# Thesis on Approval Behaviour

Kai Wang

Thesis submitted for the degree of Doctor of Philosophy in Economics

University of Sussex

September 2021

## Declaration

I hereby declare that this thesis has not been and will not be submitted in whole or in part to another University for the award of any other degree.

I also hereby declare that Chapter 1, 2 and 3 are all solo-authored.

Signature:

University of Sussex

Kai Wang

DOCTOR OF PHILOSOPHY IN ECONOMICS

Thesis on Approval Behaviour

Summary

This thesis answers the research questions in the field of decision maker (DM)'s 'approval' behaviour. A DM makes approval decisions by putting items into binary classes (authorization or rejection, approval or denial, inclusion or exclusion). This thesis introduces novel models which can be used to study approval behaviour. This thesis also draws on recent empirical frameworks that allow us to use an online commerce dataset to analyse how consumers include or exclude listings into their consideration sets (i.e., the sets of alternatives DMs actually choose between).

In *History Dependent Approval* chapter, we provide two models to study both deterministic and stochastic sequential approval behaviours that exhibit history dependency. The deterministic model is fully characterized from an Approval Consistency axiom, which states that if $x$ is approved and $y$ is disapproved given a history, then there does not exist any history given which $x$ is disapproved and $y$ is approved. The stochastic model is also characterized from a single axiom that can be seen as a stochastic version of Approval Consistency axiom.

*Logit Function in Stochastic Categorization* chapter develops a logit categorization function that can be applied to approval data. We characterize the model from a simple condition on approval data. We show that the logit categorization function can be derived from adaptations of the multinomial logit model. The logit categorization function can be useful in future empirical studies on approval behaviour.

In *Consideration Set Formation in Online Commerce* chapter, we estimate the factors that affect the probability of considering a listing on the web page using a detailed browsing dataset from eBay. We find strong evidence that consumers form their consideration sets depending on characteristics like shipping cost, total price, seller type, etc. We analyse the change of preference and consideration formation after a major platform redesign on eBay deployed in 2011.

## Acknowledgements

The last four years study, working, and writing have been a long, but exciting journey. There are many people to which I own much of my learning, and I would like to take this chance to whole-heartedly thank these people in these lines.

First and foremost, I would like to express my sincere gratitude to my main supervisor Paola Manzini, who has always been patient and supportive during my PhD study. I started my PhD with almost zero experience in economic research and academic writing, and she has always been there to guide me through. Every aspect of her supervision made it clear that she was committed to providing me with the best supervision possible that I develop the skills, the discipline, and most importantly, the perseverance to carry on. I am deeply indebted to Pawel Dziewulski, who became my main supervisor during the late stage of my PhD after Paola's leaving to the University of Bristol. Pawel has supervised me in a cool and calm way. His advice and comment on my thesis are critical. I am also very lucky to have Matthew Embrey and Giordano Mion as my supervisors. My communication with these professional people has been always open and frank, and I received timely and substantive comment on my work.

My sincere thanks also go to my thesis committee, the external examiner Valentino Dardanoni and the internal one Tom Potoms, for the time they spent reading my thesis and for the challenging and constructive inputs and comments they provided.

The economics research community at the University of Sussex is the best group I have ever experienced during my student life. I shared many lovely and memorable moments with Bernardo Caldarola, Daniele Guariso, Guillermo Larbalestier, Juan Manuel Del Pozo Segura, Nicolo Tamberi, Irene Staffieri. During the PhD studies, Anuja Suhasini Bajaj, Marta Schoch, Engenia Go, Fjolla Kondirolli, Maika Schmidt have given substantive effort in organizing numerous activities that keep our community strong and interactive. I found my working place in Jubilee building a perfect environment for research, especially with these excellent working mates Nick Jacob, Mohamed

Ali, Pierre Marion. I also enjoyed a lot during the time spent in Brighton with the Italian 'gangster', Stella Brambilla, Pier Rolla, Emanuele De Girolamo, Margherita Bove, Andrea Carboni, Filippo Bntdn, Mattia Di Ubaldo. I had a wonderful time during the lockdown with friends from my home country, Zhennan Zhu, Zhaohua Li, Yuanyuan Wu.

I own this accomplishment to my parents. Their love and belief in me helped me go through many tough moments. Their financial support allowed me to start my study in the UK. I would like to thank my uncle Xiangcai Chen, and my aunt Xiaoduo Wang, who encouraged me to boldly quit my previous job and chase what I am really after.

At the end of the day, with all those overlapping names and faces coming to my mind, there is one that I want to thank the most and express my love to. Fangyue Cui, her love and support made this happen, this thesis is for her.

# Contents

# List of Abbreviations

AR: authorization to rejection

ASC: alternative specific consideration

DM: decision maker

DHAF: deterministic history-dependent approval function

DELS: Dinerstein, Einav, Levin, and Sundaresan

IIA: independence of irrelevant alternatives

MMU: Manzini, Mariotti and Ulku

RHAF: random history-dependent approval function

RHIAF: random history-independent approval function

SAC: satisficing acceptability-continuation

TRS: top-rated seller

USAM: US automotive manufacturer

WARP: weak axiom of revealed preference

# List of Figures

# List of Tables

# Introduction

The goal of this thesis is to shed light on a new research object, approval. Throughout the thesis, approval refers to the binary classification behaviour in which a DM puts the items into binary classes. For example, a judge approves or denies parole for inmates, a doctor administers or withholds treatment to patients, a customer considers or ignores products.

Approval is distinct from 'choice', a research object that economic literature has studied extensively for years. When a choice is made, the best alternative is chosen from the menu (i.e., the set of feasible alternatives). When making approvals, however, all the items that are preferred to the threshold are approved. This difference is not trivial, as it points out that the standard choice function cannot be used to describe approval, and that the assumption of preference maximization fails in approval making. Our work draws heavily from one key paper: Manzini, Mariotti, and Ulku (2021) (henceforth, MMU). To date, MMU (2021) is the first paper studying approval by using revealed preference techniques. The novelty of approval immediately raises two questions, in the words of MMU (2021):

*When considering such examples, two issues arise. First, how can we describe approval data? Second, how can we explain (or represent) approval behaviour? The analogous questions in standard choice theory are answered, respectively, by: "a choice function" and "preference maximisation". The conditions ensuring that a choice function can be represented as preference maximisation, and the way unobserved preferences can* *be identified from observed choice, are well-understood. In this paper we aim to carry out a similar general exercise for approval, abstracting from the details of any specific context.*

In this thesis, we study approval in a variety of contexts which correspond to different scenarios. In all of the three chapters, we propose approval functions that accommodate respective decision making process. In Chapter 1 and 2, we propose new models to explain approval behaviour. We illustrate identifications of threshold and preference if the data is generated by our models. In Chapter 3, we estimate the approval function from a real world dataset using a novel econometric model from Abaluck and Adams-Prassl (2021).

In Chapter 1, we study sequential approval, which is binary classification behaviour made in a sequential manner. We relax the assumption in MMU (2021) that the threshold is independent of the history (i.e., the sequence of past examined items), and instead allow the threshold to be history dependent. The main contribution of this chapter is to provide reasonable simple models that will serve as a lens through which researchers will be able to interpret new evidence and stories that involve approval. A deterministic approval model and a stochastic approval model, both of which accommodate history dependency, are proposed. Each model is characterized with a single intuitive axiom. We build our models in the pre-existing environment of MMU (2021), and show how we can infer the preference and threshold from approval data.

In Chapter 2, we aim to 'bridge' the gap between theory and application for the study on approval. A stochastic approval model is developed for categorization behaviour, which is a type of approval whereby the order of the items does not influence the decision. We provide an empirical model that can be easily applied to real world data by incorporating the characteristics of the items and menus. The model is also axiomatized by a single postulate which is relatable and simple.

Finally, in Chapter 3, we focus on a very specific type of approval: consumers' considering or ignoring listings in online commerce. The set of alternatives that are actually considered by a DM is named the consideration set in the marketing literature. Although consideration is a very specific type of approval, it stands out amongst other types of approvals, as it is of both managerial and economic interest. From a managerial perspective, the sales of products can be largely affected by the consideration set. For example, Reebok doubled its sales by using the promoted listing on online sales platform eBay in 2016.[1] This chapter provides quantitative analysis of the formation of consideration set in online commerce. We find strong evidence that shipping cost, total price, and seller type influence the consideration set formation.

In conclusion, there is much to learn from this approval behaviour, from both theoretical and empirical point of view. The main contribution of this thesis is to provide new insights and research possibilities in economic studies that involve approval.

[1]The promoted listing is a feature that boosts number of views by up to 36%. Source of information: https://pages.ebay.com/seller-center/listing-and-marketing/promoted-listings.html#key-benefits.

# Chapter 1

# History Dependent Approval

**Abstract**

The aim of this study is to investigate the history dependency in approval making. Approval is binary classification behaviour in which a DM approves or disapproves items sequentially. In this chapter, the history is described as the sequence of items that have been examined by the DM. In our model, an agent approves an item if it is (weakly) preferred to the threshold, and disapproves it otherwise. We provide two models to study both deterministic and stochastic approval behaviours that exhibit history dependency. We provide characterizations of both models based on simple and intuitive postulates.

Keywords: approval behaviour, history dependency.

JEL codes: D00.

## 1.1 Introduction

We study the history dependency in binary classification behaviour, in which a DM decides to 'approve' or 'disapprove' an item in a sequential manner. For example, a judge sentences a crime to be felony or misdemeanor, a doctor administers or withholds treatment to a patient, an HR manager admits or rejects an interviewee. In this chapter, we name these types of sequential binary classification behaviour as *approval*, a term borrowed from the pioneering work on binary classification in MMU (2021).

In many situations, approval decisions are found to be history dependent. For example, the experiment in Pepitone and DiNubile (1976) finds an assault was judged to be less serious if it followed a narrative of a particularly egregious crime. Poses and Anthony (1991) find that a doctor was more likely to diagnose a patient to have bacteremia if he or she could recall more bacteremic patients in the past month. In a mock job interview experiment, Rowe (1967) finds that the preceding interviewees' performance have a significant effect on the decision on the candidate being considered.[1]

Throughout the chapter, a history is described as a finite sequence of items.[2] We allow the items in the history to be identical, as sometimes the number of times an item appears in the history also influences approval. For example, if the items in the history represent advertisements for movies on a streaming service platform, then not only the order of the advertisements, but also the number of repeated times can affect approvals (DM's putting movies into playlist).

We first study deterministic approval behaviour. We assume deterministic approval responses to be represented by a function $a(\lambda)$ which gives the set of approved items (out of the grand set $X$) given history $\lambda$. This deterministic approval function is novel,

[1] Various reasons can cause the menu dependency in every instance of decision making. In these examples, the history dependencies among decisions on the crimes, and that on the interviewees are caused by contrast effect; and the history dependencies in the decisions on the patients are caused by availability heuristic.

[2] In accordance with MMU (2021), we use 'item' instead of 'alternative' to stress the fact that the DM is making approval instead of choice, so that there is no 'competition' between the items.

as: (i) $a(\lambda)$ is a set rather than a single element;[3] and (ii) the approved items do not have to be in the history, as it is possible to have $x \in a(\lambda)$ but $x \notin \lambda$.[4] These two features directly illustrate the novelty of approvals compared to choices, as a DM possibly approves any item(s) from the grand set $X$ given any history, while a DM chooses one and only one alternative from the menu (i.e., the set of items that is faced by the DM).

We develop a *deterministic history-dependent approval function* (DHAF) $a_{\succsim,\eta}(\lambda) = \{y \in X : y \succsim \eta(\lambda)\}$, where $\succsim$ is a stable weak preference over $X$, and $\eta$ is a threshold rule that maps a history to an item (threshold). The approval making process here is simple: given history $\lambda$, a DM approves item $x$ if it is (weakly) preferred to the threshold $\eta(\lambda)$. The threshold $\eta(\lambda)$ is simply an item in the grand set $X$, and is dependent on the history, so that the approval decisions appear to be history dependent. Note that $\eta(\lambda)$ can be an item that has not been examined, so that we allow $\eta(\lambda) \notin \lambda$. For example, a doctor can use a patient example in the medical textbook as the threshold. Our first result provides a simple and intuitive *Approval Consistency* axiom that characterizes a DHAF. In a similar spirit to Samuelson's Weak Axiom of Revealed Preference (WARP), Approval Consistency axiom says that, if $x$ is approved and $y$ is disapproved given a history, then there does not exist any history given which $x$ is disapproved and $y$ is approved.

The approval data from many experimental and marketing research exhibit stochasticity.[5] To deal with this type of data, we then discuss the stochastic approval behaviour. We assume stochastic approval responses to be given by a function $p(x, \lambda)$ that indicates the probability that item $x$ is approved given history $\lambda$. This approval function

[3]In classical deterministic choice models, it is possible that a DM chooses more than one alternative from the menu. Still, our deterministic approval function shows some differences, which we will discuss later.

[4]Sometimes we abuse the notations and use $x \in \lambda$ if $x$ is in history $\lambda$, and $x \notin \lambda$ otherwise (note a history is a sequence, not a set).

[5] See, e.g., Elstein (1988), Gibbons and Marshall (2010), Danziger, Levav, and Avnaim-Pesso (2011).

$p(x, \lambda)$ is also new, as: (i) we do not have the adding-up constraint $\sum_{x \in X} p(x, \lambda) = 1$, as it is not a probability distribution; and (ii) we do not impose the positivity constraint $p(x, \lambda) > 0$ only if $x \in \lambda$. These two features are natural extensions from the deterministic approval responses discussed before.

Next, we offer a *random history-dependent approval function* (RHAF) that accommodates history dependency in stochastic approval behaviour. The approval probability is expressed as $p_{\succsim,t}(x, \lambda) = \sum_{y \in X : x \succsim y} t(y, \lambda)$, where $\succsim$ is a stable weak preference, and $t$ is a distribution over thresholds. The history dependency among stochastic approval behaviour is explained as the history dependency in the threshold, as captured by $t(y, \lambda)$. Given different histories, an item can be taken as the threshold with different probabilities. Our second result identifies a *Random Approval Consistency* axiom to characterize an RHAF. This axiom states that if item $x$ is approved with a *strictly* higher probability compared to item $y$ given a history, then $x$ must be approved with a *weakly* higher probability than $y$ given any other history. Random Approval Consistency axiom can be seen as a stochastic version of Approval Consistency axiom. The RHAF has a desirable property as it shows unique identification of the preference, and of the threshold distribution under a mild assumption that any item is taken as the threshold with positive probabilities.

Approval is a new topic in the literature. MMU (2021) develop a satisficing acceptability-continuation (SAC) function to study stochastic approval behaviour.[6] The *first* key feature that distinguishes the RHAF with the SAC function is that we allow the threshold distribution to be history dependent, so that the approval responses can exhibit menu dependency. The *second* key difference between the RHAF and the SAC function is the assumption on the DM's attention. The SAC function assumes imperfect attention, namely the DM stops making approval decisions with some probability given any his-

[6]MMU (2021) also develop an acceptability-continuation (AC) function which is proved to be *equivalent* to an SAC function. We focus on the comparison between our models and the SAC model because they have a similar structure.

tory.[7] In contrast, throughout this chapter, we apply the hypothesis that the DM must make approval decision for the current item at stake, i.e. perfect attention. While our perfect attention requirement may seem demanding, as for medical diagnoses, judicial trials, candidate admissions, in many cases the DM has to make decision for every item that is faced by him or her.

To better clarify these two differences, we also provide a random history-independent approval function (RHIAF) $p_{\succsim,\omega}(x,\lambda) = \sum_{y\in X: x\succsim y} \omega(y)$, where $\omega$ is a threshold distribution without history dependency. An RHIAF fails to accommodate the class of history dependent behaviour found in psychological and economic studies. For example, the study on contrast effect in Bhargava and Fisman (2014) finds that prior partner attractiveness reduces the likelihood of dating decision for the subsequent partner in a speed dating experiment. We provide characterization of an RHIAF, and compare the axiomatization of an RHIAF with that of an RHAF. We also relate the RHIAF to the SAC function: it is shown that an RHIAF (which is a special type of an RHAF) is a special type of an SAC function.

The inspiration of the definition of history is drawn from the research on choice from lists.[8] The idea of approving an item that is preferred to the threshold is similar to the work in Kovach and Ulku (2020), which studies choice behaviour, so that the first alternative which is preferred to the threshold along the list is chosen. In Rubinstein and Salant (2006), Aguiar, Boccardi and Dean (2016), the source of stochasticity of choice from lists is the randomness in lists. In our model, the history (list) is considered to be observable, and it is therefore deterministic. The stochasticity in our model comes from the randomness in the threshold. The logit categorization model in Wang (2021) focuses on binary categorization behaviour.[9] Binary categorization and approval are similar as

[7]The failure of perfect attention is the research focus of MMU (2021). In research on choice behaviour, limited attention has been more and more popular, e.g., Manzini and Mariotti (2014), Cattaneo, Ma, Masatlioglu, and Suleymanov (2020), Dardanoni, Manzini, Mariotti, and Tyson (2020).

[8]See, e.g., Horan (2011), Yildiz (2016).

[9]Wang (2021) is based on the work of Chapter 2.

a DM puts items into binary classes in both cases. However, the categorization decisions in Wang (2021) are not made sequentially, so that the order of items does not influence the approval decisions.

## 1.2 The Deterministic Model and Examples

### 1.2.1 Notation and Definitions

Let $X$ be a finite set of items, and let $\mathcal{D}$ be the domain of subsets of $X$. A history $\lambda$ is a finite sequence of elements of any subset of $X$. Let $\Lambda$ be the finite set of all histories. We allow the items in a history to be identical, for instance, we can have $\bar{\lambda} = xyzy$. We allow $\lambda$ to be empty, which we denote by $\lambda_0$. Here $\lambda_0$ represents the case for which the DM has examined no item before.

Sometimes we abuse notation, and treat histories as sets. We use $x \in \lambda$ to denote the case that $x$ appears in history $\lambda$, $x \notin \lambda$ to denote the case that $x$ does not appear in history $\lambda$. We first define a general deterministic approval rule.

**Definition 1.** A deterministic approval function $a$ is a map: $a : \Lambda \to \mathcal{D}$ such that $a(\lambda) \in \mathcal{D}$ for all $\lambda \in \Lambda$.

The deterministic approval function maps history $\lambda \in \Lambda$ to set of approved items $a(\lambda) \in \mathcal{D}$. Definition 1 is very permissive, the reason is that an item can be approved even if it has not been examined before. Definition 1 is different from a deterministic choice rule, as the approved item(s) can be any item in the grand set $X$, whereas the chosen alternative must be in the menu (i.e., the set $A \in \mathcal{D}$ that the DM is faced with).

Next, we provide a deterministic history-dependent approval function (DHAF). We denote by $x_0 \notin X$ a pseudo-item to capture the case in which none of the items in $X$ is approved ($a(\lambda) = \emptyset$). We assume that the agent has a stable weak preference ordering

$\succsim$ on $X$, and a threshold rule $\eta$. The threshold rule maps history $\lambda$ to threshold $\eta(\lambda)$, which is an item in $X \cup \{x_0\}$. In this model, item $x$ is approved if it is weakly preferred to threshold $\eta(\lambda)$ given history $\lambda$, and is disapproved otherwise. If the threshold is $x_0$, all the items in $X$ are disapproved. We state this formally as follows.

**Definition 2.** A deterministic history-dependent approval function (DHAF) is a deterministic approval function $a_{\succsim,\eta} : \Lambda \to \mathcal{D}$ for which there exists a pair $(\succsim, \eta)$, where $\succsim$ is a weak order on $X$, $\eta$ is a map $\eta : \Lambda \to X \cup \{x_0\}$, such that for all $\lambda \in \Lambda$,

$$a_{\succsim,\eta}(\lambda) = \begin{cases} \{y \in X : y \succsim \eta(\lambda)\}, & \text{if} \quad \eta(\lambda) \in X, \quad \text{and} \\ \emptyset, & \text{if} \quad \eta(\lambda) = x_0. \end{cases}$$

In this case, we say that $a$ is generated by $(\succsim, \eta)$. Note that a choice correspondence in standard utility maximization models allows multiplicity as well, if the chosen alternatives are indifferent to each other. However, in a DHAF, it is possible that $x, y \in a(\lambda)$ and $x \in a(\bar{\lambda})$, $y \notin a(\bar{\lambda})$,[10] which cannot be accommodated by the standard utility maximization models. This is because the DM does not maximize the utility given a history. Instead, the DM approves everything that meets the threshold.[11]

### 1.2.2 Examples

To demonstrate the richness of our framework, and motivate our following analysis, we describe several examples of families of approvals that exhibit history dependencies in psychological and economic studies.

[10]This happens if $x \succ y \succsim a(\lambda)$,and $x \succsim a(\bar{\lambda}) \succ y$.

[11]The DM who follows a DHAF acts similar to a 'satisficer' in Simon (1956). However, instead of choosing the first 'satisficing' alternative, the DM approves all the items from $X$ that are found to be satisficing.

**Example 1.** (Contrast Effect (Bhargava and Fisman 2014)) A DM classifies the items into 'good' class (approval) or 'bad' class (disapproval). The same item is more likely to be approved if it is preceded by a less preferred item, and vice versa. Contrast effect can happen, if, for example, the DM treats the most recently examined item as the threshold.

**Example 2.** (The first item dictates the threshold (Tversky and Kahneman 1991)) A DM's approval decisions are completely determined by the first item in the history. This can be explained if the first item in the history serves as a reference point according to which the DM forms the threshold. In other words, the first item in the history determines the threshold completely.

**Example 3.** (Availability Heuristic (Tversky and Kahneman 1973)) A DM's approval decisions are determined by the ease with which relevant instances come to mind. Availability heuristic can happen if the threshold rule defines the item which has been examined for the most times as the threshold; if several items have been examined for the same maximum times, the most recently examined one is taken as the threshold.[12]

## 1.3 Characterization and Identification of the DHAF

Our first main result is characterization of the DHAF. This characterization can serve as a benchmark for future studies on approval behaviour.

[12] In examples 1, 2 and 3, the history cannot be empty. To complete the discussion, we can assume $\eta(\lambda_0)$ be an arbitrary item in $X$.

**Axiom 1.** *(Approval Consistency) If there exists* $\lambda \in \Lambda$ *such that* $x \in a(\lambda), y \notin a(\lambda)$*, then there does not exist* $\bar{\lambda} \in \Lambda$ *such that* $x \notin a(\bar{\lambda}), y \in a(\bar{\lambda})$.

Approval Consistency axiom states that if $x$ is approved while $y$ is disapproved given history $\lambda \in \Lambda$, then it cannot be that $x$ is disapproved while $y$ is approved given any other history $\bar{\lambda} \in \Lambda$. By assuming the stable preference over $X$, we give our DHAF this 'rational' property, as in the classical deterministic choice model in Samuelson (1938). Indeed, although we are studying different objects, this axiom is similar to the standard WARP pioneered by Samuelson (1938), which says that if alternative $x$ is chosen when alternative $y$ is available, then $y$ is not chosen when $x$ is available. Our characterization result is as follows.

**Theorem 1.** *A deterministic approval function is a DHAF* $a_{\succsim,\eta}$ *if and only if Approval Consistency is satisfied.*

*Proof.* Sufficiency: For any $x, y \in X$, let $xPy$ if there does not exist $\lambda \in \Lambda$ such that $y \in a(\lambda), x \notin a(\lambda)$. We show that $P$ is complete and transitive.

For completeness, assume that there does exist $\bar{\lambda} \in \Lambda$ such that $y \in a(\bar{\lambda}), x \notin a(\bar{\lambda})$. By Approval Consistency, there does not exist $\lambda \in \Lambda$ such that $x \in a(\lambda), y \notin a(\lambda)$, so that $yPx$.

For transitivity, assume that $xPy$, and $yPz$. There does not exist $\lambda \in \Lambda$ such that $y \in a(\lambda), x \notin a(\lambda)$, or $z \in a(\lambda), y \notin a(\lambda)$. By contradiction, assume $\neg(xPz)$, there does exist $\hat{\lambda} \in \Lambda$ such that $z \in a(\hat{\lambda}), x \notin a(\hat{\lambda})$. Note that by $x \notin a(\hat{\lambda})$, we have $y \notin a(\hat{\lambda})$, as $xPy$. Once we have $y \notin a(\hat{\lambda})$, it must be $z \notin a(\hat{\lambda})$, as $yPz$. A contradiction.

Define $\succsim = P$. Next, define $\eta(\lambda) = \min(\succsim, a(\lambda))$ if $a(\lambda) \neq \emptyset$; and $\eta(\lambda) = x_0$ if

$a(\lambda) = \emptyset$. We have $a_{\succsim,\eta}(\lambda) = \{y \in X : y \succsim \eta(\lambda)\} = a(\lambda)$ if $a(\lambda) \neq \emptyset$; and $a_{\succsim,\eta}(\lambda) = \emptyset = a(\lambda)$ otherwise.

Necessity: For a DHAF $a_{\succsim,\eta}$, for all $x, y \in X$, if for some $\lambda \in \Lambda$, $x \in a(\lambda), y \notin a(\lambda)$, we can infer $x \succsim \eta(\lambda) \succ y \Rightarrow x \succ y$. Then, it cannot be $y \succsim \eta(\bar{\lambda}) \succ x$ for all $\bar{\lambda} \in \Lambda$.

□

Suppose the approval data is generated by $(\succsim, \eta)$, can we infer the preference ordering $\succsim$? What about the threshold rule $\eta$? In terms of preference, the weak order is revealed as $x \succsim y$ if there does not exist $\lambda \in \Lambda$ such that $x \notin a(\lambda), y \in a(\lambda)$. Immediately, we infer $x \sim y$ if there does not exist $\lambda \in \Lambda$ such that $x \notin a(\lambda), y \in a(\lambda)$, or $x \in a(\lambda), y \notin a(\lambda)$; and $x \succ y$ if there exists $\lambda \in \Lambda$ such that $x \in a(\lambda), y \notin a(\lambda)$.[13]

This identification property is intuitive. If there exists $\lambda \in \Lambda$ such that $x \in a(\lambda), y \notin a(\lambda)$, then $x$ is found to be weakly preferred to the threshold $\eta(\lambda)$, and $y$ is strictly preferred by $\eta(\lambda)$, then we conclude that $x \succ y$. If however, there does not exist $\lambda \in \Lambda$ such that $x \notin a(\lambda), y \in a(\lambda)$, or $x \in a(\lambda), y \notin a(\lambda)$, i.e., $x$ and $y$ are both approved or disapproved given any history $\lambda \in \Lambda$, we can only infer that $x$ and $y$ are both weakly preferred to, or strictly preferred by the threshold $\eta(\lambda)$ for all $\lambda \in \Lambda$. In this case, we arbitrarily assume $x \sim y$. After inferring $\succsim$, the threshold $\eta(\lambda)$ can be identified as any least preferred item in $a(\lambda)$ (the set of approved items), i.e., $\eta(\lambda) = \min(\succsim, a(\lambda))$. In this way, the threshold given history $\lambda$ is identified as the least preferred item from $a(\lambda)$ with respect to the preference $\succsim$.[14]

This identification is not unique. For a DHAF $a_{\succsim,\eta}$, for all $x, y \in X$, we can uniquely identify $x \succ y$ given the existence of $\lambda \in \Lambda$ with $x \in a(\lambda), y \notin a(\lambda)$. But the preference ordering over any other items remains arbitrary. This non-uniqueness in the identification of $\succsim$ leads to the non-uniqueness in the identification of $\eta$. For example, let $X = \{x_1, x_2, x_3\}$, $a(x_1x_1) = \{x_1x_2\}$ can be explained by a DHAF $a_{\succsim,\eta}$, for which: (i)$x_1 \sim x_2 \succ x_3$, and $\eta(x_1x_1) = x_1$ or $x_2$; or (ii) $x_2 \succ x_1 \succ x_3$, and $\eta(x_1x_1) = x_1$; or

[13]Note here $\succsim$ is complete, so $x \succ y$ is equivalent to $\neg(y \succsim x)$.

[14]If $a(\lambda) = \emptyset$, we immediately infer $\eta(\lambda) = x_0$.

(iii) $x_1 \succ x_2 \succ x_3$, and $\eta(x_1 x_1) = x_2$. The identification results (i) are inferred by the identification method we proposed.

## 1.4 The Stochastic Model

In Section 1.2 and 1.3, we have studied the deterministic environment. In this section, we move on to the stochastic approval domain. First, we extend the general deterministic approval function in Section 1.2 to our random approval function.

**Definition 3.** A random approval function $p$ is a map $p : X \times \Lambda \to [0,1]$, such that $p(x, \lambda) \in [0,1]$ for all $\lambda \in \Lambda$, for all $x \in X$.

In our model, an agent approves or disapproves item $x$ given history $\lambda$. Note that history $\lambda \in \Lambda$ is the list of items that has been examined by the DM. The random approval function simply generates the probability that $x$ is approved given history $\lambda$. This general rule is novel in several ways: it requires neither the positivity $p(x, \lambda) > 0$ only if $x \notin \lambda$, nor the adding-up constraint $\sum_{y \in X} p(y, \lambda) = 1$. The reason is obvious: the DM is in the process of approving, not choosing. It is possible that the DM approves an item even if it has not been examined before. Given history $\lambda$, the DM does not have to approve one and only one item from $X$, so that $\sum_{y \in X} p(y, \lambda)$ can range from 0 to $|X|$.

Next, we describe the (unobserved) distribution over thresholds.

**Definition 4.** A random threshold function is a map: $t : X \cup \{x_0\} \times \Lambda \to [0,1]$ such that $\sum_{x \in X \cup \{x_0\}} t(x, \lambda) = 1$, for all $\lambda \in \Lambda$.

A random threshold function $t(x, \lambda)$ gives the probability that item $x$ is selected as the threshold given history $\lambda$. For each approval decision, one item (among $X \cup \{x_0\}$) is selected as the threshold by the DM, therefore we have $\sum_{x \in X \cup \{x_0\}} t(x, \lambda) = 1$, for all $\lambda \in \Lambda$. This stochasticity in thresholds leads to the stochasticity in approvals.

**Definition 5.** A random history-dependent approval function (RHAF) is a random approval function $p_{\succsim,t} : X \times \Lambda \to [0,1]$ for which there exists a pair $(\succsim, t)$, where $\succsim$ is a weak order on $X$, $t$ is a random threshold function, such that for all $\lambda \in \Lambda$, for all $x \in X$,

$$p_{\succsim,t}(x, \lambda) = \sum_{y \in X : x \succsim y} t(y, \lambda).$$

In this case, we say that $p$ is generated by $(\succsim, t)$. The weak order $\succsim$ is interpreted as a standard preference relation over items, and $t(x, \lambda)$ is the threshold distribution over $X \cup \{x_0\}$ given history $\lambda$. The interpretation is that a DM (with history $\lambda$) approves item $x$ if it is weakly preferred to the threshold given history $\lambda$. If $x_0$ is taken as the threshold, no item from $X$ is approved. The threshold is generated stochastically and is history dependent, which makes the approval behaviour stochastic and history dependent.

### 1.4.1 Characterization and Identification

For the sake of completeness, we provide the characterization of an RHAF. Here, an RHAF is fully characterized from a simple axiom on the observed approval frequencies.

**Axiom 2.** *(Random Approval Consistency)* $p(x, \lambda) > p(y, \lambda) \Rightarrow p\left(x, \bar{\lambda}\right) \geq p\left(y, \bar{\lambda}\right)$, *for all* $\lambda, \bar{\lambda} \in \Lambda$, *for all* $x, y \in X$.

Along similar lines of Approval Consistency introduced in Section 1.3, Random Approval Consistency states that, if the approval probability of $x$ is strictly higher than that of $y$ given history $\lambda$, then the approval probability of $x$ is weakly higher than that of $y$ given any other history $\hat{\lambda}$. The characterization result is as follows.

**Theorem 2.** *A random approval function is an RHAF $p_{\succsim,t}$ if and only if Random Approval Consistency is satisfied.*

*Proof.* Sufficiency: Define $x\hat{P}y$ if $p(x,\lambda) \geq p(y,\lambda)$ for all $\lambda \in \Lambda$. We show $\hat{P}$ is complete and transitive.

For completeness, suppose that $p(y,\lambda) > p(x,\lambda)$ for some $\lambda \in \Lambda$. By Random Approval Consistency, $p\left(y,\bar{\lambda}\right) \geq p\left(x,\bar{\lambda}\right)$ for all $\bar{\lambda} \in \Lambda$, so that $y\hat{P}x$.

For transitivity, suppose that $p(x,\lambda) \geq p(y,\lambda)$ and $p(y,\lambda) \geq p(z,\lambda)$ for all $\lambda \in \Lambda$, then we immediately have $p(x,\lambda) \geq p(z,\lambda)$ for all $\lambda \in \Lambda$.

Define $\succsim = \hat{P}$. Fix history $\lambda \in \Lambda$. Let $\sim_1, ..., \sim_j$ be the indifference class of $\succsim$ on $X$ ordered from best to worst, which are the equality classes of $p(x,\lambda)$ for all $x \in X$. Fix $x_1, x_2, ...x_j$ such that $x_k \in \sim_k$ for every $k$. Define $t$ such that for all $\lambda \in \Lambda$, for all $y \in X$,

$$t(y,\lambda) = \begin{cases} \frac{p(x_k,\lambda)-p(x_{k+1},\lambda)}{|\sim_k|} & \text{for every} \quad y \in \sim_k, \quad k < j, \quad \text{and} \\ \frac{p(x_j,\lambda)}{|\sim_j|} & \text{for every} \quad y \in \sim_j \, . \end{cases}$$

Note that for all $\lambda \in \Lambda$, for all $y \in X$, $t(y,\lambda) \in [0,1]$; and $\sum_{y \in X} t(y,\lambda) = p(x_1,\lambda) \leq 1$. Define $t(x_0,\lambda) = 1 - p(x_1,\lambda)$. We have $t$ well-defined as a random threshold function.

For $x_j \in \sim_j$, we have $p(x_j,\lambda) = \sum_{z \in \sim_j} t(z,\lambda)$. For any $x_k \in \sim_k$ $(k < j)$, there must be

$$
\begin{aligned}
p(x_k,\lambda) &= \sum_{z\in\sim_k} t(z,\lambda) + p(x_{k+1},\lambda)\\
&= \sum_{z\in\sim_k\cup\sim_{k+1}} t(z,\lambda) + p(x_{k+2},\lambda)\\
&= \ldots\\
&= \sum_{z\in\sim_k\cup\ldots\cup\sim_j} t(z,\lambda) = \sum_{y\in X: x_k\succsim y} t(y,\lambda)
\end{aligned}
$$

Necessity: Let $p$ be generated by $(\succsim,t)$. For Random Approval Consistency, for all $\lambda,\bar{\lambda}\in\Lambda$, for all $x,y\in X$, we have

$$
\begin{aligned}
p_{\succsim,t}(x,\lambda)\geq p_{\succsim,t}(y,\lambda) &\Leftrightarrow \sum_{z\in X:x\succsim z} t(z,\lambda) \geq \sum_{z\in X:y\succsim z} t(z,\lambda)\\
&\Leftrightarrow x\succsim y\\
&\Leftrightarrow \sum_{z\in X:x\succsim z} t\left(z,\bar{\lambda}\right) \geq \sum_{z\in X:y\succsim z} t\left(z,\bar{\lambda}\right)\\
&\Leftrightarrow p_{\succsim,t}\left(x,\bar{\lambda}\right)\geq p_{\succsim,t}\left(y,\bar{\lambda}\right)
\end{aligned}
$$

□

Next, we take the perspective of an observer of a random approval function generated by an RHAF $p_{\succsim,t}$. Can we infer the primitives if the approval frequencies are observed? Comparing to a DHAF of which the identification is not unique, our first finding is that an RHAF can be substantially identified.

**Theorem 3.** *Let $p$ be a random approval function generated both by $(\succsim,t)$ and by $(\succsim',t')$. Suppose that $t(z,\lambda)>0$, $t'(z,\lambda)>0$ for all $\lambda\in\Lambda$, for all $z\in X$. Then:*

*(i) $\succsim=\succsim'$;*

*(ii) For all $\lambda\in\Lambda$, for all $x\in X$: $\sum_{z\sim x} t(z,\lambda)=\sum_{z\sim' x} t'(z,\lambda)$.*

*Proof.* (i) For all $\lambda \in \Lambda$, for all $x, y \in X$, $x \succsim y \Leftrightarrow \sum_{z \in X: x \succsim z} t(z, \lambda) \geq \sum_{z \in X: y \succsim z} t(z, \lambda) \Leftrightarrow p(x, \lambda) \geq p(y, \lambda)$.[15] The same implication holds for $\succsim'$, so $x \succsim y \Leftrightarrow p(x, \lambda) \geq p(y, \lambda) \Leftrightarrow x \succsim' y$.

(ii) Fix history $\lambda \in \Lambda$. Let $\sim_1, ..., \sim_j$ be the indifference class of $\succsim = \succsim'$ ordered from best to worst over $X$, which are the equality classes of $p(x, \lambda)$. For any $x_j \in \sim_j$, $\sum_{z \sim x_j} t(z, \lambda) = \sum_{z \sim x_j} t'(z, \lambda) = p(x_j, \lambda)$. For any $x_k \in \sim_k$ $(k < j)$ there is $\sum_{z \sim x_k} t(z, \lambda) = \sum_{x_k \succsim z} t(z, \lambda) - \sum_{x_{k+1} \succsim c} t(z, \lambda) = p(x_k, \lambda) - p(x_{k+1}, \lambda)$. Since the same implications hold with $t'$ in place of $t$, $\sum_{z \sim x_k} t(z, \lambda) = \sum_{z \sim x_k} t'(z, \lambda) = p(x_k, \lambda) - p(x_{k+1}, \lambda)$.

□

Theorem 3 shows that, the preference $\succsim$ is revealed uniquely ($x \succsim y$ iff $p(x, \lambda) \geq p(y, \lambda)$ for some $\lambda \in \Lambda$) given the observation of $p(x, \lambda)$ under a mild assumption. The stochastic threshold distribution $t$ is identified uniquely as well, since $\sum_{z \sim x} t(z, \lambda)$, the total probabilities of items that are indifferent with $x$ being the threshold given history $\lambda$ are inferred. The allocation of the probability mass among indifferent items remains free.

### 1.4.2 History Independent Approval

In this section, we specialize the RHAF and develop a random history-independent approval function (RHIAF). We then compare the axiomatization of an RHAF to that of an RHIAF.

**Definition 6.** A random history-independent approval function (RHIAF) is a random approval function $p_{\succsim,\omega} : X \times \Lambda \to [0, 1]$ if there exists a pair $(\succsim, \omega)$, where $\succsim$ is a weak order on $X$, $\omega$ is a probability distribution on $X \cup \{x_0\}$, such that for all $\lambda \in \Lambda$, for all

[15] Note that this does not hold if we allow $t(z, \lambda) = 0$ for some $\lambda \in \Lambda$, for some $z \in X$. For example, the observation of $p(x, \lambda) \geq p(y, \lambda)$ can be generated from $y \succ x$ with $\sum_{z \in X: y \succsim z \succ x} t(z, \lambda) = 0$.

$x \in X$,

$$p_{\succsim,\omega}(x,\lambda) = \sum_{y \in X : x \succsim y} \omega(y).$$

In this case we say that $p$ is generated by $(\succsim, \omega)$. It is not hard to see that an RHIAF $p_{\succsim,\omega}$ belongs to an RHAF $p_{\succsim,t}$, by defining $t(y,\lambda) = \omega(y)$, for all $\lambda \in \Lambda$ for all $y \in X \cup \{x_0\}$. What distinguishes an RHIAF from other types of RHAF is that the threshold distribution is assumed to be history independent in an RHIAF. In that sense, an RHIAF is a specialized history independent version of an RHAF.

It appears that an RHIAF generates the same approval probability of item $x$ given all $\lambda \in \Lambda$. Apparently, an RHIAF cannot accommodate history dependent approval probabilities, e.g., the contrast effect, or the availability heuristic discussed in Section 1.2. Indeed, this history independent approval property is exactly the axiom that characterizes an RHIAF.

**Axiom 3.** *(History Independent Random Approval)* $p(x,\lambda) = p\left(x,\bar{\lambda}\right)$ *for all* $\lambda, \bar{\lambda} \in \Lambda$, *for all* $x \in X$.

The characterization result is as follows.

**Proposition 1.** *A stochastic approval function is an RHIAF* $p_{\succsim,\omega}$ *if and only if History Independent Random Approval is satisfied.*

*Proof.* Sufficiency: Define $x\tilde{P}y$ if $p(x,\lambda) \geq p(y,\lambda)$ for some $\lambda \in \Lambda$. By History Independent Random Approval, $p(x,\lambda) \geq p(y,\lambda) \Rightarrow p\left(x,\bar{\lambda}\right) \geq p\left(y,\bar{\lambda}\right)$ for all $\lambda, \bar{\lambda} \in \Lambda$, for all $x, y \in X$. We immediate find that $\tilde{P}$ is complete and transitive.

Define $\succsim = \tilde{P}$. Fix history $\lambda \in \Lambda$. Let $\sim_1, ..., \sim_j$ be the indifference class of $\succsim$ on $X$ ordered from best to worst, which are the equality classes of $p(x, \lambda)$. Fix $x_1, x_2, ...x_j$ such that $x_k \in \sim_k$ for every $k$. Define $\omega$ such that for all $\lambda \in \Lambda$, for all $y \in X$,

$$\omega(y) = \begin{cases} \frac{p(x_k,\lambda)-p(x_{k+1},\lambda)}{|\sim_k|} & \text{for every } \quad y \in \sim_k, \quad k < j, \quad \text{and} \\ \frac{p(x_j,\lambda)}{|\sim_j|} & \text{for every } \quad y \in \sim_j \ . \end{cases}$$

By History Independent Random Approval axiom, $p(x, \lambda) = p\left(x, \bar{\lambda}\right)$, for all $\lambda, \bar{\lambda} \in \Lambda$, for all $x \in X$. Note here for all $\lambda \in \Lambda$, for all $y \in X$, $\omega(y) \in [0, 1]$; and $\sum_{y \in X} \omega(y) = p(x_1, \lambda) \leq 1$. Define $\omega(x_0) = 1 - p(x_1, \lambda)$, we have $\omega$ well defined as a distribution over $X \cup \{x_0\}$. For any $x_j \in \sim_j$, we have $p(x_j, \lambda) = \sum_{z \in \sim_j} \omega(z)$. For any $x_k \in \sim_k$ $(k < j)$ there must be

$$\begin{aligned} p(x_k, \lambda) &= \sum_{z \in \sim_k} \omega(z) + p(x_{k+1}, \lambda) \\ &= \sum_{z \in \sim_k \cup \sim_{k+1}} \omega(z) + p(x_{k+2}, \lambda) \\ &= ... \\ &= \sum_{z \in \sim_k \cup ... \cup \sim_j} \omega(z) \\ &= \sum_{z \in X : x_k \succsim z} \omega(z) \end{aligned}$$

Necessity: Let $p$ be generated by $(\succsim, \omega)$. It must be $p_{\succsim,\omega}(x, \lambda) = \sum_{y \in X : x \succsim y} \omega(y) = p_{\succsim,\omega}\left(x, \bar{\lambda}\right)$ for all $\lambda, \bar{\lambda} \in \Lambda$, for all $x \in X$, so that History Independent Random Approval holds.

□

Next, we compare the RHIAF to the SAC function in MMU (2021). The SAC function focuses on approval behaviour in which limited attention arises. Formally, the probability of approving item $x$ given history $\lambda$ in an SAC function is $p(x,\lambda) = \pi(\lambda)\sum_{y\in X: x\succsim y}\tau(y)$,[16] where $\pi(\lambda)$ is the probability that the DM 'considers' (makes decision for) item $x$ given history $\lambda$ (i.e. the probability of paying attention to item $x$ given history $\lambda$), and $\tau(y)$ is the probability that $y$ is taken as the threshold (which serves the same role as $\omega$). An RHIAF is a different special type of an SAC function, in which $\pi$ is restricted as follows: $\pi(\lambda) = 1$ for all $\lambda \in \Lambda$. In this case, the DM has to make approval decision for any item that is faced by him or her given any history (i.e., perfect attention). Therefore, we come to the conclusion that an RHIAF connects the RHAF and the SAC function: an RHIAF is a history independent version of an RHAF, or a perfect attention version of an SAC function. Indeed, the RHIAF emerges if we impose the perfect attention assumption ($\pi(\lambda) = 1$ for all $\lambda \in \Lambda$) on the SAC function, or the history independence assumption ($t(y,\lambda) = \omega(y)$ for all $\lambda \in \Lambda$, for all $y \in X \cup \{x_0\}$) on the RHAF.

## 1.5 Concluding Remarks

### 1.5.1 When the History Comprises Non-items

Throughout the chapter, we describe the history as a sequence of examined items. However, there are other factors that might affect the current approval decision. It is interesting to see that we can define the history as a set of elements out of $X$, without changing the main results in this chapter. For example, we can treat weather, education, working experience, etc., as the elements in history. The identification results and characterization results do not change. This property is actually in line with the difference between approval and choice: the item at stake does not have to be in the

[16] We make some tiny changes to the SAC function in MMU (2021) to simplify the comparison. MMU (2021) use $\lambda_x$ to denote the sequence of examined items (which is equivalent to $\lambda$ in this chapter).

history to be approved, so that the history can be anything that is considered relevant by a modeler. This flexibility provides new insights and interpretation of the our models.

### 1.5.2 When the History Comprises Distinct items

In our model, we allow the history to have identical items. Yet, a DM rarely faces repetitive items. For example, judges do not face the same case twice, doctors usually have new information for a return patient, HR managers rarely make decision for the same candidate repetitively. We should point out that, while it is tempting to take history with distinct items as a special type of our models, they are subtly different. For example, let $\lambda = x_1x_2$, then we cannot observe $p(x_1, x_1x_2)$ or $p(x_2, x_1x_2)$ as they lead to the history being $x_1x_2x_1$, or $x_1x_2x_2$ when making the next approval decision. However, both $x_1x_2x_1$ and $x_1x_2x_2$ comprise identical items. This constraint is non-trivial for identification, as we need to observe both $p(x_1, x_1x_2)$ and $p(x_2, x_1x_2)$ to infer $t(y, x_1x_2)$. Still, the characterization results hold, and partial identification of $t$ is feasible in this model.[17]

### 1.5.3 Logit Function in Stochastic Approvals

The logit categorization function in Wang (2021) provides a model in which the categorization probability follows a 'logit-like' form. Similarly, we can express the approval probability of item $x$ given history $\lambda$ as $p(x, \lambda) = \frac{V(x)}{V(x)+W(\lambda)}$, in which $V(x), W(\lambda) \in \mathbb{R}_{++}$ can be viewed as crude measure of the utility of $x$, and that of the threshold given history $\lambda$ respectively. This is an RHAF, as it satisfies Random Approval Consistency.[18] This model maps history $\lambda$ directly to value of threshold $W(\lambda)$. Due to its 'logit like' representation, we can consider this model from the econometric perspective and express $p(x, \lambda) = \frac{e^{v(x)}}{e^{v(x)}+e^{w(\lambda)}}$, where $v(x)$ is the utility function of the observed char-

[17]For example, for an RHAF $p_{\succsim,t}$, let $X = \{x_1, x_2\}$, we can identify all the primitives except $t(y, x_1x_2)$ and $t(y, x_2x_1)$ (i.e., the threshold distribution given history $x_1x_2$ and $x_2x_1$), from $p(x_1, \lambda_0)$, $p(x_2, \lambda_0)$, $p(x_1, x_2)$, and $p(x_2, x_1)$.

[18]We can define $\succsim, t$ in an RHAF $p_{\succsim,t}$ as: $x \succsim y \Leftrightarrow V(x) \geq V(y)$, and $\sum_{y \in X: x \succsim y} t(y, \lambda) = \frac{V(x)}{V(x)+W(\lambda)}$.

acteristics of item $x$, and $w(\lambda)$ is the threshold function of the observed characteristics of history $\lambda$. This logit function in stochastic approvals can be useful in applied studies because of its 'logit-like' expression.

### 1.5.4 Limited Recall

Many psychological and marketing research find that agents have limited ability in recalling.[19] We can interpret the randomness in the threshold distribution as *imperfect recall*, i.e. the agent's failure to recall all the items.[20] To take the limited recall into account, we can write the approval probability of item $x$ given history $\lambda$ as $p(x,\lambda) = \sum_{y\in X:x\succsim y}\left(\rho(y,\lambda)\prod_{z\in X:z\succ_s y}(1-\rho(z,\lambda))\right)$, where $\rho(y,\lambda)$ is the probability that $y$ is recalled given history $\lambda$, and $\succ_s$ is a salience relation over $X$ which orders the salience of items. This is an RHAF, by defining the random threshold as $t(y,\lambda) = \rho(y,\lambda)\prod_{z\in X:z\succ_s y}(1-\rho(z,\lambda))$.[21] The interpretation is that, the most salient recalled item is taken as the threshold. Therefore, the probability of item $y$ being the threshold given history $\lambda$ is the probability that: (i) item $y$ is recalled, and (ii) no item $z$ that is more salient than item $y$ ($z \succ_s y$) is recalled.

## Acknowledgements

I would like to thank Bram De Rock, Thomas Demuynck, Matthew Ryan, John Rehbeck, for helpful comments and suggestions. A paper based on this chapter is under revise and resubmit at *Journal of Mathematical Economics*. I am grateful to Associate Editor Juan Dubra and two anonymous referees for their constructive and insightful comments.

[19]See, e.g. Buschke (1973), Alba and Chattopadhyay (1986).

[20]Imperfect recall was first introduced in Kuhn (1953) in economics.

[21]Here $t(x_0,\lambda) = \prod_{z\in X}(1-\rho(z,\lambda))$. This is a threshold version of the stochastic choice function in Manzini and Mariotti (2014).

# Chapter 2

# Logit Function in Stochastic Categorization

**Abstract**

Binary categorization refers to the behaviour whereby a DM puts the items from a menu into binary categories (authorization category and rejection category). For example, a judge approves or denies parole for inmates, a doctor administers or withholds treatment to patients, a teacher passes or fails essays. Categorizations are observed to exhibit randomness in various contexts, but the stochasticity in categorizations has received little attention in economic studies. We consider stochastic categorization using a logit categorization function, which expresses the probability of authorizing an item from a menu in a 'logit-like' form. We characterize the model from a simple condition on authorization frequencies. Furthermore, we derive an empirical version of our model, which not only provides insight into applications of the model, but also gives an intuitive interpretation of the stochasticity in categorization.

Keywords: logit model, stochastic categorization.

JEL codes: C10, D00.

## 2.1 Introduction

Consider the following cases:

(i) A judge approves or denies parole for inmates;

(ii) A doctor administers or withholds treatment to patients;

(iii) A teacher passes or fails essays.

These are all binary categorization problems in which a DM decides which items belong to a certain category.[1] The DM assigns each item from the menu into binary categories: 'authorization' category and 'rejection' category (e.g., approval and denial, treatment and withholding, pass and fail, etc.).

Analogous to the observed stochastic choice behaviour in experimental and market settings,[2] it is common to observe categorizations that exhibit variability. In Danziger, Levav, and Avnaim-Pesso (2011), it is reported that a judge was between two and six times as likely to approve parole if the inmate was one of the first three of six considered after a food break. Elstein (1988) finds that a doctor was less likely to administer a particular treatment protocol if able to recall a case of failed treatment. Along similar lines, Gibbons and Marshall (2010) show that even with a detailed marking rubric created by the teachers themselves, the same essay was marked differently. While there is extensively rich literature on stochastic choice, the stochasticity in binary categorization has thus far received little attention in the economic literature.

In this chapter, we assume authorization responses to be given by a function $p$ that indicates the probability $p(a, B)$ that item $a$ is 'authorized' from menu $B$. This authorization probability itself is new: unlike choice probability, it does not impose the adding-up constraint $\sum_{b \in B} p(b, B) = 1$. The reason is simple and clear: when making categorizations, a DM's aim is not to maximize the preference over the menu by choosing one and only one alternative as in choice making. Instead, the DM autho-

[1] The same with Chapter 1, we use 'item' instead of 'alternative' in our models in this chapter.
[2] See the discussion in Mcfadden (2000).

rizes all the items in the menu that meet the threshold. In this chapter, we study the stochasticity in categorization by expressing the authorization probability $p(a, B)$ in a 'logit like' form which is similar to the choice probability in Luce (1959). Formally, $p_{V,W}(a, B) = \frac{V(a)}{V(a)+W(B)}$, where $V(a)$, $W(B)$ are mappings from item $a$, menu $B$ to values (of utility and threshold) correspondingly. We name our model the *logit categorization function*. We find that a simple Independence of Irrelevant Alternatives (IIA) axiom on the ratio of authorization frequency to rejection frequency can fully characterize the logit categorization function. This IIA axiom on the authorization to rejection ratio is intuitive and gives insightful interpretation of the behavioural patterns in stochastic categorization in our model.

Mcfadden (1974) develops the multinomial logit model, which is shown to be equivalent to the Luce choice model in Luce (1959), but is much more popular in econometric studies as it can incorporate the characteristics of the alternatives. Inspired by the derivation of the multinomial logit model, we further derive a full-fledged version of the logit categorization function and investigate the empirical implications of the full-fledged model. Formally, we express the probability of authorizing item $a$ from menu $B$ by DM $i$ as $p(a|i, B) = \frac{e^{v(a,i)}}{e^{v(a,i)}+e^{w(B,i)}}$, where $v(a, i), w(B, i)$ are utility function and threshold function determined by the observed characteristics of item $a$, menu $B$, and DM $i$. In the full-fledged model, we assume that a DM relies on 'thresholds' to solve categorization problems, such that he or she 'authorizes' an item from a menu if the utility of the item passes the threshold of the menu, and 'rejects' it otherwise. Following this set-up, the probability of authorizing item $a$ from menu $B$, is the probability that the utility of item $a$ exceeds the threshold of menu $B$. The full-fledged model provides intuitive explanation of the source of stochasticity in categorization, i.e., the randomness in the utilities of items and the randomness in the thresholds of menus. As with the classical Luce-Multinomial Logit Choice model, we prove that the full-fledged model is equivalent to the logit categorization function. This full-fledged model is important, as it has great potential for use in applied work by making use of the observed characteristics of the available items and menus. Our framework is not only applicable to real world dataset, but also adds value to the existing literature. To demonstrate this, we discuss how our model can be employed to study choice overload and availability bias through simple examples.

The logit categorization function is a novel model which allows the thresholds to be not only stochastic, but also menu dependent. Analogous to the findings in classical choice models,[3] non-stochastic utilities and non-stochastic thresholds would naturally generate a fully rational deterministic categorization correspondence. However, a deterministic categorization correspondence cannot explain the variability in categorizations. Then, the binary logit model, which implicitly assumes the threshold to be menu independent, can potentially solve the stochastic categorization problem.[4] Nevertheless, it is plausible to assume that in some cases, the thresholds are dependent on the menus. For example, a consumer may be less likely to consider an item if there are more options, a marker may be stricter if the average qualities of the essays are better, and a college admission team can have different thresholds in different admission rounds. The logit categorization function can accommodate the categorizations with menu-dependent thresholds. The standard logit model can be applied to study categorization if it considers the object of choice to be a set of items. For example, let $B = \{a, b\}$, a standard logit model can redefine the menu as $\mathcal{B} = \{\emptyset, \{a\}, \{b\}, \{a, b\}\}$, the power set of $B$. Then, for instance, authorizing $a$ and $b$ from menu $B$ is equivalent to choosing $\{a, b\}$ from $\mathcal{B}$.[5] However, despite its appeal, this approach is computationally demanding, as the newly defined menu sizes grow exponentially.[6] This burden on

[3] See, e.g., Samuelson (1938) and Richter (1966).

[4] In fact, the binary logit model has been used in prior research. For example, the consideration set model in Goeree (2008) essentially uses binary logit model to denote the probability that an alternative is put into the consideration set (authorization).

[5] Manski and Sherman (1980) use this method to distinguish between households with one car or two cars.

[6] Specifically, menu $B$ with $J$ number of elements is redefined as the power set $\mathcal{B}$ with $2^J$ number of elements.

the computation makes the standard logit model unapplicable if the sizes of the menus are large. In contrast, our model is compatible with large menus as the menu sizes stay at the same level.

The categorization behaviour can be essentially seen as approval. In this chapter, we decide to use the terminology 'authorization' instead of 'approval' to highlight the difference: the rankings of the items play a role in sequential approval making, whereas the rankings do not influence the decision in authorization making. Note that the models in MMU (2021) are quite different from ours, as the domain comprises lists, which are different orders of the same menu.[7] Also, MMU's models do not allow for variations of menus, as their menu is fixed at the universal set $X$. Our model does not require the observation of orders, and it is compatible for categorizations from different menus. Mandler, Manzini, and Mariotti (2012) develop a choice by checklist model, in which choices are made by a series of categorical judgments (going through checklists). The categorical judgments in their model is similar in spirit to the categorization in this chapter. The categorical judgements are lexicographic, while in our model the categorizations are not.[8]

Finally, in recent years, there has been an increasing interest in the study of machine learning models, which are heavily used for categorization problems, e.g., inferring economists' partisanship from their published papers,[9] categorizing consumers into high and low credit-risk groups,[10] or classifying Twitter user's ethnicity.[11] While both can be used for categorization, logit categorization model can retrieve analytical information of the taste (preference) and the threshold compared to the machine learning method.

The remainder of the chapter is structured as follows. Section 2.2 introduces the logit categorization function. Next, in Section 2.3, we provide a full characterization

[7]The domain in Chapter 1 also comprises lists (histories).

[8]Although the choice by checklist has this lexicographic flavour, it is proved that the final choice satisfies utility maximization.

[9]See Jelveh, Kogut, and Naidu (2018).

[10]See Khandani, Kim, and Lo (2010).

[11]See Pennacchiotti and Popescu (2011).

of the logit categorization function with a simple axiom. In Section 2.4, we derive the full-fledged version of the logit categorization function, and focus on the empirical applicability of the model. This section also provides an intuitive explanation of the source of stochasticity in our model. Finally, Section 2.5 concludes the chapter.

## 2.2 The Model

We denote by $X$ the universe of items, and by $\mathcal{D}$ the domain of subsets (i.e., the menus) of $X$. We first define a random categorization function.

**Definition 7.** A random categorization function is a map: $p : X \times \mathcal{D} \to [0, 1]$ such that $p(a, B) = 0$ for all $a \notin B$, for all $B \in \mathcal{D}$; and $p(a, B) \in (0, 1)$ for all $a \in B$, for all $B \in \mathcal{D}$.

We assume categorization responses to be given by a map $p$ that indicates the probability $p(a, B)$ item $a$ is authorized from menu $B$, which is similar to choice probability in the work of Luce (1959), and Mcfadden (1974).[12] Definition 7 is very general, and only requires $p(a, B) = 0$ for all $a \notin B$, for all $B \in \mathcal{D}$ (i.e., only items in the menu can be authorized); and $p(a, B) \in (0, 1)$ for all $a \in B$, for all $B \in \mathcal{D}$ (i.e., every item in a menu has a positive probability of being authorized). Note that, unlike standard stochastic choice function, we do not impose the constraint $\sum_{b \in B} p(b, B) = 1$. The reason is straightforward: it is not required for an agent to authorize one and only one alternative from a menu as in choice making. Then, we define a new type of random categorization function that has a 'logit-like' expression.

**Definition 8.** A random categorization function is a logit categorization function if there exists a pair $(V, W)$, where $V$ is a map $V : X \to \mathbb{R}_{++}$, and $W$ is a map $W : \mathcal{D} \to$

[12]Obviously, $1 - p(a, B)$ can be interpreted as the rejection probability of $a$ from menu $B$.

$\mathbb{R}_{++}$, such that

$$p_{V,W}(a,B) = \frac{V(a)}{V(a) + W(B)}.$$

Here $V(a)$, $W(B)$ can be interpreted as the 'response strength' associated with item $a$ and threshold of menu $B$.[13] The logit categorization function is structurally similar to the standard choice function $p(a,B) = \frac{V(a)}{\sum_{b\in B} V(b)}$ in Luce (1959). Indeed, $p_{V,W}(a,B)$ can be seen as the probability that item $a$ is preferred to the threshold of menu $B$ according to 'Luce rule'. Note that we map a menu $B$ to a value $W(B)$ by the function $W$, so that we take account of the menu dependency of the threshold.[14] The randomness in the categorizations is simply provided by logit-like probabilities in authorizations, and it does not necessitate any menu-dependency of the threshold. With menu independent thresholds $W(A) = W(B)$ for all $A, B \in \mathcal{D}$, the categorizations in our model still display stochasticity. Indeed, we provide an underlying justification for the presence of stochasticity (i.e., due to the unobserved parts of utilities and thresholds) in the discussion of the full-fledged logit categorization model in Section 2.4.

## 2.3 Characterization

In this section, we characterize a logit categorization function from a simple condition on observed authorization frequencies. The axiom is intended for all $A, B \in \mathcal{D}$, for all $a, b \in A \cap B$. Our axiom puts a constraint on the ratio

$$\gamma(a,B) = \frac{p(a,B)}{1 - p(a,B)},$$

[13]The term 'response strength' is borrowed from Gulliksen (1953), and Luce (1959).

[14]This mapping from a menu to a value is similar to the set-up in Manzini, Mariotti, and Tyson (2013) & (2016).

which is the authorization probability of item $a \in B$ from menu $B \in \mathcal{D}$ over the respective rejection probability. We name $\gamma(a, B) = \frac{p(a,B)}{1-p(a,B)}$ the authorization to rejection ratio (AR ratio) for item $a$ from menu $B$. Given Definition 7, we have $p(a, B), 1 - p(a, B) \in (0, 1)$ for all $a \in B$, for all $B \in \mathcal{D}$. Therefore, the AR ratio $\gamma$ is positive and well defined for all $a \in B$, for all $B \in \mathcal{D}$. We have an Independence of Irrelevant Alternatives (IIA) axiom on the AR ratios.

**Axiom 4.** *($\gamma$-IIA)* $\frac{\gamma(a,A)}{\gamma(a,B)} = \frac{\gamma(b,A)}{\gamma(b,B)}$

By $\gamma$-IIA, $\frac{\gamma(a,A)}{\gamma(a,B)}$ (i.e., the AR ratio for item $a$ from menu $A$, over the AR ratio for item $a$ from menu $B$) is not dependent on the item $a$ at stake, but only on the two menus $A$ and $B$ that have been considered.

Our first main result is the following.

**Theorem 4.** *A random categorization function $p$ is a logit categorization function $p_{V,W}$ if and only if $p$ satisfies $\gamma$-IIA.*

*Proof.* Sufficiency: As discussed at the beginning of this section, the AR ratio $\gamma$ is positive and well defined. For all $a \in B$, and $B \in \mathcal{D}$, define the two maps in $p_{V,W}$ such that $V(a) = \gamma(a, X)$, and $W(B) = \frac{\gamma(a,X)}{\gamma(a,B)}$. By $\gamma$-IIA, $\frac{\gamma(a,X)}{\gamma(a,B)} = \frac{\gamma(b,X)}{\gamma(b,B)}$ for all $a, b \in B$, and $B \in \mathcal{D}$. So, $W(B) = \frac{\gamma(a,X)}{\gamma(a,B)} = \frac{\gamma(b,X)}{\gamma(b,B)}$ is fixed for a menu $B \in \mathcal{D}$, with $a, b \in B$, and is well defined. Then

$$
\begin{aligned}
p_{V,W}(a,B) &= \frac{\gamma(a,X)}{\gamma(a,X)+\frac{\gamma(a,X)}{\gamma(a,B)}} \\
&= \frac{1}{1+\frac{1}{\gamma(a,B)}} = \frac{\gamma(a,B)}{\gamma(a,B)+1} \\
&= \frac{p(a,B)}{p(a,B)+1-p(a,B)} = p(a,B)
\end{aligned}
$$

Necessity: Note that $V(a)$, and $W(B)$ are strictly positive, so $p_{V,W}(a,B) = \frac{V(a)}{V(a)+W(B)} \in (0,1)$; and we have

$$
\begin{aligned}
\frac{\gamma(a,A)}{\gamma(a,B)} &= \frac{p(a,A)}{1-p(a,A)}\frac{1-p(a,B)}{p(a,B)} \\
&= \frac{V(a)}{W(A)}\frac{W(B)}{V(a)} = \frac{V(b)}{W(A)}\frac{W(B)}{V(b)} \\
&= \frac{p(b,A)}{1-p(b,A)}\frac{1-p(b,B)}{p(b,B)} = \frac{\gamma(b,A)}{\gamma(b,B)}
\end{aligned}
$$

□

The standard IIA property ($\frac{p(a,A)}{p(a,B)} = \frac{p(b,A)}{p(b,B)}$) in Luce (1959) and Mcfadden (1974) is criticized for imposing a restrictive structure on substitution patterns. The well known blue bus/red bus example in Debreu (1960) shows that IIA is inappropriate for menus containing close substitutes.[15] A logit categorization function is fully characterized by $\gamma$-IIA instead of Luce's IIA, which immediately makes it compatible for violations of IIA. However, the problem of IIA in a choice does not necessarily arise in a categorization. This is because an item does not have to 'make room' for its close substitute in the dimension of probabilities, since the authorization probabilities of items in a menu do not have to add up to one. Instead, we discuss the limitation of $\gamma$-IIA, the axiom that

[15] Gul, Natenzon, and Pesendorfer (2010), Manzini and Mariotti (2014) develop different stochastic choice models to accommodate the violation of IIA in choice making.

characterizes our model, among categorizations behaviour.

For a logit categorization function $p_{V,W}$, the odds $\frac{\gamma(a,A)}{\gamma(a,B)} = \frac{\gamma(b,A)}{\gamma(b,B)} = \frac{W(B)}{W(A)}$, can be interpreted as the ratio of menu $B$'s threshold to menu $A$'s threshold. Here, $\gamma$-IIA states that this ratio is independent on the items being considered. When this ratio is observed to be dependent on the items at stake, the $\gamma$-IIA property is violated. Here is an example.

**Example 4.** (Violation of $\gamma$-IIA) A user shares post(s) on a social media website. First assume two posts $a, b$ by influencer 1 (denoted by $a_1, b_1$) are on the website (menu $C = \{a_1, b_1\}$), and let $p(a_1, C) = p(b_1, C) = \frac{2}{3}$.[16] Next, assume another post $b$ by influencer 2 (denoted by $b_2$) is added (new menu $D = \{a_1, b_1, b_2\}$). Here $b_1$ and $b_2$ are the same post, and the only difference is that they have been uploaded by different influencers. If the agent does not care about the influencer, one might expect the agent to share (authorize) post $a_1$ with same probability as before; and split the previous authorization probability equally for $b_1$ and $b_2$, as it makes little sense sharing both of them. Therefore, it is expected that $p(a_1, D) = p(a_1, C) = \frac{2}{3}$, and $p(b_1, D) = p(b_2, D) = \frac{p(b_1,C)}{2} = \frac{1}{3}$.

This is a violation of $\gamma$-IIA, as $\frac{\gamma(a_1,C)}{\gamma(a_1,D)} = 1$, but $\frac{\gamma(b_1,C)}{\gamma(b_1,D)} = 4$. Thus, it cannot be accommodated by our model. The thresholds of menus $C$ and $D$ for the authorization of item $a_1$ are the same, as $p(a_1, C) = p(a_1, D)$; however, the thresholds of menus $C$ and $D$ for the authorization of $b_1$ are different, as $\frac{p(b_1,C)}{2} = p(b_1, D)$. The user applies the same thresholds for the authorizations of $a_1$ from menus $C$ and $D$; but uses different thresholds for the authorizations of $b_1$ from menus $C$ and $D$. In this case, the ratio of menu $C$'s threshold to menu $D$'s threshold becomes dependent on the items being

[16]Again the authorization probabilities do not add up to one, as the agent is not committed to sharing only one post.

considered, which leads to the violation of $\gamma$-IIA.

Nevertheless, we ought to point out the 'narrowness' of this example, as we implicitly make the user 'choose' between $b_1$ and $b_2$. In so doing, a problem similar to the close substitutes in blue bus/red bus example comes up again. If the user does not have to choose between $b_1$ and $b_2$ (i.e., the user shares any post he finds better than the threshold), we would expect $p(a_1, C) = p(b_1, C) = p(a_1, D) = p(a_1, C) = \frac{2}{3}$. Then, $\gamma$-IIA is still satisfied.[17]

## 2.4 Full-Fledged Logit Categorization Model

In this section, we derive a full-fledged version of logit categorization function from simple adaptations of the multinomial logit model in Mcfadden (1974). What is more, these adaptations are in line with the differences between choice and categorization behaviour. In particular, we discuss this full-fledged model from the perspective of an econometrician, and show how it can be used for empirical analysis.

The full-fledged model assumes random utilities and random thresholds. Random utilities are commonly assumed in stochastic choice models.[18] However, random categorizations can be due to not only random utilities, but also random thresholds. For example, a parole judge can be stricter if hungry, a doctor can be more prudent if recalling a failed treatment, and a teacher's grading criteria can be subjective to his or her understanding of the marking rubric. The thresholds appear to be stochastic if these factors (i.e., hungriness of the judge, the doctor's recall of a failure, and the teacher's understanding of the marking rubric) cannot be observed. Indeed, we capture the randomness of thresholds by using an error term as in Mcfadden (1974).

In empirical research, we are usually concerned with the categorizations made by

[17] For simplicity purpose, here we assume the thresholds of menus $C$ and $D$ to be the same. Note that $\gamma$-IIA can still hold if the thresholds of menus $C$ and $D$ are different.

[18] See, e.g., Block and Marschak (1960), Mcfadden (1974).

individuals indexed by $i = 1, 2, ..., n$. Let $p(a|i, B)$ denote the probability that DM $i$ from the population will authorize item $a \in B$, given that he or she faces menu $B \in \mathcal{D}$. The adaptations are naturally in line with the essential difference between a categorization problem and a choice problem: in denoting by $t(B, i)$ the threshold of menu $B$ for DM $i$, we define $p(a|i, B)$ as $\Pr\{u(a, i) \geq t(B, i)\}$ (i.e., the probability that the utility of item $a$ is (weakly) higher than the threshold of menu $B$); rather than the choice probability $p(a|i, B) = \Pr\{u(a, i) \geq u(b, i) : \forall b \in B\}$ in standard multinomial logit model in Mcfadden (1974).

As in Mcfadden (1974), we assume that the utility of an item is the sum of a representative utility and an error term. The utility of item $a$ for DM $i$ is,

$$u(a, i) = v(a, i) + \varepsilon(a, i),$$

where $v(a, i)$ is non-stochastic and is the 'representative utility' of item $a$ for DM $i$, and $\varepsilon(a, i)$ is a stochastic error term, which reflects the unobserved utility of item $a$ for DM $i$. What is more, we define

$$t(B, i) = w(B, i) + \varepsilon(B, i),$$

as the threshold of menu $B \in \mathcal{D}$ for DM $i$. Note that $w(B, i)$ is non-stochastic and is the 'representative threshold' of menu $B \in \mathcal{D}$ for DM $i$, and $\varepsilon(B, i)$ is the random error term that reflects the unobserved part of the threshold of menu $B$ for DM $i$. Note here that the unobserved error term $\varepsilon(B, i)$ is not a function of $a \in B$, the item being considered. For DM $i$, the utilities of all the items $a \in B$ in menu $B$, $u(a, i)$, are compared against the same threshold, $t(B, i)$. Once $t(B, i)$ is generated, it is used for the categorizations of all $b \in B$ among menu $B$.

As seen from the above two equations for $u(a, i)$ and $t(B, i)$, the stochasticity of authorizations comes from the unobserved error term $\varepsilon(a, i)$ in $a$'s utility for DM $i$, as well as the unobserved error term $\varepsilon(B,i)$ in menu $B$'s threshold for DM $i$. We assume that the unobserved random part $\varepsilon(a,i)$ and $\varepsilon(B,i)$ are i.i.d. with type-1 extreme distribution.

The second main result of this chapter is proving that, under the assumptions in Section 2.4, the authorization probability has a logit categorization function representation.

**Theorem 5.** *Suppose that authorization probability of item $a \in B$ from menu $B \in \mathcal{D}$ by DM $i$ is given by the function $p(a|i,B) = \Pr\{u(a,i) \geq t(B,i)\}$. Also, suppose each member $i$ of a population has a utility function $u(a,i) = v(a,i) + \varepsilon(a,i)$, and a threshold function $t(B,i) = w(B,i) + \varepsilon(B,i)$, where $v$ is a non-stochastic function reflecting 'representative' tastes for the items, $w$ is a non-stochastic function reflecting 'representative' tastes for the menus, and $\varepsilon$ is a function that varies randomly in the population, such that the values $\varepsilon(a,i)$ and $\varepsilon(B,i)$ are i.i.d. with type-1 extreme distribution. Then, the authorization probability has a logit categorization function representation.*

*Proof.* The probability that agent $i$ with menu $B \in \mathcal{D}$, will authorize $a \in B$ equals

$$\begin{aligned}
p(a|i,B) &= \Pr[u(a,i) \geq t(B,i)] \\
&= \Pr[v(a,i) + \varepsilon(a,i) \geq w(B,i) + \varepsilon(B,i)] \\
&= \Pr[\varepsilon(a,i) \geq \varepsilon(B,i) + w(B,i) - v(a,i)].
\end{aligned}$$

The values $\varepsilon(a,i)$ and $\varepsilon(B,i)$ are both i.i.d. with type-1 extreme distribution. The density is $f(\varepsilon) = e^{-\varepsilon}e^{-e^{-\varepsilon}}$, and the cumulative distribution is $F(\varepsilon) = e^{-e^{-\varepsilon}}$. If $\varepsilon(B,i)$ is given, we can write the conditional probability as $p(a|i,B)\,|\varepsilon(B,i) = 1 - e^{-e^{-(\varepsilon(B,i)+w(B,i)-v(a,i))}}$. However $\varepsilon(B,i)$ is not given, instead we get $p(a|i,B)$ from the integral of $p(a|i,B)|\varepsilon(B,i)$ over all values of $\varepsilon(B,i)$ weighted by its density:

$$p(a|i,B) = \int_{-\infty}^{\infty} \left(1 - e^{-e^{-(\varepsilon(B,i)+w(B,i)-v(a,i))}}\right) e^{-\varepsilon(B,i)} e^{-e^{-\varepsilon(B,i)}} d\varepsilon(B,i).$$

Substitutes $\varepsilon(B,i)$ by $s$, we have

$$\begin{aligned} p(a|i,B) &= \int_{-\infty}^{\infty} \left(1 - e^{-e^{-(s+w(B,i)-v(a,i))}}\right) e^{-s} e^{-e^{-s}} ds \\ &= \int_{-\infty}^{\infty} e^{-s} e^{-e^{-s}} ds - \int_{-\infty}^{\infty} e^{-e^{-s}\left(e^{-(w(B,i)-v(a,i))}+1\right)} e^{-s} ds. \end{aligned}$$

Define $x = e^{-s}$ , we get

$$\begin{aligned} p(a|i,B) &= \int_{0}^{\infty} e^{-x} dx - \int_{0}^{\infty} e^{-x\left(e^{-(w(B,i)-v(a,i))}+1\right)} dx \\ &= \left[-e^{-x}\right]_0^{\infty} - \left[\frac{e^{-x\left(e^{-(w(B,i)-v(a,i))}+1\right)}}{e^{-(w(B,i)-v(a,i))}+1}\right]_0^{\infty} \\ &= \frac{e^{v(a,i)}}{e^{v(a,i)} + e^{w(B,i)}}. \end{aligned}$$

Since $e^{v(a,i)}$, and $e^{w(B,i)}$ are strictly positive, we can define $e^{v(a,i)} = V(a,i)$, and $e^{w(B,i)} = W(B,i)$, so that $p(a|i,B) = \frac{V(a,i)}{V(a,i)+W(B,i)}$.[19]

□

We name the newly derived categorization function $p(a|i,B) = \frac{e^{v(a,i)}}{e^{v(a,i)}+e^{w(B,i)}}$ the *full-fledged logit categorization model.* Most applied work define the representative utility of items as linear in the attributes of items $v(a,i) = \beta_i x_a$, where $x_a$ is the vector of the observed attributes of item $a$, and $\beta_i$ is the coefficients vector of agent $i$. We further extend this and define the representative threshold as $w(B,i) = \alpha_i z_B$, where

[19]Here we write $V(a,i)$, $W(B,i)$ instead of $V(a)$, $W(B)$, as we are expressing the authorization probability of item $a$ from menu $B$ by DM $i$.

$z_B$ is a vector of the observed attributes of menu $B$. We can write the full-fledged logit categorization model with linear-in-parameters as,

$$p\left(a|i,B\right)=\frac{e^{\beta_i x_a}}{e^{\beta_i x_a}+e^{\alpha_i z_B}}.$$

By making use of the observed characteristics of items and menus, the full-fledged model can be useful in empirical analysis. The authorization probability has a closed form, so that the traditional maximum-likelihood method can be applied. Assume an observer of categorizations from menu $B$ by DMs indexed by $i$ is obtained for the purpose of estimation. To do the estimation, we need to maximize the probability of the DMs authorizing the items that were actually observed to be authorized and rejecting the items that were actually observed to be rejected. In a full-fledged logit categorization model, the element used for a maximum-likelihood analysis is

$$\prod_i \prod_{a\in B} p\left(a|i,B\right)^{y(a|i,B)}\left(1-p\left(a|i,B\right)\right)^{1-y(a|i,B)},$$

where $y\left(a|i,B\right)=1$ if $a$ is authorized by DM $i$ from menu $B$, and $y\left(a|i,B\right)=0$ otherwise. The log-likelihood function becomes

$$LL=\prod_i \prod_{a\in B} y\left(a|i,B\right)\ln p\left(a|i,B\right)+\prod_i \prod_{a\in B}\left(1-y\left(a|i,B\right)\right)\ln\left(1-p\left(a|i,B\right)\right).$$

A nice feature of a logit categorization function is that it allows for the exploration of how the thresholds are determined by estimating $\alpha$ (i.e., the coefficients of the characteristics of the menus).

**Example 5.** (Choice Overload) A consumer decides whether to include or exclude a jar of jam on the shelf into her consideration set (i.e., the set of products she actually considers). The consumer considers a jar of jam (authorization) if she thinks the jam is preferred to the threshold of the menu, and ignores a jar of jam (rejection) otherwise. The experiment in Iyengar and Lepper (2000) finds that, the consumers were less likely to buy a jar of jam if the sizes of the menus were larger.

Choice overload can happen if the thresholds are dependent on the size of menu. For instance, consumers may use a higher threshold with a larger menu to pre-empt regret or dissatisfaction. In this case, the consumers will be less likely to consider an item if there are more options.[20] We can define menu $B$ as the set of jam jars on the shelf, and define $w(B,i) = \alpha_i s_B + c$, where $s_B$ is the cardinality of menu $B$, and $c$ is a constant. Then, we can estimate $\alpha_i$ to see if the decreased demand in jams was due to higher thresholds, which was caused by more options in the menu.[21]

**Example 6.** (Availability Bias) A doctor decides if the patients have bacteremia. The doctor diagnoses a patient as having bacteremia (authorization) if the symptoms of the patient are more severe than the threshold; conversely, the patient is diagnosed as not having it (rejection) otherwise. The doctor's decisions are subject to availability bias. In other words, a doctor is more likely to diagnose the patient as having a disease if he can recall more relevant examples, and vice versa. The experiment in Poses and Anthony (1991) finds that doctors were more likely to diagnose a patient as having bacteremia if they could recall more bacteremic patients in the past month.

This decision process is similar to the 'threshold approach' to clinical decision making by Pauker and Kassirer (1975&1980). Availability bias can be detrimental in diagnoses, as it results in rare diseases being underdiagnosed and common diagnoses being

[20]For a review on choice overload, see Chernev, Böckenholt, and Goodman (2015).

[21]A higher threshold leads to a lower probability for a jar of jam to be considered, thus decreasing the probability of buying a jar of jam.

overdiagnosed. Our model can be used to illuminate the nature of the availability bias. We can define menu $B$ as the set of patients the doctor has diagnosed in the previous month, and define $w(B, i) = \alpha_i n_B + c$, where $n_B$ is the number of patients who were diagnosed as positive. The estimation of the coefficient $\alpha_i$ can reveal whether the doctor's threshold changes with the ease of recalling diagnosed patients who had bacteremia.

Notice that the full-fledged logit categorization model is not only an empirical view, but also provides an interpretation of the stochasticity in categorizations. The DM compares the utility of an item to the threshold of the menu, and since both utilities and thresholds are random, the DM categorizes stochastically. With the randomness of a specific form (i.i.d with type-1 extreme distribution), the logit-like representation of the authorization probabilities emerges. The simple characterization axiom in Section 2.3 is an intuitive property of the logit categorization model, which provides insight into the categorization patterns. This full-fledged model provides a proper explanation of the source of stochasticity in categorization.

## 2.5 Concluding Remarks

### 2.5.1 Categorization Precision

The categorization function in our full-fledged model clearly shows that the authorization probability is determined by the representative threshold. We model menu dependency by defining the representative threshold as a function of the menu. A natural companion of our model is that, instead of the representative threshold, the precision of the comparison is affected by the menu. For instance, the authorization probability can be defined as $p(a, B) = \frac{V(a)^{\sigma(B)}}{V(a)^{\sigma(B)} + W^{\sigma(B)}}$, in which $\sigma(B) \in \mathbb{R}$ captures the precision of the categorization. This expression is similar to the stochastic choice model in Rehbeck (2021), which uses an extended Luce choice rule where the alternative values are exponentiated by a parameter to accommodate choice overload. This is not a logit categorization model, as it does not satisfy $\gamma$-IIA. In this model, the representative threshold $W$ is menu independent, which can be taken as an 'ideal' threshold that is applied over every menu.[22] Nevertheless, the authorization probability of alternative $a$ from menu $B$ still changes with the menu, as the categorization precision is determined by the menu. This version of categorization function can be used to clarify the understanding of decision precision.

### 2.5.2 Stochastic Representative Threshold

In the full-fledged logit categorization model, we analyse the categorization behaviour in which the representative threshold $w(B,i)$ is fixed across all decisions. However, sometimes $w(B,i)$ itself can be stochastic, reflecting random 'tastes' for the thresholds of menus. For example, different doctors can be affected by availability bias to different extents. If the representative thresholds follow random distribution $f(w(B,i))$, the authorization probability of item $a$ from menu $B$ becomes

$$p(a|i,B) = \int \frac{e^{v(a,i)}}{e^{v(a,i)} + e^{w(B,i)}} f(w(B,i))\, dw(B,i).$$

The full-fledged model is a special case where the distribution of $w(B,i)$ is degenerate at fixed parameter $w^*(B,i)$: $f(w(B,i)) = 1$ for $w(B,i) = w^*(B,i)$; and 0 for $w(B,i) \neq w^*(B,i)$, where $w^*(B,i)$ is the (fixed) intrinsic representative threshold for DM $i$. With parametric assumptions (e.g., assume $w(B,i) \sim N(\mu,\sigma^2)$), we are able to estimate the corresponding parameters through simulation and identify the randomness in $w(B,i)$. This is structurally similar to the mixed logit choice probability, as we essentially use the methodology from the mixed logit model.

[22] Effects from menus on the representative threshold and on the categorization precision can both be present. In that case, we can define the authorization probability as $p(a,B) = \frac{V(a)^{\sigma(B)}}{V(a)^{\sigma(B)}+W(B)^{\sigma(B)}}$.

### 2.5.3 A Richer Dataset

A different type of dataset would be given by a stochastic authorization correspondence $p(C|i,B)$, which is the probability that all the items in set $C \subseteq B$ are authorized by DM $i$, while all the other items in set $B \backslash C$ are rejected. This type of data incorporate richer information on the correlations between authorizations of different items. While we allow for the thresholds to be menu dependent, the categorizations in our model are done by judging every item independently on its own merits. This independence property leads to

$$p(C|i,B) = \prod_{b \in C} p(b|i,B) \prod_{c \in B \backslash C} (1 - p(c|i,B)),$$

for all $C \subseteq B$, for all $B \in \mathcal{D}$. If the equation above holds, then our models can be used. If the above equation does not hold, we should expect correlations between authorizations of items, which makes our models incompatible. A potential solution is imposing parametric structure on the utilities of the subsets $C \subseteq B$ (the utility of the bundle of options).[23]

### 2.5.4 Horizontal Categorization

The categorizations in this chapter are fundamentally vertical (i.e., the quality is the key to partition the space of objects). One can further applies the methodology from Anderson and de Palma (1992) to modify logit categorization function so that it can accommodate horizontal categorizations: define the utility as $u = v + \mu\varepsilon$, in which $\mu$ is positive, and $\mu\varepsilon$ expresses the horizontal differentiation between items. This version of logit categorization function can be interesting to researchers who aim to study horizontal categorization behaviour, e.g. a consumer searches for a specific product on a shopping website and puts items listed on the pages into her shopping cart.[24]

[23] Hendel (1999) applies similar methodology in his empirical analysis on the demand of computers.

[24] The listings are usually inevitably differentiated in online shopping. For example, see the dataset from eBay in Dinerstein, Einav, Levin, and Sundaresan (2018), for which they imposed similar structure

## Acknowledgements

A paper which is based on this chapter is published at *Oxford Economic Papers* (Wang 2021). I thank the Associate Editor Alan Beggs and two anonymous referees for their useful comments and suggestions, which greatly improved the chapter. I would also like to thank Miguel Costa-Gomes for helpful comments.

on unobserved utility to accommodate differentiation.

# Chapter 3

# Consideration Set Formation in Online Commerce

**Abstract**

For pragmatical reasons, DMs often consider only a subset of the available options, namely the consideration set. The formation of consideration set can be of both managerial interest and economic interest. We use detailed browsing data from eBay to study the consideration set formation. We find strong evidence that consumers systematically decided which listings to consider, and which to ignore. We also exploit a large-scale platform redesign of eBay to evaluate the effectiveness of the redesign on both preference and consideration set formation.

Keywords: online commerce, consideration set.

JEL codes: D12.

## 3.1 Introduction

Many marketing and economic research describe the consumer choice as a two-stage (i.e., the consideration stage and the choice stage) decision process. For example, Kotler (1988), Manzini and Mariotti (2007). In the consideration stage, the consumer decides which options to consider, and which not to; and in the choice stage, the consumer picks the alternative that maximizes the utility. Current research studying online commerce assume that consumers consider an alternative if it has been displayed on the screen (web pages), e.g. De Los Santos, Hortaçsu, Wildenbeest (2012), Dinerstein, Einav, Levin, Sundaresan (2018) (hereafter, DELS). Yet, this assumption is not innocuous as it looks, as experimental evidence show that subjects failed to consider all the alternatives that were displayed on the screen.[1] It is possible that a good quality online listing shows poor sales performance, simply because it is not considered by the consumers. Therefore, the formation of the latent 'consideration set', namely the set of listings a consumer actually chooses between, can be of interest to online retailers and sellers. Our analysis on the formation of consideration set sheds light on the online users' behavioural patterns, and helps online retailers and sellers deploy their marketing strategy.

In this chapter, we use the detailed online browsing data from eBay in DELS (2018) to quantitatively analyse the formation of consideration set. In particular, we estimate the influence on consideration set from factors that affect the utilities of listings, i.e. total price, top-rated seller (TRS) indicator (if the listing is listed by a TRS);[2] and from factors irrelevant to the utilities of listings, i.e. free shipping indicator (if the listing is shipped for free), click indicator (if the listing receives a click), page type (if the listing is displayed on the product page).[3] We focus on a specific and highly

[1]See, for example, Reutskaja, Nagel, Camerer, Rangel (2011), Caplin, Dean, and Martin (2011).

[2]A TRS on eBay must have done at least 1,000 transactions and $3,000 in sales during the previous year, and score a positive feedback above 98 percent.

[3]comScore 2012 report finds that while comparison shopping, shipping charges are almost as important to consumers as product pricing (23% and 26%, respectively). Full report can be downloaded here: https://www.comscore.com/Insights/Presentations-and-Whitepapers/2012/Online-Shopping-Customer-Experience-Study. We will illustrate in detail the properties of product page in

homogeneous product, the Halo Reach video game. We find that the consumers form their consideration sets depending on total price, TRS indicator, free shipping indicator, and click indicator. This result indicates that consumers who searched for the Halo Reach video game systematically decided which listings to consider, and which to ignore.

In May, 2011, eBay launched a large-scale platform redesign. Prior to the redesign, consumers were shown individual listings ranked by a 'best match' algorithm. The redesign asked consumers to first identify an exact product, then compare listings of that product in a head-to-head manner. We exploit this platform redesign on eBay to check the difference in consumers' behavioural patterns.

Standard discrete choice models assume perfect attention, namely all the alternatives in the menu are considered in the decision process. In the literature of economics and psychology, however, it has long been known that imperfect attention plays an important role in decision making, e.g. Simon (1959), Tversky (1972), Kahneman (1973). Our chapter is related to the important literature on generalizing discrete choice models to relax the assumption that all the goods are considered. A pioneering work of Manski (1977) specifies a probability that each subset of alternatives is considered, that is, the consideration set probability. This approach is enlightening, and has become increasingly prevalent in both applied and theoretical research in economics.[4] Among them, our chapter benefits the most from the recent work by Abaluck and Adams-Prassl (2021), in which the authors prove that identifications of both preferences and considerations are feasible if (i) no auxiliary information on consideration sets is observed; and (ii) some factors affect both preferences (utilities) and consideration sets. Most previous research, if not all, rely on ancillary data on what options are consid-

Section 3.3.

[4] The theoretical work of Manzini and Mariotti (2014), Brady and Rehbeck (2016) characterize the stochastic choice models, in which the DM chooses the alternative that maximizes the utility among a randomly generated consideration set.

ered, or impose an exclusion restrictions identification, i.e. characteristics either affect utilities or considerations, but not both.[5] For example, in the applied work of Goeree (2008), an auxiliary advertising level data is used to infer the consideration sets, where the framework assumes that advertising only impacts consideration sets. By applying the 'alternative specific consideration' (ASC) approach in Abaluck and Adams-Prassl (2021) to our data, we manage to separately analyse the preferences and considerations without additional data. Our analysis on the change of preferences and considerations due to an exogenous redesign is similar to the work in Gaynor, Propper, and Seiler (2016), which assesses the effect of expanding patients' options after a reform in the English National Health Service. The reform allowed patients to choose between more hospitals, and Gaynor, Propper, and Seiler (2016) find substantial increase in patient welfare and in demand elasticity post-reform.

A great feature of the online commerce dataset in DELS (2018) is that it tracks exactly what each customer sees, namely the set of listings that are displayed on the page to individuals. In other words, the menus (the set of feasible alternatives) faced by the DMs are observed. To be clear, while this chapter uses the same data as in DELS (2018), our research is quite different to DELS (2018) in several ways. First, the assumptions on the consideration set are different. In DELS (2018), the listings on the screen (web pages) are assumed to be considered by the consumers. We relax this assumption, and instead assume that the consumers decide to include or exclude a listing into their consideration sets depending on the listing's characteristics. In DELS (2018), the consideration set is assumed to be the set of feasible listings displayed on the web pages (the menu itself); while in this chapter, the consideration set is assumed to be the latent set of alternatives that are actually considered by the consumers (a subset of the menu). Second, the focuses of the two research are different. DELS (2018) focus extensively on the market competition, and they analyse the effect from

[5]We use 'considerations' and 'consideration sets' interchangeably in this chapter.

the platform redesign on consumer welfare, markup, price elasticity, etc. They also provide an accurate out-of-sample prediction for pricing and demand. We focus more on the consumer side: the aim of this chapter is to analyse the formation of consideration set, and compare the behaviour pattern of consumers in the dimensions of preference and consideration set before and after the redesign of the display page of listings.

## 3.2 Study on Consideration Set

The consideration set formation is the key research object in this chapter. In this section, we provide a detailed literature review on the studies of consideration set.

The topic of consideration set has received much interest in economics as it sits at the intersection of marketing and economic literature. Stigler (1961) looks at the costs and benefits of gathering information from including brands into consideration sets. The formation of consideration set is described as the effort to maximize $EU\left(C\right)-\sum_{j\in C}c_j$, where $EU\left(C\right)$ is the expected maximum utility of a choice from set $C$, and $c_j$ is the marginal cost of considering $c_j$. This costs and benefits analysis assumes the formation of consideration set to be a process of searching, and it has been used widely in both marketing and economic research, e.g. Hauser and Wernerfelt (1990), Roberts and Lattin (1991), and more recently, De los Santos, Hortaçsu, and Wildenbeest (2012), Honka and Chintagunta (2017). These research assume that the price is the only source of uncertainty, and that the consumers know the distribution of prices and have rational expectations for the prices.

To relax these assumptions, there has been a growing trend of using discrete choice models to study consumer behaviour in which limited attention arises. Manski (1977) first suggests that the choice problems under limited attention to be expressed probabilistically as $P_n\left(i\right)=\sum_{C\in G(i)}P_n\left(i|C\right)P_n\left(C|G\right)$, where $P_n\left(i\right)$ is the probability of choosing alternative $i$ by individual $n$, $P_n(i|C)$ is the probability of individual $n$ choosing alternative $i$ given that the consideration set is $C$, $P_n(C|G)$ is the probability of $C$ being the consideration set of individual $n$ given menu $G$, and $G(i)$ is the set of all subsets of $G$ that contain alternative $i$. The number of possible consideration sets is very large ($2^J$,where $J$ is the cardinality of $G$), which implies a high degree of complexity. In the work of Ben-Akiva and Boccara (1995), the authors postulate explicit representation of consideration set constraints to simplify the estimation. Still, the methodology can hardly be extended to dataset with larger menus because of the substantial difficulty in estimation. To further simplify the consideration set formation, an 'independent consideration' approach, which imposes specific structure on the consideration sets has been widely used in the applied literature. Independent consideration approach expresses probability $P(C|G)$, the probability of $C$ being the consideration set given menu $G$, as $P(C|G) = \prod_{i \in C} \phi_i \prod_{k \in G \backslash C} (1 - \phi_k)$, where $\phi_i$ is the probability of considering alternative $i$ (including $i$ in the consideration set). Note that this approach assumes that the probability of considering $i$ is only determined by $i$ itself, and is independent of the other components in the menu. Independent consideration approach is formally characterized in Manzini and Mariotti (2014). Although it imposes this specific independence structure on consideration sets, which greatly simplifies the estimation, it still can be computational demanding. For that reason, empirical models usually rely on auxiliary data of what goods are considered. For example, the empirical study of the advertising effect in personal computer market in Goeree (2008) requires auxiliary dataset, i.e. the data on media exposure. What is more, the model in Goeree (2008) assumes that the advertising only affects choices via informing consumers about which goods exist, namely the advertising has no effect on the utilities of the alternatives.

The idea of eliciting preference and attention component from heterogeneous menus is not new, especially in previous theoretical work on limited attention, e.g. Masatlioglu, Nakajima, and Ozbay (2012), Manzini and Mariotti (2014), Brady and Rehbeck (2016). In these research, the characterizations are usually succinct and falsifiable, and the identification results are substantive and elegant (and in most times unique). Yet these work typically require a rich dataset. For example, Masatlioglu, Nakajima, and Ozbay (2012) require data for all menus from the universal set; Manzini and Mariotti (2014), Brady and Rehbeck (2016) impose a strong 'richness' assumption on the dataset. It is hard to have a dataset with menus that follow the imposed richness structure *exactly* without experimental environment, making those theoretical models less applicable to field data. However, it is not uncommon to have a dataset in which heterogeneous menus are observed. For example, Ebay customers browse different set of listings with the same key words; car owners search various companies seeking auto insurance; consumers buy from different online book stores.[6] As discussed in Section 3.1, the ASC model in Abaluck and Adams-Prassl (2021) allows inference of the unobserved preferences and consideration sets without ancillary data. The ASC model shows that that the identifications of preferences and considerations are possible without stringent richness assumptions on menus, and that ancillary data is not necessary. Essentially, the ASC model takes advantage of the richness in the dataset that originates from the observed covariates. Other recent works develop identification results by bridging econometric methods and behavioral theory together. Cattaneo, Ma, Masatlioglu, Suleymanov (2020) impose a monotonic attention rule on the attention pattern of agents, and develop econometric inference methods for preferences. A recent research by Dardanoni, Manzini, Mariotti, and Tyson (2020) shows that heterogeneous preferences and considerations can be uniquely identified by joint choice probabilities from three menus. Their work aims at identifying preferences and considerations with the least variation of menus, and is mostly useful if the covariates of individuals or alternatives are inaccessible.

[6]See, Dinerstein, Einav, Levin, and Sundaresan (2018), Honka and Chintagunta (2017), De los Santos, Hortaçsu, and Wildenbeest (2012) respectively.

Arguably, the consideration set formation is of managerial interest in its own right, as the sales of products can be largely affected by it. For example, the study in Hauser (2011) shows that the products of a US automotive manufacturer (disguised here as USAM) were not considered at all by half of US consumers. USAM deployed multi-million dollar programs to persuade consumers to consider their products. They offered free peer test driving (without sales pressure) and published the results of test driving on its website.[7] These strategic moves aiming at lowering consideration costs and drawing consumers' attention were proved to be successful and profit of USAM increased sharply after the programs.

The consideration set has shown to affect not only sales of the products, but also consumer welfare. DELS (2018) find empirical evidence that including more targeted products (the product that the consumer is searching) into the consideration sets significantly increased market competition and consumer welfare. The demand analysis in DELS (2018) shows that after the redesign, transaction prices fell by around 5 to 15 percent across a broad set of product categories. In an interesting theoretical research by Eliaz and Spiegler (2011), the authors show that the industry equilibrium profits increase if there are more rational agents, i.e. agents who consider everything in the market, as long as the group of rational consumers is not too big.

Finally, the consideration set is important for unbiased estimation when limited attention exists. Crawford, Griffith, and Iaria (2021) find that significant bias exists in British chocolate bar industry with full information model and the bias is larger if the share of individuals with limited cognitive capacity increases. Abaluck and Adams-Prassl (2021) also show that the failure to incorporate limited attention results in biases in estimated advertising sensitivities. Goeree (2008) shows that there exists 19 percent mark-up in demand elasticities under limited attention for top sellers in PCs industry,

[7] The participants got to drive USAM vehicles along with other vehicles (up to 100) from Acura, BMW, Buick, Cadillac, Chevrolet, Chrysler, Dodge, Ford, Honda, Lexus, Lincoln, Mercedes, Pontiac, Saab, Saturn, Toyota, Volkswagen, and Volvo.

whereas the mark-up is only 5 percent with perfect attention. Theoretical research, e.g. Masatlioglu, Nakajima, and Ozbay (2012), Manzini and Mariotti (2014), prove that the identification of preference in standard economic models can be misleading if the consideration set plays a role in choice making. Analogous to these research findings, we find that the estimates of preferences are biased if we assume all the listings on the web pages are considered.

## 3.3 Background: Platform Redesign on eBay

On May 19, 2011, eBay launched a large-scale platform redesign. Before the redesign, clients were shown individual listings drawn from a set of matches, with rankings generated by a 'best match' algorithm named by eBay. During the time the data was collected, best match algorithm was not tailored to individual consumer. In addition to that, it did not take price into account directly.[8]

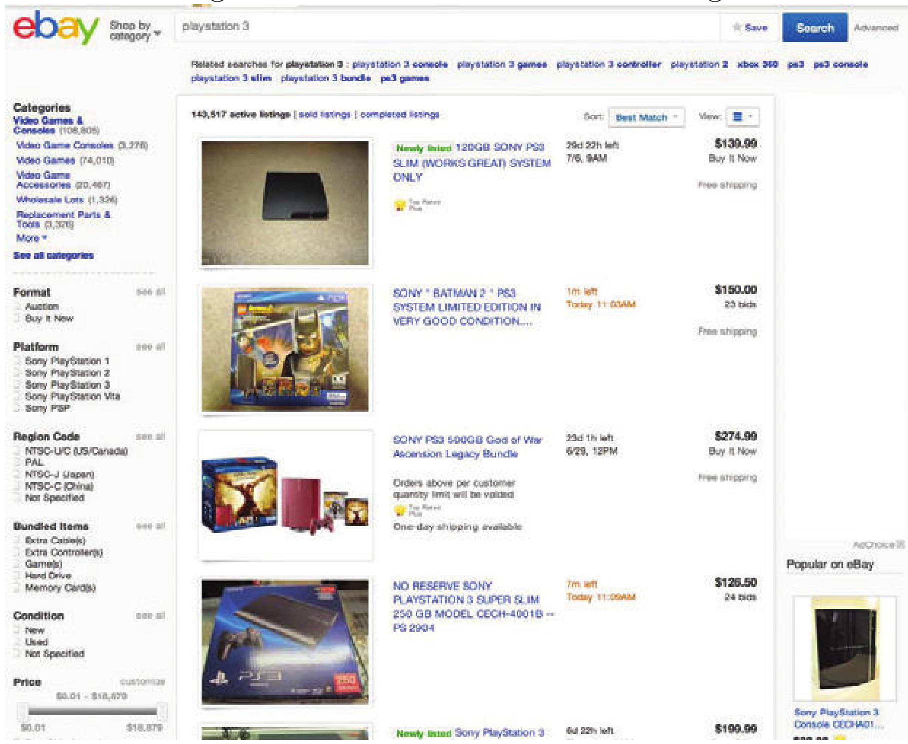
Although best match page still exists, but the design has changed a lot.[9] Figure 3.1 shows eBay's traditional listings page, generated by best match algorithm. For the sake of illustration, here we 'borrow' the graph from DELS (2018).

The redesign of online shopping platform on eBay generated an alternative two-stage search process. The consumer first sees relevant products on the basis of query terms (for example, a consumer searches 'game console' can find related game consoles 'Play Station 3', 'XBOX 360', etc. on the page). The user then clicks on the products to see the newly designed product page. The concept of a 'product page' existed on eBay earlier, but it was quite different to the product page in this chapter in terms of design and display, and it was difficult to be found by consumers. Therefore, only a tiny portion of consumers viewed it: during the before period in the dataset, in only 18

[8] At first glance, it may look strange not use price as an explicit ranking factor in best match algorithm. However, if the listings are sorted by price, most of the listings that are cheap but less related to the query terms show up. For example, if a consumer searches with query term 'game console' and uses 'price low to high' algorithm, the results are mainly cheap accessories.

[9] The best match page on eBay highlights eBay promoted listings substantively nowadays.

Figure 3.1: Traditional Best Match Page

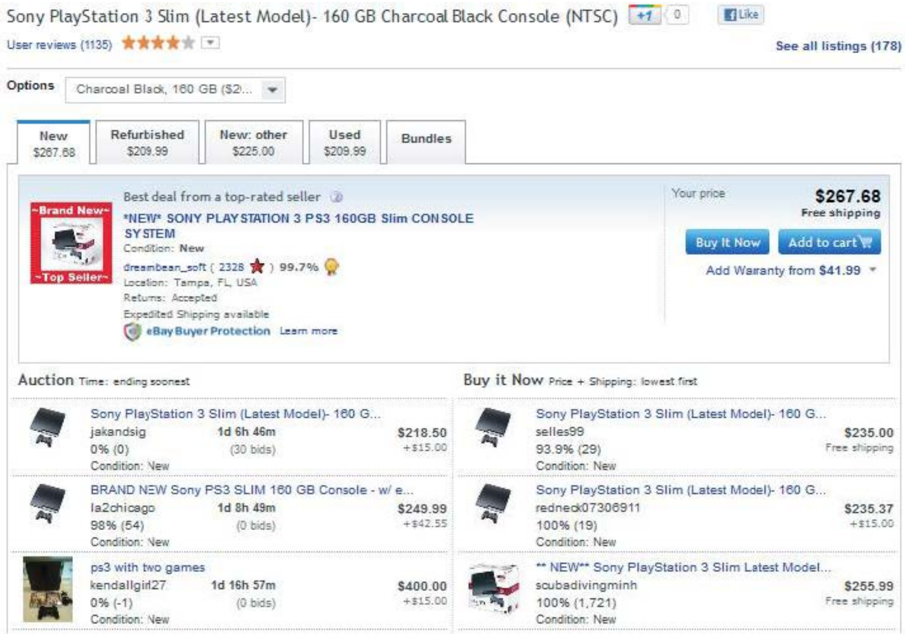



(out of 9409) search sessions the users viewed the product page.

The platform redesign on eBay has evolved substantively since introduction of the new product page in 2011. Therefore it is hard for us to replicate the product page in the up-to-date version of eBay. Again, we borrow the figure from DELS (2018) to illustrate the search results in the newly designed product page, as shown in Figure 3.2.

In the newly designed product page, a prominent 'buy box' is displayed to the buyer, which displays the listing with the lowest price (plus shipping cost) listed by a TRS. Then, there are two columns of listings displayed below the buy box, the left column comprises auctions listings, and the right column comprises listings with fixed prices. In the auction column, the listings are ranked according to the auction ending time (ending soonest on top); while in the fixed price column, the listings are ranked by the price (the cheapest on top). Note that the price of the first listing in the right column (fixed price column) can be lower than the listing in the buy box if the lowest-price seller is a non-TRS.

Figure 3.2: Newly Designed Product Page




The newly designed product page gradually became the default presentation of search results for five large categories (cell phones, digital cameras, textbooks, video games, and video game systems) from June 27, 2011 to July 2, 2011. The Halo Reach video game, which is the product we focus on, happens to be in the video games category. The traditional best match page was always accessible to consumers. After the platform redesign, the buyers were able to see two types of results, and were defaulted into the product page.[10]

## 3.4 Data and Descriptive Statistics

Our data is collected from the work of DELS (2018). We will focus on a single and well-defined product: the Halo Reach video game. The video game is one in a series of Halo video games, and was first released in Sep, 2010. The search data consists of all visits to the search results page derived from query terms that include the words 'xbox' (or 'x-box'), 'halo,' and 'reach'. We are studying the US market, and the total prices,

[10]eBay deployed an experiment in 2012, in which consumers are presented best match or product page randomly, to evaluate the redesign. After evaluating and reviewing the platform redesign for over one year, eBay made the original, best match results the default view for its users in October, 2012.

shipping cost, etc. are in US dollars. The original price set by Microsoft is \$59.99, and the price shortly dropped to \$39.99. As in DELS (2018), we use data of two periods: the period from 6/April to 18/May, 2011, during which the newly designed product page was not introduced yet; and the period of 1/Aug to 20/Sep, 2011. Note that the new product page was introduced on 19/May, 2011. The data during period 19/May, 2011 to 31/July, 2011 is dropped for two reasons: (i) the selection of periods is consistent with that in DELS (2018), so that interested researchers can make comparisons; (ii) the period from late May to late July can be taken as adjustment period, during which the sellers and consumers adjust their behaviour in response to the introduction of the newly designed product page.

We drop the search sessions that resulted in no clicks, or generated no listings (which can happen if the user exited eBay immediately after searching). Table 3.1 describes the summary statistics for the listings data of Halo Reach on eBay, before and after the platform redesign. An advantage of selecting Halo Reach as the target product is that during the observation period, this product has a large number of transactions. Unlike many other electronics on the platform which exhibit a high start and a quick fall price pattern, it had a stable price history due to the relatively stable supply and demand. The fact that this product is single and well-defined makes it easier to locate the observations, as 51 percent of Halo Reach listings have the same title. Of course, just like the work in DELS (2018), the high homogeneity of this product imposes limitation on our estimation, for example, it is not possible to analyse the effect of brands on consideration, a topic that has been studied extensively in the literature of marketing. On the other hand, it greatly simplifies the estimation of preference and consideration set as well, which helps us analyse the exclusive factors (e.g., shipping cost, page type) in online commerce.

Table 3.2 gives detailed summary statistics for the search data. This browsing dataset is rich as it has all individual-level characteristics of listings, including, total

Table 3.1: Summary Statistics for Halo Reach Listings Data

| | Before | After |
|---|---|---|
| No. of listings | 270 | 218 |
| No. of sellers | 191 | 152 |
| Percentage of sellers with multiple listings | 20 | 22 |
| Mean listing price | \$39.73 | \$37.88 |
| Median listing price | \$37.00 | \$35.00 |
| Standard deviation of listing price | \$9.2 | \$8.73 |
| Percentage of listings from TRS | 16 | 27 |

Notes: Before: April 6, 2011 to May 18, 2011. After: August 1, 2011 to September 20, 2011. TRS refers to top-rated sellers, an eBay designation that depends on a seller's volume and feedback. The difference of mean listing price is significant at 5% significance level. The difference of standard deviation of listing price is not significant. Auction listings are dropped.

price, shipping cost, search rank, page type (displayed in a product page or not), seller type (listed by a TRS or not), receives a click or not, etc. The data consists of all the visits to the product page of Halo Reach during after period, and the visits to the standard results that are generated from query terms including the words 'xbox' (or 'x-box'), 'halo,' and 'reach.' This results in 14,753 visits to the search results page (9,409 in the pre-period) and 6,733 visits to the product page (18 in the pre-period). In DELS (2018), the 'targeted listings' are defined as new Halo Reach items, listed either with a posted fixed price, or listed as an auction with a 'buy-it-now' price. The non-targeted listings are those that are derived from the terms above, but are used products, or not the Halo Reach video game itself.

Table 3.2: Summary Statistics for Search Data

| | Before | After |
|---|---|---|
| No. of 'search page' searches | 9409 | 5344 |
| No. of 'product page' searches | 18 | 6715 |
| No. of searches | 9427 | 12059 |
| Mean transaction price | \$34.56 | \$33.30 |
| Median transaction price | \$34.99 | \$34.00 |
| Standard deviation of transaction price | \$2.59 | \$3.89 |
| No. of brand new Halo Reach fixed price transactions | 97 | 148 |
| No. of other transactions | 185 | 317 |
| Percentage of listings which received a click | 6.65 | 5.23 |

Notes: The mean and median transaction prices are for the transactions of brand new Halo Reach video game at fixed price. The difference in mean transaction price, the difference in standard deviation of transaction price, and the difference in percentage of listings which received a click are all significant at 1% significance level.

## 3.5 Theoretical Framework

The model is based on the ASC model. Abaluck and Adams-Prassl (2021) use experimental data to prove that the ASC model with maximum likelihood estimation can accurately recover preferences and considerations.[11] The results also show that the ASC model can discriminate between preferences and considerations using real choice data, with some factors affecting preferences and considerations at the same time, which matches our research aim perfectly.

In addition to the maximum likelihood approach (the one used in our study), Abaluck and Adams-Prassl (2021) show that point identification in non-parametric structure is feasible with several assumptions. In Appendix, we discuss this non-parametric version of the ASC model and show that our model satisfies the assumptions. Theoretically, we should be able to recover the consideration set without parametric assumptions on the consideration set formation structure. However, the requirement

[11] Adams-Prassl (2021) run an experiment and use real choice data from the experiment to verify the model. They find that the ASC model accurately recovers the intrinsic preferences and consideration sets.

on the dataset is stringent.[12] Analogous to Abaluck and Adams-Prassl (2021), we impose parametric structure on the ASC, and use maximum likelihood approach to do the estimation.

In this section, we will first illustrate the general framework of the ASC model. Then we will explain in detail the parametric set-up in our work.

### 3.5.1 General Model

Consider a consumer $i$ who buys a listing among menu $J_i$ of $n+1$ listings. Each listing is indexed by $j$, so that $J_i = \{0, 1, ..., j, ..., n\}$, with $n \geq 1$.[13] Each listing is associated with a price, $p_j$; and is associated with additional characteristics, $\mathbf{x}_j$. The price vector $\mathbf{p} = [p_0, p_1, ..., p_n]$ is supported on $\mathbb{R}^{n+1}$. The additional characteristics vector is denoted by $\mathbf{x} = [\mathbf{x}_0, \mathbf{x}_1, ..., \mathbf{x}_n]$. The latent set of listings that is actually considered by a consumer is named the consideration set. Let $\mathcal{P}(J_i)$ be the power set of $J_i$, with any element of $\mathcal{P}(J_i)$ indexed by $C_i$. The set of consideration sets containing listing $j$ is

$$\mathbb{P}_i(j) = \{C_i : \{j, 0\} \subseteq C_i \in \mathcal{P}(J_i)\}.$$

The observed choice probabilities of consumer $i$ buying listing $j$ take the following form:

$$s_{ij}(\mathbf{p}, \mathbf{x}) = \sum_{C_i \in \mathbb{P}_i(j)} \pi_{C_i}(\mathbf{p}, \mathbf{x}) \, s^*_{ij}(\mathbf{p}, \mathbf{x} | C_i),$$

where $s_{ij}(\mathbf{p}, \mathbf{x})$ is the observed probability of listing $j$ bought by consumer $i$ given prices $\mathbf{p}$, additional characteristics $\mathbf{x}$; $\pi_{C_i}(\mathbf{p}, \mathbf{x})$ is the probability that the set of listings $C_i$ is actually considered given observable characteristics; and $s^*_{ij}(\mathbf{p}, \mathbf{x}|C_i)$ gives the probability that listing $j$ is chosen (bought) conditional on the consideration set being $C_i$. The interpretation is straightforward: the probability of choosing listing $j$ by

[12] Detailed discussion is provided in Appendix.

[13] Here the outside alternative (buying nothing) is denoted as alternative 0.

consumer $i$ from menu $J_i$, is the probability of choosing (buying) listing $j$ conditional on consideration set $C_i$, summed over all the set $C_i$ containing listing $j$.

In the ASC model, $\pi_{C_i}$ (i.e., the probability of set $C_i$ being the consideration set) is defined as

$$\pi_{C_i}(\mathbf{p}, \mathbf{x}) = \prod_{j \in C_i} \phi_{ij}(p_j, \mathbf{x}_j) \prod_{k \in J_i \setminus C_i} (1 - \phi_{ik}(p_k, \mathbf{x}_k)),$$

where the probability of listing $j$ being considered, $\phi_{ij} = \phi_{ij}(p_j, \mathbf{x}_j)$, is a function of its own characteristics only and $\phi_{i0}(p_0, \mathbf{x}_0) = 1$ for all $p_0, \mathbf{x}_0$.[14] The choice probabilities take the form:

$$s_{ij}(\mathbf{p}, \mathbf{x}) = \sum_{C_i \in \mathbb{P}_i(j)} \prod_{j \in C_i} \phi_{ij}(p_j, \mathbf{x}_j) \prod_{k \in J_i \setminus C_i} (1 - \phi_{ik}(p_k, \mathbf{x}_k)) \, s^*_{ij}(\mathbf{p}, \mathbf{x} | C_i).$$

After introducing this general framework of ASC model, we now give the parametric structure used in our work. The ASC approach from Abaluck and Adams-Prassl (2021) allow variables to impact both utilities and considerations. Still, an important issue is to decide which characteristics to be the parameters for utilities (preference parameters), and which characteristics to be the parameters for considerations (consideration parameters). Theoretically, we can make all the characteristics to be both preference parameters and consideration parameters. However, this agnosticism over what variables affect utilities and what affect considerations leads to results that can be hardly interpreted.[15] In this chapter, we try to carefully categorize the variables as preference parameters, or consideration parameters, or both. Total price and TRS indicator are treated as both preference and consideration parameters. As in DELS (2018), it is intuitive to assume that total price and TRS indicator are preference parame-

[14] The key assumption in ASC model is that the probability of considering alternative $j$ is determined by the own characteristics of $j$, and is irrelevant of the other alternatives in the menu. This a substantive assumption, although it is popular in applied work.

[15] Apart from the ASC model, the model in Crawford, Griffith, and Iaria (2021) also allows both preferences and considerations to be dependent on the same observables. However their model can hardly be applied to our dataset, as it requires panel data.

ters. What is more, it is possible that total price and TRS indicator affect consumers' consideration set formation. For instance, some consumers can possibly ignore listings that are expensive or listed by non-TRSs. Free shipping indicator is considered to be a consideration parameter only, as after controlling total price, the shipping cost should not affect utility. Then, product page indicator (if the listing is displayed on the newly designed product page) and click indicator (if the listing receives a click) are taken to be consideration parameters only, as the utilities should not be determined by the presentation of the listing, and should not change if the listings receive a click.[16]

### 3.5.2 The Consideration Functions

Manzini and Mariotti (2014) find that if the consideration set probabilities are allowed to vary arbitrarily, then the identifications of preferences and considerations become hopeless. Therefore, for identification purposes, we must place some additional restrictions on the consideration function $\pi_{C_i}$.

In each search session (both before and after the platform redesign), we treat the set of listings that appear on a respondent's screen as the menu. Note that, $\phi_{ij}$ is the probability that listing $j$ is in consumer $i$'s consideration set, which can be taken as the 'effectiveness' that listing $j$ is 'advertised' on the screen seen by consumer $i$.

Our key question is whether the consumers systematically choose what listings to consider, and what listings to ignore. More specifically, during the period before the redesign, we want to see if the total price, seller type (listed by a TRS or not), click indicator (receives a click or not), and free shipping indicator (shipped at no cost or not) influence the considerations. During the after period, eBay introduced a newly designed product page, while the traditional best match page still existed. Among the all the search sessions during the after period (12509), 55.68% (6715) used the newly designed product page, while 44.32% (5344) used the best match page. In addition to

[16] One may argue that, click indicator can be related to the utility as consumers tend to click on listings with high utility. We have another version of our model by including relevant click as a preference parameter, but the results are much harder to interpret.

the parameters we study during the pre-period, we add another product page indicator (listed on the newly designed product page or not) to study the effect from product page on the considerations.

During the before period (April 6, 2011 to May 18, 2011), $\phi_{ij}$ is given by[17]

$$\begin{aligned}\phi_{ij}\left(p_j,\mathbf{x}_{ij}\right) &= \Pr\left(\gamma_0+\gamma_1 p_j+\gamma_2 TRS_j+\gamma_3 p_j TRS_j+\gamma_4\nu_{ij}+\gamma_5\sigma_j-\eta_{ij}>0\right),\\ &= \frac{e^{\tau_{ij}}}{1+e^{\tau_{ij}}}=\frac{e^{\gamma_0+\gamma_1 p_j+\gamma_2 TRS_j+\gamma_3 p_j TRS_j+\gamma_4\nu_{ij}+\gamma_5\sigma_j}}{1+e^{\gamma_0+\gamma_1 p_j+\gamma_2 TRS_j+\gamma_3 p_j TRS_j+\gamma_4\nu_{ij}+\gamma_5\sigma_j}},\end{aligned}$$

where $\eta_{ij}$ is distributed logistic, $\nu_{ij}$ is the click indicator ($\nu_{ij}=1$ if listing $j$ receives a click by consumer $i$, and 0 otherwise), and $\sigma_j$ is the free shipping indicator ($\sigma_j=1$ if listing $j$ is shipped at no cost, and 0 otherwise).

During the after period (August 1, 2011 to September 20, 2011), $\phi_{ij}$ is given by

$$\begin{aligned}\phi_{ij}\left(p_j,\mathbf{x}_{ij}\right) = \Pr&\left(\gamma_0+\gamma_1 p_j+\gamma_2 TRS_j+\gamma_3 p_j TRS_j+\gamma_4\nu_{ij}+\gamma_5\sigma_j\right.\\ &\left.+\gamma_6 PG_{ij}+\gamma_7 p_j PG_{ij}+\gamma_8 TRS_j*PG_{ij}-\eta_{ij}>0\right)\\ = \frac{e^{\tau'_{ij}}}{1+e^{\tau'_{ij}}} &= \frac{e^{\gamma_0+\gamma_1 p_j+\gamma_2 TRS_j+\gamma_3 p_j TRS_j+\gamma_4\nu_{ij}+\gamma_5\sigma_j+\gamma_6 PG_{ij}+\gamma_7 p_j PG_{ij}+\gamma_8 TRS_j*PG_{ij}}}{1+e^{\gamma_0+\gamma_1 p_j+\gamma_2 TRS_j+\gamma_3 p_j TRS_j+\gamma_4\nu_{ij}+\gamma_5\sigma_j+\gamma_6 PG_{ij}+\gamma_7 p_j PG_{ij}+\gamma_8 TRS_j*PG_{ij}}},\end{aligned}$$

where $PG_{ij}$ is the product page indicator ($PG_{ij}=1$ if listing $j$ is displayed on the product page to consumer $i$, and 0 otherwise).

### 3.5.3 The Utility Functions

The utility of listing $j$ for consumer $i$ is denoted by $u_{ij}$. We divide the listings into two groups: (i) the targeted listings, which are fixed priced and brand new Halo Reach video game; and (ii) the non-targeted listings, which are the rest listings displayed on

[17] The definition of $\phi$ here is in line with the compensation rule in Hauser (2014).

the web pages. The non-targeted listings are possibly related to the Halo reach game. Denote by $T_i$ the set of targeted listings, and by $N_i$ the set of non-targeted listings seen by consumer $i$.

The utility of listing $j$ for consumer $i$ thus can be written as

$$u_{ij} = \tilde{u}_{ij} + \varepsilon_{ij},$$

where $\tilde{u}_{ij}$ is the representative utility, and $\varepsilon_{ij}$ is an independent Type I extreme value random variable.

The listings are described by total price $p$ (in US dollars), and TRS indicator. For targeted listings $j \in T_i$, the representative utility is given by[18]

$$\tilde{u}_{ij} = \beta_0 + \beta_1 p_j + \beta_2 TRS_j + \beta_3 p_j TRS_j.$$

For non-targeted listings $l \in N_i$, the representative utility is $\tilde{u}_{il} = \delta$. We assume that the outside listing (good 0), which represents buying nothing from eBay, has representative utility $\tilde{u}_{i0} = 0$. The same with the standard discrete choice models, the choice probability of listing $j$ conditional on consideration set $C_i \ni j$ by consumer $i$ is given as

$$\begin{aligned} s^*_{ij}(\mathbf{p}, \mathbf{x}|C_i) &= \Pr\{u_{ij} \geq u_{ik} : \forall k \in C_i\} \\ &= \frac{e^{\tilde{u}_{ij}}}{\sum_{k \in C_i} e^{\tilde{u}_{ik}}}. \end{aligned}$$

### 3.5.4 Estimation Approach

The model is estimated by the maximum likelihood approach, where the log-likelihood function should be written as

[18] On eBay the same listing is priced at the same level to different customers.

$$L(\gamma,\beta)=\sum_i\sum_j y_{ij}\log\sum_{C_i\in\mathbb{P}_i(j)}\prod_{j\in C_i}\phi_{ij}(p_j,\mathbf{x}_{ij})\prod_{k\in\mathcal{J}_i\backslash C_i}(1-\phi_{ik}(p_k,\mathbf{x}_{ik}))\,s^*_{ij}(\mathbf{p},\mathbf{x}|C_i),$$

where $y_{ij}$ is an indicator function such that $y_{ij}=1$ if consumer $i$ is observed to buy listing $j$, and $y_{ij}=0$ otherwise; $\phi_{ij}$ and $s^*_{ij}$ are given in the previous sections. An obvious problem is that, it is not computationally feasible to compute the choice probabilities by iterating over all possible $2^J$ consideration sets if $J$ is not a small number. To tackle this problem, we sample 100 consideration sets from the menu faced by consumer $i$.[19] Then, the choice probabilities are calculated by averaging over the 100 sampled consideration sets using weight given by $\pi$.[20] In this way, the consideration sets are not observed from the data, but are constructed through simulation.

## 3.6 Results and Discussion

We estimate the preferences and considerations by the ASC model, for both before and after periods. We have a standard logit demand with individual-level data and individual-specific sets of products displayed on the screen (menus).

### 3.6.1 Consideration and Preference Estimation

The estimation of the considerations is the main focus of this chapter. Table 3.3 gives the estimation results. We see that the total price negatively affects the probability of considering a listing on the screen. Note that free shipping indicator and click indicator draw attention sharply, with the free shipping indicator (before 4.30, after 4.64) having a larger effect compared to click indicator (before 2.69, after 2.35). For example, before the redesign, a listing listed by a non-TRS with free shipping at total price \$37, and has

[19] We sample a set of random uniform variables $u_0$ once, and through the optimization the consideration sets will be given by $C_i=\left\{j\in\mathcal{J}_i|\bar{\phi}_{ij}>u_0\right\}$, with $\bar{\phi}_{ij}=\frac{e^{\tau_{ij}+\kappa_{ijr}}}{1+e^{\tau_{ij}+\kappa_{ijr}}}$, and $\kappa_{ijr}$ being the log normal variable drawn for each $r=1,2,...,100$.

[20] This step is done to smooth the simulated choice probabilities.

Table 3.3: Consideration Estimation Results

| **Consideration Parameters** | Before | After |
|---|---|---|
| Total price | -0.06*** | -0.05*** |
| | (0.01) | (0.01) |
| TRS | 7.50** | 7.98*** |
| | (3.23) | (2.18) |
| Total price × TRS | -0.21*** | -0.21*** |
| | (0.07) | (0.05) |
| Click | 2.69*** | 2.35*** |
| | (0.28) | (0.18) |
| Free shipping | 4.30*** | 4.64*** |
| | (0.20) | (0.18) |
| Product page | | 0.06 |
| | | (0.09) |
| Total price × Product page | | -0.04*** |
| | | (0.01) |
| TRS × Product page | | 0.26 |
| | | (0.67) |
| Constant | -0.25*** | -0.22** |
| | (0.06) | (0.09) |

received no click is estimated to be considered by a consumer with probability 86%; a listing listed by a non-TRS without free shipping at total price \$37, and has received a click is estimated to be considered by a consumer with probability 55%; a listing listed by a non-TRS without free shipping at total price \$37, and has received no click is estimated to be considered by a consumer with probability 7.8%.

TRS indicator boosts the probability of being considered per se. On the other hand, TRS indicator amplifies the negative effect on consideration from total price, as shown by the coefficient (-0.21) for Total price × TRS. Before the redesign, a listing listed by a TRS has an equal probability of being considered with a listing listed by a non-TRS if both are priced at \$35.71 (mean transacted price is \$34.56 before the redesign). On the basis of this result, a listing listed by a TRS priced at higher than \$35.71 is estimated to be less likely considered compared to a listing listed by a non-TRS priced, ceteris paribus. After the redesign, this negative amplification effect from TRS indicator decreases: a listing listed by a TRS has an equal probability of being considered as a listing listed by a non-TRS if both are priced at \$38 (mean transacted price is \$33.30 after the redesign). Whether the listing is displayed on the product page or not has little effect on considerations, and the coefficients are not statistically significant.

In Table 3.4, we discuss the preference parameters. Again, it is not surprising to see that total price negatively affects utilities, both before and after the redesign. TRS indicator, and the coefficient of interaction term Total price × TRS are statistically not significant during the before period. Instead, after the redesign, TRS indicator is non-trivial in determining the utility: for instance, a listing listed by a TRS priced at \$35 has an equal utility as a listing listed by a non-TRS priced at \$36.08. We think the reason can be that the listings listed by TRSs are more easily identified on the product page after the redesign.

Table 3.4: Estimation Results by the ASC model and the Multinomial Logit Model

| | ASC model | Multinomial logit model | ASC model | Multinomial logit model |
|---|---|---|---|---|
| **Preference Parameters** | Before | Before | After | After |
| Constant (Halo Reach new fixed price) | 5.49*** | 4.17*** | -0.18 | 0.43 |
| | (1.73) | (1.12) | (0.73) | (0.77) |
| Total price | -0.26*** | -0.23*** | -0.13*** | -0.18*** |
| | (0.05) | (0.03) | (0.02) | (0.02) |
| TRS | 2.63 | 6.16** | 9.24*** | 11.10*** |
| | (3.41) | (2.59) | (2.06) | (1.81) |
| Total price × TRS | -0.05 | -0.15** | -0.23*** | -0.25*** |
| | (0.10) | (0.07) | (0.06) | (0.05) |
| Constant (other listings) | -5.37 | -5.57*** | -5.81*** | -6.06*** |
| | (0.09) | (0.08) | (0.07) | (0.06) |

### 3.6.2 Comparison with the Multinomial Logit Model

It is natural to make comparisons between the estimation results by the ASC model and that by the standard multinomial logit model. The standard multinomial logit model in Mcfadden (1974) does not take consideration set into account, therefore can be misleading if the consumers indeed fail to consider all the listings in the menu. We include the results by multinomial logit model in Table 3.4 (in blue) for comparison purposes.[21]

The standard multinomial logit model assumes that all the listings that are displayed on the screen are actually considered by the consumers. During the before period, the results give a TRS indicator effect of 6.16, and the null hypothesis that TRS indicator has no effect on utility is rejected at 1% significance level. However, by our baseline (ASC) model, the effect from TRS indicator on utility is estimated to be insignificant. This is because the multinomial logit model fails to discriminate the TRS indicator effect on utility against that on the consideration. The positive effect from TRS indicator on consideration is wrongly imposed on utility in the results by the multinomial logit model.

During the after period, the results from the multinomial logit model give a price effect of -0.18, 40% more than the estimated price effect (-0.13) by the ASC model; and the results by the multinomial logit model give a TRS indicator effect of 11.1, 20% more than the estimated TRS indicator effect (9.24) by the ASC model. From our baseline analysis (i.e., the results by the ASC model), there exist negative effect from total price, and positive effect from TRS indicator on the consideration. In the multinomial logit model, the price effect and the TRS indicator effect on both consideration and utility are aggregated on only utility, leading to biased estimates. As a result, the multinomial logit model exaggerates these two effects on utility.

### 3.6.3 Effect from the Platform Redesign

Another aim of our research is to find the effect from the platform redesign. To find that, we compare the marginal effects and total price elasticity during the pre-redesign period to that during the post-redesign period. As seen in Table 3.5, the marginal effect from total price decreases, so does the price elasticity. The marginal effect from seller

Table 3.5: Marginal Effects and Price Elasticity Comparison

| **Marginal Effects** | Before | After |
|---|---|---|
| Total Price | -0.002 | -0.00065 |
| | (0.005) | (0.0017) |
| TRS | 0.032 | 0.05 |
| | (0.068) | (0.13) |
| **Price Elasticity** | -1.21 | -0.77 |
| | (3.66) | (2.00) |

type (TRS indicator) increases.

Next, we use period dummy variable to indicate the change before and after the platform redesign. We rerun the regression, with extra regressors of the interactions. The result is shown in Table 3.6. The price effect on the consideration increases by 50% (0.03/0.06); and we reject the null hypothesis that there is no difference in price effects on the consideration, before and after the redesign at 1% significance level. On the other hand, the price effect on utility decreases by 50%(0.13/0.26); and we reject the null hypothesis that there is no difference in price effects on utility, before and after the redesign at 5% significance level. After the platform redesign, in the first stage (consideration stage), the price effect increases; while in the second stage (choice stage), the price effect decreases. The higher price effect on the consideration can be rationalized by the less difficulty in comparing the prices after the platform redesign. The lower price effect on utility can be resulted from the easiness in locating listings listed by TRSs, so that the consumers assigned more weight on high quality sellers and less weight on price.

[21]The results are estimated by the multinomial logit model using the utility functions defined in Sections 3.5.3.

Table 3.6: Difference in Coefficients Estimation

| | | | |
|---|---|---|---|
| **Preference Parameters** | | | |
| Constant (Halo Reach new fixed price) | 5.49*** | Constant (Halo Reach new fixed price) × Period | -5.77*** |
| | (1.73) | | (1.89) |
| Total price | -0.26*** | Total price × Period | 0.13** |
| | (0.05) | | (0.06) |
| TRS | 2.63 | TRS × Period | 6.32 |
| | (3.41) | | (4.03) |
| Total price × TRS | -0.05 | Total price × TRS × Period | -0.17 |
| | (0.10) | | (0.11) |
| Constant (other listings) | -5.36*** | Constant (other listings) × Period | -0.46*** |
| | (0.09) | | (0.12) |
| **Consideration Parameters** | | | |
| Total price | -0.06*** | Total price × Period | -0.03*** |
| | (0.01) | | (0.01) |
| TRS | 7.50** | TRS × Period | -0.50 |
| | (3.23) | | (3.72) |
| Total price × TRS | -0.21** | Total price × TRS × Period | -0.02 |
| | (0.07) | | (0.09) |
| Click | 2.69*** | Click × Period | 0.16 |
| | (0.28) | | (0.34) |
| Free shipping | 4.30*** | Free shipping × Period | 0.35 |
| | (0.20) | | (0.27) |
| Constant | -0.25*** | Period | 0.10 |
| | (0.06) | | (0.07) |

## 3.7 Conclusion

This chapter has explored the formation of consideration set by analysing the effect from total price, TRS indicator, free shipping indicator, click indicator, and page type. We find empirical evidence that even among the listings displayed on the page, not all the listings are considered. On the other hand, we do not find significant differences in considerations after the platform redesign.

The relaxation of the perfect attention assumption among the screen (i.e., all the listings displayed on the screen are considered) is non-trivial. The estimate results by the standard multinomial logit model show that TRS indicator significantly increases the utility. However, our result claims that TRS indicator insignificantly influences the utility; instead, it helps boost the probability a listing is considered under some circumstances. We thus enhance the argument that, without incorporating the consideration set properly, the estimation can be biased.

Nevertheless, we need to point out that our analysis is 'narrow', as we focus on a specific product, and the listings vary only in price, TRS indicator preference-wise; and in price, TRS indicator, free shipping indicator, click indicator, page type indicator consideration-wise. Yet, we draw a broader lesson from our analysis regarding the formation of consideration set. We find empirical evidence that consumers systematically form their consideration sets, and the effects from various factors are analysed quantitatively. The results potentially shed light on the consumers' consideration patterns and can be insightful for the online shopping platforms, sellers, and consumers.

## Acknowledgements

I thank Michael Dinerstein, Liran Einav, Jonathan Levin, and Neel Sundaresan for kindly sharing their dataset on American Economic Association website.

# Concluding Remarks

In this thesis, we have studied approval behaviour by using revealed preference techniques. In particular, Chapter 1 provides two general models that can serve as benchmarks for future research on approval. The two (deterministic and stochastic) models express approval functions in a succinct way, and explain the general decision rules in approval making. Chapter 2 provides a parametric model, which bridges the gap between theory and application, by providing a model that can be easily applied to approval data. In Chapter 3, we apply the ASC model to a detailed browsing dataset from eBay to study the approval behaviour in online commerce (consideration set formation). In Chapter 1 and 2, like most theoretical papers, we provide full characterizations of the models. The characterizations are simple and intuitive, which illustrate the properties of approvals in a neat way. We show that the preferences can be revealed fully given data generated by our models. In Chapter 3, we empirically analyse the consideration sets formation and the preferences of consumers. The quantitative analysis of the formation of consideration sets fills the gap in the existing literature.

We believe this thesis provides insightful interpretation of approval behaviour, and brings forth many possible research topics in the future. For instance, different parametric models can be extended from the general models in Chapter 1 for a variety of purposes (e.g. doing empirical analysis, explaining abnormal approval behaviour). The model in Chapter 2 is applicable to real world dataset, e.g. university admissions data, medical diagnoses data, online clicks data. Due to the constraint of our data, Chapter 3 focuses on specific product markets, where the listings vary only in a couple of covariates, and the product is highly homogeneous. We view the work in Chapter 3 as an initial step. With the increasing importance of online platforms, we think that further studies that would assess the consideration set formation in various contexts is a promising direction for further studies.

# Bibliography

[1] Alba, J. W. and Chattopadhyay, A. (1985) The effects of context and part-category cues on the recall of competing brands, *Journal of Marketing Research*, 22, 340–49.

[2] Anderson, S., de Palma, A. (1992) The logit as a model of product differentiation, *Oxford Economic Papers,* 44, 51–67.

[3] Aguiar, V.H., Boccardi, M.J., and Dean, M. (2016) Satisficing and stochastic choice, *Journal of Economic Theory*, 166, 445–82.

[4] Abaluck, J., and Adams-Prassl, A. (2021) What do consumers consider before they choose? Identification from asymmetric demand responses, *The Quarterly Journal of Economics,* 136(3), 1611–1663.

[5] Block, H.D., and Marschak, J. (1960) Random Orderings and Stochastic Theories of Responses, *Contributions to Probability and Statistics*, ed. Olkin, I., Gurye, S.G., Hoeffding, W., Madow, W.G., and Mann, H.B. Stanford, CA: Stanford University Press.

[6] Buschke, H. (1973) Selective reminding for analysis of memory and learning, *Journal of Verbal Learning & Verbal Behavior*, 12(5), 543–50.

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

[8] Bhargava, S., and Fisman, R. (2014) Contrast effects in sequential decisions: evidence from speed dating, *The Review of Economics and Statistics*, 96(3), 444–57.

[9] Brady, R.L., and Rehbeck, J. (2016) Menu-Dependent Stochastic Feasibility, *Econometrica*, 84(3), 1203–23.

[10] Berry, S., and Haile, P. (2016) Identification in Differentiated Products Markets, *Annual Review of Economics, 8*, 27–52.

[11] Caplin, A., Dean, M., and Martin, D. (2011) Search and Satisficing, *American Economic Review*, 101, 2899–2922.

[12] Chernev, A., Böckenholt, U., and Goodman, J. (2015) Choice overload: A conceptual review and meta-analysis, *Journal of Consumer Psychology*, 25(2), 333–358.

[13] Crawford, G., Griffith, R., and Iaria, A. (2021) A survey of preference estimation with unobserved choice set heterogeneity, *Journal of Econometrics*, 222(1), 4–43.

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

[15] Debreu, G. (1960) Review of "Individual Choice Behaviour" by R.D. Luce, *American Economic Review*, 50, 186–88.

[16] Danziger, S., Levav, J., and Avnaim-Pesso, L. (2011) Extraneous factors in judicial decisions, *Proceedings of the National Academy of Sciences of the United States of America*, 108(17), 6889–92.

[17] De los Santos, B., Hortaçsu, A., and Wildenbeest, M. (2012) Testing models of consumer search using data on Web browsing and purchasing behavior, *American Economic Review*, 102(1), 2455–80.

[18] Dinerstein, M., Einav, L., Levin, J. and Sundaresan, N. (2018) Consumer Price Search and Platform Design in Internet Commerce, *American Economic Review*, 108(7), 1820–59.

[19] Dardanoni, V., Manzini, P., Mariotti, M. and Tyson, C.J. (2020) Inferring cognitive heterogeneity from aggregate choices, *Econometrica*, 88(3),1269–96.

[20] Elstein A.S. (1988) Cognitive processes in clinical inference and decision making, *Reasoning, Inference, and Judgment in Clinical Psychology*, eds. Turk, D., and Salovey, P., 17–50. Free Press, New York.

[21] Eliaz, K. and Spiegler, R. (2011) Consideration Sets and Competitive Marketing, *The Review of Economic Studies*, 78(1), 235–262.

[22] Gulliksen, H. (1953) A generalization of Thurstone's learning function, *Psychometrika*, 18, 297–307.

[23] Goeree, M. S. (2008) Limited information and advertising in the US personal computer industry, *Econometrica,* 76(5), 1017–74.

[24] Gibbons, S., and Marshall, B. (2010) Assessing English: a trial collaborative standardised marking project, *English Teaching: Practice and Critique*, 9(3), 26–39.

[25] Gul, F., Natenzon, P., and Pesendorfer, W. (2010) Random Choice as Behavioral Optimization, *Econometrica*, 82(5), 1873–1912.

[26] Gaynor, M., Propper, C., and Seiler, S. (2016) Free to Choose? Reform, Choice, and Consideration Sets in the English National Health Service, *American Economic Review*, 106(11), 3521–57.

[27] Hauser, J., and Wernerfelt, B. (1990) An Evaluation Cost Model of Consideration Sets, *Journal of Consumer Research,* 16(4), 393–408.

[28] Hendel, I. (1999) Estimating Multiple-Discrete Choice Models: An Application to Computerizzation Returns, *Review of Economic Studies,* 66, 423–46.

[29] Hauser, J. (2011) A marketing science perspective on recognition-based heuristics (and the fast-and-frugal paradigm), *Judgment and Decision Making*, 6(5), 396–408.

[30] Horan, S. (2011) "Sequential Search and Choice from Lists", *working paper*, https://sites.google.com/site/smhoran/research.

[31] Hauser, J. (2014) Consideration-set heuristics, *Journal of Business Research*, 67, 1688–99.

[32] Honka, E., and Chintagunta, P. (2017) Simultaneous or Sequential? Search Strategies in the US Auto Insurance Industry, *Marketing Science*, 36(1), 21–42.

[33] Iyengar, S.S., and Lepper, M.R. (2000) When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology*, 79(6), 995–1006.

[34] Jelveh, Z., Kogut, B., and Naidu, S. (2018) Political Language in Economics, *Columbia Business School Research Paper*, No. 14–57. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2535453.

[35] Kuhn, H. W. (1953) Contributions to the Theory of Games II. Princeton, NJ: Princeton Univ. Press.

[36] Kahneman, D. (1973) Attention and Effort, Englewood Cliffs, NJ: Prentice Hall.

[37] Kotler, P. (1988) Marketing Management: Analysis, Planning, Implementation, and Control, 6th Ed. Englewood Cliffs, NJ: Prentice Hall International.

[38] Khandani, A.E., Kim, A.J., and Lo, A.W. (2010) Consumer credit-risk models via machine-learning algorithms, *Journal of Banking & Finance,* 34(11), 2767–87.

[39] Kovach, M., and Ulku, L., (2020), Satisficing with a Variable Threshold, *Journal of Mathematical Economics*, 87: 67-76.

[40] Luce, R.D. (1959) Individual Choice Behavior: A Theoretical Analysis. New York: Wiley.

[41] Lynch, J.G., Chakravarti, D., and Mitra, A. (1991) Contrast Effects in Consumer Judgments: Changes in Mental Representations or in the Anchoring of Rating Scales? *Journal of Consumer Research*, 18(3), 284–97.

[42] McFadden, D.L. (1974) Conditional logit analysis of qualitative choice behavior. in P. Zarembka, ed., *Frontiers in Econometrics*, Academic Press, New York: 105–42.

[43] Manski, C.F. (1977) The structure of random utility models, *Theory and Decision*, 8, 229–254.

[44] Manski, C., Sherman, L. (1980) An Empirical Analysis of Household Choice Among Motor Vehicles, *Transportation Research*, 14A, 349–66.

[45] McFadden, D.L. (2000) Economic Choices, *American Economic Review*, 91(3), 351–78.

[46] Manzini, P. Mariotti, M. (2007) Sequentially Rationalizable Choice, *American Economic Review*, 97(5), 1824–39.

[47] Mandler, M., Manzini, P., Mariotti, M. (2012) A million answers to twenty questions: Choosing by checklist, *Journal of Economic Theory*, 147(1), 71–92.

[48] Masatlioglu, Y., Nakajima, D. and Ozbay E.Y. (2012) Revealed Attention, *American Economic Review*, 102, 2183–2205.

[49] Manzini, P., Mariotti, M. and Tyson, C.J. (2013) Two-stage Threshold Representations, *Theoretical Economics*, 8, 875–882.

[50] Manzini, P., Mariotti, M. (2014) Stochastic choice and consideration sets, *Econometrica,* 82(3), 1153–76.

[51] Manzini, P., Mariotti, M. and Tyson, C.J. (2016) Partial Knowledge Restrictions on the Two-Stage Threshold Model of Choice, *Journal of Mathematical Economics*, 64 (4), 41–47.

[52] Manzini, P., Mariotti, M., and Ulku, L. (2021) Sequential Approval: A Model of "Likes", Paper Downloads and Other Forms of Click Behaviour, *working paper*, The University of Bristol. https://www.dropbox.com/s/bsofnkx6tugls5u/DepthAndLove_241220.pdf?dl=0.

[53] Pauker, S.G., and Kassirer, J.P. (1975) Therapeutic decision making: a cost-benefit analysis, *The New England Journal of Medicine,* 31, 293(5), 229–34.

[54] Pepitone, A., and DiNubile, M. (1976) Contrast Effects in Judgments of Crime Severity and the Punishment of Criminal Violators, *Journal of Personality and Social Psychology*, 33 (4), 448-459.

[55] Pauker, S.G., and Kassirer, J.P. (1980) The threshold approach to clinical decision making, *The New England Journal of Medicine,* 15, 302(20),1109–17.

[56] Poses, R.M., and Anthony, M. (1991) Availability, wishful thinking, and physicians' diagnostic judgments for patients with suspected bacteremia, *Medical Decision Making*, 11, 159–168.

[57] Pennacchiotti, M., and Popescu, A.M. (2011) A machine learning approach to Twitter user classification. *Proceedings of AAAI Conference on Weblogs and Social Media*, https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/download/2886/3262.

[58] Richter, M.K. (1966) Revealed Preference Theory, *Econometrica*, 34, 635–45.

[59] Rowe, P. M. (1967) Order effects in assessment decisions. *Journal of Applied Psychology*, 51(2), 170–173.

[60] Roberts, J.H., and Lattin, J.M. (1991) Development and testing of a model of consideration set composition, *Journal of Marketing Research*, 28(4), 429–440.

[61] Rubinstein, A., and Salant Y. (2006) A model of choice from lists, *Theoretical Economics,* 1(1), 3-17.

[62] Reutskaja, E., Nagel, R., Camerer, C.F. and Rangel, A. (2011) Search Dynamics in Consumer Choice under Time Pressure: An Eye-Tracking Study, *American Economic Review*, 101, 900–926.

[63] Rehbeck, J. (2021) Choice Overload via a Menu Dependent Luce Model, *working paper*, The Ohio State University. https://drive.google.com/file/d/1_-BL3A481OYawBnjMKc34ohO2plkl5ju/view.

[64] Samuelson, P.A. (1938) A Note on the Pure Theory of Consumer's Behaviour, *Economica*, 5(1), 61–71.

[65] Simon, H.A. (1956) Rational Choice and the Structure of the Environment, *Psychological Review*, 63(2), 129–138.

[66] Simon, H.A. (1959) Theories of decision-making in economics and behavioral science, *The American Economic Review*, 49(3), 253–283.

[67] Stigler, G.J. (1961) The Economics of Information, *Journal of Political Economy*, 69(3), 213–225.

[68] Tversky, A. (1972) Elimination by Aspects: A Theory of Choice, *Psychological Review*, 79(4), 281–299.

[69] Tversky, A., and Kahneman, D. (1973) Availability: A heuristic for judging frequency and probability, *Cognitive Psychology*, 5(2), 207–32.

[70] Tversky, A., and Kahneman, D. (1991) Loss aversion in riskless choice: a reference-dependent model, *Quarterly Journal of Economics*, 106, 1039–1061.

[71] Wang, K. (2021) Logit Function in Stochastic Categorizations, *Oxford Economic Papers*, forthcoming.

[72] Yildiz, K. (2016) List-rationalizable choice, *Theoretical Economics*, 11(2), 587-599.

# Appendix

## The Non-Parametric ASC model

The non-parametric ASC model is capable of eliciting $\phi_{ij}$ without imposing any parametric structures. It however requires the probability of selecting good $j$ conditional on observed $\mathbf{p}$, and all cross-derivatives as known for all possible prices $\mathbf{p} \in \mathbb{R}^{J+1}$, a standard assumption on the dataset in non-parametric identification, e.g. Berry and Haile (2016). It is most likely not feasible to have such a rich dataset. We now show that the specific model in Chapter 3 satisfies the assumptions which are sufficient for point identification in non-parametric analysis. Although we do not use this approach in Chapter 3, we believe this discussion enhances the idea that the ASC model is able to discriminate between consideration and preference, as in Abaluck and Adams-Prassl (2021).

Abaluck and Adams-Prassl (2021) show the common assumptions made in standard discrete choice models are enough for identification in the non-parametric ASC model. We will repeat these assumptions and clarify how the framework in Chapter 3 satisfies the assumptions.

**Assumption 1.** *Daly-Zachary Conditions: the unobserved choice probabilities,* $s^*_{ij}(\mathbf{p}, \mathbf{x}|C)$ *satisfy the conditions everywhere on* $\mathbf{p} \in \mathbb{R}^{J+1}$*:*

*(i) Properties:* $s^*_{ij}(\mathbf{p}, \mathbf{x}|C) \geq 0$, $\sum_{j \in C} s^*_{ij}(\mathbf{p}, \mathbf{x}|C) = 1$, *and* $\frac{\partial^J s^*_{ij}(j|C)}{\partial p_0 \ldots \partial p_{j-1} \partial p_{j+1} \ldots \partial p_n}$

*exists finitely, is* $\geq 0$*, and is continuous;*

*(ii) Symmetry: cross-price derivatives are symmetric:*
$\frac{\partial s^*_{ij}(\mathbf{p},\mathbf{x}|C)}{\partial p_{j'}} = \frac{\partial s^*_{ij'}(\mathbf{p},\mathbf{x}|C)}{\partial p_j}$;

*(iii) Absence of Nominal Illusion:*

$s^*_{ij}(\mathbf{p}+\delta,\mathbf{x}|C) = s^*_{ij}(\mathbf{p},\mathbf{x}|C)$.

It is not hard to see that the standard utility model used in this chapter satisfies all these conditions immediately. Assumption 1(ii) postulates that consumers value price equally across choices, which is also known as Slutsky symmetry. It is a substantive assumption, however it is assumed implicitly in almost all discrete choice models. Next, we prove that ASC conditions are satisfied as well. Denote by $\bar{\mathbf{p}}_{ij'}$ the price vector $\mathbf{p}$ with only $p_{ij'} \to \infty$.

**Assumption 2.** *ASC Conditions:*

*(i) As* $p_j \to -\infty, \phi_{ij}(p_j) \to 1$;[22]

*(ii)*$s_0(\mathbf{p},\mathbf{x}) - s_0(\bar{\mathbf{p}},\mathbf{x}) \neq 0$ *at all* $\mathbf{p} \in \mathbb{R}^{J+1}$.

Our estimation results in Section 3.6 show that the total price effect is negative, which makes Assumption 2 (i) immediately true. On the other hand, Assumption 2 (ii) imposes that the choice probability of default good changes if the prices of non-default goods change. It is quite weak and can be verified easily too. Now, we prove that, if Daly-Zachary Conditions and ASC conditions are satisfied, the latent probability of considering listing $j$ by consumer $i$ can be constructed by observed choice probabilities.

First, we prove that the unobserved consideration is identified with the observed $s$, which takes the form

[22]The negative price in Assumption 2 (i) is mainly made for construction reasons, as we need to pin down the constant of integration.

$$
\begin{aligned}
s_{ij}\left(\mathbf{p},\mathbf{x}\right) &= \sum_{C_i\in\mathbb{P}_i(j)}\prod_{l\in C_i}\phi_{il}\left(p_l,\mathbf{x}_l\right)\prod_{l'\in J_i\setminus C_i}\left(1-\phi_{il'}\left(p_{l'},\mathbf{x}_{l'}\right)\right)s^*_{ij}\left(\mathbf{p},\mathbf{x}|C_i\right)\\
&= \phi_{j'}\left(p_{j'},\mathbf{x}_{l'}\right)\sum_{C_i\in\mathbb{P}_i(j)\cap\mathbb{P}_i(j')}\hat{\pi}\left(C_i\right)s^*_{ij}\left(\mathbf{p},\mathbf{x}|C_i\right)\\
&+ \left(1-\phi_{j'}\left(p_{j'},\mathbf{x}_{l'}\right)\right)\sum_{C_i\in\mathbb{P}_i(j)\cap\mathbb{P}_i(j')}\hat{\pi}\left(C_i\right)s^*_{ij}\left(\mathbf{p},\mathbf{x}|C_i\setminus\{j'\}\right)\\
&= \phi_{j'}\left(p_{j'}\right)\sum_{C_i\in\mathbb{P}_i(j)\cap\mathbb{P}_i(j')}\hat{\pi}\left(C_i\right)\left(s^*_{ij}\left(\mathbf{p},\mathbf{x}|C_i\right)-s^*_{ij}\left(\mathbf{p},\mathbf{x}|C_i\setminus\{j'\}\right)\right)\\
&+ \sum_{C_i\in\mathbb{P}_i(j)\cap\mathbb{P}_i(j')}\hat{\pi}\left(C_i\right)s^*_{ij}\left(\mathbf{p},\mathbf{x}|C_i\setminus\{j'\}\right),
\end{aligned}
$$

where $\hat{\pi}\left(C_i\right)=\prod_{l\in C_i\setminus\{j'\}}\phi_{il}\left(p_l,\mathbf{x}_l\right)\prod_{l'\in J_i\setminus C_i}\left(1-\phi_{il'}\left(p_{l'},\mathbf{x}_{l'}\right)\right)$.

Note that the unobserved term can be inferred by leave-one-out observation $s_{ij}\left(\mathbf{p},\mathbf{x}|J_i\setminus\{j'\}\right)$, as

$$
\begin{aligned}
&\sum_{C_i\in\mathbb{P}_i(j)\cap\mathbb{P}_i(j')}\hat{\pi}\left(C_i\right)s^*_{ij}\left(\mathbf{p},\mathbf{x}|C_i\setminus\{j'\}\right)\\
&= \sum_{C_i\in\mathbb{P}_i(j)\cap\mathbb{P}_i(j')}\prod_{l\in C_i\setminus\{j'\}}\phi_l\left(p_l,\mathbf{x}_l\right)\prod_{l'\in J_i\setminus C_i}\left(1-\phi_{l'}\left(p_{l'},\mathbf{x}_{l'}\right)\right)s^*_{ij}\left(\mathbf{p},\mathbf{x}|C_i\setminus\{j'\}\right)\\
&= \sum_{C_i\in\mathbb{P}_i(j)\setminus\mathbb{P}_i(j')}\prod_{l\in C_i}\phi_l\left(p_l,\mathbf{x}_l\right)\prod_{l'\in J_i\setminus C_i}\left(1-\phi_{l'}\left(p_{l'},\mathbf{x}_{l'}\right)\right)s^*_{ij}\left(\mathbf{p},\mathbf{x}|C_i\right)\\
&= s_{ij}\left(\mathbf{p},\mathbf{x}|J_i\setminus\{j'\}\right).
\end{aligned}
$$

By Assumption 1(ii), the changes in latent $s^*$ cancel out within a consideration set, the cross derivative can be written as

$$
\begin{aligned}
\frac{\partial s_{ij}\left(\mathbf{p},\mathbf{x}\right)}{\partial p_{j'}}-\frac{\partial s_{ij'}\left(\mathbf{p},\mathbf{x}\right)}{\partial p_{j}} &= \frac{\partial \phi_{ij'}}{\partial p_{j'}}\sum_{C_i\in\mathbb{P}_i(j)\cap\mathbb{P}_i(j')}\hat{\pi}\left(C\right)\left(s^*_{ij}\left(\mathbf{p},\mathbf{x}|C\right)-s^*_{ij}\left(\mathbf{p},\mathbf{x}|C\setminus\{j'\}\right)\right)\\
&\quad-\frac{\partial \phi_{ij}}{\partial p_{j'}}\sum_{C_i\in\mathbb{P}_i(j)\cap\mathbb{P}_i(j')}\hat{\pi}\left(C_i\right)\left(s^*_{ij'}\left(\mathbf{p},\mathbf{x}|C_i\right)-s^*_{ij'}\left(\mathbf{p},\mathbf{x}|C_i\setminus\{j\}\right)\right)\\
&= \frac{\partial \phi_{ij'}}{\partial p_{j'}}\frac{1}{\phi_{ij'}}\left(s_{ij}\left(\mathbf{p},\mathbf{x}\right)-s_{ij}\left(\mathbf{p},\mathbf{x}|J_i\setminus\{j'\}\right)\right)\\
&\quad-\frac{\partial \phi_{ij}}{\partial p_{j}}\frac{1}{\phi_{ij}}\left(s_{ij'}\left(\mathbf{p},\mathbf{x}\right)-s_{ij'}\left(\mathbf{p},\mathbf{x}|J_i\setminus\{j\}\right)\right)\\
&= \frac{\partial \log\phi_{ij'}}{\partial p_{j'}}\left(s_{ij}\left(\mathbf{p},\mathbf{x}\right)-s_{ij}\left(\mathbf{p},\mathbf{x}|J_i\setminus\{j'\}\right)\right)\\
&\quad-\frac{\partial \log\phi_{ij}}{\partial p_{j}}\left(s_{ij'}\left(\mathbf{p},\mathbf{x}\right)-s_{ij'}\left(\mathbf{p},\mathbf{x}|J_i\setminus\{j\}\right)\right).
\end{aligned}
$$

If the leave-one-out observation is unobservable, the identification is shown to be feasible using the large support assumption on prices. Let $\mathbf{p}'_{j'}$ be the price vector which is same with $\mathbf{p}$ except that $p'_{j'} > p_{j'}$. With a weak assumption that $s^*_{ij}\left(\mathbf{p}'_{j'},\mathbf{x}|C\right)\to s^*_{ij}\left(\mathbf{p}'_{j'},\mathbf{x}|C\setminus j'\right)$ if $p_{ij'}\to\infty$ (which is satisfied by our utility function, as well as that in all parametric additive random utility models), we have

$$
\begin{aligned}
s_{ij}\left(\mathbf{p},\mathbf{x}\right)-s_{ij}\left(\mathbf{p}'_{j'},\mathbf{x}\right) &= \phi_{ij'}\left(p_{j'},\mathbf{x}_{l'}\right)\sum_{C\in\mathbb{P}(j)\cap\mathbb{P}(j')}\hat{\pi}\left(C\right)\left(s^*_{ij}\left(\mathbf{p},\mathbf{x}|C\right)-s^*_{ij}\left(\mathbf{p},\mathbf{x}|C\setminus j'\right)\right)\\
&\quad-\phi_{ij'}\left(p'_{j'},\mathbf{x}_{l'}\right)\sum_{C\in\mathbb{P}(j)\cap\mathbb{P}(j')}\hat{\pi}\left(C\right)\left(s^*_{ij}\left(\mathbf{p}'_{j'},\mathbf{x}|C\right)-s^*_{ij}\left(\mathbf{p}'_{j'},\mathbf{x}|C\setminus j'\right)\right)\\
&= \phi_{ij'}\left(p_{j'},\mathbf{x}_{l'}\right)\sum_{C\in\mathbb{P}(j)\cap\mathbb{P}(j')}\hat{\pi}\left(C\right)\left(s^*_{ij}\left(\mathbf{p},\mathbf{x}|C\right)-s^*_{ij}\left(\mathbf{p},\mathbf{x}|C\setminus j'\right)\right)\\
&= s_{ij}\left(\mathbf{p},\mathbf{x}\right)-s_{ij}\left(\mathbf{p},\mathbf{x}|J_i\setminus\{j'\}\right),
\end{aligned}
$$

recall that $\bar{\mathbf{p}}_{ij'}$ is the price vector $\mathbf{p}$ with only $p_{ij'}\to\infty$. We can write the cross-derivative difference as

$$\frac{\partial s_{ij}\left(\mathbf{p},\mathbf{x}\right)}{\partial p_{j'}}-\frac{\partial s_{ij'}\left(\mathbf{p},\mathbf{x}\right)}{\partial p_j}=\frac{\partial \log \phi_{ij'}}{\partial p_{j'}}\left(s_{ij}\left(\mathbf{p},\mathbf{x}\right)-s_{ij}\left(\bar{\mathbf{p}}_{ij'},\mathbf{x}\right)\right)-\frac{\partial \log \phi_{ij}}{\partial p_j}\left(s_{ij'}\left(\mathbf{p},\mathbf{x}\right)-s_{ij'}\left(\bar{\mathbf{p}}_{j},\mathbf{x}\right)\right),$$

and replace $j'$ by 0, we have

$$\begin{aligned}\frac{\partial s_{ij}\left(\mathbf{p},\mathbf{x}\right)}{\partial p_0}-\frac{\partial s_{i0}\left(\mathbf{p},\mathbf{x}\right)}{\partial p_j}&=-\frac{\partial \log \phi_{ij}}{\partial p_j}\left(s_{i0}\left(\mathbf{p},\mathbf{x}\right)-s_{i0}\left(\bar{\mathbf{p}}_{j},\mathbf{x}\right)\right)\Leftrightarrow\\ \frac{\partial \log \phi_{ij}}{\partial p_j}&=\frac{\frac{\partial s_{i0}(\mathbf{p},\mathbf{x})}{\partial p_j}-\frac{\partial s_{ij}(\mathbf{p},\mathbf{x})}{\partial p_0}}{s_{i0}\left(\mathbf{p},\mathbf{x}\right)-s_{i0}\left(\bar{\mathbf{p}}_{j},\mathbf{x}\right)}.\end{aligned}$$

Therefore, by integration we have

$$\phi_{ij}\left(\tilde{p}_j\right)=\exp\left(\int_{-\infty}^{\tilde{p}_j}\frac{\frac{\partial s_{i0}(\mathbf{p},\mathbf{x})}{\partial p_j}-\frac{\partial s_{ij}(\mathbf{p},\mathbf{x})}{\partial p_0}}{s_{i0}\left(\mathbf{p},\mathbf{x}\right)-s_{i0}\left(\bar{\mathbf{p}}_{j},\mathbf{x}\right)}dp_{ij}\right),$$

as $\log \phi_{ij}\left(-\infty\right)=0$ by Assumption 2 (ii).