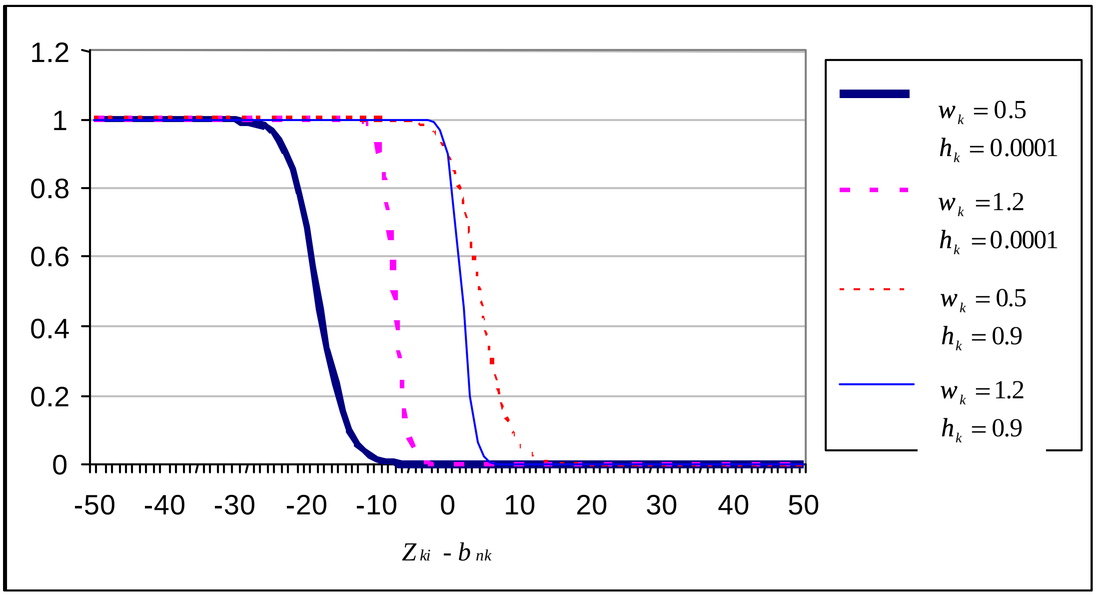


Figure 2: The effect of the scale parameter on upper cutoff factors



Also for simplicity we have considered only one pair of cutoffs (lower and upper) per attribute, but our method can include more cutoffs to represent the combined effect of more or less binding constraints. Note that in this case of multi-cutoffs for a given attribute (either upper or lower bound), the deterministic approach would eliminate all but the most binding cutoff, because the rest are zero by definition. However, in our stochastic approach even not binding cutoffs have some effect on choices, although the most violated have a larger effect. For example, consider the case of an alternative with a price close to the self-imposed maximum expenditure, which defines a first cutoff; the second, less binding, but stricter cutoff is the individual's income. The utility will tend to be reduced primarily by the fist cutoff, but some extra reduction is produced by the second cutoff; these combined effects seems to be plausible in the real context. The method may also consider cutoffs defined by a combination of attributes.

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

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

then

$$\ln f_{ni} = -\sum_{k}^{K}\left(\ln\left[1+\exp w(a_{nk} - Z_{ki} + r_k)\right] + \ln\left[1+\exp w(Z_{ki} - b_{nk} + r_k)\right]\right) \qquad (5)$$

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 are farther out the domain. Another relevant difference is that Swaits' linear penalty yields a continuous utility function but not differentiable at cutoff values of attributes. Conversely, an 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 raises 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 as capacity. Then it is natural to assume that the behavior associated to cutoffs has a random nature.

The cutoff tolerance parameter q may be understood in a dynamic context, because the tolerance to accept penalties may be specified as dependent on previous experiences, in a way that those individuals that had chosen alternatives in the vicinity of the domain limit have a better knowledge of the penalty and the benefits. Thus we can postulate that their tolerance is different compared to similar individuals in all aspects except from their experience.

THE CONSTRAINED MULTINOMIAL LOGIT MODEL (CMNL)

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

$$\underset{i \in C}{Max}\, \tilde{V}_n(Z_i) = V_n^C(Z_i) + \frac{1}{m}\ln(f_n(Z_i)) + e_{ni} \qquad (6)$$

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

$$P_{ni} = \text{Prob}\left[V_{ni}^{C} + \frac{1}{m}\ln f_{ni} + e_{ni} \geq \max_{j \in C}\left(V_{nj}^{C} + \frac{1}{m}\ln f_{nj} + e_{nj}\right)\right] \tag{7}$$

The assumption that the constrained utility is distributed identical and independent Gumbel yields the following multinomial probability function:

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

This expression represents a choice probability that complies with the feasible domain D_n, which tends asymptotically to zero if any of the alternative attributes violates any cutoff. At the very cutoff, the usual compensatory probability is multiplied by tolerance probability factors h s. This model, named Constrained Multinomial Logit (CMNL), preserves the closed expressions of the equivalent classical compensatory logit models.

Notice that the multinomial model in equation (8) can be formally derived from a joint multinomial choice model, where each upper level alternative is conditioned by a lower set of binomial models that checks if the alternative belongs to the domain.

It is now important to make same comments regarding calibration methods. First, as in the unconstrained multinomial logit model, there is no need to calibrate the parameter m, because it is embedded in the parameters of compensatory utility V^C and does not affect cutoff values. Secondly, the application of usual techniques, e.g. the maximum likelihood procedure or the least squares, in the context of parameters of the cutoff functions may not be always direct. In their usual application, the parameters adjust the model to reproduce observed choices, while cutoff parameters are associated to barely observed behavior because since they represent choices theoretically unfeasible. Nevertheless, Cascetta and Papola (2001) apply the maximum likelihood method to obtain cutoff parameters obtaining a "better" model than the unconstrained, reporting highly significant coefficients for parameters associated with cutoff variables.

Since the data required for calibrating cutoff parameters is very specific, reflecting the decision maker behavior at each edge of the choice domain, we argue that stated preferences (SP) data, specially reporting choice answers at the cutoff vicinity, is more adequate to make a consistent application of traditional calibration methods than revealed preferences (RP).

FORECASTING ISSUES

Individual choices are usually also constrained by two types of system constraints that do not affect the model calibration but have a crucial effect on forecasting choices because the total demand for alternatives is constrained. We differentiate between two system constraints according to their role in the demand model. Type I constraints occur by the saturation of the infrastructure capacity –exogenous constraints–, namely road and public transport capacity, land space, and numerous policy regulations. Type II constraints are individuals' thresholds on attributes –endogenous constraints–, with attributes defined by the outcome of all other consumers' choices, for example: neighborhood quality in residential location choice when quality is defined, for instance, by socioeconomic, racial or religious condition of neighbors; in-vehicle congestion in transport choice. In economic terminology, Type II are consumption externalities that reproduce fundamental, real and complex effects in urban markets.

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

$$\overline{a}_{ij}^{L} \leq \sum_{n} y_{ij} P_{ni} \leq \overline{b}_{ij}^{U} \tag{9}$$

where y_{ij} are exognous parameters that define the amount of the scarce resource j used if alternative i is chosen; P_{ni} is the n^{th}'s consumer probability of choosing alternative i; $\overline{a}_{ij}, \overline{b}_{ij}$ are the lower and upper system constraints for the j^{th} capacity in alternative i.

To introduce system constraints Type I 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:

$$\overline{q}_{i}^{L} = [\overline{a}_{i1}, \overline{a}_{i2}, \quad \ldots \quad \overline{a}_{iJ}], \quad \overline{q}_{i}^{U} = [\overline{b}_{i1}, \overline{b}_{i2}, \quad \ldots \quad \overline{b}_{iJ}] \tag{10}$$

which define the alternative's sub-domain $\overline{D}_i$. We also define the aggregated demand for resources j generated by alternative i, given by $\overline{Y}_{ij}(P) = \sum_{n} y_{ij} P_{ni}$.

The constrained utility function (3) is further augmented by penalties of violating the system constraints, yielding:

$$\overline{V}_{n}(Z_i) = V_n^C(Z_i) + \frac{1}{m} \ln f_{ni}(Z_i) + \frac{1}{m} \ln \Phi_i(P_i) + e_{ni} \tag{11}$$

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, $\Phi_i = \prod_{j=1}^{J} \Phi_{ij}^{L} \cdot \Phi_{ij}^{U}$ with each elemental term defined by:

$$\Phi_{ij}^{L}(P_i) = \frac{1}{1+\exp\left(\overline{w}_j\left(a_{ij} - \overline{Y}_{ij}(P_i) + \overline{r}_j\right)\right)} = \begin{cases} 1 & if \ \left(a_{ij} - \overline{Y}_i\right) \to -\infty \\ \overline{h}_j & if \ \overline{a}_{ij} = \overline{Y}_{ij} \end{cases} \tag{12a}$$

$$\Phi_{ij}^{U}(P_i) = \frac{1}{1+\exp\left(\overline{w}_j\left(\overline{Y}_{ij}(P_i) - \overline{b}_{ij} + \overline{r}_j\right)\right)} = \begin{cases} 1 & if \ \left(\overline{b}_{ij} - \overline{Y}_{ij}\right) \to \infty \\ \overline{h}_j & if \ \overline{b}_{ij} = \overline{Y}_{ij} \end{cases} \tag{12b}$$

with

$$\overline{r}_j = \frac{1}{w_j} \cdot \ln\left(\frac{1-\overline{h}_j}{\overline{h}_j}\right) \tag{12c}$$

The Type II system constraints naturally reproduce consumption externalities because they introduce interdependencies in consumption between consumer agents. These externalities may affect utilities through change in prices (pecuniary externalities) or directly changing attributes (technological externalities). The CMNL model represents these externalties by making $Z = Z(P)$, then the system constraint is represented by an endogenous cutoffs on these attributes. This type of constraint is well modeled by the cutoff factors already defined in equations (4).

The combination of all constraints restricts individual choices probability to the domain $\widetilde{D}_n = D_n \bigcap_{j}^{J} D_j$, which is defined by the augmented constraints vector $\widetilde{q}_n = q_n^L \cup q_n^U \cup \overline{q}^L \cup \overline{q}^U ; \widetilde{q}_n \in R^{2(K+I)}$. Then, the CMNL model (equation 8) can be extended to recognize system externalities as follows:

$$\widetilde{P}_{ni} = \frac{\widetilde{f}_{ni}(P_i) \cdot \exp(mV_{ni}^{C}(P_i))}{\sum_{j \in C} \widetilde{f}_{nj}(P_j) \cdot \exp(mV_{nj}^{C}(Pj))} \tag{13}$$

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

Notice that 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. Such interdependency raises numerous issues on calibration process, which are beyond the scope of this paper, but it also raises the issue of the complexity associated to the forecasting process.

The rest of this section examines the complexity issue in forecasting demand. Observe that equation (14) represents a fixed point problem $P = f(P)$, a system of $I \cdot N$ non-linear equations. 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. $$\frac{1}{l} \leq 2 \cdot \max_{mz} \left\{ \sum_{ni} \left| \frac{\partial V_{ni}^c}{\partial P_{mz}} \right| + |n| \cdot \left(\sum_{l=1}^{J} |y_{zl}| + \sum_{i} \sum_{l=1}^{K} \left| \frac{\partial Z_{li}}{\partial P_{mz}} \right| \right) \right\}$$

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

where $l = \max\{w; m\}$ is the maximum value between dispersion parameters of the binomial and multinomial functions.

Proof: See appendix.

These conditions are obtained by imposing on *f(P)* that the Banach theorem's is satisfied and that its jacobian be less than 1, which is sufficient –although not necessarycondition for contractiveness. For each condition we use the co-matrix ∞-norm and 1-norm, respectively. The convergence result requires that a minimum dispersion be present on individuals' choice behavior, that is: if the choice process is close to deterministic the convergence conditions are not guaranteed. In the theorem such condition imposes minimum values for the dispersion parameters w and m of the binomial and multinomial functions respectively.

Observe that if the number of alternatives is very large, probabilities tend to be small. Thus, in large problems local conditions are normally satisfied, which means that the dispersion parameters are likely to satisfy the bounds. In our extensive simulation exercises, with small and large problems, we have obtained a high convergence performance considering the complexity of the non-linear system of equations (14).

Nevertheless, the conditions in the above theorem are very strong, because they are sufficient condition for constractiveness of the multinomial logit function over the whole domain. In fact, violation of these bounds does not necessarily imply lack of constractiveness. Thus, a practical use of the bounds for l in operational models is as control parameters, to check if the conditions are satisfied at each iteration of the solution algorithm; if it is not, a warning flag should be raised and the following iterations must be analyzed to check if the flag has turned off and the contractiveness conditions are recovered.

The theorem constitute a fundamental advantageous property of the CMNL model for its applications to forecast urban systems performance. 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 results most likely violate constraints and miss-calculate utilities. The theorem assures that the fixed point algorithm converge to the unique solution under certain (normally satisfied) conditions. The theorem may be extended to other logit structures, for example the Nested and Mixed Logit, which also remains for further research.

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, defined as the maximum expected individuals utilities aggregated across the population. The second one measures the social cost of policies that constrains consumption (e.g. capacities and regulations), measured as the shadow price of each elemental constraint.

Consider the CMNL utility, equation (11), evaluated at the demand solution, 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 the subset $\tilde{D}_n$, which is given by the following logsum formula:

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

This equation measures the individual's maximum expected benefit obtained from the choice-set *C*, which we use to analyze the impact of urban policies on individuals' satisfaction. An aggregate utility across N consumers associated to the alternatives set C is:

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

which represents the utilitarian social measure of the consumers' benefits; this measure ignores distribution issues[5]. Notice that the domain of this social utility function is $\tilde{D}_C = \bigcup_n^N \tilde{D}_n$ defined by the augmented vector $\tilde{q}_C = \bigcup_n \tilde{q}_n$; $\tilde{q}_C \in R^{(2K+I)\cdot N}$. Notice also that the parameter m is normally unknown in applied MNL models, because it is theoretically embedded in the parameters calibrated for of compensated utility $\hat{V}^C = m\tilde{V}$, then in this case the parameter m can be correctly assumed equal to one. Equation (15) provides a measure of the social benefit yield by given urban system, which can be used for evaluating different scenarios of the urban system, for example to evaluate land regulations in location choice process and transport policies that affect demand of specific transport modes.

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

$$S_j = \frac{\partial U_C}{\partial \tilde{q}_j} = \frac{1}{m} \sum_n^N \sum_{i \in C} \tilde{P}_{ni} \left[\sum_{l \in L} w_l \frac{(1-\tilde{f}_{nl})}{\tilde{f}_{nl}} \frac{\partial \tilde{f}_{nl}}{\partial \tilde{q}_j} \right] \quad (16)$$

Again, in applied studies the scale parameter m is assumed equal to one.

This result shows that the shadow price is strictly non-negative and increases as demand for alternatives close to the edge of the domain increases, because cutoff factors tends to zero so the term in parenthesis and S_j are strictly positive in that case. Conversely, if the

[5] Distributions with different equity criteria can be introduced by adding differentiated social values for consumers' benefits.

choice pattern is sufficiently far from the cutoff in the interior of the domain, shadow prices tend to zero, which is consistent with the theoretically expected shadow price for not binding cutoffs.

The terms in brackets recognize that our model includes multiple constraints – individual thresholds and system capacities- 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 cutoff is likely to occur. Think for example on the effects of increasing the level of the individual's acceptance of travel time by car, then more users are expected to show up in congested roads, thus increasing the level of congestion and, therefore, increasing the shadow price of road capacity constraints. Another example is in land use, where a stronger zone regulation, like the minimum density, induce several effects on land values and location patterns, which may activate residents thresholds on neighbor environment.

APPLICATIONS

The application of constrained logit models covers the whole range of discrete choice processes in economic systems, 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, usual cutoffs are budget and time resources, which are assumed exogenous in the context of transport decisions. Examples of endogenous cutoffs are associated with several attributes: minimum activity level at destination for attracting trips, maximum spent on travel, waiting and access times to public transport. Another cutoff is the maximum walking limit, which may be taken as exogenous for handicapped and elderly, or as endogenous for other travelers. In vehicle route assignment models, road and vehicles capacities are exogenous cutoffs, while time at traffic jams is endogenous.

In location and land use modeling cutoffs are particularly relevant. Real estate attributes are usually numerous, then attribute thresholds are also numerous and diverse. Relevant location options is another interesting case, because agents are likely to have cognitive constraints to evaluate 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. The non-negative profit constraint for the real state production model is an economic reasonable 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.

CONCLUSIONS

Advances in discrete choice modeling has not stopped in the last three decades, but challenges to replicate the actual behavior of agents are still very open. 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.

This paper proposes a method that builds upon previous techniques to make 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, contrarily, it enhanced any function in its domain border.

Our method was applied to multinomial logit models and has the following characteristics. Physical and economical constraints (called exogenous) and attributes thresholds (endogenous constraints) are modeled as soft cutoffs controlled by a stochastic compliance tolerance. 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 results is the constrained multinomial logit model (CMNL) that 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. The paper also analyses the use of the model for the forecasting application, because several cutoffs introduce extra complexity in solving the model to find the demand. The solution problem has a fixed point whose existence and uniqueness is proved; we also prove that fixed-point iteration converges to the solution. Our empirical tests show that convergence is highly efficient for the complexity of the non-linear equations involved.

The CMNL model provides an enhanced application of the random utility model for discrete choice modeling, which constrains utility to a more realistic domain yielding also more realistic choice probabilities. The model also produces two evaluation results. One is a social benefit measure for constrained setting and the other one is the shadow price for each cutoff. These are useful tool for the economic evaluation of policies affecting perceptions of attribute cutoffs (for example by education champagnes) or system capacities.

AKNOWLEDGEMENTS

The author thanks the collaboration and fruitful comments from R. Henríquez. This research was partially funded by FONDECYT 1010422 and Milenio P01-34 projects.

REFERENCES

BEN AKIVA, M. E. and BOCCARA, B. (1995) Discrete choice models with latent choice sets. *International Journal of Research in Marketing* 12, 9-24.

CANTILLO, V. and ORTÚZAR, J. de D. (2005) A semi-compensatory discrete choice model with explicit attribute thresholds of perception. *Transportation Research B, 39-7*, 641-657.

CASCETTA, E. and PAPOLA, A. (2001) Random utility models with implicit availability/perception of choice alternatives for the simulation of travel demand. *Transportation Research C 9*, 249-263.

DOMENCIC, T.A. and Mc FADDEN, 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.

MORIKAWA, T. (1995) A hybrid probabilistic choice set model with compensatory and non-compensatory choice rules. *Proceedings 7th World Conference on Transport Research*, Volume 1, Travel Behaviour, Pergamon, Oxford.

SWAIT, J.D. (2001) A non-compensatory choice model incorporating attribute cut-offs. *Transportation Research* 35B, 903-928.

SWAIT, J.D. and BEN AKIVA, M.E. (1987). Incorporating random constraints in discrete models of choice set generation. *Transportation Research* 21B, 91-102.

TVERSKY, A. (1972) Elimination by aspects: a theory of choice. *Psychological Review* 79, 281-299.

APPENDIX

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

THEROEM 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.

THEROEM A2:(*Convergence of endogenous cutoffs fixed points)*:

The CMNL model has a unique fixed point solution, and the fixed point iteration converge to the solution if one of the following conditions holds:

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

where $l = max\{m; max_k w_k; max_j \overline{w}_j\}$ is the maximum scale factor over utility and cutoff distributions.

Proof:

If some of the conditions holds, then the jacobian of the logit function presented in equation (13) has norm less than one. On the first condition it is true for the matrix ∞-norm and on the second for the matrix 1-norm. This means that the function is contractive and an application of the Banach fixed point theorem yields the result.

Now we calculate the two jacobian norms to obtain these conditions. The two jacobian norms are the following:

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

$$\|J\|_1 = \max_{ni} \sum_{mz} \left| \frac{\partial f_{ni}}{\partial P_{mz}} \right|$$

where f is the MNL function presented in equation (13) and P is the MNL probability. The f function is such that:

$$\frac{\partial f_{ni}}{\partial P_{mz}} = P_{ni} \left\{ \begin{array}{l} m \dfrac{\partial V_{ni}^c}{\partial P_{mz}} + \sum_{l=1}^{K} w_k \dfrac{\partial Z_{li}}{\partial P_{mz}} \left(f_{nli}^U - f_{nli}^L \right) + d_z^i \sum_{l=1}^{J} \overline{w}_l y_{il} \left(f_{il}^U - f_{il}^L \right) \\ - \sum_{s \in C} P_{ns} \left(m \dfrac{\partial V_{ns}^c}{\partial P_{mz}} + \sum_{l=1}^{K} w_k \dfrac{\partial Z_{ls}}{\partial P_{mz}} \left(f_{nls}^U - f_{nls}^L \right) + d_z^s \sum_{l=1}^{J} \overline{w}_l y_{sl} \left(f_{sl}^U - f_{sl}^L \right) \right) \end{array} \right\}$$

where d_z^i equals 1 iff i=z and 0 otherwise.

Successive applications of the triangular inequality, the fact that $\left| f_\bullet^U - f_\bullet^L \right| \leq 1$ and the strict positivity of the scale factors m, w_k ,w_j and the probabilities, yields the following:

$$\left| \frac{\partial f_{ni}}{\partial P_{mz}} \right| \leq P_{ni} \left\{ \begin{array}{l} m \left| \dfrac{\partial V_{ni}^c}{\partial P_{mz}} \right| + \sum_{l=1}^{K} w_k \left| \dfrac{\partial Z_{li}}{\partial P_{mz}} \right| + d_z^i \sum_{l=1}^{J} \overline{w}_l \left| y_{il} \right| \\ + \sum_{s \in C} P_{ns} \left(m \left| \dfrac{\partial V_{ns}^c}{\partial P_{mz}} \right| + \sum_{l=1}^{K} w_k \left| \dfrac{\partial Z_{ls}}{\partial P_{mz}} \right| + d_z^s \sum_{l=1}^{J} \overline{w}_l \left| y_{sl} \right| \right) \end{array} \right\}$$

Let $l = \max\{m; \max_k w_k; \max_j \overline{w}_j\}$ be the maximum dispersion parameters over the binomial and multinomial functions. We have:

$$\left| \frac{\partial f_{ni}}{\partial P_{mz}} \right| \leq l P_{ni} \left\{ \begin{array}{l} \left| \dfrac{\partial V_{ni}^c}{\partial P_{mz}} \right| + \sum_{l=1}^{K} \left| \dfrac{\partial Z_{li}}{\partial P_{mz}} \right| + d_z^i \sum_{l=1}^{J} \left| y_{il} \right| \\ + \sum_{s \in C} P_{ns} \left(\left| \dfrac{\partial V_{ns}^c}{\partial P_{mz}} \right| + \sum_{l=1}^{K} \left| \dfrac{\partial Z_{ls}}{\partial P_{mz}} \right| + d_z^s \sum_{l=1}^{J} \left| y_{sl} \right| \right) \end{array} \right\}$$

Then for the ∞-norm we can write:

$$\|J\|_{\infty} \le l \max_{mz} \sum_{ni} P_{ni} \left\{ \begin{array}{l} \left|\frac{\partial V_{ni}^{c}}{\partial P_{mz}}\right| + \sum_{l=1}^{K} \left|\frac{\partial Z_{li}}{\partial P_{mz}}\right| + d_{z}^{i} \sum_{l=1}^{J} |y_{il}| \\ \quad + \sum_{s \in C} P_{ns} \left(\left|\frac{\partial V_{ns}^{c}}{\partial P_{mz}}\right| + \sum_{l=1}^{K} \left|\frac{\partial Z_{ls}}{\partial P_{mz}}\right| + d_{z}^{s} \sum_{l=1}^{J} |y_{sl}| \right) \end{array} \right\}$$

$$= 2\, l \max_{mz} \sum_{ni} P_{ni} \left\{ \left|\frac{\partial V_{ni}^{c}}{\partial P_{mz}}\right| + \sum_{l=1}^{K} \left|\frac{\partial Z_{li}}{\partial P_{mz}}\right| + d_{z}^{i} \sum_{l=1}^{J} |y_{il}| \right\}$$

$$\le 2\, l \max_{mz} \left\{ \sum_{ni} \left|\frac{\partial V_{ni}^{c}}{\partial P_{mz}}\right| + |n| \cdot \left(\sum_{l=1}^{J} |y_{zl}| + \sum_{i} \sum_{l=1}^{K} \left|\frac{\partial Z_{li}}{\partial P_{mz}}\right| \right) \right\}$$

and for the 1-norm:

$$\|J\|_{1} \le l \max_{ni} \left[P_{ni} \sum_{mz} \left\{ \begin{array}{l} \left|\frac{\partial V_{ni}^{c}}{\partial P_{mz}}\right| + \sum_{l=1}^{K} \left|\frac{\partial Z_{li}}{\partial P_{mz}}\right| + d_{z}^{i} \sum_{l=1}^{J} |y_{il}| \\ \quad + \sum_{s \in C} P_{ns} \left(\left|\frac{\partial V_{ns}^{c}}{\partial P_{mz}}\right| + \sum_{l=1}^{K} \left|\frac{\partial Z_{ls}}{\partial P_{mz}}\right| + d_{z}^{s} \sum_{l=1}^{J} |y_{sl}| \right) \end{array} \right\} \right]$$

$$\le l \left(\begin{array}{l} |n| \cdot \sum_{z} \sum_{l=1}^{J} |y_{zl}| + \sum_{mzs} \sum_{l=1}^{K} \left|\frac{\partial Z_{ls}}{\partial P_{mz}}\right| \\ + \max_{ni} \left[\sum_{mz} \left\{ \left|\frac{\partial V_{ni}^{c}}{\partial P_{mz}}\right| + \sum_{s \in C} \left|\frac{\partial V_{ns}^{c}}{\partial P_{mz}}\right| \right\} + \sum_{mz} \sum_{l=1}^{K} \left|\frac{\partial Z_{li}}{\partial P_{mz}}\right| + |n| \cdot \sum_{l=1}^{J} |y_{il}| \right] \end{array} \right)$$

By strictly bounding the above two norm bounds we get the theorem conditions and the proof.