Georgetown University

Essays on Consumer Heterogeneity and Personalized Discounts in an Online Market

Ph.D. Dissertation Defense

Chengjun Zhang

April 05, 2024

Introduction

In the current digital era, the vast expansion of the digital domain has made it imperative for companies to leverage available information strategically to gain a competitive edge through innovative marketing practices. With lower costs of data collection and storage, big data has become a crucial element in strategic decision-making, especially in marketing (Acquisti et al., 2016). The availability of detailed consumer clickstreams and unprecedented computing power have largely changed the way researchers model consumer behavior and their heterogeneity.

This study delves into enhancing the understanding of consumer heterogeneity and the welfare implications of marketing strategies building on top of that. Chapter 1 focuses on understanding consumer behavior, their heterogeneity, and performs detailed demand estimations of an online smartphone market. And, Chapter 2 conducts a counterfactual analysis on firms issuing personalized discounts, and the impact on profit and consumer surplus.

Introduction

In the current digital era, the vast expansion of the digital domain has made it imperative for companies to leverage available information strategically to gain a competitive edge through innovative marketing practices. With lower costs of data collection and storage, big data has become a crucial element in strategic decision-making, especially in marketing (Acquisti et al., 2016). The availability of detailed consumer clickstreams and unprecedented computing power have largely changed the way researchers model consumer behavior and their heterogeneity.

This study delves into enhancing the understanding of consumer heterogeneity and the welfare implications of marketing strategies building on top of that. Chapter 1 focuses on understanding consumer behavior, their heterogeneity, and performs detailed demand estimations of an online smartphone market. And, Chapter 2 conducts a counterfactual analysis on firms issuing personalized discounts, and the impact on profit and consumer surplus.

Introduction
Research Questions

Chapter 1:
Consumer Heterogeneity in Online Shopping: A Machine Learning Approach with Sequential Data

- Are consumers heterogeneous in preferences?
- Are the observed search behaviors related to this heterogeneity?
- Are the historical browsing behaviors related to this heterogeneity?
- How can one model this heterogeneity at the individual level in a parsimonious way?

Chapter 2:
Personalized Discounts via Coupons: Generating Gains Conditional on Observed Choices

- How can firms utilize detailed data for marketing purposes?
- What is the welfare effect of leveraging that nuanced heterogeneity for marketing?

Introduction

Research Questions

Chapter 1:

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

- Are consumers heterogeneous in preferences?
- Are the observed search behaviors related to this heterogeneity?
- Are the historical browsing behaviors related to this heterogeneity?
- How can one model this heterogeneity at the individual level in a parsimonious way?

Chapter 2:

Personalized Discounts via Coupons: Generating Gains Conditional on Observed Choices

- How can firms utilize detailed data for marketing purposes?
- What is the welfare effect of leveraging that nuanced heterogeneity for marketing?

Chapter 1

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

Table of Contents

Table of Contents

Related Literature

- Conditional Logit and Variants

Logit	:	McFadden (1974)
Nested logit	:	Ben-Akiva (1973), McFadden (1978)
Mixed logit	:	Cardell and Dunbar (1980), Erdem (1996), Boyd and Mellman (1980), Bhat (1998), Brownstone and Train (1998), Revelt and Train (1998), Bhat (2000)
Latent classes	:	Heckman and Singer (1984), Kamakura and Russell (1989), Chintagunta et al. (1991)
Non-parametric	:	Farrell et al. (2021)

- Consumer Clickstreams

Purchase conversion	:	Montgomery et al. (2004), Bigon et al. (2019), Koehn et al. (2020)
Cart abandonment	:	Rajamma et al. (2009), Kukar-Kinney and Close (2010)
Inventory management	:	Huang and Van Mieghem (2014)
Fraud detection	:	Beranek et al. (2017)
Consumer categorization	:	Moe (2003), Schellong et al. (2016), Pallant et al. (2017)

- Machine Learning

Non-sequential	:	DNN/MLP (as in Farrell et al., 2021)
Sequential	:	RNN (Rumelhart and McClelland, 1987), LSTM (Hochreiter and Schmidhuber, 1997), GRU (Cho et al., 2014), Transformers (Vaswani et al., 2017)

Related Literature

- Conditional Logit and Variants

Logit	:	McFadden (1974)
Nested logit	:	Ben-Akiva (1973), McFadden (1978)
Mixed logit	:	Cardell and Dunbar (1980), Erdem (1996), Boyd and Mellman (1980), Bhat (1998), Brownstone and Train (1998), Revelt and Train (1998), Bhat (2000)
Latent classes	:	Heckman and Singer (1984), Kamakura and Russell (1989), Chintagunta et al. (1991)
Non-parametric	:	Farrell et al. (2021)

- Consumer Clickstreams

Purchase conversion	:	Montgomery et al. (2004), Bigon et al. (2019), Koehn et al. (2020)
Cart abandonment	:	Rajamma et al. (2009), Kukar-Kinney and Close (2010)
Inventory management	:	Huang and Van Mieghem (2014)
Fraud detection	:	Beranek et al. (2017)
Consumer categorization	:	Moe (2003), Schellong et al. (2016), Pallant et al. (2017)

- Machine Learning

Non-sequential	:	DNN/MLP (as in Farrell et al., 2021)
Sequential	:	RNN (Rumelhart and McClelland, 1987), LSTM (Hochreiter and Schmidhuber, 1997), GRU (Cho et al., 2014), Transformers (Vaswani et al., 2017)

Related Literature

- Conditional Logit and Variants

Logit	:	McFadden (1974)
Nested logit	:	Ben-Akiva (1973), McFadden (1978)
Mixed logit	:	Cardell and Dunbar (1980), Erdem (1996), Boyd and Mellman (1980), Bhat (1998), Brownstone and Train (1998), Revelt and Train (1998), Bhat (2000)
Latent classes	:	Heckman and Singer (1984), Kamakura and Russell (1989), Chintagunta et al. (1991)
Non-parametric	:	Farrell et al. (2021)

- Consumer Clickstreams

Purchase conversion	:	Montgomery et al. (2004), Bigon et al. (2019), Koehn et al. (2020)
Cart abandonment	:	Rajamma et al. (2009), Kukar-Kinney and Close (2010)
Inventory management	:	Huang and Van Mieghem (2014)
Fraud detection	:	Beranek et al. (2017)
Consumer categorization	:	Moe (2003), Schellong et al. (2016), Pallant et al. (2017)

- Machine Learning

Non-sequential	:	DNN/MLP (as in Farrell et al., 2021)
Sequential	:	RNN (Rumelhart and McClelland, 1987), LSTM (Hochreiter and Schmidhuber, 1997), GRU (Cho et al., 2014), Transformers (Vaswani et al., 2017)

Table of Contents

Data

- Privacy Disclaimer

- E-commerce Platform Information
- Clickstream Data
- Product and User Information
- Assumptions on Prices and Availabilities
- Examples of Smartphone Daily Prices and Purchases

Data

E-commerce Platform Information

- a leading multi-category e-commerce giant operating exclusively within a single country
- the country's premier online shopping destination
  - 2^{nd} to 7^{th} largest online stores are travel agencies or clothing retailers
  - $\approx$ a half of all online transaction volume in the country
  - $> 10\times$ the next largest online electronics retailer by transaction revenue
- all smartphones are contract-free in this country
- uniform price across all users (no price discrimination)

Data
Clickstream Data

- clickstream data for ≈200 days (unbalanced panel)
- each row is a *click* (or *event*), encompassing:
  - timestamp
  - event type
    - *view*: enter product page, enlarge pictures, load more comments, refresh the product page
    - *cart*
    - *purchase*
  - user ID
  - product ID
  - price of the product at the timestamp
- total number of *events* is in the order of hundreds of millions (10^8)

- Data limitation:
  No data on search queries, results, filtering, sorting, or scroll clicks.

Data
Product and User Information

- $\approx$400,000 unique products each with a unique product ID and webpage
  - product name
  - categories (up to four levels of subcategories)
  - product features
  - ratings and reviews
  - etc. (everything a user would see)
- Why smartphones?
  - among about 1,000 subcategories (at the finest level) of products, the smartphones category alone occupies $\approx$25% of all the events, $\approx$45% of all the purchase events
  - yet <0.5% of all the unique products
- Users
  - focuses only on users who have at least one smartphone event, and purchased at most one smartphone during the recorded period
  - referred to as *all users* which has a size of 5,647,094
- Data limitation:
  No data on credit card info, full purchase history, account opening times, or demographics.

Data

Product and User Information

- ≈400,000 unique products each with a unique product ID and webpage
  - product name
  - categories (up to four levels of subcategories)
  - product features
  - ratings and reviews
  - etc. (everything a user would see)
- Why smartphones?
  - among about 1,000 subcategories (at the finest level) of products, the smartphones category alone occupies ≈25% of all the events, ≈45% of all the purchase events
  - yet <0.5% of all the unique products
- Users
  - focuses only on users who have at least one smartphone event, and purchased at most one smartphone during the recorded period
  - referred to as *all users* which has a size of 5,647,094
- Data limitation:
  No data on credit card info, full purchase history, account opening times, or demographics.

Data
Product and User Information

- ≈400,000 unique products each with a unique product ID and webpage
    - product name
    - categories (up to four levels of subcategories)
    - product features
    - ratings and reviews
    - etc. (everything a user would see)
- Why smartphones?
    - among about 1,000 subcategories (at the finest level) of products, the smartphones category alone occupies ≈25% of all the events, ≈45% of all the purchase events
    - yet <0.5% of all the unique products
- Users
    - focuses only on users who have at least one smartphone event, and purchased at most one smartphone during the recorded period
    - referred to as *all users* which has a size of 5,647,094
- Data limitation:
  No data on credit card info, full purchase history, account opening times, or demographics.

Data
Assumptions on Prices and Availabilities

Assumption: Prices

The price of a product at any given timestamp is assumed to be the price recorded in the latest event of that product at or before the timestamp.

Assumption: Availabilities

A smartphone is assumed to be available for purchase from the first recorded event until the last recorded event of that smartphone product.

On average, ≈1,500 smartphones are available at any given moment, contributing to a cumulative total of ≈1,700 smartphones.

Data

Assumptions on Prices and Availabilities

Assumption: Prices

The price of a product at any given timestamp is assumed to be the price recorded in the latest event of that product at or before the timestamp.

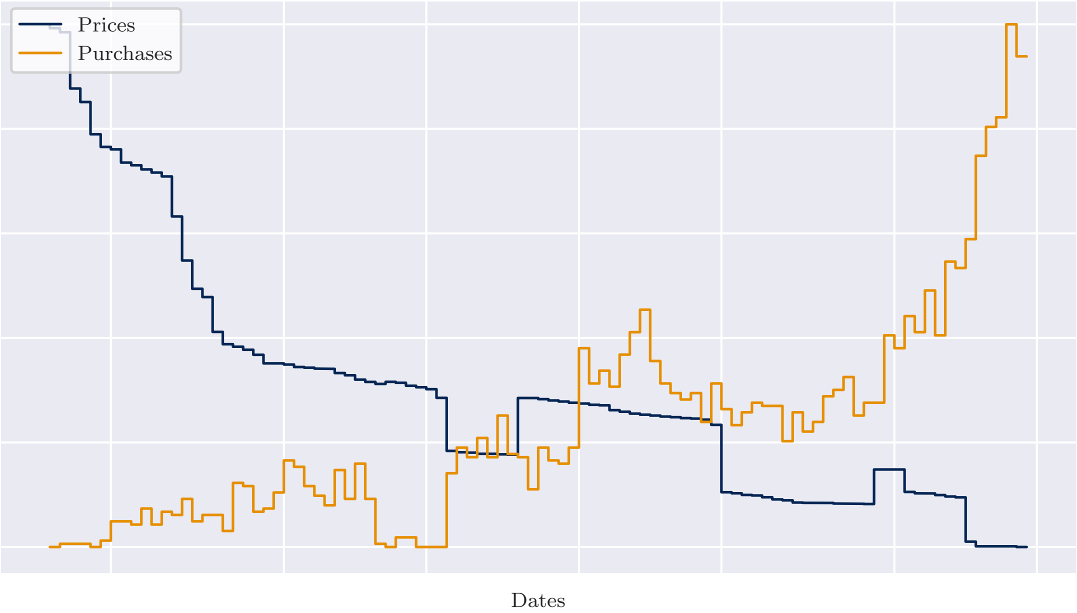
Assumption: Availabilities

A smartphone is assumed to be available for purchase from the first recorded event until the last recorded event of that smartphone product.

On average, ≈1,500 smartphones are available at any given moment, contributing to a cumulative total of ≈1,700 smartphones.

Data

Example of A Smartphone Daily Prices and Purchases (Samsung, about $350)



Data

Example of A Smartphone Daily Prices and Purchases (Xiaomi, about $190)

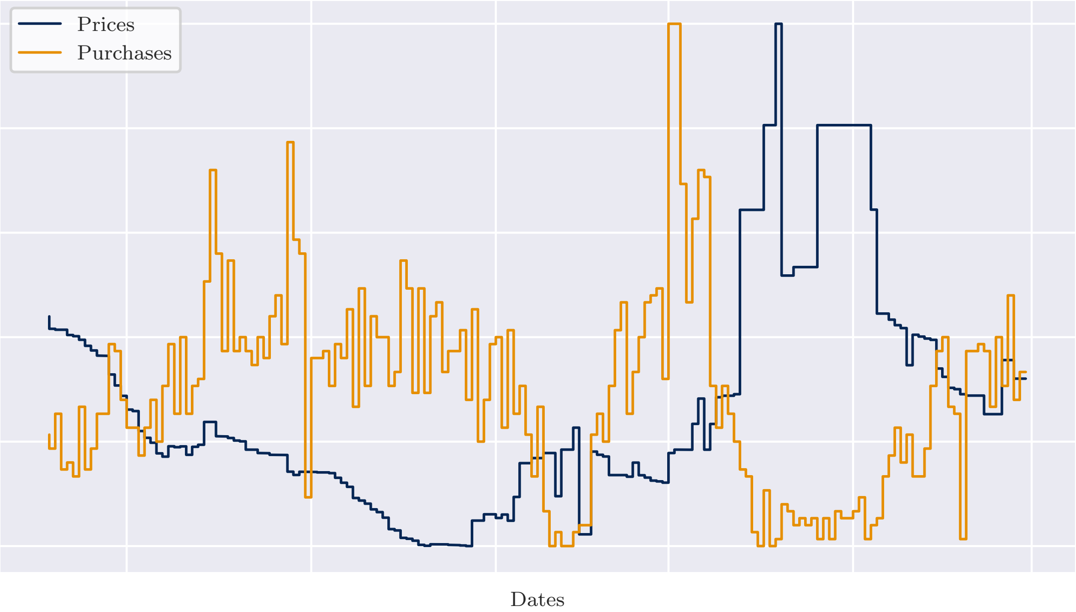


Table of Contents

Classification Based on Heuristics

- View Event Types
- Event Duration
- User Types

Classification Based on Heuristics
Some Definitions

Definition: Time Gap

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

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

Definition: Event Duration

For each user, the *event duration* of an event is calculated by the time difference between this event and its immediate subsequent event, if such an event exists.

i.e., event duration not defined for the last event of each user.

Classification Based on Heuristics
Some Definitions

Definition: Time Gap

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

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

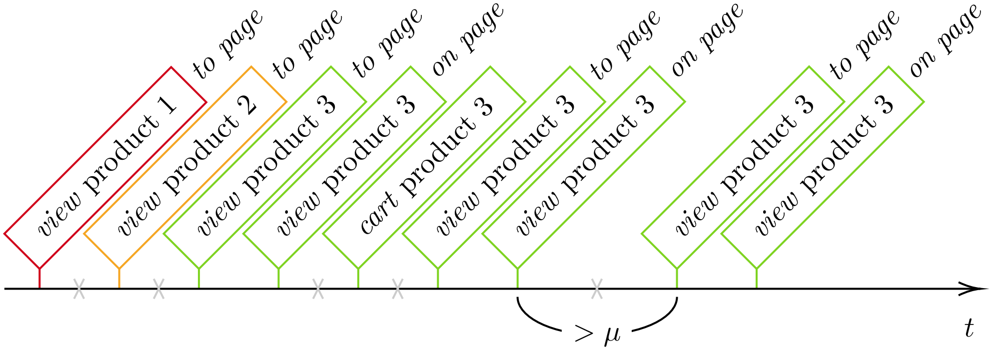
Definition: Event Duration

For each user, the *event duration* of an event is calculated by the time difference between this event and its immediate subsequent event, if such an event exists.

i.e., event duration not defined for the last event of each user.

Classification Based on Heuristics
View Event Types

- click *to page* : entails a user navigating directly to a product's webpage
- click *on page* : arises from actions taken on a product's webpage itself
  e.g., enlarging a picture, loading more comments, or refresh the page

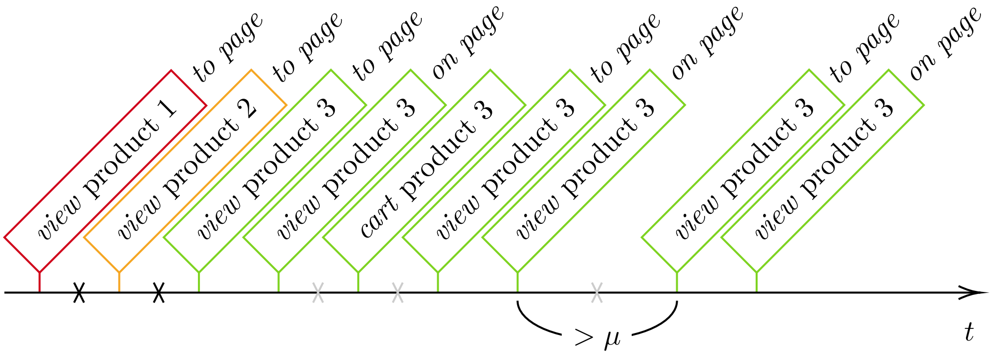


▸ Table

Classification Based on Heuristics

View Event Types

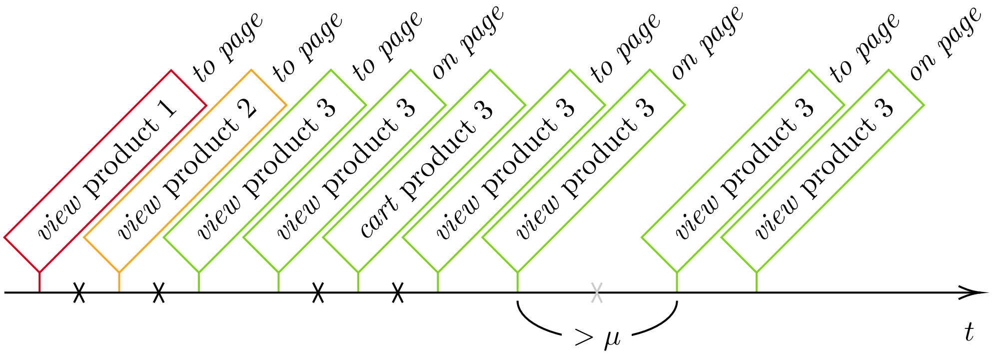
- click *to page* : entails a user navigating directly to a product's webpage
- click *on page* : arises from actions taken on a product's webpage itself
  e.g., enlarging a picture, loading more comments, or refresh the page



Table

Classification Based on Heuristics
View Event Types

- click *to page* : entails a user navigating directly to a product's webpage
- click *on page* : arises from actions taken on a product's webpage itself
  e.g., enlarging a picture, loading more comments, or refresh the page

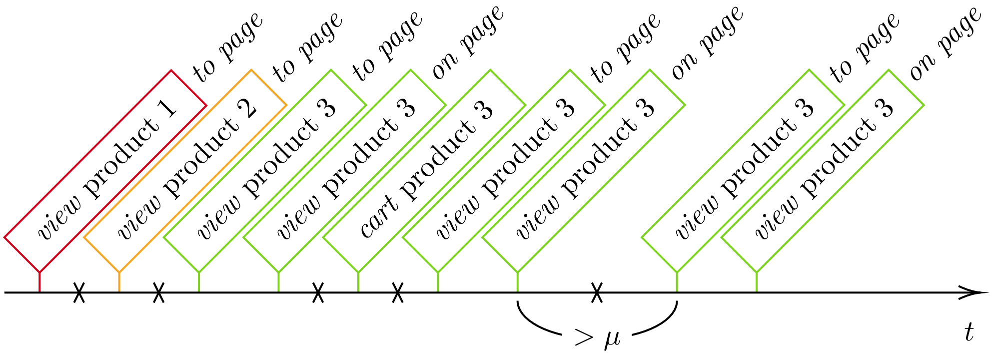


Table

Classification Based on Heuristics

View Event Types

- click *to page* : entails a user navigating directly to a product's webpage
- click *on page* : arises from actions taken on a product's webpage itself
  e.g., enlarging a picture, loading more comments, or refresh the page



▸ Table

Classification Based on Heuristics

Event Duration

Definition: Uninterrupted View Click

If an event has an *event duration* $\leq \mu$, it is presumed that the user was engaged with the product's webpage throughout this period, classifying the event as an *uninterrupted* view click.

Regressor: Avg uninterrupted view time (min)

Definition: Interrupted View Click

If an event has an *event duration* $> \mu$ or undefined, the event is considered as an *interrupted* view click.

Regressor: # of interrupted view clicks

Classification Based on Heuristics

Event Duration

Definition: Uninterrupted View Click

If an event has an *event duration* $\leq \mu$, it is presumed that the user was engaged with the product's webpage throughout this period, classifying the event as an *uninterrupted* view click.

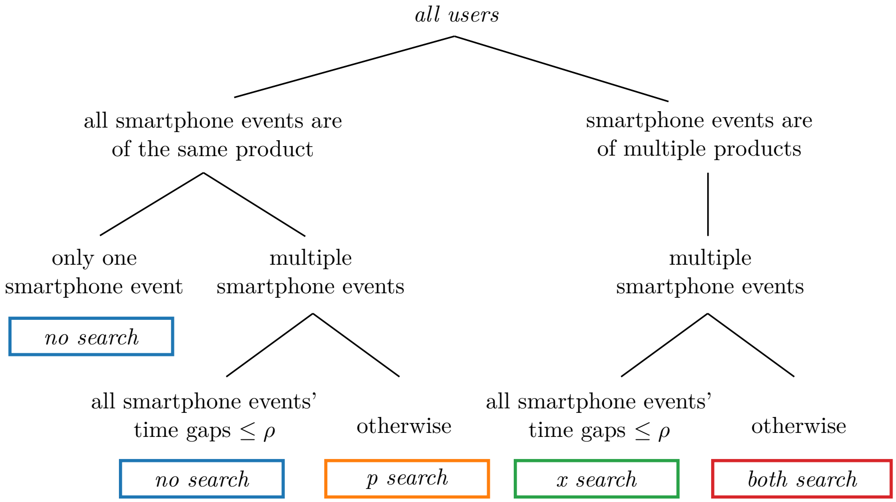
Regressor: Avg uninterrupted view time (min)

Definition: Interrupted View Click

If an event has an *event duration* $> \mu$ or undefined, the event is considered as an *interrupted* view click.

Regressor: # of interrupted view clicks

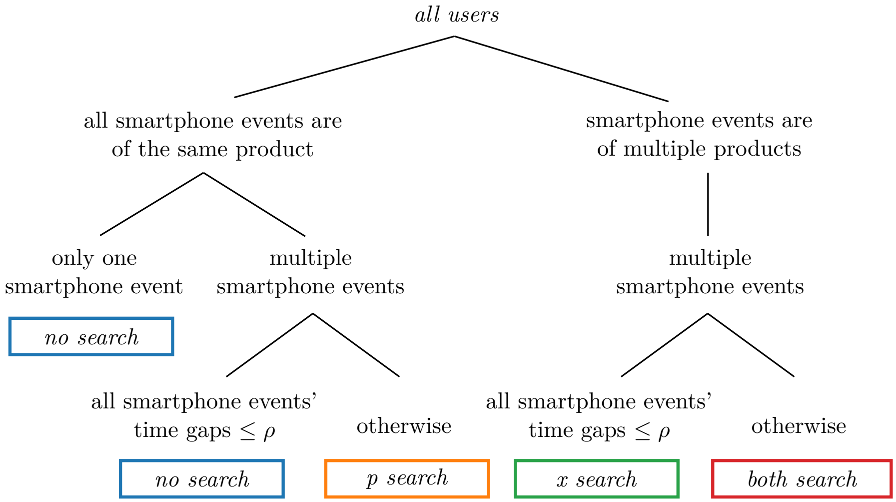
Classification Based on Heuristics
User Types



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

▸ Table

Classification Based on Heuristics
User Types



	no search	p search	x search	both search
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

▸ Table

Classification Based on Heuristics

User Types

all users

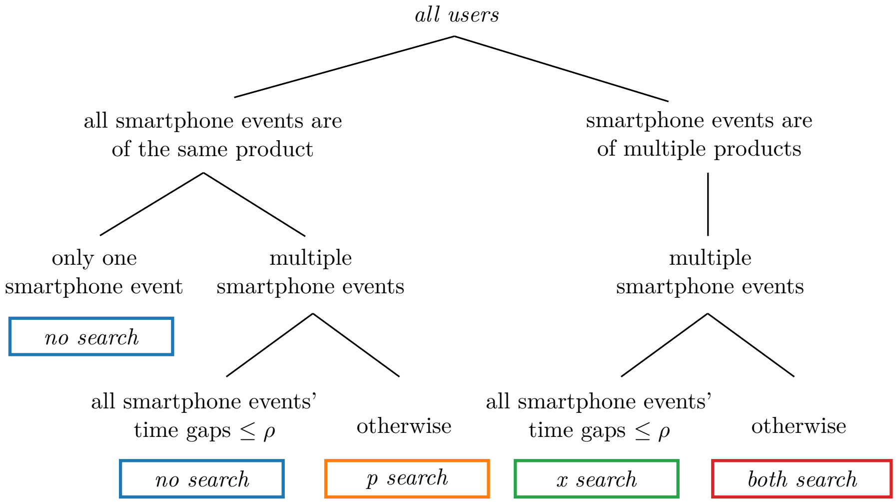
- all smartphone events are of the same product
  - only one smartphone event → *no search*
  - multiple smartphone events
    - all smartphone events' time gaps $\leq \rho$ → *no search*
    - otherwise → *p search*
- smartphone events are of multiple products
  - multiple smartphone events
    - all smartphone events' time gaps $\leq \rho$ → *x search*
    - otherwise → *both search*

	no search	*p search*	*x search*	*both search*
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

▸ Table

Classification Based on Heuristics

User Types



	no search	p search	x search	both search
Share	48.20%	5.15%	20.39%	26.25%
Conversion	6.26%	3.87%	9.73%	12.51%
Avg Δp	$-\$0.002$	$-\$0.319$	$+\$0.002$	$-\$0.702$

▸ Table

Table of Contents

Descriptive Analysis on User Types

- User Types and Indicators of Events

  Logit with user types as dependent variables and indicators of events in different categories as independent variables.

- Smartphone Events Distribution by User Types

  Empirical distribution of user activity over hours of day, and over days of week.

Descriptive Analysis on User Types
User Types and Indicators of Events

	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*.

List of Categories

Descriptive Analysis on User Types
User Types and Indicators of Events

	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*.

List of Categories

Descriptive Analysis on User Types

User Types and Indicators of Events

	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*.

List of Categories

Descriptive Analysis on User Types

User Types and Indicators of Events

	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*.

▸ List of Categories

Descriptive Analysis on User Types
User Types and Indicators of Events

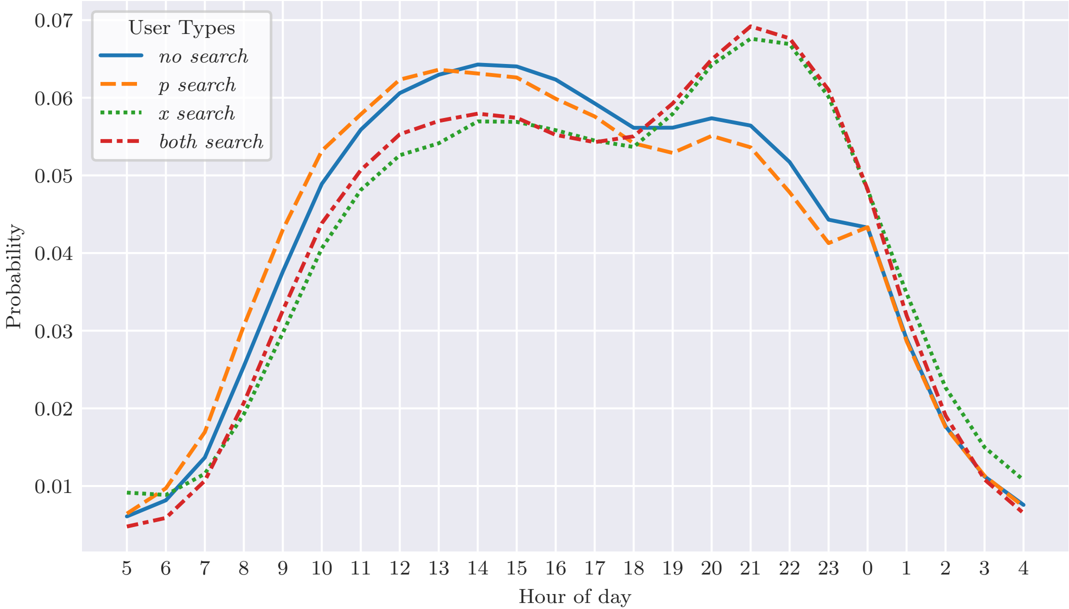
	no search		*x search*		*both search*	
Indicator of events in:						
- beauty care	0.016	(0.009)	0.122***	(0.009)	0.137***	(0.009)
- car goods	0.187***	(0.009)	0.441***	(0.009)	0.549***	(0.009)
- child goods	−0.455***	(0.007)	−0.726***	(0.007)	−0.428***	(0.007)
- computers	0.237***	(0.008)	0.602***	(0.008)	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*.

List of Categories

Descriptive Analysis on User Types

Smartphone Events Distribution by User Types



Descriptive Analysis on User Types

Smartphone Events Distribution by User Types

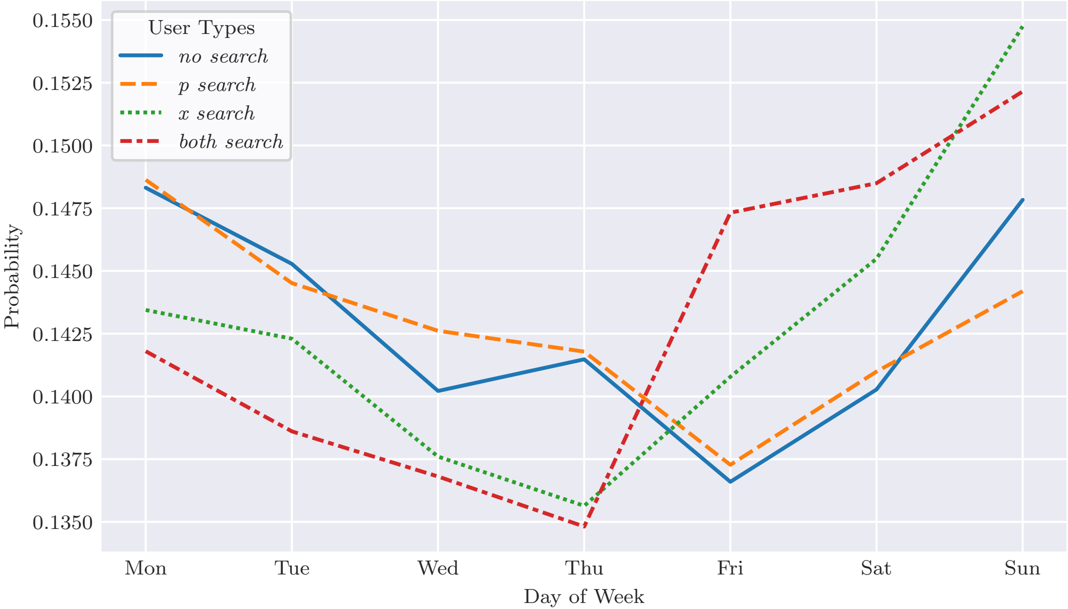


Table of Contents

Models and Results

Common Framework - Conditional Logit Model

- user $n \in \mathcal{N}$
- choice set $\mathcal{C}_n = \mathcal{A}_{t(n)} \cup \{0\}$
- active time $t(n)$
- each user makes exactly one choice from $\mathcal{C}_n$ that maximizes their utility

$$U_{n,i} \equiv u_{n,i} + \epsilon_{n,i} = \alpha \cdot p_{t(n),i} + \boldsymbol{\beta}^\mathsf{T} \mathbf{x}_{n,t(n),i} + \epsilon_{n,i}, \quad \forall i \in \mathcal{C}_n \tag{1}$$

- $u_{n,0}$ is normalized to 0
- $\epsilon_{n,i} \overset{\text{iid}}{\sim}$ Extreme Value Type-I across n and i
- choice probability

$$P_{n,i} \equiv \text{Prob}\left(U_{n,i} \geq U_{n,j}, \forall j \in \mathcal{C}_n\right) = \frac{\exp(u_{n,i})}{1 + \sum_{j \in \mathcal{A}_{t(n)}} \exp(u_{n,j})}, \quad \forall i \in \mathcal{C}_n. \tag{2}$$

- log-likelihood for MLE

$$\ell\left(\boldsymbol{\Theta}\right) = \sum_{n \in \mathcal{N}} \sum_{i \in \mathcal{C}_n} y_{n,i} \cdot \log(P_{n,i}) \tag{3}$$

Models and Results
Common Framework - Conditional Logit Model

- user $n \in \mathcal{N}$
- choice set $\mathcal{C}_n = \mathcal{A}_{t(n)} \cup \{0\}$
- active time $t(n)$
- each user makes exactly one choice from $\mathcal{C}_n$ that maximizes their utility

$$U_{n,i} \equiv u_{n,i} + \epsilon_{n,i} = \alpha \cdot p_{t(n),i} + \boldsymbol{\beta}^{\mathsf{T}} \mathbf{x}_{n,t(n),i} + \epsilon_{n,i}, \quad \forall i \in \mathcal{C}_n \tag{1}$$

- $u_{n,0}$ is normalized to 0
- $\epsilon_{n,i} \overset{\text{iid}}{\sim}$ Extreme Value Type-I across n and i
- choice probability

$$P_{n,i} \equiv \text{Prob}\left(U_{n,i} \geq U_{n,j}, \forall j \in \mathcal{C}_n\right) = \frac{\exp(u_{n,i})}{1 + \sum_{j \in \mathcal{A}_{t(n)}} \exp(u_{n,j})}, \quad \forall i \in \mathcal{C}_n. \tag{2}$$

- log-likelihood for MLE

$$\ell\left(\boldsymbol{\Theta}\right) = \sum_{n \in \mathcal{N}} \sum_{i \in \mathcal{C}_n} y_{n,i} \cdot \log(P_{n,i}) \tag{3}$$

Models and Results

Common Framework - Conditional Logit Model

- user $n \in \mathcal{N}$
- choice set $\mathcal{C}_n = \mathcal{A}_{t(n)} \cup \{0\}$
- active time $t(n)$
- each user makes exactly one choice from $\mathcal{C}_n$ that maximizes their utility

$$U_{n,i} \equiv u_{n,i} + \epsilon_{n,i} = \alpha \cdot p_{t(n),i} + \boldsymbol{\beta}^{\mathsf{T}} \mathbf{x}_{n,t(n),i} + \epsilon_{n,i}, \quad \forall i \in \mathcal{C}_n \tag{1}$$

- $u_{n,0}$ is normalized to 0
- $\epsilon_{n,i} \overset{\text{iid}}{\sim}$ Extreme Value Type-I across n and i
- choice probability

$$P_{n,i} \equiv \text{Prob}\left(U_{n,i} \geq U_{n,j}, \forall j \in \mathcal{C}_n\right) = \frac{\exp(u_{n,i})}{1 + \sum_{j \in \mathcal{A}_{t(n)}} \exp(u_{n,j})}, \quad \forall i \in \mathcal{C}_n. \tag{2}$$

- log-likelihood for MLE

$$\ell\left(\boldsymbol{\Theta}\right) = \sum_{n \in \mathcal{N}} \sum_{i \in \mathcal{C}_n} y_{n,i} \cdot \log(P_{n,i}) \tag{3}$$

Models and Results

Common Framework - Conditional Logit Model

- user $n \in \mathcal{N}$
- choice set $\mathcal{C}_n = \mathcal{A}_{t(n)} \cup \{0\}$
- active time $t(n)$
- each user makes exactly one choice from $\mathcal{C}_n$ that maximizes their utility

$$U_{n,i} \equiv u_{n,i} + \epsilon_{n,i} = \alpha \cdot p_{t(n),i} + \boldsymbol{\beta}^{\mathsf{T}} \mathbf{x}_{n,t(n),i} + \epsilon_{n,i}, \quad \forall i \in \mathcal{C}_n \tag{1}$$

- $u_{n,0}$ is normalized to 0
- $\epsilon_{n,i} \overset{\text{iid}}{\sim}$ Extreme Value Type-I across n and i
- choice probability

$$P_{n,i} \equiv \text{Prob}\left(U_{n,i} \geq U_{n,j}, \forall j \in \mathcal{C}_n\right) = \frac{\exp(u_{n,i})}{1 + \sum_{j \in \mathcal{A}_{t(n)}} \exp(u_{n,j})}, \quad \forall i \in \mathcal{C}_n. \tag{2}$$

- log-likelihood for MLE

$$\ell\left(\boldsymbol{\Theta}\right) = \sum_{n \in \mathcal{N}} \sum_{i \in \mathcal{C}_n} y_{n,i} \cdot \log(P_{n,i}) \tag{3}$$

Models and Results

All Models - with Increasing Levels of Heterogeneities

- Model I - Baseline Logit
- Model II - by User Types
- Model III - with Unobserved Heterogeneity
- Models with Individual-level Heterogeneity
  - Model IV - MLP
  - Model V - RNN

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 $\boldsymbol{\theta}_{\text{logit}} = \left[\alpha_{\text{logit}}, \boldsymbol{\beta}_{\text{logit}}{}^{\top}\right]^{\top}$ denote the vector of coefficients. Consequently, the utility $U_{n,i}$ as from the generic form (1) is hence,

$$U_{n,i} \equiv u_{n,i} + \epsilon_{n,i} = \alpha_{\text{logit}} \cdot p_{t(n),i} + \boldsymbol{\beta}_{\text{logit}}{}^{\top} \mathbf{x}_{n,t(n),i} + \epsilon_{n,i}, \quad \forall i \in \mathcal{C}_n. \tag{4}$$

The conditional choice probability is as specified in (2),

$$P_{n,i} \equiv \text{Prob}\left(U_{n,i} \geq U_{n,j}, \forall j \in \mathcal{C}_n\right) = \frac{\exp(u_{n,i})}{1 + \sum_{j \in \mathcal{A}_{t(n)}} \exp(u_{n,j})}, \quad \forall i \in \mathcal{C}_n. \tag{5}$$

And the estimation involves maximizing $\ell\left(\boldsymbol{\Theta}_{\text{I}}\right)$ as specified in (3) where $\boldsymbol{\Theta}_{\text{I}} = \boldsymbol{\theta}_{\text{logit}}$,

$$\ell\left(\boldsymbol{\Theta}_{\text{I}}\right) = \sum_{n \in \mathcal{N}} \sum_{i \in \mathcal{C}_n} y_{n,i} \cdot \log(P_{n,i}). \tag{6}$$

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) is hence,

$$U_{n,i} \equiv u_{n,i} + \epsilon_{n,i} = \alpha_{\text{logit}} \cdot p_{t(n),i} + {\boldsymbol{\beta}_{\text{logit}}}^{\mathsf{T}} \mathbf{x}_{n,t(n),i} + \epsilon_{n,i}, \quad \forall i \in \mathcal{C}_n. \tag{4}$$

The conditional choice probability is as specified in (2),

$$P_{n,i} \equiv \text{Prob}\left(U_{n,i} \geq U_{n,j}, \forall j \in \mathcal{C}_n\right) = \frac{\exp(u_{n,i})}{1 + \sum_{j \in \mathcal{A}_{t(n)}} \exp(u_{n,j})}, \quad \forall i \in \mathcal{C}_n. \tag{5}$$

And the estimation involves maximizing $\ell\left(\boldsymbol{\Theta}_{\text{I}}\right)$ as specified in (3) where $\boldsymbol{\Theta}_{\text{I}} = \boldsymbol{\theta}_{\text{logit}}$,

$$\ell\left(\boldsymbol{\Theta}_{\text{I}}\right) = \sum_{n \in \mathcal{N}} \sum_{i \in \mathcal{C}_n} y_{n,i} \cdot \log(P_{n,i}). \tag{6}$$

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) is hence,

$$U_{n,i} \equiv u_{n,i} + \epsilon_{n,i} = \alpha_{\text{logit}} \cdot p_{t(n),i} + \boldsymbol{\beta}_{\text{logit}}{}^{\mathsf{T}} \mathbf{x}_{n,t(n),i} + \epsilon_{n,i}, \quad \forall i \in \mathcal{C}_n. \tag{4}$$

The conditional choice probability is as specified in (2),

$$P_{n,i} \equiv \text{Prob}\left(U_{n,i} \geq U_{n,j}, \forall j \in \mathcal{C}_n\right) = \frac{\exp(u_{n,i})}{1 + \sum_{j \in \mathcal{A}_{t(n)}} \exp(u_{n,j})}, \quad \forall i \in \mathcal{C}_n. \tag{5}$$

And the estimation involves maximizing $\ell\left(\boldsymbol{\Theta}_{\text{I}}\right)$ as specified in (3) where $\boldsymbol{\Theta}_{\text{I}} = \boldsymbol{\theta}_{\text{logit}}$,

$$\ell\left(\boldsymbol{\Theta}_{\text{I}}\right) = \sum_{n \in \mathcal{N}} \sum_{i \in \mathcal{C}_n} y_{n,i} \cdot \log(P_{n,i}). \tag{6}$$

Result of Model I - Baseline Logit

(24 parameters)

	(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	
Ave 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 Ave LL Per User	−0.848		−0.528		−0.526	

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

Result of Model I - Baseline Logit

(24 parameters)

	(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	
Ave 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 Ave LL Per User	−0.848		−0.528		−0.526	

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

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	
Ave 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 Ave LL Per User	−0.848		−0.528		−0.526	

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

AIC and BIC

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	
Ave 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 Ave LL Per User	−0.848		−0.528		−0.526	

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

AIC and BIC

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	
Ave 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 Ave LL Per User	−0.848		−0.528		−0.526	

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

AIC and BIC

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	
Ave 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 Ave LL Per User	−0.848		−0.528		−0.526	

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

AIC and BIC

	(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)

	(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)

	(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)

	(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)

	(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)

	(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)

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	
Ave 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 Ave LL Per User	−0.848		−0.528		−0.526	

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

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	
Ave 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 Ave LL Per User	−0.848		−0.528		−0.526	

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

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	
Ave 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 Ave LL Per User	−0.848		−0.528		−0.526	

Standard errors in parentheses

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

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	
Ave 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 Ave LL Per User	−0.848		−0.528		−0.526	

Standard errors in parentheses

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

Model II - by User Types

Each user belongs to exactly one of

$$\mathcal{H} = \{no\ search, p\ search, x\ search, both\ search\}. \tag{7}$$

For each user $n \in \mathcal{N}$, the utility is assumed to have the following form

$$U_{n,i} \equiv u_{n,i} + \epsilon_{n,i} = \alpha_{h(n)} \cdot p_{t(n),i} + \boldsymbol{\beta}_{h(n)}{}^{\mathsf{T}} \mathbf{x}_{n,t(n),i} + \epsilon_{n,i}, \quad \forall i \in \mathcal{C}_n. \tag{8}$$

where

- $h(n) : \mathcal{N} \to \mathcal{H}$ represents the user type of n,
- $\boldsymbol{\theta}_m = \left[\alpha_m, \boldsymbol{\beta}_m{}^{\mathsf{T}}\right]^{\mathsf{T}}$, $\forall m \in \mathcal{H}$ is the vector of coefficients.

The conditional choice probability is as specified in (2),

$$P_{n,i} \equiv \text{Prob}\left(U_{n,i} \geq U_{n,j}, \forall j \in \mathcal{C}_n\right) = \frac{\exp(u_{n,i})}{1 + \sum_{j \in \mathcal{A}_{t(n)}} \exp(u_{n,j})}, \quad \forall i \in \mathcal{C}_n. \tag{9}$$

And the estimation involves maximizing $\ell\left(\boldsymbol{\Theta}_{\text{II}}\right)$ as specified in (3) where $\boldsymbol{\Theta}_{\text{II}} = \left[\boldsymbol{\theta}_m{}^{\mathsf{T}}\right]^{\mathsf{T}}_{m \in \mathcal{H}}$,

$$\ell\left(\boldsymbol{\Theta}_{\text{II}}\right) = \sum_{n \in \mathcal{N}} \sum_{i \in \mathcal{C}_n} y_{n,i} \cdot \log(P_{n,i}). \tag{10}$$

Model II - by User Types

Each user belongs to exactly one of

$$\mathcal{H} = \{no\ search, p\ search, x\ search, both\ search\}. \tag{7}$$

For each user $n \in \mathcal{N}$, the utility is assumed to have the following form

$$U_{n,i} \equiv u_{n,i} + \epsilon_{n,i} = \alpha_{h(n)} \cdot p_{t(n),i} + \boldsymbol{\beta}_{h(n)}{}^{\mathsf{T}} \mathbf{x}_{n,t(n),i} + \epsilon_{n,i}, \quad \forall i \in \mathcal{C}_n. \tag{8}$$

where

- $h(n) : \mathcal{N} \to \mathcal{H}$ represents the user type of n,
- $\boldsymbol{\theta}_m = \left[\alpha_m, \boldsymbol{\beta}_m{}^{\mathsf{T}}\right]^{\mathsf{T}}$, $\forall m \in \mathcal{H}$ is the vector of coefficients.

The conditional choice probability is as specified in (2),

$$P_{n,i} \equiv \text{Prob}\left(U_{n,i} \geq U_{n,j}, \forall j \in \mathcal{C}_n\right) = \frac{\exp(u_{n,i})}{1 + \sum_{j \in \mathcal{A}_{t(n)}} \exp(u_{n,j})}, \quad \forall i \in \mathcal{C}_n. \tag{9}$$

And the estimation involves maximizing $\ell\left(\boldsymbol{\Theta}_{\text{II}}\right)$ as specified in (3) where $\boldsymbol{\Theta}_{\text{II}} = \left[\boldsymbol{\theta}_m{}^{\mathsf{T}}\right]^{\mathsf{T}}_{m \in \mathcal{H}}$,

$$\ell\left(\boldsymbol{\Theta}_{\text{II}}\right) = \sum_{n \in \mathcal{N}} \sum_{i \in \mathcal{C}_n} y_{n,i} \cdot \log(P_{n,i}). \tag{10}$$

Model II - by User Types

Each user belongs to exactly one of

$$\mathcal{H} = \{no\ search, p\ search, x\ search, both\ search\}. \tag{7}$$

For each user $n \in \mathcal{N}$, the utility is assumed to have the following form

$$U_{n,i} \equiv u_{n,i} + \epsilon_{n,i} = \alpha_{h(n)} \cdot p_{t(n),i} + \boldsymbol{\beta}_{h(n)}{}^{\mathsf{T}} \mathbf{x}_{n,t(n),i} + \epsilon_{n,i}, \quad \forall i \in \mathcal{C}_n. \tag{8}$$

where

- $h(n) : \mathcal{N} \to \mathcal{H}$ represents the user type of n,
- $\boldsymbol{\theta}_m = \left[\alpha_m, \boldsymbol{\beta}_m{}^{\mathsf{T}}\right]^{\mathsf{T}}$, $\forall m \in \mathcal{H}$ is the vector of coefficients.

The conditional choice probability is as specified in (2),

$$P_{n,i} \equiv \text{Prob}\left(U_{n,i} \geq U_{n,j}, \forall j \in \mathcal{C}_n\right) = \frac{\exp(u_{n,i})}{1 + \sum_{j \in \mathcal{A}_{t(n)}} \exp(u_{n,j})}, \quad \forall i \in \mathcal{C}_n. \tag{9}$$

And the estimation involves maximizing $\ell\left(\boldsymbol{\Theta}_{\text{II}}\right)$ as specified in (3) where $\boldsymbol{\Theta}_{\text{II}} = \left[\boldsymbol{\theta}_m{}^{\mathsf{T}}\right]_{m \in \mathcal{H}}^{\mathsf{T}}$,

$$\ell\left(\boldsymbol{\Theta}_{\text{II}}\right) = \sum_{n \in \mathcal{N}} \sum_{i \in \mathcal{C}_n} y_{n,i} \cdot \log(P_{n,i}). \tag{10}$$

Result of Model II - by User Types

(96 parameters)

	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				
Ave 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 Ave LL Per User				−0.404				

Standard errors in parentheses

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

Result of Model II - by User Types

(96 parameters)

	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							
Ave 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 Ave LL Per User	−0.404							

Standard errors in parentheses

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

	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)

	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)

	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)

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							
Ave 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 Ave LL Per User	−0.404							

Standard errors in parentheses

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

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							
Ave 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 Ave LL Per User	−0.404							

Standard errors in parentheses

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

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				
Ave 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 Ave LL Per User				−0.404				

Standard errors in parentheses

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

Model III - with Unobserved Heterogeneity

Draws upon Heckman and Singer (1984), this model posits a discrete distribution of preferences within each user type in $\mathcal{H}$. For the purposes of this analysis, set $\mathcal{S} = \{1, 2\}$.

Denote the probability of user n's sub-type s as $\pi_{n,s}, \ \forall s \in \mathcal{S}$.

New parameter γ:
$$\pi_{n,1} = sigmoid(\gamma_{h(n)}) = \frac{1}{1 + \exp(-\gamma_{h(n)})} \tag{11}$$

The utility $U_{n,s,i}$ as from the generic form (1) with the additional subscript $s \in \mathcal{S}$ is hence,
$$U_{n,s,i} \equiv u_{n,s,i} + \epsilon_{n,i} = \alpha_{h(n),s} \cdot p_{t(n),i} + {\boldsymbol{\beta}_{h(n),s}}^{\top} \mathbf{x}_{n,t(n),i} + \epsilon_{n,i}, \ \ \forall i \in \mathcal{C}_n. \tag{12}$$

where $\boldsymbol{\theta}_m = \left[\alpha_{m,s}, {\boldsymbol{\beta}_{m,s}}^{\top}, \gamma_m\right]^{\top}_{s \in \mathcal{S}}$, $\forall m \in \mathcal{H}$ and $u_{n,s,i} = 0 \ \ \forall s \in \mathcal{S}$.

The conditional choice probability is a weighted average,

$$P_{n,i} = \sum_{s \in \mathcal{S}} \pi_{n,s} \cdot \frac{\exp(u_{n,s,i})}{1 + \sum_{j \in \mathcal{A}_{t(n)}} \exp(u_{n,s,j})}, \ \ \forall i \in \mathcal{C}_n. \tag{13}$$

And the estimation involves maximizing $\ell\left(\boldsymbol{\Theta}_{\text{III}}\right)$ as specified in (3) where $\boldsymbol{\Theta}_{\text{III}} = \left[{\boldsymbol{\theta}_m}^{\top}\right]^{\top}_{m \in \mathcal{H}}$,

$$\ell\left(\boldsymbol{\Theta}_{\text{III}}\right) = \sum_{n \in \mathcal{N}} \sum_{i \in \mathcal{C}_n} y_{n,i} \cdot \log(P_{n,i}). \tag{14}$$

Model III - with Unobserved Heterogeneity

Draws upon Heckman and Singer (1984), this model posits a discrete distribution of preferences within each user type in $\mathcal{H}$. For the purposes of this analysis, set $\mathcal{S} = \{1, 2\}$.

Denote the probability of user n's sub-type s as $\pi_{n,s},\ \forall s \in \mathcal{S}$.

New parameter γ:
$$\pi_{n,1} = sigmoid\left(\gamma_{h(n)}\right) = \frac{1}{1 + \exp(-\gamma_{h(n)})} \tag{11}$$

The utility $U_{n,s,i}$ as from the generic form (1) with the additional subscript $s \in \mathcal{S}$ is hence,
$$U_{n,s,i} \equiv u_{n,s,i} + \epsilon_{n,i} = \alpha_{h(n),s} \cdot p_{t(n),i} + {\boldsymbol{\beta}_{h(n),s}}^{\mathsf{T}} \mathbf{x}_{n,t(n),i} + \epsilon_{n,i},\ \ \forall i \in \mathcal{C}_n. \tag{12}$$

where $\boldsymbol{\theta}_m = \left[\alpha_{m,s}, {\boldsymbol{\beta}_{m,s}}^{\mathsf{T}}, \gamma_m\right]^{\mathsf{T}}_{s \in \mathcal{S}},\ \forall m \in \mathcal{H}$ and $u_{n,s,i} = 0\ \ \forall s \in \mathcal{S}$.

The conditional choice probability is a weighted average,
$$P_{n,i} = \sum_{s \in \mathcal{S}} \pi_{n,s} \cdot \frac{\exp(u_{n,s,i})}{1 + \sum_{j \in \mathcal{A}_{t(n)}} \exp(u_{n,s,j})},\ \ \forall i \in \mathcal{C}_n. \tag{13}$$

And the estimation involves maximizing $\ell\left(\boldsymbol{\Theta}_{\text{III}}\right)$ as specified in (3) where $\boldsymbol{\Theta}_{\text{III}} = \left[{\boldsymbol{\theta}_m}^{\mathsf{T}}\right]^{\mathsf{T}}_{m \in \mathcal{H}}$,
$$\ell\left(\boldsymbol{\Theta}_{\text{III}}\right) = \sum_{n \in \mathcal{N}} \sum_{i \in \mathcal{C}_n} y_{n,i} \cdot \log(P_{n,i}). \tag{14}$$

Model III - with Unobserved Heterogeneity

Draws upon Heckman and Singer (1984), this model posits a discrete distribution of preferences within each user type in $\mathcal{H}$. For the purposes of this analysis, set $\mathcal{S} = \{1, 2\}$.

Denote the probability of user n's sub-type s as $\pi_{n,s}, \ \forall s \in \mathcal{S}$.

New parameter γ:

$$\pi_{n,1} = sigmoid(\gamma_{h(n)}) = \frac{1}{1 + \exp(-\gamma_{h(n)})} \tag{11}$$

The utility $U_{n,s,i}$ as from the generic form (1) with the additional subscript $s \in \mathcal{S}$ is hence,

$$U_{n,s,i} \equiv u_{n,s,i} + \epsilon_{n,i} = \alpha_{h(n),s} \cdot p_{t(n),i} + \boldsymbol{\beta}_{h(n),s}{}^{\mathsf{T}} \mathbf{x}_{n,t(n),i} + \epsilon_{n,i}, \ \forall i \in \mathcal{C}_n. \tag{12}$$

where $\boldsymbol{\theta}_m = \left[\alpha_{m,s}, \boldsymbol{\beta}_{m,s}{}^{\mathsf{T}}, \gamma_m\right]^{\mathsf{T}}_{s \in \mathcal{S}}$, $\forall m \in \mathcal{H}$ and $u_{n,s,i} = 0 \ \forall s \in \mathcal{S}$.

The conditional choice probability is a weighted average,

$$P_{n,i} = \sum_{s \in \mathcal{S}} \pi_{n,s} \cdot \frac{\exp(u_{n,s,i})}{1 + \sum_{j \in \mathcal{A}_{t(n)}} \exp(u_{n,s,j})}, \ \forall i \in \mathcal{C}_n. \tag{13}$$

And the estimation involves maximizing $\ell(\boldsymbol{\Theta}_{\mathrm{III}})$ as specified in (3) where $\boldsymbol{\Theta}_{\mathrm{III}} = \left[\boldsymbol{\theta}_m{}^{\mathsf{T}}\right]^{\mathsf{T}}_{m \in \mathcal{H}}$,

$$\ell(\boldsymbol{\Theta}_{\mathrm{III}}) = \sum_{n \in \mathcal{N}} \sum_{i \in \mathcal{C}_n} y_{n,i} \cdot \log(P_{n,i}). \tag{14}$$

Model III - with Unobserved Heterogeneity

Draws upon Heckman and Singer (1984), this model posits a discrete distribution of preferences within each user type in $\mathcal{H}$. For the purposes of this analysis, set $\mathcal{S} = \{1, 2\}$.

Denote the probability of user n's sub-type s as $\pi_{n,s},\ \forall s \in \mathcal{S}$.

New parameter γ:
$$\pi_{n,1} = sigmoid\left(\gamma_{h(n)}\right) = \frac{1}{1 + \exp(-\gamma_{h(n)})} \tag{11}$$

The utility $U_{n,s,i}$ as from the generic form (1) with the additional subscript $s \in \mathcal{S}$ is hence,
$$U_{n,s,i} \equiv u_{n,s,i} + \epsilon_{n,i} = \alpha_{h(n),s} \cdot p_{t(n),i} + \boldsymbol{\beta}_{h(n),s}{}^{\mathsf{T}} \mathbf{x}_{n,t(n),i} + \epsilon_{n,i},\quad \forall i \in \mathcal{C}_n. \tag{12}$$

where $\boldsymbol{\theta}_m = \left[\alpha_{m,s}, \boldsymbol{\beta}_{m,s}{}^{\mathsf{T}}, \gamma_m\right]_{s\in\mathcal{S}}^{\mathsf{T}}$, $\forall m \in \mathcal{H}$ and $u_{n,s,i} = 0\ \ \forall s \in \mathcal{S}$.

The conditional choice probability is a weighted average,
$$P_{n,i} = \sum_{s\in\mathcal{S}} \pi_{n,s} \cdot \frac{\exp(u_{n,s,i})}{1 + \sum_{j\in\mathcal{A}_{t(n)}} \exp(u_{n,s,j})},\quad \forall i \in \mathcal{C}_n. \tag{13}$$

And the estimation involves maximizing $\ell\left(\boldsymbol{\Theta}_{\text{III}}\right)$ as specified in (3) where $\boldsymbol{\Theta}_{\text{III}} = \left[\boldsymbol{\theta}_m{}^{\mathsf{T}}\right]_{m\in\mathcal{H}}^{\mathsf{T}}$,
$$\ell\left(\boldsymbol{\Theta}_{\text{III}}\right) = \sum_{n\in\mathcal{N}} \sum_{i\in\mathcal{C}_n} y_{n,i} \cdot \log(P_{n,i}). \tag{14}$$

Result of Model III - with Unobserved Heterogeneity
(196 parameters)

	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															
Ave 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 Ave LL Per User	−0.301															

Standard errors in parentheses

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

Result of Model III - with Unobserved Heterogeneity

(196 parameters)

	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															
Ave LL Per User	−1.301															
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 Ave LL Per User	−1.301															

Standard errors in parentheses

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

	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***	−1.516***	−1.045***	−1.869*	−1.071***	−1.503***	−1.361***	−1.920***
User interaction								
# of *to page* view clicks	16.907***	1.441***	11.634***	0.461***	10.808***	1.608***	11.148***	0.688***
# of *on page* view clicks	0.046***	0.186***	0.200***	0.090***	0.123***	0.179***	0.211***	0.150***
Avg uninterrupted view time (min)	−1.255***	1.765***	0.519***	2.706***	−0.674***	1.094***	−0.684***	1.255***
# of interrupted view clicks	−20.694***	0.076	−11.955***	−0.226**	−10.880***	−0.610***	−11.073***	−0.759***
Time-dependent features								
No reviews (indicator)	−1.257***	−2.135**	−0.785	−3.001	−1.108***	−3.191***	−0.479**	−1.953***
Log # of reviews	−0.112***	1.077***	0.165***	0.978***	−0.016*	0.660***	0.039***	0.568***
Star rating	−0.013	−0.003	0.055	−0.004	−0.030	−0.032	0.048	0.040
Log days since available	−0.295***	−1.513***	−0.652***	−1.445***	−0.312***	−1.384***	−0.399***	−1.159***
Brands (indicators)								
Samsung	0.825***	1.731***	1.055***	1.762**	0.797***	1.252***	0.715***	1.639***
Apple	1.223***	3.797***	2.136***	3.919***	1.619***	3.208***	1.933***	3.693***
Xiaomi	0.653***	0.256	0.593***	0.336	0.599***	0.351***	0.425***	0.712***
Huawei	0.958***	1.301***	1.119***	0.728	0.983***	1.189***	0.730***	1.639***
Oppo	1.394***	1.184***	1.278***	1.109	0.986***	1.084***	0.861***	1.158***
Product specifications								
Log storage (G)	0.005	−0.028	0.018	0.002	−0.000	−0.038	0.035	0.007
Log RAM (G)	−0.015	−0.001	0.042	0.024	0.005	0.005	0.004	−0.007
Weight (100 g)	−1.299***	−1.199***	−1.027***	−1.523	−1.491***	−2.788***	−1.062***	−1.305***
Diagonal (in)	−0.015	0.078	0.008	0.116	−0.015	0.003	0.009	0.342***
Battery (1,000 mAh)	0.107***	0.413***	0.205*	0.236	0.245***	0.561***	0.313***	0.479***

	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***	−1.516***	−1.045***	−1.869*	−1.071***	−1.503***	−1.361***	−1.920***
User interaction								
# of *to page* view clicks	16.907***	1.441***	11.634***	0.461***	10.808***	1.608***	11.148***	0.688***
# of *on page* view clicks	0.046***	0.186***	0.200***	0.090***	0.123***	0.179***	0.211***	0.150***
Avg uninterrupted view time (min)	−1.255***	1.765***	0.519***	2.706***	−0.674***	1.094***	−0.684***	1.255***
# of interrupted view clicks	−20.694***	0.076	−11.955***	−0.226**	−10.880***	−0.610***	−11.073***	−0.759***
Time-dependent features								
No reviews (indicator)	−1.257***	−2.135**	−0.785	−3.001	−1.108***	−3.191***	−0.479**	−1.953***
Log # of reviews	−0.112***	1.077***	0.165***	0.978***	−0.016*	0.660***	0.039***	0.568***
Star rating	−0.013	−0.003	0.055	−0.004	−0.030	−0.032	0.048	0.040
Log days since available	−0.295***	−1.513***	−0.652***	−1.445***	−0.312***	−1.384***	−0.399***	−1.159***
Brands (indicators)								
Samsung	0.825***	1.731***	1.055***	1.762**	0.797***	1.252***	0.715***	1.639***
Apple	1.223***	3.797***	2.136***	3.919***	1.619***	3.208***	1.933***	3.693***
Xiaomi	0.653***	0.256	0.593***	0.336	0.599***	0.351***	0.425***	0.712***
Huawei	0.958***	1.301***	1.119***	0.728	0.983***	1.189***	0.730***	1.639***
Oppo	1.394***	1.184***	1.278***	1.109	0.986***	1.084***	0.861***	1.158***
Product specifications								
Log storage (G)	0.005	−0.028	0.018	0.002	−0.000	−0.038	0.035	0.007
Log RAM (G)	−0.015	−0.001	0.042	0.024	0.005	0.005	0.004	−0.007
Weight (100 g)	−1.299***	−1.199***	−1.027***	−1.523	−1.491***	−2.788***	−1.062***	−1.305***
Diagonal (in)	−0.015	0.078	0.008	0.116	−0.015	0.003	0.009	0.342***
Battery (1,000 mAh)	0.107***	0.413***	0.205*	0.236	0.245***	0.561***	0.313***	0.479***

Apple	1.223***	3.797***	2.136***	3.919***	1.619***	3.208***	1.933***	3.693***
Xiaomi	0.653***	0.256	0.593***	0.336	0.599***	0.351***	0.425***	0.712***
Huawei	0.958***	1.301***	1.119***	0.728	0.983***	1.189***	0.730***	1.639***
Oppo	1.394***	1.184***	1.278***	1.109	0.986***	1.084***	0.861***	1.158***
Product specifications								
Log storage (G)	0.005	−0.028	0.018	0.002	−0.000	−0.038	0.035	0.007
Log RAM (G)	−0.015	−0.001	0.042	0.024	0.005	0.005	0.004	−0.007
Weight (100 g)	−1.299***	−1.199***	−1.027***	−1.523	−1.491***	−2.788***	−1.062***	−1.305***
Diagonal (in)	−0.015	0.078	0.008	0.116	−0.015	0.003	0.009	0.342***
Battery (1,000 mAh)	0.107***	0.413***	0.205*	0.236	0.245***	0.561***	0.313***	0.479***
Log highest camera pixels (MP)	−0.002	0.003	−0.006	0.002	−0.036	−0.018	0.009	0.040*
# of cameras (back)	0.007	0.325***	−0.022	0.237	0.121***	0.250***	0.117***	0.245***
Bundled gadgets (indicators)								
Headphones	−0.612***	0.693**	0.350	−1.298	−0.138	−0.133	0.072	0.062
Other gadgets	−0.643***	0.371*	−0.802***	−0.155	0.029	0.224**	−0.019	0.017
Constant (inside options)	−11.530***	−9.822***	−9.994***	−9.831**	−8.729***	−4.308***	−9.940***	−9.586***
γ (inv-sigmoid of sub-type 1 prob)	0.783***		−0.003		−1.633***		−1.852***	
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							
Ave 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 Ave LL Per User	−0.301							

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

	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***	−1.516***	−1.045***	−1.869*	−1.071***	−1.503***	−1.361***	−1.920***
User interaction								
# of *to page* view clicks	16.907***	1.441***	11.634***	0.461***	10.808***	1.608***	11.148***	0.688***
# of *on page* view clicks	0.046***	0.186***	0.200***	0.090***	0.123***	0.179***	0.211***	0.150***
Avg uninterrupted view time (min)	−1.255***	1.765***	0.519***	2.706***	−0.674***	1.094***	−0.684***	1.255***
# of interrupted view clicks	−20.694***	0.076	−11.955***	−0.226**	−10.880***	−0.610***	−11.073***	−0.759***
Time-dependent features								
No reviews (indicator)	−1.257***	−2.135**	−0.785	−3.001	−1.108***	−3.191***	−0.479**	−1.953***
Log # of reviews	−0.112***	1.077***	0.165***	0.978***	−0.016*	0.660***	0.039***	0.568***
Star rating	−0.013	−0.003	0.055	−0.004	−0.030	−0.032	0.048	0.040
Log days since available	−0.295***	−1.513***	−0.652***	−1.445***	−0.312***	−1.384***	−0.399***	−1.159***
Brands (indicators)								
Samsung	0.825***	1.731***	1.055***	1.762**	0.797***	1.252***	0.715***	1.639***
Apple	1.223***	3.797***	2.136***	3.919***	1.619***	3.208***	1.933***	3.693***
Xiaomi	0.653***	0.256	0.593***	0.336	0.599***	0.351***	0.425***	0.712***
Huawei	0.958***	1.301***	1.119***	0.728	0.983***	1.189***	0.730***	1.639***
Oppo	1.394***	1.184***	1.278***	1.109	0.986***	1.084***	0.861***	1.158***
Product specifications								
Log storage (G)	0.005	−0.028	0.018	0.002	−0.000	−0.038	0.035	0.007
Log RAM (G)	−0.015	−0.001	0.042	0.024	0.005	0.005	0.004	−0.007
Weight (100 g)	−1.299***	−1.199***	−1.027***	−1.523	−1.491***	−2.788***	−1.062***	−1.305***
Diagonal (in)	−0.015	0.078	0.008	0.116	−0.015	0.003	0.009	0.342***
Battery (1,000 mAh)	0.107***	0.413***	0.205*	0.236	0.245***	0.561***	0.313***	0.479***
Log highest camera pixels (MP)	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Apple	1.223***	3.797***	2.136***	3.919***	1.619***	3.208***	1.933***	3.693***
Xiaomi	0.653***	0.256	0.593***	0.336	0.599***	0.351***	0.425***	0.712***
Huawei	0.958***	1.301***	1.119***	0.728	0.983***	1.189***	0.730***	1.639***
Oppo	1.394***	1.184***	1.278***	1.109	0.986***	1.084***	0.861***	1.158***
Product specifications								
Log storage (G)	0.005	−0.028	0.018	0.002	−0.000	−0.038	0.035	0.007
Log RAM (G)	−0.015	−0.001	0.042	0.024	0.005	0.005	0.004	−0.007
Weight (100 g)	−1.299***	−1.199***	−1.027***	−1.523	−1.491***	−2.788***	−1.062***	−1.305***
Diagonal (in)	−0.015	0.078	0.008	0.116	−0.015	0.003	0.009	0.342***
Battery (1,000 mAh)	0.107***	0.413***	0.205*	0.236	0.245***	0.561***	0.313***	0.479***
Log highest camera pixels (MP)	−0.002	0.003	−0.006	0.002	−0.036	−0.018	0.009	0.040*
# of cameras (back)	0.007	0.325***	−0.022	0.237	0.121***	0.250***	0.117***	0.245***
Bundled gadgets (indicators)								
Headphones	−0.612***	0.693**	0.350	−1.298	−0.138	−0.133	0.072	0.062
Other gadgets	−0.643***	0.371*	−0.802***	−0.155	0.029	0.224**	−0.019	0.017
Constant (inside options)	−11.530***	−9.822***	−9.994***	−9.831**	−8.729***	−4.308***	−9.940***	−9.586***
γ (inv-sigmoid of sub-type 1 prob)	0.783***		−0.003		−1.633***		−1.852***	
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				
Ave 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 Ave LL Per User				−0.301				

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

Models with Individual-level Heterogeneity
Models with MLP and RNN

Unlike Model III which estimates $\alpha_{m,s}$, $\boldsymbol{\beta}_{m,s}$, and γ_m for $m \in \mathcal{H}$ and $s \in \mathcal{S}$, models with individual-level heterogeneity aims to identify

$$\alpha_{n,s}, \boldsymbol{\beta}_{n,s}, \text{ and } \gamma_n \text{ for } n \in \mathcal{N} \text{ and } s \in \mathcal{S}.$$

Let $\boldsymbol{\theta}_n = \left[\alpha_{n,s}, \boldsymbol{\beta}_{n,s}{}^{\top}, \gamma_n\right]^{\top}_{s \in \mathcal{S}}$, and $\pi_{n,1} = \mathit{sigmoid}(\gamma_n)$.

Drawing inspiration from the innovative approach proposed by Farrell et al. (2021), let

$$G(\mathbf{z}_n; \boldsymbol{\Theta}) = \boldsymbol{\theta}_n, \tag{15}$$

where $\mathbf{z}_n$ is a generic notation for individual-specific inputs for the ML model, G, and $\boldsymbol{\Theta}$ is a generic notation for the set of parameters to be estimated.

This study uses two ML models for G, MLP and RNN.
Note that $\boldsymbol{\theta}_n$ are not "predictions", they are parameter with economic meanings.

Models with Individual-level Heterogeneity

Models with MLP and RNN

Unlike Model III which estimates $\alpha_{m,s}$, $\boldsymbol{\beta}_{m,s}$, and γ_m for $m \in \mathcal{H}$ and $s \in \mathcal{S}$, models with individual-level heterogeneity aims to identify

$$\alpha_{n,s}, \boldsymbol{\beta}_{n,s}, \text{ and } \gamma_n \text{ for } n \in \mathcal{N} \text{ and } s \in \mathcal{S}.$$

Let $\boldsymbol{\theta}_n = \left[\alpha_{n,s}, {\boldsymbol{\beta}_{n,s}}^{\mathsf{T}}, \gamma_n\right]^{\mathsf{T}}_{s\in\mathcal{S}}$, and $\pi_{n,1} = sigmoid(\gamma_n)$.

Drawing inspiration from the innovative approach proposed by Farrell et al. (2021), let

$$G(\mathbf{z}_n; \boldsymbol{\Theta}) = \boldsymbol{\theta}_n, \tag{15}$$

where $\mathbf{z}_n$ is a generic notation for individual-specific inputs for the ML model, G, and $\boldsymbol{\Theta}$ is a generic notation for the set of parameters to be estimated.

This study uses two ML models for G, MLP and RNN.
Note that $\boldsymbol{\theta}_n$ are not "predictions", they are parameter with economic meanings.

Models with Individual-level Heterogeneity
Models with MLP and RNN

Unlike Model III which estimates $\alpha_{m,s}$, $\boldsymbol{\beta}_{m,s}$, and γ_m for $m \in \mathcal{H}$ and $s \in \mathcal{S}$, models with individual-level heterogeneity aims to identify

$$\alpha_{n,s}, \boldsymbol{\beta}_{n,s}, \text{ and } \gamma_n \text{ for } n \in \mathcal{N} \text{ and } s \in \mathcal{S}.$$

Let $\boldsymbol{\theta}_n = \left[\alpha_{n,s}, \boldsymbol{\beta}_{n,s}{}^{\mathsf{T}}, \gamma_n\right]^{\mathsf{T}}_{s\in\mathcal{S}}$, and $\pi_{n,1} = \mathit{sigmoid}(\gamma_n)$.

Drawing inspiration from the innovative approach proposed by Farrell et al. (2021), let

$$G(\mathbf{z}_n; \boldsymbol{\Theta}) = \boldsymbol{\theta}_n, \tag{15}$$

where $\mathbf{z}_n$ is a generic notation for individual-specific inputs for the ML model, G, and $\boldsymbol{\Theta}$ is a generic notation for the set of parameters to be estimated.

This study uses two ML models for G, MLP and RNN.
Note that $\boldsymbol{\theta}_n$ are not “predictions”, they are parameter with economic meanings.

Model IV - MLP

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. The vector $\mathbf{b}_n$ contains:

- aggregated features of historical clicks
  - the # of clicks, purchases, and dollars spent in each category (level 1)
  - the # of clicks and shares across hours of the day, days of the week, and months of the year
- features of the first smartphone event
  - event type indicators
  - brand indicators
  - time indicators (hours of the day, days of the week, and months of the year)
- user type indicators

$\mathbf{b}_n$ has a size of 206, establishing the MLP function as $G_{\text{MLP}} : \mathbb{R}^{206} \to \mathbb{R}^{49}$, and (15) is effectively

$$G_{\text{MLP}}(\mathbf{b}_n; \Theta_{\text{IV}}) = \boldsymbol{\theta}_{\text{MLP},n} \tag{16}$$

where $\boldsymbol{\theta}_{\text{MLP},n} = \left[\alpha_{\text{MLP},n,s}, {\boldsymbol{\beta}_{\text{MLP},n,s}}^{\top}, \gamma_{\text{MLP},n}\right]^{\top}_{s\in\mathcal{S}}$ has a size of 49, and $\pi_{n,1} = \mathit{sigmoid}(\gamma_{\text{MLP},n})$.

Model IV - MLP

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. The vector $\mathbf{b}_n$ contains:

- aggregated features of historical clicks
  - the # of clicks, purchases, and dollars spent in each category (level 1)
  - the # of clicks and shares across hours of the day, days of the week, and months of the year
- features of the first smartphone event
  - event type indicators
  - brand indicators
  - time indicators (hours of the day, days of the week, and months of the year)
- user type indicators

$\mathbf{b}_n$ has a size of 206, establishing the MLP function as $G_{\text{MLP}} : \mathbb{R}^{206} \rightarrow \mathbb{R}^{49}$, and (15) is effectively

$$G_{\text{MLP}}(\mathbf{b}_n; \mathbf{\Theta}_{\text{IV}}) = \boldsymbol{\theta}_{\text{MLP},n} \tag{16}$$

where $\boldsymbol{\theta}_{\text{MLP},n} = \left[\alpha_{\text{MLP},n,s}, {\boldsymbol{\beta}_{\text{MLP},n,s}}^{\mathsf{T}}, \gamma_{\text{MLP},n}\right]^{\mathsf{T}}_{s\in\mathcal{S}}$ has a size of 49, and $\pi_{n,1} = \textit{sigmoid}(\gamma_{\text{MLP},n})$.

Model IV - MLP
(cont.)

The utility $U_{n,s,i}$ as from the generic form (1) with the additional subscript $s \in \mathcal{S}$ is hence,

$$U_{n,s,i} \equiv u_{n,s,i} + \epsilon_{n,i} = \alpha_{\text{MLP},n,s} \cdot p_{t(n),i} + {\boldsymbol{\beta}_{\text{MLP},n,s}}^{\mathsf{T}} \mathbf{x}_{n,t(n),i} + \epsilon_{n,i}, \quad \forall i \in \mathcal{C}_n. \tag{17}$$

The conditional choice probability is a weighted average,

$$P_{n,i} = \sum_{s \in \mathcal{S}} \pi_{n,s} \cdot \frac{\exp(u_{n,s,i})}{1 + \sum_{j \in \mathcal{A}_{t(n)}} \exp(u_{n,s,j})}, \quad \forall i \in \mathcal{C}_n. \tag{18}$$

And the estimation involves maximizing $\ell\left(\mathbf{\Theta}_{\text{IV}}\right)$ as specified in (3),

$$\ell\left(\mathbf{\Theta}_{\text{IV}}\right) = \sum_{n \in \mathcal{N}} \sum_{i \in \mathcal{C}_n} y_{n,i} \cdot \log(P_{n,i}). \tag{19}$$

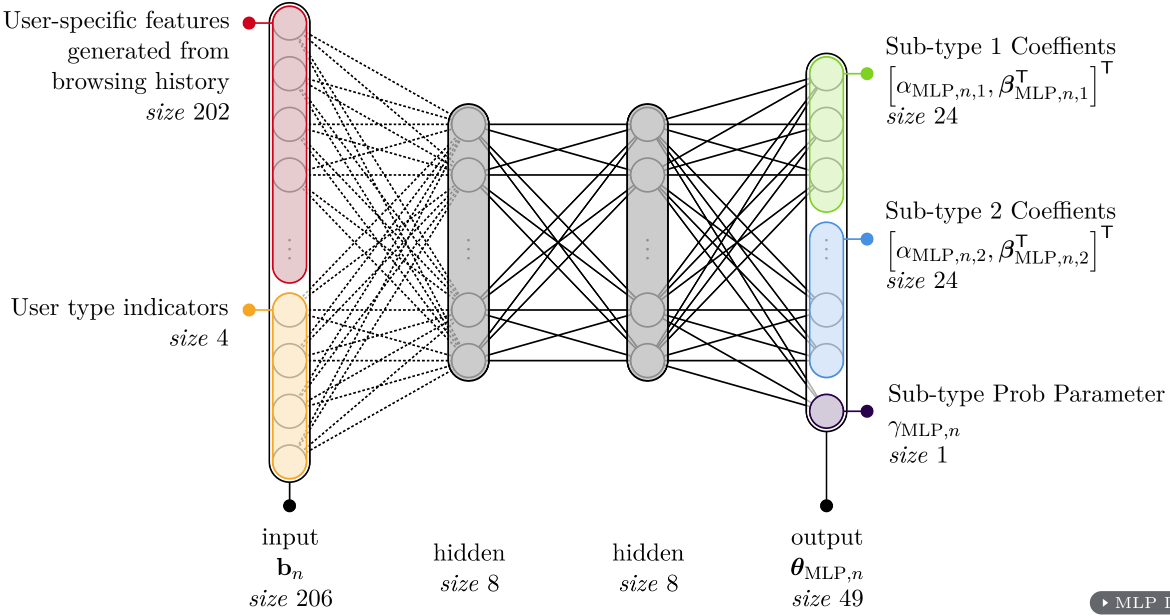
Model IV - MLP
MLP Structure

Many terms are used to refer to essentially the same model structure.

- Multi-layer perceptron (MLP)
- Deep neural network (DNN)
- Artificial neural network (ANN)
- Fully-connected feed forward neural network (FCFFNN)

Model IV - MLP

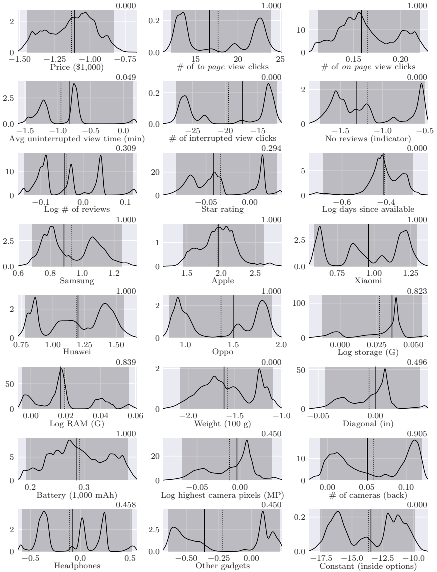
MLP Structure *(2,161 parameters)*



▸ MLP Details

Result of Model IV - MLP

(2,161 parameters)



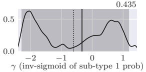
Sub-type 1

Estimation

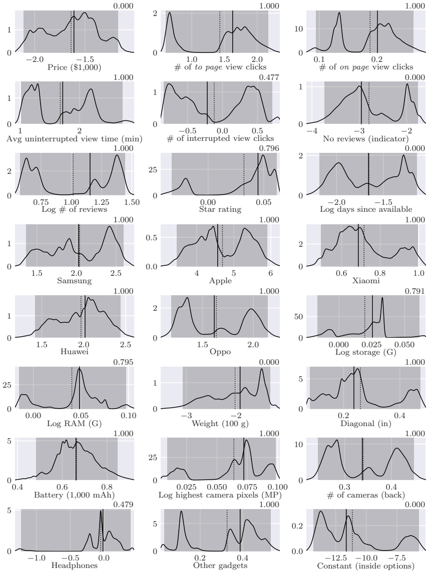
Observations	:	6,395,207,961
Users (Cases)	:	4,235,321
LL	:	−1,205,676.6
Ave LL Per User	:	−0.285
AIC	:	2,415,675.2
BIC	:	2,444,327.9

Holdout

Observations	:	2,131,744,036
Users (Cases)	:	1,411,773
LL	:	−403,102.1
Ave LL Per User	:	−0.286



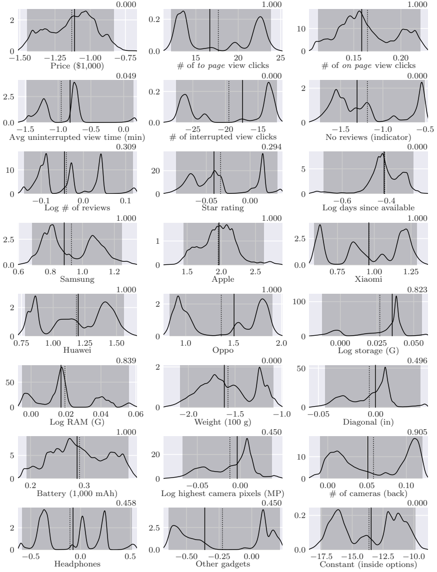
Parameter γ



Sub-type 2

Result of Model IV - MLP

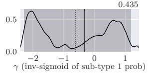
(2,161 parameters)



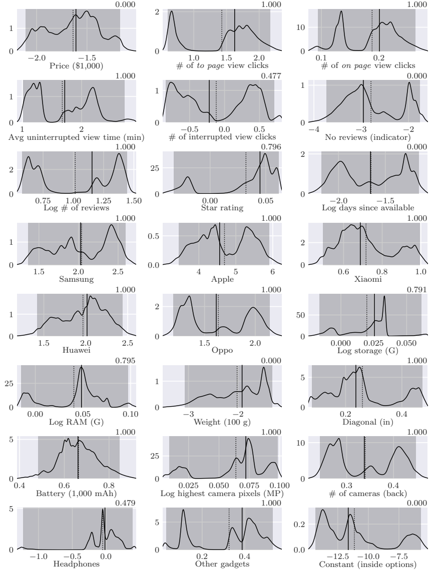
Sub-type 1

Estimation		
Observations	:	6,395,207,961
Users (Cases)	:	4,235,321
LL	:	−1,205,676.6
Ave 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
Ave LL Per User	:	−0.286

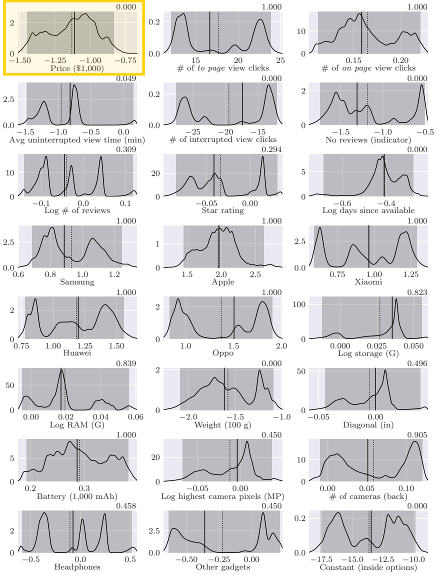


Parameter γ



Sub-type 2

Result of Model IV - MLP
(2,161 parameters)



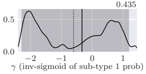
Sub-type 1

Estimation

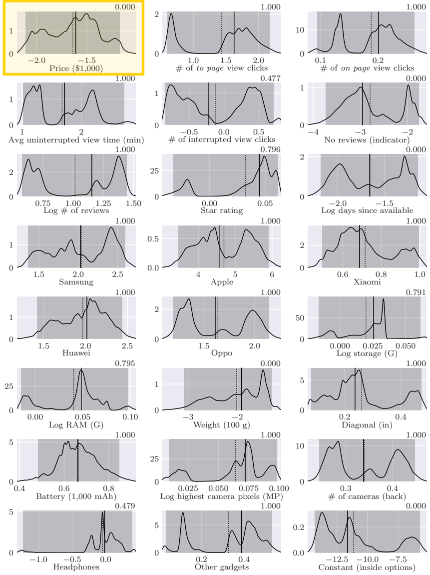
Observations	:	6,395,207,961
Users (Cases)	:	4,235,321
LL	:	−1,205,676.6
Ave LL Per User	:	−0.285
AIC	:	2,415,675.2
BIC	:	2,444,327.9

Holdout

Observations	:	2,131,744,036
Users (Cases)	:	1,411,773
LL	:	−403,102.1
Ave LL Per User	:	−0.286



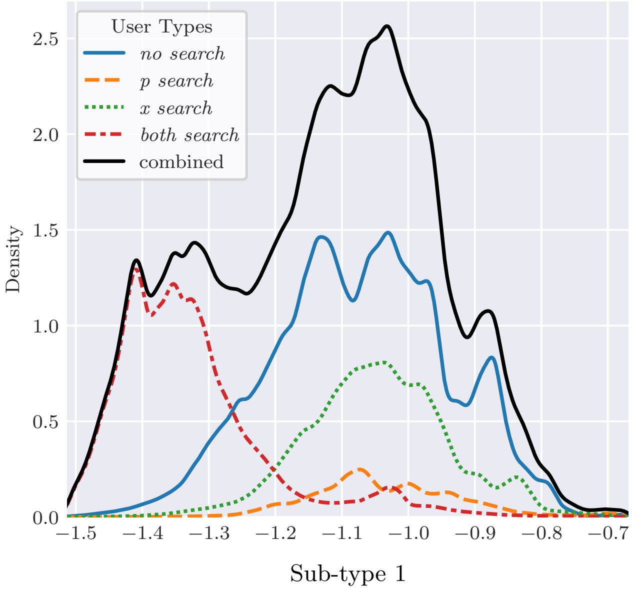
Parameter γ

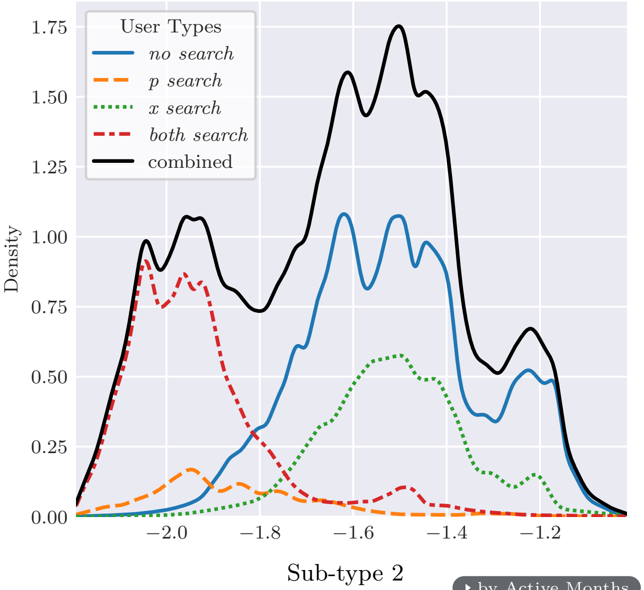


Sub-type 2

Result of Model IV - MLP

Distribution of Price ($1,000) Coefficients (by User Types)

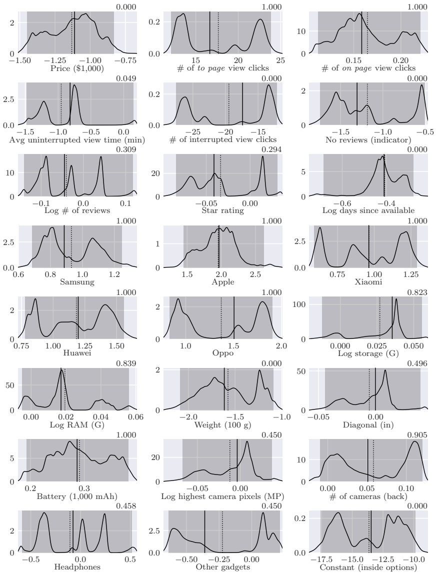




by Active Months

Result of Model IV - MLP

(2,161 parameters)



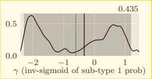
Sub-type 1

Estimation

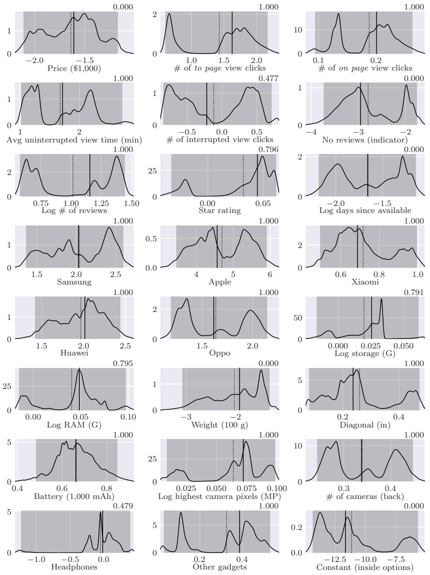
Observations	:	6,395,207,961
Users (Cases)	:	4,235,321
LL	:	$-1{,}205{,}676.6$
Ave LL Per User	:	-0.285
AIC	:	2,415,675.2
BIC	:	2,444,327.9

Holdout

Observations	:	2,131,744,036
Users (Cases)	:	1,411,773
LL	:	$-403{,}102.1$
Ave LL Per User	:	-0.286



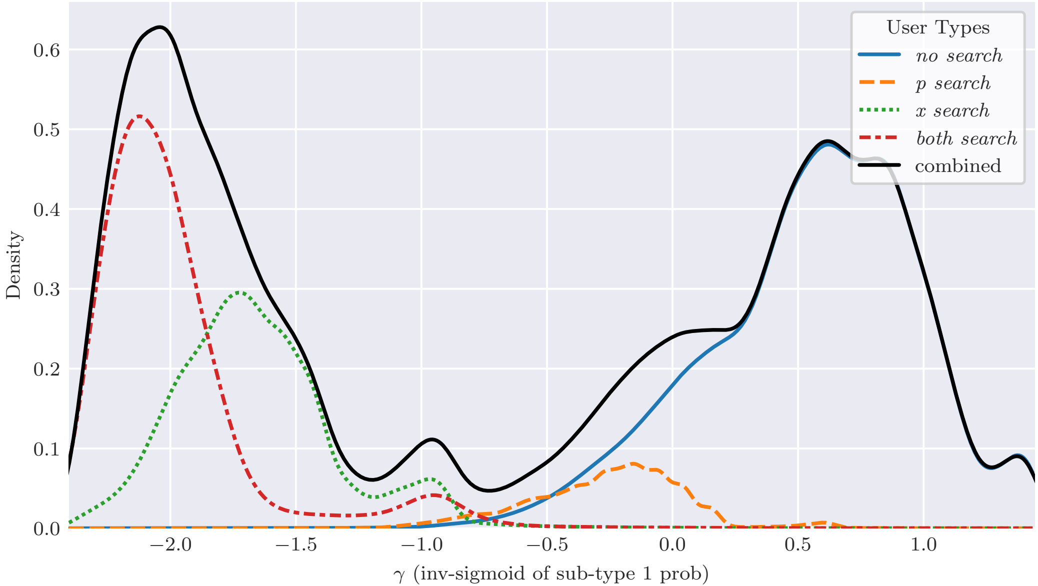
Parameter γ



Sub-type 2

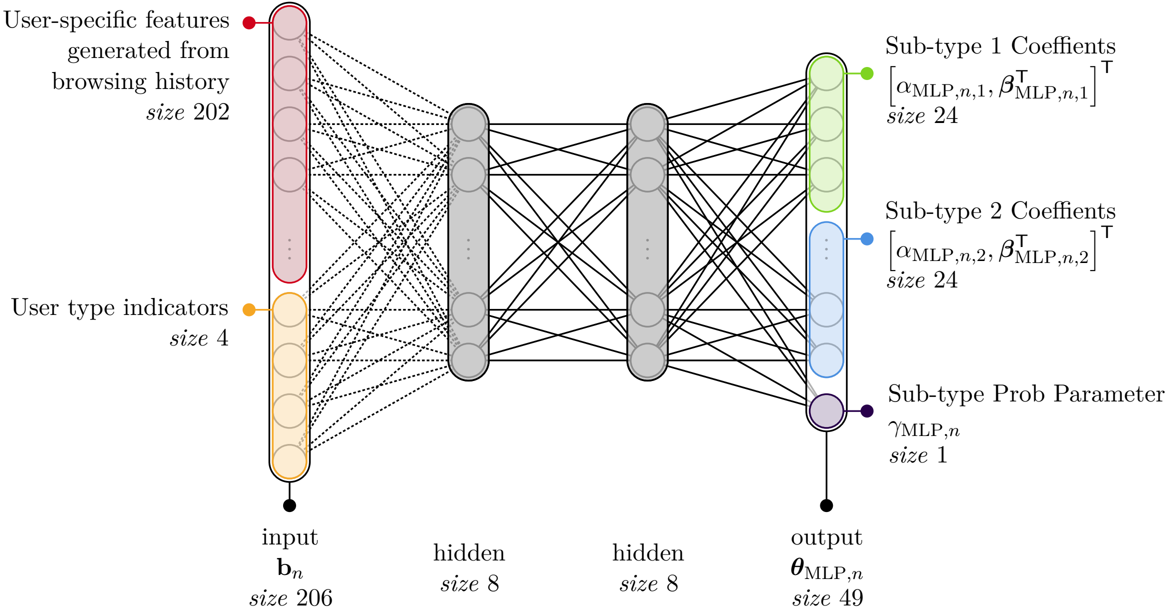
Result of Model IV - MLP

Distribution of Parameter γ



Model IV - MLP

MLP Structure *(2,161 parameters)*



Model V - RNN

For each user n, a matrix of clickstream events, denoted as $\mathbf{B}_n$, is constructed and serves as the input for the RNN. The matrix $\mathbf{B}_n$ contains:

- 50 rows
  - 50 most recent historical clicks upto the first smartphone click (truncation and padding)
- 21 columns
  - event type indicators
  - price
  - event duration and time gap
  - time indicators (days of the week, and months of the year)
  - category (level 1) embedding

$\mathbf{B}_n \in \mathbb{R}^{50} \times \mathbb{R}^{21}$ after truncation and padding, establishing the RNN model as $G_{\text{RNN}} : \mathbb{R}^{50} \times \mathbb{R}^{21} \rightarrow \mathbb{R}^{49}$,

$$G_{\text{RNN}}(\mathbf{B}_n; \boldsymbol{\Theta}_{\text{V}}) = \boldsymbol{\theta}_{\text{RNN},n} \tag{20}$$

where $\boldsymbol{\theta}_{\text{RNN},n} = \left[\alpha_{\text{RNN},n,s}, \boldsymbol{\beta}_{\text{RNN},n,s}{}^{\top}, \gamma_{\text{RNN},n}\right]_{s\in\mathcal{S}}^{\top}$ has a size of 49, and $\pi_{n,1} = \mathit{sigmoid}(\gamma_{\text{RNN},n})$.

Model V - RNN

For each user n, a matrix of clickstream events, denoted as $\mathbf{B}_n$, is constructed and serves as the input for the RNN. The matrix $\mathbf{B}_n$ contains:

- 50 rows
  - 50 most recent historical clicks upto the first smartphone click (truncation and padding)
- 21 columns
  - event type indicators
  - price
  - event duration and time gap
  - time indicators (days of the week, and months of the year)
  - category (level 1) embedding

$\mathbf{B}_n \in \mathbb{R}^{50} \times \mathbb{R}^{21}$ after truncation and padding, establishing the RNN model as $G_{\text{RNN}} : \mathbb{R}^{50} \times \mathbb{R}^{21} \rightarrow \mathbb{R}^{49}$,

$$G_{\text{RNN}}(\mathbf{B}_n; \mathbf{\Theta}_{\text{V}}) = \boldsymbol{\theta}_{\text{RNN},n} \tag{20}$$

where $\boldsymbol{\theta}_{\text{RNN},n} = \left[\alpha_{\text{RNN},n,s}, \boldsymbol{\beta}_{\text{RNN},n,s}{}^{\mathsf{T}}, \gamma_{\text{RNN},n}\right]^{\mathsf{T}}_{s\in\mathcal{S}}$ has a size of 49, and $\pi_{n,1} = sigmoid(\gamma_{\text{RNN},n})$.

Model V - RNN
(cont.)

The utility $U_{n,s,i}$ as from the generic form (1) with the additional subscript $s \in \mathcal{S}$ is hence,

$$U_{n,s,i} \equiv u_{n,s,i} + \epsilon_{n,i} = \alpha_{\mathrm{RNN},n,s} \cdot p_{t(n),i} + {\boldsymbol{\beta}_{\mathrm{RNN},n,s}}^{\mathsf{T}} \mathbf{x}_{n,t(n),i} + \epsilon_{n,i}, \quad \forall i \in \mathcal{C}_n. \tag{21}$$

The conditional choice probability is a weighted average,

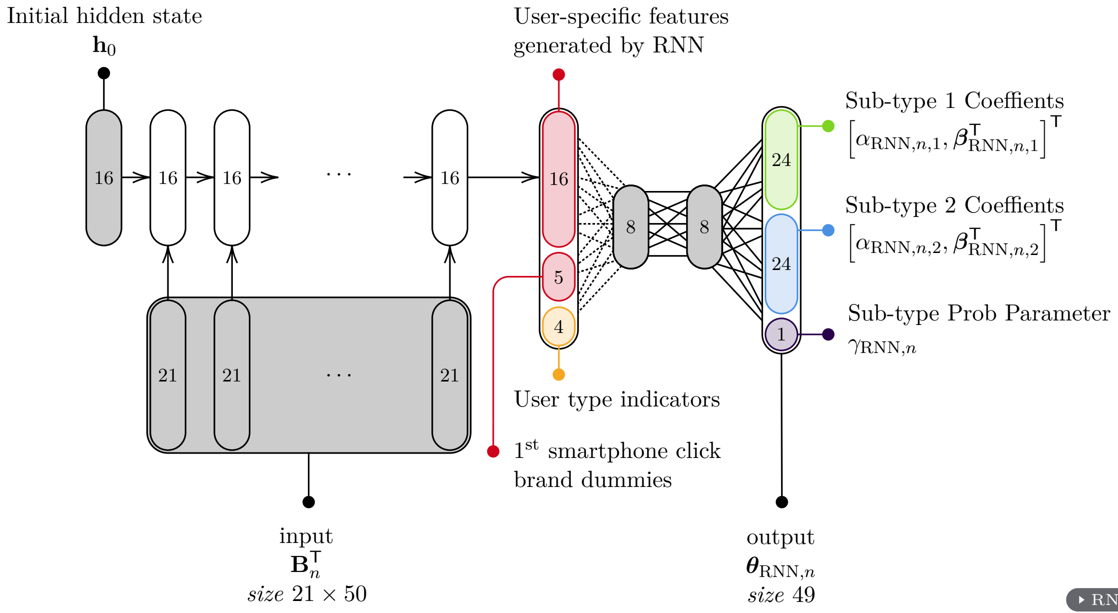
$$P_{n,i} = \sum_{s \in \mathcal{S}} \pi_{n,s} \cdot \frac{\exp(u_{n,s,i})}{1 + \sum_{j \in \mathcal{A}_{t(n)}} \exp(u_{n,s,j})}, \quad \forall i \in \mathcal{C}_n. \tag{22}$$

And the estimation involves maximizing $\ell\left(\boldsymbol{\Theta}_{\mathrm{V}}\right)$ as specified in (3),

$$\ell\left(\boldsymbol{\Theta}_{\mathrm{V}}\right) = \sum_{n \in \mathcal{N}} \sum_{i \in \mathcal{C}_n} y_{n,i} \cdot \log(P_{n,i}). \tag{23}$$

Model V - RNN

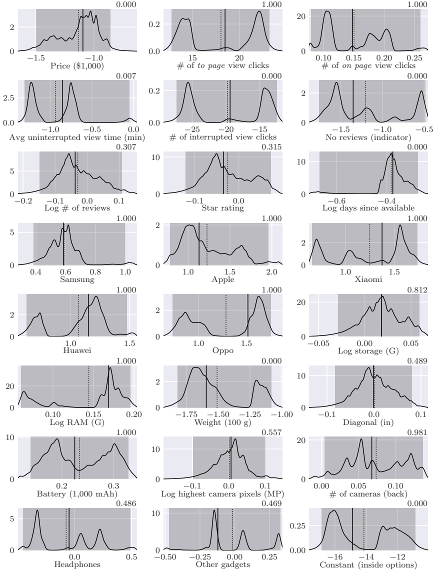
RNN Structure *(1,375 parameters)*



▸ RNN Details

Result of Model V - RNN

(1,359 parameters)



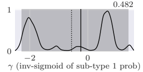
Sub-type 1

Estimation

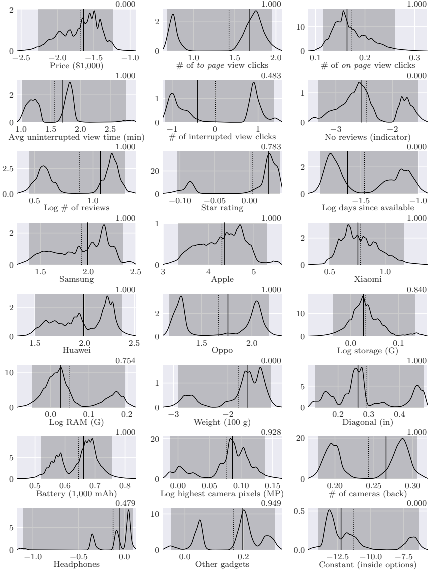
Observations	:	6,395,207,961
Users (Cases)	:	4,235,321
LL	:	−1,200,870.6
Ave LL Per User	:	−0.284
AIC	:	2,404,460.2
BIC	:	2,422,479.2

Holdout

Observations	:	2,131,744,036
Users (Cases)	:	1,411,773
LL	:	−400,974.7
Ave LL Per User	:	−0.284



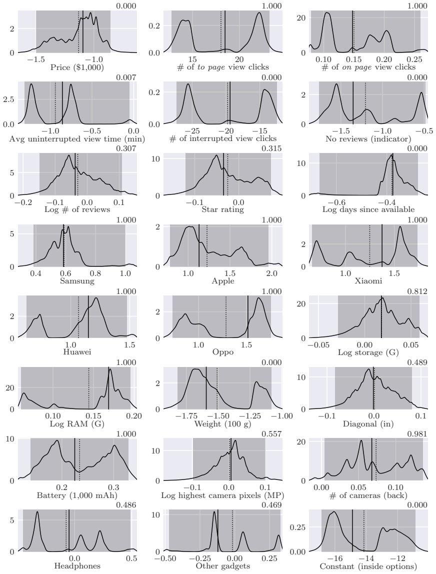
Parameter γ



Sub-type 2

Result of Model V - RNN

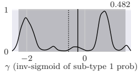
(1,359 parameters)



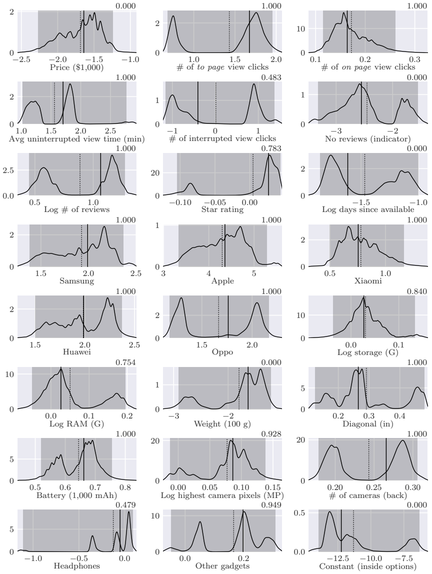
Sub-type 1

Estimation		
Observations	:	6,395,207,961
Users (Cases)	:	4,235,321
LL	:	−1,200,870.6
Ave 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
Ave LL Per User	:	−0.284



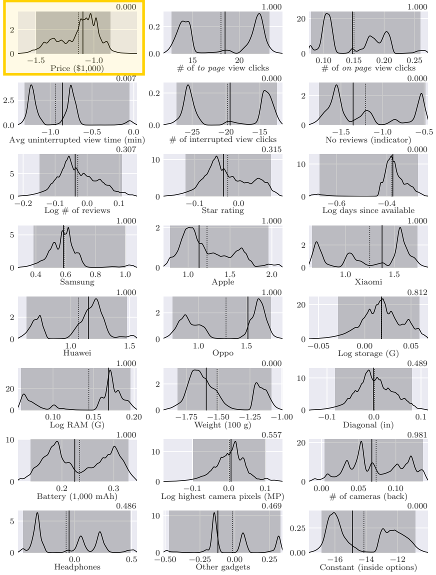
Parameter γ



Sub-type 2

Result of Model V - RNN

(1,359 parameters)



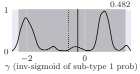
Sub-type 1

Estimation

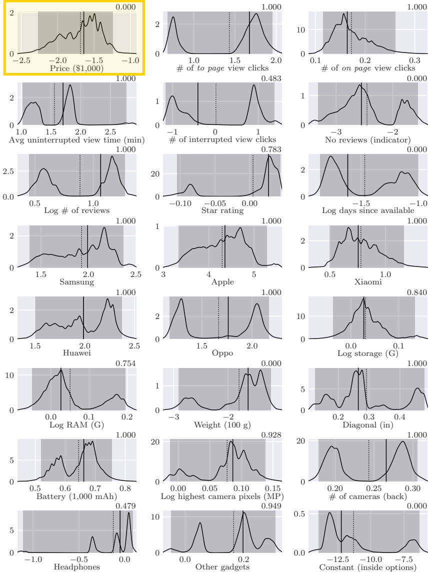
Observations	:	6,395,207,961
Users (Cases)	:	4,235,321
LL	:	−1,200,870.6
Ave LL Per User	:	−0.284
AIC	:	2,404,460.2
BIC	:	2,422,479.2

Holdout

Observations	:	2,131,744,036
Users (Cases)	:	1,411,773
LL	:	−400,974.7
Ave LL Per User	:	−0.284



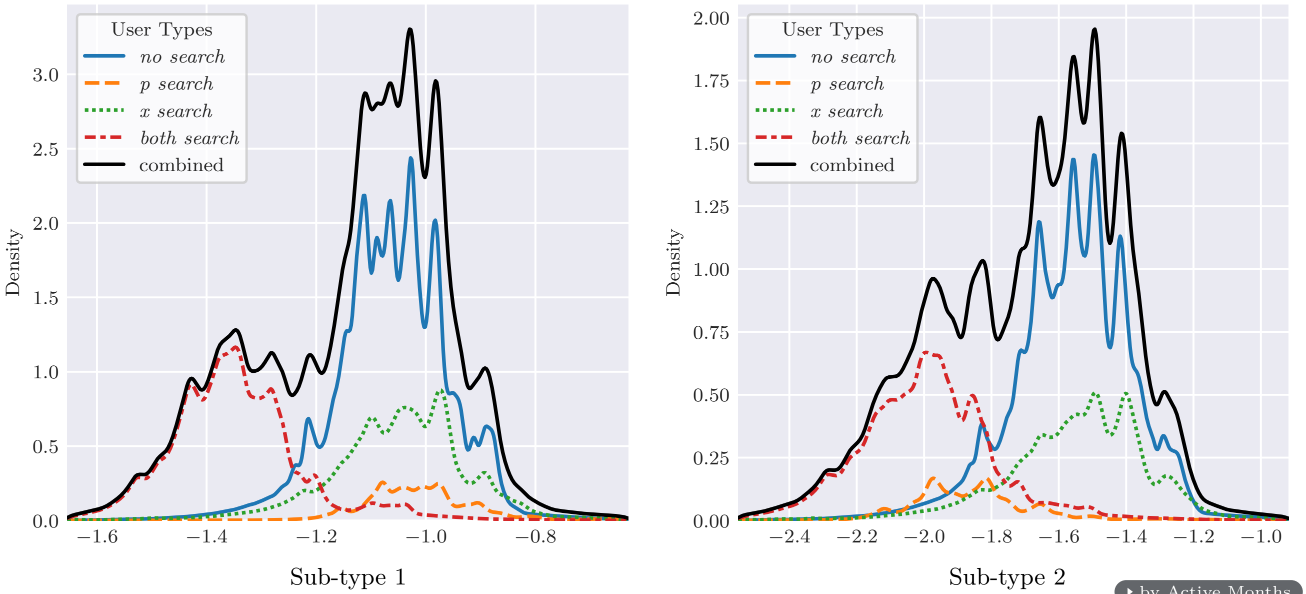
Parameter γ



Sub-type 2

Result of Model V - RNN

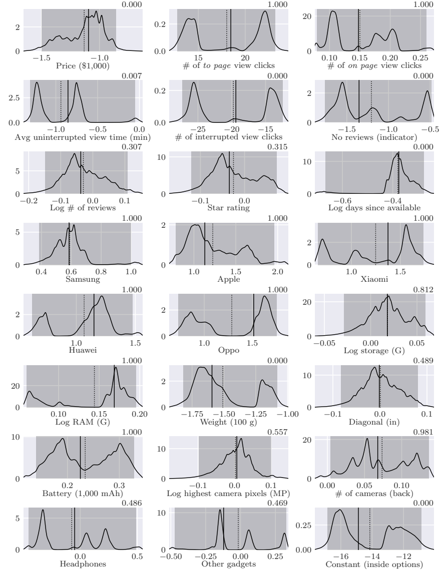
Distribution of Price ($1,000) Coefficients (by User Types)



▸ by Active Months

Result of Model V - RNN

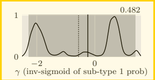
(1,359 parameters)



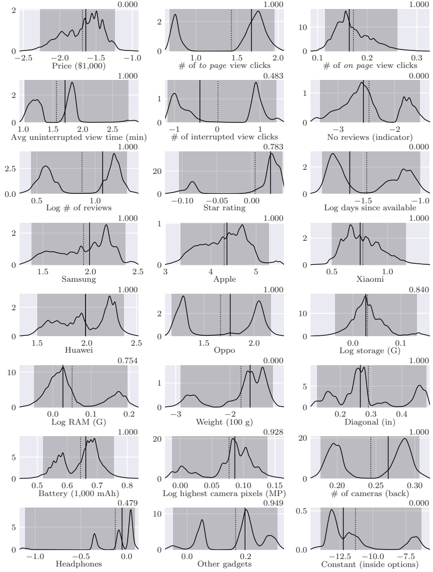
Sub-type 1

Estimation		
Observations	:	6,395,207,961
Users (Cases)	:	4,235,321
LL	:	−1,200,870.6
Ave 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
Ave LL Per User	:	−0.284



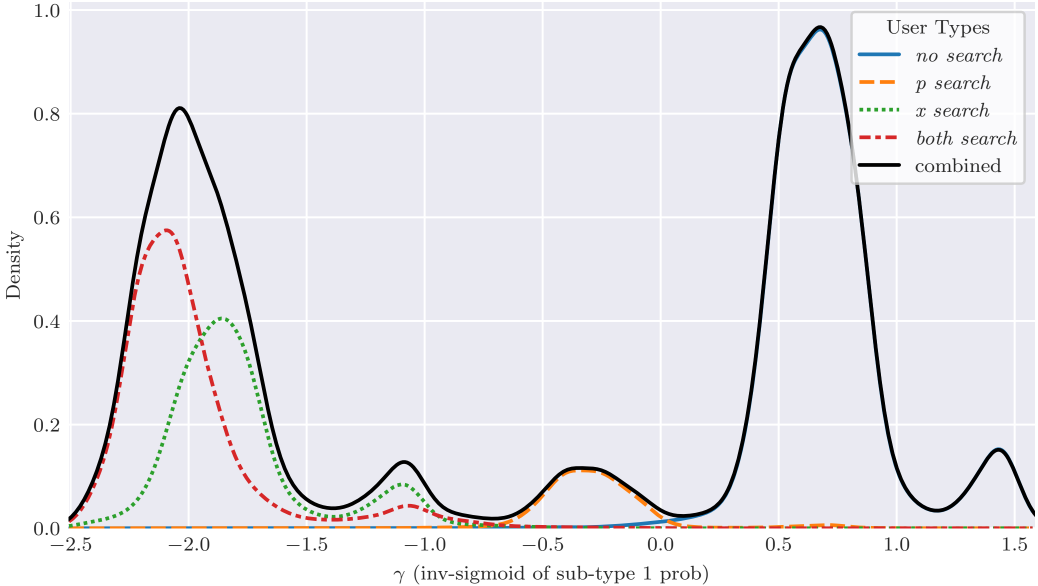
Parameter γ



Sub-type 2

Result of Model IV - RNN

Distribution of Parameter γ



Model V - RNN

RNN Structure *(1,375 parameters)*

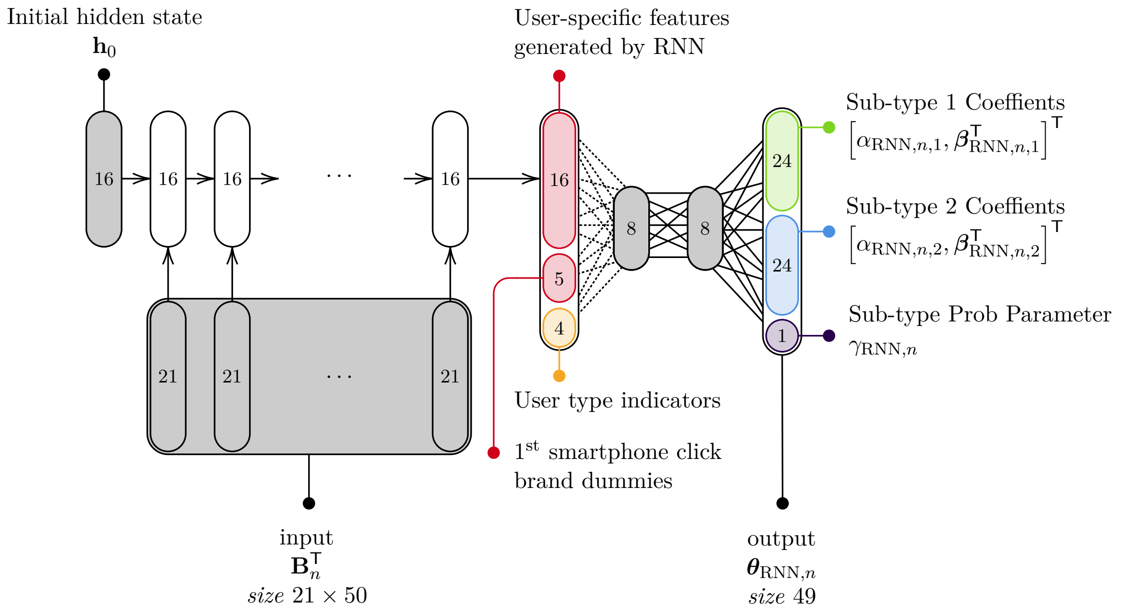
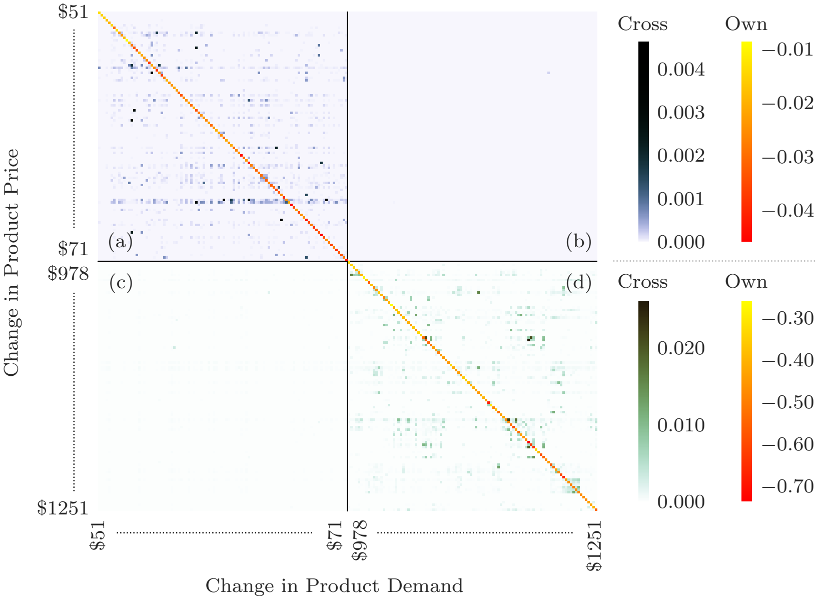


Table of Contents

Price Elasticities

Price Elasticities: Model V - RNN



Illustration

Other Models

Table of Contents

Conclusion
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
# 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.
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.

Conclusion
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
# 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.
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.

Conclusion
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
# 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.
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.

Conclusion
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
# 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.
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.

Conclusion
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
# 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.
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.

Conclusion
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
# 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.
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.

Conclusion

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
# 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.
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.

Conclusion
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
# 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.
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.

Conclusion
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
# 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.
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.

Conclusion
Contributions

- Utilizes extensive e-commerce behavioral data alongside the collected detailed product page information for a thorough analysis.
- Applies non-parametric machine learning techniques for a nuanced and efficient modeling of consumer heterogeneity in a parsimonious yet effective manner.
- Addresses the IIA limitation by integrating varied levels of consumer and product heterogeneity.
- Demonstrates the significant potential of integrating machine learning models, particularly Natural Language Processing (NLP) techniques, in empirical economic analyses of panel data.

Conclusion
Contributions

- Utilizes extensive e-commerce behavioral data alongside the collected detailed product page information for a thorough analysis.
- Applies non-parametric machine learning techniques for a nuanced and efficient modeling of consumer heterogeneity in a parsimonious yet effective manner.
- Addresses the IIA limitation by integrating varied levels of consumer and product heterogeneity.
- Demonstrates the significant potential of integrating machine learning models, particularly Natural Language Processing (NLP) techniques, in empirical economic analyses of panel data.

Conclusion
Contributions

- Utilizes extensive e-commerce behavioral data alongside the collected detailed product page information for a thorough analysis.
- Applies non-parametric machine learning techniques for a nuanced and efficient modeling of consumer heterogeneity in a parsimonious yet effective manner.
- Addresses the IIA limitation by integrating varied levels of consumer and product heterogeneity.
- Demonstrates the significant potential of integrating machine learning models, particularly Natural Language Processing (NLP) techniques, in empirical economic analyses of panel data.

Conclusion
Contributions

- Utilizes extensive e-commerce behavioral data alongside the collected detailed product page information for a thorough analysis.
- Applies non-parametric machine learning techniques for a nuanced and efficient modeling of consumer heterogeneity in a parsimonious yet effective manner.
- Addresses the IIA limitation by integrating varied levels of consumer and product heterogeneity.
- Demonstrates the significant potential of integrating machine learning models, particularly Natural Language Processing (NLP) techniques, in empirical economic analyses of panel data.

Conclusion

Takeaways

- Develops a framework that combines historical browsing data with advanced machine learning methods to pinpoint consumer heterogeneity at the individual level for flexible demand estimation.
- Provides empirical evidence that underscores the pronounced connections between consumers' search behaviors (especially concerning price and product attributes), their historical browsing behavior, and underlying preferences.
- Demonstrates that the incorporation of observed and unobserved heterogeneity, along with advanced machine learning methods, can enrich the vanilla logit model and confront the limitations posed by the IIA assumption, thereby facilitating the identification of realistic product substitution patterns.
- Emphasizes the effectiveness of RNN models in analyzing panel data, underscoring the broader applicability and potential of RNN models in economic research.
- Lays the groundwork for counterfactual analysis, enabling an examination of the impact of personalized discounts on both firm profits and consumer welfare in the next chapter.

Conclusion
Takeaways

- Develops a framework that combines historical browsing data with advanced machine learning methods to pinpoint consumer heterogeneity at the individual level for flexible demand estimation.
- Provides empirical evidence that underscores the pronounced connections between consumers' search behaviors (especially concerning price and product attributes), their historical browsing behavior, and underlying preferences.
- Demonstrates that the incorporation of observed and unobserved heterogeneity, along with advanced machine learning methods, can enrich the vanilla logit model and confront the limitations posed by the IIA assumption, thereby facilitating the identification of realistic product substitution patterns.
- Emphasizes the effectiveness of RNN models in analyzing panel data, underscoring the broader applicability and potential of RNN models in economic research.
- Lays the groundwork for counterfactual analysis, enabling an examination of the impact of personalized discounts on both firm profits and consumer welfare in the next chapter.

Conclusion
Takeaways

- Develops a framework that combines historical browsing data with advanced machine learning methods to pinpoint consumer heterogeneity at the individual level for flexible demand estimation.
- Provides empirical evidence that underscores the pronounced connections between consumers' search behaviors (especially concerning price and product attributes), their historical browsing behavior, and underlying preferences.
- Demonstrates that the incorporation of observed and unobserved heterogeneity, along with advanced machine learning methods, can enrich the vanilla logit model and confront the limitations posed by the IIA assumption, thereby facilitating the identification of realistic product substitution patterns.
- Emphasizes the effectiveness of RNN models in analyzing panel data, underscoring the broader applicability and potential of RNN models in economic research.
- Lays the groundwork for counterfactual analysis, enabling an examination of the impact of personalized discounts on both firm profits and consumer welfare in the next chapter.

Conclusion
Takeaways

- Develops a framework that combines historical browsing data with advanced machine learning methods to pinpoint consumer heterogeneity at the individual level for flexible demand estimation.
- Provides empirical evidence that underscores the pronounced connections between consumers' search behaviors (especially concerning price and product attributes), their historical browsing behavior, and underlying preferences.
- Demonstrates that the incorporation of observed and unobserved heterogeneity, along with advanced machine learning methods, can enrich the vanilla logit model and confront the limitations posed by the IIA assumption, thereby facilitating the identification of realistic product substitution patterns.
- Emphasizes the effectiveness of RNN models in analyzing panel data, underscoring the broader applicability and potential of RNN models in economic research.
- Lays the groundwork for counterfactual analysis, enabling an examination of the impact of personalized discounts on both firm profits and consumer welfare in the next chapter.

Conclusion
Takeaways

- Develops a framework that combines historical browsing data with advanced machine learning methods to pinpoint consumer heterogeneity at the individual level for flexible demand estimation.
- Provides empirical evidence that underscores the pronounced connections between consumers' search behaviors (especially concerning price and product attributes), their historical browsing behavior, and underlying preferences.
- Demonstrates that the incorporation of observed and unobserved heterogeneity, along with advanced machine learning methods, can enrich the vanilla logit model and confront the limitations posed by the IIA assumption, thereby facilitating the identification of realistic product substitution patterns.
- Emphasizes the effectiveness of RNN models in analyzing panel data, underscoring the broader applicability and potential of RNN models in economic research.
- Lays the groundwork for counterfactual analysis, enabling an examination of the impact of personalized discounts on both firm profits and consumer welfare in the next chapter.

Chapter 2

Personalized Discounts via Coupons: Generating Gains Conditional on Observed Choices

Table of Contents

Table of Contents

Related Literature

This chapter is related to the general literature about price discrimination and welfare analysis.

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.

- Firms: profit will increase

  Profitability potential : Rossi et al. (1996)
  Empirical studies : Shiller and Waldfogel (2011), Shiller (2014), Dubé and Misra (2023)

- Consumers: welfare impact may be unclear

  Concerns : Pigou (1920), Executive Office of the President (2014), Acquisti et al. (2016)
  Empirical studies : Waldfogel (2015), Shiller and Waldfogel (2011), Dubé and Misra (2023), Bergemann et al. (2015), Kehoe et al. (2022), Shiller (2014)

Related Literature

This chapter is related to the general literature about price discrimination and welfare analysis.

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.

- Firms: profit will increase

  Profitability potential : Rossi et al. (1996)
  Empirical studies : Shiller and Waldfogel (2011), Shiller (2014), Dubé and Misra (2023)

- Consumers: welfare impact may be unclear

  Concerns : Pigou (1920), Executive Office of the President (2014), Acquisti et al. (2016)
  Empirical studies : Waldfogel (2015), Shiller and Waldfogel (2011), Dubé and Misra (2023), Bergemann et al. (2015), Kehoe et al. (2022), Shiller (2014)

Related Literature

This chapter is related to the general literature about price discrimination and welfare analysis.

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.

- Firms: profit will increase

  Profitability potential : Rossi et al. (1996)
  Empirical studies : Shiller and Waldfogel (2011), Shiller (2014), Dubé and Misra (2023)

- Consumers: welfare impact may be unclear

  Concerns : Pigou (1920), Executive Office of the President (2014), Acquisti et al. (2016)
  Empirical studies : Waldfogel (2015), Shiller and Waldfogel (2011), Dubé and Misra (2023), Bergemann et al. (2015), Kehoe et al. (2022), Shiller (2014)

Table of Contents

Data
Wholesale Price Data

This study gathers wholesale price data for smartphones from a select group of wholesalers.

- wholesale price data on 7 different dates ranging across the timeframe of the clickstream data
- minimum is taken when multiple wholesale price data are available on the same date
- the wholesale price at any given date t is linearly interpolated
- denote the wholesale price of smartphone j at time t as $c_{t,j}$
- wholesale prices are available for
  - 60%+ of the unique smartphone models
  - 92.5% of all recorded smartphone clicks
  - 99.5% of all the smartphone purchases
- approximately 0.04% of the cases, the wholesale price surpassed the retail price, and the wholesale price is adjusted to match the retail level

Denote the set of smartphones with available wholesale price data at time t as $\mathcal{D}_t$, and restrict the smartphones eligible for personalized discount to that set.

Table of Contents

Model
Framework

As users conclude their online exploration, the platform acquires valuable insights into their choice decisions, i.e., for n with an unobserved realization of $s \in \mathcal{S}$ such that $y_{n,q} = 1$ where $q \in \mathcal{C}_n$,

$$U_{n,s,q} > U_{n,s,j}, \forall j \in \mathcal{C}_n \setminus \{q\}. \tag{24}$$

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. Hence, conditional on (24) with $q = 0$, revealed preference shows the user will either purchase k or continue to choose 0 with the discount.

Upcoming:

- Profit (ex-ante and ex-post)
- Platform's maximization problem
- Consumer surplus (ex-ante and ex-post)

Model
Framework

As users conclude their online exploration, the platform acquires valuable insights into their choice decisions, i.e., for n with an unobserved realization of $s \in \mathcal{S}$ such that $y_{n,q} = 1$ where $q \in \mathcal{C}_n$,

$$U_{n,s,q} > U_{n,s,j}, \forall j \in \mathcal{C}_n \setminus \{q\}. \tag{24}$$

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. Hence, conditional on (24) with $q = 0$, revealed preference shows the user will either purchase k or continue to choose 0 with the discount.

Upcoming:

- Profit (ex-ante and ex-post)
- Platform's maximization problem
- Consumer surplus (ex-ante and ex-post)

Model
Framework

As users conclude their online exploration, the platform acquires valuable insights into their choice decisions, i.e., for n with an unobserved realization of $s \in \mathcal{S}$ such that $y_{n,q} = 1$ where $q \in \mathcal{C}_n$,

$$U_{n,s,q} > U_{n,s,j}, \forall j \in \mathcal{C}_n \setminus \{q\}. \tag{24}$$

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. Hence, conditional on (24) with $q = 0$, revealed preference shows the user will either purchase k or continue to choose 0 with the discount.

Upcoming:

- Profit (ex-ante and ex-post)
- Platform's maximization problem
- Consumer surplus (ex-ante and ex-post)

Model
Conditional Expected Profit (Ex-ante)

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.,

$$\Pi_n = \sum_{j \in \mathcal{A}_{t(n)}} y_{n,j} \cdot \left(p_{t(n),j} - c_{t(n),j}\right). \tag{25}$$

Model

Conditional Expected Profit (Ex-post)

The conditional expected profit gain attributable to the discount is,

$$\overset{\triangle}{\Pi}_n(d_{n,k}) \equiv \tilde{\mathsf{P}}_{n,k}(d_{n,k}) \cdot (p_{t(n),k} - c_{t(n),k} - d_{n,k}) \tag{26}$$

where

$$\tilde{\mathsf{P}}_{n,k}(d_{n,k}) = \sum_{s \in \mathcal{S}} \boldsymbol{\pi}_{n,s} \cdot \tilde{\mathsf{P}}_{n,s,k}(d_{n,k}), \quad \boldsymbol{\pi}_{n,s} = \frac{\pi_{n,s} \cdot \text{Prob}\Big(U_{n,s,0} > U_{n,s,j},\, \forall j \in \mathcal{C}_n \setminus \{0\}\Big)}{\sum_{v \in \mathcal{S}} \pi_{n,v} \cdot \text{Prob}\Big(U_{n,v,0} > U_{n,v,j},\, \forall j \in \mathcal{C}_n \setminus \{0\}\Big)}, \tag{27}$$

$$\tilde{\mathsf{P}}_{n,s,k}(d_{n,k}) \equiv \text{Prob}\Big(\tilde{U}_{n,s,k}(d_{n,k}) > U_{n,s,i},\, \forall i \in \mathcal{C}_n \setminus \{k\} \,\Big|\, U_{n,s,0} > U_{n,s,j},\, \forall j \in \mathcal{C}_n \setminus \{0\}\Big), \tag{28}$$

and $$\tilde{U}_{n,s,k}(d_{n,k}) = U_{n,s,k} - \alpha_{n,s} \cdot d_{n,k}, \qquad \tilde{u}_{n,s,k}(d_{n,k}) = u_{n,s,k} - \alpha_{n,s} \cdot d_{n,k}. \tag{29}$$

It can be shown that $$\tilde{\mathsf{P}}_{n,s,k}(d_{n,k}) = \frac{\Delta_{n,s,k}(d_{n,k})}{\Delta_{n,s,k}(d_{n,k}) + \sum_{j \in \mathcal{C}_n} \exp(u_{n,s,j})}, \tag{30}$$

where $$\Delta_{n,s,k}(d_{n,k}) = \exp(\tilde{u}_{n,s,k}(d_{n,k})) - \exp(u_{n,s,k}). \tag{31}$$

Proof

Model
Conditional Expected Profit (Ex-post)

The conditional expected profit gain attributable to the discount is,

$$\overset{\triangle}{\Pi}_{n}(d_{n,k}) \equiv \tilde{\mathsf{P}}_{n,k}(d_{n,k}) \cdot (p_{t(n),k} - c_{t(n),k} - d_{n,k}) \tag{26}$$

where

$$\tilde{\mathsf{P}}_{n,k}(d_{n,k}) = \sum_{s\in\mathcal{S}} \boldsymbol{\pi}_{n,s} \cdot \tilde{\mathsf{P}}_{n,s,k}(d_{n,k}), \quad \boldsymbol{\pi}_{n,s} = \frac{\pi_{n,s} \cdot \mathrm{Prob}\Big(U_{n,s,0} > U_{n,s,j},\ \forall j \in \mathcal{C}_n \setminus \{0\}\Big)}{\sum_{v\in\mathcal{S}} \pi_{n,v} \cdot \mathrm{Prob}\Big(U_{n,v,0} > U_{n,v,j},\ \forall j \in \mathcal{C}_n \setminus \{0\}\Big)}, \tag{27}$$

$$\tilde{\mathsf{P}}_{n,s,k}(d_{n,k}) \equiv \mathrm{Prob}\Big(\tilde{U}_{n,s,k}(d_{n,k}) > U_{n,s,i},\ \forall i \in \mathcal{C}_n \setminus \{k\} \,\Big|\, U_{n,s,0} > U_{n,s,j},\ \forall j \in \mathcal{C}_n \setminus \{0\}\Big), \tag{28}$$

and $$\tilde{U}_{n,s,k}(d_{n,k}) = U_{n,s,k} - \alpha_{n,s} \cdot d_{n,k}, \qquad \tilde{u}_{n,s,k}(d_{n,k}) = u_{n,s,k} - \alpha_{n,s} \cdot d_{n,k}. \tag{29}$$

It can be shown that $$\tilde{\mathsf{P}}_{n,s,k}(d_{n,k}) = \frac{\Delta_{n,s,k}(d_{n,k})}{\Delta_{n,s,k}(d_{n,k}) + \sum_{j\in\mathcal{C}_n} \exp(u_{n,s,j})}, \tag{30}$$

where $$\Delta_{n,s,k}(d_{n,k}) = \exp(\tilde{u}_{n,s,k}(d_{n,k})) - \exp(u_{n,s,k}). \tag{31}$$

▸ Proof

Model

Platform's Conditional Profit Maximization Problem

For each user n whose $y_{n,0} = 1$, the e-commerce platform face the following conditional expected profit maximization problem over k and $d_{n,k}$,

$$\max_{k \in \mathcal{K}_n} \max_{\substack{d_{n,k} \in \\ \mathcal{W}\left(0, p_{t(n),k} - c_{t(n),k}\right)}} \left[\sum_{s \in \mathcal{S}} \pi_{n,s} \cdot \frac{\Delta_{n,s,k}(d_{n,k})}{\Delta_{n,s,k}(d_{n,k}) + \sum_{j \in \mathcal{C}_n} \exp(u_{n,s,j})} \right] \cdot (p_{t(n),k} - c_{t(n),k} - d_{n,k}) \tag{32}$$

where $\mathcal{W}\left(0, p_{t(n),k} - c_{t(n),k}\right)$ 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.

Three different schemes of $\mathcal{K}_n$ are considered

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,
3. all smartphone eligible for discount, i.e., $\mathcal{D}_{t(n)}$.

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.

Model

Platform's Conditional Profit Maximization Problem

For each user n whose $y_{n,0} = 1$, the e-commerce platform face the following conditional expected profit maximization problem over k and $d_{n,k}$,

$$\max_{k \in \mathcal{K}_n} \max_{\substack{d_{n,k} \in \\ \mathcal{W}\left(0, p_{t(n),k} - c_{t(n),k}\right)}} \left[\sum_{s \in \mathcal{S}} \pi_{n,s} \cdot \frac{\Delta_{n,s,k}(d_{n,k})}{\Delta_{n,s,k}(d_{n,k}) + \sum_{j \in \mathcal{C}_n} \exp(u_{n,s,j})} \right] \cdot (p_{t(n),k} - c_{t(n),k} - d_{n,k}) \tag{32}$$

where $\mathcal{W}\left(0, p_{t(n),k} - c_{t(n),k}\right)$ 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.

Three different schemes of $\mathcal{K}_n$ are considered

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,
3. all smartphone eligible for discount, i.e., $\mathcal{D}_{t(n)}$.

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.

Model

Conditional Expected Consumer Surplus (Ex-ante)

For user n with $y_{n,q} = 1$ where $q \in \mathcal{C}_n$, the consumer surplus (relative to the outside option) conditional on observing user n's choice is

$$\mathsf{CS}_n = \sum_{s \in \mathcal{S}} \pi_{n,s} \cdot \mathsf{CS}_{n,s} \tag{33}$$

where

$$\mathsf{CS}_{n,s} \equiv \frac{1}{-\alpha_{n,s}} \mathbb{E}\Big[U_{n,s,q} - U_{n,s,0} \Big| U_{n,s,q} > U_{n,s,j}, \forall j \in \mathcal{C}_n \setminus \{q\}\Big] \tag{34}$$

It can be shown that

$$\mathsf{CS}_{n,s} = \begin{cases} 0 & \text{if } y_{n,0} = 1, \\ \dfrac{1}{-\alpha_{n,s}} & \text{if } y_{n,0} = 0, u_{n,s,q} = 0, \\ \dfrac{1}{-\alpha_{n,s}} \cdot \dfrac{u_{n,s,q}}{1 - e^{-u_{n,s,q}}} & \text{if } y_{n,0} = 0, u_{n,s,q} \neq 0. \end{cases} \tag{35}$$

Proof see B.4.1

Model

Conditional Expected Consumer Surplus (Ex-ante)

For user n with $y_{n,q} = 1$ where $q \in \mathcal{C}_n$, the consumer surplus (relative to the outside option) conditional on observing user n's choice is

$$\mathsf{CS}_n = \sum_{s \in \mathcal{S}} \pi_{n,s} \cdot \mathsf{CS}_{n,s} \tag{33}$$

where

$$\mathsf{CS}_{n,s} \equiv \frac{1}{-\alpha_{n,s}} \mathbb{E}\Big[U_{n,s,q} - U_{n,s,0} \Big| U_{n,s,q} > U_{n,s,j}, \forall j \in \mathcal{C}_n \setminus \{q\}\Big] \tag{34}$$

It can be shown that

$$\mathsf{CS}_{n,s} = \begin{cases} 0 & \text{if } y_{n,0} = 1, \\ \dfrac{1}{-\alpha_{n,s}} & \text{if } y_{n,0} = 0, u_{n,s,q} = 0, \\ \dfrac{1}{-\alpha_{n,s}} \cdot \dfrac{u_{n,s,q}}{1 - e^{-u_{n,s,q}}} & \text{if } y_{n,0} = 0, u_{n,s,q} \neq 0. \end{cases} \tag{35}$$

Proof see B.4.1

Model

Conditional Expected Consumer Surplus (Ex-post)

The conditional consumer surplus gain attributable to the discount is,

$$\overset{\triangle}{\mathsf{CS}}_n = \sum_{s \in S} \pi_{n,s} \cdot \overset{\triangle}{\mathsf{CS}}_{n,s} \tag{36}$$

where

$$\overset{\triangle}{\mathsf{CS}}_{n,s} = \tilde{\mathsf{P}}_{n,s,k}(d_{n,k}) \cdot \frac{1}{-\alpha_{n,s}} E_{n,s,k}(d_{n,k}), \tag{37}$$

$$E_{n,s,k}(d_{n,k}) \equiv \mathbb{E}\left[\tilde{U}_{n,s,k}(d_{n,k}) - U_{n,s,0} \middle| \tilde{U}_{n,s,k}(d_{n,k}) > U_{n,s,0} \wedge U_{n,s,0} > U_{n,s,j}, \forall j \in \mathcal{C}_n \setminus \{0\}\right]. \tag{38}$$

Let

$$E_{n,s,k}(d_{n,k}, \epsilon_{n,0}) \equiv \mathbb{E}\left[\tilde{U}_{n,s,k}(d_{n,k}) - U_{n,s,0} \middle| \tilde{U}_{n,s,k}(d_{n,k}) > U_{n,s,0} \wedge U_{n,s,0} > U_{n,s,j}, \forall j \in \mathcal{C}_n \setminus \{0\}, \epsilon_{n,0}\right], \tag{39}$$

and hence,

$$E_{n,s,k}(d_{n,k}) = \int_{-\infty}^{\infty} E_{n,s,k}(d_{n,k}, \epsilon_{n,0}) \cdot e^{-s} \cdot e^{-e^{-s}}\, ds. \tag{40}$$

Model
Conditional Expected Consumer Surplus (Ex-post)

The conditional consumer surplus gain attributable to the discount is,

$$\overset{\triangle}{\mathsf{CS}}_n = \sum_{s \in S} \pi_{n,s} \cdot \overset{\triangle}{\mathsf{CS}}_{n,s} \tag{36}$$

where

$$\overset{\triangle}{\mathsf{CS}}_{n,s} = \tilde{\mathsf{P}}_{n,s,k}(d_{n,k}) \cdot \frac{1}{-\alpha_{n,s}} E_{n,s,k}(d_{n,k}), \tag{37}$$

$$E_{n,s,k}(d_{n,k}) \equiv \mathbb{E}\left[\tilde{U}_{n,s,k}(d_{n,k}) - U_{n,s,0} \middle| \tilde{U}_{n,s,k}(d_{n,k}) > U_{n,s,0} \wedge U_{n,s,0} > U_{n,s,j}, \forall j \in \mathcal{C}_n \setminus \{0\}\right]. \tag{38}$$

Let

$$E_{n,s,k}(d_{n,k}, \epsilon_{n,0}) \equiv \mathbb{E}\left[\tilde{U}_{n,s,k}(d_{n,k}) - U_{n,s,0} \middle| \tilde{U}_{n,s,k}(d_{n,k}) > U_{n,s,0} \wedge U_{n,s,0} > U_{n,s,j}, \forall j \in \mathcal{C}_n \setminus \{0\}, \epsilon_{n,0}\right], \tag{39}$$

and hence,

$$E_{n,s,k}(d_{n,k}) = \int_{-\infty}^{\infty} E_{n,s,k}(d_{n,k}, \epsilon_{n,0}) \cdot e^{-s} \cdot e^{-e^{-s}} \, ds. \tag{40}$$

Model

Conditional Expected Consumer Surplus (Ex-post, cont.)

$E_{n,s,k}(d_{n,k})$ does not have a closed form, but $E_{n,s,k}(d_{n,k}, \epsilon_{n,0})$ has a closed form, although not ordinary.

$$E_{n,s,k}(d_{n,k}, \epsilon_{n,0}) = \frac{1}{\exp\left[-e^{u_{n,s,k}-\epsilon_{n,0}}\right] - \exp\left[-e^{\tilde{u}_{n,s,k}(d_{n,k})-\epsilon_{n,0}}\right]} \times \left[-\alpha_{n,s,k} \cdot d_{n,k} \cdot \exp\left(-e^{u_{n,s,k}-\epsilon_{n,0}}\right) + \mathrm{Ei}\left(-e^{u_{n,s,k}-\epsilon_{n,0}}\right) - \mathrm{Ei}\left(-e^{\tilde{u}_{n,s,k}(d_{n,k})-\epsilon_{n,0}}\right)\right] \quad (41)$$

where $\mathrm{Ei}(\cdot)$ is the exponential integral function,

$$\mathrm{Ei}(x) = \int_{-\infty}^{x} \frac{e^t}{t}\, dt$$

which can be easily evaluated.

Then, $E_{n,s,k}(d_{n,k})$ can be calculated numerically, e.g., via Gaussian quadrature or Monte Carlo.

Proof see B.4.2

Table of Contents

Numerical Stability

For Conditional Expected Consumer Surplus (Ex-post)

Direct evaluation of the following expression can be numerically unstable, due to cases when

1. $\exp\left[-e^{u_{n,s,k}-\epsilon_{n,0}}\right] \approx \exp\left[-e^{\tilde{u}_{n,s,k}(d_{n,k})-\epsilon_{n,0}}\right]$ which leads to division by zero,
2. $\mathrm{Ei}\left(-e^{u_{n,s,k}-\epsilon_{n,0}}\right)$ and $\mathrm{Ei}\left(-e^{\tilde{u}_{n,s,k}(d_{n,k})-\epsilon_{n,0}}\right)$ diverge to $-\infty$.

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

Note that $E_{n,s,k}(d_{n,k},\epsilon_{n,0})$ can be regarded as a function of

$$u_{n,s,k} - \epsilon_{n,0} \quad \text{and} \quad -\alpha_{n,s,k}\cdot d_{n,k},$$

since

$$\tilde{u}_{n,s,k}(d_{n,k}) - \epsilon_{n,0} = u_{n,s,k} - \epsilon_{n,0} - \alpha_{n,s,k}\cdot d_{n,k}$$

Numerical Stability

For Conditional Expected Consumer Surplus (Ex-post)

Direct evaluation of the following expression can be numerically unstable, due to cases when

1. $\exp\left[-e^{u_{n,s,k}-\epsilon_{n,0}}\right] \approx \exp\left[-e^{\tilde{u}_{n,s,k}(d_{n,k})-\epsilon_{n,0}}\right]$ which leads to division by zero,
2. $\mathrm{Ei}\left(-e^{u_{n,s,k}-\epsilon_{n,0}}\right)$ and $\mathrm{Ei}\left(-e^{\tilde{u}_{n,s,k}(d_{n,k})-\epsilon_{n,0}}\right)$ diverge to $-\infty$.

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

Note that $E_{n,s,k}(d_{n,k},\epsilon_{n,0})$ can be regarded as a function of

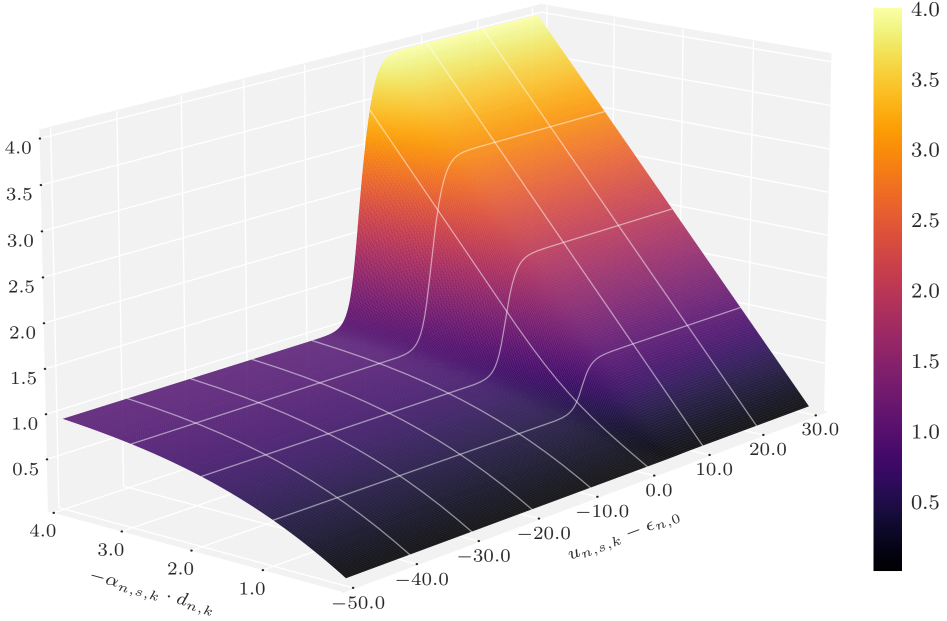
$$u_{n,s,k} - \epsilon_{n,0} \quad \text{and} \quad -\alpha_{n,s,k}\cdot d_{n,k},$$

since

$$\tilde{u}_{n,s,k}(d_{n,k}) - \epsilon_{n,0} = u_{n,s,k} - \epsilon_{n,0} - \alpha_{n,s,k}\cdot d_{n,k}$$

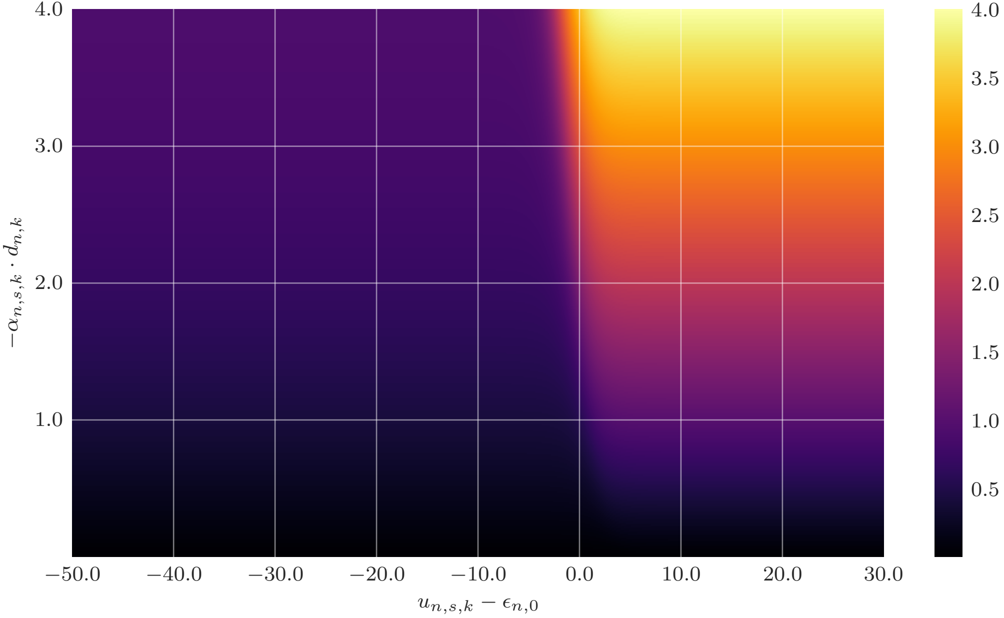
Numerical Stability

Conditional Expected Consumer Surplus (Surface)



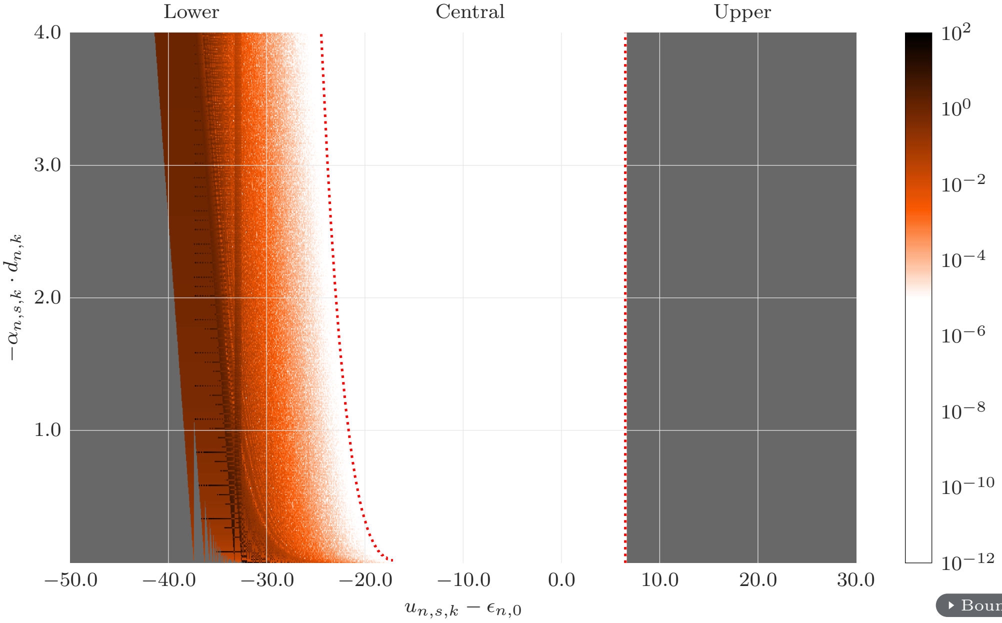
Numerical Stability

Conditional Expected Consumer Surplus (Heatmap)



Numerical Stability

Absolute Error of Numerically Unstable Algorithm



▸ Boundary Conditions

Numerical Stability

Stable Algorithm

- **Lower region** : approximate the value using the limit as $u_{n,s,k} - \epsilon_{n,0} \to -\infty$,

$$E_{n,s,k}(d_{n,k}, \epsilon_{n,0}) \approx 1 - \alpha_{n,s,k} \cdot d_{n,k} \cdot \frac{e^{\alpha_{n,s,k} \cdot d_{n,k}}}{e^{\alpha_{n,s,k} \cdot d_{n,k}} - 1}. \tag{42}$$

- **Central region** : already stable.
- **Upper region** : rewrite to

$$E_{n,s,k}(d_{n,k}, \epsilon_{n,0}) = \frac{-\alpha_{n,s,k} \cdot d_{n,k}}{1 - \exp\left[e^{u_{n,s,k} - \epsilon_{n,0}}\left(1 - e^{-\alpha_{n,s,k} \cdot d_{n,k}}\right)\right]} + \underbrace{\frac{\mathrm{Ei}\left(\log(y_1)\right) - \mathrm{Ei}\left(\log(y_2)\right)}{y_1 - y_2}}_{\approx \frac{d}{dy} \mathrm{Ei}(\log(y))\Big|_{y=y_1} = \frac{1}{\log(y_1)}} \tag{43}$$

where $$y_1 = \exp\left(-e^{u_{n,s,k} - \epsilon_{n,0}}\right), \text{ and } y_2 = \exp\left(-e^{\tilde{u}_{n,s,k} - \epsilon_{n,0}}\right). \tag{44}$$

Hence, $$E_{n,s,k}(d_{n,k}, \epsilon_{n,0}) \approx \frac{-\alpha_{n,s,k} \cdot d_{n,k}}{1 - \exp\left[e^{u_{n,s,k} - \epsilon_{n,0}}\left(1 - e^{-\alpha_{n,s,k} \cdot d_{n,k}}\right)\right]} + \frac{1}{-e^{u_{n,s,k} - \epsilon_{n,0}}}. \tag{45}$$

Proof see C.3

Numerical Stability

Stable Algorithm

- **Lower region :** approximate the value using the limit as $u_{n,s,k} - \epsilon_{n,0} \to -\infty$,

$$E_{n,s,k}(d_{n,k}, \epsilon_{n,0}) \approx 1 - \alpha_{n,s,k} \cdot d_{n,k} \cdot \frac{e^{\alpha_{n,s,k} \cdot d_{n,k}}}{e^{\alpha_{n,s,k} \cdot d_{n,k}} - 1}. \tag{42}$$

- **Central region :** already stable.
- **Upper region :** rewrite to

$$E_{n,s,k}(d_{n,k}, \epsilon_{n,0}) = \frac{-\alpha_{n,s,k} \cdot d_{n,k}}{1 - \exp\left[e^{u_{n,s,k} - \epsilon_{n,0}}\left(1 - e^{-\alpha_{n,s,k} \cdot d_{n,k}}\right)\right]} + \underbrace{\frac{\mathrm{Ei}\left(\log(y_1)\right) - \mathrm{Ei}\left(\log(y_2)\right)}{y_1 - y_2}}_{\approx \frac{d}{dy} \mathrm{Ei}(\log(y))\Big|_{y=y_1} = \frac{1}{\log(y_1)}} \tag{43}$$

where
$$y_1 = \exp\left(-e^{u_{n,s,k} - \epsilon_{n,0}}\right), \text{ and } y_2 = \exp\left(-e^{\tilde{u}_{n,s,k} - \epsilon_{n,0}}\right). \tag{44}$$

Hence,
$$E_{n,s,k}(d_{n,k}, \epsilon_{n,0}) \approx \frac{-\alpha_{n,s,k} \cdot d_{n,k}}{1 - \exp\left[e^{u_{n,s,k} - \epsilon_{n,0}}\left(1 - e^{-\alpha_{n,s,k} \cdot d_{n,k}}\right)\right]} + \frac{1}{-e^{u_{n,s,k} - \epsilon_{n,0}}}. \tag{45}$$

Proof see C.3

Numerical Stability
Stable Algorithm

- **Lower region** : approximate the value using the limit as $u_{n,s,k} - \epsilon_{n,0} \to -\infty$,

$$E_{n,s,k}(d_{n,k}, \epsilon_{n,0}) \approx 1 - \alpha_{n,s,k} \cdot d_{n,k} \cdot \frac{e^{\alpha_{n,s,k} \cdot d_{n,k}}}{e^{\alpha_{n,s,k} \cdot d_{n,k}} - 1}. \tag{42}$$

- **Central region** : already stable.
- **Upper region** : rewrite to

$$E_{n,s,k}(d_{n,k}, \epsilon_{n,0}) = \frac{-\alpha_{n,s,k} \cdot d_{n,k}}{1 - \exp\left[e^{u_{n,s,k} - \epsilon_{n,0}}\left(1 - e^{-\alpha_{n,s,k} \cdot d_{n,k}}\right)\right]} + \underbrace{\frac{\mathrm{Ei}\left(\log(y_1)\right) - \mathrm{Ei}\left(\log(y_2)\right)}{y_1 - y_2}}_{\approx \frac{d}{dy} \mathrm{Ei}(\log(y))\Big|_{y=y_1} = \frac{1}{\log(y_1)}} \tag{43}$$

where $$y_1 = \exp\left(-e^{u_{n,s,k} - \epsilon_{n,0}}\right), \text{ and } y_2 = \exp\left(-e^{\tilde{u}_{n,s,k} - \epsilon_{n,0}}\right). \tag{44}$$

Hence, $$E_{n,s,k}(d_{n,k}, \epsilon_{n,0}) \approx \frac{-\alpha_{n,s,k} \cdot d_{n,k}}{1 - \exp\left[e^{u_{n,s,k} - \epsilon_{n,0}}\left(1 - e^{-\alpha_{n,s,k} \cdot d_{n,k}}\right)\right]} + \frac{1}{-e^{u_{n,s,k} - \epsilon_{n,0}}}. \tag{45}$$

Proof see C.3

Numerical Stability
Stable Algorithm

- **Lower region** : approximate the value using the limit as $u_{n,s,k} - \epsilon_{n,0} \to -\infty$,

$$E_{n,s,k}(d_{n,k}, \epsilon_{n,0}) \approx 1 - \alpha_{n,s,k} \cdot d_{n,k} \cdot \frac{e^{\alpha_{n,s,k} \cdot d_{n,k}}}{e^{\alpha_{n,s,k} \cdot d_{n,k}} - 1}. \tag{42}$$

- **Central region** : already stable.
- **Upper region** : rewrite to

$$E_{n,s,k}(d_{n,k}, \epsilon_{n,0}) = \frac{-\alpha_{n,s,k} \cdot d_{n,k}}{1 - \exp\left[e^{u_{n,s,k} - \epsilon_{n,0}} \left(1 - e^{-\alpha_{n,s,k} \cdot d_{n,k}}\right)\right]} + \underbrace{\frac{\mathrm{Ei}\left(\log(y_1)\right) - \mathrm{Ei}\left(\log(y_2)\right)}{y_1 - y_2}}_{\approx \frac{d}{dy} \mathrm{Ei}(\log(y))\Big|_{y=y_1} = \frac{1}{\log(y_1)}} \tag{43}$$

where $$y_1 = \exp\left(-e^{u_{n,s,k} - \epsilon_{n,0}}\right), \text{ and } y_2 = \exp\left(-e^{\tilde{u}_{n,s,k} - \epsilon_{n,0}}\right). \tag{44}$$

Hence, $$E_{n,s,k}(d_{n,k}, \epsilon_{n,0}) \approx \frac{-\alpha_{n,s,k} \cdot d_{n,k}}{1 - \exp\left[e^{u_{n,s,k} - \epsilon_{n,0}} \left(1 - e^{-\alpha_{n,s,k} \cdot d_{n,k}}\right)\right]} + \frac{1}{-e^{u_{n,s,k} - \epsilon_{n,0}}}. \tag{45}$$

Proof see C.3

Numerical Stability

Absolute Error of Numerically Stable Algorithm

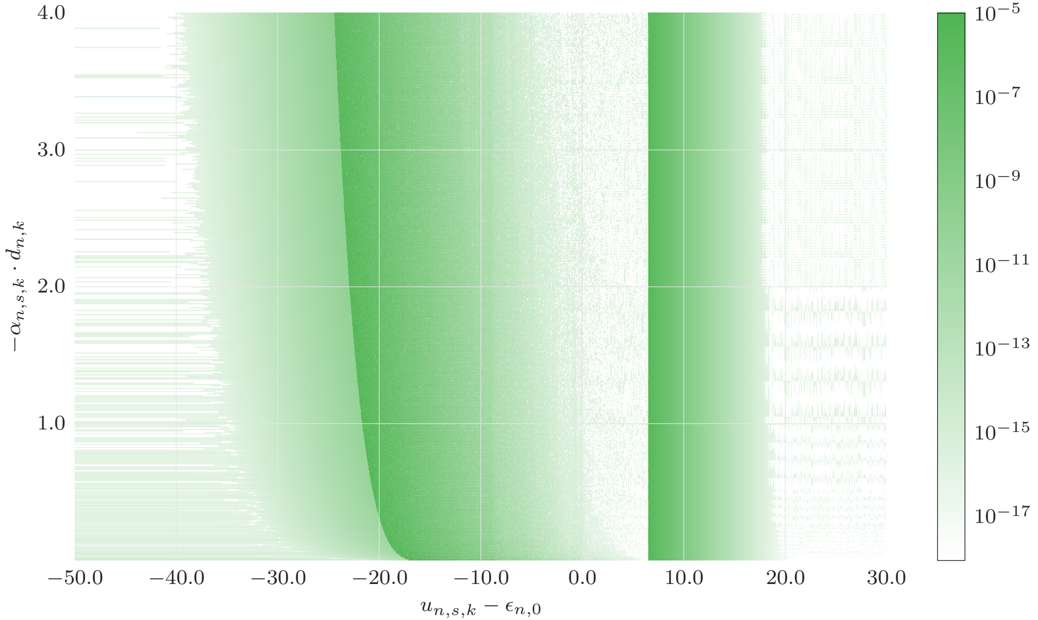


Table of Contents

Results

Estimation Results of Personalized Discount

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.

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			42,782,814.57		
Conditional CS before coupon ($)	4,486,714.16			3,268,146.22		
$\mathcal{K}_n$ schemes	**Last click**	**All clicked**	**All $\mathcal{D}_{t(n)}$**	**Last click**	**All clicked**	**All $\mathcal{D}_{t(n)}$**
Gain in conditional profit ($)	331,929.23	1,077,208.15	1,081,419.93	388,962.47	1,286,091.25	1,289,161.04
Gain in conditional CS ($)	145,054.18	160,178.22	219,572.72	144,972.20	158,648.06	205,844.97
Gain in sales units	3,940.39	11,675.92	11,677.82	4,816.79	14,334.33	14,333.35
Gain in conditional profit (%)	0.81	2.64	2.65	0.95	3.15	3.15
Gain in conditional CS (%)	3.23	3.57	4.89	4.44	4.86	6.30
Gain in sales units (%)	0.82	2.44	2.44	1.00	2.99	2.99
Share of discounted smartphone was clicked (%)	100.00	100.00	55.05	100.00	100.00	65.15
Average coupon value ($)	61.22	65.93	87.99	61.52	65.59	83.22
# 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

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			42,782,814.57		
Conditional CS before coupon ($)	4,486,714.16			3,268,146.22		
$\mathcal{K}_n$ **schemes**	**Last click**	**All clicked**	**All** $\mathcal{D}_{t(n)}$	**Last click**	**All clicked**	**All** $\mathcal{D}_{t(n)}$
Gain in conditional profit ($)	331,929.23	1,077,208.15	1,081,419.93	388,962.47	1,286,091.25	1,289,161.04
Gain in conditional CS ($)	145,054.18	160,178.22	219,572.72	144,972.20	158,648.06	205,844.97
Gain in sales units	3,940.39	11,675.92	11,677.82	4,816.79	14,334.33	14,333.35
Gain in conditional profit (%)	0.81	2.64	2.65	0.95	3.15	3.15
Gain in conditional CS (%)	3.23	3.57	4.89	4.44	4.86	6.30
Gain in sales units (%)	0.82	2.44	2.44	1.00	2.99	2.99
Share of discounted smartphone was clicked (%)	100.00	100.00	55.05	100.00	100.00	65.15
Average coupon value ($)	61.22	65.93	87.99	61.52	65.59	83.22
# 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

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			42,782,814.57		
Conditional CS before coupon ($)	4,486,714.16			3,268,146.22		
$\mathcal{K}_n$ **schemes**	**Last click**	**All clicked**	**All** $\mathcal{D}_{t(n)}$	**Last click**	**All clicked**	**All** $\mathcal{D}_{t(n)}$
Gain in conditional profit ($)	331,929.23	1,077,208.15	1,081,419.93	388,962.47	1,286,091.25	1,289,161.04
Gain in conditional CS ($)	145,054.18	160,178.22	219,572.72	144,972.20	158,648.06	205,844.97
Gain in sales units	3,940.39	11,675.92	11,677.82	4,816.79	14,334.33	14,333.35
Gain in conditional profit (%)	0.81	2.64	2.65	0.95	3.15	3.15
Gain in conditional CS (%)	3.23	3.57	4.89	4.44	4.86	6.30
Gain in sales units (%)	0.82	2.44	2.44	1.00	2.99	2.99
Share of discounted smartphone was clicked (%)	100.00	100.00	55.05	100.00	100.00	65.15
Average coupon value ($)	61.22	65.93	87.99	61.52	65.59	83.22
# 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

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			42,782,814.57		
Conditional CS before coupon ($)	4,486,714.16			3,268,146.22		
$\mathcal{K}_n$ schemes	**Last click**	**All clicked**	**All $\mathcal{D}_{t(n)}$**	**Last click**	**All clicked**	**All $\mathcal{D}_{t(n)}$**
Gain in conditional profit ($)	331,929.23	1,077,208.15	1,081,419.93	388,962.47	1,286,091.25	1,289,161.04
Gain in conditional CS ($)	145,054.18	160,178.22	219,572.72	144,972.20	158,648.06	205,844.97
Gain in sales units	3,940.39	11,675.92	11,677.82	4,816.79	14,334.33	14,333.35
Gain in conditional profit (%)	0.81	2.64	2.65	0.95	3.15	3.15
Gain in conditional CS (%)	3.23	3.57	4.89	4.44	4.86	6.30
Gain in sales units (%)	0.82	2.44	2.44	1.00	2.99	2.99
Share of discounted smartphone was clicked (%)	100.00	100.00	55.05	100.00	100.00	65.15
Average coupon value ($)	61.22	65.93	87.99	61.52	65.59	83.22
# 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

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			42,782,814.57	
Conditional CS before coupon ($)		4,486,714.16			3,268,146.22	
$\mathcal{K}_n$ schemes	**Last click**	**All clicked**	**All $\mathcal{D}_{t(n)}$**	**Last click**	**All clicked**	**All $\mathcal{D}_{t(n)}$**
Gain in conditional profit ($)	331,929.23	1,077,208.15	1,081,419.93	388,962.47	1,286,091.25	1,289,161.04
Gain in conditional CS ($)	145,054.18	160,178.22	219,572.72	144,972.20	158,648.06	205,844.97
Gain in sales units	3,940.39	11,675.92	11,677.82	4,816.79	14,334.33	14,333.35
Gain in conditional profit (%)	0.81	2.64	2.65	0.95	3.15	3.15
Gain in conditional CS (%)	3.23	3.57	4.89	4.44	4.86	6.30
Gain in sales units (%)	0.82	2.44	2.44	1.00	2.99	2.99
Share of discounted smartphone was clicked (%)	100.00	100.00	55.05	100.00	100.00	65.15
Average coupon value ($)	61.22	65.93	87.99	61.52	65.59	83.22
# 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

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			42,782,814.57		
Conditional CS before coupon ($)	4,486,714.16			3,268,146.22		
$\mathcal{K}_n$ **schemes**	**Last click**	**All clicked**	**All** $\mathcal{D}_{t(n)}$	**Last click**	**All clicked**	**All** $\mathcal{D}_{t(n)}$
Gain in conditional profit ($)	331,929.23	1,077,208.15	1,081,419.93	388,962.47	1,286,091.25	1,289,161.04
Gain in conditional CS ($)	145,054.18	160,178.22	219,572.72	144,972.20	158,648.06	205,844.97
Gain in sales units	3,940.39	11,675.92	11,677.82	4,816.79	14,334.33	14,333.35
Gain in conditional profit (%)	0.81	2.64	2.65	0.95	3.15	3.15
Gain in conditional CS (%)	3.23	3.57	4.89	4.44	4.86	6.30
Gain in sales units (%)	0.82	2.44	2.44	1.00	2.99	2.99
Share of discounted smartphone was clicked (%)	100.00	100.00	55.05	100.00	100.00	65.15
Average coupon value ($)	61.22	65.93	87.99	61.52	65.59	83.22
# 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

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			42,782,814.57	
Conditional CS before coupon ($)		4,486,714.16			3,268,146.22	
$\mathcal{K}_n$ schemes	**Last click**	**All clicked**	**All $\mathcal{D}_{t(n)}$**	**Last click**	**All clicked**	**All $\mathcal{D}_{t(n)}$**
Gain in conditional profit ($)	331,929.23	1,077,208.15	1,081,419.93	388,962.47	1,286,091.25	1,289,161.04
Gain in conditional CS ($)	145,054.18	160,178.22	219,572.72	144,972.20	158,648.06	205,844.97
Gain in sales units	3,940.39	11,675.92	11,677.82	4,816.79	14,334.33	14,333.35
Gain in conditional profit (%)	0.81	2.64	2.65	0.95	3.15	3.15
Gain in conditional CS (%)	3.23	3.57	4.89	4.44	4.86	6.30
Gain in sales units (%)	0.82	2.44	2.44	1.00	2.99	2.99
Share of discounted smartphone was clicked (%)	100.00	100.00	55.05	100.00	100.00	65.15
Average coupon value ($)	61.22	65.93	87.99	61.52	65.59	83.22
# 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

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			42,782,814.57		
Conditional CS before coupon ($)	4,486,714.16			3,268,146.22		
$\mathcal{K}_n$ schemes	**Last click**	**All clicked**	**All $\mathcal{D}_{t(n)}$**	**Last click**	**All clicked**	**All $\mathcal{D}_{t(n)}$**
Gain in conditional profit ($)	331,929.23	1,077,208.15	1,081,419.93	388,962.47	1,286,091.25	1,289,161.04
Gain in conditional CS ($)	145,054.18	160,178.22	219,572.72	144,972.20	158,648.06	205,844.97
Gain in sales units	3,940.39	11,675.92	11,677.82	4,816.79	14,334.33	14,333.35
Gain in conditional profit (%)	0.81	2.64	2.65	0.95	3.15	3.15
Gain in conditional CS (%)	3.23	3.57	4.89	4.44	4.86	6.30
Gain in sales units (%)	0.82	2.44	2.44	1.00	2.99	2.99
Share of discounted smartphone was clicked (%)	100.00	100.00	55.05	100.00	100.00	65.15
Average coupon value ($)	61.22	65.93	87.99	61.52	65.59	83.22
# 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

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			42,782,814.57		
Conditional CS before coupon ($)	4,486,714.16			3,268,146.22		
$\mathcal{K}_n$ schemes	**Last click**	**All clicked**	**All $\mathcal{D}_{t(n)}$**	**Last click**	**All clicked**	**All $\mathcal{D}_{t(n)}$**
Gain in conditional profit ($)	331,929.23	1,077,208.15	1,081,419.93	388,962.47	1,286,091.25	1,289,161.04
Gain in conditional CS ($)	145,054.18	160,178.22	219,572.72	144,972.20	158,648.06	205,844.97
Gain in sales units	3,940.39	11,675.92	11,677.82	4,816.79	14,334.33	14,333.35
Gain in conditional profit (%)	0.81	2.64	2.65	0.95	3.15	3.15
Gain in conditional CS (%)	3.23	3.57	4.89	4.44	4.86	6.30
Gain in sales units (%)	0.82	2.44	2.44	1.00	2.99	2.99
Share of discounted smartphone was clicked (%)	100.00	100.00	55.05	100.00	100.00	65.15
Average coupon value ($)	61.22	65.93	87.99	61.52	65.59	83.22
# 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

Table of Contents

Conclusion
Contributions

- Outlines a marketing approach that achieves a mutually beneficial outcome for both sellers and buyers.
- Develops a numerically stable algorithm that ensures both efficiency and numerical stability in calculating conditional expected consumer surplus.
- Lays the groundwork for subsequent studies that endogenize more strategic behaviors of consumers and firms.

Conclusion
Contributions

- Outlines a marketing approach that achieves a mutually beneficial outcome for both sellers and buyers.
- Develops a numerically stable algorithm that ensures both efficiency and numerical stability in calculating conditional expected consumer surplus.
- Lays the groundwork for subsequent studies that endogenize more strategic behaviors of consumers and firms.

Conclusion
Contributions

- Outlines a marketing approach that achieves a mutually beneficial outcome for both sellers and buyers.
- Develops a numerically stable algorithm that ensures both efficiency and numerical stability in calculating conditional expected consumer surplus.
- Lays the groundwork for subsequent studies that endogenize more strategic behaviors of consumers and firms.

Conclusion
Takeaways

- Demonstrates that personalized discounts as a marketing strategy can enable the firm to achieve an immediate profit increase and, at the same time, boost consumer surplus, in conditional expected terms.
- Offers further numerical insights into the welfare effects arising from personalized discounts, by leveraging the demand model which intricately captures individual-level heterogeneity and the foundation laid by a numerically stable algorithm.
- Showcases that 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.
- Emphasizes that the firm's profit and consumer welfare are not necessarily opposing objectives, promoting the analysis of policy frameworks that could benefit both firms and consumers and promote economic efficiencies.

Conclusion
Takeaways

- Demonstrates that personalized discounts as a marketing strategy can enable the firm to achieve an immediate profit increase and, at the same time, boost consumer surplus, in conditional expected terms.
- Offers further numerical insights into the welfare effects arising from personalized discounts, by leveraging the demand model which intricately captures individual-level heterogeneity and the foundation laid by a numerically stable algorithm.
- Showcases that 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.
- Emphasizes that the firm's profit and consumer welfare are not necessarily opposing objectives, promoting the analysis of policy frameworks that could benefit both firms and consumers and promote economic efficiencies.

Conclusion
Takeaways

- Demonstrates that personalized discounts as a marketing strategy can enable the firm to achieve an immediate profit increase and, at the same time, boost consumer surplus, in conditional expected terms.
- Offers further numerical insights into the welfare effects arising from personalized discounts, by leveraging the demand model which intricately captures individual-level heterogeneity and the foundation laid by a numerically stable algorithm.
- Showcases that 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.
- Emphasizes that the firm's profit and consumer welfare are not necessarily opposing objectives, promoting the analysis of policy frameworks that could benefit both firms and consumers and promote economic efficiencies.

Conclusion

Takeaways

- Demonstrates that personalized discounts as a marketing strategy can enable the firm to achieve an immediate profit increase and, at the same time, boost consumer surplus, in conditional expected terms.
- Offers further numerical insights into the welfare effects arising from personalized discounts, by leveraging the demand model which intricately captures individual-level heterogeneity and the foundation laid by a numerically stable algorithm.
- Showcases that 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.
- Emphasizes that the firm's profit and consumer welfare are not necessarily opposing objectives, promoting the analysis of policy frameworks that could benefit both firms and consumers and promote economic efficiencies.

Georgetown University

Essays on Consumer Heterogeneity and Personalized Discounts in an Online Market

Ph.D. Dissertation Defense

Chengjun Zhang

April 05, 2024

Appendix

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
	perc.	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
	perc.	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
	perc.	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
	perc.	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
	perc.	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
	perc.	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
	perc.	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
	perc.	48.93%	4.42%	22.05%	24.59%	100.00%

Back

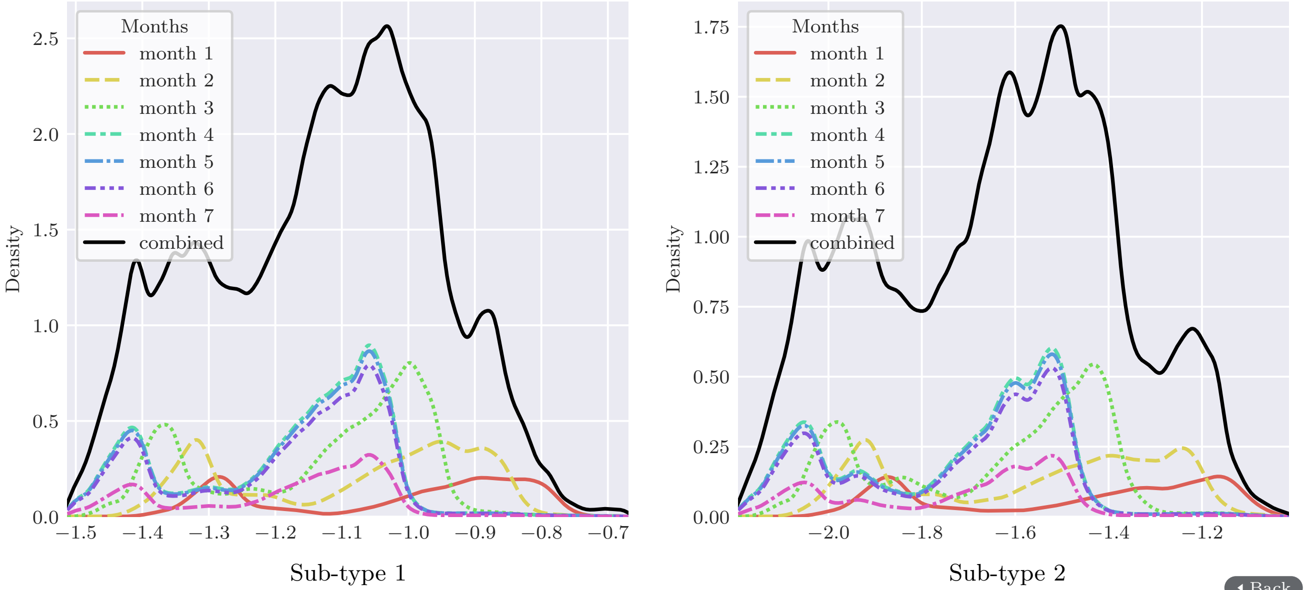
Appendix
View Events Classification Results (by μ)

	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%

Back

Result of Model IV - MLP

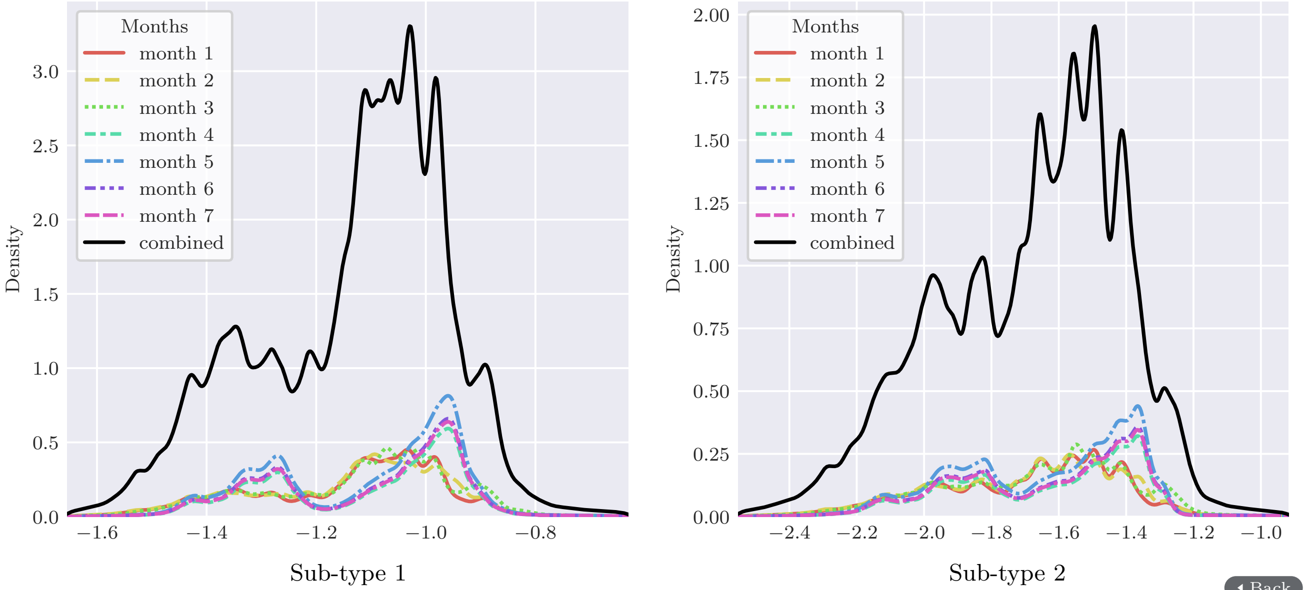
Distribution of Price ($1,000) Coefficients (by Active Months)



◂ Back

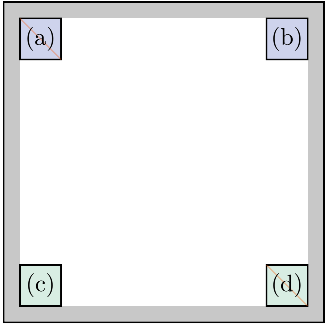
Result of Model V - RNN

Distribution of Price ($1,000) Coefficients (by Active Months)



◂ Back

Appendix
Illustration of Cross-price Elasticity Heat Map

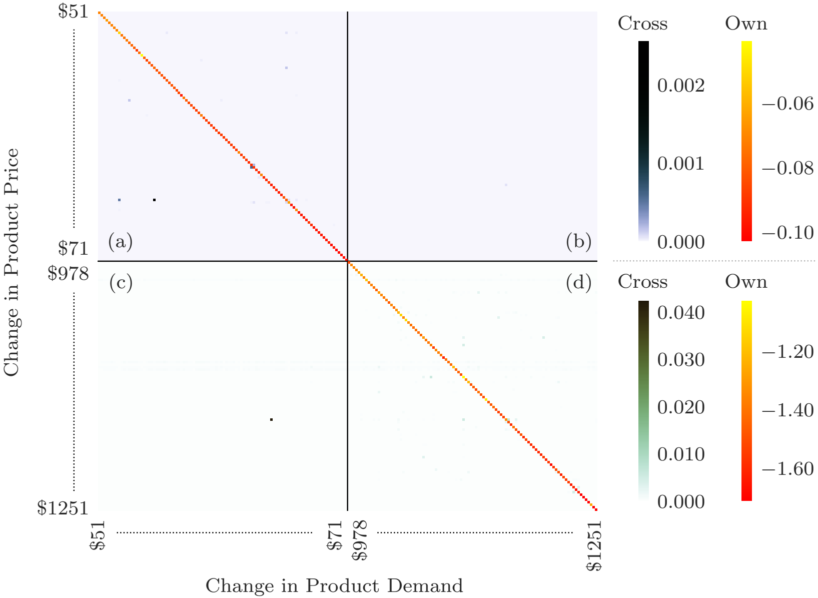


The outermost square depicted outlines the full spectrum of cross- and own-price elasticities, organized with both the rows and columns arranged in ascending order according to the prices of smartphones. Given that a product's price can fluctuate over time, the median price per event is utilized for 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 figures.

◂ Back

Appendix

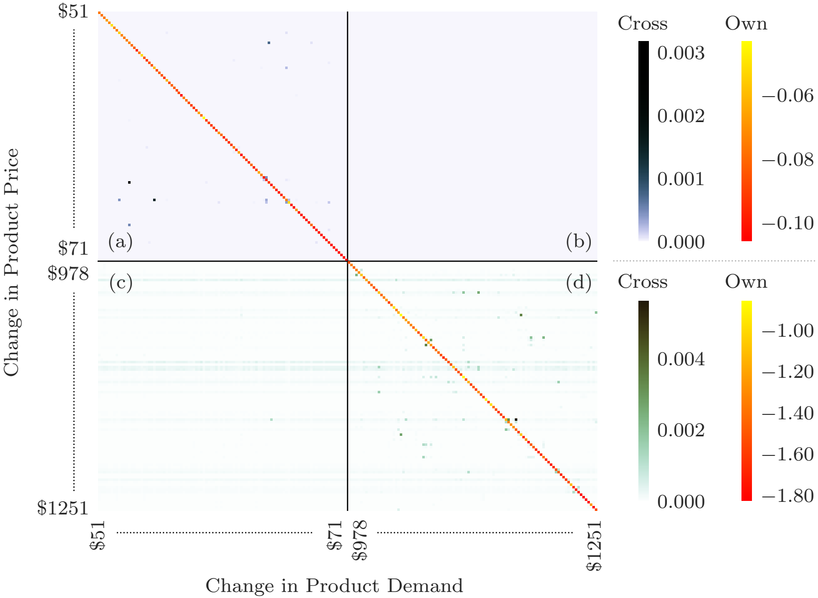
Price Elasticities: Model I - Baseline Logit



Back

Appendix

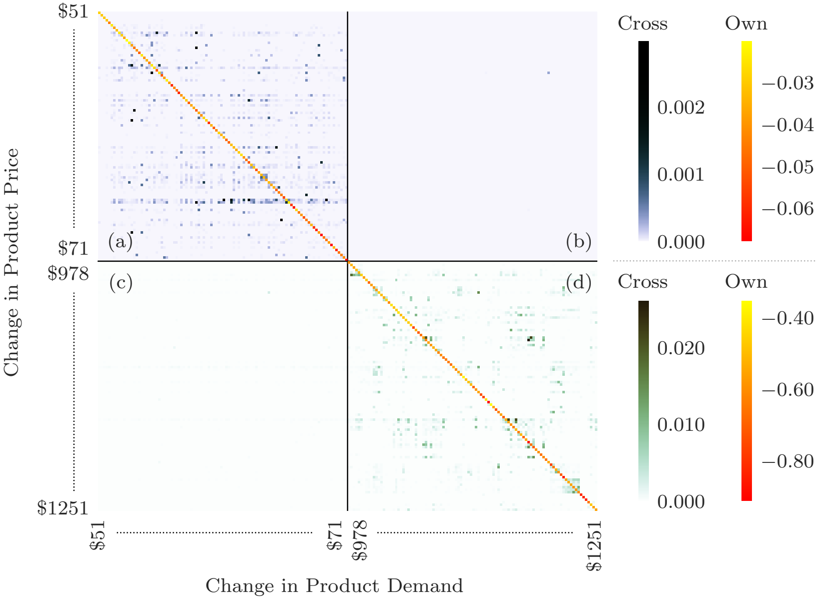
Price Elasticities: Model II - by User Types



Back

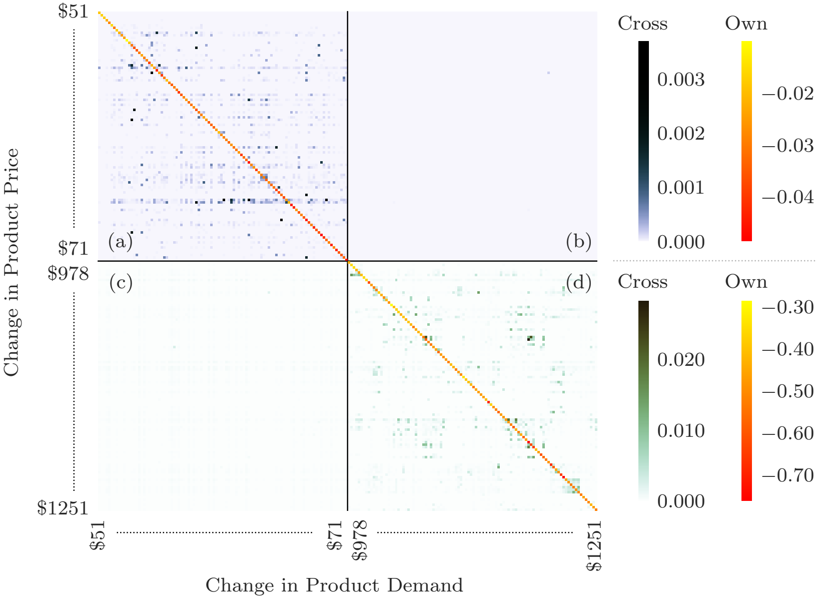
Appendix

Price Elasticities: Model III - with Unobserved Heterogeneity



◂ Back

Appendix
Price Elasticities: Model IV - MLP



◂ Back

Appendix
Formulas of AIC and BIC

$$\text{AIC} = 2k - 2\ln(\hat{\ell}\,)$$

$$\text{BIC} = k\ln(n) - 2\ln(\hat{\ell}\,)$$

Back

Appendix

Model IV - MLP *(2,161 parameters)*

Let $\sigma(\cdot)$ denote the function that applies element-wise ReLU, where $\text{ReLU}(x) = \max\{0, x\}$.

For user n, let the input $\mathbf{x} = \mathbf{b}_n^\top \in \mathbb{R}^1 \times \mathbb{R}^{206}$.

1. $\text{Linear}^{(1)}(\mathbf{x})$	$\mathbf{h}^{(1)} = \mathbf{x}\mathbf{W}^{(1)}$	$\in \mathbb{R}^1 \times \mathbb{R}^8$
2. $\text{Activation}\left(\mathbf{h}^{(1)}\right)$	$\mathbf{h}^{(2)} = \sigma\left(\mathbf{h}^{(1)}\right)$	$\in \mathbb{R}^1 \times \mathbb{R}^8$
3. $\text{Linear}^{(2)}\left(\mathbf{h}^{(2)}\right)$	$\mathbf{h}^{(3)} = \mathbf{h}^{(2)}\mathbf{W}^{(2)} + \mathbf{b}^{(2)}$	$\in \mathbb{R}^1 \times \mathbb{R}^8$
4. $\text{Activation}\left(\mathbf{h}^{(3)}\right)$	$\mathbf{h}^{(4)} = \sigma\left(\mathbf{h}^{(3)}\right)$	$\in \mathbb{R}^1 \times \mathbb{R}^8$
5. $\text{Linear}^{(3)}\left(\mathbf{h}^{(4)}\right)$	$\boldsymbol{\theta}_{\text{MLP},n}{}^\top = \mathbf{h}^{(4)}\mathbf{W}^{(3)} + \mathbf{b}^{(3)}$	$\in \mathbb{R}^1 \times \mathbb{R}^{49}$

Hence, $\boldsymbol{\Theta}_{\text{IV}} = \left\{\mathbf{W}^{(1)}, \mathbf{W}^{(2)}, \mathbf{b}^{(2)}, \mathbf{W}^{(3)}, \mathbf{b}^{(3)}\right\}$.

Back

Appendix

Model V - RNN *(1,359 parameters)*

Let $\sigma(\cdot)$ denote the function that applies element-wise ReLU, where $\text{ReLU}(x) = \max\{0, x\}$.

Let $\mathbf{e} \in \mathbb{R}^{22}$ denote an embedding vector for 21 product categories (level 1) and the padded entry.

For user n, let the input $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_{50}] = \mathbf{B}_n^\intercal \in \mathbb{R}^{21} \times \mathbb{R}^{50}$. Exactly one element in each of $\mathbf{x}_1, \ldots, \mathbf{x}_{50}$ contains the element of the corresponding product category (or padded entry) from $\mathbf{e}$.

1. Initial Hidden State	$\mathbf{h}_0$	$\in \mathbb{R}^1 \times \mathbb{R}^{16}$
2. For $t = 1, \ldots, 50:$ $\{$		
RNN Hidden State	$\mathbf{h}_t = \tanh(\mathbf{x}_t \mathbf{W}_{\mathbf{xh}} + \mathbf{h}_{t-1} \mathbf{W}_{\mathbf{hh}} + \mathbf{b}_{\mathbf{h}})$	$\in \mathbb{R}^1 \times \mathbb{R}^{16}$
$\}$		
3. Subsequent MLP Input	$\mathbf{q}^{(1)} = [\mathbf{h}_{50}, 5 \text{ brand dummies}, 4 \text{ user type dummies}]$	$\in \mathbb{R}^1 \times \mathbb{R}^{25}$
4. $\text{Linear}^{(1)}\left(\mathbf{q}^{(1)}\right)$	$\mathbf{q}^{(2)} = \mathbf{q}^{(1)} \mathbf{W}^{(1)}$	$\in \mathbb{R}^1 \times \mathbb{R}^8$
5. $\text{Activation}\left(\mathbf{q}^{(2)}\right)$	$\mathbf{q}^{(3)} = \sigma\left(\mathbf{q}^{(2)}\right)$	$\in \mathbb{R}^1 \times \mathbb{R}^8$
6. $\text{Linear}^{(2)}\left(\mathbf{q}^{(3)}\right)$	$\mathbf{q}^{(4)} = \mathbf{q}^{(3)} \mathbf{W}^{(2)} + \mathbf{b}^{(2)}$	$\in \mathbb{R}^1 \times \mathbb{R}^8$
7. $\text{Activation}\left(\mathbf{q}^{(4)}\right)$	$\mathbf{q}^{(5)} = \sigma\left(\mathbf{q}^{(4)}\right)$	$\in \mathbb{R}^1 \times \mathbb{R}^8$
8. $\text{Linear}^{(3)}\left(\mathbf{q}^{(5)}\right)$	$\boldsymbol{\theta}_{\text{RNN},n} = \mathbf{q}^{(5)} \mathbf{W}^{(3)} + \mathbf{b}^{(3)}$	$\in \mathbb{R}^1 \times \mathbb{R}^{49}$

Hence, $\boldsymbol{\Theta}_{\text{V}} = \left\{\mathbf{e}, \mathbf{h}_0, \mathbf{W}_{\mathbf{xh}}, \mathbf{W}_{\mathbf{hh}}, \mathbf{b}_{\mathbf{h}}, \mathbf{W}^{(1)}, \mathbf{W}^{(2)}, \mathbf{b}^{(2)}, \mathbf{W}^{(3)}, \mathbf{b}^{(3)}\right\}$.

Back

Appendix
List of Level 1 and 2 Categories

Detailed list of subcategories is hidden for confidentiality reasons in this shared presentation.

Back

Appendix

Proof of Conditional Probability (Page 1/2)

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 $d > 0$ given for smartphone k is as follows,

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

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 $d > 0, \alpha_{n,s} < 0$. 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

Back

Appendix

Proof of Conditional Probability (Page 2/2)

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

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

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

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

$$=\frac{\Delta_{n,s,k}(d_{n,k})}{\Delta_{n,s,k}(d_{n,k}) + \sum_{j\in\mathcal{C}_n} \exp(u_{n,s,j})} \tag{30}$$

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

Back

Appendix

Boundary Conditions for Stable Algorithm

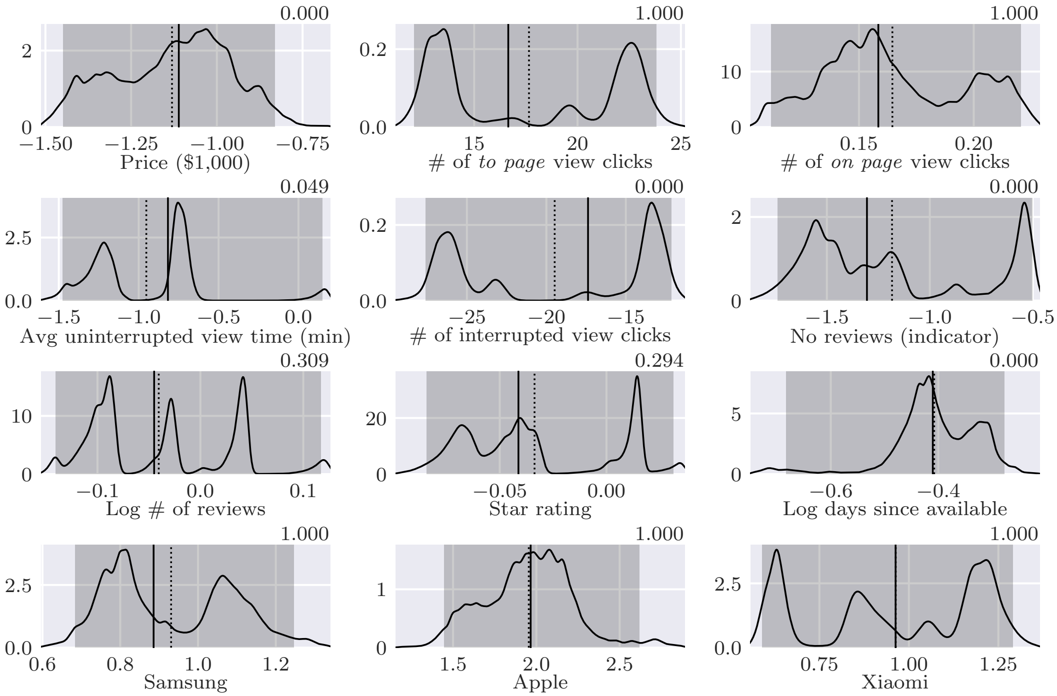
- **Lower region**:

$$-\alpha_{n,s,k} \cdot d_{n,k} \leq 5 \times 10^{-4} \times (u_{n,s,k} - \epsilon_{n,0} + 15)^4 \wedge u_{n,s,k} - \epsilon_{n,0} \leq 15 \tag{A.7}$$

- **Upper region**:

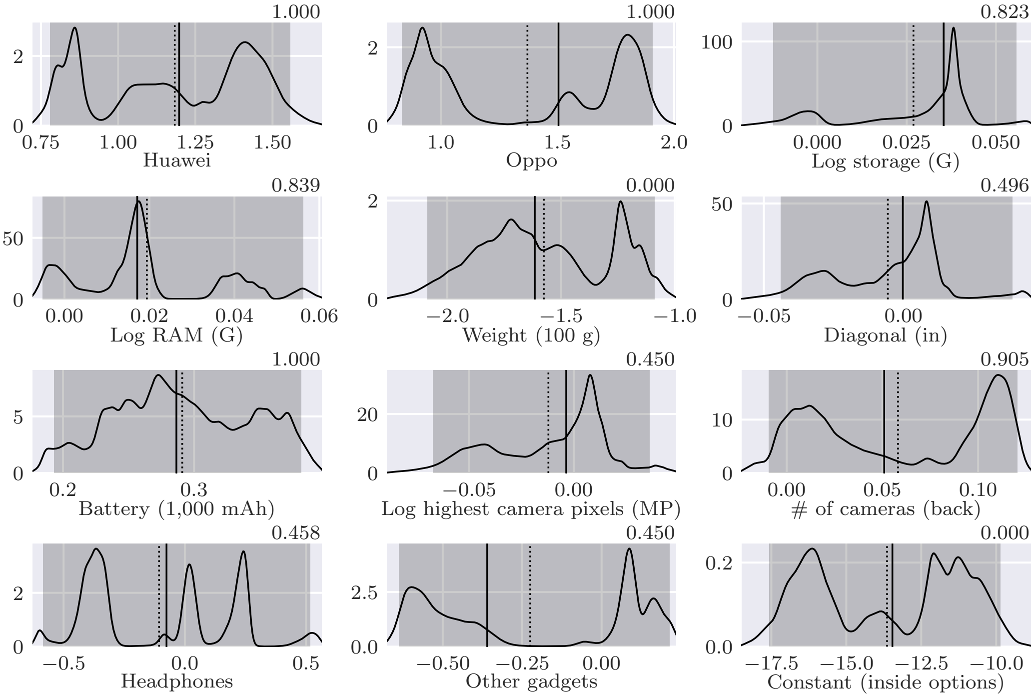
$$u_{n,s,k} - \epsilon_{n,0} \geq 6.5 \tag{A.8}$$

Back



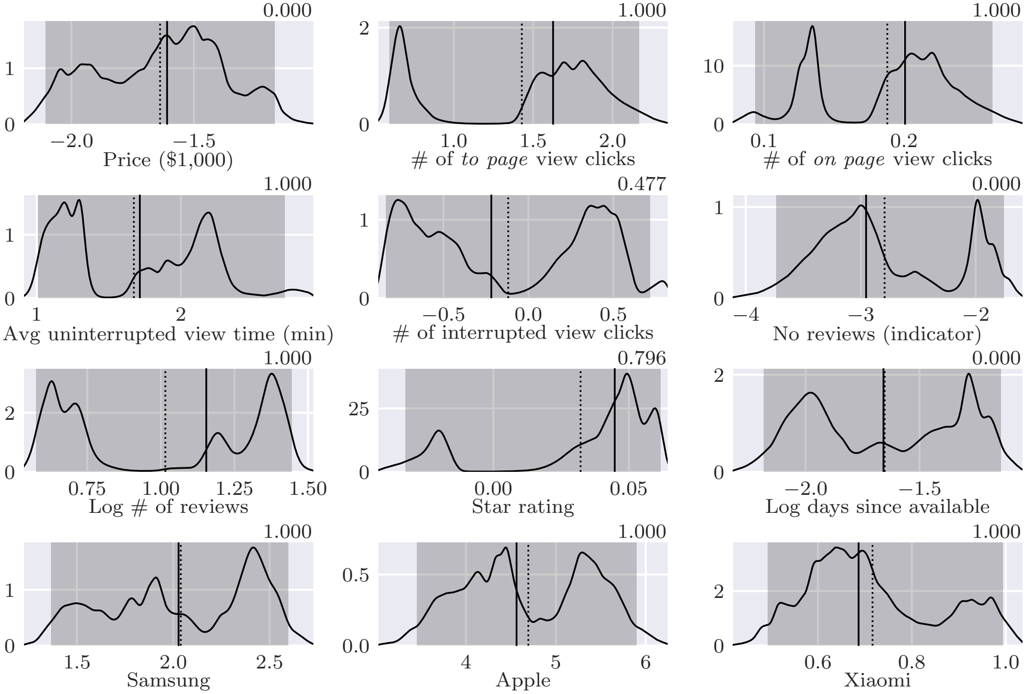
Result of Model IV - MLP: Sub-type 1 (Top half)

Back



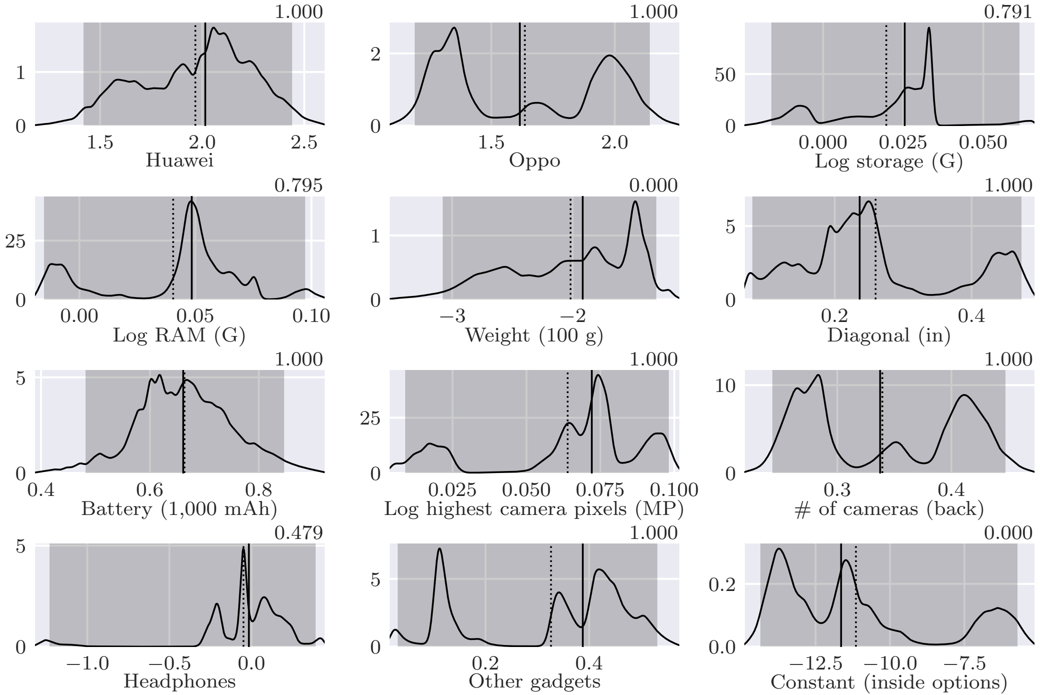
Result of Model IV - MLP: Sub-type 1 (Bottom half)

Back



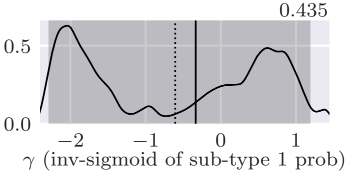
Result of Model IV - MLP: Sub-type 2 (Top half)

Back

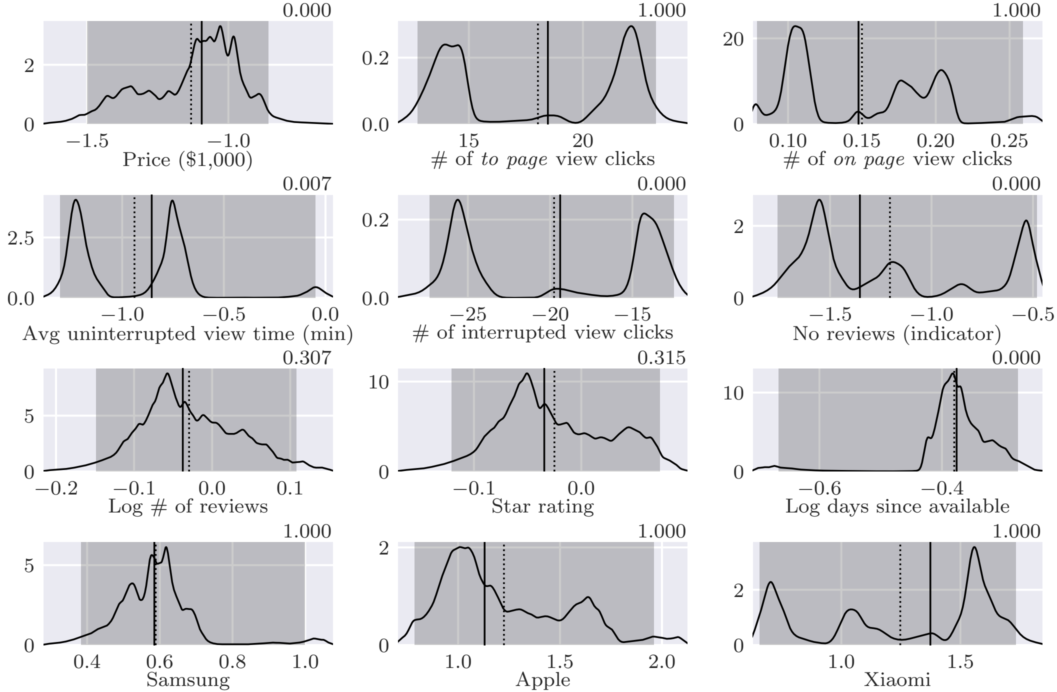


Result of Model IV - MLP: Sub-type 2 (Bottom half)

Back

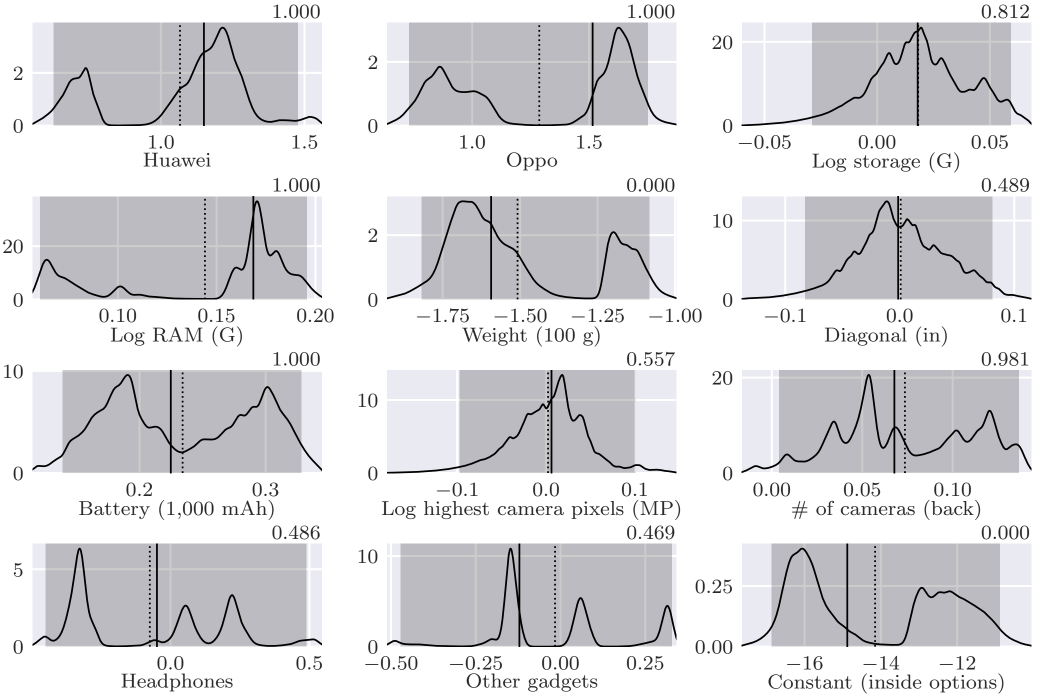


Result of Model IV - MLP: Parameter γ



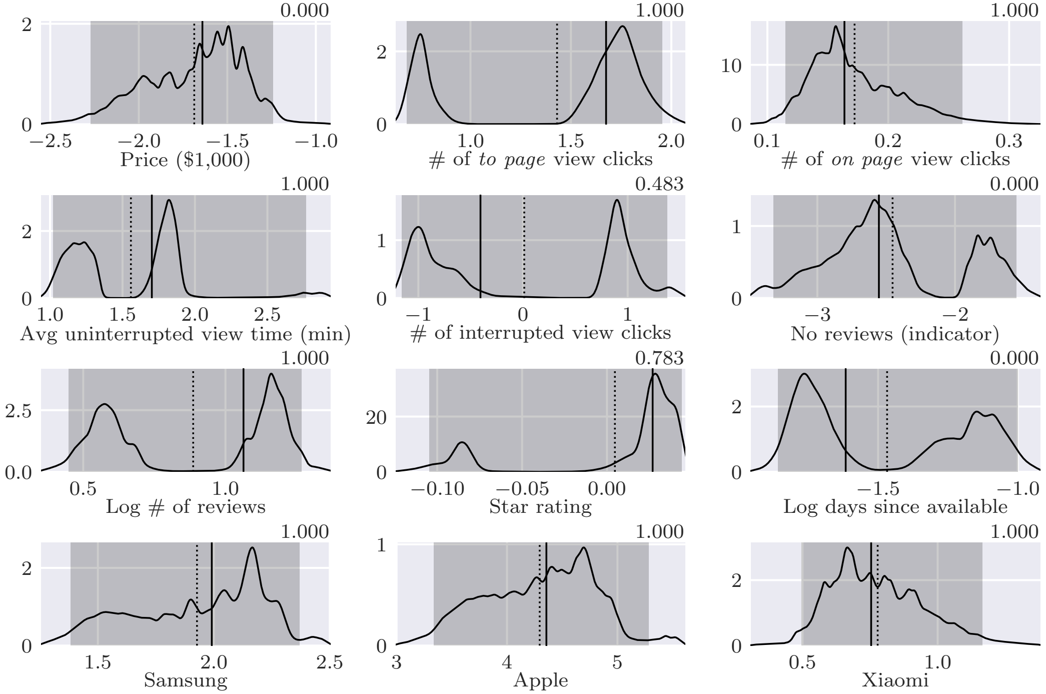
Result of Model V - RNN: Sub-type 1 (Top half)

Back



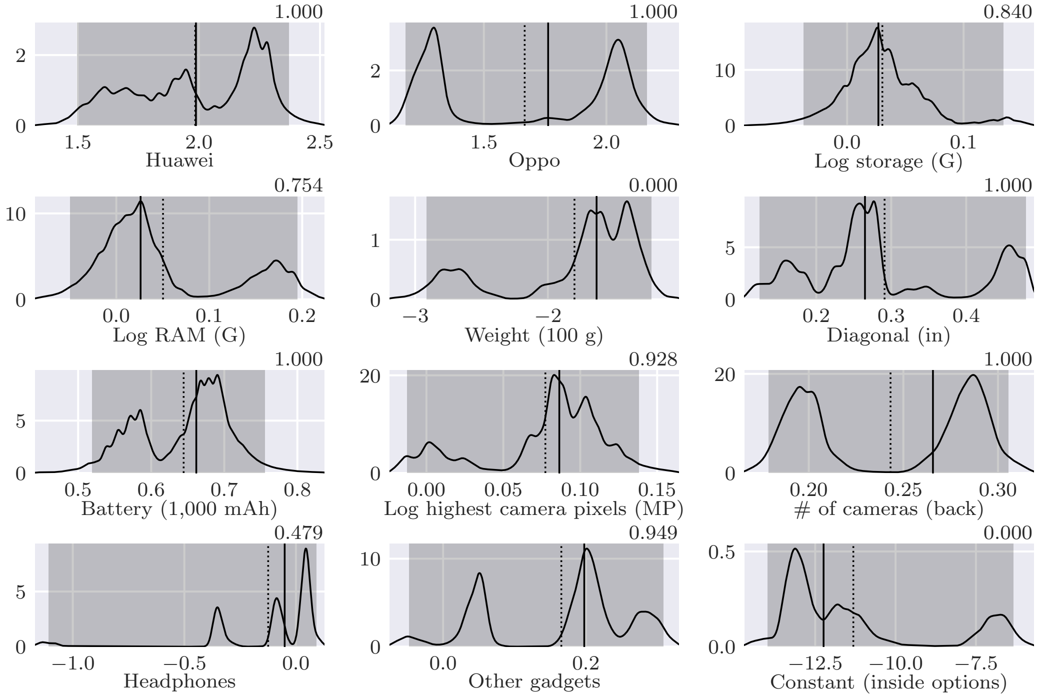
Result of Model V - RNN: Sub-type 1 (Bottom half)

Back



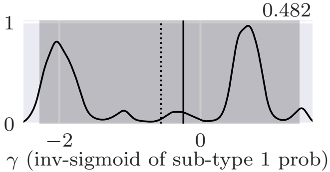
Result of Model V - RNN: Sub-type 2 (Top half)

Back



Result of Model V - RNN: Sub-type 2 (Bottom half)

Back



Result of Model V - RNN: Parameter γ

Back