Essays on Consumer Heterogeneity and Personalized Discounts in an Online Market

A Thesis
submitted to the Faculty of the
Graduate School of Arts and Sciences
of Georgetown University
in partial fulfillment of the requirements for the
degree of
Doctor of Philosophy
in Economics

By

Chengjun Zhang, B.S.

Washington, DC
April 16, 2024

Essays on Consumer Heterogeneity and Personalized Discounts in an Online Market

Chengjun Zhang, B.S.

Thesis Advisor: John Rust, Ph.D.

Abstract

This thesis delves into consumer heterogeneity in an online marketplace from an empirical lens on business practices. Furthermore, it evaluates the welfare consequences of employing personalized discounts as a strategic marketing approach.

The first chapter utilizes comprehensive consumer clickstream data to construct and refine demand models for smartphones on an e-commerce platform. The narrative unfolds through the exploration of increasing levels of consumer heterogeneity, built upon the conditional logit framework. The last model directly leverages consumer historical clickstreams with a recurrent neural network (RNN), offering detailed individual-level preferences and realistic product substitution patterns. This model excels by outperforming other models in both in-sample and out-of-sample fit.

The second chapter, building upon the demand model established in the first, conducts a counterfactual analysis that enables the issuance of personalized discounts tailored to individual consumer preference parameters. Using a numerically stable algorithm, this chapter presents empirical evidence that highlights the welfare implications. The findings illuminate a mutually beneficial scenario for firm profitability and consumer welfare, in conditional expected terms.

Index words: Consumer Clickstream, Consumer Heterogeneity, E-commerce, Machine Learning, Multi-layer Perceptron, Recurrent Neural Network, Personalized Discount, Price Discrimination

Dedication

To the past, the present, and the future.

To my family, my friends, and myself.

Acknowledgments

I wish to express my profound gratitude to my advisor, Professor John Rust, for his unwavering support, insightful guidance, and invaluable mentorship throughout this journey. I am also immensely thankful to the members of my dissertation committee, Professors Sharat Ganapati and Nathan Miller, for their constructive feedback, thought-provoking questions, and valuable suggestions.

I owe a deep appreciation to my parents and family for their love, patience, and steadfast belief in me. Their encouragement and emotional support have been my stronghold through the challenges and moments of doubt.

A special thank you goes to my fellow classmates for their camaraderie, stimulating discussions, and helping hands in times of need.

Lastly, I would like to acknowledge my own dedication, resilience, and perseverance on this journey. This endeavor has been a significant personal and professional growth experience, teaching me the value of hard work, persistence, and the courage to face challenges head-on.

This journey has been transformative, and my gratitude extends to every individual who has contributed to it.

Table of Contents

List of Figures

List of Tables

Chapter 1

Consumer Heterogeneity in Online Shopping: A Machine Learning Approach with Sequential Data

1.1 Introduction

In an era where the digital realm is more interconnected than ever, firms are increasingly compelled to harness the potential of available information to carve out a competitive advantage through marketing strategies. The declining costs of data collection and storage have elevated the role of "big data" to a critical component in strategic decision-making. This phenomenon has underscored the immense potential of big data across various domains, particularly in marketing, as highlighted by Acquisti et al. (2016). Central to this discourse are two pivotal questions: How can firms effectively utilize the granular data for marketing, and what implications does this have for the welfare of both firms and consumers? This paper endeavors to unravel these questions in two chapters respectively. The first chapter introduces a novel approach for identifying consumer heterogeneity using clickstream data, while the second chapter offers a counterfactual analysis on the welfare impacts of personalized discounts, informed by the identified heterogeneity.

The transition of modern economies toward production of knowledge and recent radical advancements in information technology (in particular, the rise of the Internet) have vastly enlarged the amount of individual information that can be collected, stored, analyzed, and repurposed for new uses (Acquisti et al., 2016). This digital revolution has paved the way for firms to adopt personalized pricing strategies to tailor prices according to individual consumers based on their preferences. Since the firm can always charge every consumer the same uniform price, revealed preference indicates monopoly personalized pricing will always weakly increase the firm's profits (Dubé and Misra, 2023). However, the implementation of such a marketing strategy ignites a complex debate surrounding its implications on consumer welfare, and consequently, social welfare. On the one hand, firms have the potential to transform nearly the entire consumer surplus into profit, thereby raising concerns about the equitable distribution of economic benefits. On the other hand, it offers a nuanced advantage by making products accessible to a broader demographic, including those who might otherwise be priced out of the market. This dual-edged nature of personalized pricing presents an intricate puzzle regarding its ultimate impact on consumer surplus, warranting further explorations of its wider economic impact.

To navigate the complexities and assess the impact of personalized marketing strategies on welfare, it is crucial to first pinpoint and understand consumer heterogeneity. A significant concern among consumers revolves around the utilization of personal data for targeted marketing efforts, including sensitive demographic information such as age, gender, race, location, and other characteristics. Shiller (2014) posits that such demographic details could have been used in face-to-face transactions, informed by observable cues like physical appearance, accent, and attire. In response to consumer apprehensions regarding the sharing of personal demographic information, this chapter deliberately avoids employing such data, instead providing valuable insights into scenarios where consumers are hesitant to divulge personal demographics to firms. The approach adopted herein capitalizes on the analysis of consumers' clickstream data to identify the heterogeneity in consumer preferences in the e-commerce smartphone market. To achieve this, I employ a range of machine learning algorithms, selected for their flexible and non-parametric nature, while maintaining the interpretability of traditional economics models.

The first strand of literature draws upon clickstream data, or browsing and purchase histories in general, to examine consumer behavior in the realms of marketing and management. A substantial body of literature is dedicated to recommender systems and product search refinement, with the primary aim of enhancing the consumer's product search experience. For a thorough overview and a comprehensive exploration of various tasks within this domain, the reader is referred to Quadrana et al. (2018). Additionally, Montgomery et al. (2004) and Bigon et al. (2019) delve into consumer purchase conversion, while Rajamma et al. (2009) and Kukar-Kinney and Close (2010) address the issue of shopping cart abandonment in online markets. An investigation by Huang and Van Mieghem (2014) centers on inventory management. Beranek et al. (2017) focus on fraud detection. Furthermore, Moe (2003), Schellong et al. (2016), and Pallant et al. (2017) distinguish consumers across various categories, including buying, browsing, searching, and knowledge-building behaviors. Numerous other research topics harness the potential of clickstream data, such as consumer retention and lifetime value analysis. However, among these topics, the analysis of the relationship between consumer clickstream behavior and their preferences aligns most closely with the subject of this chapter. Rossi et al. (1996) underscore the considerable potential for enhancing the profitability of direct marketing efforts by leveraging household purchase histories. They conclude that even short purchase histories can yield a net gain in revenue through targeted couponing, which outperforms blanket couponing by 2.5 times. Montgomery et al. (2004) suggest that clickstream data may contain valuable insights into a user's goals, knowledge, and interests, which can aid in predicting their future actions on a website.

The second strand of related literature is about the conditional logit model established by McFadden (1974), and its variants incorporating parametric or non-parametric estimation of heterogeneity. The conditional logit model is one of the workhorse models in the empirical economics literature analyzing consumer purchase decisions. Although, the limitation of the conditional logit model in producing realistic substitution patterns due to the independence of irrelevant alternatives (IIA) assumption is apparent, it continues to be the backbone of modeling choices based on random utilities in the literature. Many variants of the vanilla logit model have been introduced and applied to produce more realistic substitution patterns, at the same time, avoiding the computational difficulty of multi-normal integral when estimating probit models. For example, the nested logit model Ben-Akiva (1973) and McFadden (1978) allows for a broader set of substitution patterns by partitioning the choice set into smaller subsets (nests) and maintaining the IIA assumption only within nests instead of across all alternatives. The mixed logit model, also known as the random-parameter logit model was introduced by Cardell and Dunbar (1980) and Boyd and Mellman (1980). With the advancement of numerical integration using computers, Bhat (1998), Brownstone and Train (1998), Erdem (1996), Revelt and Train (1998), Bhat (2000) and many others had leveraged the power of mixed logit models. The latent class models are often similar or special cases of the mixed logit models with discrete distributions of coefficients. Heckman and Singer (1984) relax the parametric assumption of coefficients and propose a non-parametric method of estimating preferences with mass points. Kamakura and Russell (1989) and Chintagunta et al. (1991) also estimate the parameters by creating segments of the population. Farrell et al. (2021) is an example of non-parametric estimations of heterogeneity at the individual level using deep neural networks. This is closely related to the strand of literature on machine learning discussed later.

The third strand of related literature encompasses machine learning (ML) and artificial intelligence (AI), with a particular focus on advanced methodologies such as deep neural networks (DNNs), recurrent neural networks (RNNs), and other sequence-oriented models. Within this chapter, the multi-layer perceptron (MLP) is employed as a non-parametric tool for estimating consumer preference parameters, drawing parallels with the approach of Farrell et al. (2021). Given the sequential nature of consumer clickstreams, this chapter further extends the analysis by incorporating an RNN to model consumer heterogeneity based on browsing history. One of the most common applications of RNN and related models such as gated recurrent units (GRU) (Cho et al., 2014) and long short-term memory (LSTM) (Hochreiter and Schmidhuber, 1997) is natural language processing (NLP). Interestingly, consumer clickstreams and panel data in general bear a resemblance to language models, a connection that may not be immediately apparent. The advent and rapid proliferation of the Transformer model (Vaswani et al., 2017), along with large language models (LLMs), and the broader domain of generative artificial intelligence (GenAI), are likely to accelerate the development of analytical tools and methodologies for analyzing large and high-dimensional panel data in the field of economics.

This chapter introduces five models that build upon the vanilla conditional logit framework, each incorporating progressively more intricate levels of consumer heterogeneity to model and estimate smartphone demand on an e-commerce platform. Some heterogeneity among users is observable to econometricians, and heuristic-based classifications can yield insights into user preference parameters. The reduced-form results reveal a significant correlation between users' browsing histories and the outcomes of heuristic-based user type classifications. Furthermore, demand estimations that include user types uncover a distinct association between user types and their preference parameters. Consequently, this chapter employs machine learning models to harness the observed heterogeneity in user types and their historical clickstream data for the estimation of individual-level preference parameters. The novel application of historical browsing behavior not only enhances model fit but also achieves this with a reduced number of parameters compared to the MLP model. This approach results in a demand model that captures a rich tapestry of consumer heterogeneity and realistic substitution patterns.

The structure of the ensuing sections is as follows. Section 1.2 delves into the behavioral data of users used in this study. Section 1.3 introduces the heuristics applied to classify users, view events, and event durations, while Section 1.4 demonstrates the outcomes of descriptive analyses highlighting the connection between heuristic-based user types and their browsing behavior. Section 1.5 elaborates on the models and discusses the estimation results, taking into account varying degrees of heterogeneities. Section 1.6 presents the price elasticities derived from the estimated models. Finally, Section 1.7 rounds off the chapter with the conclusions.

1.2 Data

1.2.1 Privacy Disclaimer

The data source for this study will be kept confidential to protect the company's privacy. As such, I will refrain from discussing specific details that could lead to the identification of the company, in order to maintain confidentiality and protect proprietary information.

1.2.2 Clickstream Dataset

The clickstream dataset offers a detailed chronicle of consumer interactions over an extensive period of approximately 200 days, capturing the essence of digital footprints left by users. Each entry in the dataset corresponds to a *click* (or *event*), encompassing details including a timestamp, event type, product ID, product price at the time of the event, and user ID. This dataset categorizes user clicks into three event types: *view*, *cart*, and *purchase*. A *view* event encompasses four types of user engagements, navigating to the product's webpage (e.g., from a product list page), enlarging a product image (on the product webpage), loading additional comments (on the product webpage), or refreshing the product webpage. However, these types of view events cannot be directly distinguished from each other in the data. To refine this, Section 1.3.2 introduces an inventive heuristic method for classifying view events with enhanced specificity. A *cart* event is recorded when a product is added to the shopping cart, and a *purchase* event signifies the completion of a product purchase by the user.

Unique identifiers trace the trajectory of users and products throughout this voluminous dataset, which is composed of events tallying in the hundreds of millions.

1.2.3 Product and User Information

The dataset presents an expansive array of approximately 400,000 unique products, each linked to a unique URL showcasing extensive details as seen by users, product name, categories (up to four levels of subcategories), features, ratings, reviews, images, and other facets.

This study zooms in on the smartphone sector, a pivotal category within this collection. Despite representing a mere fraction, less than 0.5%, of the entire product range, smartphones command a disproportionate share of activity, accounting for a quarter of all recorded events. This pronounced emphasis on smartphones sets an optimal scenario for demand analysis, leveraging the considerable volume of data against a relatively narrow selection of product choices.

The analysis is particularly concerned with users who have made no more than a single smartphone purchase within the study's timeframe, constituting about 60% of all participants who have interacted with this category. This filtering aims to exclude potential resellers, or those buying for gifting purposes. This defined group, which amounts to 5,647,094 users, is collectively referred to as *all users*. It is acknowledged, however, that some within this group may be purchasing smartphones on behalf of others. In such instances, these transactions are considered representative of the purchasing patterns for another individual.

Comprehensive specifications for each smartphone are meticulously collected from their individual product pages, capturing essential details such as brand, storage capacity, RAM size, battery capacity, screen size, and camera resolution, among others.[1]

1.2.4 Assumptions on Prices and Availabilities

The assumptions on prices and availabilities play a pivotal role in shaping the choice sets for individual consumers, which are essential for the subsequent model estimation and inference processes. Firstly, due to the high volume of user interactions on a daily basis, it is assumed that the price of a product at any given moment mirrors the price recorded in the latest event for that product, up to and including that specific time point. Secondly, the availability window for purchasing a smartphone is defined by the span from its initial event in the dataset to the final recorded event concerning that device.

Consequently, a user's choice set comprises the array of all smartphones available at the time the user is active on the platform. On average, approximately 1,500 smartphones are available at any given moment, contributing to a cumulative total

[1] Users who authored product reviews cannot be linked to their click events within the dataset, rendering them effectively anonymous.

of approximately 1,700 smartphones.

1.2.5 E-commerce Platform Information

The dataset is sourced from a leading multi-category e-commerce giant operating exclusively within a single country and offers a wide range of product categories. A notable characteristic of the smartphone market in this country is the sales without any contractual obligations, avoiding the practice of bundling these devices with service plans. Additionally, the retailer adheres to a policy of uniform pricing, ensuring equitable charges for all customers purchasing identical items at the same time, thus precluding any form of price differentiation.

Dominating the national e-commerce landscape, this platform significantly outperforms its rivals, securing its position as the country's premier online shopping destination. The nearest e-commerce competitors, ranking from second to seventh in size, predominantly engage in sectors like travel and apparel, which are unrelated to this study's focus on smartphones. This leading e-commerce platform accounts for nearly half of the entire online transaction volume within the country and exceeds the transaction revenue of its closest competitor in the online electronics segment by a margin of over ten times.

1.3 Classification Based on Heuristics

The dataset reveals significant unrecorded variations, which could significantly shift econometricians' approach to analyzing and understanding observed consumer actions. Specifically, varied search patterns among users may signal their distinct intentions prior to making decisions. Moreover, despite being uniformly documented, *view* events correspond to a broad spectrum of online activities, potentially obscuring valuable insights if aggregated indiscriminately. The prolonged intervals between clicks also pique curiosity, leading to speculation about user engagement during these periods. To navigate through this complexity, heuristic methodologies have been adopted to systematically delineate user types, view clicks, and event durations, with further explanations as follows.

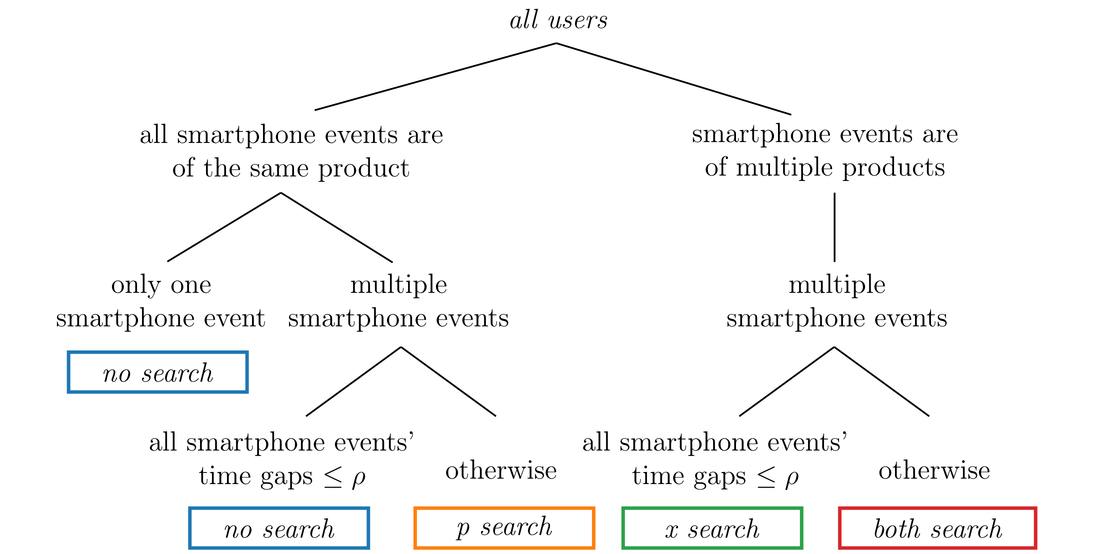
1.3.1 User Types

Among all product attributes, price stands out as a focal point of economic analysis of demand. Its potential to fluctuate throughout the day drives the examination of consumer behaviors related to prices. Although this study abstracts from modeling the dynamics of prices and treats such changes as exogenous, it is these very fluctuations that cast light on fascinating patterns of consumer search intentions. Notably, it is observed that a subset of consumers exhibits a strategic patience, opting to delay purchases in anticipation of potential price reductions. This behavior has informed a heuristic framework for categorizing consumers into four distinct groups based on their engagement in the smartphone market: *no search*, *p search*, *x search*, or *both search*. This classification schema is visually encapsulated in Figure 1.1, where ρ denotes a threshold of choice.

For each user, the *time gap* for every event is calculated by the time difference between the timestamp of this event and its immediate successor within the same subcategory (at the most detailed level), if such an event exists.[2]

Users categorized as *no search* engage with only one smartphone product, potentially in repeated instances, where the time gaps for all smartphone events are within ρ. Their activities may involve viewing the product, refreshing the page, or other

[2] i.e., *time gap* is not defined for the last event in each subcategory (at the most detailed level) for each user.



Share	48.20%	5.15%	20.39%	26.25%
Conversion	6.26%	3.87%	9.73%	12.51%
Avg Δp	$-\$0.002$	$-\$0.319$	$+\$0.002$	$-\$0.702$

Figure 1.1: User Types Classification Tree

forms of engagement on the product page, yet the time between these interactions remains within the ρ threshold, ultimately leading to a decision to purchase or not.

Users defined as *p search* are focused on monitoring price changes, either in anticipation of a reduction or due to unintentional delays that might lead to encountering a price change. These individuals repeatedly interact with a single smartphone model, with at least one of their interactions separated by a time gap greater than ρ, indicating a period of waiting that could influence their purchasing decision due to potential price fluctuations.

In contrast, *x search* user are interested in examining various smartphone features aside from price. These users explore multiple smartphone models, with each of their interactions closely followed by another, keeping all time gaps below the ρ threshold.

This indicates a consistent and uninterrupted exploration of options.

Users identified as *both search* display characteristics of both *x search* and *p search* behaviors, interacting with several models and experiencing at least one time gap that surpasses the ρ threshold. This suggests a deliberate or accidental period of waiting for potential price adjustments.

The parameter ρ, chosen as a hyperparameter, signifies a duration deemed sufficient to classify a user's pause as indicative of waiting for a price change. The categorization of users into different types based on varying ρ values is detailed in Table A.1, with a 30-minute interval chosen as a reasonable value for this study. As a result, the shares, conversion rates, and Avg Δp for each user type are shown in Figure 1.1, where Avg Δp quantifies the variation in prices paid for purchases relative to the prices at the initial click of the products on average.

The categorization of user types, while insightful, may not capture the full complexity of user behavior, particularly activities beyond product pages, e.g., interactions on search result pages. Therefore, this framework should be regarded as a heuristic approximation, reflecting the user searching characteristic as observed from the available data.

1.3.2 View Event Types

As previously highlighted, a view event encompasses a range of actions. While it's challenging to differentiate among these actions directly, they can be classified based on the nature of user interaction. A view event is categorized as a click *to page* when it entails a user navigating directly to a product's webpage. Conversely, it is deemed a click *on page* if it arises from actions taken on the webpage itself, such as enlarging a product picture, loading more comments, or refreshing the page. To effectively segregate these types of view events, a heuristic approach is utilized.

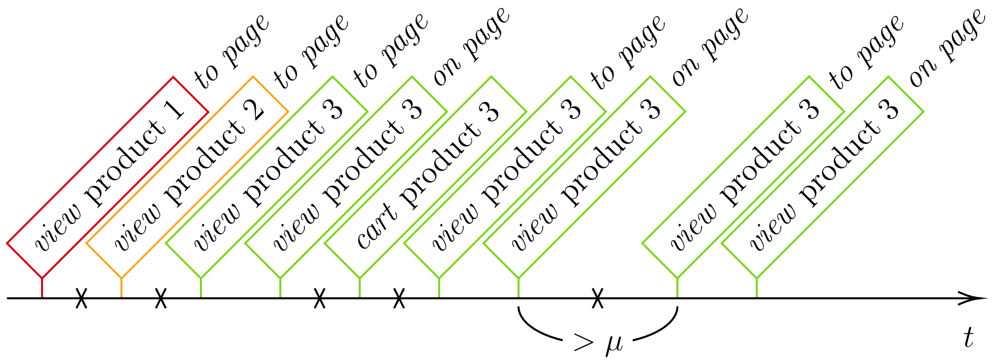


Figure 1.2: View Type Classification

Let μ be a tunable hyperparameter. For any given user, the clickstream is a sequence of events. Consider breaking the sequence into disjoint subsequences before every event that has a change (compared to the previous event, if any) in type (*view*, *cart*, or *purchase*) or in product, and after every event that has an event duration $> \mu$. Figure 1.2 shows a simple example. Any disjoint subsequence that consists only of *view* events is called a *view subsequence*. By construction, a *view subsequence* is the longest possible contiguous and uninterrupted subsequence of *view* events of the same product.[3] Hence, the first event in a *view subsequence* is classified as click *to page*, and the rest are classified as click *on page*. The intuition is simple, if a user has consecutive *view* events on the same product with no pause longer than μ, then the user is considered to be clicking *on page* for all the events except the very first one, which is a click *to page*. Table A.2 details the breakdown of view event types across various μ values. For the purposes of this study, a μ value of 5 minutes is selected as a suitable cut-off.

[3]A period of inactivity longer than μ is considered as an interruption.

A user may pause for a period longer than μ and *revisit* a previously clicked product. Setting μ at 5 minutes, 32.26% of users exhibit at least one revisit to a smartphone product page. This suggests a trend where users return to product pages, possibly to delve deeper into product details or to refresh their view on a product they have previously considered.

1.3.3 Event Durations

Determining the duration each user spends on a product is essential, as it reflects the user's engagement level with a particular option in the choice set, thereby offering insights into user preferences. The timestamps of the events enables the creation of such measure. For an individual user, the *event duration* of an event is calculated by the time difference between the event in question and the subsequent event, if there is one.[4]

Challenges arise when users take prolonged breaks, leading to inflated event durations that incorporate these inactive periods. To address this, I employ a heuristic approach: for any given event, if the event duration is $\leq \mu$, it is presumed that the user was engaged with the product page throughout this period, classifying the event as an *uninterrupted* view click. Conversely, if the event duration exceeds μ or if it's undefined (as in the case of a user's final event), the event is considered as an *interrupted* view click. It's important to note that this doesn't cap a user's total viewing time for a product at μ, as multiple interactions with the product, each spaced within μ of the previous, could accumulate to surpass μ.

[4]The event duration is undefined for a user's final event. And for each event, if the subsequent event by the same user pertains to a product within the same subcategory, the event duration equals the time gap. Conversely, if the next event involves a different subcategory, the event duration is less than the time gap.

1.4 Descriptive Analysis on User Types

While the primary focus of this study is on consumer choices within the smartphone category, observing users' historical clickstreams across other categories contributes to a more comprehensive understanding of them. This raises the question of whether there's a link between how users engage with the smartphone category and their activities in other segments. To explore this, a series of descriptive analyses are undertaken.

1.4.1 User Types and Indicators of Events

The heuristic classification of user types described in Section 1.3.1 offers a basic framework for understanding user behavior within the smartphone category. Utilizing a simple logit regression can shed light on the relationship between users' smartphone selection behaviors and their historical clickstream data. Table 1.1 presents the results from a multinomial logit model, where the dependent variable is the user type with *p search* as the base outcome, and the independent variables are indicators of events across various product categories, excluding smartphones.[5]

The categories of child goods and home stand out as having negative and statistically significant coefficients, indicating that users active in these categories tend towards the *p search* type (the base outcome) more than those who aren't. This trend suggests that users engaging with these categories may have familial responsibilities, prioritizing cost over other factors, thus making them more inclined to wait for price reductions and less likely to extensively explore the smartphone offerings.

Conversely, categories such as car goods, computers, construction and repair, fashion, furniture, gifts and party supplies, home equipment, leisure, pharmacy, phone gadgets, shoes, and sports and outdoors exhibit positive coefficients for all non-base

[5] Events within the office and school supplies, as well as food categories, are also excluded from consideration due to their infrequency among users ($\leq 0.05\%$).

Table 1.1: Estimation Result of Multinomial Logit

	no search		*x search*		*both search*	
Indicator of events in:						
- beauty care	0.016	(0.009)	0.122***	(0.009)	0.137***	(0.009)
- car goods	0.187***	(0.009)	0.441***	(0.009)	0.549***	(0.009)
- child goods	−0.455***	(0.007)	−0.726***	(0.007)	−0.428***	(0.007)
- computers	0.237***	(0.008)	0.602***	(0.008)	0.691***	(0.008)
- construction and repair	0.120***	(0.011)	0.198***	(0.011)	0.239***	(0.011)
- fashion	0.348***	(0.012)	0.564***	(0.012)	0.543***	(0.012)
- fashion accessories	−0.010	(0.008)	0.063***	(0.008)	0.192***	(0.008)
- furniture	0.197***	(0.010)	0.317***	(0.010)	0.339***	(0.010)
- gifts and party supplies	0.224***	(0.033)	0.274***	(0.033)	0.297***	(0.033)
- home	−0.114***	(0.009)	−0.145***	(0.010)	−0.167***	(0.010)
- home equipment	0.219***	(0.007)	0.478***	(0.008)	0.571***	(0.007)
- leisure	0.101***	(0.013)	0.131***	(0.013)	0.167***	(0.013)
- pet goods	0.152**	(0.049)	0.079	(0.050)	−0.002	(0.050)
- pharmacy	0.082**	(0.025)	0.085**	(0.026)	0.086***	(0.025)
- phone gadgets	0.213***	(0.009)	0.781***	(0.009)	1.160***	(0.009)
- shoes	0.105***	(0.011)	0.178***	(0.011)	0.256***	(0.011)
- sports and outdoors	0.086***	(0.010)	0.153***	(0.011)	0.150***	(0.011)
- tv audio	0.006	(0.006)	0.264***	(0.007)	0.523***	(0.007)
Constant	2.174***	(0.002)	1.068***	(0.002)	1.064***	(0.002)
Observations		5,647,094	incl. base outcome			

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Base outcome is *p search*.

outcomes. This implies that users interested in these categories, regardless of their purchasing decisions on this platform, generally have more leisure time or exhibit lower price sensitivity, making them less prone to fall into the *p search* category and more likely to be other types. Additionally, beauty care, fashion accessories, and tv audio categories show significant positive coefficients for the *x search* and *both search* outcomes, indicating a user preference for quality of life or electronics, which translates to a more thorough search for the ideal smartphone. Therefore, users with activities in these three categories are less likely to belong to the *no search* or *p search* types and more inclined towards attribute-based exploration.

Users engaged in the pet goods category tend to fall predominantly into the *no search* group, showing a lesser inclination towards attribute or price comparisons. This trend likely reflects pet owners who, due to their commitments to pet care, might exhibit less sensitivity to prices and have more limited free time for extensive product exploration.

Furthermore, the analysis reveals through its constant terms the differences in shares among user types for those who have no recorded interactions in categories outside of smartphones.

This straightforward regression analysis reveals a significant connection between user types, which are determined exclusively based on their interactions within the smartphone category, and their conduct across a variety of other product domains. This relationship serves to illuminate how a user's general shopping habits and preferences across different categories can profoundly impact their engagement and decision-making processes when it comes to selecting smartphones. This relationship underscores the value of using historical clickstream data to model consumer heterogeneity discussed in later sections.

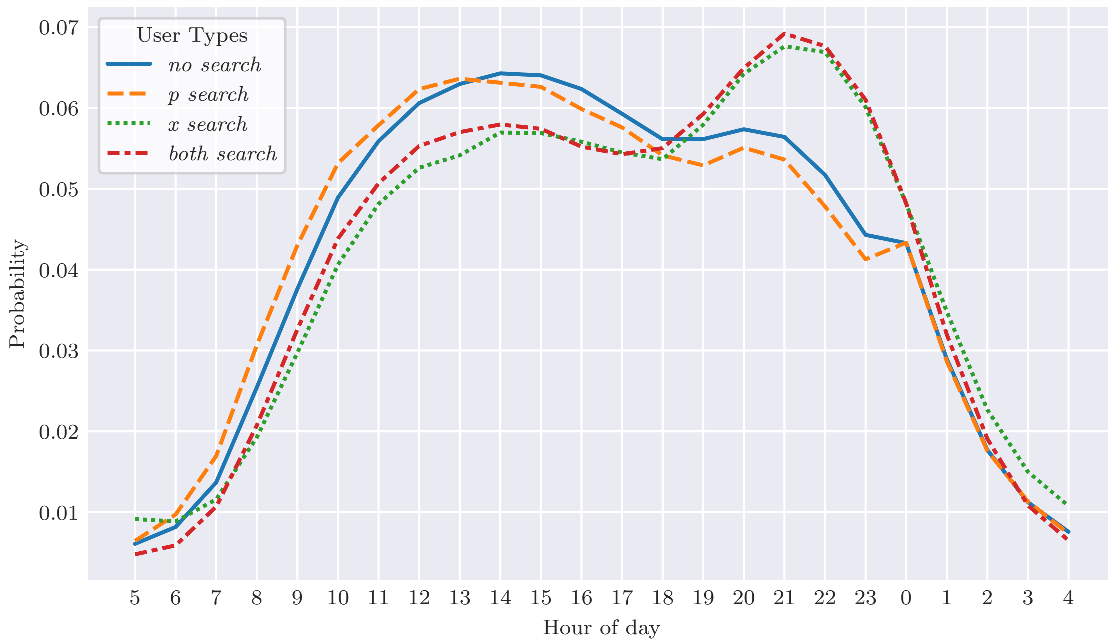
1.4.2 Smartphone Events Distribution by User Types

Exploring the variance among user types can also be achieved by examining their activity distribution throughout the day or week. Different user types may exhibit distinct behaviors at various times for numerous reasons. For instance, users categorized under *x search* and *both search* typically require extended periods to explore a range of smartphone options. Figure 1.3 illustrates the distribution of all smartphone-related activities across the hours of the day, segmented by user type. For clarity, the vertical axis represents the probability mass, and connecting the data points enhances visualization, even though the probability mass function isn't applicable to non-integer values. This visualization approach is similarly applied in Figure 1.4, which maps the activity distribution across the week.

Figure 1.3 reveals some fascinating trends. The activity patterns among different user types are similar from midnight through the early morning but diverge significantly during daylight hours. Each user type exhibits a bimodal distribution pattern, with the *p search* type showing a unique tri-modal distribution. Commonly, all types have an activity peak in the early afternoon (around 1 - 2 p.m.) and another in the evening (around 8 - 9 p.m.), correlating with post-lunch and post-dinner times, respectively. However, *x search* and *both search* users display notably higher peaks in the evening, reflecting their need for substantial time to compare an array of smartphones, unlike other users who only needs to quickly visit and exit the site after viewing a single smartphone. Intriguingly, *p search* users exhibit a third peak at midnight, unlike other types whose activity tends to decline from 11 p.m to midnight. This indicates *p search* users' anticipation of potential price changes with the onset of a new day, even though price adjustments do not necessarily occur at midnight.

Figure 1.4, which illustrates the distribution throughout the week, highlights an interesting crossover where *x search* and *both search* types overtake *no search* and *p search* types on Fridays and weekends. This trend accentuates the increased availability of longer time periods for exploration as the weekend approaches, further highlighting the temporal nuances in user engagement with browsing smartphones.

The identified differences in user activity patterns across times of day and days of the week are pivotal, laying a foundational basis for a more granular analyses of consumer heterogeneity.



Hour of day indicates the “hour” value of the event in 24-hour format, i.e., 5 indicates any event between 5:00:00 a.m. to 5:59:59 a.m. inclusively.

Figure 1.3: Distribution of Smartphone Events Over Hours of Day

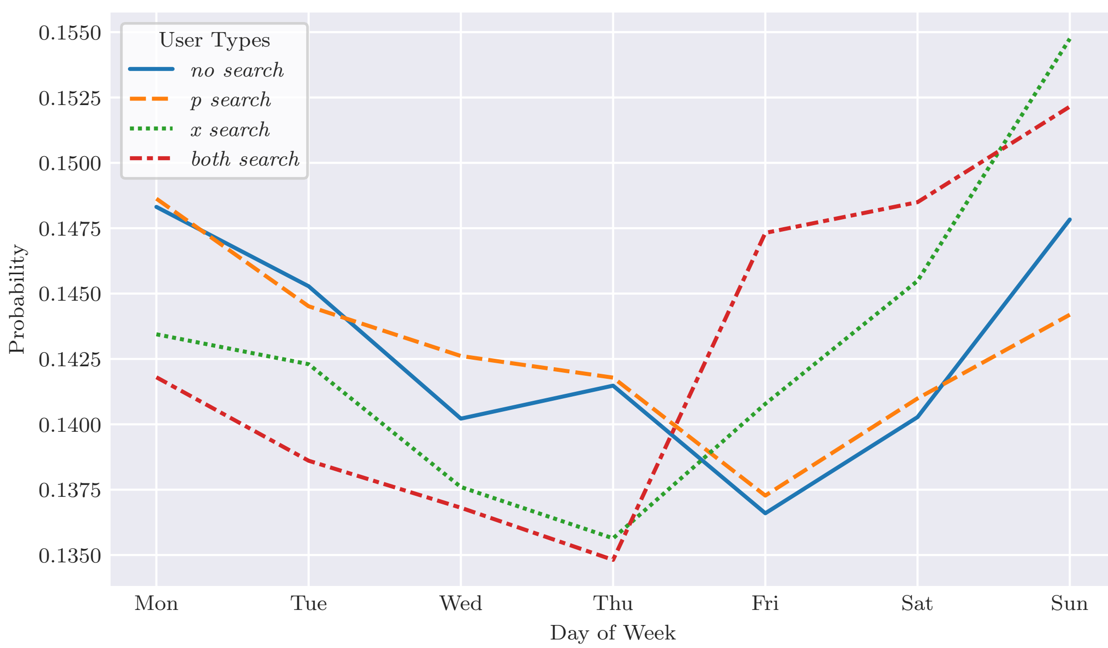


Figure 1.4: Distribution of Smartphone Events Over Days of Week

1.5 Models and Results

This section delves into the conditional logit choice model (McFadden, 1974), to analyze users' smartphone purchasing decisions, alongside four variant models designed to accommodate varying levels of heterogeneity. The baseline model is introduced in 1.5.1, employing the standard conditional logit choice framework. Subsequent models are developed to enhance this foundational model by integrating both parametric and non-parametric methods to account for increasing level of heterogeneity. In Section 1.5.3, the model is expanded to include user types based on observed user behaviors. Section 1.5.5 further advances the model by incorporating unobserved heterogeneity alongside the pre-classified user types in a probabilistic manner. Lastly, Section 1.5.7 explores the application of two machine learning techniques aimed at estimating the logit model with consumer heterogeneity at individual-level, showcasing the model's adaptability to capture the intricate variations in consumer preferences, and establish the foundation for counterfactual analysis of personalized discounts.

In the following, fundamental concepts and notation shared among the models are outlined. Each user, represented as $n \in \mathcal{N}$, confronts a choice set $\mathcal{C}_n = \mathcal{A}_{t(n)} \cup \{0\}$, with $t(n)$ signifying the user's time of activity. This choice set $\mathcal{C}_n$ includes all smartphones available on the platform at time $t(n)$, represented by $\mathcal{A}_{t(n)}$, along with the outside option, denoted as $\{0\}$. The outside option is assumed selected when the user does not finalize a smartphone purchase within the observed timeframe. This choice is interpreted as an indistinct amalgamation of either opting not to buy a smartphone or deciding to purchase from an alternative source. Each user makes exactly one choice from $\mathcal{C}_n$ that maximizes their utility.

To lay the groundwork for a thorough examination and comparison of each model's distinctive features and modifications, I begin with the foundational structure of the utility function, as delineated below. This serves as the conceptual cornerstone for the ensuing discussions on various models within this section:

$$U_{n,i} \equiv u_{n,i} + \epsilon_{n,i} = \alpha \cdot p_{t(n),i} + \boldsymbol{\beta}^{\mathsf{T}} \mathbf{x}_{n,t(n),i} + \epsilon_{n,i}, \quad \forall i \in \mathcal{C}_n \tag{1.1}$$

where the deterministic component $u_{n,i}$ is a linear combination of the price, $p_{t(n),i}$, and a set of attributes represented by the column vector, $\mathbf{x}_{n,t(n),i}$. The price faced by user n for option i is contingent on the user's active time on the platform, hence the subscript $t(n)$. The attribute vector $\mathbf{x}_{n,t(n),i}$ consists of four sets of variables: (1) the user's interaction with choice i, intrinsically tied to n; (2) time-dependent features of choice i, varying with $t(n)$; (3) invariant features of choice i, including brands, product specifications, and bundled gadgets; and (4) a constant intercept.[6] For the outside option, $u_{n,0}$ is normalized to 0. The random component $\epsilon_{n,i}$ follows a i.i.d. standard Extreme Value Type-I distribution across n and i, potentially encapsulating unobserved attributes that could influence the user's utility, such as the tactile experience of the smartphone and its design aesthetics.

Building on the utility function framework, the choice probability for $i \in \mathcal{C}_n$ is hence:

$$P_{n,i} \equiv \text{Prob}\left(U_{n,i} \geq U_{n,j}, \forall j \in \mathcal{C}_n\right) = \frac{\exp(u_{n,i})}{1 + \sum_{j \in \mathcal{A}_{t(n)}} \exp(u_{n,j})}. \tag{1.2}$$

And the log-likelihood for maximum likelihood estimation (MLE) is expressed as:

$$\ell\left(\boldsymbol{\Theta}\right) = \sum_{n \in \mathcal{N}} \sum_{i \in \mathcal{C}_n} y_{n,i} \cdot \log(P_{n,i}) \tag{1.3}$$

where $\boldsymbol{\Theta}$ is the parameter of estimates to be specified later, and $y_{n,i}$ is the indicator variable of user n chooses i.

[6] Refer to Table 1.2 for a detailed list of variables.

As we delve into each model in the subsequent discussions, additional subscripts for $\mathbf{\Theta}$, and variations of $U_{n,i}$, $u_{n,i}$, and $P_{n,i}$ accounting for consumer sub-types will be introduced. Despite these enhancements, it is vital to recognize that the core structure of the utility function, the conditional choice probability, and the log-likelihood framework remain consistent across all model variants. Furthermore, for each of following models, three-quarters of users are randomly allocated for estimation, with the remaining quarter serving as a consistent holdout sample to validate performance.

1.5.1 Model I - Baseline Logit

The vanilla version of the conditional logit model is absent of consumer heterogeneity, serving as a foundational benchmark for the comparison of subsequent, more intricate models. Let $\boldsymbol{\theta}_{\text{logit}} = \left[\alpha_{\text{logit}}, {\boldsymbol{\beta}_{\text{logit}}}^{\mathsf{T}}\right]^{\mathsf{T}}$ denote the vector of coefficients. Consequently, the utility $U_{n,i}$ as from the generic form (1.1) is hence,

$$U_{n,i} \equiv u_{n,i} + \epsilon_{n,i} = \alpha_{\text{logit}} \cdot p_{t(n),i} + {\boldsymbol{\beta}_{\text{logit}}}^{\mathsf{T}} \mathbf{x}_{n,t(n),i} + \epsilon_{n,i}, \quad \forall i \in \mathcal{C}_n. \tag{1.4}$$

The conditional choice probability is as specified in (1.2), and the estimation involves maximizing $\ell\left(\mathbf{\Theta}_{\text{I}}\right)$ as specified in (1.3) where $\mathbf{\Theta}_{\text{I}} = \boldsymbol{\theta}_{\text{logit}}$.

1.5.2 Estimation Result of Model I - Baseline Logit

Table 1.2 presents the MLE estimation results of the vanilla conditional logit models, employing various sets of regressors to examine robustness. The estimated coefficients are predominantly statistically significant. Notably, the coefficient associated with price is negative, indicative of a downward-sloping demand curve. In analyzing user interaction with products, it is observed that an increase in the # of *to page* view clicks positively influences the choice probability, while the # of *on page* view clicks shows no significant effect, underscoring the critical distinction between types of view events as discussed in Section 1.3.2. Furthermore, a user's increased average time spent per view click suggests a higher level of engagement with the product, consequently leading to a higher likelihood of purchase. Conversely, the negative coefficient for interrupted views suggests that interruptions in browsing diminish the likelihood of purchase.

The estimation also reveals that products lacking reviews are generally less favored, starkly contrasts with the positive and significant coefficients on log of # of reviews and star ratings. This contrast emphasizes the importance of peer evaluations on consumer decisions. Moreover, the negative coefficient associated with the log days since available suggests a pronounced preference for newer smartphone models. This finding emphasizes the critical need to account for the age of products in analyses, as older products might accrue more reviews over time, yet this does not inherently equate to a higher consumer preference, thereby highlighting the nuanced relationship between product age, consumer reviews, and their perceived desirability.

The examination of brand effects, focusing on the top five brands[7] which represent a significant portion of the dataset, reveals Apple as having the most substantial brand impact. As for product specifications, except for weight, where a negative coefficient is observed, all other variables positively correlate with preference whenever significant, aligning with expectations.

Smartphones are occasionally bundled with gadgets, with wireless headphones being the most commonly included item. Other gadgets, such as power banks and fitness trackers, are less frequently bundled and are thus aggregated into a single category due to their scarcity. While headphone bundles do not necessarily indicate higher preference, bundles featuring these less common gadgets seem to garner more

[7] The dataset encompasses 43 brands, excluding the outside option. The foremost five brands, ranked by the volume of events, collectively represent 52.85% of all unique smartphone products and account for 93.89% of total smartphone-related activities.

Table 1.2: Estimation Results of Model I - Baseline Logit

	(1)		(2)		(3)	
Price						
Price (\$1,000)	-2.162^{***}	(0.007)	-1.019^{***}	(0.010)	-1.386^{***}	(0.017)
User interaction						
# of *to page* view clicks			1.689^{***}	(0.002)	1.685^{***}	(0.002)
# of *on page* view clicks			0.002	(0.002)	-0.000	(0.002)
Avg uninterrupted view time (min)			1.150^{***}	(0.002)	1.137^{***}	(0.002)
# of interrupted view clicks			-2.145^{***}	(0.005)	-2.148^{***}	(0.005)
Time-dependent features						
No reviews (indicator)			-1.244^{***}	(0.058)	-0.980^{***}	(0.065)
Log # of reviews			0.802^{***}	(0.002)	0.767^{***}	(0.002)
Star rating			0.150^{***}	(0.012)	0.115^{***}	(0.014)
Log days since available			-1.357^{***}	(0.003)	-1.167^{***}	(0.003)
Brands (indicators)						
Samsung	4.499^{***}	(0.010)	1.536^{***}	(0.012)	1.519^{***}	(0.013)
Apple	4.959^{***}	(0.011)	2.408^{***}	(0.015)	3.282^{***}	(0.018)
Xiaomi	2.863^{***}	(0.011)	0.297^{***}	(0.013)	0.191^{***}	(0.014)
Huawei	3.206^{***}	(0.012)	1.169^{***}	(0.013)	1.308^{***}	(0.014)
Oppo	3.499^{***}	(0.013)	1.224^{***}	(0.014)	1.051^{***}	(0.015)
Product specifications						
Log storage (G)					0.015^{**}	(0.005)
Log RAM (G)					-0.001	(0.010)
Weight (100 g)					-0.893^{***}	(0.022)
Diagonal (in)					0.200^{***}	(0.011)
Battery (1,000 mAh)					0.391^{***}	(0.007)
Log highest camera pixels (MP)					0.013^{*}	(0.006)
# of cameras (back)					0.215^{***}	(0.003)
Bundled gadgets (indicators)						
Headphones					0.003	(0.029)
Other gadgets					0.168^{***}	(0.018)
Constant (inside options)	-12.141^{***}	(0.010)	-7.290^{***}	(0.057)	-9.630^{***}	(0.077)
Observations	6,395,207,961		6,395,207,961		6,395,207,961	
Users (Cases)	4,235,321		4,235,321		4,235,321	
LL	−3,569,875.8		−2,230,203.0		−2,221,949.5	
Avg LL Per User	−0.843		−0.527		−0.525	
AIC	7,139,765.5		4,460,436.0		4,443,947.0	
BIC	7,139,858.3		4,460,634.9		4,444,265.2	
Holdout Observations	2,131,744,036		2,131,744,036		2,131,744,036	
Holdout Users (Cases)	1,411,773		1,411,773		1,411,773	
Holdout LL	−1,197,658.6		−745,080.6		−742,270.8	
Holdout Avg LL Per User	−0.848		−0.528		−0.526	

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

interest.

Additionally, the model incorporates a intercept term for all inside options. This inclusion serves two key purposes: it firstly adjusts for potential discrepancies in the measurement units of log-transformed variables by normalizing an additive constant. Secondly, it reflects the broader cost-benefit dynamics of choosing a smartphone over the outside option, which might include considerations like the expense of transferring data to a new device. Due to the compound nature of this constant, its sign is not informative.

Finally, Column (3) in Table 1.2 rejects both Column (1) and (2) by the likelihood ratio test, highlighting the critical need to include all groups of variables for a comprehensive understanding of consumer choices.[8]

1.5.3 Model II - by User Types

The classification of user types outlined in Section 1.3.1 lays the groundwork for the observed heterogeneity among consumers, which is manifested through their smartphone search behaviors. This heterogeneity is not merely incidental but fundamental to our analysis, providing a detailed lens through which consumer behavior can be dissected. Integrating this heterogeneity into the baseline logit model introduces a refined approach that captures the essence of consumer diversity. As aforementioned, each user belongs to exactly one of the following user types:

$$\mathcal{H} = \{\textit{no search}, \textit{p search}, \textit{x search}, \textit{both search}\}. \tag{1.5}$$

Define $h : \mathcal{N} \to \mathcal{H}$ as the mapping from n to their user type, and the vector of coefficients as $\boldsymbol{\theta}_m = \left[\alpha_m, \boldsymbol{\beta}_m{}^{\intercal}\right]^{\intercal}$ $\forall m \in \mathcal{H}$, i.e., $\alpha_{h(n)}$ and $\boldsymbol{\beta}_{h(n)}$ are tailored to specific user types, instead of uniform for all users. Consequently, the utility $U_{n,i}$ as from the

[8] Column (3) is referred to as Model I - Baseline Logit.

generic form (1.1) is hence,

$$U_{n,i} \equiv u_{n,i} + \epsilon_{n,i} = \alpha_{h(n)} \cdot p_{t(n),i} + {\boldsymbol{\beta}_{h(n)}}^{\top} \mathbf{x}_{n,t(n),i} + \epsilon_{n,i}, \quad \forall i \in \mathcal{C}_n. \tag{1.6}$$

The conditional choice probability is as specified in (1.2), and the estimation involves maximizing $\ell\left(\boldsymbol{\Theta}_{\mathrm{II}}\right)$ as specified in (1.3) where $\boldsymbol{\Theta}_{\mathrm{II}} = \left[{\boldsymbol{\theta}_m}^{\top}\right]^{\top}_{m \in \mathcal{H}}$, both with the new $U_{n,i}$ and $u_{n,i}$ in (1.6).

1.5.4 Estimation Result of Model II - by User Types

Incorporating user-type-specific preference parameters effectively tailors the baseline model to distinct segments within the user population, as if conducting separate vanilla logit models for each user type identified in $\mathcal{H}$. The estimation result in Table 1.3 demonstrates significant variations in consumer preferences. This variation not only corroborates the user type classification framework established in Section 1.3.1, but also highlights the presence of significant heterogeneity among users.

Particularly noteworthy is the differential price coefficients among the user types. Users categorized under *p search* and *both search* exhibit more pronounced negative coefficients when compared to *no search* and *x search*, indicating a higher price sensitivity. This correlation between the way consumers search for smartphones and their price responsiveness offers a key insight into the study, suggesting that search behaviors are reflective of underlying consumer preferences.

Table 1.3 shows that the coefficient for the # of *to page* view clicks is positive across all user types, implying that direct visits to a product page generally enhance the likelihood of making a purchase. In contrast, the impact of the # of *on page* view clicks and the average uninterrupted view time vary significantly. Notably, for *no search* users, both coefficients are negative and statistically significant, suggesting that such interactions may reflect frustration or hesitancy rather than intent to buy,

Table 1.3: Estimation Result of Model II - by User Types

	no search		*p search*		*x search*		*both search*	
Price								
Price ($1,000)	-1.211^{***}	(0.035)	-1.690^{***}	(0.141)	-1.267^{***}	(0.035)	-1.713^{***}	(0.027)
User interaction								
# of *to page* view clicks	9.352^{***}	(0.012)	2.968^{***}	(0.018)	2.546^{***}	(0.005)	1.073^{***}	(0.002)
# of *on page* view clicks	-0.024^{**}	(0.009)	0.069^{***}	(0.015)	0.137^{***}	(0.005)	0.082^{***}	(0.003)
Avg uninterrupted view time (min)	-0.293^{***}	(0.005)	2.657^{***}	(0.035)	0.809^{***}	(0.004)	1.162^{***}	(0.003)
# of interrupted view clicks	-15.394^{***}	(0.054)	-2.646^{***}	(0.021)	-2.496^{***}	(0.017)	-1.017^{***}	(0.004)
Time-dependent features								
No reviews (indicator)	-1.316^{***}	(0.113)	-0.834	(0.566)	-0.689^{***}	(0.115)	-1.533^{***}	(0.084)
Log # of reviews	0.435^{***}	(0.004)	0.742^{***}	(0.016)	0.515^{***}	(0.004)	0.566^{***}	(0.003)
Star rating	0.002	(0.023)	0.136	(0.118)	0.283^{***}	(0.024)	0.039^{*}	(0.017)
Log days since available	-0.771^{***}	(0.008)	-1.189^{***}	(0.029)	-0.991^{***}	(0.007)	-1.057^{***}	(0.005)
Brands (indicators)								
Samsung	1.625^{***}	(0.029)	1.500^{***}	(0.111)	1.286^{***}	(0.023)	1.321^{***}	(0.018)
Apple	2.908^{***}	(0.038)	3.344^{***}	(0.147)	2.805^{***}	(0.034)	3.351^{***}	(0.027)
Xiaomi	0.671^{***}	(0.031)	0.333^{**}	(0.118)	0.264^{***}	(0.025)	0.446^{***}	(0.019)
Huawei	1.651^{***}	(0.031)	1.301^{***}	(0.119)	1.278^{***}	(0.025)	1.274^{***}	(0.020)
Oppo	1.642^{***}	(0.033)	1.357^{***}	(0.128)	1.031^{***}	(0.028)	1.001^{***}	(0.022)
Product specifications								
Log storage (G)	-0.002	(0.011)	-0.007	(0.044)	0.007	(0.011)	0.018^{*}	(0.008)
Log RAM (G)	-0.006	(0.022)	0.167^{*}	(0.085)	0.006	(0.021)	-0.004	(0.016)
Weight (100 g)	-0.598^{***}	(0.047)	-0.555^{**}	(0.184)	-1.169^{***}	(0.046)	-1.205^{***}	(0.036)
Diagonal (in)	0.000	(0.024)	0.064	(0.093)	0.054^{*}	(0.022)	0.458^{***}	(0.018)
Battery (1,000 mAh)	0.247^{***}	(0.015)	0.182^{**}	(0.059)	0.345^{***}	(0.014)	0.333^{***}	(0.010)
Log highest camera pixels (MP)	0.046^{**}	(0.015)	0.061	(0.056)	-0.003	(0.013)	0.020^{*}	(0.010)
# of cameras (back)	0.157^{***}	(0.007)	0.187^{***}	(0.028)	0.220^{***}	(0.007)	0.210^{***}	(0.005)
Bundled gadgets (indicators)								
Headphones	-0.045	(0.064)	0.245	(0.240)	-0.036	(0.057)	0.120^{**}	(0.043)
Other gadgets	0.278^{***}	(0.034)	-0.125	(0.161)	0.233^{***}	(0.036)	0.059^{*}	(0.029)
Constant (inside options)	-9.996^{***}	(0.145)	-10.058^{***}	(0.659)	-8.597^{***}	(0.142)	-9.268^{***}	(0.108)
Observations	3,085,424,936		332,669,512		1,296,969,925		1,680,143,588	
Users (Cases)	2,040,986		217,930		864,294		1,112,111	
LL	$-1{,}712{,}007.5$							
Avg LL Per User	-0.404							
AIC	3,424,207.0							
BIC	3,425,479.9							
Holdout Observations	1,029,616,116		111,439,128		431,092,034		559,596,758	
Holdout Users (Cases)	681,099		72,982		287,253		370,439	
Holdout LL	$-570{,}629.7$							
Holdout Avg LL Per User	-0.404							

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

given their nature of not actively seeking information. In contrast, for other user categories, these coefficients are positive, suggesting that greater engagement with page content is associated with an increased probability of purchase.

The findings remain consistent with the baseline model regarding time-dependent features. However, the significance of star ratings diverges across user types. Unlike the first two user types, the latter two display positive and significant coefficients for star ratings, indicating their reliance on such evaluations in their purchasing decisions which aligns with their attribute-searching characteristics.

The coefficients for brands and product specifications exhibit the anticipated signs when they are statistically significant, and reveals variances in magnitudes across different user types. This highlights considerable heterogeneity in consumer preferences. Additionally, bundled gadgets do not yield significant coefficients for *p search* users, indicating that these users prioritize finding acceptable prices, and product bundling does not notably influence their purchase decisions.

By accounting for the observed heterogeneity via user types, this model achieves a markedly improved log-likelihood and decisively rejects Model I in the likelihood ratio test, despite incorporating four times the number of parameters. This underscores the deep-seated heterogeneity across user types, highlighting the strong correlation between the heuristic-based classification of users and their preferences. It emphasizes the importance of integrating this observed heterogeneity into demand modeling to capture the intricate consumer choice decisions more accurately.

1.5.5 Model III - with Unobserved Heterogeneity

Recognizing the inherent differences among consumers classified under the same user type is crucial, as there likely exists unobserved heterogeneity within each category. To address these nuances not captured by the user types classification, this model introduces an element of uncertainty into consumer preferences, adopting a probabilistic approach. In an effort to reduce the distributional assumptions concerning this unobserved heterogeneity, the model draws upon the methodology proposed by Heckman and Singer (1984), which posits a discrete distribution of preferences within each user type in $\mathcal{H}$ index by the set $\mathcal{S} = \{1, 2, \dots\}$. Specifically, a user n belonging to a user type has a random draw of preferences from a discrete distribution of sub-types, indexed by $s \in \mathcal{S}$, and assume this random draw from $\mathcal{S}$ is independent of $\epsilon_{n,i}$. For the purposes of this analysis, the study delineates two discrete sub-types for each of the four user types in $\mathcal{H}$, thus setting $\mathcal{S} = \{1, 2\}$.[9]

In addition to the coefficients in utility functions, it is necessary to introduce a new notation, γ, to delineate the probability distribution of sub-types within each user category. To ensure that the probabilities are appropriately bounded between 0 and 1, the probability of user n belonging to sub-type 1 within their user type $h(n) \in \mathcal{H}$ is defined as

$$\pi_{n,1} = sigmoid\left(\gamma_{h(n)}\right) = \frac{1}{1 + \exp(-\gamma_{h(n)})}. \tag{1.7}$$

And hence, $\pi_{n,2} = 1 - \pi_{n,1}$. Without loss of generality, sub-type 1 is assumed to be the sub-type with lesser price sensitivity among the two. This leads to the new definition of the parameter vector $\boldsymbol{\theta}_m = \left[\alpha_{m,s}, \boldsymbol{\beta}_{m,s}{}^{\mathsf{T}}, \gamma_m\right]_{s \in \mathcal{S}}^{\mathsf{T}}$ $\forall m \in \mathcal{H}$. Consequently, the utility $U_{n,s,i}$ as from the generic form (1.1) with the additional subscript $s \in \mathcal{S}$ is hence,

$$U_{n,s,i} \equiv u_{n,s,i} + \epsilon_{n,i} = \alpha_{h(n),s} \cdot p_{t(n),i} + \boldsymbol{\beta}_{h(n),s}{}^{\mathsf{T}} \mathbf{x}_{n,t(n),i} + \epsilon_{n,i}, \quad \forall i \in \mathcal{C}_n \tag{1.8}$$

where $u_{n,s,i}$ is normalized to 0 for all $s \in \mathcal{S}$. And the probability of user n choosing choice i is a weighted average of the probabilities conditional on s,

$$P_{n,i} = \sum_{s \in \mathcal{S}} \pi_{n,s} \cdot \frac{\exp(u_{n,s,i})}{1 + \sum_{j \in \mathcal{A}_{t(n)}} \exp(u_{n,s,j})} \tag{1.9}$$

[9] The choice to assign an equal number of sub-types across user types is based on analytical preferences rather than a requirement.

The estimation, hence, involves maximizing the log-likelihood function $\ell\left(\boldsymbol{\Theta}_{\mathrm{III}}\right)$ as specified in (1.3) where $\boldsymbol{\Theta}_{\mathrm{III}} = \left[\boldsymbol{\theta}_m{}^{\mathsf{T}}\right]_{m\in\mathcal{H}}^{\mathsf{T}}$ as defined by (1.3), but with the updated $\boldsymbol{\theta}_m$ above and probability $P_{n,i}$ in (1.9).

1.5.6 Estimation Result of Model III - with Unobserved Heterogeneity

Table 1.4 presents the estimation results, wherein each user type in $\mathcal{H}$ is composed of two sub-types, following a probabilistic approach as detailed in Section 1.5.5. This approach, which incorporates both observed and unobserved heterogeneities, reveals additional layers of variability among users within each user type, extending beyond Model II, which accounted only for observed heterogeneity.

As previously stated, sub-type 1 is assumed to be the less price sensitive sub-type within each user types without loss of generality. Consistent with the correlation identified in Section 1.5.4, it is insightful to observe the price coefficient being more negative for *p search* and *both search* compared to *no search* and *x search* respectively by sub-types, indicative of higher price sensitivity, except for sub-type 1 of *p search* when compared to *no search*. However, this exception does not undermine the insight as this trend is further substantiated by the parameter γ, which governs the distribution of probabilities between sub-types 1 and 2. The γ values associated *p search* and *both search* users are also more negative, relative to *no search* and *x search* respectively, suggesting a higher likelihood of encountering the more price-sensitive sub-type. This nuanced differentiation underscores the conclusion drawn in Section 1.5.4 affirming that search behaviors reflect the fundamental preferences of consumers, further accentuates the significance and efficacy of the user type classification in delineating consumer heterogeneity.

For all sub-types across the four user types, both the # of *to page* and *on page* view

Table 1.4: Estimation Result of Model III - with Unobserved Heterogeneity

	no search				*p search*				*x search*				*both search*			
	Sub-type 1		Sub-type 2		Sub-type 1		Sub-type 2		Sub-type 1		Sub-type 2		Sub-type 1		Sub-type 2	
Price																
Price ($1,000)	−1.057***	(0.056)	−1.516***	(0.150)	−1.045***	(0.219)	−1.869*	(0.843)	−1.071***	(0.053)	−1.503***	(0.082)	−1.361***	(0.058)	−1.920***	(0.050)
User interaction																
# of *to page* view clicks	16.907***	(0.067)	1.441***	(0.025)	11.634***	(0.125)	0.461***	(0.073)	10.808***	(0.057)	1.608***	(0.008)	11.148***	(0.071)	0.688***	(0.003)
# of *on page* view clicks	0.046***	(0.012)	0.186***	(0.029)	0.200***	(0.029)	0.090***	(0.026)	0.123***	(0.010)	0.179***	(0.011)	0.211***	(0.008)	0.150***	(0.004)
Avg uninterrupted view time (min)	−1.255***	(0.014)	1.765***	(0.026)	0.519***	(0.066)	2.706***	(0.109)	−0.674***	(0.013)	1.094***	(0.007)	−0.684***	(0.013)	1.255***	(0.004)
# of interrupted view clicks	−20.694***	(0.079)	0.076	(0.055)	−11.955***	(0.142)	−0.226**	(0.078)	−10.880***	(0.065)	−0.610***	(0.028)	−11.073***	(0.073)	−0.759***	(0.007)
Time-dependent features																
No reviews (indicator)	−1.257***	(0.215)	−2.135**	(0.650)	−0.785	(0.691)	−3.001	(2.996)	−1.108***	(0.138)	−3.191***	(0.194)	−0.479**	(0.172)	−1.953***	(0.144)
Log # of reviews	−0.112***	(0.008)	1.077***	(0.017)	0.165***	(0.027)	0.978***	(0.077)	−0.016*	(0.006)	0.660***	(0.008)	0.039***	(0.007)	0.568***	(0.005)
Star rating	−0.013	(0.045)	−0.003	(0.135)	0.055	(0.144)	−0.004	(0.582)	−0.030	(0.029)	−0.032	(0.039)	0.048	(0.036)	0.040	(0.030)
Log days since available	−0.295***	(0.014)	−1.513***	(0.029)	−0.652***	(0.050)	−1.445***	(0.143)	−0.312***	(0.012)	−1.384***	(0.015)	−0.399***	(0.014)	−1.159***	(0.009)
Brands (indicators)																
Samsung	0.825***	(0.051)	1.731***	(0.151)	1.055***	(0.168)	1.762**	(0.627)	0.797***	(0.042)	1.252***	(0.050)	0.715***	(0.046)	1.639***	(0.034)
Apple	1.223***	(0.065)	3.797***	(0.182)	2.136***	(0.222)	3.919***	(0.820)	1.619***	(0.056)	3.208***	(0.075)	1.933***	(0.063)	3.693***	(0.051)
Xiaomi	0.653***	(0.057)	0.256	(0.158)	0.593***	(0.178)	0.336	(0.681)	0.599***	(0.046)	0.351***	(0.053)	0.425***	(0.049)	0.712***	(0.035)
Huawei	0.958***	(0.058)	1.301***	(0.159)	1.119***	(0.182)	0.728	(0.714)	0.983***	(0.045)	1.189***	(0.053)	0.730***	(0.049)	1.639***	(0.036)
Oppo	1.394***	(0.066)	1.184***	(0.166)	1.278***	(0.208)	1.109	(0.745)	0.986***	(0.052)	1.084***	(0.058)	0.861***	(0.055)	1.158***	(0.039)
Product specifications																
Log storage (G)	0.005	(0.021)	−0.028	(0.043)	0.018	(0.076)	0.002	(0.216)	−0.000	(0.018)	−0.038	(0.024)	0.035	(0.020)	0.007	(0.015)
Log RAM (G)	−0.015	(0.040)	−0.001	(0.083)	0.042	(0.143)	0.024	(0.407)	0.005	(0.034)	0.005	(0.045)	0.004	(0.037)	−0.007	(0.029)
Weight (100 g)	−1.299***	(0.081)	−1.199***	(0.191)	−1.027***	(0.308)	−1.523	(0.906)	−1.491***	(0.074)	−2.788***	(0.111)	−1.062***	(0.081)	−1.305***	(0.064)
Diagonal (in)	−0.015	(0.042)	0.078	(0.096)	0.008	(0.153)	0.116	(0.440)	−0.015	(0.038)	0.003	(0.049)	0.009	(0.042)	0.342***	(0.031)
Battery (1,000 mAh)	0.107***	(0.029)	0.413***	(0.061)	0.205*	(0.094)	0.236	(0.304)	0.245***	(0.023)	0.561***	(0.030)	0.313***	(0.024)	0.479***	(0.017)
Log highest camera pixels (MP)	−0.002	(0.026)	0.003	(0.057)	−0.006	(0.095)	0.002	(0.296)	−0.036	(0.022)	−0.018	(0.029)	0.009	(0.023)	0.040*	(0.017)
# of cameras (back)	0.007	(0.013)	0.325***	(0.028)	−0.022	(0.048)	0.237	(0.143)	0.121***	(0.011)	0.250***	(0.015)	0.117***	(0.012)	0.245***	(0.009)
Bundled gadgets (indicators)																
Headphones	−0.612***	(0.098)	0.693**	(0.255)	0.350	(0.348)	−1.298	(3.632)	−0.138	(0.094)	−0.133	(0.146)	0.072	(0.097)	0.062	(0.075)
Other gadgets	−0.643***	(0.055)	0.371*	(0.160)	−0.802***	(0.235)	−0.155	(1.264)	0.029	(0.051)	0.224**	(0.085)	−0.019	(0.056)	0.017	(0.051)
Constant (inside options)	−11.530***	(0.278)	−9.822***	(0.728)	−9.994***	(0.897)	−9.831**	(3.211)	−8.729***	(0.209)	−4.308***	(0.259)	−9.940***	(0.245)	−9.586***	(0.189)
γ (inv-sigmoid of sub-type 1 prob)	0.783***	(0.008)			−0.003	(0.020)			−1.633***	(0.007)			−1.852***	(0.005)		
Observations	3,085,424,936				332,669,512				1,296,969,925				1,680,143,588			
Users (Cases)	2,040,986				217,930				864,294				1,112,111			
LL	−1,274,381.4															
Avg LL Per User	−0.301															
AIC	2,548,954.8															
BIC	2,550,227.6															
Holdout Observations	1,029,616,116				111,439,128				431,092,034				559,596,758			
Holdout Users (Cases)	681,099				72,982				287,253				370,439			
Holdout LL	−425,437.3															
Holdout Avg LL Per User	−0.301															

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

clicks demonstrate positive coefficients, indicating a consistent pattern of engagement leading to increased likelihood of purchase. Notably, the coefficients of the # of *to page* view clicks exhibit significant variation across sub-types within each user category, underscoring the critical importance of accounting for unobserved heterogeneity in the analysis. However, the average uninterrupted view time presents a subtle picture: in addition to the negative coefficient for sub-type 1 of *no search* users, the coefficients turn negative for sub-type 1 of both *x search* and *both search* users, despite the relatively low probabilities associated with sub-type 1 within these user type. This unexpected finding sheds light on the shortcomings of the uniform user types classification rule, which is intended to identify consumer heterogeneity, being itself homogeneously applied to all users, resulting in an imperfect classification. This precisely motivates the relaxed dependence of this rule-base classification result in later models. This insight motivates more flexible approaches to capture the approaches towards observed heterogeneity in later models to reduce the reliance on this one-size-fits-all rule, allowing for a better accommodation to the diverse behaviors and preferences within the consumer base.

Regarding time-dependent features, the coefficients generally align with their anticipated signs when significant. However, it's important to note an exception: the coefficient of log # of reviews becomes negative for sub-type 1 of the *no search* users. This could be indicative of frustration from an overload of reviews cluttering the webpage. A similar pattern is observed for sub-type 1 of the *x search* users further underscores the potential shortcomings of the rule-based classification approach previously mentioned.

The coefficients for brands and product specifications retain their qualitative significance, albeit with notable variance in magnitude. Interestingly, the coefficients for bundled gadgets shift to negative and are significant for sub-type 1 of *no search* and *p search* users, diverging from the findings of the previous model. This change implies that for certain consumer segments, bundling a smartphone with additional items might detract from its appeal, even when price and other factors are held constant. This phenomenon could stem from some consumers having a clear idea of their intended purchase, and the presence of a bundle could potentially complicate their decision-making process and eventually fail to convert. They might perceive the bundle as less desirable, possibly under the assumption that the smartphone alone could be acquired at a lower price. This intriguing observation underscores the critical need to account for unobserved heterogeneity in the analysis to prevent skewed interpretations.

The inclusion of unobserved heterogeneity significantly enhances the model's fit, conclusively outperforming all previous models in the likelihood ratio test. Although heuristic-based classification of user types as a form of observed heterogeneity may have its drawbacks, the findings reiterate a clear relationship between user types and their preference parameters. Moreover, the results reveal substantial heterogeneity within each user type that remains unaccounted for. This revelation advocates for the development and use of more sophisticated and flexible models, aimed at capturing the broad spectrum of consumer preferences with greater accuracy.

1.5.7 Models with Individual-level Heterogeneity

Models II and III introduce a degree of consumer heterogeneity but are constrained by their reliance on a heuristic classification of user types, which inherently simplifies the rich tapestry of consumer behavior. Moreover, these models overlook a crucial aspect of user diversity: their clickstreams. While the classification of user types offers valuable insights into consumer heterogeneity, this section proposes a shift away from an over-reliance on this broad and oversimplified categorization of users towards harnessing the vast variations in consumers' historical clickstream data to capture observed heterogeneity more precisely. This approach taps into a wealth of information in the data that Models II and III overlook, offering a more sophisticated approach to consumer heterogeneity at an individual level. As evidenced in Section 1.4.1, there exists a significant correlation between users' activities outside the smartphone category and their behavior during smartphone searches, which in turn reflects their preferences. Hence, this approach mitigates the limitation posed by the absence of demographic data for each user by capitalizing on the rich insights provided by detailed historical browsing behavior. Building on the ideas of Farrell et al. (2021), this section introduces two models employing machine learning algorithms to facilitate the estimation of individual-level preference parameters, marking a significant advancement in capturing the intricacies of consumer behavior.

The extension of the conditional logit framework to accommodate individual-level heterogeneity mirrors the approach outlined in Model III from Section 1.5.5. However, the model parameters, previously denoted as $\alpha_{m,s}, \boldsymbol{\beta}_{m,s}$, and γ_m for $m \in \mathcal{H}$ and $s \in \mathcal{S}$ are now tailored to each individual, represented as $\alpha_{n,s}, \boldsymbol{\beta}_{n,s}$, and γ_n for $n \in \mathcal{N}$ and $s \in \mathcal{S}$. With this change, $\boldsymbol{\theta}_n = \left[\alpha_{n,s}, {\boldsymbol{\beta}_{n,s}}^{\mathsf{T}}, \gamma_n\right]^{\mathsf{T}}_{s\in\mathcal{S}}$ presents a formidable challenge due to the vast dimensionality involved, rendering direct estimation infeasible. To navigate this complexity, I draw inspiration from the innovative approach proposed by Farrell et al. (2021), which utilizes machine learning algorithms as highly flexible and non-parametric functions of individual-specific inputs, generically expressed as

$$G(\mathbf{z}_n; \boldsymbol{\Theta}) = \boldsymbol{\theta}_n, \tag{1.10}$$

where $\mathbf{z}_n$ denotes the generic individual-specific inputs and $\boldsymbol{\Theta}$ is the set of parameters to be estimated. While machine learning models are predominantly recognized for their predictive capabilities, Farrell et al. (2021) emphasize that $\boldsymbol{\theta}_n$ are not predictions, they are parameter with economic meanings. For a more detailed elaboration on the interpretation of $\boldsymbol{\theta}_n$, readers are encouraged to consult the work of Farrell et al. (2021).

1.5.7.1 Model IV - MLP

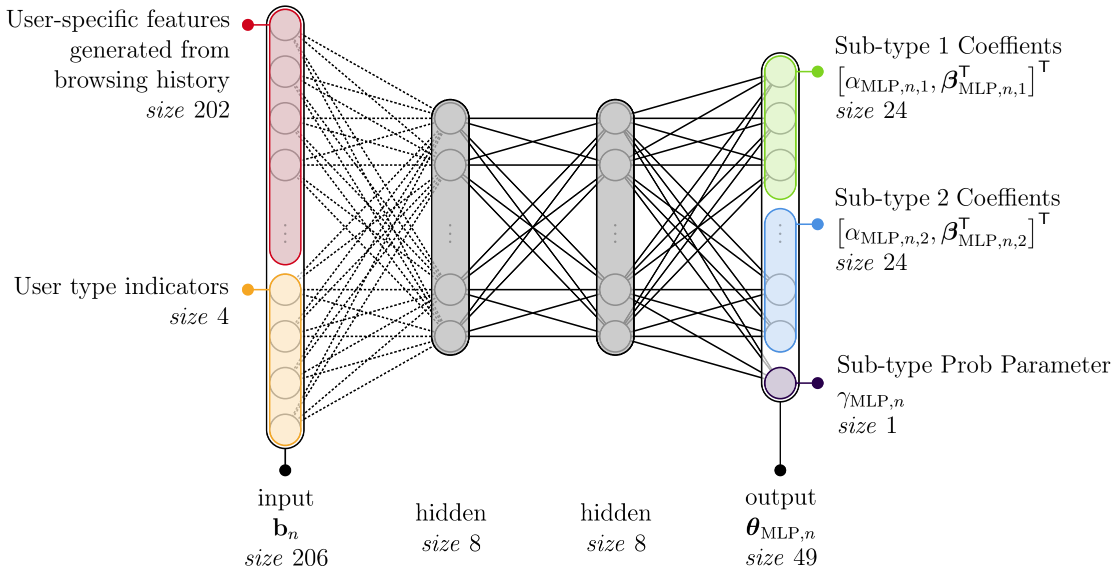
The first model with individual level heterogeneity employs a multi-layer perceptron (MLP) on a rich array of user characteristics gleaned from their clickstream data. For each user n, a vector of user-specific features, denoted as $\mathbf{b}_n$, is constructed and serves as the input for the MLP. This vector encapsulates a wide range of user-specific attributes generated from user's historical clickstream up to and including their first smartphone click, including the total number of clicks, purchases, and expenditure across various product categories (level 1 categories), the temporal distribution of clicks (spanning hours of the day, days of the week, and months), the price, brand, event type, and time indicators of the first-click smartphone, along with indicators of observed user types as supplementary information. Consequently, $\mathbf{b}_n$ has a size of 206, establishing the MLP function as $G_{\text{MLP}} : \mathbb{R}^{206} \to \mathbb{R}^{49}$, and (1.10) is effectively

$$G_{\text{MLP}}(\mathbf{b}_n; \boldsymbol{\Theta}_{\text{IV}}) = \boldsymbol{\theta}_{\text{MLP},n} \tag{1.11}$$

where $\boldsymbol{\theta}_{\text{MLP},n} = \left[\alpha_{\text{MLP},n,s}, {\boldsymbol{\beta}_{\text{MLP},n,s}}^{\intercal}, \gamma_{\text{MLP},n}\right]^{\intercal}_{s\in\mathcal{S}}$ has a size of 49.

To constrain the parameter set $\boldsymbol{\Theta}_{\text{IV}}$ and mitigate the risk of overfitting, the MLP model is architecturally configured with two hidden layers, each comprising 8 nodes. These layers employ the Rectified Linear Unit (ReLU) as their activation function, which has been proven effective in the machine learning literature, ensuring non-linear transformations within the network (Nair and Hinton, 2010). As a result, $\boldsymbol{\Theta}_{\text{IV}}$ contains 2,161 scalar parameters. The structural design of the MLP is depicted in Figure 1.5.

Consequently, the utility $U_{n,s,i}$ as from the generic form (1.1) with the additional



Dashed lines indicate that the layer has no bias due to multicollinearity among the user-type indicators.

Figure 1.5: Model IV - MLP Structure

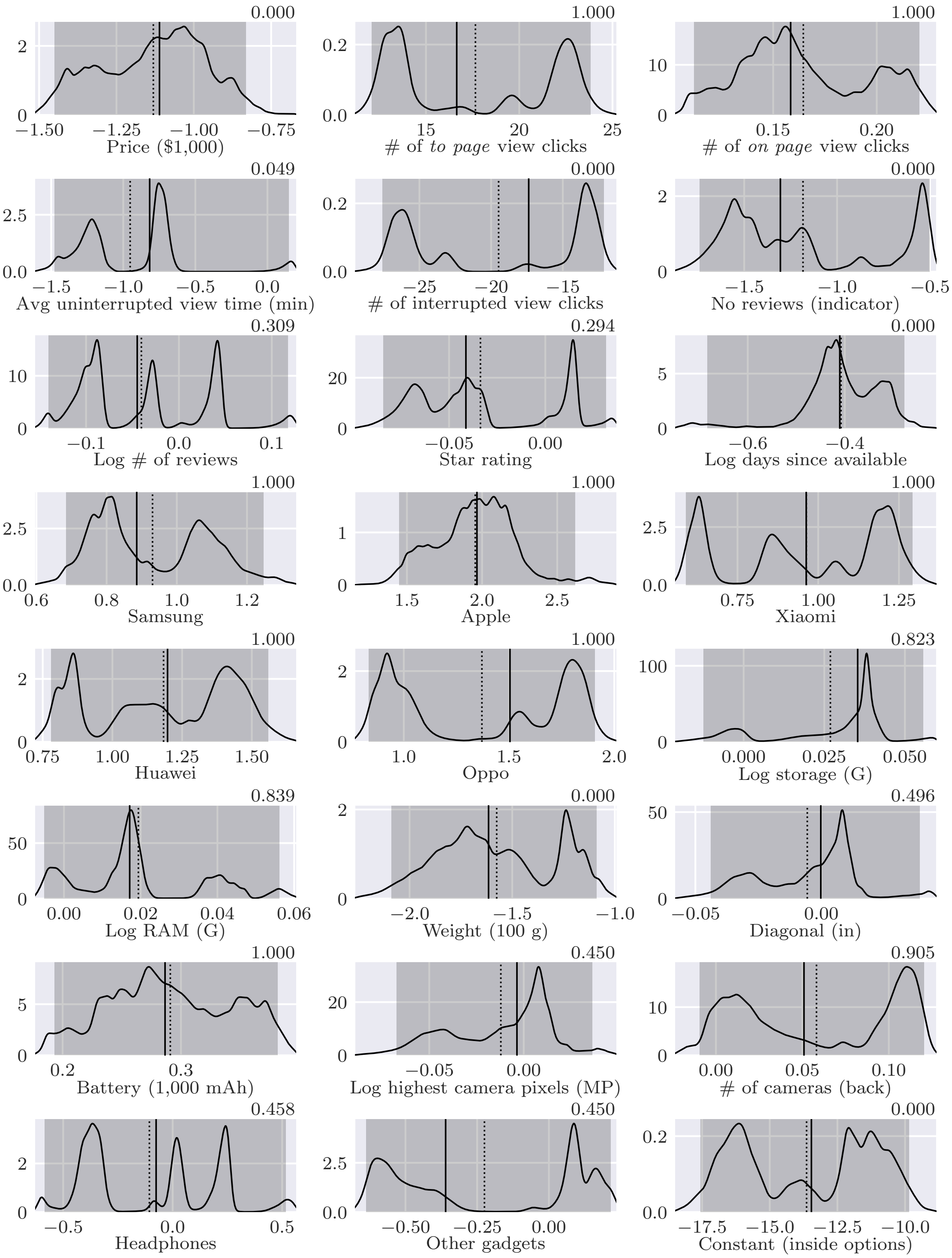
subscript $s \in \mathcal{S}$ is hence,

$$U_{n,s,i} \equiv u_{n,s,i} + \epsilon_{n,i} = \alpha_{\text{MLP},n,s} \cdot p_{t(n),i} + {\boldsymbol{\beta}_{\text{MLP},n,s}}^{\mathsf{T}} \mathbf{x}_{n,t(n),i} + \epsilon_{n,i}, \quad \forall i \in \mathcal{C}_n. \tag{1.12}$$

The conditional choice probability mirrors that presented in (1.9) incorporating the updated specifications of $U_{n,s,i}$ and $u_{n,s,i}$ as outlined in (1.12). The log-likelihood function is $\ell\left(\boldsymbol{\Theta}_{\text{IV}}\right)$ articulated in accordance with equation (1.3), adapting to this updated probability framework.

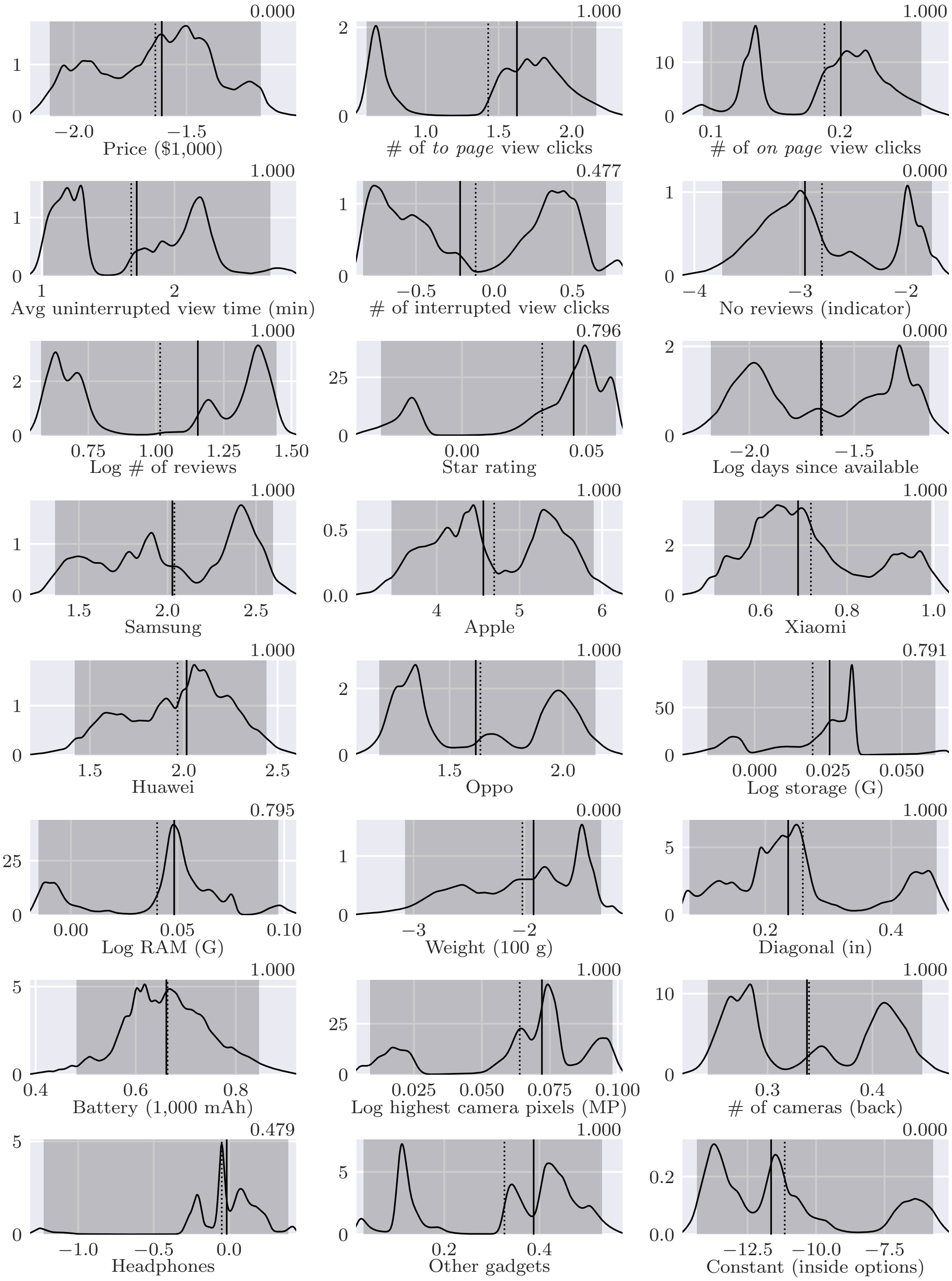
1.5.7.2 Estimation Result of Model IV - MLP

Figure 1.6 delineates the distributions of MLP output coefficients, $\boldsymbol{\theta}_{\text{MLP},n}$, for $n \in \mathcal{N}$, with sub-figures (a), (b), and (c) illustrating the coefficients for sub-type 1, sub-type 2, and γ_{MLP}, respectively.

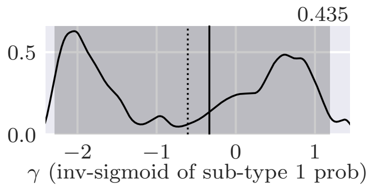


(a) Distributions of Coefficients (Sub-type 1)

Each plot in sub-figures (a), (b), and (c) displays the central 99.5% density of each distribution, with the median and mean indicated by solid and dashed lines respectively, central 95% densities indicated by the shaded region, and the probability of > 0 indicated on the top right corner.



(b) Distributions of Coefficients (Sub-type 2)



Estimation			
Observations	: 6,395,207,961	Users (Cases)	: 4,235,321
LL	: −1,205,676.6	Avg LL Per User	: −0.285
AIC	: 2,415,675.2	BIC	: 2,444,327.9
Holdout			
Observations	: 2,131,744,036	Users (Cases)	: 1,411,773
LL	: −403,102.1	Avg LL Per User	: −0.286

(c) Distribution of γ and Estimation Statistics

Figure 1.6: Estimation Result of Model IV - MLP

It becomes immediately apparent that incorporating a MLP to model consumer heterogeneity at the individual level significantly enhances the heterogeneity of consumer preferences. These results, while bearing qualitative similarities to Model III, demonstrate a more pronounced variation across user for every coefficient. This non-parametric method, utilizing the flexibility of machine learning models, enables arbitrary variations with no distributional assumptions.

The estimation reveals some new insights previously unexplored. Notably, the coefficient corresponding to the # of interrupted view clicks is consistently negative across all users' sub-type 1, suggesting interruptions as indications of aversion. However, this pattern shifts for a significant share of users' sub-type 2, where the coefficient turns positive, hinting that prolonged breaks may not diminish the likelihood of a purchase for certain users. While the majority of users exhibit positive coefficients for log storage and log RAM, indicating a general preference for higher specifications, a minority appears less influenced by these attributes. Furthermore, the more price-sensitive user sub-type, sub-type 2, displays a significantly higher probability of positive coefficients for both screen diagonal size and log highest camera pixels, underscoring the significant of incorporating unobserved heterogeneity.

Regarding coefficients on price, Figure 1.7 and 1.8 offer a decomposition of densities by user types for the two sub-types respectively, acknowledging compelling patterns despite the acknowledged imperfections in user type classification. Specifically,

Figure 1.7 illustrates that the distribution for the sub-type 1 of consumers predominantly skews towards more negative values for the *p search* users, compared to other three user types. Furthermore, Figure 1.8 highlights a similar trend for the sub-type 2 of consumers, with a marked inclination towards negative values among both the *p search* and *both search* users. Supplementary Figure A.1 and A.2, delineating the same distributions broken down by active months, in contrast, do not present a clear segregation pattern. As shown in Figure 1.9, the parameter $\gamma_{\mathrm{MLP},n}$ is also distributed toward more negative values for *p search* and *both search* users compared to *no search* and *x search* users respectively, indicating higher probabilities of being the more price-sensitive sub-type, sub-type 2, in general. These findings not only echo the insights gleaned from Model III but also refine them by providing individual-level estimates instead of aggregate point estimates. The discernible link between user types, particularly those engaged in price searches, and their heightened price sensitivity, emerges as a fascinating aspect of this research, offering deeper understanding into consumer behaviors and preferences.

Model IV significantly enhances our ability to identify nuanced heterogeneity across a broad spectrum of consumers. By strategically constraining the parameter set $\boldsymbol{\Theta}_{\mathrm{IV}}$, Model IV can be transformed into each of Models I through III, and likelihood ratio tests firmly reject the preceding models in favor of Model IV, despite it encompassing over 2,000 scalar parameters. This method underscores the limitations inherent in oversimplified categorizations, such as user types, for capturing observed heterogeneity, and highlights the potential of machine learning models in economic analysis.

However, it's crucial to acknowledge that applying the MLP approach involves critical human decisions in the creation of user-specific features $\mathbf{b}_n$, which can significantly affect model performance. This process, which translates a user's historical

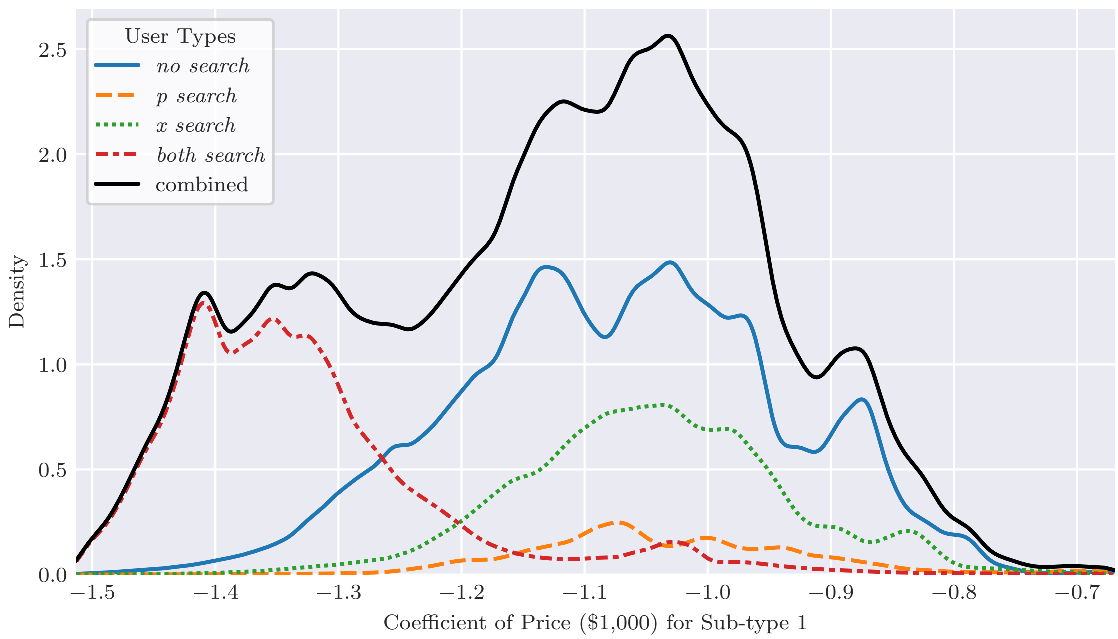


Figure 1.7: Distribution of Price Coefficient of Model IV - MLP (Sub-type 1, by User Types)

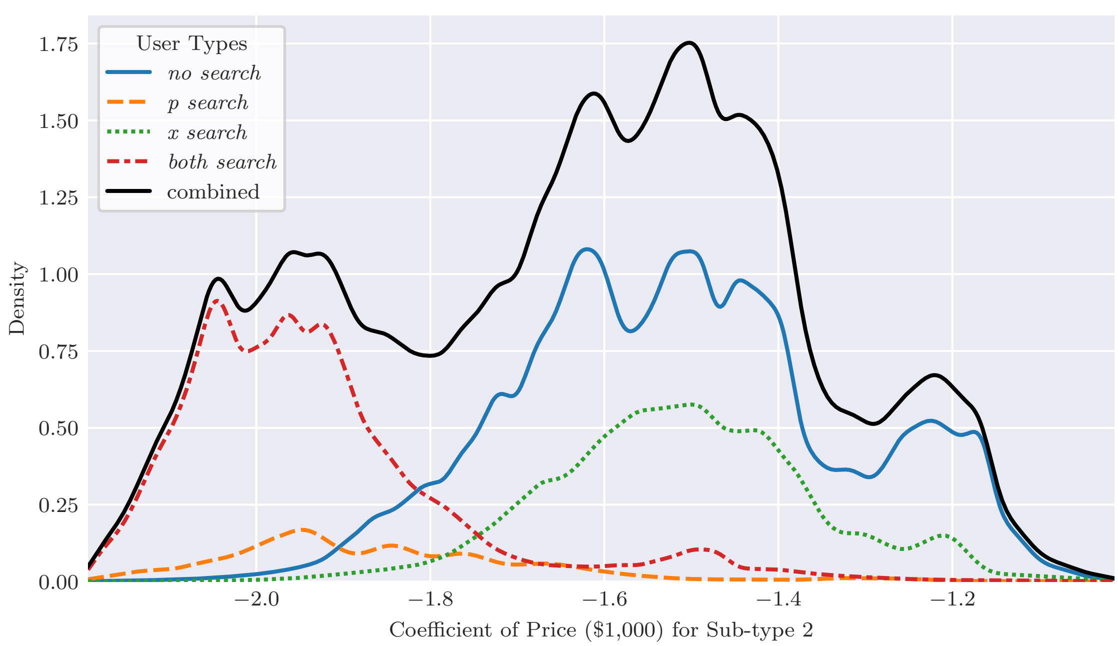


Figure 1.8: Distribution of Price Coefficient of Model IV - MLP (Sub-type 2, by User Types)

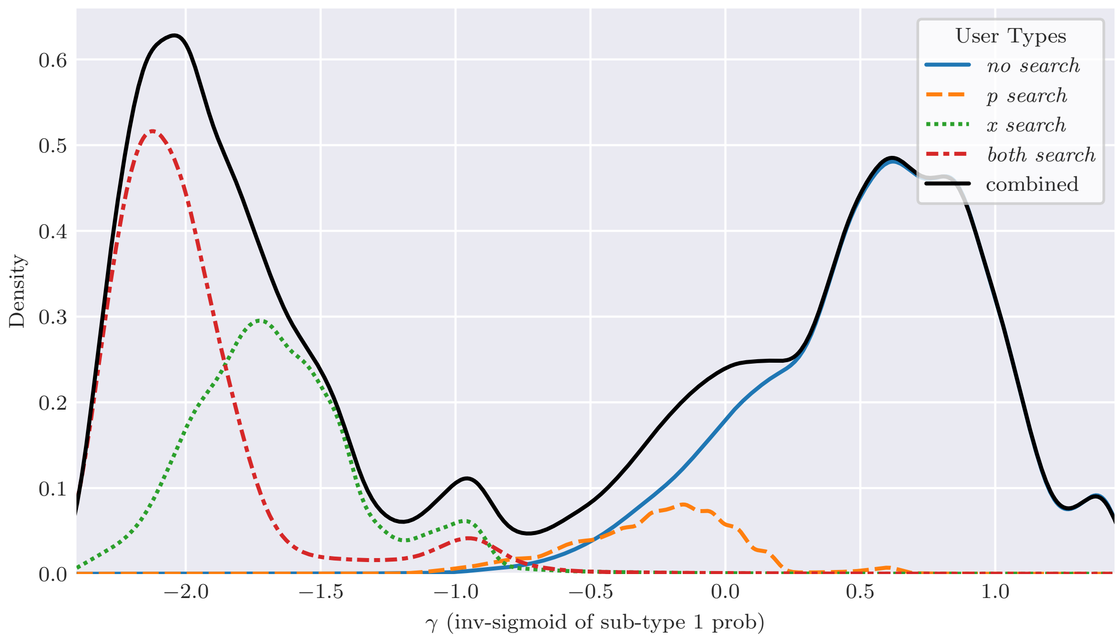


Figure 1.9: Distribution of Parameter γ of Model IV - MLP (by User Types)

browsing behavior into a vector representation, inevitably leads to some information loss—for instance, the sequence of clicks. To retain the sequence information, one might consider concatenating the features of each click to form an exceptionally lengthy vector for every user. This approach, while comprehensive, results in an exceedingly large parameter set, even with a minimal number of hidden layers and sizes. The challenge then becomes how to directly utilize a user's historical clickstream data, without manual intervention, and account for the click sequence order in a manner that is both parsimonious and capable of capturing individual-level heterogeneity. The subsequent section aims to address these critical issues.

1.5.7.3 Model V - RNN

To circumvent the need for human intervention in generating user-specific feature vectors and to minimize the information loss inherent in aggregating clickstream data into a singular vector, a model capable of processing raw clickstream data directly is essential. Recurrent Neural Networks (RNNs) (Rumelhart and McClelland, 1987), as introduced by are renowned for their ability to handle sequential data, showcasing impressive performance across various domains. This chapter innovatively advocates for the application of RNNs to sequential user clickstream data, aiming to autonomously generate user-specific features. While these generated features may lose their interpretability, this methodology offers an automated and efficient means of capturing consumer heterogeneity, allowing the model to discern and highlight pertinent features on its own.

Echoing Model IV, this section harnesses the historical clickstream data and the first smartphone click as the inputs to discern observed consumer heterogeneity. However, diverging from the previous model that distilled these inputs into a feature vector, this approach retains the data in its original form, with the sole modification being the introduction of dummy variables. For each user, the input materializes as a two-dimensional matrix, wherein rows represent click events, and columns encapsulate event attributes such as price, event type dummies, event duration, temporal dummies, and an embedded scalar denoting product category.[10]

It is important to note that due to varying lengths of historical clickstreams among users, the matrices may differ in row count for each user within the set $\mathcal{N}$. To standardize the dimensions of these matrices across all users, the model employs truncation and padding techniques, a common practice in machine learning literature.

[10] Embedding is explained in the notes accompanying Figure 1.10.

Specifically, the clickstream is truncated to 50 events, with any shortfall being compensated by padding zeros, resulting in a uniform matrix, denoted as $\mathbf{B}_n$ for $n \in \mathcal{N}$. Consequently, $\mathbf{B}_n \in \mathbb{R}^{50} \times \mathbb{R}^{21}$ serves as the input for the RNN model. The RNN model's output is subsequently amalgamated with the brand indicators from the first smartphone click and user type indicators. This combined data is then fed into a MLP to generate the preference coefficients and the γ parameter.[11]

For notational convenience, denote the composition of RNN and the subsequent MLP as $G_{\text{RNN}} : \mathbb{R}^{50} \times \mathbb{R}^{21} \rightarrow \mathbb{R}^{49}$, and (1.10) is effectively

$$G_{\text{RNN}}(\mathbf{B}_n; \mathbf{\Theta}_{\text{V}}) = \boldsymbol{\theta}_{\text{RNN},n} \tag{1.13}$$

where $\boldsymbol{\theta}_{\text{RNN},n} = \left[\alpha_{\text{RNN},n,s}, \boldsymbol{\beta}_{\text{RNN},n,s}{}^{\mathsf{T}}, \gamma_{\text{RNN},n}\right]_{s\in\mathcal{S}}^{\mathsf{T}}$ has a size of 49.

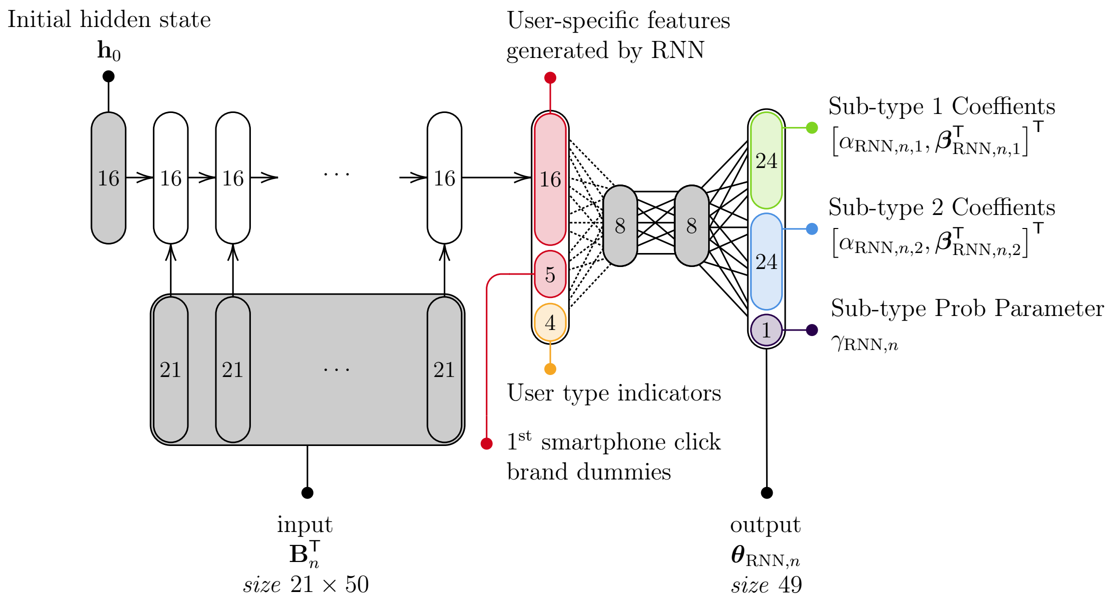
The hidden state of the RNN component is set to a dimension of 16, strategically constraining the number of parameters to showcase the efficiency of RNNs in automatically crafting machine-interpretable, user-specific features, a marked advancement over the human-created counterparts in Model IV. To ensure comparability with Model IV, the MLP components of this model mirror the same structure, featuring two hidden layers, each with a hidden size of 8, and employing the ReLU activation function. As a result, $\mathbf{\Theta}_{\text{V}}$ contains 1,359 scalar parameters. The structural design of the RNN is depicted in Figure 1.10.

Consequently, the utility $U_{n,s,i}$ as from the generic form (1.1) with the additional subscript $s \in \mathcal{S}$ is hence,

$$U_{n,s,i} \equiv u_{n,s,i} + \epsilon_{n,i} = \alpha_{\text{RNN},n,s} \cdot p_{t(n),i} + \boldsymbol{\beta}_{\text{RNN},n,s}{}^{\mathsf{T}} \mathbf{x}_{n,t(n),i} + \epsilon_{n,i}, \quad \forall i \in \mathcal{C}_n. \tag{1.14}$$

The conditional choice probability mirrors that presented in (1.9) incorporating the

[11] This MLP is not the same MLP model in discussed in Section 1.5.7.1, albeit sharing a remarkably similar architecture.



The structure also incorporates an additional embedding layer with dimensions $(22, 1)$, an one-to-one mapping from the 21 product categories (level 1) and the padded entry to 22 scalars. Each column (*size* 21) of $\mathbf{B}_n^{\intercal}$ contains the scalar value associated with the product category (or padded entry) pertinent to the respective click event. Dashed lines indicate that the layer has no bias due to multicollinearity among the user-type indicators.

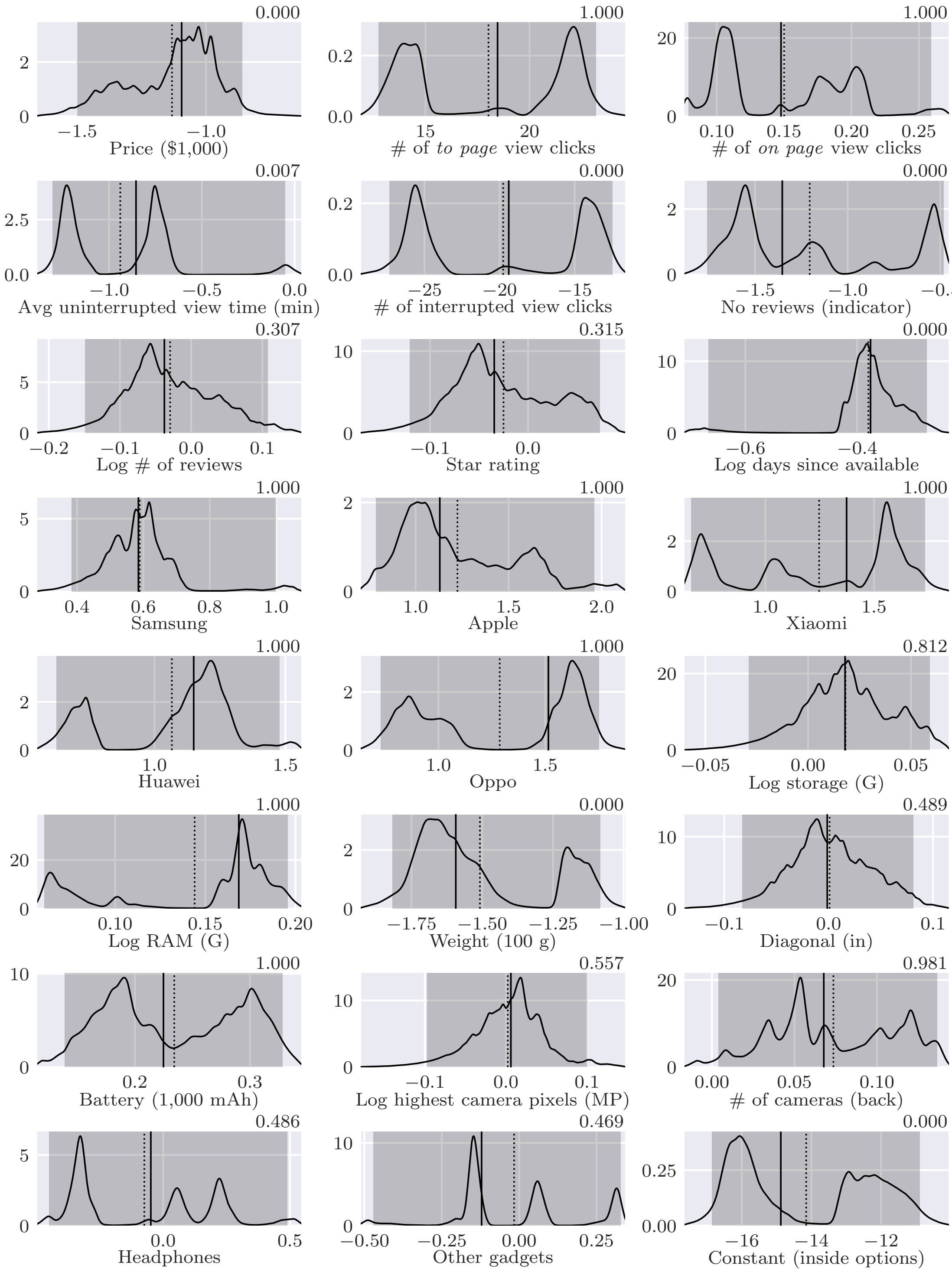
Figure 1.10: Model IV - RNN Structure

updated specifications of $U_{n,s,i}$ and $u_{n,s,i}$ as outlined in (1.14). Consistently, the log-likelihood function is $\ell(\boldsymbol{\Theta}_{\mathrm{V}})$ in accordance with equation (1.3), adapting to this updated probability framework.

1.5.7.4 Estimation Result of Model V - RNN

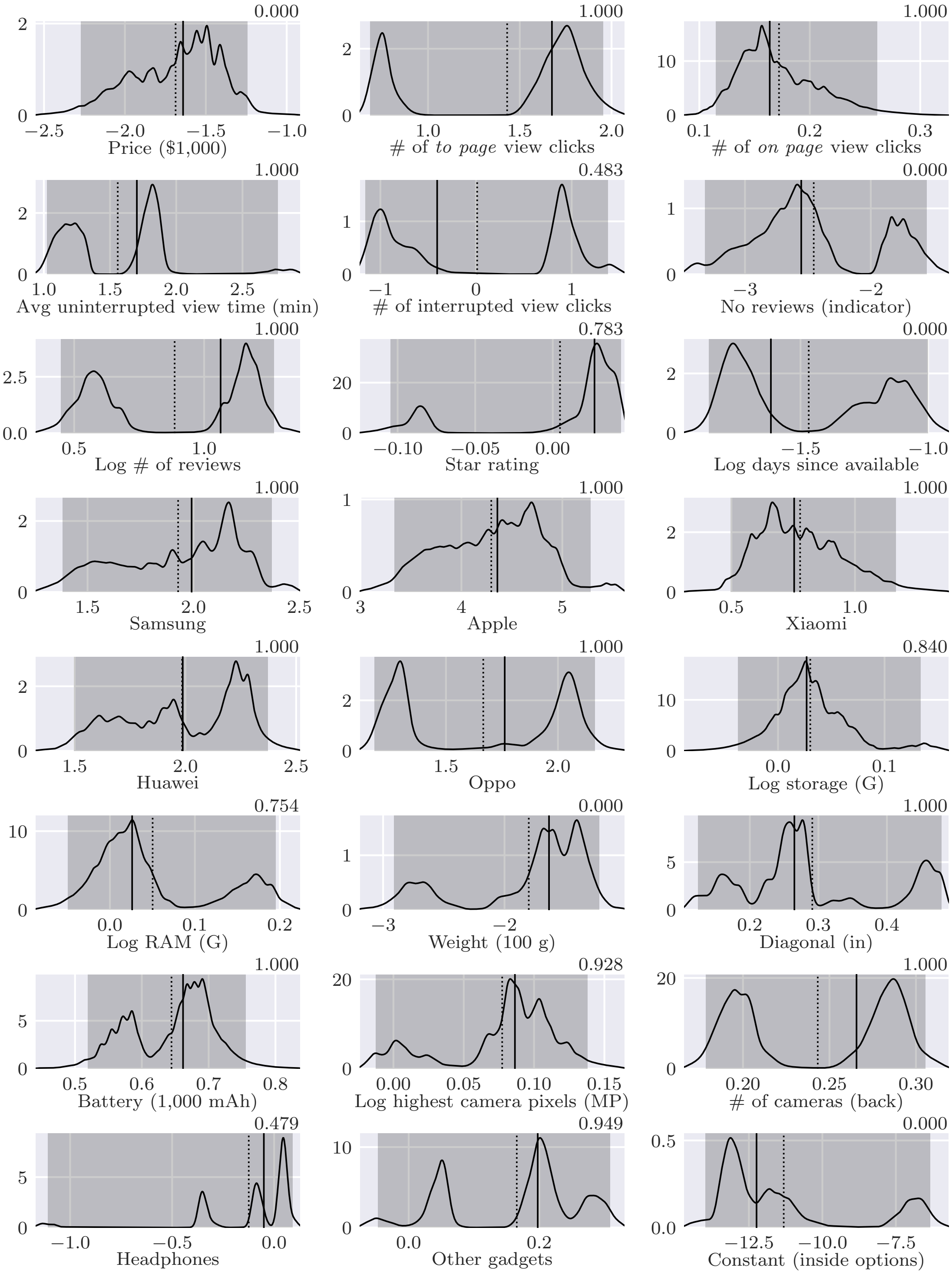
Figure 1.11 delineates the distributions of MLP output coefficients, $\boldsymbol{\theta}_{\mathrm{RNN},n}$, for $n \in \mathcal{N}$, with sub-figures (a), (b), and (c) illustrating the coefficients for sub-type 1, sub-type 2, and γ_{RNN}, respectively.

The estimation results bear a qualitative resemblance to those of Model IV, although the distributions for some coefficients show diverging shapes from those

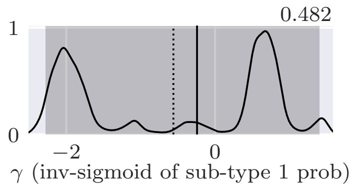


(a) Distributions of Coefficients (Sub-type 1)

Each plot in sub-figures (a), (b), and (c) displays the central 99.5% density of each distribution, with the median and mean indicated by solid and dashed lines respectively, central 95% densities indicated by the shaded region, and the probability of > 0 indicated on the top right corner.



(b) Distributions of Coefficients (Sub-type 2)



Estimation					
Observations	:	6,395,207,961	Users (Cases)	:	4,235,321
LL	:	−1,200,870.6	Avg LL Per User	:	−0.284
AIC	:	2,404,460.2	BIC	:	2,422,479.2
Holdout					
Observations	:	2,131,744,036	Users (Cases)	:	1,411,773
LL	:	−400,974.7	Avg LL Per User	:	−0.284

(c) Distribution of γ and Estimation Statistics

Figure 1.11: Estimation Result of Model V - RNN

in Model IV. For instance, when examining the distributions of coefficients for the log # of review, star rating, and Samsung in Figure 1.6 (a) and contrasting them with their counterparts in Figure 1.11 (a), it's evident that the distributions from the former exhibit a multi-modal nature. This divergence could be attributed to the inability of $\mathbf{b}_n$ to encapsulate the full spectrum of variations as effectively as $\mathbf{B}_n$, likely a consequence of both the aggregation process and the human intervention involved in constructing $\mathbf{b}_n$. Consequently, the estimation outcomes might disproportionately depend on the user type indicators in the previous model, leading to segregated support across different user types.

Figure 1.12 and 1.13 provide a detailed breakdown of densities by user types for the two respective sub-types, closely mirroring the patterns observed in Figure 1.7 and 1.8, and contrasting Figure A.3 and A.4 with no straightforward separation by active months. Figure 1.14 shows that the parameter $\gamma_{\text{RNN},n}$ by user types shares similar patterns as discussed in the previous model. These resemblances underscore the connection between users' search behaviors and their price sensitivities even at the individual level, demonstrating the extent to which a platform can infer about a user's preferences based on their interactions. This insight not only reaffirms the observed patterns but also sheds light on the potential for platforms to tailor experiences and offers to individual users.

The estimation statistics deliver pivotal insights into this model's efficacy. Model

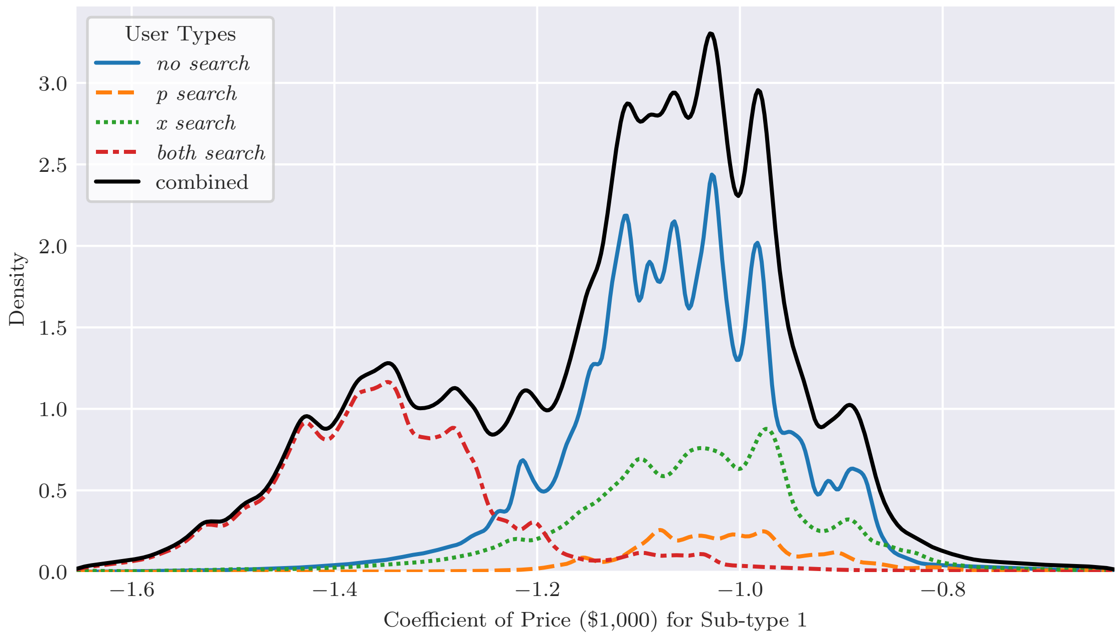


Figure 1.12: Distribution of Price Coefficient of Model V - RNN (Sub-type 1, by User Types)

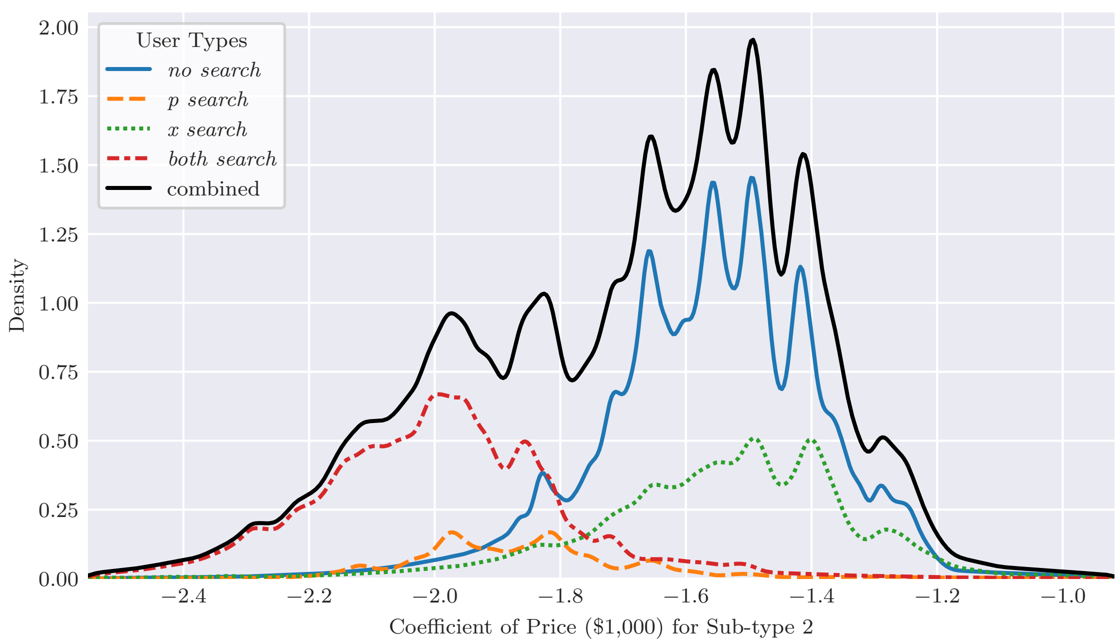


Figure 1.13: Distribution of Price Coefficient of Model V - RNN (Sub-type 2, by User Types)

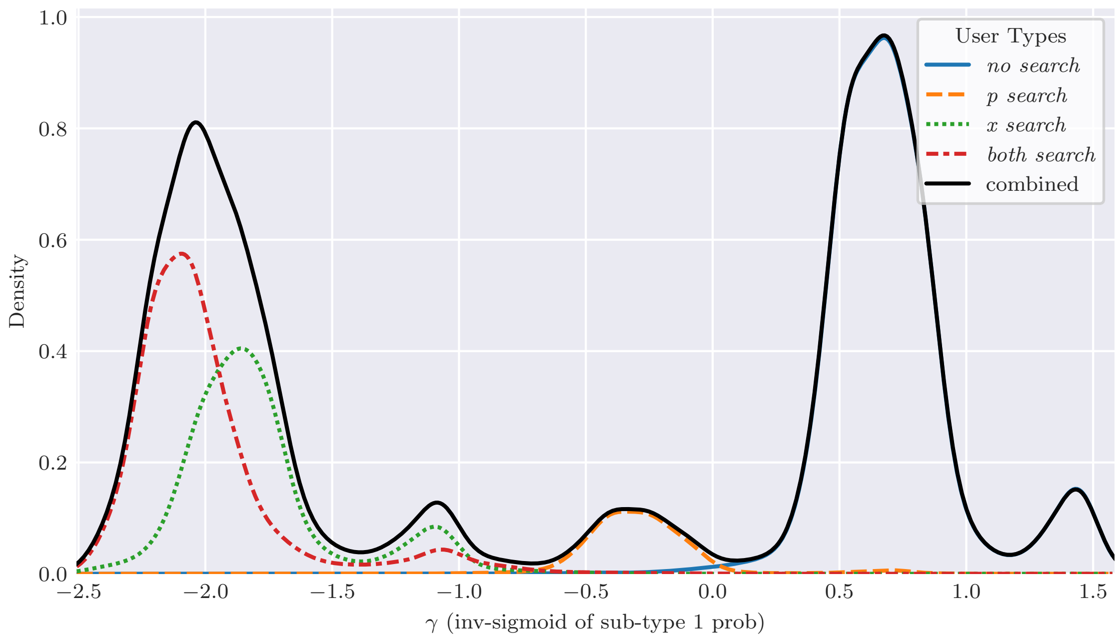


Figure 1.14: Distribution of Parameter γ of Model V - RNN (by User Types)

V rejects Model I through III by the likelihood ratio test. However, the likelihood ratio test is not applicable for comparisons between Model IV and Model V due to their fundamentally divergent architectures, precluding the characterization of one as a restricted variant of the other. It is precisely this distinction in structural design that enables Model V to outperform Model IV in terms of model fit with fewer parameters. This observation does not imply a deficiency of the MLP structure relative to RNN; rather, it underscores that for the dataset and application at hand, the RNN framework emerges as a more suitable choice for capturing heterogeneity efficiently and parsimoniously, with minimal need for manual intervention.

Other machine learning models are also viable for sequential data applications, such as GRU (Cho et al., 2014) and LSTM (Hochreiter and Schmidhuber, 1997), alongside attention-based models like Transformers (Vaswani et al., 2017). These alternatives could enhance estimation accuracy and may be better suited for analyzing larger datasets and more extensive browsing histories. The innovative application of RNNs transcends clickstream data analysis, offering broad applicability to panel-structured data in general. Amidst the burgeoning interest in natural language processing (NLP), the advancements in large language models (LLMs) and generative artificial intelligence (GenAI), the machine learning domain continues to unveil sophisticated techniques for non-parametric analysis on panel data within the economic sphere, paving the way for more nuanced and insightful empirical research.

Table A.3 provides a comprehensive comparison of all the demand models discussed in this chapter.

1.6 Price Elasticities

In the realm of sophisticated demand modeling, price elasticity emerges as a pivotal metric for evaluation. The conditional logit choice model, however, is hampered by the Independent Irrelevant Alternative (IIA) property. Introducing heterogeneity into the model mitigates this limitation, yielding more accurate substitution patterns among products. Figure 1.15 presents a heatmap of the price elasticities derived from Model V with RNN, and Figure A.5 illustrates its construction methodology.

The patterns observed in Figure 1.15 reveal that a price change in one product does not uniformly affect the demand for other products. Specifically, the near-diagonal regions (a) and (d) exhibit numerous non-zero values, indicating significant cross-price elasticity. Conversely, the off-diagonal regions, (b) and (c), predominantly display zero values, underscoring minimal cross-price elasticity. This outcome intuitively suggests that a price shift in a smartphone significantly impacts the demand for other smartphones within similar price bracket, while exerting negligible influence on those in the opposite price spectrum.

Within each of Figures 1.15, A.8, and A.9, the (a) region highlights two rows with pronounced cross-price elasticities for Meizu smartphones, suggesting a high likelihood of consumer migration to similarly priced alternatives in response to even minor price adjustments. Additionally, the apparent cluster of elevated cross-price elasticities observed in the lower right of region (d) corresponds to a cohort of Apple iPhone 11 models, reflecting coherent substitution patterns. These observed substitution dynamics underscore the enhanced realism and efficacy of the demand models.

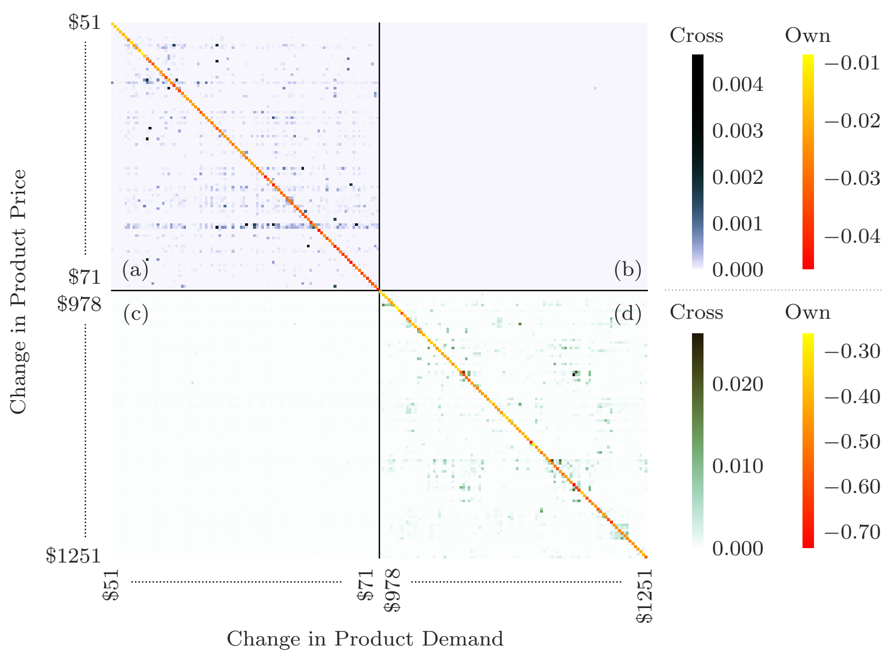


Figure 1.15: Price Elasticities from Model V -RNN

1.7 Conclusion

This chapter has developed a framework for demand estimation that captures the intricate heterogeneity of consumer behavior at the individual level, utilizing a novel integration of historical browsing data with advanced machine learning techniques. This non-parametric approach to understanding consumer diversity is substantiated by empirical findings that highlight significant variability in consumer preferences during online smartphone purchases.

Empirical evidence underscores a pronounced link between consumers' search behaviors, especially concerning price and product attributes, and their underlying preferences. This correlation not only validates the methodological approach employed but also deepens the understanding of consumer decision-making processes. By transcending the simplified heuristic-based classifications, the chapter reveals a complex portrait of consumer preferences, offering a detailed and realistic view of market demand.

Additionally, this analysis confronts the limitations posed by the IIA assumption, thereby facilitating the identification of realistic product substitution patterns. This advancement improves the precision of demand estimations and yields valuable insights for strategic product positioning and competitive market analysis.

RNNs, with their ability to process and interpret the sequential patterns of consumer behavior, stand out as significant analytical assets. The incorporation of RNNs into this study underscores the capabilities of sequence-based machine learning models for examining panel data in economics. Such advancements in methodology indicate a promising direction for more nuanced and dynamic interpretations of economic data.

While leveraging user interaction variables as covariates furnishes profound insights into consumer behavior, it inherently assumes the exogeneity of consumers' strategic search behaviors on smartphones. This aspect potentially limits the scope for counterfactual analyses regarding certain policy shifts, for example, the introduction of personalized pricing at the onset or during the consumer search process.

Nevertheless, this constraint does not diminish the model's capacity to conduct counterfactual analyses of post-search personalized discounts presented as definitive offers in a "take-it-or-leave-it" format. Thus, this comprehensive and detailed demand model lays the groundwork for examination of the impact of personalized discounts on both firm profit and consumer welfare in the next chapter. This extension not only enhances the model's applicability to real-world scenarios but also provides critical insights into the efficacy of targeted marketing strategies in optimizing economic outcomes.

Chapter 2

Personalized Discounts via Coupons: Generating Gains Conditional on Observed Choices

2.1 Introduction

The practice of companies monitoring individuals' online activities to enhance their marketing approaches is well-documented (Madrigal, 2012). The detailed observations of online behavior grant firms deep insights into consumer preferences, thereby enabling them to enhance their product and service offerings. The advent of individual-level preference (e.g., from Chapter 1) has paved the way for the introduction of personalized pricing strategies. Consequently, this sophisticated understanding of consumer preferences holds profound implications for both corporate profits and consumer welfare.

Price discrimination, as a strategic facet of marketing, encompasses various levels. First-degree price discrimination, theoretically, enables firms to maximize profits by capturing the entire consumer surplus (Pigou, 1920). And personalized pricing is often recognized as an exemplar of first-degree price discrimination in the literature (e.g., Shiller and Waldfogel, 2011; Shiller, 2014; Kehoe et al., 2022). However, as Dubé and Misra (2023) contend that, particularly in scenarios reliant on random utility models, firms are inherently incapable of discerning the random component of the utility, ϵ_i, even with repeated observations on the same consumer. This inherent limitation suggests that such practices align more closely with third-degree price discrimination.

This chapter adopts this perspective, acknowledging that while firms might be able to infer the random component to some extent based on the observed individual consumer choice, this component is not fully resolved. As a result, firms are unable to distinguish between individuals exhibiting identical online behaviors and making the same choices within identical choice sets, rendering these consumers identical from the firm's perspective.

This chapter contributes to the extensive literature on price discrimination and its implications for welfare. The advent of personalized pricing has emerged as a pivotal factor influencing market dynamics, particularly concerning firm profitability and consumer welfare. Rossi et al. (1996) suggest that there is a tremendous potential for improving the profitability of direct marketing efforts by more fully utilizing household purchase histories using random coefficient model. Given that uniform pricing represents a constrained version of the price discrimination strategy, personalized pricing should, in theory, at least match uniform pricing in profitability, assuming the firm can effectively optimize its strategies and has a thorough understanding of customer behavior and demand drivers. The positive impact of personalized pricing on profit margins is well-established in the literature. For instance, Shiller and Waldfogel (2011) delve into the digital music industry, revealing that pricing tailored to individual consumers could potentially boost revenues by over 50%, in stark contrast to the minimal revenue impact of third-degree price discrimination based on observable characteristics such as gender, ethnicity, resident alien status, or age. Moreover, Shiller (2014) employs a probit model for Netflix subscription prediction, leveraging consumer purchase histories to estimate heterogeneous demand for the formulation of personalized prices. Findings reveal that personalized pricing models with even basic web behavior can lead to a revenue uptick exceeding 12%, and suggests Web browsing data make personalized pricing more appealing to firms and is likely to be implemented, thus impacting consumers.[1] Additionally, Dubé and Misra (2023) explore the efficacy of personalized pricing in a field experiment with ZipRecruiter employing a Bayesian framework, and uncovers a substantial profit enhancement through personalized pricing strategies, surpassing those predicted in past work.

Amid these results, the discourse surrounding personalized pricing extends beyond profit maximization, encompassing broader implications for consumer surplus and societal welfare. Acquisti et al. (2016) reviews the economic literature on privacy, concluding that both sharing and protecting personal data can lead to positive and negative outcomes at the individual and societal levels. They note that the commercial exploitation of data might reduce private utility and, in some instances, negatively impact social welfare. This gives consumers valid concerns about the unauthorized commercial use of their private information. In a similar vein, the Executive Office of the President (2014) highlights concerns that differential pricing could shift value from consumers to shareholders. In contrast, Dubé and Misra (2023) acknowledge that while price discrimination may typically benefit firms in a monopolistic setting, its effects on consumer welfare remain uncertain, lacking a definitive consensus.

Waldfogel (2015) examines personalized pricing within higher education, revealing that personalization of prices based on test scores decrease consumer surplus more significantly than it boosts revenue. Similarly, Shiller and Waldfogel (2011) observe that personalized pricing negatively affects consumer surplus, suggesting that profit gains come at consumers' expense. On the contrary, Dubé and Misra (2023), consistent with Bergemann et al. (2015), find a nonmonotonic relationship between consumer surplus and the quantity of consumer data available to the firm for personalization. They

[1] Basic web behavior included: total website visits, total unique transactions (excluding Netflix), percent of online browsing by time of day and by day of week, and broadband indicator.

argue that personalized pricing, unlike perfect price discrimination which entirely shifts consumer surplus to the firm, could potentially enhance consumer surplus compared to standard uniform pricing due to inherent classification errors. Furthermore, they highlight the risk of stringent regulations on data-driven pricing strategies could inadvertently diminish social welfare and harm consumers under certain conditions. In the domain of durable goods, Kehoe et al. (2022) study a duopoly scenario between Apple and Samsung in the smartphone and tablet markets, noting that a significant fraction of consumers are better off under price discrimination than from uniform pricing. This chapter shares a close affinity with the work of Shiller (2014), which similarly explores the potential for firms to implement targeted discount strategies. Within this framework, Shiller (2014) discovers that utilizing extensive online browsing data beyond mere demographic details can lead to a concurrent uplift in both firm profits and overall consumer surplus, which aligns with the results of this study.

This chapter examines a counterfactual scenario in which the e-commerce platform can offer a personalized discount (coupon) on a single smartphone to each consumer who ultimately did not make a smartphone purchase. Importantly, this analysis circumvents the limitation of the exogeneity of user interactions with the product, since the discounts are extended as definitive, take-it-or-leave-it offers post the consumer's completion of their search process and their decision to choose the outside option. The introduction of a discount after the search process concludes does not alter the user's prior interactions, and the nature of the take-it-or-leave-it offer precludes any further engagement with the product. The impact of this marketing strategy on both profit and consumer surplus is assessed. Two models equipped with individual-level preference parameters, namely MLP and RNN, are each evaluated under three different schemes: limiting the firm's choice of the discounted smartphone for each consumer to the last clicked smartphone, any clicked smartphone, or any smartphone irrespective of click history.

Conditional on the observed user choices prior to the application of discounts, the analysis demonstrates that personalized discounts can concurrently boost profit by 3.15% and consumer surplus by an impressive 6.30% in expected terms. Notably, the RNN model, which capitalizes on the sequential insights from clickstream data, outperforms the MLP model, which relies on aggregated user-specific features vector, by generating 50 basis points higher and over 141 basis points higher percentage gains in conditional expected profit and consumer surplus respectively. These pieces of evidence support the insights from Shiller (2014) and challenge the notion that personalized pricing invariably benefits firms at consumers' expense under universal circumstances. Moreover, this chapter reveals that as firms are afforded greater discretion in selecting smartphones for discounts across the three examined strategies, there is a corresponding rise in the incremental gains in both conditional expected profit and consumer surplus.

The remaining of this chapter is structured as follows. Section 2.2 introduces the wholesale price data employed in the study. Section 2.3 describes the counterfactual settings, the firm's optimization problem, and the approach to calculating relevant welfare measures. Section 2.4 presents the estimation results. The chapter concludes with Section 2.5, summarizing the main insights and their broader implications.

2.2 Data

This study gathers wholesale price data for smartphones from a select group of wholesalers. Although it remains ambiguous whether those wholesalers are part of the supply chain linked to the e-commerce platform in this study, the similarity in the quoted wholesale prices suggests a highly competitive market. The data collection effort focused on seven distinct dates, ranging across the timeframe of the clickstream data. In instances of multiple price points for the same smartphone on a given date, the minimum value was adopted to maintain consistency.

The wholesale prices for each smartphone model on any given date is linearly interpolated, and denoted as $c_{t,j}$. As a result, wholesale prices are available for over 60% of the unique smartphone models captured in the clickstream dataset, encompassing 92.5% of all recorded smartphone clicks, and 99.5% of the purchases. It's noteworthy that in a minuscule fraction of smartphone clicks (approximately 0.04%), the wholesale price surpassed the retail price. In these rare instances, the wholesale price was pragmatically adjusted to match the retail level, i.e., resulting in a markup of 0.

This chapter restricts that only the set of smartphones with available wholesale price data, denoted as $\mathcal{D}_t$ at time t, are considered eligible for personalized discount offers. And hence, should provide a lower bound on profit gains for less restricted situations.

2.3 Model

This section outlines a strategic framework enabling e-commerce platforms to enhance profitability through personalized discounts (coupons) upon the conclusion of a user's search session. As users conclude their online exploration, the platform acquires valuable insights into their choice decisions, i.e., when $y_{n,q} = 1$ for some $q \in \mathcal{C}_n$, then q is the optimal choice for user n without discounts. Although, the exact utility levels of different choices are unobserved, this observation gives invaluable insights about the relationship between them.

Leveraging the intricate understanding of consumer heterogeneity as delineated in

Chapter 1 and consumers' choice decisions before discounts, the platform can tailor coupons to individual preferences. This counterfactual analysis of personalized discounts fosters a captivating exploration of personalized marketing's impact on both profitability and consumer surplus.

Given the observed clickstream data, this study delves into the situation where the platform can issue a personalized discount via a coupon for a specific smartphone at the end of user's search sessions. Assume the platform will strategically issue a personalized discount $d_{n,k}$ for user n on smartphone k contingent upon the absence of a purchase, i.e., $y_{n,0} = 1$. Conversely, no discount is extended if a purchase has been made, i.e., $y_{n,0} \neq 1$. Base on the model assumptions, the user who receives the discount experiences an increase in the utility of smartphone k as their price coefficient is negative, and hence, has a higher probability of purchasing smartphone k relative to other options.

The observation that whether user n has purchased a smartphone or not helps the platform and the econometricians to establish more informed analysis on profit and the resulting consumer surplus.

2.3.1 Conditional Expected Profit Maximization

Prior to the issuance of any discounts, the expected profit derived from user n conditional on observing user n's choice, is simply the realized profit generated by user n, denoted as Π_n. i.e.,

$$\Pi_n = \sum_{j \in \mathcal{A}_{t(n)}} y_{n,j} \cdot (p_{t(n),j} - c_{t(n),j}). \tag{2.1}$$

With the introduction of the discount $d_{n,k}$ for user n who ends up not purchasing, smartphone k becomes more appealing than before, due to the fact that $\alpha_{n,s,k} < 0$ for all s and $d_{n,k} > 0$. Given that the personalized discount is exclusively associated with smartphone k, the perceived utility of other smartphones remains unchanged. According to the principle of revealed preference, user n will either maintain their preference for the outside option if the utility of smartphone k, even after applying the discounts $d_{n,k}$, falls short of the utility offered by the outside option, or opt to purchase smartphone k if its utility, after the discount, surpasses that of the outside option. The conditional expected profit gain attributable to the discount for user n with $y_{n,0} = 1$ is therefore,

$$\overset{\triangle}{\mathsf{\Pi}}_n(d_{n,k}) \equiv \sum_{s \in \mathcal{S}} \boldsymbol{\pi}_{n,s} \cdot \overset{\triangle}{\mathsf{\Pi}}_{n,s}(d_{n,k}) \tag{2.2}$$

$$= \sum_{s \in \mathcal{S}} \boldsymbol{\pi}_{n,s} \cdot \tilde{\mathsf{P}}_{n,s,k}(d_{n,k}) \cdot (p_{t(n),k} - c_{t(n),k} - d_{n,k}) \tag{2.3}$$

where $\boldsymbol{\pi}_{n,s}$ denotes the probability for user n being sub-type s conditional on choosing the outside option prior to any discounts, i.e.,

$$\boldsymbol{\pi}_{n,s} = \frac{\pi_{n,s} \cdot \text{Prob}\Big(U_{n,s,0} > U_{n,s,j},\ \forall j \in \mathcal{C}_n \setminus \{0\}\Big)}{\sum_{v \in \mathcal{S}} \pi_{n,v} \cdot \text{Prob}\Big(U_{n,v,0} > U_{n,v,j},\ \forall j \in \mathcal{C}_n \setminus \{0\}\Big)}. \tag{2.4}$$

And $\tilde{\mathsf{P}}_{n,s,k}(d_{n,k})$ denotes user n's probability of purchasing smartphone k after the discount $d_{n,k}$ is given, conditional on choosing the outside option without the discount and sub-type s. Hence,

$$\tilde{\mathsf{P}}_{n,s,k}(d_{n,k}) \equiv \text{Prob}\Big(\tilde{U}_{n,s,k}(d_{n,k}) > U_{n,s,i},\ \forall i \in \mathcal{C}_n \setminus \{k\}\ \Big|\ U_{n,s,0} > U_{n,s,j},\ \forall j \in \mathcal{C}_n \setminus \{0\}\Big). \tag{2.5}$$

In the aforementioned equation, $\tilde{U}_{n,s,k}(d_{n,k})$ denotes the updated utility for smartphone k with the coupon $d_{n,k}$, i.e.,

$$\tilde{U}_{n,s,k}(d_{n,k}) = U_{n,s,k} - \alpha_{n,s} \cdot d_{n,k}. \tag{2.6}$$

Appendix B.1 shows that

$$\tilde{\mathsf{P}}_{n,s,k}(d_{n,k}) = \frac{\Delta_{n,s,k}(d_{n,k})}{\Delta_{n,s,k}(d_{n,k}) + \sum_{j \in \mathcal{C}_n} \exp(u_{n,s,j})} \tag{2.7}$$

$$\text{where} \quad \tilde{u}_{n,s,k}(d_{n,k}) = u_{n,s,k} - \alpha_{n,s} \cdot d_{n,k},$$
$$\Delta_{n,s,k}(d_{n,k}) = \exp(\tilde{u}_{n,s,k}(d_{n,k})) - \exp(u_{n,s,k}).$$

The selection of $d_{n,k}$ is naturally limited to the open interval $\left(0, p_{t(n),k} - c_{t(n),k}\right)$, because selecting values outside this interval would result in the objective function having zero or negative values, whereas any $d_{n,k}$ within this range ensures a strictly positive outcome for the objective function. Let $\mathcal{W}\left(0, p_{t(n),k} - c_{t(n),k}\right)$ denotes the set of discrete integer numbers (in local currency) that is strictly between 0 and $p_{t(n),k} - c_{t(n),k}$, evenly spaced for approximately every \$1.29. The platform is assumed to optimize $d_{n,k}$ over this discrete set $\mathcal{W}\left(0, p_{t(n),k} - c_{t(n),k}\right)$, rather than the continuous interval $\left(0, p_{t(n),k} - c_{t(n),k}\right)$. This assumption is underpinned by several considerations. Firstly, discrete coupons align more closely with practical applications in real-world settings. Secondly, the first-order condition for maximizing over $d_{n,k}$ continuously results in an implicit function that lacks a closed-form solution. Moreover, no specific condition guarantees a unique fixed point solution for this implicit function. Additionally, the concavity of the objective function is contingent upon the values of the conditional probability, rendering the existence of a unique maximum uncertain, although the existence of a maximum is assured by the Extreme Value Theorem. Lastly, the absence of a guaranteed unique fixed point for the implicit function or a unique maximum for the objective function may lead to poor convergence of numerical algorithms aimed at approximating the solution.

Three schemes under consideration impact the set $\mathcal{K}_n$ within which k is optimized, (1) the last clicked smartphone by user n if eligible for discount, (2) all smartphones that have been clicked by user n and are eligible for discount, or (3) all smartphones eligible for discount, $\mathcal{D}_{t(n)}$, regardless of whether they have been clicked or not. It is important to highlight that the three schemes are presented in decreasing order of restrictions. Specifically, in the first scheme, $\mathcal{K}_n$ has a max cardinality of 1, and hence, the optimization over k is trivial. If ever $\mathcal{K}_n$ is the empty set, it implies the absence of eligible discounts for user n, resulting in the objective function assuming a value of zero. Consequently, for each user n whose $y_{n,0} = 1$, the e-commerce platform face the following conditional expected profit maximization problem over k and $d_{n,k}$,

$$\max_{k \in \mathcal{K}_n} \max_{\substack{d_{n,k} \in \\ \mathcal{W}\left(0, p_{t(n),k} - c_{t(n),k}\right)}} \sum_{s \in \mathcal{S}} \pi_{n,s} \cdot \frac{\Delta_{n,s,k}(d_{n,k})}{\Delta_{n,s,k}(d_{n,k}) + \sum_{j \in \mathcal{C}_n} \exp(u_{n,s,j})} \cdot (p_{t(n),k} - c_{t(n),k} - d_{n,k}), \tag{2.8}$$

and the sum of the objective functions at the optimal values over $n \in \mathcal{N}$ whose $y_{n,0} = 1$ is the optimal gain in conditional expected profit attributable to the personalized discounts, which is a key parameter of interest.

2.3.2 Conditional Expected Consumer Surplus

In light of the platform's strategic issuance of personalized discounts to optimize conditional expected profit, there is a consequent change in consumer surplus due to the ensuing price drops. Given that utilities are inherently comparative, the consumer surplus discussed herein is benchmarked against the outside option 0 which has a utility of $U_{n,s,0} = 0 + \epsilon_{n,0}$. To assess the shifts in consumer surplus, let CS_n denote the conditional expected consumer surplus prior to the application of discounts.

For a user who did not make a purchase prior to any discount, i.e., $y_{n,0} = 1$, the consumer surplus, contingent upon opting for the outside option is effectively 0. Conversely, for a user who made a smartphone purchase prior to any discount, i.e., $y_{n,0} \neq 1$, the consumer surplus conditional on the choice $y_{n,q} = 1$ where $q \in \mathcal{A}_{t(n)}$ is therefore $U_{n,s,q} - U_{n,s,0}$. While this specific surplus is not directly observable by econometricians, the model's assumptions and the estimated parameters allow for the calculation of its conditional expectation. Note that

$$\mathsf{CS}_n = \sum_{s \in \mathcal{S}} \boldsymbol{\pi}_{n,s} \cdot \mathsf{CS}_{n,s} \tag{2.9}$$

where $\mathsf{CS}_{n,s}$ is the same conditional expected consumer surplus for user n but of sub-type s prior to the application of discounts. i.e., for $y_{n,q} = 1$ where $q \in \mathcal{A}_{t(n)}$,

$$\mathsf{CS}_{n,s} = \frac{1}{-\alpha_{n,s}} \mathbb{E}\Big[U_{n,s,q} - U_{n,s,0} \Big| U_{n,s,q} > U_{n,s,j}, \forall j \in \mathcal{C}_n \setminus \{q\}\Big] \tag{2.10}$$

Appendix B.2.1 shows that

$$\mathsf{CS}_{n,s} = \begin{cases} 0 & \text{if } y_{n,0} = 1 \\ \dfrac{1}{-\alpha_{n,s}} & \text{if } y_{n,0} = 0, u_{n,s,q} = 0, \\ \dfrac{1}{-\alpha_{n,s}} \cdot \dfrac{u_{n,s,q}}{1 - e^{-u_{n,s,q}}} & \text{if } y_{n,0} = 0, u_{n,s,q} \neq 0. \end{cases} \tag{2.11}$$

Upon receiving the personalized discount $d_{n,k}$, user n will either persist in selecting the outside option 0 which carries a conditional expected consumer surplus of zero, or opt to purchase smartphone k at the discounted price, thereby attaining a strictly positive conditional expected consumer surplus, as previously explained. Consequently, the gain in conditional expected consumer surplus attributable to the discount for user n is

$$\overset{\triangle}{\mathsf{CS}}_n = \sum_{s \in S} \boldsymbol{\pi}_{n,s} \cdot \overset{\triangle}{\mathsf{CS}}_{n,s} \tag{2.12}$$

where

$$\overset{\triangle}{\mathsf{CS}}_{n,s} = \tilde{\mathsf{P}}_{n,s,k}(d_{n,k}) \cdot \frac{1}{-\alpha_{n,s}} E_{n,s,k}(d_{n,k}), \tag{2.13}$$

$$E_{n,s,k}(d_{n,k}) \equiv \mathbb{E}\Big[\tilde{U}_{n,s,k}(d_{n,k}) - U_{n,s,0} \Big| \tilde{U}_{n,s,k}(d_{n,k}) > U_{n,s,0} \wedge U_{n,s,0} > U_{n,s,j}, \forall j \in \mathcal{C}_n \setminus \{0\}\Big]. \tag{2.14}$$

Unfortunately, $E_{n,s,k}(d_{n,k})$ does not have a closed form expression, but Appendix B.2.2 shows that

$$E_{n,s,k}(d_{n,k}) = \int_{-\infty}^{\infty} E_{n,s,k}(d_{n,k}, \epsilon_{n,0}) \cdot e^{-s} \cdot e^{-e^{-s}} \, ds. \tag{2.15}$$

where

$$\begin{aligned} E_{n,s,k}(d_{n,k}, \epsilon_{n,0}) = & \frac{1}{\exp\left[-e^{u_{n,s,k}-\epsilon_{n,0}}\right] - \exp\left[-e^{\tilde{u}_{n,s,k}(d_{n,k})-\epsilon_{n,0}}\right]} \\ & \times \Bigg[-\alpha_{n,s,k} \cdot d_{n,k} \cdot \exp\Big(-e^{u_{n,s,k}-\epsilon_{n,0}}\Big) + \mathrm{Ei}\Big(-e^{u_{n,s,k}-\epsilon_{n,0}}\Big) \\ & \quad - \mathrm{Ei}\Big(-e^{\tilde{u}_{n,s,k}(d_{n,k})-\epsilon_{n,0}}\Big)\Bigg]. \end{aligned} \tag{2.16}$$

Note that $E_{n,s,k}(d_{n,k}, \epsilon_{n,0})$ can be evaluated, and hence $E_{n,s,k}(d_{n,k})$ can be approximated with numerical methods, such as Gaussian quadrature or Monte Carlo. Appendix C illustrates a numerically stable algorithm for approximating $E_{n,s,k}(d_{n,k}, \epsilon_{n,0})$.

2.4 Results

The counterfactual analysis delineated in Section 2.3 is applied to Model IV and Model V in Chapter 1, employing MLP and RNN to capture heterogeneity respectively, and each with the three aforementioned schemes of $\mathcal{K}_n$. Estimation results, along with standard errors, are systematically reported in Table 2.1.

Farrell et al. (2021) proposes a methodology for calculating robust standard errors on inference parameter of interest for structures with DNNs and ReLU activation functions, as seen in Model IV. However, this chapter refrains from employing this approach for several reasons. Firstly, the method's applicability to Model V, which employs RNNs, is questionable due to the specific convergence rate requirements associated with DNN models. Secondly, the conditional expected profit maximization problem involving discrete optimizations, for $k \in \mathcal{K}_n$, results in the non-differentiability of the optimal solution function with respect to $\boldsymbol{\theta}_n$, and consequently, the non-differentiability of the function from $\boldsymbol{\theta}_n$ to the inference parameter of interest, namely, conditional expected profit gain and conditional expected consumer surplus gain. This issue renders the Farrell et al. (2021) methodology inapplicable. Thirdly, even if $|\mathcal{K}_n| = 1$ in (2.8), optimizing $d_{n,k}$ over the open interval $\big(0, p_{t(n),k} - c_{t(n),k}\big)$ and finding its derivative with respect to $\boldsymbol{\theta}_n$ encounters the challenges explained in Section 2.3.1, leading to the analogous non-differentiability concern. In light of these reasons, this chapter employs the bootstrap method for standard errors. Attention is directed towards Model V in Table 2.1, as the results from Model IV demonstrate similar qualitative features and Model V achieves better log likelihood with fewer parameters.

Access to the wholesale data reveals that the realized profit, also known as the conditional expected profit before the issuance of discounts, amounts to \$40.86 million. Furthermore, utilizing the estimated parameters of Model V, the conditional expected consumer surplus prior to the issuance of discounts is determined to be \$3.27 million. These numbers establish the baseline values against which the increases in conditional expected profit or consumer surplus, attributable to personalized discounts, are compared.

The analysis unveils notable percentage gains in conditional expected profit for the three distinct schemes, which stand at 0.95%, 3.15%, and 3.15%, respectively. With the entirety of $\mathcal{D}_{t(n)}$ at the platform's disposal for coupon issuance, an anticipated immediate profit surge exceeding \$1.28 million is on the horizon. Furthermore, the gains in conditional expected consumer surplus are substantial, at 4.44%, 4.86%, and 6.30%, respectively for the three different $\mathcal{K}_n$'s. These positive gains across all schemes underscore the effectiveness of personalized discounts in not only enhancing profit

Table 2.1: Estimation Results of Personalized Discounts

Expected Values	Model IV - MLP			Model V - RNN		
Realized profit ($)	40,862,045.55			40,862,045.55		
Unconditional expected profit ($)	43,256,228.89 (41,702,601.81, 44,754,111.50)			42,782,814.57 (40,179,774.64, 45,613,013.07)		
Conditional CS before coupon ($)	4,486,714.16 (4,388,407.52, 4,574,966.35)			3,268,146.22 (3,179,677.16, 3,367,325.48)		
$\mathcal{K}_n$ schemes	**Last click**	**All clicked**	**All $\mathcal{D}_{t(n)}$**	**Last click**	**All clicked**	**All $\mathcal{D}_{t(n)}$**
Gain in conditional profit ($)	331,929.23 (321,553.94, 344,953.76)	1,077,208.15 (1,033,741.21, 1,125,119.91)	1,081,419.93 (1,037,958.54, 1,129,546.00)	388,962.47 (370,730.73, 409,135.44)	1,286,091.25 (1,223,441.31, 1,345,573.82)	1,289,161.04 (1,226,686.10, 1,348,812.80)
Gain in conditional CS ($)	145,054.18 (144,988.44, 145,125.50)	160,178.22 (159,983.66, 160,337.38)	219,572.72 (218,752.07, 220,285.53)	144,972.20 (144,871.81, 145,079.47)	158,648.06 (158,395.91, 158,908.35)	205,844.97 (203,217.92, 208,489.37)
Gain in sales units	3,940.39 (3,827.47, 4,088.32)	11,675.92 (11,242.89, 12,119.74)	11,677.82 (11,244.88, 12,123.62)	4,816.79 (4,583.80, 5,105.58)	14,334.33 (13,599.98, 15,101.54)	14,333.35 (13,600.62, 15,098.13)
Gain in conditional profit (%)	0.81 (0.79, 0.84)	2.64 (2.53, 2.75)	2.65 (2.54, 2.76)	0.95 (0.91, 1.00)	3.15 (2.99, 3.29)	3.15 (3.00, 3.30)
Gain in conditional CS (%)	3.23 (3.17, 3.31)	3.57 (3.50, 3.65)	4.89 (4.79, 5.01)	4.44 (4.31, 4.56)	4.86 (4.72, 4.99)	6.30 (6.12, 6.48)
Gain in sales units (%)	0.82 (0.80, 0.85)	2.44 (2.35, 2.53)	2.44 (2.35, 2.53)	1.00 (0.96, 1.07)	2.99 (2.84, 3.15)	2.99 (2.84, 3.15)
Share of discounted smartphone was clicked (%)	100.00	100.00	55.05 (54.22, 55.72)	100.00	100.00	65.15 (63.84, 66.20)
Average coupon value ($)	61.22 (61.16, 61.29)	65.93 (65.85, 65.99)	87.99 (87.67, 88.29)	61.52 (61.42, 61.64)	65.59 (65.48, 65.69)	83.22 (82.18, 84.30)
# of grid points searched per model	441,354,463	1,385,892,713	552,873,774,813	441,354,463	1,385,892,713	552,873,774,813

95% confidence intervals in parentheses, based on 120 bootstrap estimations.

margins and conversion rates but also in promoting consumer surplus. Compared to the original uniform pricing on the platform, these findings shed light on the effectiveness of personalized discount strategies, revealing opportunities where both sellers and consumers can mutually benefit. This mutual advantage suggest a fresh perspective on personalized pricing by proposing a scenario where the well-being of consumers and the profitability of sellers are not necessarily opposing objectives.

As $\mathcal{K}_n$ transitions from being highly constrained—focusing initially on the last click, then expanding to include all clicked items, and finally encompassing all items in $\mathcal{D}_{t(n)}$—an increase in the gains in conditional expected profit is naturally observed. Intriguingly, this expansion is paralleled by an uptick in the average value of the coupons issued, with the average coupon value escalating from \$61.52 to \$83.22 alongside the expansion of $\mathcal{K}_n$. This phenomenon hints at a strategy by the platform that transcends the simplistic exploitation of a broader selection to offer smartphones with higher markups alongside less generous coupons. Rather, it reflects a nuanced approach that considers the balance between the conditional probability of purchase and the post-coupon markup, as detailed in (2.8). This trend of increasing gains in conditional expected profit, however, should not be misconstrued as the platform gaining more leeway in selecting smartphones for coupon issuance to maximize profit extraction from consumer surplus. Partially due to the issuance of higher-valued coupons, it is captivating to see that the increasing gains in conditional expected profit coincides with increasing gains in conditional expected consumer surplus. This challenges the concern that often portrays personalized pricing strategies as tools for sellers to enhance profits at the expense of consumers. Armed with a rich estimation of consumer heterogeneity, the platform's maximization problem, though primarily focused on the gain in conditional expected profit, inadvertently leads to an improvement in consumer welfare. This underscores the notion that the dynamics between profit and consumer surplus are not inherently antagonistic; and additionally, in fact, they may even be mutually reinforcing.

In the third scheme, when $\mathcal{K}_n = \mathcal{D}_{t(n)}$ for each n where $y_{n,0} = 1$, coupons are distributed for over 1,000 distinct smartphones across all users. This highlights the rich heterogeneity of consumer preferences, and showcases the capability of machine learning-enhanced models to capture this diversity effectively. Table 2.1 reveals a fascinating insight that when $\mathcal{K}_n = \mathcal{D}_{t(n)}$, a mere 65.15% of the coupons are issued on a smartphone that has been clicked by the user. This unveils that approximately one-third of the users are receiving coupons on smartphones they haven't interacted with during their search process. Despite this phenomenon being a consequence of the platform's strategic selection of k and $d_{n,k}$ to optimize conditional expected profits, it intriguingly highlights the platform's role in uncovering new options for users—options they might not have considered otherwise. The issuance of a coupon boosts the likelihood of a consumer purchasing the smartphone they would otherwise overlook. Thus this illustrates the strategic advantage of personalized discounts in unveiling potential preferences and expanding consumer choice.

Lastly, the comparison between the outcomes of Model IV - DNN and those of Model V - RNN, reveals significant insights. Model V capitalizes on the RNN's architecture to directly leverage raw clickstream data with minimal preprocessing. This approach stands in stark contrast to Model IV, which relies on consumer features distilled from the raw clickstream through aggregation, a method that inevitably leads to a loss of valuable information. This fundamental difference in information retention is precisely why Model V, despite its leaner parameter set, outperforms in terms of log likelihood. The impact of these differences in information processing transcends mere performance indicators, manifesting in substantial implications for the observed enhancements in both profit and consumer surplus across various scenarios. Notably, in the third scenario where $\mathcal{K}_n = \mathcal{D}_{t(n)}$, Model V with RNN delivers an estimated percentage gains in conditional expected profit that are 50 basis points higher, and conditional expected consumer surplus that are over 141 basis points higher, all achieved with a lower average coupon value. The detailed capture of consumer behavior, which facilitates a deeper insight into consumer heterogeneity, does not invariably translate to an increase profit at the expense of consumer welfare. The availability of granular data affords a more refined comprehension of consumer preferences and behaviors, enabling the design of personalized offerings that align more closely with individual needs and desires. This underscores the principle that acquiring more detailed consumer data can set the stage for outcomes that are mutually advantageous for businesses and consumers alike.

2.5 Conclusion

By leveraging a demand model that intricately captures individual-level heterogeneity, this chapter delves into the implications for firm profitability and consumer welfare within a counterfactual framework, empowering the platform to offer personalized discounts on a specific smartphone model to each consumer who opt not to make a purchase at the conclusion of their search session.

The analysis presented herein suggests the potential of a mutually beneficial scenario for both firms and consumers, conditional on the observed user choices prior to discounts. It outlines a strategy through which firms can not only realize immediate gains in profitability and conversion rates but also amplify consumer surplus. Leveraging the foundation laid by a numerically stable algorithm, which guarantees the efficient and accurate computation of conditional expected consumer surplus, this analysis further offers numerical estimates of the welfare effects and examines the diverse impacts of a firm's flexibility in selecting smartphone products for discount implementation.

This analysis also emphasize that firm's profit and consumer welfare are not necessary opposing objectives, promote the analysis on policy frameworks that could benefit both firms and consumer and promote economic efficiencies. This chapter lays a robust foundation for future research endeavors. It invites subsequent investigations to incorporate more sophisticated structural models to endogenize strategic consumer and firm behavior, thereby enabling a deeper exploration of the intricate market structures and dynamics.

Appendix A

Supplementary Tables and Figures

Table A.1: User Types Classification Results (by ρ)

		User types				Total
		no search	*p search*	*x search*	*both search*	
$\rho = 10$ min	count	2,693,867	319,130	1,061,114	1,572,983	5,647,094
	percentage	47.70%	5.65%	18.79%	27.85%	100.00%
$\rho = 20$ min	count	2,713,295	299,702	1,125,683	1,508,414	5,647,094
	percentage	48.05%	5.31%	19.93%	26.71%	100.00%
$\rho = 30$ min	count	2,722,085	290,912	1,151,547	1,482,550	5,647,094
	percentage	48.20%	5.15%	20.39%	26.25%	100.00%
$\rho = 40$ min	count	2,728,129	284,868	1,167,688	1,466,409	5,647,094
	percentage	48.31%	5.04%	20.68%	25.97%	100.00%
$\rho = 50$ min	count	2,732,897	280,100	1,179,200	1,454,897	5,647,094
	percentage	48.39%	4.96%	20.88%	25.76%	100.00%
$\rho = 1$ hour	count	2,736,748	276,249	1,188,517	1,445,580	5,647,094
	percentage	48.46%	4.89%	21.05%	25.60%	100.00%
$\rho = 2$ hour	count	2,752,736	260,261	1,223,647	1,410,450	5,647,094
	percentage	48.75%	4.61%	21.67%	24.98%	100.00%
$\rho = 3$ hour	count	2,763,396	249,601	1,245,288	1,388,809	5,647,094
	percentage	48.93%	4.42%	22.05%	24.59%	100.00%

Table A.2: View Events Types Classification Results (by μ)

	All view events				Smartphone view events			
	click *to page*		click *on page*		click *to page*		click *on page*	
	count	perc.	count	perc.	count	perc.	count	perc.
$\mu = 1$ min	89,946,979	84.05%	17,065,502	15.94%	28,023,170	83.70%	5,455,032	16.29%
$\mu = 2$ min	86,170,926	80.52%	20,841,555	19.47%	26,635,352	79.56%	6,842,850	20.43%
$\mu = 3$ min	85,034,369	79.46%	21,978,112	20.53%	26,181,641	78.20%	7,296,561	21.79%
$\mu = 4$ min	84,528,525	78.98%	22,483,956	21.01%	25,974,143	77.58%	7,504,059	22.41%
$\mu = 5$ min	84,243,261	78.72%	22,769,220	21.27%	25,855,117	77.22%	7,623,085	22.77%
$\mu = 6$ min	84,059,033	78.55%	22,953,448	21.44%	25,777,277	76.99%	7,700,925	23.00%
$\mu = 7$ min	83,930,074	78.43%	23,082,407	21.56%	25,721,970	76.83%	7,756,232	23.16%
$\mu = 8$ min	83,833,792	78.34%	23,178,689	21.65%	25,680,236	76.70%	7,797,966	23.29%
$\mu = 9$ min	83,757,718	78.26%	23,254,763	21.73%	25,647,023	76.60%	7,831,179	23.39%
$\mu = 10$ min	83,696,085	78.21%	23,316,396	21.78%	25,619,873	76.52%	7,858,329	23.47%
$\mu = 11$ min	83,644,784	78.16%	23,367,697	21.83%	25,597,240	76.45%	7,880,962	23.54%
$\mu = 12$ min	83,601,180	78.12%	23,411,301	21.87%	25,578,141	76.40%	7,900,061	23.59%
$\mu = 13$ min	83,563,842	78.08%	23,448,639	21.91%	25,561,906	76.35%	7,916,296	23.64%
$\mu = 14$ min	83,530,932	78.05%	23,481,549	21.94%	25,547,854	76.31%	7,930,348	23.68%
$\mu = 15$ min	83,501,853	78.03%	23,510,628	21.96%	25,535,677	76.27%	7,942,525	23.72%
$\mu = 16$ min	83,475,138	78.00%	23,537,343	21.99%	25,524,164	76.24%	7,954,038	23.75%
$\mu = 17$ min	83,450,358	77.98%	23,562,123	22.01%	25,513,346	76.20%	7,964,856	23.79%
$\mu = 18$ min	83,427,609	77.96%	23,584,872	22.03%	25,503,553	76.17%	7,974,649	23.82%
$\mu = 19$ min	83,406,856	77.94%	23,605,625	22.05%	25,494,577	76.15%	7,983,625	23.84%
$\mu = 20$ min	83,387,402	77.92%	23,625,079	22.07%	25,486,091	76.12%	7,992,111	23.87%

Table A.3: Comparison of Models

	Model I	Model II	Model III	Model IV	Model V
Model specification	Vanilla Logit	w/ user type indicators	w/ user type indicators and sub-types	MLP at individual level w/ sub-types	RNN at individual level w/ sub-types
Observed heterogeneity	None	4 user types	4 user types	User-specific features generated from browsing history	User-specific sequential browsing history
Unobserved heterogeneity	None	None	2 sub-types at user type level	2 sub-types at user level	2 sub-types at user level
Result table/figure	Table 1.2	Table 1.3	Table 1.4	Figure 1.6	Figure 1.11
# of parameters	24	96	196	2,161	1,359
LL	−2,221,949.5	−1,712,007.5	−1,274,381.4	−1,205,676.6	−1,200,870.6
LL per case	−0.525	−0.404	−0.301	−0.285	−0.284
AIC	4,443,947.0	3,424,207.0	2,548,954.8	2,415,675.2	2,404,460.2
BIC	4,444,265.2	3,425,479.9	2,550,227.6	2,444,327.9	2,422,479.2
Holdout LL	−742,270.8	−570,629.7	−425,437.3	−403,102.1	−400,974.7
Holdout LL per case	−0.526	−0.404	−0.301	−0.286	−0.284
LR test	Baseline	Rejects Model I	Rejects Model I and II	Rejects Model I, II, and III	Rejects Model I, II, and III. Not applicable to test against Model IV.
Elasticity result	Figure A.6	Figure A.7	Figure A.8	Figure A.9	Figure 1.15
Produce realistic elasticities	No	No	Yes	Yes	Yes

All LR test rejections are at 0.001 significance level. Realistic elasticities determined based on smell test.

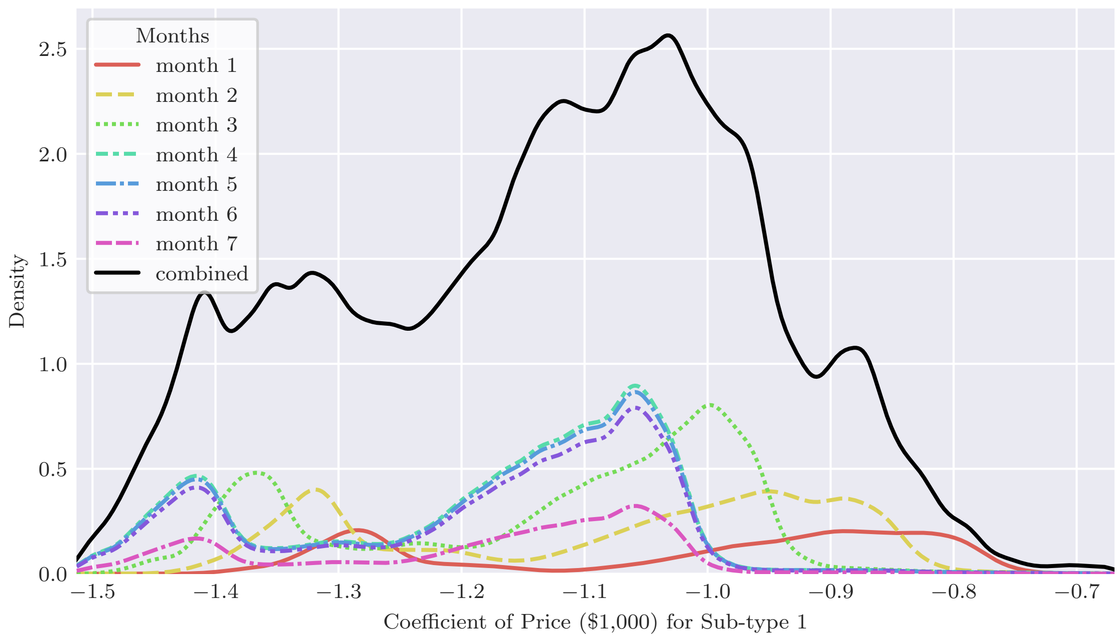


Figure A.1: Distribution of Price Coefficient of Model IV - MLP (Sub-type 1, by Active Months)

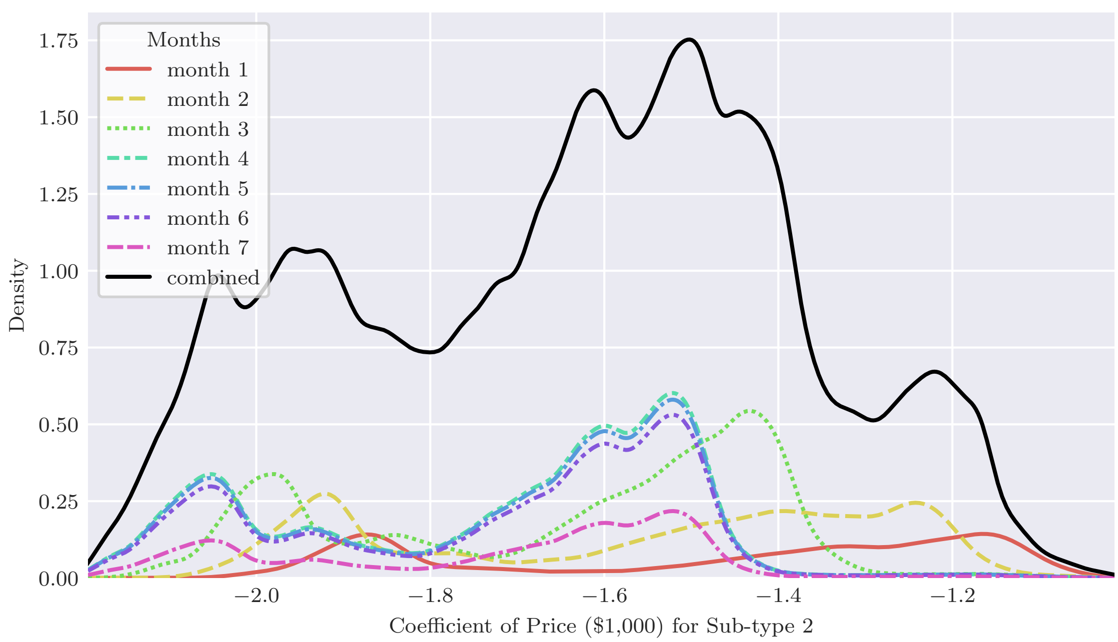


Figure A.2: Distribution of Price Coefficient of Model IV - MLP (Sub-type 2, by Active Months)

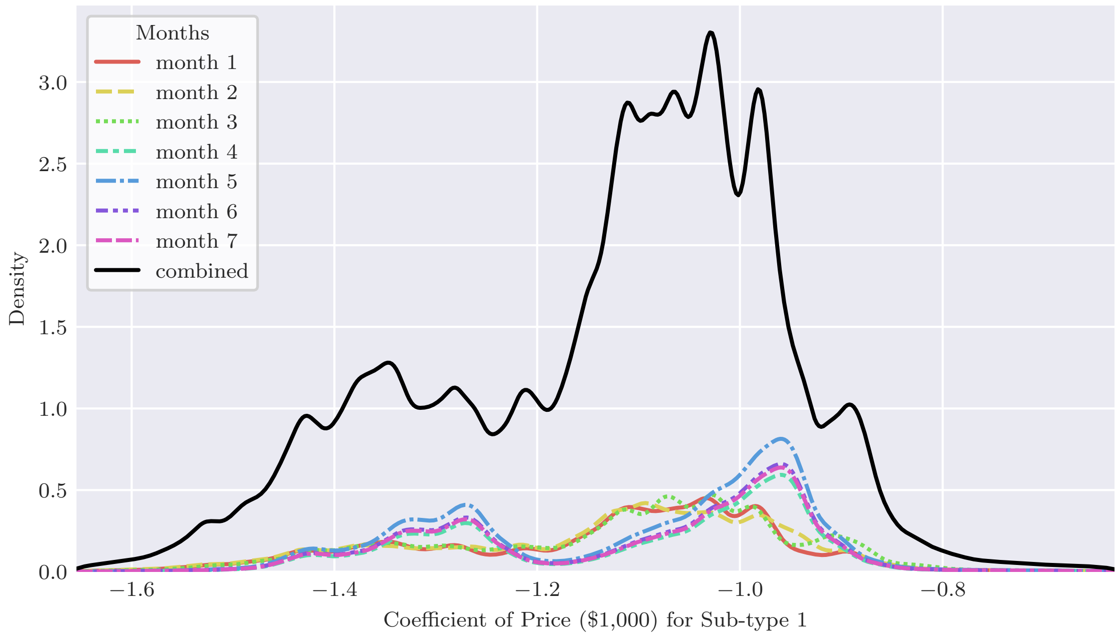


Figure A.3: Distribution of Price Coefficient of Model V - RNN (Sub-type 1, by Active Months)

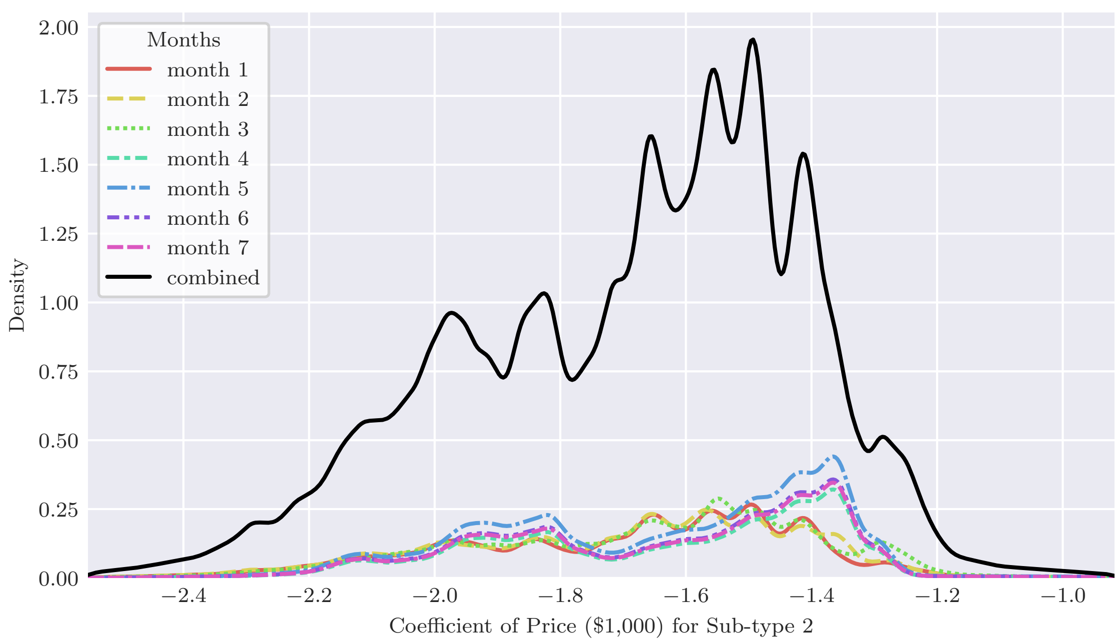
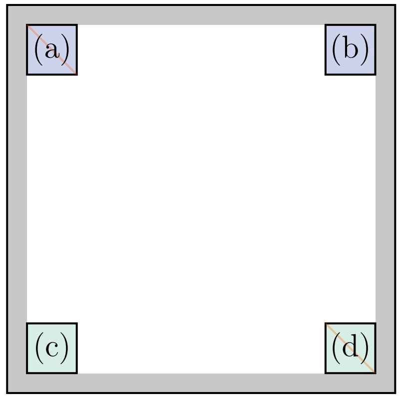


Figure A.4: Distribution of Price Coefficient of Model V - RNN (Sub-type 2, by Active Months)



The outermost square depicted outlines the full spectrum of cross- and own-price elasticities, organized with both the rows and columns arranged in ascending order according to the prices of smartphones. Given that a product's price can fluctuate over time, the median price per event is utilized for this sorting process. This outermost matrix, measuring approximately 1,700 by 1,700, represents the entirety of smartphones included in the dataset. A shaded area within this matrix highlights the exclusion of the 50 most affordable and 50 most expensive smartphones. Enclosed within this overarching square are four smaller squares, labeled from (a) to (d), each showcasing the subsequent 100 most affordable and 100 most expensive smartphones. The direct concatenation of them yields the price elasticities presented in Figure 1.15, A.6, A.7, A.8, and A.9.

Figure A.5: Illustration of Price Elasticities Heatmap

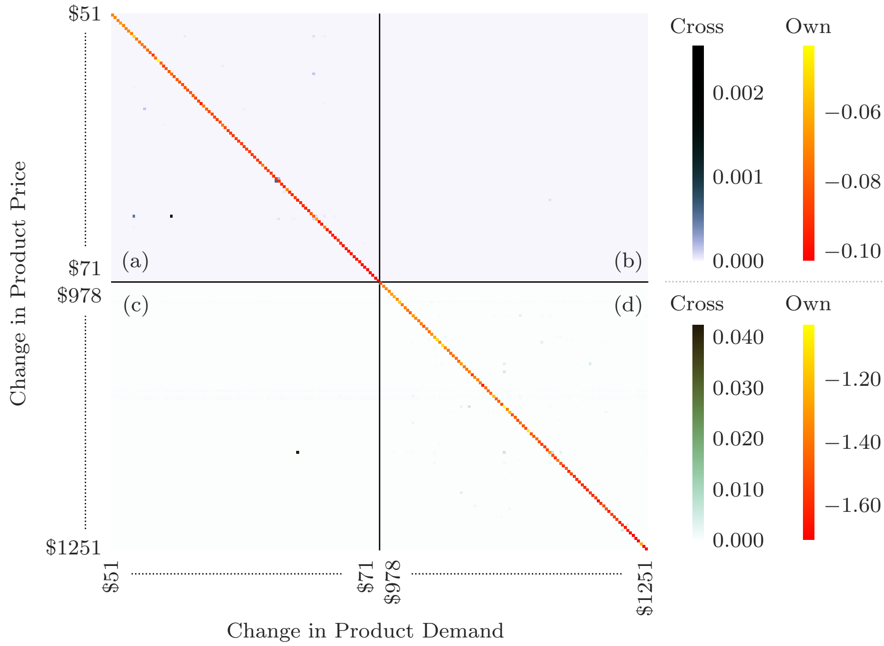


Figure A.6: Price Elasticities from Model I - Baseline Logit

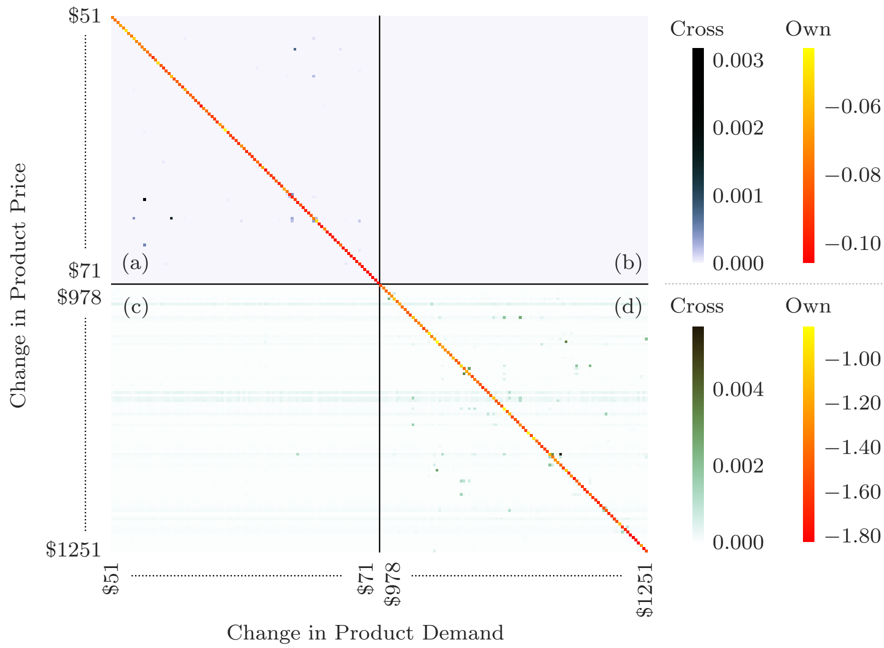


Figure A.7: Price Elasticities from Model II - by User Types

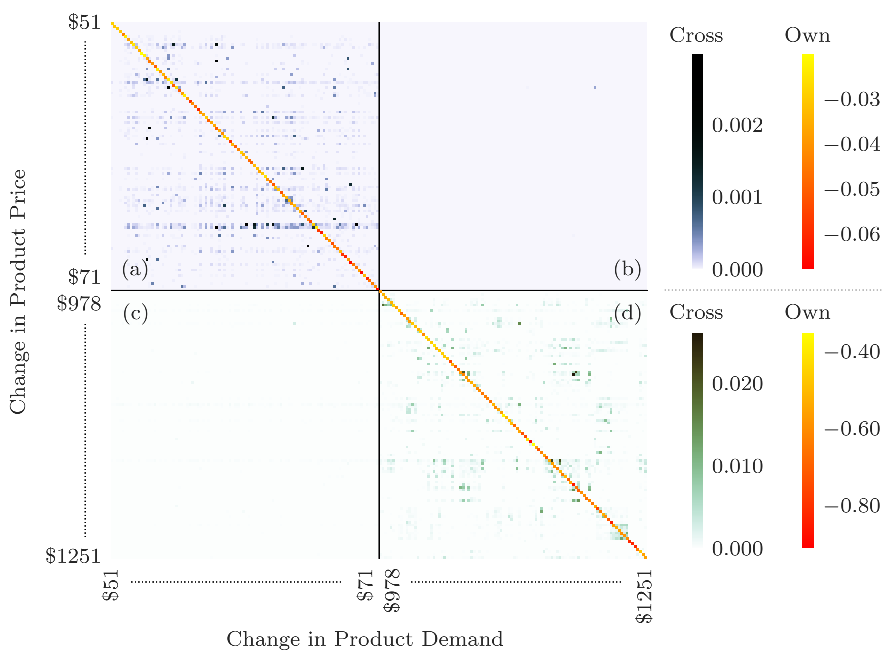


Figure A.8: Price Elasticities from Model III - with Unobserved Heterogeneity

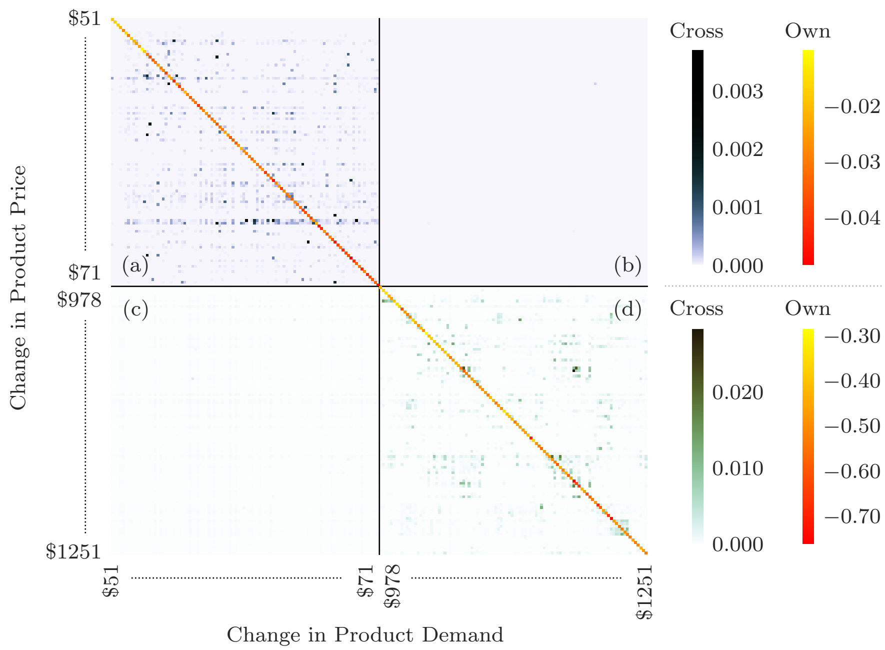


Figure A.9: Price Elasticities from Model IV - MLP

Appendix B

Derivation of Conditional Probability of Purchase After Discount and Expected Consumer Surplus

B.1 Conditional Probability of Purchase After Discount

For user n with sub-type s, conditional on choosing the outside option before any discount, the probability of choosing smartphone $k \in \mathcal{A}_{t(n)}$ with a discount $d_{n,k} > 0$ given for smartphone k is as follows,

$$\begin{aligned}
&\tilde{\mathsf{P}}_{n,s,k}(d_{n,k}) \\
\equiv&\text{Prob}\Big(\tilde{U}_{n,s,k}(d_{n,k}) > U_{n,s,i},\ \forall i \in \mathcal{C}_n \setminus \{k\} \,\Big|\, U_{n,s,0} > U_{n,s,j},\ \forall j \in \mathcal{C}_n \setminus \{0\}\Big) \quad \text{(B.1)}\\
=&1 - \text{Prob}\Big(U_{n,s,0} > U_{n,s,i},\ \forall i \in \mathcal{C}_n \setminus \{0\} \wedge U_{n,s,0} > \tilde{U}_{n,s,k}(d_{n,k}) \\
&\qquad \Big|\, U_{n,s,0} > U_{n,s,j},\ \forall j \in \mathcal{C}_n \setminus \{0\}\Big) \quad \text{(B.2)}
\end{aligned}$$

where $\tilde{U}_{n,s,k}(d_{n,k}) = U_{n,s,k} - \alpha_{n,s} \cdot d_{n,k}$ is the new utility from smartphone k after the discount $d_{n,k}$.

Note that $\tilde{U}_{n,s,k}(d_{n,k}) > U_{n,s,k}$, since $d_{n,k} > 0$ and $\alpha_{n,s} < 0$. Hence, $U_{n,s,0} > \tilde{U}_{n,s,k}(d_{n,k})$ implies $U_{n,s,0} > U_{n,s,k}$. The above equation reduces to

$$=1 - \frac{\text{Prob}\Big(U_{n,s,0} > U_{n,s,i},\ \forall i \in \mathcal{C}_n \setminus \{0,k\} \wedge U_{n,s,0} > \tilde{U}_{n,s,k}(d_{n,k})\Big)}{\text{Prob}\Big(U_{n,s,0} > U_{n,s,j},\ \forall j \in \mathcal{C}_n \setminus \{0\}\Big)} \quad \text{(B.3)}$$

$$=1 - \frac{\dfrac{\exp(u_{n,s,0})}{\exp(\tilde{u}_{n,s,k}(d_{n,k})) + \sum_{j \in \mathcal{C}_n \setminus \{k\}} \exp(u_{n,s,j})}}{\dfrac{\exp(u_{n,s,0})}{\sum_{j \in \mathcal{C}_n} \exp(u_{n,s,j})}} \quad \text{(B.4)}$$

$$=1-\frac{\sum_{j\in\mathcal{C}_n}\exp(u_{n,s,j})}{\exp(\tilde{u}_{n,s,k}(d_{n,k}))+\sum_{j\in\mathcal{C}_n\setminus\{k\}}\exp(u_{n,s,j})} \tag{B.5}$$

$$=\frac{\exp(\tilde{u}_{n,s,k}(d_{n,k}))-\exp(u_{n,s,k})}{\exp(\tilde{u}_{n,s,k}(d_{n,k}))+\sum_{j\in\mathcal{C}_n\setminus\{k\}}\exp(u_{n,s,j})} \tag{B.6}$$

$$=\frac{\Delta_{n,s,k}(d_{n,k})}{\Delta_{n,s,k}(d_{n,k})+\sum_{j\in\mathcal{C}_n}\exp(u_{n,s,j})} \tag{2.7}$$

where $\Delta_{n,s,k}(d_{n,k}) = \exp(\tilde{u}_{n,s,k}(d_{n,k}))-\exp(u_{n,s,k}) = \exp(u_{n,s,k})\cdot[\exp(-\alpha_{n,s}\cdot d_{n,k})-1]$.

B.2 Conditional Expected Consumer Surplus

Derivation of the conditional expected consumer surplus (relative to choosing the outside option) is shown below.

B.2.1 Before Discount

The firm observes user's choice decisions before any discount is provided. Either user n purchased smartphone $q \in \mathcal{A}_{t(n)}$ or chose the outside option 0.

If the outside option 0 was chosen, then $\mathsf{CS}_{n,s} = 0$.

If smartphone q was purchased, then it must be that q yields the highest utility level. Hence, the expected conditional expected consumer surplus for user n of subtype s is

$$\mathsf{CS}_{n,s} = \frac{1}{-\alpha_{n,s}}\mathbb{E}\Big[U_{n,s,q}-U_{n,s,0}\Big|U_{n,s,q}>U_{n,s,j},\forall j\in\mathcal{C}_n\setminus\{q\}\Big] \tag{2.10}$$

$$= \frac{1}{-\alpha_{n,s}}\mathbb{E}\Big[u_{n,s,q}+\epsilon_{n,q}-u_{n,s,0}-\epsilon_{n,0}\Big|u_{n,s,q}+\epsilon_{n,q}>u_{n,s,j}+\epsilon_{n,j},\forall j\in\mathcal{C}_n\setminus\{q\}\Big] \tag{B.7}$$

$$= \frac{1}{-\alpha_{n,s}}\big(u_{n,s,q}-u_{n,s,0}\big)+ \frac{1}{-\alpha_{n,s}}\mathbb{E}\Big[\epsilon_{n,q}-\epsilon_{n,0}\Big|\epsilon_{n,j}<u_{n,s,q}+\epsilon_{n,q}-u_{n,s,j},\forall j\in\mathcal{C}_n\setminus\{q\}\Big]. \tag{B.8}$$

To derive the expectation term, first assume that $\epsilon_{n,q}$ is known. Hence,

$$\text{Prob}\Big(\epsilon_{n,q} - \epsilon_{n,0} < x \Big| \epsilon_{n,j} < \epsilon_{n,q} + u_{n,s,q} - u_{n,s,j}, \forall j \in \mathcal{C}_n \setminus \{q\}, \epsilon_{n,q}\Big) \tag{B.9}$$

$$= \frac{\text{Prob}\Big(\epsilon_{n,q} - x < \epsilon_{n,0} \wedge \epsilon_{n,j} < \epsilon_{n,q} + u_{n,s,q} - u_{n,s,j}, \forall j \in \mathcal{C}_n \setminus \{q\}, \epsilon_{n,q}\Big)}{\text{Prob}\Big(\epsilon_{n,j} < \epsilon_{n,q} + u_{n,s,q} - u_{n,s,j}, \forall j \in \mathcal{C}_n \setminus \{q\} \Big| \epsilon_{n,q}\Big)} \tag{B.10}$$

$$= \frac{\big[F(\epsilon_{n,q} + u_{n,s,q} - u_{n,s,0}) - F(\epsilon_{n,q} - x)\big] \cdot \prod_{j \in \mathcal{C}_n \setminus \{0,q\}} F(\epsilon_{n,q} + u_{n,s,q} - u_{n,s,j})}{\big[F(\epsilon_{n,q} + u_{n,s,q} - u_{n,s,0})\big] \cdot \prod_{j \in \mathcal{C}_n \setminus \{0,q\}} F(\epsilon_{n,q} + u_{n,s,q} - u_{n,s,j})} \tag{B.11}$$

$$= \frac{F(\epsilon_{n,q} + u_{n,s,q} - u_{n,s,0}) - F(\epsilon_{n,q} - x)}{F(\epsilon_{n,q} + u_{n,s,q} - u_{n,s,0})} \tag{B.12}$$

$$= 1 - \frac{F(\epsilon_{n,q} - x)}{F(\epsilon_{n,q} + u_{n,s,q} - u_{n,s,0})} \tag{B.13}$$

for $x > u_{n,s,0} - u_{n,s,q}$ and 0 otherwise, where $F(t) = \exp\big(-\exp(-t)\big)$ is the cdf of a standard Extreme Value Type-I distribution. Hence, integrating out $\epsilon_{n,q}$, yields

$$\int_{-\infty}^{\infty} \left[1 - \frac{F(s-x)}{F(s + u_{n,s,q} - u_{n,s,0})}\right] \cdot e^{-s} \cdot e^{-e^{-s}}\, ds \tag{B.14}$$

$$= 1 - \int_{-\infty}^{\infty} \exp\left[-e^{-(s-x)} + e^{-(s+u_{n,s,q}-u_{n,s,0})}\right] \cdot e^{-s} \cdot e^{-e^{-s}}\, ds \tag{B.15}$$

$$= 1 - \int_{-\infty}^{\infty} \exp\left[-e^{-(s-x)} + e^{-(s+u_{n,s,q}-u_{n,s,0})} - e^{-s}\right] \cdot e^{-s}\, ds \tag{B.16}$$

$$= 1 - \int_{-\infty}^{\infty} \exp\left[-e^{-s}\big(e^{x} - e^{-u_{n,s,q}+u_{n,s,0}} + 1\big)\right] \cdot e^{-s}\, ds. \tag{B.17}$$

Substitute $t = e^{-s}$, $dt = -e^{-s}ds$, and $\lambda = e^{x} - e^{-u_{n,s,q}+u_{n,s,0}} + 1$, yields

$$= 1 + \int_{\infty}^{0} \exp[-t \cdot \lambda]\, dt = 1 + \left[-\frac{e^{-t\cdot\lambda}}{\lambda}\right]_{t=\infty}^{0} \tag{B.18}$$

$$= 1 - \frac{1}{\lambda} = 1 - \frac{1}{e^{x} - e^{-u_{n,s,q}+u_{n,s,0}} + 1}. \tag{B.19}$$

Hence, the conditional pdf of $\epsilon_{n,q} - \epsilon_{n,0}$ is

$$\frac{e^{x}}{\left(e^{x} - e^{-u_{n,s,q}+u_{n,s,0}} + 1\right)^2} \quad \text{for } x > u_{n,s,0} - u_{n,s,q} \text{ and 0 otherwise.} \tag{B.20}$$

Therefore,

$$\mathsf{CS}_{n,s} = \frac{1}{-\alpha_{n,s}}\big(u_{n,s,q} - u_{n,s,0}\big) + \frac{1}{-\alpha_{n,s}} \int_{u_{n,s,0}-u_{n,s,q}}^{\infty} x \cdot \frac{e^x}{\left(e^x - e^{-u_{n,s,q}+u_{n,s,0}} + 1\right)^2}\, dx. \tag{B.21}$$

If $-u_{n,s,q} + u_{n,s,0} = 0$, then the above reduces to

$$\mathsf{CS}_{n,s} = \frac{1}{-\alpha_{n,s}} \int_0^{\infty} x \cdot \frac{1}{e^x}\, dx = \frac{1}{-\alpha_{n,s}}. \tag{B.22}$$

Otherwise,

$$\begin{aligned}\mathsf{CS}_{n,s} &= \frac{1}{-\alpha_{n,s}}\big(u_{n,s,q} - u_{n,s,0}\big) \\ &\quad + \frac{1}{-\alpha_{n,s}}\left[\frac{\log(|e^x - e^{-u_{n,s,q}+u_{n,s,0}} + 1|) - x}{e^{-u_{n,s,q}+u_{n,s,0}} - 1} - \frac{x}{e^x - e^{-u_{n,s,q}+u_{n,s,0}} + 1}\right]_{x=u_{n,s,0}-u_{n,s,q}}^{\infty}\end{aligned} \tag{B.23}$$

$$= \frac{1}{-\alpha_{n,s}}\big(u_{n,s,q} - u_{n,s,0}\big) + \frac{1}{-\alpha_{n,s}}\left[0 + \frac{u_{n,s,0} - u_{n,s,q}}{e^{-u_{n,s,q}+u_{n,s,0}} - 1} + \big(u_{n,s,0} - u_{n,s,q}\big)\right] \tag{B.24}$$

$$= \frac{1}{-\alpha_{n,s}} \cdot \frac{u_{n,s,q} - u_{n,s,0}}{1 - e^{-u_{n,s,q}+u_{n,s,0}}}. \tag{B.25}$$

Since $u_{n,s,0}$ is normalized to 0, hence,

$$\mathsf{CS}_{n,s} = \begin{cases} 0 & \text{if } y_{n,0} = 1 \\ \dfrac{1}{-\alpha_{n,s}} & \text{if } y_{n,0} = 0, u_{n,s,q} = 0, \\ \dfrac{1}{-\alpha_{n,s}} \cdot \dfrac{u_{n,s,q}}{1 - e^{-u_{n,s,q}}} & \text{if } y_{n,0} = 0, u_{n,s,q} \neq 0. \end{cases} \tag{2.11}$$

B.2.2 After Discount

Let user n be a user who choose the outside option 0 when no discounts are given. The choice outcome for user n is uncertain after the discount $d_{n,k} > 0$ on smartphone k, since econometricians do not observe that. Conditional on the observed data, either user n of sub-type s will choose smartphone k with a probability of $\tilde{\mathsf{P}}_{n,s,k}(d_{n,k})$, or continues to choose the outside option 0 with a probability of $1 - \tilde{\mathsf{P}}_{n,s,k}(d_{n,k})$. The former yields a gain in conditional expected consumer surplus, whereas the latter does not. Hence, the gain in the conditional expected consumer surplus for user n of sub-type s is

$$\begin{aligned}
&\overset{\triangle}{\mathsf{CS}}_{n,s}\\
&\equiv \tilde{\mathsf{P}}_{n,s,k}(d_{n,k})\cdot\frac{1}{-\alpha_{n,s}}\mathbb{E}\Big[\tilde{U}_{n,s,k}(d_{n,k})-U_{n,s,0}\\
&\qquad\qquad\Big|\tilde{U}_{n,s,k}(d_{n,k})>U_{n,s,0}\wedge U_{n,s,0}>U_{n,s,j},\forall j\in\mathcal{C}_n\setminus\{0\}\Big] \qquad \text{(B.26)}\\
&= \tilde{\mathsf{P}}_{n,s,k}(d_{n,k})\cdot\frac{1}{-\alpha_{n,s}}\mathbb{E}\Big[\tilde{u}_{n,s,k}(d_{n,k})+\epsilon_{n,k}-u_{n,s,0}-\epsilon_{n,0}\Big|\tilde{u}_{n,s,k}(d_{n,k})+\epsilon_{n,k}>u_{n,s,0}+\epsilon_{n,0}\\
&\qquad\qquad\wedge\ u_{n,s,0}+\epsilon_{n,0}>u_{n,s,j}+\epsilon_{n,j},\forall j\in\mathcal{C}_n\setminus\{0\}\Big] \qquad \text{(B.27)}\\
&= \tilde{\mathsf{P}}_{n,s,k}(d_{n,k})\cdot\frac{1}{-\alpha_{n,s}}E_{n,s,k}(d_{n,k}) \qquad \text{(2.13)}
\end{aligned}$$

where $E_{n,s,k}(d_{n,k})$ denotes the expectation term in (B.26), or (B.27).

To derive the expectation term, first assume that $\epsilon_{n,0}$ is known, and derive the conditional cdf as follows,

$$\begin{aligned}
&\mathrm{Prob}\Big(\tilde{u}_{n,s,k}(d_{n,k})+\epsilon_{n,k}-u_{n,s,0}-\epsilon_{n,0}<x\Big|\tilde{u}_{n,s,k}(d_{n,k})+\epsilon_{n,k}>u_{n,s,0}+\epsilon_{n,0}\\
&\qquad\qquad\wedge\ u_{n,s,0}+\epsilon_{n,0}>u_{n,s,j}+\epsilon_{n,j},\forall j\in\mathcal{C}_n\setminus\{0\},\epsilon_{n,0}\Big) \qquad \text{(B.28)}\\
&= \mathrm{Prob}\Big(\epsilon_{n,k}<x+u_{n,s,0}-\tilde{u}_{n,s,k}(d_{n,k})+\epsilon_{n,0}\Big|u_{n,s,0}-\tilde{u}_{n,s,k}(d_{n,k})+\epsilon_{n,0}<\epsilon_{n,k}\\
&\qquad\qquad\wedge\ \epsilon_{n,j}<u_{n,s,0}-u_{n,s,j}+\epsilon_{n,0},\forall j\in\mathcal{C}_n\setminus\{0\},\epsilon_{n,0}\Big). \qquad \text{(B.29)}
\end{aligned}$$

Note that, the above is 0 if $x \le 0$, and is 1 if $\tilde{u}_{n,s,k}(d_{n,k}) - u_{n,s,k} = -\alpha_{n,s,k}\cdot d_{n,k} \le x$. Hence, for $0 < x < -\alpha_{n,s,k}\cdot d_{n,k}$, the conditional cdf assuming a known $\epsilon_{n,0}$ is

$$= \frac{\begin{aligned}&\mathrm{Prob}\Big(u_{n,s,0}-\tilde{u}_{n,s,k}(d_{n,k})+\epsilon_{n,0}<\epsilon_{n,k}<x+u_{n,s,0}-\tilde{u}_{n,s,k}(d_{n,k})+\epsilon_{n,0}\\&\qquad\wedge\ \epsilon_{n,j}<u_{n,s,0}-u_{n,s,j}+\epsilon_{n,0},\forall j\in\mathcal{C}_n\setminus\{0,k\},\epsilon_{n,0}\Big)\end{aligned}}{\begin{aligned}&\mathrm{Prob}\Big(u_{n,s,0}-\tilde{u}_{n,s,k}(d_{n,k})+\epsilon_{n,0}<\epsilon_{n,k}<u_{n,s,0}-u_{n,s,k}+\epsilon_{n,0}\\&\qquad\wedge\ \epsilon_{n,j}<u_{n,s,0}-u_{n,s,j}+\epsilon_{n,0},\forall j\in\mathcal{C}_n\setminus\{0,k\},\epsilon_{n,0}\Big)\end{aligned}} \qquad \text{(B.30)}$$

$$= \frac{\begin{array}{r}[F(x+u_{n,s,0}-\tilde{u}_{n,s,k}(d_{n,k})+\epsilon_{n,0})-F(u_{n,s,0}-\tilde{u}_{n,s,k}(d_{n,k})+\epsilon_{n,0})] \\ \cdot\prod_{j\in\mathcal{C}_n\setminus\{0,k\}}F(\epsilon_{n,0}+u_{n,s,0}-u_{n,s,k})\end{array}}{\begin{array}{r}[F(u_{n,s,0}-u_{n,s,k}+\epsilon_{n,0})-F(u_{n,s,0}-\tilde{u}_{n,s,k}(d_{n,k})+\epsilon_{n,0})] \\ \cdot\prod_{j\in\mathcal{C}_n\setminus\{0,k\}}F(\epsilon_{n,0}+u_{n,s,0}-u_{n,s,k})\end{array}} \tag{B.31}$$

$$= \frac{F(x+u_{n,s,0}-\tilde{u}_{n,s,k}(d_{n,k})+\epsilon_{n,0})-F(u_{n,s,0}-\tilde{u}_{n,s,k}(d_{n,k})+\epsilon_{n,0})}{F(u_{n,s,0}-u_{n,s,k}+\epsilon_{n,0})-F(u_{n,s,0}-\tilde{u}_{n,s,k}(d_{n,k})+\epsilon_{n,0})} \tag{B.32}$$

where $F(t) = \exp\big(-\exp(-t)\big)$ is the cdf of a standard Extreme Value Type-I distribution. Hence, integrating out $\epsilon_{n,0}$ yields the conditional cdf

$$\int_{-\infty}^{\infty} \frac{\exp\left[-e^{-(x+u_{n,s,0}-\tilde{u}_{n,s,k}+s)}\right]-\exp\left[-e^{-(u_{n,s,0}-\tilde{u}_{n,s,k}(d_{n,k})+s)}\right]}{\exp\left[-e^{-(u_{n,s,0}-u_{n,s,k}+s)}\right]-\exp\left[-e^{-(u_{n,s,0}-\tilde{u}_{n,s,k}(d_{n,k})+s)}\right]}\cdot e^{-s}\cdot e^{-e^{-s}}\,ds. \tag{B.33}$$

Unfortunately, this integral does not have a closed form expression. Denote

$$\begin{aligned} &E_{n,s,k}(d_{n,k},\epsilon_{n,0}) \\ &\equiv \mathbb{E}\Big[\tilde{U}_{n,s,k}(d_{n,k})-U_{n,s,0}\Big|\tilde{U}_{n,s,k}(d_{n,k})>U_{n,s,0}\wedge U_{n,s,0}>U_{n,s,j},\forall j\in\mathcal{C}_n\setminus\{0\},\epsilon_{n,0}\Big] \end{aligned} \tag{B.34}$$

To find $E_{n,s,k}(d_{n,k})$, one can either find the pdf corresponding the expectation (B.34) which assumes a known $\epsilon_{n,0}$ by differentiating (B.32), then apply the law of iterated expectations (LIE) to $E_{n,s,k}(d_{n,k},\epsilon_{n,0})$ to account for the unknown $\epsilon_{n,0}$; or alternatively, one can apply the Leibniz integral rule on (B.33) with proper substitution, e.g., $t=\frac{1}{1+e^{-s}}$, to find the pdf corresponding to $E_{n,s,k}(d_{n,k})$, directly without assuming a known $\epsilon_{n,0}$, and take the expectation with the Fubini's theorem. Both methods yield the same resulting integral. The following shows the first method.

Differentiating (B.32) yields

$$\frac{\partial}{\partial x}\left(\frac{F(x+u_{n,s,0}-\tilde{u}_{n,s,k}(d_{n,k})+\epsilon_{n,0})-F(u_{n,s,0}-\tilde{u}_{n,s,k}(d_{n,k})+\epsilon_{n,0})}{F(u_{n,s,0}-u_{n,s,k}+\epsilon_{n,0})-F(u_{n,s,0}-\tilde{u}_{n,s,k}(d_{n,k})+\epsilon_{n,0})}\right) \tag{B.35}$$

$$= \frac{\partial}{\partial x}\left(\frac{\exp\left[-e^{-(x+u_{n,s,0}-\tilde{u}_{n,s,k}(d_{n,k})+\epsilon_{n,0}}\right]-\exp\left[-e^{-(u_{n,s,0}-\tilde{u}_{n,s,k}(d_{n,k})+\epsilon_{n,0})}\right]}{\exp\left[-e^{-(u_{n,s,0}-u_{n,s,k}+\epsilon_{n,0})}\right]-\exp\left[-e^{-(u_{n,s,0}-\tilde{u}_{n,s,k}(d_{n,k})+\epsilon_{n,0})}\right]}\right) \tag{B.36}$$

$$= \frac{\exp\left[-e^{-(x+u_{n,s,0}-\tilde{u}_{n,s,k}(d_{n,k})+\epsilon_{n,0})}\right]\cdot e^{-(x+u_{n,s,0}-\tilde{u}_{n,s,k}(d_{n,k})+\epsilon_{n,0})}}{\exp\left[-e^{-(u_{n,s,0}-u_{n,s,k}+\epsilon_{n,0})}\right]-\exp\left[-e^{-(u_{n,s,0}-\tilde{u}_{n,s,k}(d_{n,k})+\epsilon_{n,0})}\right]}. \tag{B.37}$$

Take the expectation with the above pdf, yields

$$\begin{aligned}
&\int_0^{-\alpha_{n,s,k}\cdot d_{n,k}} x\cdot \frac{\exp\left[-e^{-(x+u_{n,s,0}-\tilde{u}_{n,s,k}(d_{n,k})+\epsilon_{n,0})}\right]\cdot e^{-(x+u_{n,s,0}-\tilde{u}_{n,s,k}(d_{n,k})+\epsilon_{n,0})}}{\exp\left[-e^{-(u_{n,s,0}-u_{n,s,k}+\epsilon_{n,0})}\right]-\exp\left[-e^{-(u_{n,s,0}-\tilde{u}_{n,s,k}(d_{n,k})+\epsilon_{n,0})}\right]}\,dx \\
&= \frac{1}{\exp\left[-e^{-(u_{n,s,0}-u_{n,s,k}+\epsilon_{n,0})}\right]-\exp\left[-e^{-(u_{n,s,0}-\tilde{u}_{n,s,k}(d_{n,k})+\epsilon_{n,0})}\right]} \\
&\quad \cdot \int_0^{-\alpha_{n,s,k}\cdot d_{n,k}} x\cdot\exp\left[-e^{-(x+u_{n,s,0}-\tilde{u}_{n,s,k}(d_{n,k})+\epsilon_{n,0})}\right]\cdot e^{-(x+u_{n,s,0}-\tilde{u}_{n,s,k}(d_{n,k})+\epsilon_{n,0})}\,dx. \qquad (\text{B.38})
\end{aligned}$$

The integral term on the RHS of (B.38) can be solved as follows

$$\int_a^b x\cdot\exp\left[-e^{-(x+m)}\right]\cdot e^{-(x+m)}\,dx \tag{B.39}$$

$$= \int_{e^a}^{e^b} \log(t)\cdot\exp\left[-\frac{1}{t}\cdot e^{-m}\right]\cdot e^{-m}\cdot\frac{1}{t^2}\,dt \tag{B.40}$$

$$\text{where } t=e^x, dt = t\,dx, m = u_{n,s,0}-\tilde{u}_{n,s,k}(d_{n,k})+\epsilon_{n,0}$$

$$= \left[\log(t)\cdot\exp\Big(-\frac{1}{t}\cdot e^{-m}\Big)\right]_{t=e^a}^{e^b} - \int_{e^a}^{e^b}\frac{1}{t}\cdot\exp\Big(-\frac{1}{t}\cdot e^{-m}\Big)\,dt \tag{B.41}$$

$$\text{integration by parts with } f=\log(t), dg=\exp\left[-\frac{1}{t}\cdot e^{-m}\right]\cdot e^{-m}\cdot\frac{1}{t^2}\,dt$$

$$= \left[\log(t)\exp\Big(-\frac{1}{t}\cdot e^{-m}\Big)\right]_{t=e^a}^{e^b} + \int_{-e^{-a-m}}^{-e^{-b-m}}\frac{e^v}{v}\,dv \tag{B.42}$$

$$\text{where } v=-\frac{1}{t}\cdot e^{-m}, dv = v^2\cdot e^m\,dt$$

$$= b\cdot\exp\Big(e^{-b-m}\Big) - a\cdot\exp\Big(e^{-a-m}\Big) + \mathrm{Ei}\left(-e^{-b-m}\right) - \mathrm{Ei}\left(-e^{-a-m}\right) \tag{B.43}$$

$$\text{where } \mathrm{Ei}(x) = \int_{-\infty}^{x}\frac{e^t}{t}\,dt \text{ is the exponential integral.}$$

Hence,

$$\begin{aligned}
&E_{n,s,k}(d_{n,k},\epsilon_{n,0}) \\
&\quad = \frac{1}{\exp\left[-e^{-(u_{n,s,0}-u_{n,s,k}+\epsilon_{n,0})}\right]-\exp\left[-e^{-(u_{n,s,0}-\tilde{u}_{n,s,k}(d_{n,k})+\epsilon_{n,0})}\right]}
\end{aligned}$$

$$\times \left[-\alpha_{n,s,k} \cdot d_{n,k} \cdot \exp\left(-e^{-(u_{n,s,0}-u_{n,s,k}+\epsilon_{n,0})}\right) + \mathrm{Ei}\left(-e^{-(u_{n,s,0}-u_{n,s,k}+\epsilon_{n,0})}\right) \right.$$

$$\left. - \mathrm{Ei}\left(-e^{-(u_{n,s,0}-\tilde{u}_{n,s,k}(d_{n,k})+\epsilon_{n,0})}\right) \right]. \tag{B.44}$$

Note that since $u_{n,s,0}$ is normalized to 0, the above expression simplifies to

$$\begin{aligned} E_{n,s,k}(d_{n,k}, \epsilon_{n,0}) = & \frac{1}{\exp\left[-e^{u_{n,s,k}-\epsilon_{n,0}}\right] - \exp\left[-e^{\tilde{u}_{n,s,k}(d_{n,k})-\epsilon_{n,0}}\right]} \\ & \times \left[-\alpha_{n,s,k} \cdot d_{n,k} \cdot \exp\left(-e^{u_{n,s,k}-\epsilon_{n,0}}\right) + \mathrm{Ei}\left(-e^{u_{n,s,k}-\epsilon_{n,0}}\right) \right. \\ & \left. - \mathrm{Ei}\left(-e^{\tilde{u}_{n,s,k}(d_{n,k})-\epsilon_{n,0}}\right) \right]. \end{aligned} \tag{2.16}$$

And the expectation term $E_{n,s,k}(d_{n,k})$ can be expressed as

$$E_{n,s,k}(d_{n,k}) = \int_{-\infty}^{\infty} E_{n,s,k}(d_{n,k}, \epsilon_{n,0}) \cdot e^{-s} \cdot e^{-e^{-s}} \, ds. \tag{2.15}$$

Note that $E_{n,s,k}(d_{n,k})$ does not have a closed form, but can be efficiently calculated via numerical methods, e.g., Monte Carlo or Gaussian quadrature. The expectation exists and is finite since $\tilde{U}_{n,s,k}(d_{n,k}) - U_{n,s,0}$ has a compact support, $\left(0, -\alpha_{n,s,k} \cdot d_{n,k}\right)$, which justifies using numerical methods. And, the gain in conditional expected consumer surplus attributable to the discount on smartphone k, can be calculated as

$$\overset{\triangle}{\mathsf{CS}}_{n,s} = \tilde{\mathsf{P}}_{n,s,k}(d_{n,k}) \cdot \frac{1}{-\alpha_{n,s}} E_{n,s,k}(d_{n,k}). \tag{B.45}$$

Appendix C

Numerical Stability of Calculating the Gain in Conditional Expected Consumer Surplus

When evaluating $E_{n,s,k}(d_{n,k}, \epsilon_{n,0})$ directly using (2.16), the calculation of the fraction term and the Ei function can be numerically unstable. This appendix provides a numerically stable and computationally efficient (with `float64`) algorithm to approximate the expression. Note that $E_{n,s,k}(d_{n,k}, \epsilon_{n,0})$ can be regarded as a function of $u_{n,s,k} - \epsilon_{n,0}$ and $-\alpha_{n,s,k} \cdot d_{n,k}$, since $\tilde{u}_{n,s,k}(d_{n,k}) - \epsilon_{n,0} = u_{n,s,k} - \epsilon_{n,0} - \alpha_{n,s,k} \cdot d_{n,k}$.

C.1 True Values

The true values of $E_{n,s,k}(d_{n,k}, \epsilon_{n,0})$ for a range of $u_{n,s,k} - \epsilon_{n,0}$ and $-\alpha_{n,s,k} \cdot d_{n,k}$ is plotted in Figure C.2 and C.1 evaluated using algorithms that allows mathematical calculations for arbitrary precision.

C.2 Unstable Algorithm

Directly evaluating $E_{n,s,k}(d_{n,k}, \epsilon_{n,0})$ using (2.16) may face numerical instability for certain values of $u_{n,s,k} - \epsilon_{n,0}$ and $-\alpha_{n,s,k} \cdot d_{n,k}$. Figure C.3 shows the absolute error of directly using `float64` and (2.16) compared with the true values. The gray shaded area indicates infinity values or division by zero errors that occur when $\exp\left[-e^{u_{n,s,k}-\epsilon_{n,0}}\right] \approx \exp\left[-e^{\tilde{u}_{n,s,k}(d_{n,k})-\epsilon_{n,0}}\right]$ or when $\mathrm{Ei}\left(-e^{u_{n,s,k}-\epsilon_{n,0}}\right)$ and $\mathrm{Ei}\left(-e^{\tilde{u}_{n,s,k}(d_{n,k})-\epsilon_{n,0}}\right)$ diverges to $-\infty$. This precisely matches the areas where

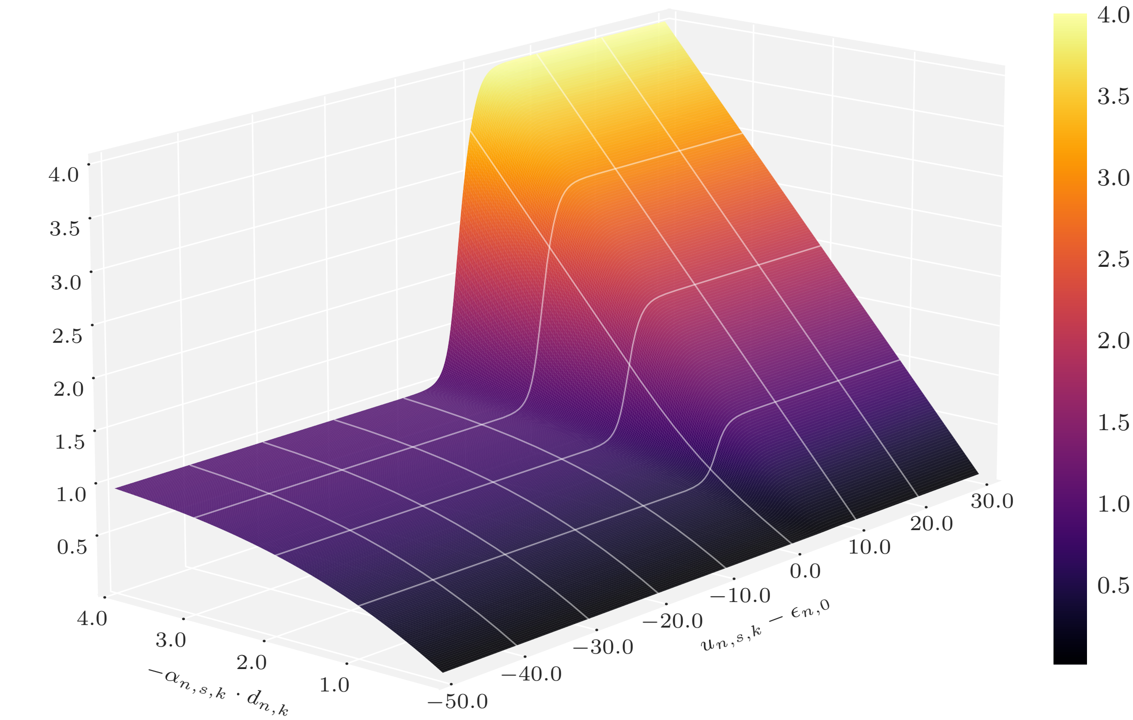


Figure C.1: Plot of $E_{n,s,k}(d_{n,k}, \epsilon_{n,0})$ (Surface)

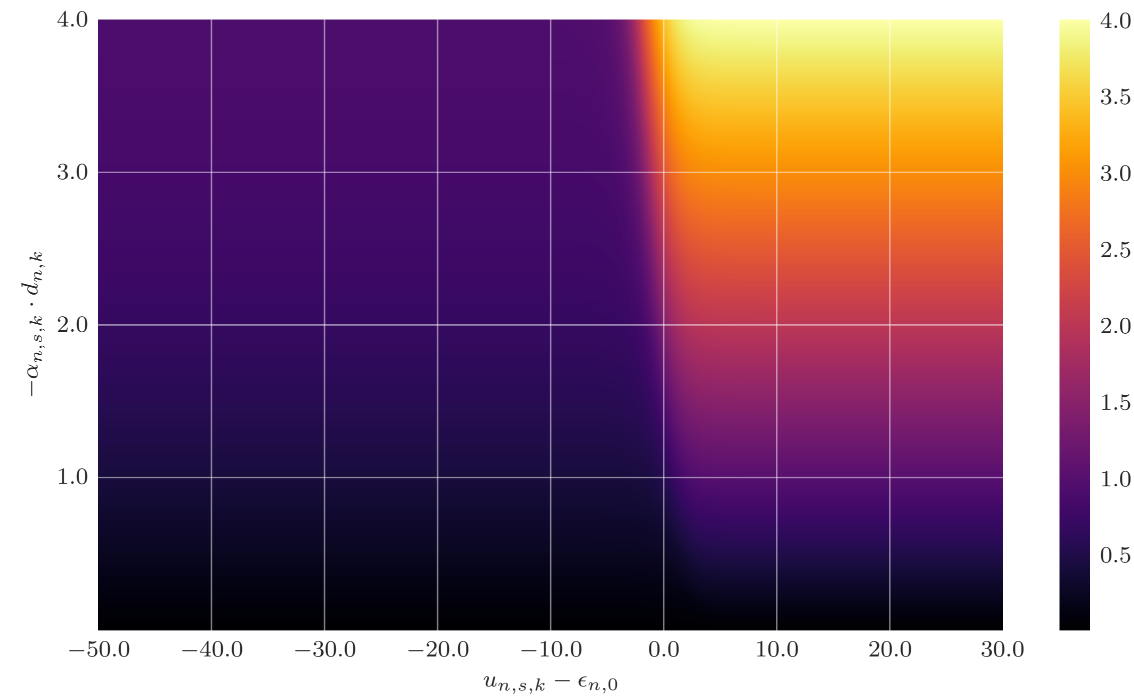


Figure C.2: Plot of $E_{n,s,k}(d_{n,k}, \epsilon_{n,0})$ (Heatmap)

$u_{n,s,k} - \epsilon_{n,0}$ is sufficiently away from 0.

Based on the absolute error, Figure C.3 is split into three regions by the red dotted lines, lower, central, and upper regions. The condition for the lower region is

$$-\alpha_{n,s,k} \cdot d_{n,k} \leq 5 \times 10^{-4} \times (u_{n,s,k} - \epsilon_{n,0} + 15)^4 \wedge u_{n,s,k} - \epsilon_{n,0} \leq 15, \tag{C.1}$$

the condition for upper region is

$$u_{n,s,k} - \epsilon_{n,0} \geq 6.5. \tag{C.2}$$

The rest are considered as the central region. And the distinct absolute errors across those three regions motivates the stable algorithm discussed below.

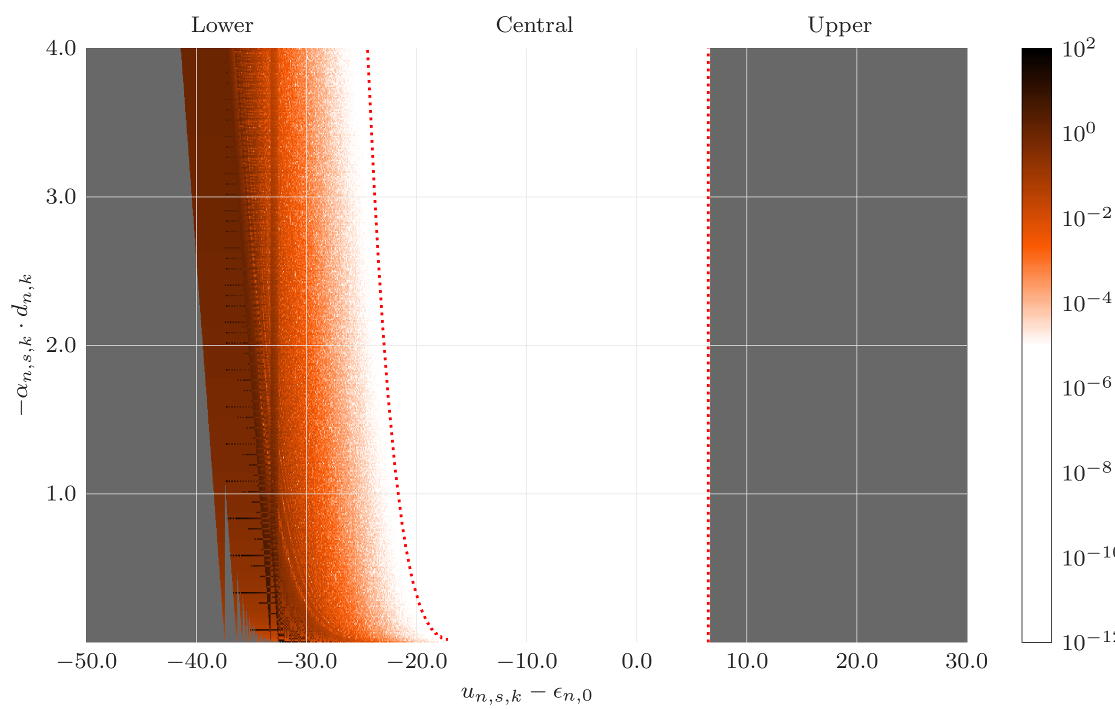


Figure C.3: Absolute Error of Numerically Unstable Algorithm

C.3 Stable Algorithm

C.3.1 Central Region

In the central region of Figure C.3, `float64` provides accurate values to the expectation, and is already numerically stable.

C.3.2 Lower Region

For the lower region, Figure C.2 and C.1 shows that the expectation has already converged for any given value of $-\alpha_{n,s,k} \cdot d_{n,k}$. Hence, the following limit provides accurate approximations.

$$\lim_{u_{n,s,k}-\epsilon_{n,0}\to-\infty} E_{n,s,k}(d_{n,k},\epsilon_{n,0}) = 1 - \alpha_{n,s,k} \cdot d_{n,k} \cdot \frac{e^{\alpha_{n,s,k}\cdot d_{n,k}}}{e^{\alpha_{n,s,k}\cdot d_{n,k}} - 1} \tag{C.3}$$

Proof. Let $y = \exp\left(-e^{u_{n,s,k}-\epsilon_{n,0}}\right)$, then

$$\begin{aligned}
&\lim_{u_{n,s,k}-\epsilon_{n,0}\to-\infty} E_{n,s,k}(d_{n,k},\epsilon_{n,0}) \\
&\quad = \lim_{y\to1} \frac{-\alpha_{n,s,k}\cdot d_{n,k}\cdot y + \mathrm{Ei}\left(\log\left(y\right)\right) - \mathrm{Ei}\left(\log\left(y^{e^{-\alpha_{n,s,k}\cdot d_{n,k}}}\right)\right)}{y - y^{e^{-\alpha_{n,s,k}\cdot d_{n,k}}}}.
\end{aligned} \tag{C.4}$$

By 6.2.6 and 6.6.2 of Olver et al. (2010), for $x < 0$,

$$\mathrm{Ei}(x) = -E_1(-x) = \gamma + \log(-x) + \sum_{n=1}^{\infty} \frac{x^n}{n!\,n} \tag{C.5}$$

where $\gamma \approx 0.57722$ is the Euler's Constant. Hence, the numerator of (C.4) has a limit of

$$\begin{aligned}
&\lim_{y\to1} -\alpha_{n,s,k}\cdot d_{n,k}\cdot y + \mathrm{Ei}\left(\log\left(y\right)\right) - \mathrm{Ei}\left(\log\left(y^{e^{-\alpha_{n,s,k}\cdot d_{n,k}}}\right)\right) \\
= &-\alpha_{n,s,k}\cdot d_{n,k} + \lim_{y\to1}\Bigg[\log\left(-\log\left(y\right)\right) + \sum_{n=1}^{\infty}\frac{\log\left(y\right)^n}{n!\,n} \\
&\qquad - \log\left(-\log\left(y^{e^{-\alpha_{n,s,k}\cdot d_{n,k}}}\right)\right) - \sum_{n=1}^{\infty}\frac{\log\left(y^{e^{-\alpha_{n,s,k}\cdot d_{n,k}}}\right)^n}{n!\,n}\Bigg]
\end{aligned} \tag{C.6}$$

$$= -\alpha_{n,s,k} \cdot d_{n,k} + \lim_{y\to 1}\left[-\log\left(e^{-\alpha_{n,s,k}\cdot d_{n,k}}\right) + \sum_{n=1}^{\infty} \frac{\log(y)^n \left(1 - e^{-\alpha_{n,s,k}\cdot d_{n,k}}\right)}{n!\, n}\right] \tag{C.7}$$

$$= \alpha_{n,s,k} \cdot d_{n,k} - \alpha_{n,s,k} \cdot d_{n,k} \tag{C.8}$$

$$= 0. \tag{C.9}$$

Note that the denominator of (C.4) also has a limit of 0, hence, apply L'Hôpital's rule yields

$$= \lim_{y\to 1} \frac{-\alpha_{n,s,k} \cdot d_{n,k} + \log(y)^{-1} + \log\left(y^{e^{-\alpha_{n,s,k}\cdot d_{n,k}}}\right)^{-1} \cdot e^{-\alpha_{n,s,k}\cdot d_{n,k}} \cdot y^{e^{-\alpha_{n,s,k}\cdot d_{n,k}}-1}}{1 - e^{-\alpha_{n,s,k}\cdot d_{n,k}} \cdot y^{e^{-\alpha_{n,s,k}\cdot d_{n,k}}-1}} \tag{C.10}$$

$$= \frac{-\alpha_{n,s,k} \cdot d_{n,k}}{1 - e^{-\alpha_{n,s,k}\cdot d_{n,k}}} + \lim_{y\to 1} \frac{\log(y)^{-1} + \log(y)^{-1} \cdot y^{e^{-\alpha_{n,s,k}\cdot d_{n,k}}-1}}{1 - e^{-\alpha_{n,s,k}\cdot d_{n,k}} \cdot y^{e^{-\alpha_{n,s,k}\cdot d_{n,k}}-1}} \tag{C.11}$$

$$= \frac{-\alpha_{n,s,k} \cdot d_{n,k}}{1 - e^{-\alpha_{n,s,k}\cdot d_{n,k}}} + \frac{1}{(1 - e^{-\alpha_{n,s,k}\cdot d_{n,k}})} \lim_{y\to 1} \frac{1 + y^{e^{-\alpha_{n,s,k}\cdot d_{n,k}}-1}}{\log(y)}. \tag{C.12}$$

Applying L'Hôpital's rule again to the limit term above,

$$= \frac{-\alpha_{n,s,k} \cdot d_{n,k}}{1 - e^{-\alpha_{n,s,k}\cdot d_{n,k}}} + \frac{1}{(1 - e^{-\alpha_{n,s,k}\cdot d_{n,k}})} \lim_{y\to 1} -\left(e^{-\alpha_{n,s,k}\cdot d_{n,k}} - 1\right) y^{e^{-\alpha_{n,s,k}\cdot d_{n,k}}-1} \tag{C.13}$$

$$= \frac{-\alpha_{n,s,k} \cdot d_{n,k}}{1 - e^{-\alpha_{n,s,k}\cdot d_{n,k}}} + \frac{-\left(e^{-\alpha_{n,s,k}\cdot d_{n,k}} - 1\right)}{(1 - e^{-\alpha_{n,s,k}\cdot d_{n,k}})} \tag{C.14}$$

$$= 1 - \frac{\alpha_{n,s,k} \cdot d_{n,k}}{1 - e^{-\alpha_{n,s,k}\cdot d_{n,k}}} \tag{C.15}$$

$$= 1 - \alpha_{n,s,k} \cdot d_{n,k} \cdot \frac{e^{\alpha_{n,s,k}\cdot d_{n,k}}}{e^{\alpha_{n,s,k}\cdot d_{n,k}} - 1}. \tag{C.16}$$

□

Therefore, for the lower region,

$$E_{n,s,k}(d_{n,k}, \epsilon_{n,0}) \approx 1 - \alpha_{n,s,k} \cdot d_{n,k} \cdot \frac{e^{\alpha_{n,s,k}\cdot d_{n,k}}}{e^{\alpha_{n,s,k}\cdot d_{n,k}} - 1} \tag{C.17}$$

and can be calculated in an efficient and numerically stable fashion with `float64`.

C.3.3 Upper Region

Rewrite (2.16) to

$$E_{n,s,k}(d_{n,k}, \epsilon_{n,0}) = \frac{(\tilde{u}_{n,s,k}(d_{n,k}) - u_{n,s,k}) \cdot \exp\left(-e^{u_{n,s,k}-\epsilon_{n,0}}\right)}{\exp\left[-e^{u_{n,s,k}-\epsilon_{n,0}}\right] - \exp\left[-e^{\tilde{u}_{n,s,k}(d_{n,k})-\epsilon_{n,0}}\right]} + \frac{\mathrm{Ei}\left(-e^{u_{n,s,k}-\epsilon_{n,0}}\right) - \mathrm{Ei}\left(-e^{\tilde{u}_{n,s,k}(d_{n,k})-\epsilon_{n,0}}\right)}{\exp\left[-e^{u_{n,s,k}-\epsilon_{n,0}}\right] - \exp\left[-e^{\tilde{u}_{n,s,k}(d_{n,k})-\epsilon_{n,0}}\right]}. \quad (C.18)$$

The first term in (C.18) can be rewritten as

$$= \frac{-\alpha_{n,s,k} \cdot d_{n,k}}{1 - \exp\left[e^{u_{n,s,k}-\epsilon_{n,0}} - e^{\tilde{u}_{n,s,k}(d_{n,k})-\epsilon_{n,0}}\right]} \quad (C.19)$$

$$= \frac{-\alpha_{n,s,k} \cdot d_{n,k}}{1 - \exp\left[e^{u_{n,s,k}-\epsilon_{n,0}}\left(1 - e^{-\alpha_{n,s,k} \cdot d_{n,k}}\right)\right]} \quad (C.20)$$

which is numerically stable to calculate for positive $u_{n,s,k} - \epsilon_{n,0}$. For the second term in (C.18), let

$$y_1 = \exp\left(-e^{u_{n,s,k}-\epsilon_{n,0}}\right), \text{ and } y_2 = \exp\left(-e^{\tilde{u}_{n,s,k}-\epsilon_{n,0}}\right). \quad (C.21)$$

Note that $y_1 \approx y_2$ for large positive values of $u_{n,s,k} - \epsilon_{n,0}$. Hence, rewriting the second term in (C.18) yields,

$$= \frac{\mathrm{Ei}\left(\log(y_1)\right) - \mathrm{Ei}\left(\log(y_2)\right)}{y_1 - y_2} \approx \frac{d}{dy}\,\mathrm{Ei}\left(\log(y)\right)\bigg|_{y=y_1} = \frac{1}{\log(y_1)} \quad (C.22)$$

which gives a good approximation.

Therefore, for the upper region,

$$E_{n,s,k}(d_{n,k}, \epsilon_{n,0}) \approx \frac{-\alpha_{n,s,k} \cdot d_{n,k}}{1 - \exp\left[e^{u_{n,s,k}-\epsilon_{n,0}}\left(1 - e^{-\alpha_{n,s,k} \cdot d_{n,k}}\right)\right]} + \frac{1}{-e^{u_{n,s,k}-\epsilon_{n,0}}} \quad (C.23)$$

and this can be calculated in an efficient and numerically stable fashion with `float64`. Note that the above is not a limit, albeit the limit exists. The derivation is as follows,

$$\lim_{u_{n,s,k}-\epsilon_{n,0} \to \infty} E_{n,s,k}(d_{n,k}, \epsilon_{n,0})$$

$$= \lim_{y_1 \to 0} \frac{-\alpha_{n,s,k} \cdot d_{n,k} \cdot y_1 + \mathrm{Ei}\left(\log\left(y_1\right)\right) - \mathrm{Ei}\left(\log\left(y_1^{e^{-\alpha_{n,s,k} \cdot d_{n,k}}}\right)\right)}{y_1 - y_1^{e^{-\alpha_{n,s,k} \cdot d_{n,k}}}} \tag{C.24}$$

$$= \lim_{y_1 \to 0} \frac{-\alpha_{n,s,k} \cdot d_{n,k}}{1 - y_1^{e^{-\alpha_{n,s,k} \cdot d_{n,k}} - 1}} + \lim_{y_1 \to 0} \frac{\mathrm{Ei}\left(\log\left(y_1\right)\right) - \mathrm{Ei}\left(\log\left(y_1^{e^{-\alpha_{n,s,k} \cdot d_{n,k}}}\right)\right)}{y_1 - y_1^{e^{-\alpha_{n,s,k} \cdot d_{n,k}}}} \tag{C.25}$$

$$= -\alpha_{n,s,k} \cdot d_{n,k} + \lim_{y_1 \to 0} \frac{\log\left(y_1\right)^{-1} - \log\left(y_1\right)^{-1} y_1^{e^{-\alpha_{n,s,k} \cdot d_{n,k}} - 1}}{1 - e^{-\alpha_{n,s,k} \cdot d_{n,k}} \cdot y_1^{e^{-\alpha_{n,s,k} \cdot d_{n,k}} - 1}} \tag{C.26}$$

$$= -\alpha_{n,s,k} \cdot d_{n,k} + \lim_{y_1 \to 0} \log\left(y_1\right)^{-1} \cdot \left(1 - y_1^{e^{-\alpha_{n,s,k} \cdot d_{n,k}} - 1}\right) \tag{C.27}$$

$$= -\alpha_{n,s,k} \cdot d_{n,k} + \lim_{y_1 \to 0} \log\left(y_1\right)^{-1} \tag{C.28}$$

$$= -\alpha_{n,s,k} \cdot d_{n,k}. \tag{C.29}$$

C.3.4 Approximation Error

By applying (C.17) and (C.23) to the lower and upper region of Figure C.3 respectively, the resulting absolute error is significantly reduced and numerically stable everywhere as shown in Figure C.4. The errors are the largest near the boarders of the three regions, but are still in the order of 10^{-5} or smaller. This shows that `float64` is capable to approximate $E_{n,s,k}(d_{n,k}, \epsilon_{n,0})$ sufficiently well using the stable algorithm described above, and is computationally efficiently. SciPy, CuPy, Cephes and many other packages all provide interfaces to accurately evaluate the exponential integral function with `float64` for the central region.

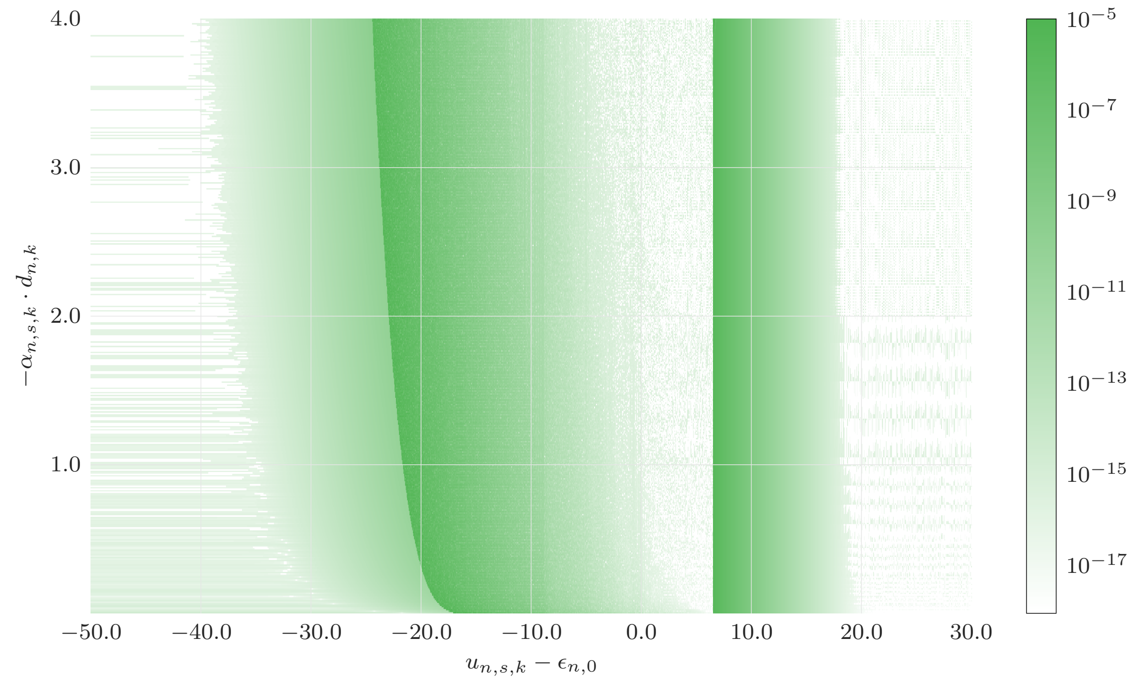


Figure C.4: Absolute Error of Numerically Stable Algorithm

Bibliography

Acquisti, A., C. Taylor, and L. Wagman (June 1, 2016). "The Economics of Privacy." *Journal of Economic Literature* 54.2, pp. 442–492. ISSN: 0022-0515. DOI: 10.1257/jel.54.2.442.

Ben-Akiva, M. E. (1973). "Structure of passenger travel demand models." PhD thesis. Massachusetts Institute of Technology.

Beranek, L., V. Nýdl, and R. Remes (2017). "Click Stream Data Analysis for Online Fraud Detection in E-Commerce."

Bergemann, D., B. Brooks, and S. Morris (Mar. 2015). "The Limits of Price Discrimination." *American Economic Review* 105.3, pp. 921–57. DOI: 10.1257/aer.20130848.

Bhat, C. R. (1998). "Accommodating variations in responsiveness to level-of-service measures in travel mode choice modeling." *Transportation Research Part A: Policy and Practice* 32.7, pp. 495–507. ISSN: 0965-8564. DOI: 10.1016/S0965-8564(98)00011-1.

Bhat, C. R. (2000). "Incorporating Observed and Unobserved Heterogeneity in Urban Work Travel Mode Choice Modeling." *Transportation Science* 34.2. Publisher: INFORMS, pp. 228–238. ISSN: 0041-1655.

Bigon, L. et al. (2019). "Prediction is very hard, especially about conversion. Predicting user purchases from clickstream data in fashion e-commerce." arXiv: 1907.00400 [cs.IR].

Boyd, J. H. and R. E. Mellman (1980). "The effect of fuel economy standards on the U.S. automotive market: An hedonic demand analysis." *Transportation Research Part A: General* 14.5, pp. 367–378. ISSN: 0191-2607. DOI: 10.1016/0191-2607(80)90055-2.

Brownstone, D. and K. Train (1998). "Forecasting new product penetration with flexible substitution patterns." *Journal of Econometrics* 89.1, pp. 109–129. ISSN: 0304-4076. DOI: 10.1016/S0304-4076(98)00057-8.

Cardell, N. S. and F. C. Dunbar (1980). "Measuring the societal impacts of automobile downsizing." *Transportation Research Part A: General* 14.5, pp. 423–434. ISSN: 0191-2607. DOI: 10.1016/0191-2607(80)90060-6.

Chintagunta, P. K., D. C. Jain, and N. J. Vilcassim (1991). "Investigating Heterogeneity in Brand Preferences in Logit Models for Panel Data." *Journal of Marketing Research* 28.4. Publisher: American Marketing Association, pp. 417–428. ISSN: 0022-2437. DOI: 10.2307/3172782.

Cho, K. et al. (2014). "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation." arXiv: 1406.1078 [cs.CL].

Dubé, J.-P. and S. Misra (2023). "Personalized Pricing and Consumer Welfare." *Journal of Political Economy* 131.1, pp. 131–189. DOI: 10.1086/720793.

Erdem, T. (1996). "A Dynamic Analysis of Market Structure Based on Panel Data." *Marketing Science* 15.4. Publisher: INFORMS, pp. 359–378. ISSN: 0732-2399.

Executive Office of the President (2014). *Big Data: Seizing Opportunities, Preserving Values.*

Farrell, M. H., T. Liang, and S. Misra (2021). "Deep Learning for Individual Heterogeneity: An Automatic Inference Framework." arXiv: 2010.14694 [econ.EM].

Heckman, J. and B. Singer (1984). "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data." *Econometrica* 52.2, pp. 271–320.

Hochreiter, S. and J. Schmidhuber (1997). "Long Short-Term Memory." *Neural Computation* 9.8, pp. 1735–1780. DOI: `10.1162/neco.1997.9.8.1735`.

Huang, T. and J. A. Van Mieghem (2014). "Clickstream Data and Inventory Management: Model and Empirical Analysis." *Production and Operations Management* 23.3, pp. 333–347. ISSN: 1937-5956. DOI: `10.1111/poms.12046`.

Kamakura, W. A. and G. J. Russell (1989). "A Probabilistic Choice Model for Market Segmentation and Elasticity Structure." *Journal of Marketing Research* 26.4. Publisher: American Marketing Association, pp. 379–390. ISSN: 0022-2437. DOI: `10.2307/3172759`.

Kehoe, P. J., B. J. Larsen, and E. Pastorino (2022). "Dynamic Competition in the Era of Big Data." *Hoover Institution Economics Working Paper* 22102.

Kukar-Kinney, M. and A. G. Close (2010). "The determinants of consumers' online shopping cart abandonment." *Journal of the Academy of Marketing Science* 38, pp. 240–250.

Madrigal, A. C. (2012). "I'm Being Followed: How Google—and 104 Other Companies—Are Tracking Me on the Web." *The Atlantic*.

McFadden, D. (1974). "Conditional logit analysis of qualitative choice behavior." *Frontiers in Econometrics*. Ed. by P. Zarembka. New York: Academic Press, pp. 105–142.

McFadden, D. (1978). “Modeling the choice of residential location.” *Spatial Interaction Theory and Planning Models*. Ed. by A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull. Amsterdam: North-Holland, pp. 75–96.

Moe, W. W. (2003). “Buying, Searching, or Browsing: Differentiating between Online Shoppers Using In-Store Navigational Clickstream.” *Journal of Consumer Psychology* 13.1/2, pp. 29–39. ISSN: 10577408.

Montgomery, A. L., S. Li, K. Srinivasan, and J. C. Liechty (2004). “Modeling Online Browsing and Path Analysis Using Clickstream Data.” *Marketing Science* 23.4, pp. 579–595. DOI: 10.1287/mksc.1040.0073.

Nair, V. and G. E. Hinton (2010). “Rectified Linear Units Improve Restricted Boltzmann Machines.” *ICML*.

Olver, F., D. Lozier, R. Boisvert, and C. Clark (May 2010). *The NIST Handbook of Mathematical Functions*. Cambridge University Press, New York, NY.

Pallant, J. I., P. J. Danaher, S. J. Sands, and T. S. Danaher (2017). “An empirical analysis of factors that influence retail website visit types.” *Journal of Retailing and Consumer Services* 39, pp. 62–70. ISSN: 0969-6989. DOI: 10.1016/j.jretconser.2017.07.003.

Pigou, A. C. (1920). *The Economics of Welfare*. Macmillan and co., Ltd.

Quadrana, M., P. Cremonesi, and D. Jannach (2018). “Sequence-Aware Recommender Systems.” arXiv:1802.08452. DOI: 10.48550/arXiv.1802.08452. arXiv: 1802.08452[cs].

Rajamma, R. K., A. Paswan, and M. M. Hossain (2009). “Why do shoppers abandon shopping cart? Perceived waiting time, risk, and transaction inconvenience.” *Journal of Product & Brand Management* 18, pp. 188–197.

Revelt, D. and K. Train (1998). "Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level." *The Review of Economics and Statistics* 80.4. Publisher: The MIT Press, pp. 647–657. ISSN: 0034-6535.

Rossi, P. E., R. E. McCulloch, and G. M. Allenby (1996). "The Value of Purchase History Data in Target Marketing." *Marketing Science* 15.4, pp. 321–340. ISSN: 0732-2399.

Rumelhart, D. E. and J. L. McClelland (1987). "Learning Internal Representations by Error Propagation." *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations*, pp. 318–362.

Schellong, D., J. Kemper, and M. Brettel (2016). "Clickstream Data as a Source to Uncover Consumer Shopping Types in a Large-Scale Online Setting." *European Conference on Information Systems.*

Shiller, B. and J. Waldfogel (2011). "Music for a Song: An Empirical Look at Uniform Pricing and Its Alternatives." *The Journal of Industrial Economics* 59.4, pp. 630–660. ISSN: 0022-1821, 1467-6451. DOI: 10.1111/j.1467-6451.2011.00470.x.

Shiller, B. (2014). "First Degree Price Discrimination Using Big Data." Working Paper No. 58, Brandeis University, Department of Economics and International Business School.

Vaswani, A. et al. (2017). "Attention Is All You Need." arXiv: 1706.03762 [cs.CL].

Waldfogel, J. (2015). "First Degree Price Discrimination Goes to School." *The Journal of Industrial Economics* 63.4, pp. 569–597. ISSN: 00221821, 14676451.