Salience-Biased Nested Logit*

Caluz, A. D. — Insper

Faro, J. H. — Insper

Sanches, F. M. — Sao Paulo School of Economics - FGV

This draft: July 2024

Abstract

This paper proposes an axiomatic-based foundation for salience-driven random choice in a nested logit (NL) framework. In this context, the model features a partition of the set of alternatives (the collection of nests or categories) and utility (or Luce) values assigned to each option. We call it salience-biased nested logit (SBNL), which differs from the standard NL model due to the fact that once the agent faces a menu, salient alternatives, represented in each category, draw the attention of the decision-makers. This modifies an implicit assumption in the usual NL that all payoffs are considered in the category choice. Once the category is chosen, the agent picks one alternative according to the traditional Luce rule, a feature shared by the SBNL and the standard NL model. We compare our axiomatic foundation for this class of random choice functions with the axioms underlying the NL model. We show that this class of models does not obey regularity. We find an "inverse" decoy effect that cannot be captured by the usual Nested Logit. Moreover, under mild assumptions, SBNL satisfies the moderate stochastic transitivity. Beyond the axiomatic approach, we show a derivation of our model in a framework with random utility, presenting the behavioral axioms underlying the SBNL probabilities under modified conditions used in the usual NL. Using a parametric version of the utility, we propose an estimation procedure to estimate the model's parameters using both individual-level and aggregate market data. In this framework, we provide computations for a rich set of price elasticities accounting for the possibility of different marginal effects from changes in salient or common options.

Keywords: Discrete choice, Nested Logit, Stochastic Choice, Salience-Bias.

*We thank P. Natenzon, M. Kovach, G. Teserenjigmid, and seminar/conference participants at EPGE-FGV, FEA-RP/USP, TUS 2021 (U. Paris 1), RUD 2022 (PSE), Lames 2022 (U. del Pacífico) for helpful discussions and comments. Caluz is grateful for financial support from FAPESP (grant 2020/11460-6). José Heleno Faro is also grateful for financial support from CNPq-Brazil (Grant no. 306515/2022-9).

1 Introduction

There is a broad literature on modeling and applying discrete choice and logit-type models to deal with individual choice behavior. The seminal model is due to Luce (1959), which proposed the Luce model, with the behavioral foundation for what is also known as Multinomial Logit (MNL). This primer model has well-known limitations due to unrealistic substitution patterns implied by the model, specifically by the Independence of Irrelevant Alternatives (IIA) axiom. The IIA problem motivated many model variations, such as the prominent nested logit (NL) and Mixed Logit models, which are grounded on more flexible assumptions than the IIA.

In the standard NL model, nests partition the set of alternatives into "similar" goods, leading to more flexible substitution patterns than in the MNL by dealing with the correlation between similar alternatives. The NL model works as if it were a two-stage choice, where the decision-maker first chooses a nest (or category) and then chooses an alternative within the category. One implicit assumption in the NL is that all alternatives within the category contribute positively to the likelihood of the nest. In other words, when comparing categories, the entire payoff into the category matters. However, it could be costly to pay attention to all alternatives of each category when comparing them. Related to this attention issue, some models aim to study salience bias in the choice procedures, meaning that agents focus on prominent information, payoffs, or alternatives.

Salience bias or perceptual salience describes the tendency of individuals who are more likely to focus on alternatives or information that are more prominent and ignore those less so, which creates a bias in favor of striking options or data that are more noteworthy. Agent's behavior may present this bias due to inattention or cognitive limitations in the decision process (Cosemans and Frehen, 2021; Kahneman, 1973). There is robust literature showing the importance of introducing salience in the modeling of consumer choice to rationalize striking evidence of decision theory and behavioral economics. In the context of lotteries, Bordalo et al. (2012), henceforth referred to as BGS, proposes a model in which salience grabs the decision-maker's attention. This behavior in the choice under risk has direct implications for asset prices (Bordalo et al., 2013a; Cosemans and Frehen, 2021). The BGS model is also built for the deterministic framework, where the consumer choice is affected by salient attributes, e.g., prices (Bordalo et al., 2013b).

In this paper, we propose a salience-biased nested logit (SBNL) model of stochastic choice consistent with the idea that the decision-maker categorizes choice alternatives into disjoint sets, called nests, thus, grounded on the same idea of the standard NL model. The decision-maker starts picking, at random, one category, and then, to reach a final decision, she randomly chooses an option within the selected category.

As well-known in the applied literature, some models already fit this two-stage random process, where similar alternatives are aggregated into categories, notably the usual nested logit model[1]. Our interest relies on a two-stage stochastic process in which the decision-maker wishes to avoid stressful/demanding cognitive strain when deciding the category. For the sake of cognitive ease, each nest's likelihood depends on the most prominent alternatives, characterizing salience bias.

Assume a population choosing cars or an agent choosing (stochastically) a car. Initially, each agent selects a car brand. If the agent were to consider all payoffs/utilities of each brand, she would need to process information about all available cars. For instance, if there are two brands denoted by $\{A_1, A_2\}$, with cars $A_1 = \{x, y, z\}$ and $A_2 = \{a, b\}$, in the standard model such as the NL, the agent first compares the payoffs of $v_x + v_y + v_z$ with $v_a + v_b$, thus she needs to know all car payoffs. After selecting a brand, say brand A_2, she compares v_a with v_b. Due to sales campaigns, advertising on social media, and television, certain options within each brand are likely more prominent to decision-makers, offering a higher net payoff in the choice process. To choose a brand (here, assumed to define nests), the agent may initially consider only these salient options of each brand, let's say that they are x and a. After that, she visits the car dealership/store, and learns about other available payoffs within the chosen brand. Consequently, this results in a (probabilistic) choice where the prominent alternatives are selected with a higher probability. How-

[1]Kovach and Tserenjigmid (2022a) proposes a more general model, called nested stochastic choice function. Faro (2023a) proposed a model where alternatives in the same nest are stochastically identical.

ever, since the agent has other options available after the initial nest choice, other similar options are also chosen with some probability. This process underlies the SBNL, where initially, the agent compares $v_x = \max_{k \in A_1} v_k$ with $v_a = \max_{k \in A_2} v_k$, and after choosing a brand, she compares the options within that chosen brand, essentially selecting nests by initially favoring the more prominent options.

Formally, we consider a finite set of choice alternatives X and a decision-maker (DM) that ranks the choice alternatives according to a Luce value mapping $v : X \to (0, \infty)$ and then categorize each one in a different nest described by a partition $\mathbb{A} = \{A_1, ..., A_n\}$ of X. Each element of the partition is associated with a strictly increasing map $f_i : (0, +\infty) \to (0, +\infty)$. Given a menu $A \subseteq X$, the DM's attention is drawn to salient choice alternatives within each nest so that the likelihood of the nest $i \in \{1, ..., n\}$ is given by

$$\frac{f_i\left(\max\limits_{y \in A \cap A_i} v_y\right)}{\sum\limits_{k: A \cap A_k \neq \emptyset} f_k\left(\max\limits_{y \in A \cap A_k} v_y\right)}$$

where the main difference to the usual nested logit is that in the standard model we have that the attention to nests is a function of all alternatives, $f_i\left(\sum\limits_{y \in A \cap A_i} v_y\right)$ instead only the best option, that we call salient alternative.

Using this framework, we define the Salience-Biased Nested Logit (SBNL) as a special case given by the function $f_i(v) = v^{\lambda_i}$, where $\lambda_i > 0$. In the usual nested logit model we have that the probability of choosing an element of the nest i is a function of $\sum\limits_{z \in A \cap A_i} v_z$ instead depending only on the salient option.

Once a nest $i \in \{1, ..., n\}$ is selected, in the second stage, a choice alternative $x \in A \cap A_i$ is chosen with probability given by a Multinomial Logit (MNL) – also known as Luce Model – within alternatives in the same nest

$$\frac{v_x}{\sum\limits_{y \in A \cap A_{i_x}} v_y}.$$

In this paper, we provide minimal behavioral axioms over (stochastic) choice functions to represent them as a salience-biased nested logit, following the Luce (1959) and Kovach and Tserenjigmid (2022a) approach. Moreover, since most applied work is grounded on random utility frameworks, we also show how to derive the model in a framework in which utility is random, and agents are somehow maximizers, following the McFadden et al. (1973), Cardell (1997), and Berry et al. (1995) tradition[2].

In the axiomatic approach, in a nested stochastic choice structure (Kovach and Tserenjigmid, 2022a), we show that if the probability of a category represented in a menu A is not affected by weakly dominated alternatives (within the cattegory) then the probability of an option in the category being chosen is dependent only the best payoff available. Thus, we model and provide axioms to a context-dependent salience in the random choice framework.

In the random utility estimation framework, beyond the derivation of the model using behavioral and distributional assumptions, we propose a parametric version that one can estimate with real data, presenting a simple two-step estimation procedure – and applying it in a simulation – to individual-level choices. We demonstrate how to use the model to get standard parameters of interest in economics and industrial organization applications, such as price elasticity. We show how the implications and conclusions of the salience-biased case differ from the usual nested logit. To be more specific, we present six cases of price elasticities: own-price elasticity, within-nest price elasticity, cross nest, but for each of these three types we have to consider if the option changing its price is a salient option of the nest or not. To understand how salience bias works in this context, we show that if the price of an option of a nest n is changing, but it is not the price of a salient option, then its effect on other nests is null. Moreover, another effect not captured by the standard nested logit is the within-nest cross-price elasticity. The effect of a price change in an option in the same nest always reduces the conditional (to an option in the nest) probability. However, the unconditional probability may affect positively or nega-

[2]Luce (1959) and Kovach and Tserenjigmid (2022a) approach are based on decision theory characterization, with axioms over random choice functions, while McFadden et al. (1973), Cardell (1997), and Berry et al. (1995) are based on a random utility characterization, which constructions are important to match the model with real choice data. In our paper, we build the model in both traditions.

tively. This happens because the option changing its price may be a salient option or not. Thus, an increase in a price may even reduce the share of other options in the same nest if this increase occurs in a salient option. This happens because the overall attraction to the nest is reduced. The final direction of the effect depends on nest parameters as shown in section 5.

To illustrate how the model works and what it implies to applications we show two additional examples. First, consider a scenario where a decision-maker is selecting among different beaches for a vacation, categorizing beaches according to the regions they are in, each served by a common airport. Suppose two beaches at the north, $A_1 = \{x_1, x_2\}$, and three beaches at the south, $A_2 = \{x_3, x_4, x_5\}$ forming nests. She chooses A_1 with probability 1/3 and $\{A_2\}$ with 2/3. But, once she arives at the south, she randomizes between the two beaches going 5 times more frequently to x_1 than to x_2. Similarly, in the north, beach x_5 stands out as the most salient, visited 50% of the time, with beaches x_4 and beach x_3 being chosen 30% and 20% of the time, respectively. This selection pattern aligns with the Salience-Biased Nested Logit model (SBNL) with lambdas equals one, and Luce values (utilities) $v = (5, 1, 4, 6, 10)$, respectively. Thus, beach one is salient in the south, and beach five is salient in the north. To choose between south and north, she only compares $v_1 = 5$ to $v_5 = 10$. [3]

Now, consider the same example, but let's assume that option values are dependent on choice characteristics, such as prices. Denote $v_j = \delta_j + \alpha p_j$, where δ_j is the portion of the utility that depends on characteristics other than prices and α represents the impact of price j on the utility v_j. Thus, the SBNL can be viewed as the demand for beaches, when the menu available is the set of all five beaches. Notice that we can compute the impact of any price change on goods' demand. If a discount or sales campaign reduces the price of the salient beach in south p_5, its value (v_5) increases (assuming $\alpha < 0$, i.e., a usual demand). Thus, the probability of choosing x_5 increases. This price change may affect other goods in both nests. Since x_5 is the salient option, this increase in v_5 increases the overall probability of an option in A_2 being chosen, decreasing the probability of A_1, and then, of all

[3] And since lambdas are one, $Pr(\text{south}) = 5/(5 + 10) = 1/3$ and $Pr(\text{south}) = 2/3$.

options in A_1. However, notice that if one promotes the good x_4 by decreasing its price (i.e., increasing v_4)[4], the price change will affect the demand for x_5, but this does not affect the overall probability of A_2. This occurs since x_4 is not a salient option, and then, it is not taken into account in the first stage of choice. Moreover, changes in x_4 also have a null effect on good in A_1. This result is formally stated in the results of the paper.

Extending to this random utility structure, we provide a simple application using aggregate market data, a frequent context in applications of the usual nested logit model, once it is possible to transform it into simple linear regressions using market shares. The salience-biased nested logit has an analogue transformation. We use this transformation to apply the model to Brazilian flight destination choice, using five country regions as nests, and price tariff and living cost (of the destination) as explanatory variables. We deal with endogeneity by estimating each step using instrumental variables.

Importantly, the salience-biased nested logit (SBNL) model can capture non-regularity in stochastic choice models even when the nest utility function is concave (which the traditional nested logit cannot). Thus, we rationalize the existing empirical evidence of non-regularity through a salience effect (Huber et al., 1982). We provide an example where a product introduction can increase probabilities (demand) of similar alternatives, using a salience-driven probability that differs from the traditional decoy effect concept.

Another significant result is that, under mild conditions (on nest parameters), SBNL satisfies moderate stochastic transitivity, a form of choice transitivity to the random choice context. To show this, we fit the SBNL model on the framework of He and Natenzon (2022), which established a general model called the moderate utility model and showed that any stochastic choice function is a moderate utility model if and only if it satisfies moderate stochastic transitivity. We show that the model is a moderate utility model when for any pair of nests, the nest-specific parameters

[4] This result holds provided that the marginal change in p_4 is sufficiently small such that v_4 is still lower than v_5. A sharp change that modifies the salient option of a nest may affect the demand for all goods available.

satisfy $\lambda_i + \lambda_k \leq 2$.[5] Hence, under this condition, the salience-biased nested logit satisfies moderate stochastic transitivity.

This work relates to the literature modeling bounded rationality in the stochastic choice framework, particularly inattention models. Matějka and McKay (2015) study the presence of rational inattention in discrete choice, providing a new foundation for the Multinomial Logit when the full knowledge of the menu is costly. Kovach and Tserenjigmid (2022b) proposes a focal Luce model weakening the controversial assumption of Independence of Irrelevant Alternatives, capturing bounded rationality for agents that have a bias for focal sets.

Our model differs from the previous by describing a salience weighting in a two-stage choice in which similar alternatives belong to common nests. The behavioral foundation of this type of a two-stage decision – or nested stochastic choice – was provided by Kovach and Tserenjigmid (2022a). The authors built a class of models called Nested Stochastic Choice (NSC). We incorporate salience bias in an NSC structure where attention to categories is driven by context-dependent salience. Faro (2023b) proposes a Luce model with replicas that fit the NSC structure. In this work, nests are composed of identical (in utility terms) alternatives. The Luce model with replicas is a limit case of the nested logit where no salience can emerge from each nest since no variation exists within each nest. Our model allows for this variation with the salience attribute given by the best alternative of each nest, producing non-regularity in a stochastic choice model, which usual Random Utility Models cannot do.

In the next section, we detail the stochastic choice theory with bounded rationality and the salience theory. Then, we present the framework of our model in section (iii). We move to the axiom-based characterization and show how the model can capture non-regularity in section (iv). Section (v) presents the derivation of the model in the random-utility framework, showing its usage and an estimation procedure. We finish discussing the practical implications of applied economic issues.

[5] The NL model constantly asks λs to be lower or equal to one in order to be a random utility model, which is verified by all usual applications in the IO context.

2 Literature Review

There is theoretical and experimental evidence on the behavioral economics literature arguing that agents may not satisfy the traditional rationality concept or that, in some frameworks, one should consider restrictions in agents' capacity. The theory developed different frameworks to incorporate bounded rationality. Acquiring information can be costly (Stigler, 1961; Caplin and Dean, 2015), that is, it can be challenging, or even impossible, to process all information available. The context/restrictions (as the time to choose) may affect how choices are made Gul et al. (2014). Individuals may have cognitive limitations (Camerer et al., 2004), thinking aversion (Ergin and Sarver, 2010; Ortoleva, 2013), and there can exist attention grabbers in the choice problems faced by individuals (Barber and Odean, 2008). These limitations in the individuals' behavior are associated with the bounded rationality concept.

Empirical evidence shows that agents are not fully attentive, simplifying the choice process. This type of behavior leads agents to limit the search scope to an extensive menu of options using some criteria. De los Santos et al. (2012) used a big-data analysis of internet search-shopping data to investigate how deep is the consumers' search in the choice of books to buy. Usually, people go into a few bookstores, and due to this poor search, people often do not purchase in the lowest price bookstore. Indeed, a skewed distribution is observed favoring the most popular stores (e.g., Amazon). Notice that, at least partially, salience can drive this effect. This choice simplification is also observed in experimental frameworks Dean and Neligh (2017).

The economic literature developed new theories aiming to capture this simplification of the consumer's choice process. Tversky and Kahneman (1992) developed the concept of rank-dependent utility in which utility weights are higher for top-ranked alternatives. This theory is known as cumulative prospect theory and incorporates the idea that there is a rank effect on consumer choice. This phenomenon has been documented in the behavioral finance literature (Frydman and Wang, 2020; D'Acunto et al., 2019; Hartzmark, 2015). For instance, individuals tend to trade more top and worst stocks than what is predicted by an optimal portfolio allocation. Hartzmark (2015) argues that the rank effect is driven by salience. Some alternatives are more protuberant in the menu faced by the individual, grabbing individual attention. Moreover, as discussed by Bordalo et al. (2013b) and Cosemans and Frehen (2021), the distortion towards the top-ranked alternatives depends on how salient these alternatives are relative to the others available. That is, the salience attribute is context-dependent, which is a key property in the economic/decision theory of salience.

Studies like Bordalo et al. (2013b) and Cosemans and Frehen (2021) incorporate this salience pattern in the choice model. With salience in agents' utility, the authors provide a unified rationalization to other different behavioral puzzles documented by the decision theory and behavioral economics literature. Some examples are the decoy effect, context-dependent willingness to pay, preference reversal, and others. These models are more general than Tversky and Kahneman (1992) in the way that the utility weights favor salient alternatives, attributes, or payoffs.

Ellis and Masatlioglu (2019) provide minimal behavioral axioms for BGS' utility functions. Moreover, they provide a salience version of the Strong Axiom of Revealed Preference (salience-SARP), which is weaker than the usual SARP assumption. This allows testing whether some empirically observed behavior is consistent with the Salience Choice Model of BGS, showing that it's essential to have a clean behavioral axiomatization of choice models and that those axioms provide testable implications for empirical tests.

Meanwhile, there is scant work on random choice models of salience. Data is noisy, and individuals usually respond differently in similar situations, even in experimental settings, as highlighted by McFadden et al. (1973); Agranov and Ortoleva (2017). Thus, stochastic choice models are considered as the natural approach to fill this gap between theories and real data (Manzini and Mariotti, 2014). Our paper contributes to the literature by incorporating salience preferences in stochastic choice models.

Apart from salience, other general types of puzzling behaviors stand in random choice settings. In Manzini and Mariotti (2014), agents' choice is made through consideration sets, where agents reduce the set of alternatives to a smaller subset. This reduction in the originally available menus is what defines inattention. Moreover, agents may be rationally inattentive by using optimal consideration sets, which defines the concept of rational inattention of Caplin et al. (2019). Cattaneo et al. (2020) modelled this inattention as random. Finally, Aguiar et al. (2021) unifies the random attention framework with rational inattention models. More related to salience, Kovach and Tserenjigmid (2022b) proposes a version of Luce (1959) model in which there is a focal set that imbalances the probabilities in favor of the focal alternatives.

Finally, our model fits in a recent structure of Nested Stochastic Choice (NSC). In this framework, the choice set is composed by a partition of similar alternatives. The elements in the partition are called *Nests*. The most prominent model in the class of models is the nested logit. Kovach and Tserenjigmid (2022a) proposes a behavioral axiomatization to NSC models, weakening the axiom in Luce (1959). Faro (2023b) developed a model – the Luce model with replicas – where alternatives in a nest have identical payoffs, also called replicas. The Luce model with replicas is an extreme case of the nested logit where no variation is allowed within each nest. We allow this variation and model the nest attractiveness by the best alternative of each nest, generating a non-regular behavior that fits the NSC structure. We show that the proposed model can be applied in different settings – as in a random utility framework – or different types of data, contributing to the econometric literature. The next section presents the setup and the model defining how we are introducing salience bias in a nested logit context.

3 Framerwork and Model

Let X be a nonempty set of choice objects. We assume that X is finite. The elements of $2^X \setminus \{\emptyset\}$ are called choice sets (also called menus or choice problems). The mapping $p : 2^X \times \left(2^X \setminus \{\emptyset\}\right) \to [0, 1]$ is a random choice function if for all

$A \in 2^X \setminus \{\emptyset\}$, $p(\cdot, A) : X \to [0,1]$ satisfies $p(x,A) := p(\{x\}, A) = 0 \ \forall x \notin A$ and

$$\sum_{x \in A} p(x, A) = 1.$$

The value $p(x,A)$ is interpreted as the probability of an individual agent (decision-maker), on whose behavior we focus, choosing the alternative x from the menu A. Another common interpretation is that $p(x,A)$ captures the fraction of a given population choosing x from A. A random choice function p is non-degenerate if $p(x,A) \in (0,1)$ for all $A \in 2^X \setminus \{\emptyset\}$ with $card\,(A) \geq 2$ and $x \in A$ (this property is also called full-support assumption).

Recall that a family $\{A_i : i \in I\}$ of subsets of X is a partition if: $A_i \neq \emptyset$ for each $i \in I$; $\cup_{i \in I} A_i = X$; and $A_i \cap A_j = \emptyset$ for all $i \neq j$. Since X is finite, we have only finite partition, denoted by $\mathbb{A} = \{A_1, ..., A_n\}$, where $n \leq card\,(X)$. In this setup, we present our first definition:

Definition 1 *(Salience-Biased Nested Stochastic Choice): We say that a stochastic choice rule p is a salience-biased nested stochastic choice rule if there exists a partition $\mathbb{A} = \{A_1, ..., A_n\}$ of X (note that each $x \in X$ can be associated with a particular $A_{i_x} \in \mathbb{A}$ in which x lies), a set of n increasing maps $f_i : (0, +\infty) \to (0, +\infty)$, and a Luce value function $v : X \to (0, +\infty)$ such that for all $A \in 2^X \setminus \{\emptyset\}$, $x \in A$*

$$p\,(x, A) = \frac{f_{i_x}\left(\max\limits_{y \in A \cap A_{i_x}} v_y\right)}{\sum\limits_{i : A \cap A_i \neq \emptyset} f_i\left(\max\limits_{y \in A \cap A_i} v_y\right)} \frac{v_x}{\sum\limits_{y \in A \cap A_{i_x}} v_y}.$$

A particular case of interest is the salience-biased nested logit model, similar to the usual nested logit model.

Definition 2 *(Salience-Biased Nested Logit) The salience-biased nested logit model is a salience-biased nested stochastic choice with $f_i(v) = v^{\lambda_i}$ for each $i \in \{1, ..., n\}$, where $\lambda_i \geq 0$.*

The main difference between the salience-biased nested logit and the Nested Logit is how agents evaluate the nest utility. In the usual Nested Logit Model, the dependence is not on $\max_{y \in A \cap A_{i_x}} v_y$ but on $\sum_{y \in A \cap A_{i_x}} v_y$. This difference between the SBNL and the NL allows us to capture non-regularities caused by salience that the NL is not able to do even in the case [6].

To illustrate this behavior, suppose an agent choosing among three alternatives x, y, z in different menus $A \subseteq \{x, y, z\}$. Table 1 presents one possible scenario showing how salience bias, as defined by our model, arises in probability choices.

First, notice that when the agent chooses among $\{x, y\}$, she chooses more frequently x than y. In the second line, we see that x and z are both equally probable. However, when introducing z to the menu $\{x, y\}$, we see two important facts: firstly, $p(x, \{x, y, z\})/p(y, \{x, y, z\}) = p(x, \{x, y\})/p(y, \{x, y\})$, which is a characteristic of a nested choice where a nest A_1 is composed by x and y and another nest $A_2 = \{z\}$. Secondly, from the menu $\{x, z\}$ to $\{x, y, z\}$ the probability of alternatives in A_1 is invariant to the introduction of y (Table 1, lines 2 to 4).

Table 1: Example of a Salience Biased Nested Logit behavior

A	p(x,A)	p(y,A)	p(z,A)
$\{x, y\}$	0.800	0.200	0.000
$\{x, z\}$	0.500	0.000	0.500
$\{y, z\}$	0.000	0.333	0.667
$\{x, y, z\}$	0.400	0.100	0.500

Example of a choice with salience bias in the nest probabilities

Indeed, these probabilities can be represented as an SBNL model with $(v_x = 4, v_y = 1, v_z = 8)$ and $(\lambda_1 = 1/2, \lambda_2 = 1/3)$. One possible explanation for this behavior is that from the menu $\{x, z\}$ to $\{x, y, z\}$, we add an alternative that is not salient relative to nest 1. However, introducing x to $\{y, z\}$ increases the probability of an alternative in the first nest being chosen. This happens because $p(x, \{x, y\}) > p(y, \{x, y\})$, that is, x can be viewed as the salient option of nest A_1 in the menu $\{x, y, z\}$. In the usual nested logit, both alternatives would increase the probability of the first nest, provided that a regularity condition, that we explore later, holds.

We note that we obtain the Luce's model (Luce, 1959) when the partition $\mathbb{A}$ is given

[6] The nested logit is regular (set-monotone) when $\lambda_i \leq 1$ for all nest i.

by

$$\mathbb{A} = \{\{x_1\}, ..., \{x_n\}\},$$

where $X = \{x_1, ..., x_n\}$. In this case, it follows that, given $A \in 2^X \setminus \{\emptyset\}$, for all $x \in A$,

$$p(x, A) = \frac{v_x}{\sum\limits_{y \in A} v_y}.$$

A related model is the Luce model with replicas, developed by Faro (2023b). In the Luce model with replicas, nests are formed by alternatives with identical Luce utilities, and then, an agent chooses uniformly within each nest. In this case, introducing a new option in some nest to a menu will not increase the probability of choosing that nest because no change occurs in the nest's salience.

The Luce model with replicas is also characterized by a partition $\mathbb{A} = \{A_1, ..., A_n\}$ where for each nest $\exists v^*_{A_i} \in \mathbb{R}_{++}$ such that for all $A \subseteq X$ and $x \in A$

$$p(x, A) = \frac{v^*_{A_{i_x}}}{card\left(A \cap A_{i_x}\right) \sum\limits_{i: A \cap A_i \neq \emptyset} v^*_{A_i}}.$$

In our model, considering nests with non-identical options, we have the opposite case since only the best option matters to the nest probability. However, even an agent with salience bias would behave exactly as in the Luce model with replicas in a context where nests are composed of identical utilities. Our following proposition shows that both models intersect in this case.

Proposition 1 *Given a salience-biased nested stochastic choice function p represented by with a partition $\mathbb{A} = \{A_1, ..., A_n\}$ $i \in \{1, 2, ..., n\}$, if for all $x \in A_i$ $v_x = v_i$ (i.e., all utilities in the same nests are equivalent), then the salience-biased nested logit is a Luce model with replicas.*

We note that if for each $i \in \{1, ..., n\}$, there exist $v_{A_i} > 0$ s.t. $v_z = v_{A_i}$ for all $z \in A_i$

then

$$p(x,A)=\frac{f_i\left(v_{A_{i_x}}\right)}{\sum\limits_{i:A\cap A_i\neq\emptyset} f_i\left(v_{A_i}\right)}\frac{v_{A_{i_x}}}{\sum\limits_{y\in A\cap A_{i_x}} v_{A_{i_x}}}$$
$$=\frac{v^*_{A_{i_x}}}{card\left(A\cap A_{i_x}\right)\sum\limits_{i:A\cap A_i\neq\emptyset} v^*_{A_i}},$$

where $v^*_{A_i} := f_i\left(v_{A_i}\right) > 0$, which is the Luce model with replicas characterized by Faro (2023b).

Both models belong to a larger class of stochastic choice functions called nested stochastic choice, characterized by Kovach and Tserenjigmid (2022a). A positive stochastic choice rule p is a nested stochastic choice (NSC) if there exists a partition $\mathbb{A}=\{A_1,...,A_n\}$ of X (note that each $x\in X$ can be associated with a particular $A_{i_x}\in\mathbb{A}$ in which x lies), a utility $v:X\to(0,+\infty)$, and a set-function $V:2^X\to[0,+\infty)$, with $V(\emptyset)=0$ such that for all $A\in 2^X\backslash\{\emptyset\}$, $x\in A$

$$p(x,A)=\frac{V\left(A\cap A_{i_x}\right)}{\sum\limits_{i:A\cap A_i\neq\emptyset} V\left(A\cap A_i\right)}\frac{v_x}{\sum\limits_{y\in A\cap A_{i_x}} v_y}.$$

We note that the model we characterize is an NSC with n nests where for each nest $i\in\{1,...,n\}$

$$V\left(A\cap A_i\right)=f_i\left(\max_{y\in A\cap A_{i_x}} v_y\right)$$

4 Axioms and Main Results

To build an axiomatic-based characterization for the SBNL, we start defining the Independence of Irrelevant Alternatives (IIA), the core axiom used by Luce (1959) to characterize the Luce model, also known as Multinomial Logit model:

Definition 3 *(Independence of Irrelevant Alternatives): A stochastic choice function p satisfies IIA between $x,y\in X$ if for all $A\subseteq X$ such that $x,y\in A$*

$$\frac{p(x,\{x,y\})}{p(y,\{x,y\})}=\frac{p(x,A)}{p(y,A)}$$

We say that p satisfies IIA when it holds for all pairs in $x,y\in X$.

From IIA we define the *similarity relation* commonly used in the literature (Li and Tang, 2016; Kovach and Tserenjigmid, 2022a):

Definition 4 *For any alternatives $a, b \in X$, we say that a and b are similar, denoted by $a \sim_p b$, if p satisfies IIA at a and b. We also say that a and b are dissimilar if $a \not\sim_p b$.*

As stated previously, the most remarked property of the NL model is that IIA only holds within each nest. The similarity relation above is extremely useful for defining nests.

Given an alternative $x \in X$, we denote $[x] = \{y \in X : y \sim_p x\}$. As it is well-known that a collection of sets $\{[x] : x \in X\}$ forms a partition of X if and only if $\sim_p$ is an equivalence relation. Li and Tang (2016) show that, in general, the relation $\sim_p$ is not an equivalence relation because it may fail to satisfy transitivity.

Kovach and Tserenjigmid (2022a) introduce an axiom called *Independence of Symmetric Alternatives* (**ISA**) that guarantees the transitivity of $\sim_p$ (so, it is an equivalence relation under ISA) that we present below:

Axiom 1 *(Independence of Symmetric Alternatives) For any $A \subseteq X$, $x, y \in A$, and $z \notin A$:*

$$\begin{array}{c} x \sim_p z \wedge y \sim_p z \\ or \\ x \not\sim_p z \wedge y \not\sim_p z \end{array} \quad \Rightarrow \quad \frac{p(x, A)}{p(y, A)} = \frac{p(x, A \cup \{z\})}{p(y, A \cup \{z\})}.$$

Kovach and Tserenjigmid (2022a) show that a positive stochastic choice function satisfies ISA if and only if it can be represented by a nested stochastic choice (NSC) model, defined in the previous section. Thus, we can see that this is the minimal model where IIA holds within each nest, not the nested logit, which requires more axioms[7].

As said previously, our model is an NSC, and then we need axiom 1 plus other assumptions to build our model with salience bias. Next, we present the main axiom characterizing salience bias in the nested choice function:

Axiom 2 *(Neutrality of weakly dominated alternatives within categories) Let x, y be two alternatives such that $x \sim_p y$ and $p(x, [x]) \geq p(y, [x])$. For all $A \subseteq [x]$ s.t. $x \in A$ and $B \subseteq [x]^c$,*

$$p(A, A \cup B) = p(A \setminus \{y\}, (A \setminus \{y\}) \cup B).$$

[7]See Kovach and Tserenjigmid (2022a) for details.

Intuitively, the salience option of a nest A_i relative to a menu is the only one that matters to predict the attention of individuals to options in that nest. We can characterize the generalized salience-biased nested stochastic choice with a flexible functional form for nest probabilities with this axiom.

Theorem 1 *(salience-biased Nested Stochastic Choice) A positive stochastic choice function p satisfies **Independence of Symmetric Alternatives** and **Neutrality of weakly dominated alternatives within categories** if and only if there exist a partition $\mathbb{A} = \{A_1, ..., A_n\}$ of X ($A_{i_x} \in \mathbb{A}$ is s.t. $x \in A_{i_x}$), a collection of n strictly increasing maps $f_i : (0, +\infty) \to (0, +\infty)$, and a Luce value function $v : X \to (0, +\infty)$ such that for all $A \in 2^X \setminus \{\emptyset\}$ and $x \in A$*

$$p(x, A) = \frac{f_{i_x}\left(\max_{y \in A \cap A_{i_x}} v_y\right)}{\sum_{i: A \cap A_i \neq \emptyset} f_i\left(\max_{y \in A \cap A_i} v_y\right)} \frac{v_x}{\sum_{y \in A \cap A_{i_x}} v_y}.$$

Finally, we need to obtain a specific shape for the nest attractiveness function in order to obtain the salience-biased nested logit. The axiomatic foundation to the SBNL, from definition 2, is based on an additional technical condition in the same spirit of the log ratio invariance proposed by Kovach and Tserenjigmid (2022a). We present this axiom in Appendix A and call it *salience log-ratio invariance.*

Alternatively, section 5 presents a foundation to the SBNL in the Random Utility framework. In this case, even not imposing the axiom 3, we can obtain a random choice model equivalent to the SBNL. To do this, we use our central axiom, the axiom 2, and use distributional and optimization assumptions similar to the standard nested logit model.

The salience-biased nested logit model differs from the Nested Logit changing the nest utility function. In the nested logit the nest utility is given by $(\sum_{y \in A \cap A_i} v_y)^{\lambda_i}$, that is, all utility $\sum_{y \in A \cap A_i} v_y$ in the same nest count for the attractiveness of the nest, while in our model only $\max_{y \in A \cap A_i} v_y$ matters. Similarly, the General salience-biased nested logit changes the Exteded Nest Logit model, provided by Kovach and Tserenjigmid (2022a), in which the attractiveness of the nest is given by a more general function $f_i(\sum_{y \in A \cap A_i} v_y)$, while in the salience-biased version, only $f_i(\max_{y \in A \cap A_i} v_y)$ is important. This change is able to capture some behavioral phenomena that we discuss next.

4.1 On Regularity

A basic property satisfied by the Multinomial Logit Model is regularity. Roughly speaking, regularity states that if we add new options to the menu, the prior probabilities of each good must decrease. The usual Nested Logit model satisfies regularity when the nest parameters are all lower (or equal) to one.

More formally, a stochastic choice function p satisfies regularity when $\forall x \in A$,

$$A \subseteq B \implies p(x, B) \leq p(x, A) \tag{1}$$

In the salience-biased nested logit, after the choice of the nest, the agent chooses among alternatives within that nest as in an MNL, and then this second-stage probability decreases with the addition of new similar options. However, the nest probability is non-decreasing and has a non-smooth dependence on the Luce value of the salient option, which can move choice probability upwards when we add a new salient option in that nest.

There is evidence that the decision-maker's behavior can violate regularity. The classical violation of the regularity is known as the decoy effect. This phenomenon is due to Huber et al. (1982) and occurs when an *inferior* alternative can increase the probability of the dominant option relative to its competitor. The SBNL presents a different type of violation by allowing the probability of some alternatives to increase with more options because those options can sharply increase the attractiveness of their nests. Indeed, if we add a new salient option to the menu, that is, the new option has a higher relative choice probability relative to its similar options, the probability of the nest will increase. As we can show, this can lead to an increase in the probability of some alternative x.

The Nested Logit Model is regular provided that $\lambda_i \leq 1, \forall i$ which is the case that $(\sum_{y \in A \cap A_i} v_y)^{\lambda_i}$ is concave. Analogously, in the Extended Nested Logit[8] the regularity holds when $f_i(\sum_{y \in A \cap A_i} v_y)$ is concave, as shown by Kovach and Tserenjigmid (2022a). As we argue, there are cases in which the regularity is violated in the salience-biased nested logit, and more importantly, this may happen even with concave nest utilities. To make it more concrete, we provide an example.

Example 1 *Let $X = \{x_1, x_2, x_3, x_4, x_5, y_1, y_2\}$ and suppose a Stochastic Choice Function*

[8] Kovach and Tserenjigmid (2022a) define the extended nested logit by changing the nest attractiveness $(u)^{\lambda_i}$ by a general increasing function $f_i(u)$.

p over a consumption set . Suppose that p can be represented by a salience-biased nested logit with a partition $\{A_1, A_2\}$ *with* $A_1 = \{x_1, x_2, x_3, x_4, x_5\}$ *with respective Luce Values* $\{1/2, 1, 3/2, 2, 10\}$, $A_2 = \{y_1, y_2\}$ *with respective Luce Values* $\{100, 400\}$, *and Nest specific functions* $f_1(x) = x$ *and* $f_2(x) = x^{1/2}$.

Suppose a menu $A = X \setminus \{x_5\} = \{x_1, x_2, x_3, x_4, y_1, y_2\}$. *Notice that* $\max\limits_{u \in A_1 \cap A} v_u = v_{x_4} = 2$. *Now, computing* $p(x_2, A)$ *we get:*

$$\begin{aligned} p(x_2, A) &= \frac{2}{(\sqrt{400} + 2)} \frac{1}{(1/2 + 1 + 3/2 + 2)} = \\ &= \frac{2}{22} \frac{1}{5} = \\ &= \frac{1}{55} \end{aligned}$$

Now, adding x_5 *to the menu we have that, the salience of the nest* 1 *changes by* $\max\limits_{u \in A_1 \cap A} v_u = v_{x_5} = 5$, *then:*

$$\begin{aligned} p(x_2, A \cup \{x_5\}) &= \frac{10}{(20 + 10)} \frac{1}{(1/2 + 1 + 3/2 + 2 + 10)} = \\ &= \frac{10}{30} \frac{1}{15} = \\ &= \frac{1}{45} > \frac{1}{55} = p(x_2, A) \end{aligned}$$

Hence, p does not satisfy regularity.

The example shows that a salience-biased nested logit can capture regularity violations even in the concave case, while the Nested Logit model does not.

To understand which type of regularity this model can capture, we use two types of regularity defined in Kovach and Tserenjigmid (2022a):

Definition 5 *Given* $A \subseteq X$, $a \in A$, *we say that p satisfies Dissimilar Regularity if for all* $b \notin A$ *such that* $b \notin [a]$, $p(a, A \cup \{b\}) \leq p(a, A)$. *Moreover, it satisfies Similar Regularity when for all* $\hat{b} \notin A$ *such that* $\hat{b} \in [a]$, $p(a, A \cup \{\hat{b}\}) \leq p(a, A)$.

Notice that the SBNL model satisfies *Dissimilar Regularity.* However, as we show in the example 1 the model does not satisfy *Similar Regular.* Intuitively, we may have this violation is when we add an option y that becomes the salient option of the nest $[x]$. In this case, y decreases the conditional probability of options in $[x]$ but increases the nest $[x]$ probability.

The Regularity property is also important because all Random Utility Models are regular (or monotonic). Then, in general, the SBNL is not a RUM. The traditional Nested Logit model is a Random Utility Model when it is regular, i.e., $\lambda_i \leq 1$.

A strand of the literature constructs menu-dependent stochastic choice models to different model information brought by different menus. Thus, there exist models that are menu-dependent Random Utility. One important stochastic model that is equivalent to a menu-dependent utility model is the Cross-nested logit[9].

We can show that the SBNL is a cross-nested logit and then a menu-dependent Random Utility model with a specific type of dependence. The nest parameters depend precisely on the way it causes salience bias in nest probabilities. That is, each menu brings different information with its alternatives.

4.2 SBNL and Stochastic Transitivity

An important property of the usual choice framework is transitivity. This consistency assumption is one of the properties characterizing rationality in the deterministic choice framework. Researchers have been working in different types of transitivity in the random choice set-up, called stochastic transitivity (Rieskamp et al., 2006; He and Natenzon, 2022). Recently, He and Natenzon (2022) proposed a model that entails all stochastic choice functions satisfying the moderate version of stochastic transitivity, this model is called Moderate Utility Model (MUM).

A stochastic choice function p satisfies moderate transitivity when for all x, y, z, $\min\{p(x,y), p(y,z)\} \geq 1/2 \implies p(x,z) > \min\{p(x,y), p(y,z)\}$ or $p(x,z) = p(x,y) = p(y,z)$. He and Natenzon (2022) show that a random choice p satisfies moderate transitivity if and only if it is a Moderate Utility Model, i.e., if there exists $v : X \to \mathbb{R}$, a metric $d : X \times X \to \mathbb{R}_+$, and a strictly increasing F, such that for every $x \neq y \in X$, we can write $p(x,y) = F\left(\frac{v_x - v_y}{d(x,y)}\right)$.

Assume a salience-biased nested logit model with $u_x = e^{v_x}$, and let $\tilde{v}_x = v_x/\lambda_k$, if $x \in A_k$ (i.e. x belongs to the nest k). Then:

$$p(x,y) = \frac{e^{\tilde{v}_x/\lambda_k}}{e^{\tilde{v}_x/\lambda_k} + e^{\tilde{v}_y/\lambda_k}}$$

when $x, y \in A_k$, and

$$\frac{e^{\tilde{v}_x}}{e^{\tilde{v}_x} + e^{\tilde{v}_y}}$$

[9]See Kovach and Tserenjigmid (2022a) to details.

when $x \notin [y]$.

Thus, after few manipulations, we can show that

$$p(x,y) = \frac{1}{1 + e^{-\frac{(\tilde{v}_x - \tilde{v}_y)}{d(x,y)}}}$$

where $d(x,y) = \lambda_k$ when $x, y \in A_k$ and $d(x,y) = 1$ when $x \notin [y]$. Since $F(t) = \frac{1}{1+e^{-at}}$ is increasing in t (for $a \geq 0$) then we can apply the characterization of Moderate Utility Models (MUM) of He and Natenzon (2022) to state that if $p(\cdot)$ is a SBNL then it is a MUM and, hence, it satisfies the moderate transitivity, provided that $d(\cdot)$ is a metric. Thus, although our model is not regular which does not impose restrictions on λs, transitity may impose it. To see this, take $x, y, z \in X$, then if $x \in [z] = A_k$ and $y \notin [z]$ then $d(x,z) = \lambda_k$, and $d(x,y) + d(y,z) = 2$. Thus, in order to $d(\cdot)$ be a metric, we need to impose $\lambda_k \leq 2, k = 1, 2, ..., n$ [10].

Thus, $d(\cdot)$ is a metric and our model satisfies moderate transitivity when $\lambda_k \leq 2$ for all nest k. Notice that this is a weak restriction if the researcher is interested in transitivity. The usual nested logit imposes $\lambda \leq 1$, thus moderate transitivity is satisfied. Moreover, as shown by He and Natenzon (2022), the EBA and the usual Luce rule are also MUM.

5 Salience-Biased Nested Logit in Practice

This section discusses the practical application of the nested logit with salience bias. To do this, we consider the model as a modified Random Utility Maximization, in a framework similar to the derivation of the standard nested logit problem. We show how choice probabilities with salience can be derived within the nested logit framework by slightly modifying optimization assumptions. We then develop a simple estimation strategy to a parametric version of the model and present a Monte Carlo simulation example. Most of the results shown in this section are standard and widely used in the Random Utility literature. We refer the readers to this literature for proofs and other technical details behind our derivations. The derivation of the model can be done similarly to the standard nested logit model. However, as shown previously, the main difference between both models relies on nest probabilities. In the usual nested logit, all payoffs are taken into

[10] all other cases the triangular inequality $d(x,z) \leq d(x,y) + d(y,z)$ can be easily verified. Moreover, for binary choices, our model can be written as a usual nested logit, which satisfies the moderate transitivity, as shown by He and Natenzon (2022).

account to compute the probability of a nest being chosen against others. In the salience-biased nested logit, only the option with the highest payoff plays a role in this first-stage choice. In the second stage, the choice within the nest, both models are identical and work as the Luce rule.

Thus, we can view this model as if the decision-maker first chooses a nest, simplifying his choice by using only options that are salient to her. Then she chooses as a usual logit (Luce) model between options in the surviving nest. in this section, we first show the step-by-step derivation of the usual nested logit. Thus, we modify an optimization assumption, based on our previous axioms, to characterize the model in a random utility estimation framework. After that, we provide practical procedures to apply the model using both individual-level (or purchase-level) and aggregate market-level data, as in common Industrial Organization applications. Moreover, we derive a rich set of price elasticities implied by the model, highlighting the role of salience in the substitution patterns.

5.1 Salience Bias in a Random Utility Estimation Framework

The Random Utility framework is the backbone of empirical models of discrete choice and is widely used in many fields (McFadden, 1977; 2001). Typically, in empirical studies involving the analysis of discrete choices, researchers derive choice probabilities – or individual demand functions – from some basic assumptions defining (i) the individual choice set, (ii) the indirect utility individuals attribute to each alternative contained in the choice set, and (iii) a behavioral assumption that describes how agents make choices. Different discrete choice models such as the Multinomial Logit, the standard Nested Logit, and the Random Coefficients Logit are derived from variations of these three assumptions (Anderson et al., 1992; Train, 2009; Ben-Akiva and Lerman, 2018).

Choice Set. Following this literature, we start by defining decision-marker choice sets. Employing a common notation used in empirical studies of discrete choices, let C be a set of alternatives available to decision maker i.[11] We assume that C exhibits the following properties:

- Assumption C1: C is finite, discrete, and exhaustive (i.e. all possible alternatives

[11] In principle, C can vary across individuals. To simplify the notation, we assume that the choice set is the same for all individuals.

available to individual i are included in C);

- Assumption C2: C can be partitioned into $M < |C|$ subsets C_m (also called nests), $m \in \{1, 2, 3, ..., M\}$, such that $\bigcup_{m=1}^{M} C_m = C$ and $C_m \cap C_{m'} = \emptyset$ for any m, $m' \in \{1, 2, 3, ..., M\}$, $m \neq m'$.

Indirect Utility Function. Individual i derives his random utility U_{ij_m} from any alternative $j_m \in C_m$. We also assume that:

- Assumption U1: $U_{ij_m} = V_{j_m} + \xi_m + \xi_{ij_m}$, where V_{j_m} is a choice specific utility index, ξ_m is a nest specific index and ξ_{ij_m} is an individual-choice specific index;
- Assumption U2: ξ_{ij_m} is independent and identically distributed across individuals and choices and has distribution Extreme Value type I with scale parameter $\mu_m > 0$.

We highlight that only agent observes U_{ij_m}, but the econometrician only observes V_{j_m} and know the distribution of the shocks ξ_m and ξ_{j_m}.

Individual Behavior. Assumptions C1-C2 and U1-U2 are commonly used for the derivation of the standard nested logit within the Random Utility Maximization framework – see, for instance, Ben-Akiva and Lerman (2018). The main difference between the nested logit and the Salience-Biased nested logit is the assumption defining individuals' choice behavior. To illustrate this point, we decompose individuals' decision processes into two steps. In the first step, individual i chooses a nest C_m; in the second, given the choice of the nest, she selects an alternative within that nest.

Now, notice that we can write

$$P(a_i = j_m) = P(a_i \in C_m) \times P(a_i = j_m | a_i \in C_m)$$

which allows the reasearcher to solve the two problems separetely.

Solving this problem backwards, we assume that in each nest C_m, individual i chooses the alternative $j_m \in C_m$ that maximizes her utility. In other words, conditional on the choice of the nest, we assume that individual i chooses alternative $j_m \in C_m$ if the alternative gives her the highest alternative among all the alternatives in nest m. More formally, let $a_i \in C$ represent individual i's choice. Our assumption B1 defines the set of events where agent i chooses $j_m \in C_m$:

- Assumption B1: Given an event such that $a_i \in C_m$, then $a_i = j_m$ if and only if $U_{ij_m} \geq U_{ij'_m}$ for any $j'_m \in C_m$, $j'_m \neq j_m$.

Under assumptions B1, C1, and C2, the choice probability, conditional to the choice of the nest, has the standard Multinomial Logit formula (McFadden et al., 1973):

$$P(a_i = j_m | a_i \in C_m) = \frac{e^{\mu_m V_{j_m}}}{\sum\limits_{k_m \in C_m} e^{\mu_m V_{k_m}}}, \tag{2}$$

where, $P(a_i = j_m | a_i \in C_m)$ represents the probability of choosing alternative j_m given that the choice set is restricted to nest C_m.

Based on this result, we derive now $P(a_i \in C_m)$ – i.e., the probability of choosing an alternative that is in nest C_m – consistent with salience bias. But first, we present the standard nested logit derivation and we restrict the assumptions to derive the SBNL model.

In the standard nested logit framework (Ben-Akiva and Lerman, 2018), the set of events such that individual i chooses any alternative in nest C_m are defined by the events where $\max\limits_{j_m \in C_m} U_{ij_m} \geq \max\limits_{j_{m'} \in C_{m'}} U_{ij_{m'}}$ for or any $C_{m'} \in C$, $m' \neq m$ or, substituting the utility function, $a_i \in C_m$ if and only if:

$$\xi_m + \max_{j_m \in C_m} \{V_{j_m} + \xi_{ij_m}\} \geq \xi_{m'} + \max_{j_{m'} \in C_{m'}} \{V_{j_{m'}} + \xi_{ij_{m'}}\}, \ \forall m' \neq m. \tag{3}$$

Given assumption U2, it is easy to show that $\max\limits_{j_m \in C_m} \{V_{j_m} + \xi_{ij_m}\}$ has distribution Extreme Value with scale parameter μ_m and location parameter given by:

$$\tilde{V}_m = \frac{1}{\mu_m} ln \left[\sum_{j_m \in C_m} exp\left(\mu_m V_{j_m}\right) \right].$$

The probability $P(a_i = j_m | a_i \in C_m)$ consistent with the nested logit is thus derived assuming that ξ_m has a probability distribution such that $\xi_m + \max\limits_{j_m \in C_m} \{V_{j_m} + \xi_{ij_m}\}$ has distribution Extreme Value with location parameter $\hat{V}_m$ and scale parameter equal to 1 – see Cardell (1997) for a proof of the existence of this distribution. Using these assumptions and the set of inequalities (3), the probability of choosing an alternative in nest m that is consistent with the nested logit Model is given by – see, for example, Ben-Akiva and

Lerman (2018):

$$G\left(a_i \in C_m\right) = \frac{exp\left(\tilde{V}_m\right)}{\sum_{k=1}^{M} exp\left(\tilde{V}_k\right)}. \tag{4}$$

Now, let's consider the SBNL case. First, define $V_m^* = \max_{j_m \in C_m} V_{j_m}$ (and remember that C is finite, so is C_m). Suppose that:

- Assumption SB1: $|V_m^*| = 1$, i.e., the highest payoff of each nest is unique.

Let $j_m^* = arg \max_{j_m \in C_m} P(a_i = j_m | a_i \in C_m)$ be the salient option in nest C_m. Notice that, given SB1, $arg \max_{j_m \in C_m} P(a_i = j_m | a \in C_m) = arg \max_{j_m \in C_m} V_{j_m}$ (and we denote $V_m^* = V_{j_m^*}$), which means that the salient option in nest m is the option with the best-observed characteristics in that nest.

Notice that, using axiom 2 we can write

$$P(a_i \in C_m) = P(j_m^*, \{j_1^*, j_2^*, ..., j_M^*\}) = P(V_{j_m^*} + \varepsilon_{j_m^*} + \xi_m \geq V_{j_k^*} + \varepsilon_{k_n} + \xi_n \ : \ \forall n \neq m)$$

This, toguether with distributional assumption, lead to the SBNL representation in the RU estimation framework:

- Assumption SB2: an event satisfies $a_i \in C_m$ if and only if

$$\xi_m + \max_{j_m \in E(C_m)} \{V_{j_m} + \xi_{ij_m}\} \geq \xi_{m'} + \max_{j_{m'} \in E(C_{m'})} \{V_{j_{m'}} + \xi_{ij_{m'}}\}, \ \forall m' \neq m.$$

  where $E(C_m) = \arg \max_{j_m \in C_m} V_{j_m}$.

- SB3: ξ_m is such that $\xi_m + \xi_{j_m^*} \sim EV(1,0)$, where $\xi_{j_m^*}$ is the good specific shock associated with the salient option of nest m[12].

Thus, under SB1-SB3:

$$\begin{aligned} P(a_i \in C_m) &= P(\xi_m + \max_{j_m \in E(C_m)} \{V_{j_m} + \xi_{ij_m}\} \geq \xi_{m'} + \max_{j_{m'} \in E(C_{m'})} \{V_{j_{m'}} + \xi_{ij_{m'}}\}, \ \forall m' \neq m) \\ &= \frac{e^{V_m^*}}{\sum\limits_k e^{V_k^*}} \end{aligned} \tag{5}$$

Clearly, the key difference between the nested logit and the Salience-Biased nested logit is summarized by assumption SB2. Intuitively, considering that choice is a two-step

[12]See Cardell (1997) to the existence of such ξ_m.

decision problem, assumption SB2 implies that in our model, individuals exhibit a specific form of bounded rationality: before visiting a nest C_m, individuals only pay attention to the subset of salient options of each nest, $E(C_1) = j_1^*, ...E(C_m) = j_m^*$. After choosing the nest, individuals are able to select an option as a random utility maximizer within each nest. Hence, the assumption in the nest choice is in the spirit of the inattention and consideration sets literature (Manzini and Mariotti, 2014; Matějka and McKay, 2015; Cattaneo et al., 2020).

Finally, using the Bayes rule and equations (2) and (5) we can derive the Salience-Biased nested logit choice probability:

$$P(a_i = j_m) = \frac{\left(\max\limits_{j_m \in C_m} e^{V_{j_m}}\right)}{\sum\limits_k \left(\max\limits_{j_k \in C_k} e^{V_{j_k}}\right)} \frac{e^{\mu_m V_{j_m}}}{\sum\limits_{j'_m \in C_m} e^{\mu_m V_{j'_m}}} \tag{6}$$

$$= \frac{\left(\max\limits_{j_m \in C_m} e^{\bar{V}_{j_m}}\right)^{\lambda_m}}{\sum\limits_k \left(\max\limits_{j_k \in C_k} e^{\bar{V}_{j_k}}\right)^{\lambda_k}} \frac{e^{\bar{V}_{j_m}}}{\sum\limits_{j'_m \in C_m} e^{\bar{V}_{j'_m}}} \tag{7}$$

in which we transform $\lambda_m = \frac{1}{\mu_m}$ and $\bar{V}_{j_m} = \frac{V_{j_m}}{\lambda_m}$.

This derivation of the SBNL under the Random Utility framework can be summarized by the following proposition:

Proposition 2 *Under the conditions above, if $\varepsilon_{j_m} \sim EV(0, \mu_m)$ for all alternatives j_m and ξ_m is such that $\varepsilon_{j_m^*} + \xi_m \sim EV(0, 1)$ then:*

$$P(a_i = j_m) = \frac{\left(\max\limits_{j_m \in C_m} e^{\bar{V}_{j_m}}\right)^{\lambda_m}}{\sum\limits_k \left(\max\limits_{j_k \in C_k} e^{\bar{V}_{j_k}}\right)^{\lambda_k}} \frac{e^{\bar{V}_{j_m}}}{\sum\limits_{j'_m \in C_m} e^{\bar{V}_{j'_m}}} \tag{8}$$

where, $\lambda_m = \frac{1}{\mu_m}$ and $\bar{V}_{j_m} = \frac{V_{j_m}}{\lambda_m}$.

In the next subsection, we discuss the estimation of the SBNL model, where one needs to estimate $\{\mu_m\}_{m=1}^M$ and $\{V_{j_m}\}_{j_m \in C_m, m=1,2,...M}$, when Vs are function of product/choice caracteristics.

5.2 Estimation of Salience-Biased Nested Logit Models

This section develop a simple estimation strategy for the Salience-Biased nested logit probabilities represented by equation (7). We base our estimation procedure on the same procedure of two-step estimation for the NL, but with particular differences due to the construction of the SBNL.

We consider a simple case where

$$V_{j_m} = x'_{j_m}\beta$$

where, x_{j_m} is a L-dimensional vector of observed characteristics of alternative j_m and β is a L-dimensional vector of parameters. We further assume that the researcher observes a cross-section $D_N = \{a_i, x'_i\}_{i=1}^N$ of N individual choices ($a_i \in C$) and the K-dimensional vector of characteristics of the choices made by each individual (x'_i).[13] The goal of the researcher is to consistently estimate β and μ_m, $m = 1, 2, ...M$ from the data D_N.

The key issue behind the estimation of the Salience-Biased nested logit is to identify the salient option from choices, input this information into the estimation procedure, and then, estimate the model parameters. Notice that the option with the highest value in the nest (the salient option) must also be the option chosen with the highest probability. Thus, we identify salient options as the options with the highest probability in each nest. Using this idea, we present a step-by-step method to recover β and $\{\mu_m\}_{m=1}^M$. Since equation (7) is not differentiable in a set of points, in principle, the Maximum Likelihood estimator commonly employed to estimate the standard nested logit cannot be directly used to estimate the nested logit with salience. Our proposed estimation strategy is outlined below:

1) First, for each nest, select $j_m^* = arg \max P(j_m|C_m)$. Using these salient options, run a logit model for $P(a \in C_m)$, recovering β from the first stage:

$$P(a_i \in C_m) = \frac{e^{x'_{j_m^*}\beta}}{\sum_{n=1}^M e^{x'_{j_n^*}\beta}}.$$

2) Using the estimate $\hat{\beta}$, compute $\hat{V}_{j_m} = x'_{j_m}\hat{\beta}$.

[13]For simplicity, we focus here on cross-sectional data. The procedures we develop here can be trivially extended to accommodate other types of data structures.

3) For each nest m, run a logit model for $P(a_i = j_m|C_m)$ to recover μ_m, using $\hat{V}_{j_m}$ a explanatory variable:

$$P(a_i = j_m|a_i \in C_m) = \frac{e^{\mu_m \hat{V}_{j_m}}}{\sum\limits_{k_m \in C_m} e^{\mu_m \hat{V}_{k_m}}}.$$

This procedure can be easily implemented using common Logit commands available in many statistical software. However, in this two-step estimation, since we are using estimated $\hat{\beta}$ from the first stage, it can carry an inefficiency in the estimation. Moreover,

To illustrate the properties of this estimator, we conducted a simple Monte Carlo experiment.[14] We assumed that:

$$U_{i,j_m} = \beta_0 + \beta_1 x_{i,j_m,1} + \beta_2 x_{i,j_m,2} + \xi_{i,m} + \xi_{i,j_m},$$

where, $x_{j_m,1}$ and $x_{j_m,2}$ are 2 observed (simulated) characteristics of choices, $\beta \equiv (\beta_0, \beta_1, \beta_2)$ is the vector of parameters to be estimated joint to the nest specific parameters, ξ_{j_m} is iid Extreme Value across individuals with scale parameter μ_m, ξ_m satisfies the previous assumption in this section. We further assume that each nest has three products and the number of nests is equal to 2. Using the SBNL probabilities in equation (7) we simulated 1000 samples of N observations and estimated the parameters of the model using the procedure described above. We compute the sample mean squared error (MSE) of each parameter using the SBNL probabilities.

Table 2 shows the results for samples with $N = 100$, $N = 500$, $N = 1000$ and $N = 10000$ observations. The first column has the true value of the parameters. The columns in the table show the MSE results of the estimation using the SBNL estimator. Moreover, we present the resulting MSE of the second stage using the true β to recover $\hat{V}$. The MSE decreases quickly when the sample size increases. However, notice that the parameters μ_1 and μ_2 have a significantly lower MSE when the estimation does not carry the first stage error. Thus, there is an inefficiency in the two-stage procedure proposed. Overall, the results in the table suggest that the estimation procedure developed in this paper works well even when the sample size is moderate.

[14] This two-step estimator is analogous to the two-step estimator for the traditional NL model as proposed in McFadden (1981). The estimator is consistent and $\sqrt{N}$-asymptotically normal under general conditions (McFadden, 1981). In the second step, however, the variances of μ_m have to be corrected to account for the first step estimates of β – see the appendix in McFadden (1981) for details.

Table 2: Monte Carlo Experiment Results

True Parameter		N=100	N=500	N=1000	N=10000
		Mean Squared Error			
β_1	-2	0.537	0.100	0.044	0.004
β_2	3	0.746	0.118	0.054	0.005
		Sample MSE			
μ_1	1	0.096	0.018	0.009	< 0.001
μ_2	1	0.092	0.019	0.008	< 0.001
		Sample MSE using the true β			
μ_1	1	0.066	0.009	0.005	< 0.001
μ_2	1	0.056	0.009	0.005	< 0.001

Note: Monte Carlo with $H = 1000$ simulations of samples of size $N \in \{100, 1000, 10000, 100000\}$. Two-step estimation of the model for the SBNL model. The numbers in each row represent the Mean Square Error of each parameter averaged across 1000 simulations.

5.3 Aggregate Market Data and Applications

A critical issue in the applications of logit-type models to economics is how to deal with aggregate market data. In the nested logit model, there is a concern regarding the use of an individual choice model to the usually available data, as data on market shares and aggregate prices Berry et al. (1995). The traditional models, such as the Multinomial Logit and nested logit, can be linearized by a log transformation, where its reduced form can be estimated by the usual OLS or by GMM when dealing with endogeneity issues or structural estimations. This section takes the parallel of the SBNL model with the usual NL, showing how one can compute price elasticities, an important object in models of demand, and how to log-transform the model so that it can be used in an aggregate market data context.

5.3.1 Price elasticities

Recall that price elasticities refer to how a change in one good price affects the demand (the choice probability) for another good. There are two elasticities: the own price elasticity and the cross-price elasticity. Additionally, since alternatives in the same nest are similar, the cross-price elasticity is different for a price change in an alternative in the same nest than for a price change of an alternative in a different nest. In the SBNL, we must also take into account if the price change refers to a salient option or not since the salient option is the driver of the nest probability. This led to several special cases in presenting price elasticities.

The demand for a good is dependent on its characteristics, which is important to model product differentiation, mainly to its price. To simplify notation, we denote α_m as the price coefficient of a good in nest m, hence

$$V_{j_m} = x'_{j_m}\beta + \alpha \cdot p_{j_m},$$

where we expect $\alpha < 0$, since a usual demand is negatively related to price.

To denote the market share of good j_m we use the notation s_{j_m}. We can also have the conditional market (to nest m) share of j_m by $s_{j_m|C_m}$, and the category m market share by $s_m = \sum_{j_m \in C_m} s_{j_m}$.

In Appendix B we compute the derivative $(\partial s_{j_m}/\partial p_{k_n})$ to get the price elasticity of good j in nest m with respect to a change in the price of good k in nest n, ϵ_{j_m,k_n}, where:

$$\epsilon_{j_m,k_n} = \frac{\partial s_{j_m}}{\partial p_{k_n}} \frac{p_{k_n}}{s_{j_m}}$$

This difference between the response of a good to a salient good leads the model to six cases of price elasticity, which we present next.

Own-price elasticity:

$$\epsilon_{j_m} = \begin{cases} \alpha\mu_m(1 - s_{j_m|C_m})p_{j_m}, & \text{if } j_m \text{ is not a salient option} \\ \alpha\left((1 - s_m) + \mu_m(1 - s_{j_m|C_m})\right)p_{j_m}, & \text{if } j_m = j_m^* \text{ is a salient option} \end{cases}$$

Since only characteristics in nest m are changing, the first term is the same as multinomial logit models. The second term takes into account that if $j_m = j_m^*$ (j_m is the salient option of m), an increase in p_{j_m} reduces the attractiveness of alternatives in m relative to other nests.

Within nest cross-price elasticity:

Now we consider a price p_{k_m} in the same nest changing

$$\epsilon_{j_m,k_m} = \begin{cases} -\alpha\mu_m s_{k_m|C_m} p_{k_m}, & \text{if } k_m \text{ is not a salient option} \\ \alpha\left((1 - s_m) - \mu_m s_{k_m|C_m}\right)p_{k_m}, & \text{if } k_m = j_m^* \text{ is a salient option} \end{cases}$$

Again, in the first term, we are considering only changes within nest m. In the second term, we consider both the change within the nest and the change of the salient option of nest m relative to the others.

Cross nest cross-price elasticity:

Finally, considering changes in k_n from a different nest we get

$$\epsilon_{j_m,k_n} = \begin{cases} 0, \text{ if } k_n \text{ is not a salient option} \\ -\alpha s_n p_{k_n}, \text{ if } k_n \text{ is a salient option} \end{cases}$$

Notice that if we change some price of a non-salient good of nest n, nothing is expected to happen with options in nest m. In the usual nest logit, any change in n will affect nest m probability.

Following the previous observation on the regularity of section 4, it is important to notice that increasing the price of a similar option may both increase or decrease the market share of j_m depending on how salient k_m is within the nest. If k_m is not a salient option, increasing its price will decrease its value V_{k_m}, increasing the demand for good j_m. However, when k_m is salient in that nest, the attractiveness of the nest m will decrease, and the final effect is ambiguous.

This result implies that a firm may want to capture individuals' welfare by focusing on fire sales and discounts on the more salient goods. Moreover, with a product introduction, the firm can increase a category demand strategically with releases of prominent options. In the usual nested logit, any product introduction or discount to a product in a category would increase its demand. Additionally, this salience effect is not equal to the decoy effect. Introducing a good will increase the category probability only if this is the good chosen with the highest probability within the nest, i.e., it is a better good relative to its similar. In contrast, in a decoy effect, usually, a dominated but similar alternative is introduced.

There are well-known advantages to using the nested logit model instead of the multinominal logit. The IIA problem also reflects unrealistic substitution patterns between different goods[15]. Addressing the IIA assumption also implies a more unrestricted substitution pattern between alternatives, which the nested logit does. Not only nested logit allow for more flexible substitution patterns, but other nested stochastic choice models generally have more flexible substitution patterns implied by the model. Our model has

[15]See Train (2009), ch. 6, for examples and discussion.

the advantage of a more general substitution pattern while simultaneously capturing how goods in different nests respond to a change in the salience of some nest. In the next subsection, we discuss other practical implications of the model using aggregate data and comparing equation estimation of the multinomial logit and the nested logit model.

5.3.2 Fitting the SBNL model using aggregate data

Several applications of stochastic choice models, as nested logit models, explore aggregate market data, such as data on market-shares and product characteristics (e.g. prices), to estimate model parameters – and consequently, the demand – instead of observing individual-level choices. In particular, researchers may face a time-product data structure to fit the demand model. A usual procedure is to transform the model via log-linearization so linear panel data techniques can estimate it. In this section, we propose a transformation of the model to apply it on aggregate market data context, comparing the resulting estimable equation to the usual NL and with the multinomial logit model.

In this context, the probabilities $P(a_i = j_m)$ are usually interpreted as market shares, which we simplify the notation to s_{j_m}. Also, common to this framework, assume that the utility of an outside option is $V_0 = 0$ and let s_0 be its market share. In Appendix C we present a log transformation to the SBNL model that is useful when researchers have repeated observations of market shares and product characteristics:

$$\log(s_{j_m}/s_0) - \log(s_{j_m^*|C_m}) = x'_{j_m^*}\beta + \mu_m(x_{j_m} - x_{j_m^*})'\beta + \zeta_{j_m} \quad (9)$$

which makes explicit the non-linearity in parameters from the equation.

Equation (9) shows the importance of the salience in defining market shares. We can see that both a marginal change in a characteristic of x_{j_m} or a change in the characteristic of the salient option, $x^*_{j_m}$ of the nest, may directly change the market share of the good. Moreover, notice that the left-hand side of the equation (9) depends not only on s_{j_m} but also on $s_{j_m^*|C_m}$ and s_0. Thus, if, for example, one of the characteristics is the price, we need to use the procedure in section (5.3.1) and Appendix B, deriving the marginal effects from the structural model. For instance, if we are interested in the marginal effect of changing the price of the salient option of a nest n in demand for a good j_m in another nest m, they have $\partial s_{j_m}/\partial p_{k_n^*} = -\alpha s_n s_{j_m}$, where s_n denotes the demand to goods in nest

n, $\sum_{k_n \in C_n} s_{k_n}$ and α is the price coefficient associated with alternatives in the nest C_n, which we assume equal for all nests in the econometric model with aggregate market data, which is also a common modeling choice in the usual nested logit.

Moreover, with this equation, one can estimate the model with common linear panel data techniques. To see this, notice that

$$s_m/s_0 = e^{V_m^*}$$

thus, it is possible to estimate β using the linear equation

$$\ln(s_m/s_0) = x'_{j_m^*}\beta + \xi_{j_m^*}$$

then, we can get $\hat{\beta}$ by running a linear model using the observations of market-shares of salient options of each nest. Then, plugging $\hat{\beta}$ into equation (9) we can estimate μ_m using the linear model, with nest-specific parameters

$$\log(s_{j_m}/s_0) - \log(s_{j_m^*|C_m}) = x'_{j_m^*}\hat{\beta} + \mu_m(x_{j_m} - x_{j_m^*})'\hat{\beta} + \zeta_{j_m}.$$

Notice that most applications would have a panel data of market-shares and product characteristics, i.e., variables indexed by time $\{s_{j_m,t},\ x_{j_m,t}\}$, $t = 1, 2, ..., T$. We omit this subscript t above for the sake of simplification.

To compare with the Multinomial Logit model and the usual nested logit, an estimation equation in the nested logit case would be:

$$\log(s_{j_m}/s_0) = (1 - \lambda_m)\log(s_{j_m|C_m}) + x_{j_m} \cdot \beta + \zeta_{j_m}. \tag{10}$$

with $1 - \lambda = 0$ in the Multinomial Logit case[16]. Thus in the SBNL both LHS and RHS of the estimation equation are taking into account the salient option when computing the model parameters.

Example 2 presents the fit of the model with real data on flight destination and how to use the fitted parameters to understand elasticities.

Example 2 *(Fitting the model with aggregate data on flight-destination) To give one example and estimate the probability and utility parameters in real data, we collected Brazil-*

[16]As we show in the previous section, in our model μ_m is analogous to the inverse of the λ_m used in the nested logit, i.e., $\lambda_m = \frac{1}{\mu_m}$.

ian data on domestic flights starting from the State of Sao Paulo from any airport to any other airport (including others in the same state). In this data, we have the origin and destination of the air-passages, the number of flight seats sold, and the price paid. We have monthly observations from 2002 until 2019 for the 27 Brazilian states. We also collected a proxy for the living cost, which is the average m^2*-cost of construction in each Brazilian state, computed and publicly available at IBGE (Brazilian Institute of Geography and Statistics). We use January, July, and December to be vacation seasons and February due to the carnival, a period with lots of tourist travel in Brazil.*

In this framework, we build nests naturally as regions. Thus we have five nests m*, where each nest corresponds to a collection of Brazilian States,* $m = 1$ *is Southeast,* $m = 2$ *corresponds to South,* $m = 3$ *to Midwest,* $m = 4$ *to Northeast, and* $m = 5$ *to North.*

Hence, we parametrize the utility of travel to state j *in the region* m*,* $V_{j_m,t}$*, as a function of the price* $p_{j_m,t}$ *and living cost* $c_{j_m,t}$ *of destination* j *in the region* m *bought in period* t*. Our estimated equation following equation (7) is*

$$\log(s_{j_m,t}/s_0)-\log(s_{j_m^*|C_m}) = \beta_0+\beta_1 p_{j_m^*,t}+\beta_2 c_{j_m^*,t}+\mu_m(\beta_1(p_{j_m,t}-p_{j_m^*,t})+\beta_2(c_{j_m,t}-c_{j_m^*}))+\zeta_{j_m,t}.$$

To compute market shares using the number of seats sold each period, we need to define the potential market. Here, we took the estimated population of the State of Sao Paulo at each month t*, also available at IBGE, and consider the population in the first income quintile as the potential market for flight travels* [17]*. Then, we divide the number of seats sold in month* t *to travel for state* j_m *by this potential market,* s_0 *is the "residual" of this market-shares such that shares add up to a unit.*

In particular, we did two exercises fitting the model parameters. First, we estimated by Non-Linear Least Squares. After that, we fit a GMM using instruments to the moment conditions. Results are presented in Table 3. We did not restrict λ *to be positive, although the researcher can include this restriction to ensure this constraint in the estimates.*

[17] We restricted the sample for flights from the State of Sao Paulo because it is the most populated state in Brazil. Moreover, flight services are considered very expansive in the country, not accessible to the whole population.

Table 3: Model Estimation with Aggregate Market Data

First-Equation (Estimated by TSLS)		Second-Equation (Estimated by GMM)	
Dependent Variable	$ln(s_m/s_0)$	Dependent Variable	$ln(s_{j_m}/s_0) - ln(s_{j_m^*})$
β_0	-3.023***	μ_1	16.539
	(0.149)		(23.079)
β_1	-0.005***	μ_2	10.256
	(0.000)		(18.165)
β_2	0.001***	μ_3	14.995***
	(0.000)		(18.165)
		μ_4	1.847***
			(0.061)
		μ_5	3.895***
			(0.140)
F-Statistic 1st-stage IV F-statistic	469.68		
Overid. Test	0.167	Overid. Test	0.309
p-value	0.6825	p-value	0.857

***$p < .01$, **$p < .05$, *$p < .1$. Estimation of utility and probability parameters of SBNL model. Robust standard errors in parentheses. The instruments for the salient price in the first-equation are the total cost of salient options in other nests and the squared cost of the salient option of the nest. Instruments for p_{j_m} are the living cost of options in other nests, the living cost of salient option, and the squared cost of the option. Bootstrap standard-error in the GMM equation. 1080 observations in the first-equation (salient-options) and 4340 valid observations in the second-equation.

The price coefficient is negative, as expected. The control variable, living cost, is positive but very small. With no need for restriction in the estimation procedure, we got positive nest coefficients in the estimation.

The example above has price-elasticity implications and also implies stochastic transitivity. The estimates of $\lambda_m = 1/\mu_m$ are all lower than two, the condition we obtain to the MUM representation of He and Natenzon (2022). Moreover, one can plug the nest parameters and the price coefficient into the price-elasticity formulas to derive the response behavior of the Brazilian flight market to price changes (see section 5.3.1). To exemplify, considering the aggregate market data on dec/2019, a price change in a non-salient state of south would not change the flight tariff in the State of Sao Paulo (in southeast, $m = 1$), however, we can compute that a change in the price of the salient state in south ($m = 2$), by 1%, would change the demand for the State of Sao Paulo flights in 0.02%. Moreover, an increase in the salient option of southeast would have a positive net effect on the demand for the State of Sao Paulo destination. The higher μ_1, the higher is the net effect of the salient option on the demand for the State of Sao Paulo.

6 Conclusion and Discussion

This paper presents a novel foundation to the nested logit model to incorporate salience bias. We designed an axiom-based characterization in the spirit of Kovach and Tserenjigmid (2022a), providing testable conditions defining the model. We show how this model is able to capture a new type of violation on regularity, a common property of stochastic choice models, in special random utility models. The SBNL model can be non-regular even when $\lambda_i < 1$, which corresponds to the concave case of the usual NL model.

The violation of regularity shows how a salient option can discontinuously increase the attractiveness of a nest, increasing the demand for similar options, which the traditional nested logit is not able to predict. One can notice that this characterizes a new type of violation of regularity different from the decoy effect. In the decoy effect, a similar but dominated alternative attracts attention to the dominant one. Our central axiom defines an invariance of nest probabilities to the inclusion of non-salient options to a menu.

In a more structural setting, we show how the model can be derived in a random utility framework, in a similar choice structure to the usual nested logit, but modifying the behavioral maximization of agents. Thus, our model can also be derived in the random utility tradition of McFadden et al. (1973). One can add common parametric assumptions in the utility, usual in econometric applications of multinomial choice (as MNL and NL). Using these assumptions, the model can be estimated with real data using standard logit estimation routines in a two-step logit estimation.

One of the potential applications of the model is the estimation of demand considering salience bias. In many settings, researchers have access to aggregate market data. We show how to apply a log transformation in the model to achieve an estimable equation with aggregate market share and market characteristics. Finally, one important object in the Industrial Organization literature is price elasticity. We derive the price elasticity of the SBNL, considering the own price elasticity, cross-price elasticity, and if the price change comes from a salient option of a nest, which leads our model to more elasticity cases than in the usual nested logit, showing that we can capture more market movements using our model. We present a simple application to show how to fit an estimable equation using flight demand data.

Additionally, due to the capacity to capture non-regularity, our model also fits in the evaluation of the effects of product innovation and mergers and acquisitions using our implied demand. In these frameworks, the salient option of each firm may change, generating a possible incentive (or disincentive) for these actions. Moreover, it is interesting to investigate whether firms capture more consumer welfare due to the salience bias. Thus, there is an opportunity to revisit some empirical applications comparing the usual nested logit with the salience-biased nested logit. As highlighted by Ellison (2006), the literature lacks investigation on the interplay of supply and demand in a structural model with bounded rationality. Since our model directly describes the demand, one can apply it to structural empirical estimation in an adequate context.

Appendix A - Proofs

In the appendix, we provide proof of our results.

Proof. (Proof of Theorem 1)

Sufficiency: From Kovach and Tserenjigmid (2022a) we know that there exists a partition $\{A_1, ..., A_n\}$, a luce value $v_x : X \to \mathbb{R}_{++}$ and a nest utility $V : \cup_{i=1}^{n} 2^{A_i} \to \mathbb{R}_+$ representing p as a Nested Stochastic Choice. Indeed, we can take $v(A \cap A_i) = p(A \cap A_i, A)$ and $u(x) = p(x, A_{i_x}) = p(x, [x])$.

Now, from the axiom 2, given $x \in A$, let x^* be the salient option of $[x]$ relative to the menu A. since $A = (A \cap A_{i_x}) \cup (A \cap A_{i_x}^c)$. Then, for all $y \in A_{i_x} \cap A$

$$p(A_{i_x} \cap A, A) = V(A \cap A_{i_x}) = V(A \cap A_{i_x} \setminus \{y\})$$

Then, $V(A \cap A_{i_x}) = V(A \cap \{x^*\}) = V(x^*)$.

Since $V(x^*)$ is a constant function of the salient option of A_{i_x} (relative to A), we can take $V(x^*) = f_i\big(p(x^*, [x])\big) = f_i(v_{x^*})$.

Necessity: Take $x \neq y$ such that $x \sim_p y$ and $p(x, [x]) \geq p(y, [y])$. Take $A \subseteq [x] = A_{i_x}$ with $\{x, y\} \subseteq A_{i_x}$ and $B \subseteq [x]^c$. Let x^* be the salient option of $[x]$ relative to A. Then notice that $v_{x^*} \geq v_x \geq v_y$ hence

$$p(A, A \cup B) = \frac{f_i(v_{x^*})}{\sum_{k=1}^{n} f_k \left(\max\limits_{z \in A \cap A_k} v_z \right)}$$

Moreover, since either y is not a salient option or there exists more than one salient option (relative to $A \cup B$), we get

$$p(A \setminus \{y\}, A \setminus \{y\} \cup B) = \frac{f_i(\max_{z \in A \cap A_{i_x}} v_z)}{\sum_{k=1}^{n} f_k(\max_{z \in A \cap A_k} v_z)} = $$

$$= \frac{f_i(v_{x^*})}{\sum_{k=1}^{n} f_k(\max_{z \in A \cap A_k} v_z)} = $$

$$= p(A, A \cup B)$$

Thus, Axiom 2 holds. Axiom 1 is straightforward since p is an NSC function. ■

Axiom 3 *(Salience Log-Ratio Invariance) Given $x \in X$, for all $A, B \subseteq [x], \forall y \in [x]^c$. If $p(a^*, A) \geq p(a, A)$ for all $a \in A$ and $p(b^*, B) \geq p(b, B)$ for all $b \in B$, then, for all $\hat{a}$ s.t. $p(\hat{a}, [x]) < \min\{p(a^*, [x]), p(b^*, [x])\}$ then*

$$\frac{\ln\left(\frac{p(A, A\cup\{y\})}{p(y, A\cup\{y\})} / \frac{p(x,\{x,y\})}{p(y,\{x,y\})}\right)}{\ln\left(\frac{p(a^*, A\cup\{\hat{a}\})}{p(\hat{a}, A\cup\{\hat{a}\})}\right)} = \frac{\ln\left(\frac{p(B, B\cup\{y\})}{p(y, B\cup\{y\})} / \frac{p(x,\{x,y\})}{p(y,\{x,y\})}\right)}{\ln\left(\frac{p(b^*, B\cup\{\hat{a}\})}{p(\hat{a}, B\cup\{\hat{a}\})}\right)}$$

Theorem 2 *A positive stochastic choice function p satisfies* ***Independence of Symmetric Alternatives****,* ***Neutrality of Weakly Dominated Alternatives within Categories*** *and* ***Salience Log-Ratio Invariance*** *if and only it is a salience-biased nested logit, i.e., $\exists \mathbb{A} = \{A_1, A_2, ...A_n\}$ partitioning X, a set of parameters $\{\lambda_1, \lambda_2, ..., \lambda_n\} \subseteq \mathbb{R}^n_+$ and a luce value $v : X \to \mathbb{R}_{++}$ such that for all $A \in 2^X \setminus \{\emptyset\}$, and for all $x \in A$*

$$p(x, A) = \frac{\left(\max_{y \in A \cap A_{i_x}} v_y\right)^{\lambda_{i_x}}}{\sum_{i: A \cap A_i \neq \emptyset} \left(\max_{y \in A \cap A_i} v_y\right)^{\lambda_i}} \frac{v_x}{\sum_{y \in A \cap A_{i_x}} v_y}$$

Proof. (Proof of Theorem 2)

Sufficiency: Given a partition $\{A_1, A_2, ..., A_n\}$ and $i \in \{1, 2, ..., n\}$ take $A \subseteq A_i$ and take $\hat{a}$ s.t. $v_{\hat{a}} < \max_{z \in A} v_z$[18]. Define λ_i as following

$$\lambda_i := \frac{\ln\left(\frac{f_i(\max_{z \in A_i} v_z)}{f_i(v_{\hat{a}})}\right)}{\ln\left(\frac{\max_{z \in A_i} v_z}{v_{\hat{a}}}\right)} = \frac{\ln\left(\frac{p(A_i, A_i \cup \{y\})}{p(y, A_i \cup \{y\})} \Big/ \frac{p(x,\{x,y\})}{p(y,\{x,y\})}\right)}{\ln\left(\frac{p(a_i^*, A_i \cup \{\hat{a}\})}{p(\hat{a}, A_i \cup \{\hat{a}\})}\right)}$$

Then, in axiom 3, we can take $B = A_i$ to get that:

$$\lambda_i = \frac{\ln\left(\frac{p(A, A \cup \{y\})}{p(y, A \cup \{y\})} \Big/ \frac{p(\hat{a},\{x,y\})}{p(y,\{\hat{a},y\})}\right)}{\ln\left(\frac{p(a^*, A \cup \{\hat{a}\})}{p(\hat{a}, A \cup \{\hat{a}\})}\right)} = \frac{\ln\left(\frac{f_i(\max_{z \in A} v_z)}{f_i(v_y)} \Big/ \frac{f_i(v_{\hat{a}})}{f_i(v_y)}\right)}{\ln\left(\frac{\max_{z \in A} v_z}{v_{\hat{a}}}\right)} = \frac{\ln\left(\frac{f_i(\max_{z \in A} v_z)}{f_i(v_{\hat{a}})}\right)}{\ln\left(\frac{\max_{z \in A} v_z}{v_{\hat{a}}}\right)}$$

Now, taking $\delta_i = \frac{(v_{\hat{a}})^{\lambda_i}}{f_i(v_{\hat{a}})}$ we get:

$$f_i\left(\max_{z \in A} v_z\right) = \delta_i \left(\max_{z \in A} v_z\right)^{\lambda_i}$$

Thus, transforming the Luce values to $\delta_i v_z$ if $z \in A_i$ we get the salience-biased nested logit representation.

Necessity: Given $x \in X$ $A \subseteq [x]$, $y \notin [x]$ and $\hat{a}$ s.t. $v_{\hat{a}} < v_{a_A^*}$, notice that

(i)

$$\frac{p(A, A \cup \{y\})}{p(y, A \cup \{y\})} = \frac{(v_{a_A^*})^{\lambda_i}}{(v_y)^{\lambda_i}}$$

(ii)

$$\frac{p(x, y)}{p(y, x)} = \frac{v_x^{\lambda_i}}{v_y^{\lambda_i}}$$

(iii)

$$\frac{p(a^*, A \cup \{\hat{a}\})}{p(\hat{a}, A \cup \{\hat{a}\})} = \frac{v_{a_A^*}}{v_{\hat{a}}}$$

Then:

$$\frac{\ln\left(\frac{p(A, A \cup \{y\})}{p(y, A \cup \{y\})} \Big/ \frac{p(x,\{x,y\})}{p(y,\{x,y\})}\right)}{\ln\left(\frac{p(a^*, A \cup \{\hat{a}\})}{p(\hat{a}, A \cup \{\hat{a}\})}\right)} = \lambda_i \frac{\ln\left(\frac{v_{a_A^*}}{v_{\hat{a}}}\right)}{\ln\left(\frac{v_{a_A^*}}{v_{\hat{a}}}\right)} = \lambda_i$$

i.e., axiom 3 holds. ■

[18]When $|\{v_y : y \in A_i\}| = 1$ we can take w.l.o.g. $\lambda_i = 1$.

Appendix B - Market-share responses to prices

In this appendix, we provide market share derivatives of good j_m with respect to price of good k_n. Here, we keep the assumption $|V_m^*| = 1$, that is, each nest has a single salient option. Moreover, derivatives are valid locally so that the salient option of each nest remain the same after the price change.

We denote:

- $s_m = P(a_i \in C_m)$: the share of nest m
- $V_{j_m} = x'_{j_m}\beta + \alpha p_{j_m}$: thus, the scalar α is the price coefficient, following the common notation in Industrial Organization.

Own-price response:

To compute the own-price response we need to consider the cases when $j_m \neq j_m^*$ and $j_m = j_m^*$, i.e., when j_m is or is not the salient option in nest m

- Case 1: $j_m \neq j_m^*$.

  In this case, s_m does not depend on p_{j_m} and using the fact that $s_m s_{j_m|C_m} = s_{j_m}$ then:

$$\begin{aligned}\frac{\partial s_{j_m}}{\partial p_{j_m}} &= s_m \frac{\partial}{\partial p_{j_m}} \left(\frac{e^{\mu_m V_{j_m}}}{\sum_{k_m \in C_m} e^{\mu_m V_{k_m}}} \right) \\ &= s_m \left(\alpha \mu_m s_{j_m|C_m} (1 - s_{j_m|C_m}) \right) \\ &= \alpha \mu_m s_{j_m} (1 - s_{j_m|C_m})\end{aligned}$$

- Case 2: $j_m = j_m^*$. In this case, both s_m and $s_{j_m|C_m}$ dependent on p_{j_m}, then a new term composes the derivative:

$$\begin{aligned}\frac{\partial s_{j_m}}{\partial p_{j_m}} = \frac{\partial s_{j_m}}{\partial p_{j_m}^*} &= \frac{\partial s_m}{\partial p_{j_m^*}} s_{j_m|C_m} + s_m \frac{\partial s_{j_m|C_m}}{\partial p_{j_m}} = \\ &= \alpha s_m (1 - s_m) s_{j_m|C_m} + s_m (\alpha \mu_m s_{j_m|C_m} (1 - s_{j_m|C_m})) \\ &= \alpha s_{j_m} [1 - s_m + \mu_m (1 - s_{j_m|C_m})]\end{aligned}$$

Within nest cross-price response:

Now, we compute the response of s_{j_m} to a change in price of $k_m \in C_m$.

- Case 3: $k_m \neq j_m^*$ (and $k_m \neq j_m$ with $k_m \in C_m$).

$$\begin{aligned}\frac{\partial s_{j_m}}{\partial p_{k_m}} &= s_m \frac{\partial s_{j_m|C_m}}{\partial p_{k_m}} \\ &= s_m \left(-\alpha \mu_m \frac{e^{\mu V_{j_m}}}{(\sum_{\tilde{k}_m \in C_m} e^{\mu_m V_{\tilde{k}_m}})^2} e^{\mu_m V_{k_m}} \right) \\ &= -s_m s_{j_m|C_m} s_{k_m|C_m} \alpha \mu_m \\ &= -\alpha \mu_m s_{j_m} s_{k_m|C_m}.\end{aligned}$$

- Case 4: $k_m = j_m^*$ (and $k_m \neq j_m$ with $k_m \in C_m$)

$$\begin{aligned}\frac{\partial s_{j_m}}{\partial p_{k_m}} = \frac{\partial s_{j_m}}{\partial p_{j_m^*}} &= \frac{\partial s_m}{\partial p_{j_m^*}} s_{j_m|C_m} + s_m \frac{\partial s_{j_m|C_m}}{\partial p_{k_m}} \\ &= (\alpha s_m - \alpha s_m^2) s_{j_m|C_m} + s_m(-\alpha \mu_m s_{j_m|C_m} s_{k_m|C_m}) \\ &= \alpha s_{j_m}(1 - s_m) - \alpha \mu_m s_{j_m} s_{k_m|C_m} \\ &= \alpha s_{j_m}[(1 - s_m) - \mu_m s_{k_m|C_m}]\end{aligned}$$

where we use that $s_{j_m} = s_m s_{j_m|C_m}$.

Cross-nest price responses

- Case 5: $k_n \neq j_m$, $k_n \in C_n \neq C_m$ and $k_n \neq k_n^*$ (alternatives in different nests, and k_s is not a salient option of nest s)

$$\frac{\partial s_{j_m}}{\partial p_{k_s}} = 0.$$

- Case 6: $k_n \neq j_m$, $k_n \in C_n \neq C_m$ and $k_n = k_n^*$ (alternatives in different nests, and k_s is the salient option of nest s, so we denote k_s^*)

$$\frac{\partial s_{j_m}}{\partial p_{k_n}} = \frac{\partial s_m}{\partial p_{k_n^*}} s_{j_m|C_m} = -\alpha s_n s_{j_m},$$

  where, $s_n = Pr(a_i \in C_n)$.

Appendix C - Log-transformation of the SBNL

In this appendix we show how to do a log-transformation type to the individual choices in the SBNL to accommodate aggregate market data. In other words, suppose that the researcher observes $D_N = \left\{ \{s_{i,j_m}\}_{j_m \in C}, x_i \right\}_{i=1}^{N}$, where i here represents a market (or time periods) and s_{i,j_m} is the market share of product j_m at market i, and x_i represents the product characteristics face by agent i. As in many applications, assume that $V_0 = 0$ the utility of the outside option that is assumed to be the only alternative at nest $m = 0$, with market share s_0. Let s_{j_m} be the share of the good j that belongs to nest m, and analogously, $s_{j_m^*}$ the share of the salient option in nest m. Using the SBNL probabilities in equation (7) we have that:

$$\frac{P(a_i = j_m)}{P(a_i = 0)} = e^{V_{j_m^*}} \frac{e^{\mu_m V_{j_m}}}{\sum\limits_{k_m \in C_m} e^{\mu_m V_{k_m}}} = e^{V_m^*} \left(\frac{e^{\mu_m V_{j_m^*}}}{\sum\limits_{k_m \in C_m} e^{\mu_m V_{k_m}}} \right) \frac{e^{\mu_m V_{j_m}}}{e^{\mu_m V_{j_m^*}}}.$$

Notice that $\frac{e^{\mu_m V_{j_m^*}}}{\sum\limits_{k_m \in C_m} e^{\mu_m V_{k_m}}} = s_{j_m^*|C_m}$. Then, taking the log on both sides we get

$$\log(s_{j_m}/s_0) - \log(s_{j_m^*|C_m}) = V_{j_m^*} + \mu_m(V_{j_m} - V_{j_m^*}).$$

In this framework its usual to assume a utility shock ν_{j_m} observed by the agent, not the econometrician so that $V_{j_m} = x'_{j_m}\beta + \nu_{j_m}$, then we get

$$\log(s_{j_m}/s_0) - \log(s_{j_m^*|C_m}) = x'_{j_m^*}\beta + \mu_m(x_{j_m} - x_{j_m^*})'\beta + \nu_{j_m^*} + \mu_m(\nu_{j_m} - \nu_{j_m^*})$$

where we finally define $\zeta_{j_m} := \nu_{j_m^*} + \mu_m(\nu_{j_m} - \nu_{j_m^*})$, an error term that is not observed by researchers.

$$\log(s_{j_m}) - \log(s_{j_m^*|C_m}) - \log(s_0) = x'_{j_m^*}\beta + \mu_m(x_{j_m} - x_{j_m^*})'\beta + \zeta_{j_m}$$

which is an estimable equation by usual methods as Non-linear Least Squares or GMM, when addressing endogeneity in the model.

References

Agranov, M. and Ortoleva, P. (2017). Stochastic choice and preferences for randomization. *Journal of Political Economy*, 125(1):40–68.

Aguiar, V. H., Boccardi, M. J., Kashaev, N., and Kim, J. (2021). Random utility and limited consideration. *Mimeo, Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3871256.*

Anderson, S. P., De Palma, A., and Thisse, J.-F. (1992). *Discrete choice theory of product differentiation.* MIT press.

Barber, B. M. and Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The review of financial studies*, 21(2):785–818.

Ben-Akiva, M. and Lerman, S. R. (2018). *Discrete choice analysis: theory and application to travel demand.* Transportation Studies.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, pages 841–890.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2012). Salience theory of choice under risk. *The Quarterly journal of economics*, 127(3):1243–1285.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2013a). Salience and asset prices. *American Economic Review*, 103(3):623–28.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2013b). Salience and consumer choice. *Journal of Political Economy*, 121(5):803–843.

Camerer, C. F., Ho, T.-H., and Chong, J.-K. (2004). A cognitive hierarchy model of games. *The Quarterly Journal of Economics*, 119(3):861–898.

Caplin, A. and Dean, M. (2015). Revealed preference, rational inattention, and costly information acquisition. *American Economic Review*, 105(7):2183–2203.

Caplin, A., Dean, M., and Leahy, J. (2019). Rational inattention, optimal consideration sets, and stochastic choice. *The Review of Economic Studies*, 86(3):1061–1094.

Cardell, N. S. (1997). Variance components structures for the extreme-value and logistic distributions with application to models of heterogeneity. *Econometric Theory*, pages 185–213.

Cattaneo, M. D., Ma, X., Masatlioglu, Y., and Suleymanov, E. (2020). A random attention model. *Journal of Political Economy*, 128(7):2796–2836.

Cosemans, M. and Frehen, R. (2021). Salience theory and stock prices: Empirical evidence. *Journal of Financial Economics*, 140(2):460–483.

De los Santos, B., Hortaçsu, A., and Wildenbeest, M. R. (2012). Testing models of consumer search using data on web browsing and purchasing behavior. *American economic review*, 102(6):2955–80.

Dean, M. and Neligh, N. L. (2017). Experimental tests of rational inattention. Technical report, National Bureau of Economic Research (NBER).

D'Acunto, F., Prabhala, N., and Rossi, A. G. (2019). The promises and pitfalls of robo-advising. *The Review of Financial Studies*, 32(5):1983–2020.

Ellis, A. and Masatlioglu, Y. (2019). A behavioral foundation for endogenous salience. *mimmeo*.

Ellison, G. (2006). Bounded rationality in industrial organization. *Econometric Society Monographs*, 42:142.

Ergin, H. and Sarver, T. (2010). A unique costly contemplation representation. *Econometrica*, 78(4):1285–1339.

Faro, J. H. (2023a). The luce model with replicas. *Journal of Economic Theory*, page 105596.

Faro, J. H. (2023b). The luce model with replicas. *Journal of Economic Theory*, 208:105596.

Frydman, C. and Wang, B. (2020). The impact of salience on investor behavior: Evidence from a natural experiment. *The Journal of Finance*, 75(1):229–276.

Gul, F., Natenzon, P., and Pesendorfer, W. (2014). Random choice as behavioral optimization. *Econometrica*, 82(5):1873–1912.

Hartzmark, S. M. (2015). The worst, the best, ignoring all the rest: The rank effect and trading behavior. *The Review of Financial Studies*, 28(4):1024–1059.

He, J. and Natenzon, P. (2022). Moderate utility. *Unpublished manuscript*.

Huber, J., Payne, J. W., and Puto, C. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of consumer research*, 9(1):90–98.

Kahneman, D. (1973). *Attention and effort*, volume 1063. Citeseer.

Kovach, M. and Tserenjigmid, G. (2022a). Behavioral foundations of nested stochastic choice and nested logit. *Journal of Political Economy*, 130(9):2411–2461.

Kovach, M. and Tserenjigmid, G. (2022b). The focal luce model. *American Economic Journal: Microeconomics*, 14(3):378–413.

Li, J. and Tang, R. (2016). Associationistic luce rule. *Available at SSRN 2770773*.

Luce, R. D. (1959). Individual choice behavior: a theoretical analysis.

Manzini, P. and Mariotti, M. (2014). Stochastic choice and consideration sets. *Econometrica*, 82(3):1153–1176.

Matějka, F. and McKay, A. (2015). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review*, 105(1):272–98.

McFadden, D. (1977). Modelling the choice of residential location.

McFadden, D. (1981). Econometric models of probabilistic choice. *Structural analysis of discrete data with econometric applications*, 198272.

McFadden, D. (2001). Economic choices. *American economic review*, 91(3):351–378.

McFadden, D. et al. (1973). *Conditional logit analysis of qualitative choice behavior*. Institute of Urban and Regional Development, University of California . . . .

Ortoleva, P. (2013). The price of flexibility: Towards a theory of thinking aversion. *Journal of Economic Theory*, 148(3):903–934.

Rieskamp, J., Busemeyer, J. R., and Mellers, B. A. (2006). Extending the bounds of rationality: Evidence and theories of preferential choice. *Journal of Economic Literature*, 44(3):631–661.

Stigler, G. J. (1961). The economics of information. *Journal of political economy*, 69(3):213–225.

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, 5(4):297–323.