Probit and nested logit models based on fuzzy measure

M. Aggarwal

1 Information Systems Area, IIM Ahmedabad, India
manish@iima.ac.in

Abstract
Inspired by the interactive discrete choice logit models [Aggarwal, 2019], this paper presents the advanced families of discrete choice models, such as nested logit, mixed logit, and probit models to consider the interaction among the attributes. Besides the DM’s attitudinal character is also taken into consideration in the computation of choice probabilities. The proposed choice models make use of Choquet integral and the recent attitudinal Choquet integral. The models are termed as Choquet nested logit (CNL), Choquet mixed multinomial logit (CMMNL), and Choquet multinomial probit (CMP). These are further extended to represent the DM’s attitudinal character, and termed as attitudinal CNL, attitudinal CMMNL, and attitudinal CMP.

Keywords: Choice models, probabilistic, decision-making, attitudinal, attributes interaction.

1 Introduction
Multinomial logit (MNL) [60] model is by far the most widely used discrete choice model. It postulates that each alternative can be seen as a bundle of attributes desired by a decision-maker (DM) who makes choices among various alternatives so as to maximize his utility by choosing the alternative whose attributes collectively yield more utility than any of the other alternatives in the given choice set.

The popularity of the discrete choice models can be gauged through their applications in diverse domains in the recent times. They are applied in severity analysis [14, 41, 68], price optimization [20], revenue optimization [25], location planning [30, 57], choice analysis problems [12, 57, 67], risk analysis [11, 67], demand analysis [11, 24], data analytics [27, 13, 6], regression analysis [21, 19, 59], causal inference in medicine [37, 34, 5], and forecasting [33], to name a few.

The simplicity, and intuitiveness of MNL model have inspired many researchers to develop its extensions such as nested logit (NL) [63, 20], GEV [61], multinomial probit (MNP) [10], paired combinatorial logit (PCL) [17, 58], cross-nested logit (CNL) [65], continuous CNL, generalized nested logit (GNL) [67], generalized MNL (GenMNL) [67], mixed multinomial logit (MMNL) [52] and fuzzy integral MNL [62] models. Recently, MNL model has been extended in [3] to develop Choquet MNL (CMNL) model that takes into consideration the interaction among the attributes. The CMNL model is further extended to consider the DM’s attitude in attitudinal CMNL (ACMNL).

1.1 Motivation
Though the proposed models undoubtedly lead to better modelling of a DM’s choices, they still suffer from the inherent drawbacks of the conventional MNL model, which are as follows:

- MNL considers the unobservable utility component as independent and identically distributed (i i d) extreme random variable for all i, with a Gumbel distribution. This hardly stands true in practice.
• Also, due to this, it suffers from undesirable IIA (independence from irrelevant alternatives) property. That is, the ratio of choice probabilities for any two alternatives $a_i$ and $a_j$ does not depend on any alternatives other than $a_i$ and $a_j$. Hence, the relative odds of choosing $a_i$ over $a_j$ remains the same regardless of the other alternatives. This is uninintuitive in most of the real world situations.

• Non-linear utility: It is very difficult to model the utility through a non-linear relationship. This restricts the modelling capabilities of MNL models.

The advanced models such as nested logit, mixed logit, probit, generalized MNL, etc. address these limitations of MNL model. However, unlike CMNL and ACMNL models, they are unable to consider the degree of interaction among the attributes, or the agent’s attitude, or both. In this paper, we are motivated to extend state-of-the-art models such as nested logit, mixed logit, and others to take into account the degree of interaction, and the agent’s attitude at the same time.

1.2 The proposed work

Recently, research in aggregation operators has evoked a great interest, as noted from many recent operators [11, 12, 1, 13, 14, 15, 16, 17]. The popular studies such as [8] and Grabisch [29] in the area of aggregation operators have played a seminal role in highlighting the crucial role of the aggregation operators in various real world applications. The aggregation operators hold significant potential in having the choice models more closely portraying the reality. With this motivation, we make use of the Choquet integral (see [15, 28]), and the recent attitudinal Choquet integral [1] to develop new choice models with interactive attributes. The proposed models represent the situations where different DMs, who may be having the same utility values for the given attributes, would still have the different choices, on account of their different perceived degree of interaction among the attributes, or their individualistic attitudes, or both. The paper is organized as follow: Section II builds the background for the paper. An interactive nested logit model is introduced in Section III. Section IV presents interactive probit models. Section V is dedicated to the general properties of the proposed interactive models. Using these properites, Section VI extends the proposed interactive logit and probit models as interactive mixed logit and interactive mixed probit models. Section VII concludes the paper.

2 Background

In this section, we give a brief discussion of state-of-the-art choice models. We shall begin with the MNL model, followed by its recent variants that consider the DM’s attitude and interaction among the attributes.

2.1 Multinomial logit model

Let us consider a MCDM situation with a choice set of $K$ alternatives $A = (a_1, \ldots, a_K)$, each of which is defined by a set of $M$ attributes $(c_1, \ldots, c_M)$. The set of values that $a_i$ takes is shown as:

\[ a_i = (a_i^{(1)}, \ldots, a_i^{(M)}) \]

where each of the $a_i^{(i)}$, $i = 1, \ldots, M$, refer to the normalized values in the range of 0 to 1. Each of these attribute values holds a different utility that is modelled by the vector:

\[ \beta = (\beta^{(1)}, \ldots, \beta^{(M)}) \]

where $\beta^{(i)}$ refers to the utility coefficient. $\beta^{(i)} > 0$ for desirable attributes, and $\beta^{(i)} < 0$ for non-desirable attributes. The individual utility corresponding to the attribute value $a_i^{(m)}$ is modelled as :

\[ v_i^{(m)} = \beta^{(m)} a_i^{(m)} \]  \hspace{1cm} (1)

and the net utility by virtue of the observable attributes, commonly referred to as representative utility, is computed through a scalar product of the vectors $a_i$ and $\beta$, shown as:

\[ V_i = \sum_{m=1}^{M} v_i^{(m)} = \sum_{m=1}^{M} \beta^{(m)} a_i^{(m)} \]  \hspace{1cm} (2)
MNL model postulates that a DM derives from an alternative $A_i$ a utility value $U_i$, given as:

$$U_i = V_i + \epsilon_i$$

where, $\epsilon_i$ is the unestimated utility.

Since $\epsilon_i$ is unknown, it is treated as random with a joint density $f(\epsilon)$. Due to this element in $U_i$, a researcher is never sure about a DM’s choice, and the researcher could only give the chance (probability) of an alternative $a_i$ to be chosen. Since, $f(\epsilon)$ is unknown, it can at best be assumed. MNL model assumes $\epsilon$ as an i i d (independent and identically distributed) extreme random variable for all $i$, with a Gumbel distribution. Accordingly, the choice probability can be derived from (2) and (3), and shown as:

$$P_i = \frac{\exp(V_i)}{\sum_{k=1}^{K} \exp(V_k)}$$

$$P_i = \frac{\exp(\sum_{m=1}^{M} \beta(m) a_i^{(m)})}{\sum_{k=1}^{K} \exp(\sum_{m=1}^{M} \beta(m) a_k^{(m)})} \tag{5}$$

### 2.2 Choquet multinomial logit model

The linear relationship in (4) guarantees monotonicity i.e. for a given $\beta(m)$, the increase (decrease) of a certain attribute value, say $a_i^{(m)}$, can only result in an increase in the total utility $V_i$. For example, increase in weight of a part is monotonically related to the increase in the total weight of the whole. Another advantage is its interpretability, i.e. the direction and strength of the influence of an attribute value $a_i^{(m)}$ is directly reflected by sign of its coefficient $\beta(m)$. A positive $\beta(m)$ indicates that $V_i^{(m)}$ is monotonically increasing with $a_i^{(m)}$, and the reverse is true for a negative $\beta(m)$.

However, such a model is unable to take into consideration any interaction among the attributes. That is the effect of unit increase (or decrease) of $a_i^{(m)}$ is always the same on $V_i$, given as $\frac{dV_i}{da_i^{(m)}} = \beta(m)$, regardless of the other attribute values. Also, the other shortcomings of MNL, as discussed in Section 4, make it untenable for many real world applications. In this regard, Choquet multinomial logit (CMNL) model is proposed in [3] by redefining $V_i$ as a function of fuzzy measure $\mu$, computed through the Choquet integral, and shown as:

$$V_i = \sum_{m=1}^{M} v_i^{(\sigma(m))}(\mu(B^{(m)}) - \mu(B^{(m+1)})) \tag{6}$$

where, $\mu : 2^{\mathbb{V}} \rightarrow [0,1]$ is a fuzzy measure on the set of utility values $(v_i^{(1)}, \ldots, v_i^{(M)}); \sigma(m)$ denotes a permutation on $M$ such that $v_i^{(\sigma(1))} \leq \ldots \leq v_i^{(\sigma(M))}$; and $B^{(m)} = \{v_i^{(\sigma(m))}, \ldots, v_i^{(\sigma(M))}\};$ and $B^{(M+1)} = \emptyset$. Here since the given fuzzy measure is used to indicate the degree of interaction among various attributes, the various permutations $\sigma(1), \ldots, \sigma(M)$ remain the same across the alternatives, and are specific to the DM. They can also be considered to be a representative of the rationale behind the choices made by the DM.

Replacing (3) as $V_i$ in the logit model, the choice probability so obtained is shown as:

$$P_i = \frac{\exp(V_i)}{\sum_{k=1}^{K} \exp(V_k)} = \frac{\exp\left(\sum_{m=1}^{M} v_i^{(\sigma(m))}(\mu(B^{(m)}) - \mu(B^{(m+1)}))\right)}{\sum_{k=1}^{K} \exp\left(\sum_{m=1}^{M} v_k^{(\sigma(m))}(\mu(B^{(m)}) - \mu(B^{(m+1)}))\right)} \tag{7}$$

(7)

The fuzzy measure $\mu(\cdot)$ helps to take account of the extent of influence, a particular attribute utility value, exerts on the other such utility values. When the attributes are not interactive, $V_i$ reduces to simple weighted averaging (see [2, 6]) of the utility values, as shown:

$$P_i = \frac{\exp\left(\sum_{m=1}^{M} w_m v_i^{(m)}\right)}{\sum_{k=1}^{K} \exp\left(\sum_{m=1}^{M} w_m v_k^{(m)}\right)}, \tag{8}$$

where $w_m$ gives the weight of $c_m$. 


2.3 Attitudinal choquet multinomial logit model

The proposed CMNL does not take into consideration the varying attitudes of the DMs. To this end, CMNL model has been extended as attitudinal CMNL (ACMNL) in [29]. Here, \( V_i \) is computed through the attitudinal Choquet integral as:

\[
V_i = \log \left( \sum_{m=1}^{M} \left( \mu(B^{(m)}) - \mu(B^{(m+1)}) \right) \lambda^{(m)} \right), \tag{9}
\]

and the ACMNL probability is shown as:

\[
P_i = \frac{\exp \left( \log \sum_{m=1}^{M} \left( \mu(B^{(m)}) - \mu(B^{(m+1)}) \right) \lambda^{(m)} \right)}{\sum_{k=1}^{K} \exp \left( \log \sum_{m=1}^{M} \left( \mu(B^{(m)}) - \mu(B^{(m+1)}) \right) \lambda^{(m)} \right)}, \tag{10}
\]

where \( \lambda \in (0, \infty), \lambda \neq 1 \) indicates a DM’s level of disjunctiveness. ACMNL model helps to cater to the situations with interactive attributes and the varying attitudinal characters of the DMs.

ACMNL gives the choice probabilities in accordance with a DM’s attitudinal character, modelled through \( \lambda \). As \( \lambda \in (0, \infty) \) moves through near 0 to near \( \infty \), the choice probability varies between two extremes of (24) and (31).

3 Interactive nested logit models

The i.i.d assumption in MNL model implies that a choice probability for an alternative is independent of the others. This leads to a closed-form, which is computationally convenient and shown as in (9). However, it is restrictive, as it hardly stands true in practice. To this end, a few variants of the logit model have appeared in the literature, which makes a more practical assumption. The nested logit (NL) model assumes a type of generalized extreme value. In this section, we extend the NL model to consider the DM’s attitude and the interaction among the attributes.

3.1 Choquet nested logit models

We extend the nested logit model with CI as Choquet nested logit (CNL) models to relax the restriction posed by the independence from irrelevant alternatives (IIA) property of the proposed CMNL and ACMNL models. The nested logit model belongs to the class of generalized extreme value (GEV) models exhibiting a variety of substitution patterns, with the assumption that the unobserved portions of the utility for the various alternatives are jointly distributed as a GEV. CNL model allows for correlations over alternatives and interaction among the attribute utilities at the same time. It reduces to MNL model when all the correlations are zero, and when the attributes are additive.

The set of alternatives is first partitioned into “nests” such that IIA property holds for any two alternatives in the same nest, but not true for the alternatives in different nests. Let a set of \( K \) alternatives, labeled \( j \), be partitioned into \( L \) non-overlapping subsets denoted by \( R_1, \ldots, R_L \), called as nests. The vector of unobserved utility \( \epsilon = (\epsilon_1, \ldots, \epsilon_K) \) is assumed to have a cumulative distribution:

\[
\exp \left( \sum_{n=1}^{L} \sum_{j \in R_n} \exp(-\epsilon_j/\gamma_n) \right)^{\gamma_n}, \tag{11}
\]

where, \( \gamma_n \) is a measure of the degree of independence in the unobserved utility among the alternatives in nest \( n \). A higher value of \( \gamma \) indicates a greater degree of independence. The statistic \( 1 - \gamma \) can be used to indicate the degree of correlation. With this distribution of the unobserved components of utility, the choice probability for alternative \( i \in R_n \) obtained with CNL model is:

\[
P_i = \frac{\exp(V_i/\gamma_n) \left( \sum_{j \in R_n} \exp(V_j/\gamma_n) \right)^{\gamma_n^{-1}} \cdots \sum_{l=1}^{L} \exp(V_l/\gamma_l) \right)^{\gamma_l^{-1}}}{\sum_{n=1}^{L} \exp(V_n/\gamma_n) \right)^{\gamma_n^{-1}}}, \tag{12}
\]

where \( V_i = \sum_{m=1}^{M} v^{(m)}(\mu(B^{(m)}) - \mu(B^{(m+1)})). \) When \( \gamma_n = 1 \), (or \( 1 - \gamma_n = 0 \) for all \( n \), it indicates no correlation among the unobserved components of utility for alternatives within a nest.
3.2 Attitudinal choquet nested logit models

We extend CNL as *attitudinal CNL* (ACNL) by replacing CI with ACI for computing representative utility. ACNL takes CNL a step ahead by also considering a DM’s attitudinal character. ACNL probability at $\lambda$ is given as:

$$P_i = \int \frac{\exp(V_i(\lambda)/\gamma_n)}{\sum_{\ell=1}^{L}\exp(V_j(\lambda)/\gamma_n)} \gamma_n^{-1} \, f(\lambda) \, d\lambda$$

where $V_i = \log_{\lambda} \left( \sum_{m=1}^{M} \left( \mu(B^{(m)}) - \mu(B^{(m+1)}) \right) \lambda^{u_i^{(m)}} \right)$.

ACNL model takes account of a DM’s attitudinal character through the adjustable parameter $\lambda$. However, in the real world situations, it may be difficult to have a crisp value of $\lambda$ known for a DM. Instead, a broad distribution of $\lambda$ may be more likely to be known. In this regard, we propose a ACNL variant that considers a broad distribution of $\lambda$, instead of a crisp value. We term this model as *stochastic ACNL* (SACNL). If we denote ACNL probability obtained at $\lambda$ as $N_i(\lambda)$, then SACNL probability is obtained as:

$$P_i = \int N_i(\lambda)f(\lambda) \, d\lambda$$

$$= \int \frac{\exp(V_i(\lambda)/\gamma_n)}{\sum_{\ell=1}^{L}\exp(V_j(\lambda)/\gamma_n)} \gamma_n^{-1} \, f(\lambda) \, d\lambda$$

SACNL is a mixture of the ACNL models arrived at with all possible $\lambda$ values with $f(\lambda)$ as the density function. A variety of SACNL models can be conceived in accordance with the form of $f(\lambda)$, which can take normal, lognormal, exponential, uniform, triangular, gamma, distributions, to name a few. It can be easily noted that the ACNL model is a special case of the SACNL model.

When $\lambda$ takes a finite set of possible distinct values and $f(\lambda)$ becomes discrete, we obtain a discrete SACNL:

$$P_i = \sum_{e=1}^{E} s_e \left( \frac{\exp(V_i(\lambda_e)/\gamma_n)}{\sum_{\ell=1}^{L}\exp(V_j(\lambda_e)/\gamma_n)} \gamma_n^{-1} \right)$$

With the provision of adjusting IIA property (through “nesting”), discrete SACNL provides a flexible model to give the different choice behaviours of individuals categorized by their respective $\lambda$ value.

4 Interactive probit models

Probit model arises under the assumption that $f(.)$ is a multivariate normal. We extend the same to consider the interaction among the attributes in the proposed Choquet multinomial logit model, and also to take into account the agent’s attitude in attitudinal Choquet multinomial logit model.

4.1 Choquet multinomial probit model

The proposed CMNL and ACMNL models, suffer from three drawbacks:

- cannot represent the random variation in the unobservable utility.
- exhibit the restrictive substitution patterns due to IIA property.
- difficult to apply when unobserved factors are correlated over time for each DM.

CNL is an useful extension of CMNL which relaxes the second restriction. In this section, we present *Choquet multinomial probit* (CMNP) by extending MNP to address all the three drawbacks.

It is based on the premise that the observed utility values interact with each other, and the unobserved utility takes a joint normal distribution. We consider the vector of the unobserved utility values for $(a_1, \ldots, a_K)$ as $\epsilon = \{\epsilon_1, \ldots, \epsilon_K\}$. It is assumed that $\epsilon$ is distributed normal with a mean vector of zero, covariance matrix $\Omega$, and density:

$$\varphi(\epsilon) = \frac{1}{(2\pi)^{\frac{K}{2}} |\Omega|^\frac{1}{2}} \exp\left(-\frac{1}{2} \epsilon^T \Omega^{-1} \epsilon\right)$$
where, \( \Omega \) depends on the variables faced by a DM, and is specific to the DM. The choice probability is arrived at as:

\[
P_i = P(U_i > U_j, \forall j \neq i)
\]

\[
= \int I(V_i + \epsilon_i > V_j + \epsilon_j, \forall j \neq i) \varphi(\epsilon) d\epsilon
\]

(16)

where, \( V_i = \sum_{m=1}^{M} v_i^{(\sigma(m))}(\mu(B^{(m)}) - \mu(B^{(m+1)})) \), \( I(.) \) is an indicator function yielding 1 or 0 depending upon if the statement in the parentheses holds or not, and the integral is over all the values of \( \epsilon \).

4.2 Attitudinal Choquet multinomial probit model

We extend CMNP with ACI to compute the representative utility. We term the resulting model as *attitudinal Choquet multinomial probit* (ACMNP), and it is expressed as:

\[
P_i = P(U_i > U_j, \forall j \neq i)
\]

\[
= \int I(V_i + \epsilon_i > V_j + \epsilon_j, \forall j \neq i) \varphi(\epsilon) d\epsilon
\]

(17)

where \( V_i = \log_\lambda \left( \sum_{m=1}^{M} (\mu(B^{(m)}) - \mu(B^{(m+1)})) \lambda^{v_i^{(\sigma(m))}} \right) \).

We now present the stochastic variant of ACMNP by considering \( \lambda \) as a random variable with density \( f(\lambda) \). We term this model as *stochastic ACMNP* (SACMNP). It is an integral of the ACMNP probabilities obtained at different \( \lambda \) values with density \( f(\lambda) \). It is expressed as:

\[
P_i = \int_{\lambda} \int_\epsilon I(V_i + \epsilon_i > V_j + \epsilon_j, \forall j \neq i) \varphi(\epsilon) d\epsilon f(\lambda) d\lambda
\]

(18)

SACMNP is more flexible than SACMNL and SACNL. When \( \lambda \) takes distinct values with probability \( s_\epsilon \) that \( \lambda = \lambda_\epsilon \), we obtain discrete SACMNP shown as:

\[
P_i = \sum_{\epsilon=1}^{E} s_\epsilon \int I(V_i + \epsilon_i > V_j + \epsilon_j, \forall j \neq i) \varphi(\epsilon) d\epsilon
\]

(19)

The proposed CMNP and ACMNP models, however, while adding to the usefulness of the classical MNP model, also inherit its following limitations that are addressed in Choquet mixed multinomial logit model, presented next.

- It is restricted to the normal distribution for all unobserved portions of utility.
- It does not have a closed form and hence is computationally difficult to solve.

5 General properties of the proposed interactive models

The proposed models are distinguished with the existing models primarily on account of their following abilities:

- consideration of the interaction among the attributes
- adaptability in accordance with the DM’s attitude

This section investigates in depth the general properties that hold good for the proposed choice models. In order to prove these properties, we consider the Möbius transform for the proposed representative utility \( V_i \), which is shown as follows:

\[
V_i = ACI(a_i^1, \ldots, a_i^M) = \log_\lambda \left( \sum_{T \subseteq N} M(T) \lambda^{\min \left\{ a_i^m \mid m \in T \right\}} \right),
\]

(20)

where \( \min \left\{ a_i^m \mid m \in T \right\} := \min_{\{m|m \in T\}} \{a_i^m\}_m \), and \( M_\mu \) refers to Möbius transform. It is interesting to note that the yield of ACI operator varies in accordance with \( \lambda \). All other values remaining the same, the more the \( \lambda \), the more is \( V_i \). Shortly, we would show that at \( \lambda \to 1 \), we obtain the conventional form of Choquet integral, i.e. Lovász extension.
In this regard, the form in (21) could be seen as a generalization of Lovász extension or more specifically the Choquet integral.

For all \( B \subseteq N \), Möbius transform \( \mathcal{M}_\mu \) of measure \( \mu \) is defined as follows:

\[
\mathcal{M}_\mu(A) = \sum_{B \subseteq A} (-1)^{|A| - |B|} \mu(B)
\]

(21)

The value \( \mathcal{M}_\mu(A) \) can be interpreted as the weight that is exclusively allocated to the subset of attributes \( A \), instead of being indirectly connected with \( A \) due to the interaction with other subsets.

**Theorem 5.1.** The representative utility as \( V_i \) as used in the proposed choice models sets a preference order on the alternatives based on the attribute values.

**Proof.** Let us consider a preference tuple, where \( a_i \succ a_j \). In such a case, we show that:

\[
V_i > V_j
\]

where \( V_i \) and \( V_j \) are obtained through ACI operator. Since \( a_i \succ a_j \), \( \min\{a_i^{(m)} \mid m \in T\} \geq \min\{a_j^{(m)} \mid m \in T\} \). We also consider the following cases:

- When \( \lambda > 1 \), then \( \lambda^\min\{a_i^{(m)} \mid m \in T\} \geq \lambda^\min\{a_j^{(m)} \mid m \in T\} \) and also \( \sum_{T \subseteq N} \mathcal{M}(T) \lambda^\min\{a_i^{(m)} \mid m \in T\} \geq \sum_{T \subseteq N} \mathcal{M}(T) \lambda^\min\{a_j^{(m)} \mid m \in T\} \).

Finally, the logarithm operator holds the direction of inequality.

- When \( 0 < \lambda < 1 \), then \( \lambda^\min\{a_i^{(m)} \mid m \in T\} \leq \lambda^\min\{a_j^{(m)} \mid m \in T\} \) and also \( \sum_{T \subseteq N} \mathcal{M}(T) \lambda^\min\{a_i^{(m)} \mid m \in T\} \leq \sum_{T \subseteq N} \mathcal{M}(T) \lambda^\min\{a_j^{(m)} \mid m \in T\} \).

Resultantly, the logarithm operator changes the direction of inequality (since it is smaller than 1).

Let us consider a case, when all the attributes of an alternative take the same value, i.e., \( a_i = (a, \ldots, a) \). Since \( \min\{a_i^{(m)} \mid m \in T\} = a \), therefore:

\[
\log_\lambda \left( \sum_{T \subseteq N} \mathcal{M}(T) \lambda^\min\{a_i^{(m)} \mid m \in T\} \right) = \log_\lambda \left( \lambda^a \sum_{T \subseteq N} \mathcal{M}(T) \right) = \log_\lambda (\lambda^a) + \log_\lambda \left( \sum_{T \subseteq N} \mathcal{M}(T) \right) = a.
\]

\( \square \)

**Theorem 5.2.** The attitudinal representative utility varies in accordance with the attitudinal parameter \( \lambda \).

**Proof.** Consider an alternative \( a_i \) described by a set of attributes \( (a_i^{(1)}, \ldots, a_i^{(m)}) \), such that \( a_i^{(m)} \in \mathbb{R} \). Also, \( \lambda \in \mathbb{R}^+ \), \( \lambda \neq 1 \). Then \( V_i \) can be written as:

\[
V_i = \log_\lambda \left( \sum_{T \subseteq N} \mathcal{M}(T) \lambda^\min\{a_i^{(m)} \mid m \in T\} \right) = \log_\lambda \left( \sum_{T \subseteq N} \mathcal{M}(T) \min\left\{ \lambda^{a_i^{(m)}} \mid m \in T\right\} \right).
\]

(22)

On the other hand in the case of \( 0 < \lambda < 1 \) and \( 0 \leq a_i^{(m)} \leq 1 \),

\[
\log_\lambda \left( \sum_{T \subseteq N} \mathcal{M}(T) \lambda^\min\{a_i^{(m)} \mid m \in T\} \right) = \log_\lambda \left( \sum_{T \subseteq N} \mathcal{M}(T) \max\left\{ \lambda^{a_i^{(m)}} \mid m \in T\right\} \right).
\]

(23)

Since

\[
\max\left\{ \lambda^{a_i^{(m)}} \mid m \in T\right\} = \sum_{K \subseteq T} \left(-1\right)^{|K|+1} \min\left\{ \lambda^{a_i^{(m)}} \mid m \in K\right\},
\]

(24)

we could extend (23) as:

\[
\log_\lambda \left( \sum_{T \subseteq N} \mathcal{M}(T) \max\left\{ \lambda^{a_i^{(m)}} \mid m \in T\right\} \right) = \log_\lambda \left( \sum_{T \subseteq N} \mathcal{M}(T) \left( \sum_{K \subseteq T} \left(-1\right)^{|K|+1} \min\left\{ \lambda^{a_i^{(m)}} \mid m \in K\right\} \right) \right).
\]

(25)
Equation (26) can be simplified as follows:

$$\log_{\lambda} \left( \sum_{T \subseteq N} M(T) \max \left\{ \lambda^{(m)} \mid m \in T \right\} \right) = \log_{\lambda} \left( \sum_{T \subseteq N} (-1)^{|T|+1} \left( \sum_{L \subseteq N \setminus T} M(T \cup L) \min \left\{ \lambda^{(m)} \mid m \in T \right\} \right) \right).$$

(26)

By assuming $M^*(T) = \sum_{L \subseteq N \setminus T} M(T \cup L)$, the equation in (26) can be reformulated as:

$$\log_{\lambda} \left( \sum_{T \subseteq N} M(T) \max \left\{ \lambda^{(m)} \mid m \in T \right\} \right) = \log_{\lambda} \left( \sum_{T \subseteq N} (-1)^{|T|+1} M^*(T) \min \left\{ \lambda^{(m)} \mid m \in T \right\} \right).$$

(27)

The attitudinal parameter $\lambda$ in $V_i$ rescales the input values. Resultantly, for the same set of attribute values, the representative utility $V_i$ varies with $\lambda$. Hence, for an alternative, a range of representative utility values, and accordingly different choice probabilities can be obtained with different attitudes represented by the different values for $\lambda$. A few significant values of the representative utility obtained with different values of $\lambda$ are shown as follows:

- When $\lambda \to 0$, then
  $$V_i \to \min\{v^{(m)}\}_{m=1}^M$$
  (28)

Accordingly, the proposed models yield the following choice probability:

$$P_i = \frac{\exp \left( \min_{m=1}^M \{v^{(m)}\} \right)}{\exp \left( \sum_{m=1}^K \min_{m=1}^M \{v_k^{(m)}\} \right)}$$

(29)

It demonstrates the choice behaviour of a pessimistic DM who takes his decisions based on the minimum of the attribute utility values, regardless of the fuzzy measure (attribute interaction).

- When $\lambda \to 1$, then
  $$V_i \to \sum_{m=1}^M v_i^{(\sigma(m))}(\mu(B^{(m)}) - \mu(B^{(m+1)}))$$
  (30)

Resultantly, the following choice probability is obtained:

$$P_i = \frac{\exp \left( \sum_{m=1}^M v_i^{(\sigma(m))} (\mu(B^{(m)}) - \mu(B^{(m+1)})) \right)}{\sum_{k=1}^K \exp \left( \sum_{m=1}^M v_k^{(\sigma(m))} (\mu(B^{(m)}) - \mu(B^{(m+1)})) \right)}$$

The choice of $\lambda \to 1$ indicates a neutral attitudinal character of the DM, and $V_i$ is reduced to Choquet integral.

- When $\lambda \to \infty$ (i.e., a large number), then $V_i \to \max\{v^{(m)}\}_{m=1}^M$. This leads to the following choice probability:

$$P_i = \frac{\exp \left( \max_{m=1}^M \{v^{(m)}\} \right)}{\exp \left( \sum_{k=1}^K \max_{m=1}^M \{v_k^{(m)}\} \right)}$$

(31)

This represents the case of an optimistic DM, who arrives at his choice probability based on only the maximum of the attribute utility values.

- When $\lambda$ is unknown, with only some knowledge about its broad distribution $f(\lambda)$ (as mean $\mu$ and covariance $\sigma$), then the stochastic choice model is obtained. If $f(\lambda)$ is considered as discrete, and $\lambda$ takes suppose $E$ distinct values, $\lambda_1, \ldots, \lambda_E$ with probability $s_e$ that $\lambda = \lambda_e$, then:

$$P_i = \sum_{e=1}^E s_e \frac{\exp \left( \log_{\lambda} \left( \sum_{m=1}^M (\mu(B^{(m)}) - \mu(B^{(m+1)})) \lambda_e^{(\sigma(m))} \right) \right)}{\sum_{k=1}^K \exp \left( \log_{\lambda} \left( \sum_{m=1}^M (\mu(B^{(m)}) - \mu(B^{(m+1)})) \lambda_k^{(\sigma(m))} \right) \right)}.$$  

(32)
6 Interactive mixed logit models

The mixed logit is based on the assumption that the unobserved portion of utility is comprised of two parts, one is the distribution as specified by the modeler and another is a i i d extreme value. We present the Choquet integral and attitudinal Choquet integral based variants of the mixed logit model.

6.1 Choquet mixed multinomial logit model

Mixed multinomial logit (MMNL) is a highly flexible model that approximates any random utility model without the restrictions of MNL, NL, and MNP. That is, it allows for a random variation in the unobserved portion of utility, unrestricted substitution patterns, and correlation in the unobserved factors. It is also not restricted to the normal distributions like MNP.

The mixed logit is based on the assumption that the unobserved portion of utility is comprised of two parts, one is a i i d extreme value. We present the Choquet integral and

\[ P_i = \int C_i(\beta, \lambda) \psi(\beta) d\beta \]

\[ \int \sum_{m=1}^{M} \frac{v_i^{(\sigma(m))} \left( \mu(B^{(m)}) - \mu(B^{(m+1)}) \right)}{\sum_{k=1}^{K} \sum_{m=1}^{M} \psi^{(\sigma(m))} \left( \mu(B^{(m)}) - \mu(B^{(m+1)}) \right)} \psi(\beta) d\beta \]

6.2 Attitudinal choquet mixed multinomial logit model

Attitudinal Choquet mixed multinomial logit (ACMMNL) model extends CMMNL model by considering the agent’s varying attitudinal character through an adjustable parameter and is shown as:

\[ P_i = \int C_i(\beta, \lambda) \psi(\beta) d\beta \]

where \( C_i(\beta, \lambda) \) is ACMNL probability for \( i^{th} \) alternative obtained at fixed \( \beta \) and \( \lambda \). The choice probability is given as:

\[ P_i = \int \sum_{m=1}^{M} \frac{\left( \mu(B^{(m)}) - \mu(B^{(m+1)}) \right) \lambda_i^{(\sigma(m))}}{\sum_{k=1}^{K} \sum_{m=1}^{M} \left( \mu(B^{(m)}) - \mu(B^{(m+1)}) \right) \lambda_i^{(\sigma(m))}} \psi(\beta) d\beta \]

Remark: This integral yields a weighted average of ACMNL probabilities, arrived at with random attribute utilities, with weights provided by \( \psi(\beta) \). A range of choice probabilities can be obtained through ACMMNL in accordance with \( \psi(\beta) \) and \( \lambda \).

We now add to the flexibility of ACMMNL by considering \( \lambda \) as a random variable with a distribution \( f(\lambda) \). We term this variant of ACMMNL as stochastic ACMMNL (SACMMNL) that is a mixture of the ACMMNL models arrived at with distributions \( \psi(\beta) \) and \( f(\lambda) \).

\[ P_i = \int \lambda \int C_i(\beta, \lambda) \psi(\beta) f(\lambda) d\beta d\lambda \]

SACMMNL is ideal for representing complex situations, for which \( \beta \) and \( \lambda \) are not known crisply but only their distributions are known.

The mixing of \( \psi(\beta) \) and \( f(\lambda) \) leads to a very fine range of choice probabilities.

Remark: SACMMNL has as its special cases ACMNL, ACNL, CMMNL, CMNL, and MNL models. We now delve upon a few of them, and in the process, we also examine the various SACMMNL variants.

When \( f(\lambda) \) is discrete and \( \lambda \) takes a finite set of \( E \) distinct values labeled \( \lambda_1, \ldots, \lambda_E \) with probability \( s_e \) that \( \lambda = \lambda_e \), then the choice probability is:

\[ P_i = \sum_{e=1}^{E} s_e \int C_i(\beta, \lambda_e) \psi(\beta) d\beta \]

\[ = \sum_{e=1}^{E} s_e \int \frac{\sum_{m=1}^{M} \left( \mu(B^{(m)}) - \mu(B^{(m+1)}) \right) \lambda_i^{(\sigma(m))}}{\sum_{k=1}^{K} \sum_{m=1}^{M} \left( \mu(B^{(m)}) - \mu(B^{(m+1)}) \right) \lambda_i^{(\sigma(m))}} \psi(\beta) d\beta \]

\[ \int C_i(\beta, \lambda) \psi(\beta) d\beta \]
The special cases of this form are investigated as following.

**Theorem 6.1.** At $\lambda \to 0$, it reduces to the pessimistic probability, as obtained in (44). Interestingly, at $\lambda \to 0$, $\mu$ is masked off such that it has no effect on $P_i$.

**Theorem 6.2.** At $\lambda_c \to 1$ with $s_c = 1$, it reduces to CMNL, shown in (6).

**Theorem 6.3.** At $\lambda \to \infty$, it gives the choice probability of an optimistic DM, as shown in (43).

**Proof.** The proofs of Theorems 6.1-6.3 follow directly from Theorem 6.2.

**Corollary:** When $\psi(\beta)$ is discrete with $\beta$ taking a finite set of $Z$ values labeled $\beta_1, \ldots, \beta_Z$, with probability $s_z$ that $\beta = \beta_z$, then:

$$P_i = \sum_{z=1}^{Z} s_z \int_{\lambda} C_i(\beta_z, \lambda) f(\lambda) d\lambda$$

$$= \sum_{z=1}^{Z} s_z \int_{\lambda} \sum_{m=1}^{M} \lambda^{v_{i,z}}_{m} (\mu(B^{(m)}) - \mu(B^{(m+1)})) f(\lambda) d\lambda$$

At a fixed $\beta$, say $\beta_z$, with $s_z = 1$, it reduces to SACMNL, as shown in (63).

**Corollary:** When both $\psi(\beta)$ and $f(\lambda)$ are discrete, then the choice probability is:

$$P_i = \sum_{e=1}^{E} \sum_{z=1}^{Z} s_es_z C_i(\beta_z, \lambda_e)$$

$$= \sum_{e=1}^{E} \sum_{z=1}^{Z} s_es_z \frac{\exp \left( \sum_{m=1}^{M} \lambda^{v_{i,z}}_{m} (\mu(B^{(m)}) - \mu(B^{(m+1)})) \right)}{\sum_{k=1}^{K} \exp \left( \sum_{m=1}^{M} \lambda^{v_{k,z}}_{m} (\mu(B^{(m)}) - \mu(B^{(m+1)})) \right)}$$

It has as the following special cases:

- Discrete SACMNL, as shown in (62) at $\beta = \beta_z$ with $s_z = 1$.
- ACMNL, shown in (60), at $\beta = \beta_z$ with $s_z = 1$, and $\lambda = \lambda_e$ with $s_e = 1$.
- All special cases of ACMNL, given in Section 6.6, at $\beta = \beta_z$ with $s_z = 1$.
- CMNL, shown in (60), at $\lambda_e \to 1$ with $s_e = 1$.
- MMNL variant at $\lambda_e \to 1$ with $s_e = 1$, and additive attributes shown as:

$$P_i = \sum_{z=1}^{Z} s_z \frac{\sum_{m=1}^{M} b^{(m)}_{i,z} v_{i,z}^{(m)} + \sum_{k=1}^{K} \sum_{m=1}^{M} b^{(m)}_{k,z} v_{k,z}^{(m)}}{\sum_{k=1}^{K} \sum_{m=1}^{M} b^{(m)}_{k,z} v_{k,z}^{(m)}}$$

where $s_z$ is the probability that $b^{(m)}$ is the weight of $m^{th}$ attribute.

### 7 Conclusions

The advanced discrete choice models such as nested logit, probit, and mixed logit are extended to consider the interaction among the attributes, and the decision-maker’s (DM’s) unique attitude at the same time. The Choquet integral at the root of the models impart its interesting capabilities to the proposed models. Also, the each of the proposed models have the convolution logit model as one of its very large number of special cases. The adjustable parameters dedicated to represent the attributes interaction, and the attitudinal character facilitate to take into account the real world factors in determining the choice probabilities, which otherwise remain unconsidered in the covnetional models.

As a future work, we intend to apply the proposed models in a real world choice problem, and compare it with other state-of-the-art approaches. Besides, it would be interesting to learn these parameter values for the given choices of a DM. Once the parameter values are known, they can be used in the proposed models to predict the DM’s choices for another set of alternatives. The established machine learning techniques, such as preference learning, offer a lot of potential in this regard. The work also has many interesting applications in consumer behaviour, supplier selection, and medical diagnosis. Also, the proposed models find significance in representing mass choice behaviour, such as in national population behaviour, or stock market dynamics.
References


D. McFadden, Modeling the choice of residential location, Transportation Research Record, 672 (1978), 72-77.


D. McFadden, Model of the choice of residential location, Transportation Research Record, 672 (1978), 72-77.


S. Pulugurta, A. Arun, M. Errampalli, Use of artificial intelligence for mode choice analysis and comparison with traditional multinomial logit model, Procedia-Social and Behavioral Sciences, 104 (2013), 583-592.


P. Vovsha, Application of cross-nested logit model to mode choice in Tel Aviv, Israel, metropolitan area, Transportation Research Record, 1607 (1997), 6-15.


H. C. W. L. Williams, On the formation of travel demand models and economic evaluation measures of user benefit, Environment and Planning A: Economy and Space, 9(3) (1977), 285-344.


F. Ye, D. Lord, Comparing three commonly used crash severity models on sample size requirements: Multinomial logit, ordered probit and mixed logit models, Analytic Methods in Accident Research, 1 (2014), 72-85.