1-2018

On substitutability and complementarity in discrete choice models

Guiyun FENG
Singapore Management University, gyfeng@smu.edu.sg

Xiaobo LI

Zizhuo WANG

On substitutability and complementarity in discrete choice models

Article *in* Operations Research Letters · December 2017

DOI: 10.1016/j.orl.2017.11.016

3 authors, including:

Guiyun Feng
University of Minnesota Twin Cities

Xiaobo Li
University of Minnesota Twin Cities

Operations Research Letters 46 (2018) 141–146

Contents lists available at ScienceDirect

Operations Research Letters

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

On substitutability and complementarity in discrete choice models

Guiyun Feng *, Xiaobo Li, Zizhuo Wang

Department of Industrial and Systems Engineering, University of Minnesota, 111 Church St SE, Minneapolis, MN 55455, USA

ARTICLE INFO

Article history:
Received 15 March 2017
Received in revised form 27 November 2017
Accepted 27 November 2017
Available online 6 December 2017

Keywords:
Discrete choice models
Substitutability
Complementarity
Halo effect
Random utility model
Representative agent model

ABSTRACT

In this paper, we propose the concepts of substitutability and complementarity in discrete choice models. These concepts concern whether the choice probability of one alternative in a choice model increases or decreases with the utility of another alternative, and they play important roles in capturing certain practical choice patterns, such as the halo effect. We study conditions on discrete choice models that will lead to substitutability and complementarity. We also present ways of constructing choice models that exhibit complementary property.

1. Introduction

In this paper, we propose and study the concepts of substitutability and complementarity in discrete choice models. Discrete choice models are useful tools to model choices made by individuals when facing a finite set of alternatives. For instance, they can be used to model consumers' choices among a set of products, passengers' choices among a set of transportation modes, and many other choice scenarios. Because of the adaptability, flexibility and analytical convenience, discrete choice models have gained a lot of attention in the economics, marketing, operations research and management science communities in the last few decades. In particular, such models have been used as the underlying behavioral model for many operational decision-making problems, such as transportation planning, assortment optimization, and multi-product pricing.

A variety of discrete choice models have been proposed in the literature. In this paper, we focus on those that map a vector of utilities of each alternative to a vector of choice probabilities. Many important classes of discrete choice models take such a form, including the random utility models, the representative agent models, and the recently proposed welfare-based choice models [12]. We will provide a more detailed review of these models and their relations in the end of this section.

In this paper, we define two useful properties in choice models – *substitutability and complementarity* – and study how such properties can be reflected in choice modeling. The two properties concern whether the choice probability of one alternative will increase or decrease when the utility of another alternative increases. We show that the random utility models only allow substitutability between alternatives. However, in certain applications, it is desirable to allow some alternatives to exhibit complementarity, in order to explain certain phenomenon observed in practice, such as the halo effect (or the synergistic effect). For that, we derive conditions under which a choice model exhibits substitutable/complementary properties. In addition, we show a few examples of choice models that allow complementarity between alternatives and propose a few ways to construct choice models with complementary patterns. As far as we know, this is the first formal study of such properties in choice models. We believe that this study will open new possibilities in the design of choice models by enlarging its horizon and capturing more practical choice patterns.

In the remainder of this section, we review several classes of discrete choice models that are related to the discussions in our paper, including the random utility model, the representative agent model and the welfare-based choice model. Before that, we first introduce the notation. Throughout the paper, the following notations will be used. We use notation $\mathcal{R}$ to denote the set of real numbers, and $\bar{\mathcal{R}} = \mathcal{R} \cup \{-\infty, +\infty\}$ to denote the set of extended real numbers. We use $\mathbf{e}$ to denote a vector of all ones, $\mathbf{e}_i$ to denote a vector of zeros except 1 at the ith entry, and $\mathbf{0}$ to denote a vector of all zeros (the dimension of these vectors will be clear from the context). Also, we write $\boldsymbol{x} \geq \boldsymbol{y}$ to denote

* Corresponding author.
E-mail addresses: fengx421@umn.edu (G. Feng), lixx3195@umn.edu (X. Li), zwang@umn.edu (Z. Wang).

https://doi.org/10.1016/j.orl.2017.11.016
0167-6377/

a componentwise relationship and Δ_{n-1} to denote the $(n-1)$-dimensional simplex, i.e., $\Delta_{n-1} = \{\boldsymbol{x}|\mathbf{e}^T\boldsymbol{x} = 1, \boldsymbol{x} \geq \mathbf{0}\}$. In our discussions, ordinary lowercase letters $x, y, \ldots$ denote scalars, boldfaced lowercase letters $\boldsymbol{x}, \boldsymbol{y}, \ldots$ denote vectors.

1.1. Random utility model

One most popular class of discrete choice models is the random utility model (RUM) (see [2] for a comprehensive review). In the random utility model, a random utility is assigned to each alternative, and each individual picks the alternative with the highest realized utility. Here, the randomness in the utilities could originate from the lack of information of the alternatives for a particular individual or the idiosyncrasies of preferences within a population. As an output, the random utility model predicts a vector of choice probabilities among the alternatives. Mathematically, suppose there are n alternatives denoted by $\mathcal{N} = \{1, 2, \ldots, n\}$, then in the random utility model, the utility of alternative i is $u_i = \pi_i + \epsilon_i, \forall i \in \mathcal{N}$, where $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_n)$ is the deterministic part of the utility and $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)$ is the random part, following a joint distribution θ. Then the probability of alternative i being chosen is (we assume θ is absolutely continuous in our discussion, which ensures that the choice probabilities are well-defined):

$$q_i(\boldsymbol{\pi}) = \mathbb{P}_{\boldsymbol{\epsilon}\sim\theta}\left(i = \underset{k\in\mathcal{N}}{\operatorname{argmax}}\,(\pi_k + \epsilon_k)\right). \tag{1}$$

And the expected utility an individual can get is:

$$w(\boldsymbol{\pi}) = \mathbb{E}_{\boldsymbol{\epsilon}\sim\theta}\left[\max_{i\in\mathcal{N}}\ \pi_i + \epsilon_i\right]. \tag{2}$$

By choosing different distributions for the random components, one can obtain different random utility models. Among them, the most widely used one is the multinomial logit (MNL) model, first proposed in [18]. In the MNL model, $(\epsilon_1, \ldots, \epsilon_n)$ follows i.i.d. Gumbel distributions with scale parameter η, and the choice probability can be written as:

$$q_i^{\text{mnl}}(\boldsymbol{\pi}) = \frac{\exp(\pi_i/\eta)}{\sum_{k\in\mathcal{N}}\exp(\pi_k/\eta)}.$$

The existence of a closed-form formula for the MNL model makes it a very popular choice model. We refer to [2,4,28] for more discussions about the MNL model. In addition to the MNL model, there are other random utility models that are studied in the literature, including the probit model (see, e.g., [9]), the nested logit model (see, e.g., [20]) and the exponomial choice model (see, e.g., [1]).

1.2. Representative agent model

Another way to model choices is through a representative agent model (RAM). In the representative agent model, a representative agent makes a choice among n alternatives on behalf of the entire population. In particular, this agent may choose any fractional amount of each alternative, or equivalently, a vector $\boldsymbol{x} = (x_1, \ldots, x_n)$ on the $(n-1)$-dimensional simplex Δ_{n-1}. When making his/her choice, the agent maximizes the expected utility while preferring some degree of diversification. More precisely, the representative agent solves the following optimization problem:

$$w^r(\boldsymbol{\pi}) = \text{maximize}_{\boldsymbol{x}\in\Delta_{n-1}}\quad \boldsymbol{\pi}^T\boldsymbol{x} - V(\boldsymbol{x}). \tag{3}$$

Here $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_n)$ is the deterministic utility of each alternative, which is similar to that in the random utility model. $V(\boldsymbol{x}) : \Delta_{n-1} \mapsto \mathcal{R}$ is a regularization term such that $-V(x)$ rewards diversification. In this paper, without loss of generality, we assume $V(\boldsymbol{x})$ is convex and lower semi-continuous. This assumption is without loss of generality because if $V(\boldsymbol{x})$ is not convex or lower semi-continuous, then we can replace $V(\boldsymbol{x})$ with a convex and lower semi-continuous function $V^{**}(\boldsymbol{x}) = \sup_{\boldsymbol{y}}\{\boldsymbol{y}^T\boldsymbol{x} - w^r(\boldsymbol{y})\}$ and Eq. (3) still holds (see, e.g. [8,12]). Moreover, if for any $\boldsymbol{\pi}$, there is a unique solution to (3), then we define

$$\boldsymbol{q}^r(\boldsymbol{\pi}) = \arg\max\left\{\boldsymbol{\pi}^T\boldsymbol{x} - V(\boldsymbol{x})\middle|\boldsymbol{x} \in \Delta_{n-1}\right\} \tag{4}$$

to be the choice probability vector given by the representative agent model.

In [15], the authors show that for any random utility model with continuously distributed random utility, there exists a representative agent model that gives the same choice probability (i.e., the same mapping from $\boldsymbol{\pi}$ to the choice probability vector $\boldsymbol{q}$). Furthermore, they show that the reverse statement is not true when $n \geq 4$. Thus the representative agent model strictly subsumes the random utility model.

1.3. Welfare-based choice model

Recently, Feng et al. [12] propose a welfare-based choice model after noticing that both the RUM and the RAM allow a welfare function $w(\boldsymbol{\pi})$ that captures the expected utility an individual can get from the choice model, and the choice probability vector can be viewed as the gradient of $w(\boldsymbol{\pi})$ with respect to $\boldsymbol{\pi}$. They first define the choice welfare function as follows:

Definition 1 (*Choice Welfare Function*)**.** Let $w(\boldsymbol{\pi})$ be a mapping from $\mathcal{R}^n$ to $\bar{\mathcal{R}} = \mathcal{R}\bigcup\{-\infty, \infty\}$. $w(\boldsymbol{\pi})$ is called a choice welfare function if $w(\boldsymbol{\pi})$ satisfies the following properties:

1. (Monotonicity): For any $\boldsymbol{\pi}_1, \boldsymbol{\pi}_2 \in \mathcal{R}^n$ and $\boldsymbol{\pi}_1 \geq \boldsymbol{\pi}_2$, $w(\boldsymbol{\pi}_1) \geq w(\boldsymbol{\pi}_2)$;
2. (Translation Invariance): For any $\boldsymbol{\pi} \in \mathcal{R}^n$, $t \in \mathcal{R}$, $w(\boldsymbol{\pi} + t\mathbf{e}) = w(\boldsymbol{\pi}) + t$;
3. (Convexity): For any $\boldsymbol{\pi}_1, \boldsymbol{\pi}_2 \in \mathcal{R}^n$ and $0 \leq \lambda \leq 1$, $\lambda w(\boldsymbol{\pi}_1) + (1-\lambda)w(\boldsymbol{\pi}_2) \geq w(\lambda\boldsymbol{\pi}_1 + (1-\lambda)\boldsymbol{\pi}_2)$.

For a differentiable choice welfare function $w(\boldsymbol{\pi})$, the welfare-based choice model derived from it is defined by $\boldsymbol{q}(\boldsymbol{\pi}) = \nabla w(\boldsymbol{\pi})$. Note that the monotonicity and translation invariance properties guarantee that $q(\boldsymbol{\pi})$ is a valid probability vector, which follows by differentiating the identity in the translation invariance property with respect to t. Feng et al. [12] prove the equivalence between the welfare-based choice model and the representative agent model as shown in Proposition 1: (For detailed definition of essentially strictly convex function, see [23]. Note that any strictly convex function is essentially strictly convex.)

Proposition 1 (*Theorem 2 from [12]*)**.** *For a choice model $\boldsymbol{q} : \mathcal{R}^n \mapsto \Delta_{n-1}$, the following statements are equivalent:*

1. *There exists a differentiable choice welfare function $w(\boldsymbol{\pi})$ such that $\boldsymbol{q}(\boldsymbol{\pi}) = \nabla w(\boldsymbol{\pi})$;*
2. *There exists an essentially strictly convex function $V(\boldsymbol{x})$ such that*

$$\boldsymbol{q}(\boldsymbol{\pi}) = \arg\max\left\{\boldsymbol{\pi}^T\boldsymbol{x} - V(\boldsymbol{x})\middle|\boldsymbol{x} \in \Delta_{n-1}\right\}.$$

They further prove that when there are only two alternatives, the class of random utility models is equivalent to the class of welfare-based choice models (thus also the representative agent models). However, when the number of alternatives $n \geq 3$, the welfare-based choice model (thus also the representative agent model) strictly subsumes the random utility model. In contrast to [12] which studies the relation between different types of choice models, in the present paper, we propose the notion of substitutability and complementarity in choice models and study the conditions under which these properties would hold. We also provide guidelines on how to construct choice models that can exhibit complementarity property.

Before we end this section, we comment that there are other types of choice models studied in the literature beyond the aforementioned ones. Some examples include the Markov chain-based choice model (see [6]), the two-stage choice model (see [16]), the generalized attraction model (see [13]) and the non-parametric model (see [11]). However, they do not take the form of mapping a utility vector to a choice probability vector. Thus they are less related to this work. For the sake of space, we choose not to include a detailed review for those choice models in this paper.

2. Substitutability and complementarity in choice models

In this section, we propose two concepts in choice models, *substitutability* and *complementarity*, and discuss their practical implications. We show that if a choice model is derived from a random utility model, then the alternatives can only exhibit substitutability. However, the representative agent model and the welfare-based choice model allow for more flexible substitutability/complementarity patterns. We also show how these properties can be reflected through the choice welfare function in a welfare-based choice model or through the regularization term in a representative agent model. Before we formally define these two concepts, we first introduce the definition of local monotonicity:

Definition 2 (*Local Monotonicity*)**.** A function $f(x) : \mathcal{R} \mapsto \mathcal{R}$ is locally increasing at x if there exists $\delta > 0$ such that $f(x-h) \le f(x) \le f(x+h),\ \forall\, 0 < h < \delta$. Similarly, $f(x)$ is locally decreasing at x if there exists $\delta > 0$ such that $f(x-h) \ge f(x) \ge f(x+h),\ \forall\, 0 < h < \delta$.

Now we introduce the definition of substitutability and complementarity in choice models:

Definition 3. Consider a choice model $\boldsymbol{q}(\boldsymbol{\pi}) : \mathcal{R}^n \mapsto \Delta_{n-1}$. For any fixed $\boldsymbol{\pi}$ and $i, j \in \mathcal{N}$:

1. (Substitutability) If $q_j(\boldsymbol{\pi})$ is locally decreasing in π_i at $\boldsymbol{\pi}$, then we say alternative i is substitutable to alternative j at $\boldsymbol{\pi}$. Furthermore, if $q_j(\boldsymbol{\pi})$ is decreasing in π_i for all $\boldsymbol{\pi}$, then we say alternative i is substitutable to alternative j;
2. (Complementarity) If $q_j(\boldsymbol{\pi})$ is locally increasing in π_i at $\boldsymbol{\pi}$, then we say alternative i is complementary to alternative j at $\boldsymbol{\pi}$. Furthermore, if $q_j(\boldsymbol{\pi})$ is increasing in π_i for all $\boldsymbol{\pi}$, then we say alternative i is complementary to alternative j.
3. (Substitutable and Non-Substitutable Choice Model) If alternative i is substitutable to alternative j for all $i \ne j$, then we say $\boldsymbol{q}(\boldsymbol{\pi})$ is a substitutable choice model. Otherwise, we say $\boldsymbol{q}(\boldsymbol{\pi})$ is a non-substitutable choice model.

Remark. Note that for a choice model, if not all pairs of alternatives are substitutable, then the model is called a non-substitutable choice model. Moreover, it is not possible for an alternative to be complementary with all other alternatives because the sum of choice probabilities must be equal to 1. Therefore, even in a non-substitutable choice model, at any $\boldsymbol{\pi}$, there must be at least one pair of alternatives that are substitutable at $\boldsymbol{\pi}$.

We note that the complementary property is closely related to the halo effect, which is first conceptualized in [26] (this effect is also called the synergistic effect, see [10]). The halo effect is a cognitive bias in which an observer's overall impression of a person, company, brand or product influences the observer's feelings and thoughts about that entity's character or properties [21]. For a comprehensive review and discussion about the halo effect, we refer the readers to [24]. In the context of consumer theory and marketing, the halo effect is the phenomenon that the choice probabilities of certain existing products increase after a new product (usually of the same brand) is introduced. In a choice model that maps a vector of utilities to a vector of choice probabilities, introducing a new product can be viewed as increasing the utility of that product from negative infinity to some finite value. Therefore, the notion of complementarity defined in Definition 3 provides an alternative characterization of the halo effect in the context of choice modeling. We have the following example illustrating this situation:

Example 1. Suppose a customer is considering to buy a camera from the following three alternatives: a Canon-A model, a Canon-B model and a Sony-C model. On a certain website, there are customer review scores for each model, which we denote by v_1, v_2 and v_3, respectively. We assume that the customer's choice is solely based on those review scores (suppose other factors are fixed). That is, the choice probability $\boldsymbol{q}$ is a function of $\boldsymbol{v} = (v_1, v_2, v_3)$. Suppose at a certain time, a new review for the Canon-A model comes in, rating it favorably. How would it change the purchase probability of the Canon-B model?

The answer to the above question may depend. There might be two forces. On one hand, due to a new favorable rating given to the Canon brand, the probability of choosing the Canon-B model might increase. On the other hand, the favorable rating for the Canon-A model might switch some customers from the Canon-B model to the Canon-A model. Either force might be dominant in practice. If the former force is stronger, then it is plausible that one additional favorable rating for the Canon-A model might increase the choice probability of the Canon-B model (this scenario can be viewed as a case of the halo effect).

The above example illustrates that sometimes it might be desirable to have a choice model in which a certain pair of alternatives exhibit complementarity. One may notice that the above example may be reminiscent of the nested logit model, in which the customers first choose a nest (in this case, the brand), and then choose a particular product. However, we note that the nested logit model with dissimilarity parameters within $(0, 1]$ is a random utility model (see, e.g., [2]). Therefore, it is impossible to capture complementarity between alternatives through such a nested logit model (see Proposition 3). In Section 3, we show that complementarity of alternatives can be captured through a general nested logit model, in which the dissimilarity parameters are allowed to be greater than one. We will also show other ways to construct choice models with complementary alternatives in the next section.

In the following, we investigate some basic facts about substitutability and complementarity.

Proposition 2. *Consider a choice model $\boldsymbol{q}(\boldsymbol{\pi}) : \mathcal{R}^n \mapsto \Delta_{n-1}$ that is derived from a differentiable choice welfare function $w(\boldsymbol{\pi})$. For any i, alternative i must be complementary to itself. Furthermore, if $w(\boldsymbol{\pi})$ is second-order continuously differentiable and alternative i is substitutable (complementary, resp.) to alternative j at $\boldsymbol{\pi}$, then alternative j must be substitutable (complementary, resp.) to alternative i at $\boldsymbol{\pi}$.*

Proof. Since $w(\boldsymbol{\pi})$ is convex and differentiable, for any $\boldsymbol{\pi} \in \mathcal{R}^n$ and any $t > 0$, we have $w(\boldsymbol{\pi} + t\boldsymbol{e_i}) - w(\boldsymbol{\pi}) \ge t\boldsymbol{e_i^T}\nabla w(\boldsymbol{\pi}) = tq_i(\boldsymbol{\pi})$ and $w(\boldsymbol{\pi}) - w(\boldsymbol{\pi}+t\boldsymbol{e_i}) \ge -t\boldsymbol{e_i^T}\nabla w(\boldsymbol{\pi}+t\boldsymbol{e_i}) = -tq_i(\boldsymbol{\pi}+t\boldsymbol{e_i})$. From these two inequalities, we have $q_i(\boldsymbol{\pi}+t\boldsymbol{e_i}) - q_i(\boldsymbol{\pi}) \ge 0$, for all $t > 0$ and $\boldsymbol{\pi}$. Thus, alternative i is complementary to itself.

Furthermore, if $w(\boldsymbol{\pi})$ is second-order continuously differentiable, then we have $\frac{\partial q_i}{\partial \pi_j} = \frac{\partial^2 w}{\partial \pi_i \partial \pi_j} = \frac{\partial^2 w}{\partial \pi_j \partial \pi_i} = \frac{\partial q_j}{\partial \pi_i}$. Thus, if alternative i is substitutable (complementary, resp.) to alternative j at $\boldsymbol{\pi}$, then alternative j is substitutable (complementary, resp.) to alternative i at $\boldsymbol{\pi}$. □

Proposition 2 implies that when $w(\boldsymbol{\pi})$ is second-order continuously differentiable, the substitutability (complementarity, resp.) property is a reciprocal property. In these cases, we shall say i and j are substitutable (complementary, resp.) in the following discussions.

In the following, we investigate substitutability and complementarity in choice models. First, by the definition of random utility model and Proposition 6 in [12], random utility models are all substitutable. We formalize it in the following proposition:

Proposition 3. *Any random utility model $\boldsymbol{q}(\boldsymbol{\pi})$ is a substitutable choice model.*

According to Proposition 3, in a random utility model, if the utility of one alternative increases while the utilities of all other alternatives stay the same, then it must be that the choice probabilities of all other alternatives decrease. This is certainly plausible in practice, especially if $\boldsymbol{\pi}$ is interpreted as how much a consumer values each product. However, as Example 1 has shown, sometimes it might be desirable to allow certain alternatives to exhibit certain degrees of complementarity. This is especially true if we allow more versatile meanings of the utility $\boldsymbol{\pi}$.

Now we present conditions for a choice model to be substitutable or non-substitutable. In the following discussion, we consider choice models $\boldsymbol{q}(\boldsymbol{\pi})$ that are derived from differentiable choice welfare functions $w(\boldsymbol{\pi})$ (thus equivalently they can be derived from representative agent models). We provide necessary (sufficient, resp.) conditions for a choice model to be substitutable, and consequently also obtain sufficient (necessary, resp.) conditions for a choice model to be non-substitutable. We have the following theorem:

Theorem 1. *Consider a choice model $\boldsymbol{q}(\boldsymbol{\pi}) : \mathcal{R}^n \mapsto \Delta_{n-1}$ that is derived from a differentiable choice welfare function $w(\boldsymbol{\pi})$. Then*

1. *$\boldsymbol{q}(\boldsymbol{\pi})$ is a substitutable choice model if and only if $w(\boldsymbol{\pi})$ is submodular.*
2. *If $\boldsymbol{q}(\boldsymbol{\pi})$ is a substitutable choice model, then there exists an essentially strictly convex $V(\cdot)$ with $\bar{V}_i(\cdot)$ supermodular on $\mathcal{R}^{n-1}$ for all i, such that $\boldsymbol{q}(\boldsymbol{\pi}) = \arg\max\left\{\boldsymbol{\pi}^T\boldsymbol{x} - V(\boldsymbol{x})\middle|\boldsymbol{x} \in \Delta_{n-1}\right\}$, where*

$$\bar{V}_i(\boldsymbol{z}) = \begin{cases} V\left(z_1, z_2, \ldots, z_{i-1}, 1 - \displaystyle\sum_{j=1}^{n-1} z_j, z_i, \ldots, z_{n-1}\right), \\ \qquad \text{if } \boldsymbol{e}^T\boldsymbol{z} \le 1 \text{ and } \boldsymbol{z} \ge 0, \\ +\infty, \quad \text{otherwise.} \end{cases} \tag{5}$$

Furthermore, the reverse is true if $n = 3$.

Proof. In this proof, we use the following lemma from [22].

Lemma 1 (*Theorem 8.1 and Proposition 8.2 of [22]*). *Let $f : \mathcal{R}^n \mapsto \mathcal{R} \cup \{\infty\}$ be a function such that there exists at least one $\boldsymbol{\pi}$ such that $f(\boldsymbol{\pi}) < \infty$. Let $g(\boldsymbol{x}) = \max_{\boldsymbol{\pi}}\{\boldsymbol{\pi}^T\boldsymbol{x} - f(\boldsymbol{\pi})\}$ be the convex conjugate of f. We have*

1. *If f is submodular, then g is supermodular.*
2. *If $n = 2$ and f is supermodular, then g is submodular.*

Now we use this lemma to prove the theorem. To prove the first part, by [25], a differentiable function $w(\boldsymbol{\pi})$ is submodular in $\boldsymbol{\pi}$ if and only if $\frac{\partial w(\boldsymbol{\pi})}{\partial \pi_i}$ is decreasing in π_j for all $i \neq j$. By the definition of $\boldsymbol{q}(\boldsymbol{\pi}) = \nabla w(\boldsymbol{\pi})$, the result holds.

For the second part, let $V(\boldsymbol{x}) = \max_{\boldsymbol{\pi}}\{\boldsymbol{\pi}^T\boldsymbol{x} - w(\boldsymbol{\pi})\}$ be the convex conjugate of $w(\boldsymbol{\pi})$. Therefore, $V(\boldsymbol{x})$ is essentially strictly convex and $\boldsymbol{q}(\boldsymbol{\pi}) = \arg\max\left\{\boldsymbol{\pi}^T\boldsymbol{x} - V(\boldsymbol{x})\middle|\boldsymbol{x} \in \Delta_{n-1}\right\}$. For any $\boldsymbol{y} \in \mathcal{R}^{n-1}$ and $i \in \mathcal{N}$, define $f_i(\boldsymbol{y}) = w(y_1, y_2, \ldots, y_{i-1}, 0, y_i, \ldots, y_{n-1})$. Also define $\boldsymbol{\pi}_{-i} = (\pi_1, \ldots, \pi_{i-1}, \pi_{i+1}, \ldots, \pi_n)$, then we have

$$\begin{aligned}\bar{V}_i(\boldsymbol{z}) &= \max_{\boldsymbol{\pi}}\{\boldsymbol{\pi}_{-i}^T\boldsymbol{z} + \pi_i(1 - \boldsymbol{e}^T\boldsymbol{z}) - w(\boldsymbol{\pi})\} \\ &= \max_{\boldsymbol{\pi}, \pi_i = 0}\{\boldsymbol{\pi}_{-i}^T\boldsymbol{z} + \pi_i(1 - \boldsymbol{e}^T\boldsymbol{z}) - w(\boldsymbol{\pi})\} = \max_{\boldsymbol{y}}\{\boldsymbol{y}^T\boldsymbol{z} - f_i(\boldsymbol{y})\},\end{aligned}$$

where the second equality is due to the translation invariance property of $w(\boldsymbol{\pi})$. The submodularity of $w(\boldsymbol{\pi})$ implies the submodularity of $f_i(\boldsymbol{y})$ for all $i \in \mathcal{N}$. Thus $\bar{V}_i(\boldsymbol{z})$, as the convex conjugate of $f_i(\boldsymbol{y})$, is supermodular by Lemma 1.

For the last statement, since $V(\cdot)$ is an essentially strictly convex function, $\boldsymbol{q}(\boldsymbol{\pi}) = \arg\max\left\{\boldsymbol{\pi}^T\boldsymbol{x} - V(\boldsymbol{x})\middle|\boldsymbol{x} \in \Delta_{n-1}\right\}$ is well-defined. By Theorem 2 in [12], $\boldsymbol{q}(\boldsymbol{\pi}) = \nabla w(\boldsymbol{\pi})$ where $w(\boldsymbol{\pi}) = \sup\left\{\boldsymbol{\pi}^T\boldsymbol{x} - V(\boldsymbol{x})\middle|\boldsymbol{x} \in \Delta_{n-1}\right\}$. For any $\boldsymbol{y} \in \mathcal{R}^{n-1}$ and $i \in \mathcal{N}$, define $f_i(\boldsymbol{y}) = w(y_1, y_2, \ldots, y_{i-1}, 0, y_i, \ldots, y_{n-1})$. Also define $\boldsymbol{x}_{-i} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$, then we have

$$\begin{aligned}f_i(\boldsymbol{y}) &= \max_{\boldsymbol{x} \in \Delta_{n-1}}\left\{\boldsymbol{x}_{-i}^T\boldsymbol{y} + 0(1 - \boldsymbol{e}^T\boldsymbol{x}_{-i}) - V(\boldsymbol{x})\right\} \\ &= \max_{\boldsymbol{x} \in \Delta_{n-1}}\left\{\boldsymbol{x}_{-i}^T\boldsymbol{y} - \bar{V}_i(\boldsymbol{x}_{-i})\right\} \\ &= \max_{\boldsymbol{e}_{n-1}^T\boldsymbol{x}_{-i} \le 1,\ \boldsymbol{x}_{-i} \ge 0}\left\{\boldsymbol{x}_{-i}^T\boldsymbol{y} - \bar{V}_i(\boldsymbol{x}_{-i})\right\} = \max_{\boldsymbol{z}}\left\{\boldsymbol{y}^T\boldsymbol{z} - \bar{V}_i(\boldsymbol{z})\right\},\end{aligned}$$

where the last equality holds since $\bar{V}_i(\boldsymbol{z}) = +\infty$ for all $\boldsymbol{z} \notin \{\beta \in \mathcal{R}^{n-1} | \boldsymbol{e}_{n-1}^T\beta \le 1,\ \beta \ge 0\}$. From Lemma 1, given that $n = 3$ and thus $\boldsymbol{y} \in \mathcal{R}^2$, $f_i(\boldsymbol{y})$ is submodular. It remains to show that $w(\boldsymbol{\pi})$ is also submodular. According to the definition of submodularity, it suffices to show that $q_i(\boldsymbol{\pi})$ is locally decreasing with π_j for all $j \neq i$ for all $\boldsymbol{\pi}$. Fix i, j and let $k \neq i, j$. We assume $i > j$ without loss of generality. We have $q_i(\boldsymbol{\pi} - \pi_k\boldsymbol{e}) = q_i(\boldsymbol{\pi})$ from the translation invariance property. But $q_i(\boldsymbol{\pi} - \pi_k\boldsymbol{e}) = \frac{\partial f_k(\pi_i - \pi_k, \pi_j - \pi_k)}{\partial \pi_i}$ is non-decreasing with π_j due to the submodularity of f_k. Thus $w(\boldsymbol{\pi})$ is submodular and $\boldsymbol{q}(\boldsymbol{\pi}) = \nabla w(\boldsymbol{\pi})$ is a substitutable choice model. □

Theorem 1 provides some sufficient and necessary conditions for $\boldsymbol{q}(\boldsymbol{\pi})$ to be substitutable. We note that the supermodularity of $\bar{V}_i(\cdot)$ has nothing to do with the supermodularity of $V(\cdot)$. In fact, since $V(\boldsymbol{x})$ is only defined on Δ_{n-1}, it can always be extended to a supermodular function in $\mathcal{R}^n$ by defining $V(\boldsymbol{x}) = +\infty$ for all $\boldsymbol{x} \notin \Delta_{n-1}$. The definition of $\bar{V}_i(\cdot)$ reduces a redundant variable in $V(\cdot)$, making the condition meaningful.

Next we provide an easy-to-check sufficient condition for a choice model to be substitutable. The following theorem shows that choice models derived from separable $V(\cdot)$s are always substitutable:

Theorem 2. *If $V(\boldsymbol{x}) = \sum_{i \in \mathcal{N}} V_i(x_i)$ on Δ_{n-1} where $V_i(x_i) : [0, 1] \mapsto \mathcal{R}$ is a strictly convex function for all $i \in \mathcal{N}$, then $\boldsymbol{q}(\boldsymbol{\pi})$ defined by $\boldsymbol{q}(\boldsymbol{\pi}) = \arg\max\left\{\boldsymbol{\pi}^T\boldsymbol{x} - V(\boldsymbol{x})\middle|\boldsymbol{x} \in \Delta_{n-1}\right\}$ is a substitutable choice model.*

Proof. We first consider the case where $V_i(x_i)$ is differentiable for all $i \in \mathcal{N}$. Let $\lambda(\boldsymbol{\pi})$ be the Lagrangian multiplier of the constraint $\sum_i x_i = 1$. The KKT conditions (see [5]) for problem $\max\left\{\boldsymbol{\pi}^T\boldsymbol{x} - V(\boldsymbol{x})\middle|\boldsymbol{x} \in \Delta_{n-1}\right\}$ can be written as:

$$\begin{aligned}&\pi_i - V_i'(q_i(\boldsymbol{\pi})) - \lambda(\boldsymbol{\pi}) \le 0, \quad \forall i \in \mathcal{N}; \\ &\pi_i - V_i'(q_i(\boldsymbol{\pi})) - \lambda(\boldsymbol{\pi}) = 0, \quad \forall i \text{ s.t. } q_i(\boldsymbol{\pi}) \neq 0; \\ &\qquad q_i(\boldsymbol{\pi}) \ge 0, \qquad\qquad\quad \forall i \in \mathcal{N}; \\ &\sum_{i \in \mathcal{N}} q_i(\boldsymbol{\pi}) = 1.\end{aligned}$$

Now we consider any two points $\boldsymbol{\pi}_0$ and $\boldsymbol{\pi}_0 + t\boldsymbol{e_i}$ where $\boldsymbol{e_i}$ is a unit vector along the ith coordinate axis and $t > 0$. Suppose that there exists a $j \neq i$ such that $q_j(\boldsymbol{\pi}_0 + t\boldsymbol{e_i}) > q_j(\boldsymbol{\pi}_0)$. Since V_j is strictly convex, $V_j'(q_j(\boldsymbol{\pi}_0+t\boldsymbol{e_i}))>V_j'(q_j(\boldsymbol{\pi}_0))$. There are two possible cases for $q_j(\boldsymbol{\pi}_0)$:

- $q_j(\boldsymbol{\pi}_0)>0$: In this case, we have $\pi_j-V_j'(q_j(\boldsymbol{\pi}_0+t\boldsymbol{e_i}))-\lambda(\boldsymbol{\pi}_0+t\boldsymbol{e_i})=0$ and $\pi_j-V_j'(q_j(\boldsymbol{\pi}_0))-\lambda(\boldsymbol{\pi}_0)=0$, therefore, we have $\lambda(\boldsymbol{\pi}_0+t\boldsymbol{e_i})<\lambda(\boldsymbol{\pi}_0)$.
- $q_j(\boldsymbol{\pi}_0)=0$: In this case, $\pi_j-V_j'(q_j(\boldsymbol{\pi}_0))-\lambda(\boldsymbol{\pi}_0)\le 0$, which implies that $\pi_j-V_j'(q_j(\boldsymbol{\pi}_0+t\boldsymbol{e_i}))-\lambda(\boldsymbol{\pi}_0)<0$. But $\pi_j-V_j'(q_j(\boldsymbol{\pi}_0+t\boldsymbol{e_i}))-\lambda(\boldsymbol{\pi}_0+t\boldsymbol{e_i})=0$, we have $\lambda(\boldsymbol{\pi}_0+t\boldsymbol{e_i})<\lambda(\boldsymbol{\pi}_0)$.

In both cases, $\lambda(\boldsymbol{\pi}_0+t\boldsymbol{e_i})<\lambda(\boldsymbol{\pi}_0)$. This implies that $q_j(\boldsymbol{\pi}_0+t\boldsymbol{e_i})\ge q_j(\boldsymbol{\pi}_0)$ for all $j\ne i$. Note that we also have $q_i(\boldsymbol{\pi}_0+t\boldsymbol{e_i})>q_i(\boldsymbol{\pi}_0)$ by Proposition 2. Therefore, we have $\sum_{j\in\mathcal{N}}q_j(\boldsymbol{\pi}_0+t\boldsymbol{e_i})>\sum_{j\in\mathcal{N}}q_j(\boldsymbol{\pi}_0)=1$, which contradicts with that $\boldsymbol{q}(\boldsymbol{\pi}_0+t\boldsymbol{e_i})\in\Delta_{n-1}$. Thus we have $q_j(\boldsymbol{\pi}_0+t\boldsymbol{e_i})\le q_j(\boldsymbol{\pi}_0)$ for all $j\ne i$. Since this is true for all $\boldsymbol{\pi}_0$ and $t>0$, $\boldsymbol{q}$ is substitutable.

If $V_i(x_i)$ is not differentiable, we need to replace the derivative with the subgradient in the above argument. Since V_i is strictly convex, $g_1>g_2$ for all $g_1\in\partial V_i(x_1)$ and $g_2\in\partial V_i(x_2)$ if $x_1>x_2$, the above argument is still valid. □

3. Examples and constructions of non-substitutable choice models

3.1. General nested logit model

The nested logit model, first proposed in [3], is perhaps the most widely used choice model other than the MNL model. In this model, it is assumed that the set of alternatives is partitioned into K subsets (*nests*) denoted by $B_1, B_2, \dots, B_K$. The probability of choosing alternative i, given that $i\in B_k$ is

$$q_i^{\mathrm{nl}}(\boldsymbol{\pi})=\frac{\exp(\pi_i/\lambda_k)\left(\sum_{j\in B_k}\exp(\pi_j/\lambda_k)\right)^{\lambda_k-1}}{\sum_{l=1}^K\left(\sum_{j\in B_l}\exp(\pi_j/\lambda_l)\right)^{\lambda_l}},\tag{6}$$

where λ_k is called the *dissimilarity parameter* for the kth nest. [27] interpret the dissimilarity parameters as a measure of substitutability among alternatives: if $\lambda_k\in(0,1)$, then the substitution is greater within nests than across nests, while if $\lambda_k>1$, then the substitution is greater across nests than within nests. In [19], the authors show that the nested logit model is consistent with the RUM for all $\boldsymbol{\pi}\in\mathcal{R}^n$ if and only if $\lambda_k\in(0,1]$ for all $k=1,\dots,K$.

There have been many studies on the case when λ_k is greater than one for some k. Most of those studies focus on how to relate this case with the RUM. For example, [7] shows that $\boldsymbol{q}^{\mathrm{nl}}(\boldsymbol{\pi})$ is consistent with the RUM for certain ranges of $\boldsymbol{\pi}$. [17] and [14] provide tests for consistency of the nested logit model with utility maximization.

When $\lambda_k>1$ for some k, the nested logit model can possess some interesting properties. In particular, when a new product is introduced to a nest k with $\lambda_k>1$, the probability of choosing certain existing products in that nest may increase in some circumstances, thus certain pairs of products may exhibit complementarity relationship [10]. In fact, as the next proposition shows, complementarity property exists in any nested logit model with certain dissimilarity parameter greater than one.

Proposition 4. *Consider a nested logit model with at least two nests. For any nest k with dissimilarity parameter $\lambda_k>1$ and any two distinct alternatives i and j in that nest, there always exists $\boldsymbol{\pi}\in\mathcal{R}^n$ such that $\frac{\partial q_i^{\mathrm{nl}}(\boldsymbol{\pi})}{\partial\pi_j}>0$.*

Proof. By simple algebra, we have:

$$\frac{\partial q_i^{\mathrm{nl}}(\boldsymbol{\pi})}{\partial\pi_j}=K(\boldsymbol{\pi})\left(-\frac{1}{\lambda_k}\left(\sum_{s\in B_k}\exp(\pi_s/\lambda_k)\right)^{\lambda_k}+\frac{\lambda_k-1}{\lambda_k}\sum_{l\ne k}\left(\sum_{s\in B_l}\exp(\pi_s/\lambda_l)\right)^{\lambda_l}\right),\tag{7}$$

where

$$K(\boldsymbol{\pi})=\frac{\exp\left((\pi_i+\pi_j)/\lambda_k\right)\left(\sum_{s\in B_k}\exp(\pi_s/\lambda_k)\right)^{\lambda_k-2}}{\left(\sum_{l=1}^K\left(\sum_{s\in B_l}\exp(\pi_s/\lambda_l)\right)^{\lambda_l}\right)^2}>0.$$

Clearly, $-\frac{1}{\lambda_k}\left(\sum_{s\in B_k}\exp(\pi_s/\lambda_k)\right)^{\lambda_k}<0$ and $\frac{\lambda_k-1}{\lambda_k}\sum_{l\ne k}\left(\sum_{s\in B_l}\exp(\pi_s/\lambda_l)\right)^{\lambda_l}>0$. Therefore, when $\boldsymbol{\pi}$ is chosen such that $\left(\sum_{s\in B_k}\exp(\pi_s/\lambda_k)\right)^{\lambda_k}\le(\lambda_k-1)\sum_{l\ne k}\left(\sum_{s\in B_l}\exp(\pi_s/\lambda_l)\right)^{\lambda_l}$, $\frac{\partial q_i^{\mathrm{nl}}(\boldsymbol{\pi})}{\partial\pi_j}\ge 0$. Finally we note that one can always choose $\boldsymbol{\pi}$ such that the inequality holds. This is because we can choose π_s large enough for some $s\in B_l$, $l\ne k$. □

3.2. Quadratic regularization

Another way to generate non-substitutable choice models is to start from the representative agent model and to choose $V(\cdot)$ as a quadratic function. Remember that $V(\cdot)$ has to be a convex function in the representative agent model. Thus, a quadratic function could be used as an approximation. We have the following proposition about the substitutability and complementarity in such models.

Proposition 5. *Consider a choice model $\boldsymbol{q}(\boldsymbol{\pi})=\arg\max\left\{\boldsymbol{\pi}^T\boldsymbol{x}-V(\boldsymbol{x})\middle|\boldsymbol{x}\in\Delta_{n-1}\right\}$ with $V(\boldsymbol{x})=\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}$ where $\boldsymbol{A}$ is a positive definite matrix. If the choice model is substitutable, then $A_{jk}-A_{ik}-A_{ij}+A_{ii}\ge 0$ for all distinct $i,j,k\in\mathcal{N}$, where A_{ij} is the (i,j) th entry of $\boldsymbol{A}$. Furthermore, the reverse is true if $n=3$.*

Proof. According to Theorem 1, it suffices to prove that

$$\bar V_i(\boldsymbol{z})=\begin{cases}V\left(z_1,z_2,\dots,z_{i-1},1-\sum_{j=1}^{n-1}z_j,z_i,\dots,z_{n-1}\right), & \\ \qquad\text{if } \boldsymbol{e}^T\boldsymbol{z}\le 1 \text{ and } \boldsymbol{z}\ge 0, & \\ +\infty, \quad \text{otherwise} & \end{cases}$$

is supermodular if and only if $A_{j,k}-A_{i,k}-A_{i,j}+A_{i,i}\ge 0$ for all distinct $i,j,k\in\mathcal{N}$. For $i\in\mathcal{N}$, $\bar V_i$ is an $n-1$ variate quadratic function. Let H^i denote the Hessian matrix of $\bar V_i(\boldsymbol{z})$. For $j,k\in\{1,2,\dots,n-1\}$ and $j\ne k$, the off-diagonal element $H^i_{j,k}=A_{\tilde j,\tilde k}-A_{i,\tilde k}-A_{i,\tilde j}+A_{i,i}$, where

$$\tilde j=\begin{cases} j, & \text{if } j<i,\\ j+1, & \text{if } j\ge i;\end{cases}\quad\text{and}\quad \tilde k=\begin{cases} k, & \text{if } k<i,\\ k+1, & \text{if } k\ge i.\end{cases}$$

Thus, $\bar V_i(\boldsymbol{z})$ is supermodular if and only if $H^i_{j,k}\ge 0$ for all $j,k\in\{1,2,\dots,n-1\}$ and $j\ne k$, which is equivalent to $A_{j,k}-A_{i,k}-A_{i,j}+A_{i,i}\ge 0$ for all distinct $i,j,k\in\mathcal{N}$. □

By Proposition 5, we know that when $n=3$ and $V(\boldsymbol{x})=\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}$ with $\boldsymbol{A}\succ\boldsymbol{0}$, the choice model defined by $\boldsymbol{q}(\boldsymbol{\pi})=\arg\max\left\{\boldsymbol{\pi}^T\boldsymbol{x}-V(\boldsymbol{x})\middle|\boldsymbol{x}\in\Delta_{n-1}\right\}$ is substitutable if and only if

$A_{12}+A_{33}\ge A_{13}+A_{23}$, $\quad A_{13}+A_{22}\ge A_{12}+A_{23}$ and $A_{23}+A_{11}\ge A_{12}+A_{13}$.

Note that the above condition is different from $\boldsymbol{A}$ being positive semidefinite. Indeed, the following example shows a case where the choice model is not substitutable even if $V(\boldsymbol{x})$ is strictly convex and supermodular (this example was also shown in [12] for showing that the representative model strictly subsumes the random utility model even when there are only three alternatives):

Example 2. Consider $\boldsymbol{q}(\boldsymbol{\pi}) = \arg\max\left\{\boldsymbol{\pi}^T\boldsymbol{x} - V(\boldsymbol{x})\middle|\boldsymbol{x} \in \Delta_{n-1}\right\}$, where $V(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}$ with $\boldsymbol{A} = \begin{bmatrix} 3 & 2 & 0 \\ 2 & 3 & 2 \\ 0 & 2 & 3 \end{bmatrix} \succ \boldsymbol{0}$. It is easy to see that $V(\boldsymbol{x})$ is strictly convex and supermodular. However, it does not satisfy that $A_{13}+A_{22} \geq A_{12}+A_{23}$. By some further calculations, we obtain that

$$\bar{V}_2(\boldsymbol{z}) = \boldsymbol{z}^T\begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}\boldsymbol{z} - [-2; -2]^T\boldsymbol{z} + 3,$$

which is not supermodular. Therefore $\boldsymbol{q}(\boldsymbol{\pi})$ is not a substitutable choice model by Theorem 1. □

3.3. Crossing transformation

Crossing transformation is a systematic way to generate non-substitutable choice models from existing substitutable choice models by using the welfare-based approach. Let $\boldsymbol{A}$ be an $m \times n$ matrix with $A_{ij} \geq 0$ and $\boldsymbol{A}\boldsymbol{e}_n = \boldsymbol{e}_m$, where $\boldsymbol{e}_\ell$ refers to an ℓ-dimensional column vector of ones. Given an existing choice welfare function $\bar{w}(\cdot) : \mathcal{R}^m \mapsto \bar{\mathcal{R}}$ and its choice probabilities $\bar{\boldsymbol{q}}(\cdot)$, we can easily verify that $w(\boldsymbol{\pi}) = \bar{w}(\boldsymbol{A}\boldsymbol{\pi})$ is still a choice welfare function that maps $\mathcal{R}^n$ to $\bar{\mathcal{R}}$ and the corresponding welfare-based choice model is $\boldsymbol{q}(\boldsymbol{\pi}) = \nabla_{\boldsymbol{\pi}} w(\boldsymbol{\pi}) = \boldsymbol{A}^T\nabla\bar{w}(\boldsymbol{A}\boldsymbol{\pi}) = \boldsymbol{A}^T\bar{\boldsymbol{q}}(\boldsymbol{A}\boldsymbol{\pi})$. By some calculation, we have $\nabla^2 w(\boldsymbol{\pi}) = \boldsymbol{A}^T\nabla^2\bar{w}(\boldsymbol{A}\boldsymbol{\pi})\boldsymbol{A}$.

Even if $\bar{w}(\boldsymbol{\pi})$ is submodular, i.e., the off-diagonal entries of $\nabla^2\bar{w}(\boldsymbol{\pi})$ are all negative, it is still possible to construct matrix $\boldsymbol{A}$ such that $\boldsymbol{A}^T\nabla^2\bar{w}(\boldsymbol{A}\boldsymbol{\pi})\boldsymbol{A}$ has positive off-diagonal entries. Therefore, by choosing some proper matrix $\boldsymbol{A}$, we can construct non-substitutable choice model $w(\boldsymbol{\pi})$ from substitutable choice model $\bar{w}(\boldsymbol{\pi})$. We call this method the *crossing transformation* and the corresponding matrix $\boldsymbol{A}$ the crossing matrix. Particularly, it is possible that the original choice model $\bar{\boldsymbol{q}}(\cdot)$ is a substitutable choice model, while $\boldsymbol{q}(\cdot)$ is no longer substitutable.

In the following, we give an example of constructing a non-substitutable choice model from the MNL model using the crossing transformation. Note that this example was also in Feng et al. [12] as an example to show that the welfare-based choice model strictly subsumes the random utility model even when there are only three alternatives.

Example 3. Let $\bar{w}(\boldsymbol{x}) = \log(e^{x_1} + e^{x_2} + e^{x_3} + e^{x_4})$ be the choice welfare function for an MNL model for 4 alternatives. Let the crossing matrix $\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0.5 & 0.5 & 0 \end{bmatrix}$. Then the new welfare-based choice model becomes $w(\boldsymbol{\pi}) = \log\left(e^{\pi_1} + e^{\pi_2} + e^{\pi_3} + e^{0.5(\pi_1+\pi_2)}\right)$ with choice probability $\boldsymbol{q}(\boldsymbol{\pi}) = \frac{1}{e^{\pi_1}+e^{\pi_2}+e^{\pi_3}+e^{0.5(\pi_1+\pi_2)}}\left(e^{\pi_1} + \frac{1}{2}e^{0.5(\pi_1+\pi_2)},\right.$ $\left. e^{\pi_2} + \frac{1}{2}e^{0.5(\pi_1+\pi_2)}, e^{\pi_3}\right)$. It is easy to check that $\frac{\partial q_1(\boldsymbol{\pi})}{\partial \pi_2} = \frac{\partial q_2(\boldsymbol{\pi})}{\partial \pi_1}$ is positive if and only if $e^{\pi_3} \geq 4e^{0.5\pi_1+0.5\pi_2} + e^{\pi_1} + e^{\pi_2}$. Therefore, under this choice model, when both π_1 and π_2 are small enough (compared to π_3), alternatives 1 and 2 are complementary. Otherwise, they are substitutable.

4. Concluding remarks

In this paper, we propose and study the concepts of substitutability and complementarity in choice models. Such concepts are fundamental for choice models and are very useful in capturing practical choice patterns. We show how substitutability and complementarity can be reflected in the construction of choice models and thus how to construct choice models with complementarity property. We believe our work is useful for future studies of choice models.

An important future research direction is the estimation problem of the choice model using real data, especially choice models with complementarity property. We expect the common estimation methods would still work, but will leave the details for future study.

References

[1] A. Alptekinoğlu, J. Semple, The exponomial choice model: A new alternative for assortment and price optimization, Oper. Res. 64 (1) (2016) 79–93.
[2] S.P. Anderson, A.D. Palma, J.F. Thisse, Discrete Choice Theory of Product Differentiation, The MIT Press, 1992.
[3] M. Ben-Akiva, Structure of Passenger Travel Demand Models (Ph.D. thesis), Massachusetts Institute of Technology, 1973.
[4] M. Ben-Akiva, S.R. Lerman, Discrete Choice Analysis: Theory and Application to Travel Demand, The MIT Press, 1985.
[5] D. Bertsekas, Convex Analysis and Optimization, Athena Scientific, 2003.
[6] J. Blanchet, G. Gallego, V. Goyal, A Markov chain approximation to choice modeling, Oper. Res. 64 (4) (2016) 886–905.
[7] A. Börsch-Supan, On the compatibility of nested logit models with utility maximization, J. Econometrics 43 (3) (1990) 373–388.
[8] J.M. Borwein, A.S. Lewis, Convex Analysis and Nonlinear Optimization: Theory and Examples, Springer Science & Business Media, 2010.
[9] C. Daganzo, Multinomial Probit: The Theory and Its Application to Demand Forecasting, Academic Press, 1980.
[10] J.M. Davis, G. Gallego, H. Topaloglu, Assortment optimization under variants of the nested logit model, Oper. Res. 62 (2) (2014) 250–273.
[11] V.F. Farias, S. Jagabathula, D. Shah, A nonparametric approach to modeling choice with limited data, Manage. Sci. 59 (2) (2013) 305–322.
[12] G. Feng, X. Li, Z. Wang, Technical note - on the relation between several discrete choice models, Oper. Res. 65 (6) (2017) 1516–1525.
[13] G. Gallego, R. Ratliff, S. Shebalov, A general attraction model and sales-based linear program for network revenue management under customer choice, Oper. Res. 63 (1) (2014) 212–232.
[14] J.A. Herriges, C.L. Kling, Testing the consistency of nested logit models with utility maximization, Econom. Lett. 50 (1) (1996) 33–39.
[15] J. Hofbauer, W.H. Sandholm, On the global convergence of stochastic fictitious play, Econometrica 70 (6) (2002) 2265–2294.
[16] S. Jagabathula, P. Rusmevichientong, A nonparametric joint assortment and price choice model, Manage. Sci. 63 (9) (2017) 3128–3145.
[17] C.L. Kling, J.A. Herriges, An empirical investigation of the consistency of nested logit models with utility maximization, Am. J. Agric. Econ. 77 (4) (1995) 875–884.
[18] D. McFadden, Conditional logit analysis of qualitative choice behavior, in: Frontiers in Econometrics, Academic Press, 1974, pp. 105–142.
[19] D. McFadden, Quantitative Methods for Analyzing Travel Behaviour of Individuals: Some Recent Developments. Tech. Rep., Cowles Foundation for Research in Economics, Yale University, 1977.
[20] D. McFadden, Econometric models for probabilistic choice among products, J. Bus. 53 (3) (1980) 13–29.
[21] S. McShane, M. Von Glinow, Organizational Behavior, McGraw-Hill Higher Education, 2015.
[22] K. Murota, DIscrete Convex Analysis, Society for Industrial and Applied Mathematics, 2003.
[23] T. Rockafellar, Conjugate Duality and Optimization, Society for Industrial and Applied Mathematics, 1974.
[24] P. Rosenzweig, The Halo Effect:... and the Eight Other Business Delusions That Deceive Managers, Simon and Schuster, 2014.
[25] D. Simchi-Levi, X. Chen, J. Bramel, Convexity and supermodularity, in: The Logic of Logistics, Springer, 2014, pp. 15–44.
[26] E.L. Thorndike, A constant error in psychological ratings, J. Appl. Psychol. 4 (1) (1920) 25–29.
[27] K. Train, D. McFadden, M. Ben-Akiva, The demand for local telephone service: A fully discrete model of residential calling patterns and service choices, Rand J. Econ. 18 (1) (1987) 109–123.
[28] K.E. Train, Discrete Choice Methods with Simulation, Cambridge University Press, 2009.