Do I Really Want to Buy This? Preference Discovery and Consumer Search *

Tobias J. Klein† Christoph Walsh‡ Tinghan Zhang§

This draft: June 9, 2025

Abstract

One of the most invoked assumptions in economics is that consumers know their preferences when making choices. Although theories and experiments in psychology and behavioral economics suggest that this may be unrealistic, there is relatively little evidence from the field on this question. In this paper, we use detailed clickstream data from a large Central Asian online platform to study the extent to which consumers learn about their preferences while searching for a smartphone. To quantify the speed at which this takes place and account for other factors, most notably that consumers obtain additional product information when they inspect product pages, we estimate a rich search model in which consumers learn about their willingness to pay each time they visit the checkout page. Consumers initially underestimate their price sensitivity and update it along the way. Taking this into account shows that consumers are more price sensitive than a standard search model would predict, and an intervention that prompts consumers to end their search early can lead to potential welfare loss.

JEL-classification: D83, D91, L81, M31.
Key words: preference discovery, consumer search, e-commerce.

*We would like to thank Jaap Abbring, Paulo Albuquerque, Bart Bronnenberg, and participants of EMAC 2025, Tilburg-Rotterdam WDM 2024 and Tilburg Structural Econometrics Group, and for their helpful comments.

†Tilburg University and CEPR, T.J.Klein@tilburguniversity.edu
‡Tilburg University and CEPR, C.B.T.Walsh@tilburguniversity.edu
§Tilburg University School of Economics and Operations Research, T.Zhang_2@tilburguniversity.edu

1 Introduction

A fundamental assumption commonly invoked in modern economics is the utility-maximizing behavior. This implicitly assumes that decision makers know their preferences. However, this assumption is frequently challenged in behavioral and psychological studies. Preference discovery may be particularly likely to occur when decision-makers lack experience or are unfamiliar with the decision environment (Ariely et al., 2003). For instance, consider a consumer who wants to buy an expensive and durable product, such as a smartphone, a car, or a fridge. This is a decision she takes infrequently, perhaps every few years. She needs to choose from a range of differentiated and complex options that evolve quickly over time. Even with complete information on all products, accurately evaluating each option is a difficult task. As a result, the choices she makes may not align with her true preferences.

On the other hand, as consumers gain decision-making experience, they reflect on past choices to better understand their true preferences—a process known as *preference discovery* (Plott, 1996). Preference discovery helps consumers evaluate the utility of alternatives more accurately in subsequent decisions, aligning these decisions more closely to true utility maximization (Kahneman et al., 1997; Keller and Rady, 1999). While experimental evidence increasingly highlights the prevalence of preference discovery in consumer behavior, its empirical impact in real-world markets remains underexplored.

This paper studies preference discovery within the context of consumer search and highlights its importance in evaluating market competition, strategies, and policies. In online retail, consumers typically begin their shopping journey with limited information on available options and search for additional product details to make more informed decisions. We propose that the search process enables consumers to accumulate experience and reflect on it, fostering preference discovery that, in turn, influences subsequent search and purchase decisions. Failing to consider potential preference discovery in the search process can lead to misinterpreting the motivations behind consumer behavior at different stages on shopping platforms, resulting in inaccurate market strategy predictions. For example, if consumers prioritize product quality when searching but shift their focus to price when purchasing, firms that advertise higher-quality, more expensive products might yield lower returns than those that implement discount strategies. Moreover, since preference discovery often stems from reflecting on search decisions, regulators should be mindful of potential consumer surplus losses when sellers solicit consumers to purchase too quickly. For instance, flash sale discounts may prompt hasty purchases and cut short the search process, which not only reduces the range of products they might have otherwise searched but also leads to purchases that may be worse within the narrowed search set of options. Understanding and quantifying preference discovery in consumer search is thus vital for businesses and regulators to address these issues effectively.

Using clickstream data from a Central Asian online platform that tracks mobile phone sales and consumer checkout behavior, we exploit institutional features to separate the effect of preference discovery from that of information acquisition in consumer search. The platform employs a streamlined checkout design: no cart, no extra transaction details, and only a few clicks to complete a purchase. Yet in 59% of cases, we observe consumers enter the checkout page but not immediately pay, either inspecting other products or leaving the site.[1]

Our interpretation suggests that consumers tend to reflect at the checkout stage, asking themselves, "Do I really want to buy this?" and make the decision to abandon their checkouts. In the absence of additional information input, we argue that the decision is triggered by the consumer's behavioral reaction to the decision-making, with preference discovery, particularly regarding price sensitivity, playing a key role. During the search stage, consumers primarily compare various attributes across alternatives, such as brand or quality. At the checkout stage, however, they are prompted to reflect on their choice behavior by reassessing the product's benefits relative to its monetary cost. This reflection enables consumers to gain a deeper understanding of their preferences and more accurately determine their true willingness to pay. As a result, adjustments in perceived price sensitivity are manifested in checkout abandonment and in subsequent search and purchase decisions.

Several existing behavioral and experimental studies support our interpretation. The "pain of paying" theory (Zellermayer, 1996; Prelec and Loewenstein, 1998) posits that psychological discomfort associated with spending increases price sensitivity. Prospect theory (Kahneman and

[1] A related concept that includes checkout abandonment is cart abandonment. However, cart abandonment does not necessarily indicate a clear intention to purchase. It may occur for research, organizing, bookmarking, waiting for sales, etc (Kukar-Kinney and Close, 2010; Huang et al., 2018).

Tversky, 1979) suggests that individuals are more sensitive to potential losses, such as a payment, than to equivalent gains. Mental accounting theory (Thaler, 1985) argues that consumers treat payments as the opening of a new mental account, prompting a reassessment of their decision under a different evaluative standard. In addition, a field experiment by Cao and Zhang (2021) shows that consumers are more likely to engage in preference learning when decisions are more likely to be realized, as in the case of checkout versus search. Together, these studies offer conceptual support for the emergence of preference discovery at the checkout page.

The model-free evidence from the data also supports our interpretation. We observe that many consumers engage in the checkout process multiple times. On average, products brought to the checkout page become cheaper as the number of checkout attempts increases, while we observe no similar trend for other product attributes after controlling for their correlation with price. Moreover, when dividing the search process into intervals based on checkout behaviors, the price divergence between searched and purchased products decreases with the number of intervals experienced. However, this trend is not significant between clicks into product pages. These findings suggest that preference changes, considered mainly in price sensitivity, are more evident before and after checkout behaviors than during the search process between checkouts.

Based on the above findings, we propose a novel sequential search model that jointly captures consumer search and (price) preference discovery. This model differs from the classical search model in two key aspects. First, we assume that consumers base their search decisions on beliefs but not knowledge about their true price sensitivity. Second, instead of proceeding directly to payment after selecting a product, consumers are directed to a checkout page, where they undergo a cognitive shift that prompts reflection on their perceived price sensitivity and leads to an update in their beliefs. This setting introduces the possibility that decisions optimal under prior beliefs may become suboptimal after belief updates, prompting regret and checkout abandonment.[2] The model distinguishes between information acquisition and preference discovery by treating them as alternating stages in the search process. During the information acquisition stage, belief-based price sensitivity is held constant, while preference discovery occurs exclusively at the checkout stage. As checkouts introduce no new product information,

[2] We assume myopic learning, meaning consumers do not anticipate belief updating or search cost changes at the checkout stage, consistent with existing theoretical and experimental research.

whether the consumer abandons the checkout serves to identify and quantify changes in perceived price sensitivity, thereby capturing the process of preference discovery.

Our study shows that, on average, consumers become increasingly sensitive to price as the search process progresses. Before their first checkout, consumers underestimate their price sensitivity by 29.8% compared to their true preferences. After two checkouts, this underestimation decreases to 19%. By the time of purchase, consumers' perceived price sensitivity is more than 10.5% higher than when they initially begin searching.

Compared to the classic sequential search model, we find that incorporating preference discovery raises the estimated own-price elasticity of purchase by approximately 1.5 times, indicating that the classic model significantly underestimates the intensity of market price competition. This arises because the classic model overlooks consumers' purchase attempts before the final transaction and portrays them as more patient than they actually are. In reality, consumers frequently attempt to purchase during the search process, but also often abandon their checkout decisions. Although checkout abandonment may lead consumers to view additional high-priced options, they tend to select lower-priced products when making a final purchase. As a result, the effectiveness of market strategies that reduce search costs may be overestimated by as much as 30%. In addition, we evaluate the impact of the "one-click purchase" mechanism, which encourages consumers to buy quickly and thereby reduces checkout abandonment. However, it also limits consumers' ability to reflect before making a final decision, potentially leading to more impulsive purchases. Our results show that while one-click purchase reduces market exit, retains more consumers, and increases total revenue, it lowers the surplus of those who would have purchased even without the mechanism. Specifically, when the adoption rate of one-click purchase exceeds 50%, the average utility per buyer declines by 5% to 9%.

Our model poses substantial challenges to existing estimation approaches in the empirical search literature. The structural complexity introduced by a multi-round sequential search process, linked by repeated checkouts, renders traditional numerical methods for standard sequential search models largely inapplicable. To address this, we first adopt the result from Zhang (2025) to recast the standard sequential search process into a fully equivalent ranking model with a more tractable probability structure. Then we evaluate the probability that the change in rankings before and after checkout conforms to a Bayesian updating process. This approach ultimately allows us to estimate the full model using simulated maximum likelihood while maintaining a manageable computational burden.

This paper relates to three strands of literature. First, it contributes to the empirical study of preference discovery. In a seminal work, Plott (1996) proposes the Discovered Preference Hypothesis, arguing that individuals initially exhibit myopic decision-making when facing unfamiliar tasks and gradually form stable preferences through repeated choices. This hypothesis has inspired a large body of theoretical (e.g., Cooke, 2017; Cerreia-Vioglio et al., 2023) and experimental research (e.g., Plott and Zeiler, 2005; Delaney et al., 2020; Cao and Zhang, 2021). However, empirical evidence has been relatively scarce until recent years. Narita (2018) uses data from the New York City school enrollment system to show that families revise their acceptance choices as they learn more about schools. Similarly, Grenet et al. (2022) find in the context of German university admissions that students are more likely to accept early offers than later ones. These studies focus on final decisions made in an exogenously structured, fully informed environment, shaped by the institutional features of the application process. In contrast, our study examines a more realistic setting in which preference discovery not only affects final choices but also shapes the search process and the formation of consideration sets.

Second, our paper is related to the research on consumer learning. This strand of literature assumes consumers' incomplete knowledge about products and assuming that choices are based on perceived rather than actual utilities, while utility beliefs can be updated from learning. Among this literature, Erdem and Keane (1996) establish a general model exploring how consumers form beliefs about brands with advertisements and repeat purchases. Subsequent studies extended this model to further delve into consumer learning about individual brands or products (e.g., Ackerberg, 2003; Crawford and Shum, 2005; Ching, 2010). Though some papers examine spillover learning within a brand across products or categories (e.g., Erdem, 1998; Balachander and Ghose, 2003), and others explored cross-brand learning (e.g., Szymanowski and Gijsbrechts, 2012), the learning impacts are usually limited to the perceived utility of a specific product or a group of related products. In contrast, our study examines preference discovery, a process that impacts perceived utilities of all alternatives simultaneously.

Last but not least, our paper is closely related to learning in consumer search. Many empirical search studies (e.g., Kim et al., 2010; Chen and Yao, 2017; Ursu, 2018; Jolivet and Turon, 2019) assume consumers only obtaining product information in search. Some theoretical papers investigate optimal search when considering further learning (e.g., Branco et al., 2012; Chick and Frazier, 2012). Recent empirical research incorporates both consumer search and learning about different aspects of the market, such as the distribution of product-level uncertainty (e.g., Koulayev, 2014; De Los Santos et al., 2017), the distribution of product attributes (e.g., Jindal and Aribarg, 2021; Gardete and Hunter, 2024; Wu et al., 2024), and beliefs over consumer-product-specific match values (e.g., Ursu et al., 2020). These papers assume learning occurs with consumer search to derive optimal responses or strategies by identifying prior beliefs with additional information and estimating the learning process. Of the many papers, Dzyabura and Hauser (2019) is the only one that considers consumers learning about their preferences. With a theoretical framework that allows perceived preference change, they show the optimal recommendation system should encourage learning by suggesting products with diverse attributes at the early part of the search process. Our paper follows the idea of this paper, providing a tractable structural framework, and estimating the model with observational data.

For the remainder of this paper, we introduce our context, data, and reduced-form evidence in Section 2. In Section 3, we present our model. Section 4 explains our estimation strategy. Section 5 contains our main empirical results. We summarize and conclude in Section 6.

2 Data

2.1 Data Source and Platform Design

Our data originates from a prominent online marketplace in a central Asian country. The clickstream dataset is anonymized, collected by the Open CDP project, and publicly shared on Kaggle.com[3]. It comprises chronological data of all consumers' clicks spanning seven months, from October 2019 to April 2020. The clicks are categorized into three types:

1. *View*: A consumer clicks on a product on the list page and approaches the product page.

[3]The clickstream dataset is available at https://www.kaggle.com/datasets/mkechinov/ecommerce-behavior-data-from-multi-category-store.

2. *Checkout*: A consumer clicks on ‘purchase’ on one of the sellers on the product page and approaches the checkout page[4].

3. *Purchase*: A consumer clicks on ‘purchase’ on the product page and confirms payment.

Each click is recorded with a user ID for a consumer account, an SKU ID for the product page, a timestamp for the slot when the click happens, and the product price related to the click. Our analysis focuses on cellphone sales within this marketplace. Because a cellphone can only communicate with a phone number, consumers usually seek to buy an optimal choice but not multiple complementary alternatives. This allows consumers to apply a consistent comparison standard across different products. Moreover, cellphones are durable and relatively expensive, prompting consumers to invest time and effort in knowing both product information and their preferences better before purchasing rather than choosing randomly.

The marketplace interface encompasses three primary web pages, corresponding to three stages of the consumers’ shopping process. Firstly, consumers browse through a list page displaying product alternatives, each accompanied by an outline picture and a product name with essential information. For example, ‘Smartphone Samsung Galaxy S10 4GB/128GB black’ provides information about the brand (Samsung), RAM (4GB), storage (128GB), and color. Subsequently, consumers can access product pages that provide more comprehensive information, including sellers, delivery periods and fees, reviews on products and sellers, and more detailed specifications. We collect the information in product names as the list-page attributes, together with part of the product-page information, as our attribute dataset of all 1708 cellphone product pages on the platform. We merge the attribute dataset with the click-stream dataset.

Finally, the checkout page shows the payment interface, allowing consumers to finalize their purchase decisions seamlessly. We show the illustration of the checkout page in Figure 1.

We highlight several key features of the marketplace’s checkout page that support our analysis. First, the checkout process is highly streamlined. Consumers choose between delivery and in-store pickup, provide an address if they select delivery, and then select a payment method. All relevant information displayed during checkout, such as the product name and total pay-

[4]The original dataset labeled the type of clicks as ‘cart’, while we found that the website does not have a virtual shopping cart design within the data span.

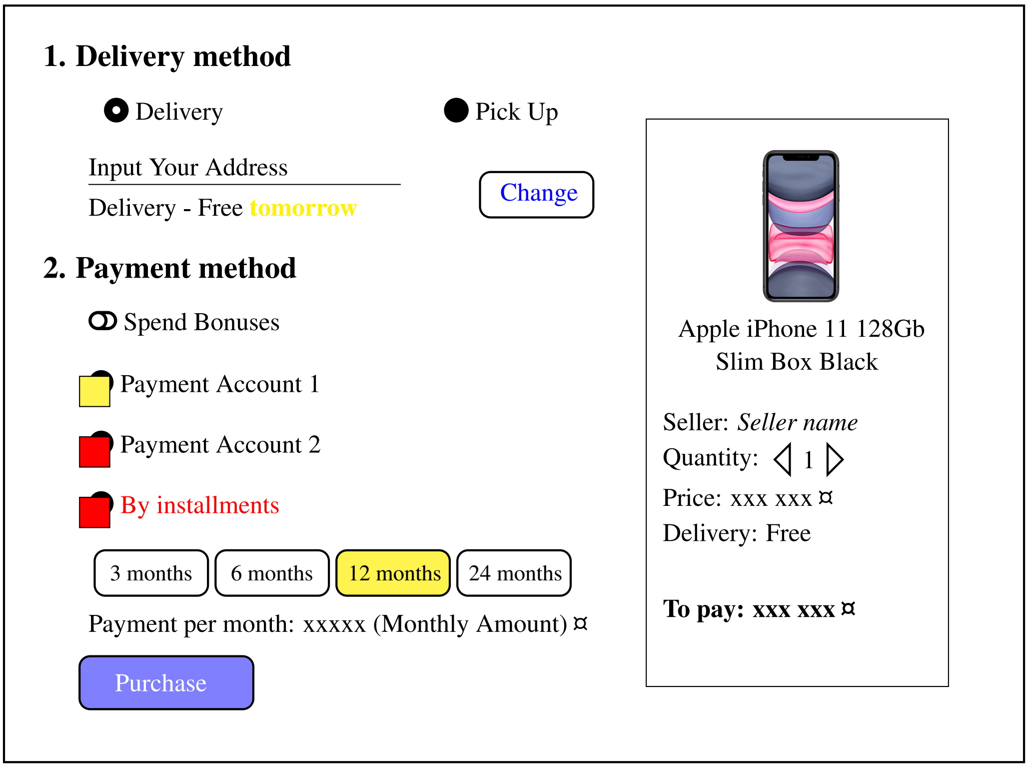

FIGURE 1 – Checkout Page Illustration

Notes: The figure shows the layout of a page with high similarity to the marketplace in our data.

ment amount (in lump-sum or installments), is already available on the list or product pages. After reviewing the details, consumers confirm the purchase with a single click, and the entire checkout process can typically take less than 30 seconds. Second, the platform does not use a virtual shopping cart. Consumers can check out only one product at a time, so the checkout decision is specific to a single product and unaffected by the other alternatives. Third, the checkout page presents no new information. Delivery times and fees are shown on the product page, and no additional attributes beyond those visible in the product name are introduced. The process involves no further steps or verification procedures. The platform is operated by a FinTech company and is the top-ranked online trading platform in the country, accounting for 70 percent of internet-based transactions. Around 60 percent of the national population holds an active account. Although an account is required for purchase, this has become a widely accepted norm. Consumer accounts are linked to bank-issued debit or credit cards, resulting in minimal friction at checkout and limited credit-related concerns. These features ensure a consistent and straightforward checkout experience, minimizing external influences and ruling out commonly cited factors for checkout abandonment.

2.2 Descriptive Evidence

The clicks a consumer makes over a seven-month period constitute the observed sequence for that consumer in the dataset. Our sample comprises 1,671,791 consumers, with summary statistics shown in Table 1 [5]. Using this sample, we conduct a descriptive analysis to explore consumers' checkout behavior and their potential roles in preference discovery.

TABLE 1 – Summary Statistics of the Descriptive Sample

Consumer stats	
No. of consumers	1,671,791
No. of consumers who do not checkout	1,074,018
No. of consumers who checkout but leave	218,971
No. of consumers who purchase	378,802
Steps stats	
No. of steps	13,984,172
No. of steps with checkout abandonment	528,474
No. of steps with purchase	378,802
Checkout abandonment rate	58.93%

In our sample, the consumer checkout abandonment rate reaches approximately 59%, with checkouts either not proceeding to payment or interrupted by consumers clicking on other products before completing payment. Previous research typically does not examine checkout abandonment independently but includes it within cart abandonment, for which survey-based estimates indicate a rate of 70% (Baymard Institute, 2024) [6]. Since our website does not include a shopping cart feature, and most reasons mentioned in surveys are rarely applicable to our setup, the persistently noticeable checkout abandonment rate requires a more plausible explanation.

We then turn to the preference change by comparing the list-page attributes of the checkout products and viewed products. We collect our main results in Figures 2 and 3.

Figure 2 shows the relationship between the average attributes of checkout products and the number of checkouts experienced. We focus on three continuous list-page attributes: price, RAM, and storage. In the top left sub-figure, we observe that, without interference from additional views, the price of products brought to checkout decreases as the number of checkouts increases (black line), a trend that remains robust across consumers with different total checkout numbers (red, blue and emerald lines). This indicates that, on average, consumers tend to

[5] See Appendix A for detailed data cleaning process.

[6] Although survey data on checkout abandonment rates is limited, industry experts suggest an ideal rate should be around 20% (Saleh and Shukairy, 2010).

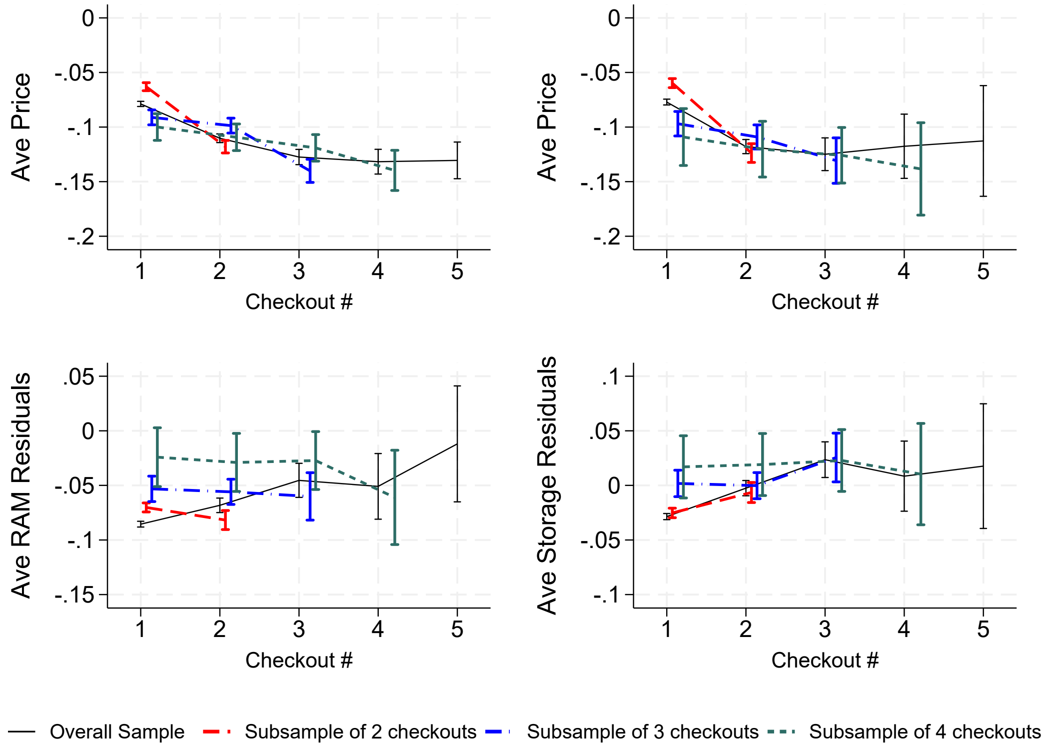

FIGURE 2 – Attributes of Checkout Products on Experienced Numbers of Checkouts
Notes: This graph shows the trends of average checkout product attributes as the times of checkout grows. The top left subfigure uses all consumers in Table 1, while the rest three subfigures use consumers who have viewed all checkout products before the first checkout.

take higher-priced products to checkout in earlier decisions, but after abandoning their checkout, they tend to check out products at a lower average price. To ensure that consumer choices at checkout are unaffected by those additional choices introduced by later views, we turn to a subsample that includes only those consumers who view all checkout products before their first checkout. We find a similar pattern in the top-right sub-figure. Our findings clearly show that consumers become more price-sensitive in each subsequent checkout. Notably, this trend does not appear in the residual plots for RAM and storage (bottom sub-figures), which suggests that among list-page attributes, adjustments in price sensitivity serve as the primary driver of preference discovery between checkouts.

Next, we turn to the deviation of product attributes to illustrate how search preference converges to purchase preference. To show this, following Bronnenberg et al. (2016), we use consumers who purchase and calculate the log deviation by taking the differences between the log attributes of the viewed products and the purchased product.

We plot the distribution of the log deviation by the decile position of the click in the search process (Appendix Figures B.1 and B.2). We observe a convergence on all attributes throughout the search process. The distribution of the log of relative prices is roughly symmetric around 0, showing large heterogeneity in consumers' search process. We use the absolute value of the log deviation to measure the extent of search deviation from the purchased product and regress it on the view and checkout dummies. We show the coefficient plots in Figure 3.

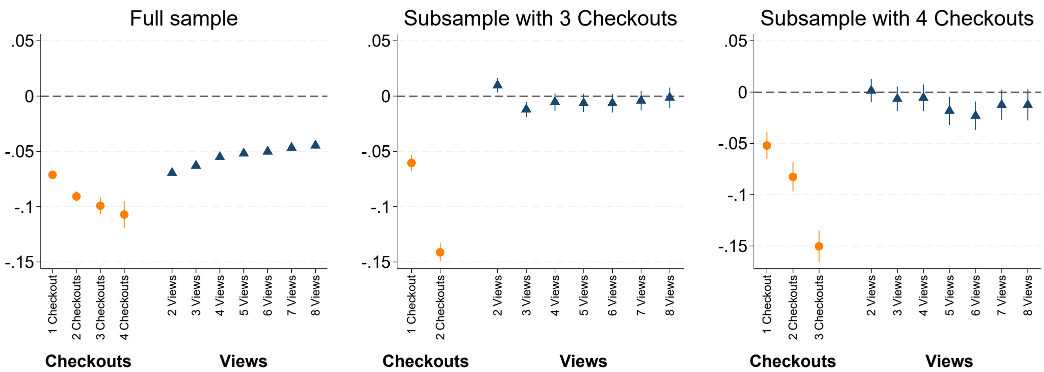

FIGURE 3 – View Step and Checkout Fixed Effects on Absolute Log Deviations
Notes: The regressions control dummies for the first 30 view steps.

We use the absolute log deviation of the first viewed product as the baseline. Our results show that as consumers experience more checkouts, they tend to view products with less deviated attributes. Specifically, the fixed effects of checkout steps on the absolute log deviations exhibit a downward trend, while the effects of view steps move in the opposite direction and are less pronounced (Subfigure 1). To account for potential preference heterogeneity across consumers with different total checkout counts, we replicate the analysis using only sequences with exactly three and four checkouts (Subfigures 2 and 3). In both cases, the convergence in product attributes is almost entirely driven by the checkout fixed effects, whereas the view fixed effects remain practically insignificant. Similar patterns hold across other list-page attributes, as shown in Appendix Table B.1.[7]

We observe in the data, consistent with Bronnenberg et al. (2016), that the attributes of viewed products exhibit a stable convergence toward the attributes of the purchased products, with no systematic drifts or sudden changes. The original study thus concludes that preference learning does not play a substantial role during search. However, our further analysis suggests

[7]These findings are robust to reversing the order of checkouts and inspection steps.

that such stability may mask the presence of actual preference discovery: attribute shifts are reflected in consumers' checkout choices, and the convergence path of inspected products is closely linked to checkout behavior. We insist that preference discovery does not occur uniformly throughout the search process but is instead concentrated at checkouts. Because the timing of checkout varies across search sequences, its aggregate effects are difficult to identify from the attribute patterns of viewed products. Our findings highlight the critical role of checkouts in preference discovery, suggesting that this process can be viewed as a single-dimensional one, in which consumers gradually learn their willingness to pay through repeated checkouts. This mechanism provides theoretical support for modeling preference discovery as a discrete updating process on perceived price sensitivity, which primarily occurs at the checkout stage.

3 Model

In this section, we propose a structural model that integrates preference discovery into sequential search.[8] Our model differs from the classic Weitzman-style sequential search model (Weitzman, 1979) in two key respects. First, consumers hold a belief about their price sensitivity rather than knowing it with certainty. Second, selection is provisional: after selecting a product, the consumer proceeds to the corresponding checkout page but not immediate payment, receives a signal about her true price sensitivity, updates her belief, and decides whether to purchase.

Hence, the model describes an alternating process of information acquisition and preference discovery. Information acquisition follows a standard sequential search process, where each selection leads to checkout. Preference discovery takes place at the checkout stage, where the consumer updates her belief and reconsiders her search decision based on the revised perception of price sensitivity. A purchase is observed if the consumer selects the same product again without further inspection, given the posterior belief.

3.1 The Information Acquisition Process

Consider a consumer i who intends to purchase a cellphone on our platform. Her shopping process begins on the cellphone list page, where she observes all available products, forming her

[8] We adopt a sequential search model because preference discovery is identified through changes in consumer decisions, which standard discrete-choice or simultaneous search models cannot capture. These models typically assume that search behavior is determined by perceived preference *a priori*.

choice set $\mathcal{M}_i$ with size $|\mathcal{M}_i|$. For each product $j \in \mathcal{M}_i$, the list page displays partial information—such as price, brand, RAM, and storage—referred to as *list-page attributes*. We assume these attributes are freely accessible to the consumer upon opening the list page. Additional attributes that are not fully displayed on the list page but appear on each product's detail page are referred to as *product-page attributes*.

Let x_j denote the non-price attributes shown on the list page, and let p_{ij} represent the displayed price of product j, which may vary across consumers due to differences in purchasing dates. We assume that consumer i does not have full knowledge of her true price sensitivity and introduce the concept of *perceived price sensitivity*, which governs how she evaluates prices in her decision-making. This perceived sensitivity varies across consumers and may evolve throughout the search process. When consumer i enters the market, her perceived price sensitivity is β_{i1}; after $r-1$ updates, it becomes β_{ir}.

Define *perceived utility* as the utility that consumer i expects to obtain from purchasing product j under full information, given the perceived price sensitivity β_{ir}. The utility function is specified as a linear combination of list-page attributes, expressed as:

$$u_{ijr} = \underbrace{\gamma^\top x_j + \beta_{ir} p_{ij}}_{v_{ijr}} + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma_\varepsilon^2), \quad j \in \mathcal{M}_i$$

Here, v_{ijr} denotes consumer i's valuation of the list-page attributes under the perceived price sensitivity β_{ir}. The term ε_{ij} represents a one-dimensional index capturing consumer i's unobserved idiosyncratic evaluation of the product-page attributes. Following prior empirical studies (e.g., Kim et al., 2010; Chen and Yao, 2017; Ursu, 2018), we model ε_{ij} as a consumer-product-specific match value drawn from an i.i.d. normal distribution with mean zero and standard deviation σ_ε. This distribution is assumed to be known to consumer i.

Given β_{ir}, the only component of u_{ijr} that is unknown to consumer i is ε_{ij}. To resolve this uncertainty and determine the perceived utility, consumer i must click into the product page of j to obtain its product-page information. This action is referred to as the consumer's *inspection* of product j. Through inspection, ε_{ij} is fully revealed, and the perceived utility u_{ijr} becomes known. Importantly, a product cannot be purchased until u_{ijr} is determined.

Inspection is modeled as a step-by-step sequential process, where the consumer must incur a search cost to obtain the full information associated with each product. This cost includes both objective factors (such as time and physical effort) and subjective ones (such as cognitive effort, attentional demands, patience, and the emotional stress in decision-making). We assume that consumer i knows the search cost prior to inspection, but this cost is unobservable to the researcher. Following Moraga-González et al. (2023) and Chung et al. (2024), we model the search cost as heterogeneous across consumer-product pairs and allow it to vary randomly when perceived price sensitivity is updated (the rationale is discussed in Section 3.3). These random search costs are assumed to follow an i.i.d. log-normal distribution.

$$\ln c_{ijr} \sim \mathcal{N}(\bar{c}, \sigma_c^2).$$

We assume that inspections are 1) exhaustive, inspection leaves no hidden information related to ε_{ij}; and 2) permanent, consumer i can recall the value of ε_{ij} without any additional cost.

Consumer i can stop inspecting additional products after any inspection. This stopping decision leads to one of two outcomes: either leaving the market completely, or proceeding to the checkout page with a selected product, referred to as a *checkout*. In the former case, the consumer selects the outside option ($j = 0$) and does not enter the checkout page. In the latter case, she initiates a checkout process. Checking out incurs minimal cost and does not provide additional information, implying that the consumer will only proceed to checkout if she genuinely intends to purchase the product.

We segment the entire search process into multiple stages, each bounded by a checkout. Each stage consists of a sequence of consecutive inspections followed by a checkout, and is defined as an *inspection round*. All actions in inspection round r are based on a fixed perceived price sensitivity β_{ir} and search costs c_{ijr}. In each inspection round, the consumer may inspect any number of previously uninspected products or choose not to inspect at all. At checkout, she may select from all products inspected in the current or any previous inspection round. [9]

[9] In the empirical analysis, we focus only on consumers who inspect at least one product in inspection round 1, as our data does not include those who browse the list page without clicking on any products.

3.2 Optimal Search within Inspection Rounds

Within an inspection round, the consumer follows a standard sequential search process. Assuming that ε_{ij} is independently and identically distributed across products, the optimal strategy is governed by the Optimal Search Rules of Weitzman (1979). To introduce these rules, we first define the value of an inspection. Suppose the consumer i has access to a fallback option that provides utility $\bar{v}$. Then, the expected benefit from inspecting another product j is given by:

$$H_{ijr}(\bar{v}) = \int_{\varepsilon_{ij} > \bar{v} - v_{ijr}} (\varepsilon_{ij} - (\bar{v} - v_{ijr}))\, dF(\varepsilon_{ij}) = (1 - F(\bar{v})) \cdot \mathbb{E}(\varepsilon_{ij} + v_{ijr} - \bar{v} \mid \varepsilon_{ij} + v_{ijr} > \bar{v})$$

with $F(\cdot)$ denoting the pdf of ε_{ij}. Hence, the consumer i inspects product j only if the expected gain $H_{ijr}(\bar{v})$ exceeds the search cost c_{ijr}. We assume consumer i is risk-neutral, meaning she is indifferent between resolving and reserving a product's uncertainty when expected benefits are identical. The *reservation value* z_{ijr} is defined as the value of $\bar{v}$ that equates the expected gain and search cost:

$$c_{ijr} = H_{ijr}(z_{ijr})$$

Following Appendix 1 in Kim et al. (2010), z_{ijr} has a linear specification:

$$z_{ijr} = v_{ijr} + \sigma_\varepsilon \cdot m\left(\frac{c_{ijr}}{\sigma_\varepsilon}\right) = v_{ijr} + \sigma_\varepsilon \cdot m_\varepsilon\left(c_{ijr}\right),$$
$$\text{where } m^{-1}(\eta) = (1 - \Phi(\eta))\left[\frac{\phi(\eta)}{1 - \Phi(\eta)} - \eta\right].$$

where $\phi(.)$ and $\Phi(.)$ the pdf and cdf of the Gaussian distribution.[10] The reservation value determines whether consumer i inspects a product given an alternative value, and can be interpreted as the equivalent value of an inspection. The function $m_\varepsilon(c_{ijr})$ is a strictly decreasing bijection that maps the search cost to the value of an option with unresolved ε_{ij}, capturing the informational rent due to product-level uncertainty. This value depends only on c_{ijr} and the distribution of ε_{ij}, both of which are assumed to be known to the consumer. Therefore, given β_{ir} and c_{ijr}, consumer i knows the reservation values of all products at the beginning of each inspection

[10] For the detailed derivation of z_{ijr}, see Appendix C.

round, referred to as the perceived reservation values in inspection round r.

Suppose consumer i undergoes R inspection rounds and inspects a total of J_R products. Following the inspection order, we index the rounds from 1 to R and the inspected products from 1 to J_R. Uninspected products are indexed by $k \in \{J_R+1, J_R+2, \ldots, |\mathcal{M}i|\}$. In each round r, the checkout product is denoted by h_r, and the last inspected product by J_r. Given the perceived utilities and reservation values in round r, we apply the result of Weitzman (1979) to characterize consumer i's optimal search decisions using the following rules.[11]

1. Optimal Ranking: Inspect products in decreasing order of perceived reservation values:

$$z_{i,J_{r-1}+1,r} \geq z_{i,J_{r-1}+2,r} \geq \ldots \geq z_{i,J_r,r} > \max_{k>J_r}\{z_{i,k,r}\}.$$

2. Optimal Continuing: Continue searching when the maximum perceived reservation value exceeds the perceived utility of any inspected product:

$$z_{ijr} \geq \max_{\ell=0}^{j-1}\{u_{i\ell r}\}, \quad \forall j \in \{J_{r-1}+1, J_{r-1}+2, \cdots, J_r\}.$$

3. Optimal Stopping: Stop searching when the maximum perceived utility of inspected products exceeds the perceived reservation value of any uninspected product:

$$\max_{j\leq J_r}\{u_{ijr}\} \geq \max_{k\geq J_r+1}\{z_{ikr}\}.$$

4. Optimal Selecting (Checkout): When search stops, check out the product with the maximum perceived utility among inspected products.

$$u_{i,h_r,r} \geq \max_{j=1, j\neq h_r}^{J_r}\{u_{ijr}\}.$$

[11]These optimal search rules apply when search in inspection round r is independent of other inspection rounds, that is, when βir and c_{ijr} are held as given.

3.3 Preference Discovery between Inspection Rounds

This subsection explains how the perceived price sensitivity β_{ir} is determined in each inspection round. We assume that, in round r, consumer i holds a prior belief about their true price sensitivity, modeled as a normally-distributed belief $\mathcal{N}(\beta_{ir}, \omega_r^2)$.

We assume that consumer i bases her search decisions on the mean of her prior belief β_{ir} under two key premises. First, the consumer is risk-neutral with respect to information uncertainty (consistent with our reservation value setup), so the strength of her belief does not affect her decisions. Second, consumers are myopic, meaning they do not anticipate preference discovery or search cost changes across inspection rounds. Although the second assumption may appear strong, we offer several justifications. First, our reduced-form analysis indicates that preference discovery occurs primarily around checkouts. If consumers were forward-looking, it would be unclear why belief updating would be triggered upon entering the checkout page. Second, forward-looking consumers should attempt to identify their preferences as early as possible (Dzyabura and Hauser, 2019). If preference discovery requires entering checkout, consumers would be expected to initiate checkouts frequently in the early stages of search. However, data shows that checkout events tend to occur more frequently in the later part of the search process. Third, theoretically, preference discovery is triggered by the anticipated consequences of decision-making (Cao and Zhang, 2021). Since the consequences of a purchase are neither realized nor observable in inspections, it is less likely to assume forward-looking learning when purchase intentions are not explicitly involved.[12]

Consumer i's initial prior mean is determined by:

$$\beta_{i1} \sim \mathcal{N}(\beta_i + \delta, \sigma_0^2), \quad \text{where } \beta_i \sim \mathcal{N}(\bar{\beta}, \sigma_{\bar{\beta}}^2),$$

Here, β_i represents consumer i's true price sensitivity, $\bar{\beta}$ is the overall mean of consumers' true price sensitivity in the sample, δ denotes the population-level deviation between consumers'

[12] From a modeling perspective, solving the optimal solution of a forward-looking process that jointly incorporates information acquisition, preference discovery, and purchase decisions would be highly complex, subject to the curse of dimensionality, difficult to identify, and unlikely to be supported by available data. To our knowledge, we are aware of no simplification making this problem tractable, nor do we expect consumers to solve such a cognitively demanding optimization in their online shopping.

prior mean and their true price sensitivity, and σ_0^2 is the variance of the prior mean, conditional on the true price sensitivity and population deviation. In this setup, we assume that the perception deviation is consistent across consumers, while each consumer has a unique true price sensitivity and starts their search from a different position relative to the deviated preference. The parameters δ and σ_0 jointly capture the extent of preference discovery: when both equal zero, no preference discovery occurs; when δ is 0 but σ_0 is positive, the average perceived price sensitivity remains stable, yet discovery still takes place at the individual level.

Each time consumer i enters the checkout page, she receives a signal about her true price sensitivity, drawn from a normal distribution with mean β_i.

$$\beta_i^s \sim \mathcal{N}(\beta_i, \sigma_s^2)$$

Here, σ_s^2 denotes the variance of the perceived signal, which is assumed to be known to consumers. It reflects the precision with which consumer i perceives her true price sensitivity—a smaller σ_s^2 implies a more accurate signal. Following Bayesian updating, when consumer i reaches the checkout page, her belief is updated from a prior belief $\mathcal{N}(\beta_{ir}, \omega_r^2)$ to a posterior $\mathcal{N}(\beta_{i,r+1}, \omega_{r+1}^2)$, with the updated parameters given by:

$$\begin{aligned}
\beta_{i,r+1} &= \beta_{i,r} + \frac{\tau_r^2}{1+\tau_r^2}(\beta_i^s - \beta_{i,r}), \\
\omega_{r+1}^2 &= \omega_r^2 \cdot \frac{1}{1+\tau_r^2} = \omega_1^2 \cdot \frac{1}{1+\tau_1^2 \cdot (r-1)}, \\
\tau_r &= \frac{\omega_r}{\sigma_s} = \frac{\tau_1}{\sqrt{1+\tau_1^2 \cdot (r-1)}}.
\end{aligned}$$

Here, τ_r is the ratio of the prior belief's standard deviation to that of the signal, reflecting how much consumers trust the signal over their prior beliefs. It is also an indicator the consumer's learning speed: as τ_r approaches infinity, consumers fully trust the signal received; when $\tau_r = 0$, no learning occurs. For any $\tau_1 > 0$, the consumer's perceived price sensitivity converges to her true price sensitivity β_i with the variance approaches zero. At the sample level, perceived price sensitivity converges to $\bar{\beta}$, and the variance to σ_β^2.

We conclude by explaining why consumer i is assumed to redraw her search costs at the checkout stage. This is because these costs include subjective emotional components, which often become unstable upon entering checkout. For instance, the consumer may grow more concerned about missing uninspected products or begin to doubt whether she is truly ready to make a purchase. In such cases, search costs may fluctuate due to behavioral shocks unrelated to product information. We model these shocks as stochastic components within c_{ijr}, allowing us to distinguish abandonment driven by preference discovery from that caused by exogenous randomness. In contrast, we assume the utility term revealed during inspection (i.e., ε_{ij}) remains constant throughout the search process, since it depends on product-page information that does not change and is not explicitly presented on the checkout page.

3.4 The Full Model

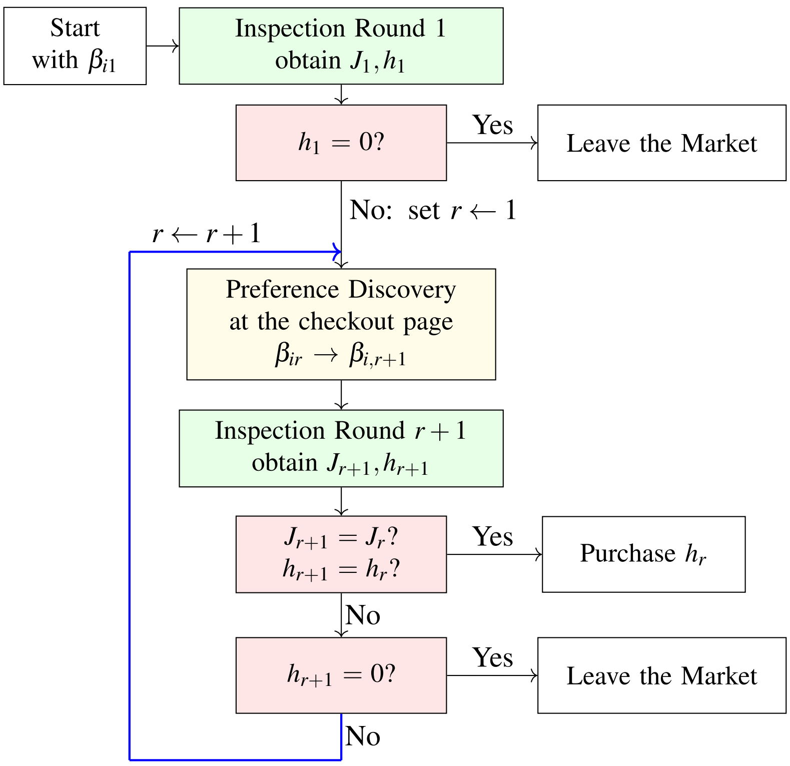

FIGURE 4 – Search Process with Preference Discovery

We present the full model in Figure 4. After entering the market, consumer i begins her first inspection round and makes decisions based on her perceived price sensitivity β_{i1}. When she decides to stop inspecting, she either exits the market ($h_1 = 0$) or selects an inspected product to proceed to the checkout stage ($h_1 \in \{1, 2, \ldots, J_1\}$). In the latter case, she receives a signal on the checkout page, updates her perceived price sensitivity to β_{i2}, and redraws search costs c_{ij2}

for all uninspected products. She then enters the next round and reconsiders her prior decisions. If she inspects other products or changes her checkout selection, she exits the current checkout page and clicks on another product. This behavior is interpreted as checkout abandonment.

The entire search process takes the form of an iterative cycle, in which information acquisition within each inspection round alternates with preference discovery between inspection rounds. This process continues until the consumer no longer enters a new checkout page in an inspection round. This occurs in one of two actions:

1. Leave ($h_R = 0$): The consumer exits the market in inspection round R without entering the checkout page. As a result, no further preference discovery occurs, and the search terminates without a purchase.

2. Purchase ($J_R = J_{R-1}$ and $h_R = h_{R-1}$): After preference discovery, the consumer decides to purchase the product checked out in round $R-1$. This implies that in round R: (1) she does not inspect any uninspected products; (2) she confirms the checkout choice made in round $R-1$. These two conditions jointly ensure that she remains on the checkout page of product h_{R-1} and completes the purchase without receiving any new preference signals.

The model concludes when either of these two conditions is met.

The outside option is the last component to be specified. We assume a round-dependent outside option value, which takes the following functional form:

$$u_{ir}^{outside} = \begin{cases} -\infty & \text{if } r = 1; \\ \mu_0^{outside} + \log(1 + r \cdot \xi^{outside}) + \varepsilon_i^{outside} & \text{if } r > 1. \end{cases}$$

Consumers are assumed to know the outside option value at the beginning of each inspection round. We further assume they are myopic with respect to its evolution across rounds, as prolonging search in expectation of a better outside option is not economically rational. The specification of the outside option is motivated by two key observations. First, 77% of customers in our data never complete checkout. Their search intent is unclear, and many may be casual browsers or window-shoppers. Including such consumers in estimation could confound our results. Since they never reach the checkout page, they are also irrelevant to our analysis of preference discovery around checkout. We therefore exclude them from estimation and set the outside option value to negative infinity before the first checkout, ensuring that all sampled consumers complete at least one checkout. Second, we assume the outside option value increases with the number of inspection rounds. Beyond the high exit rate in the first round, over 25% of consumers exit in each subsequent round. These exits are unlikely to be driven by product information or perceived price sensitivity alone, suggesting a nontrivial role for external factors. For example, 60% of U.S. consumers compare prices across platforms (van Gelder, 2023), and others may face time constraints (Greminger, 2024) or search fatigue (Ursu et al., 2023). Although these factors are not directly observed in our data, their influence of prompting leaving likely increases over time. To capture this, we model the outside option value as increasing across inspection rounds to better align with the observed exit behavior.

Our model provides flexible explanations for consumer behavior on the checkout page. After abandoning checkout, consumers may proceed with a previously inspected product, inspect a new one, or exit the market. For the latter two cases, our model considers alternative motivations. Consumers may inspect further due to changes in subjective search costs or exit due to a higher outside option value. However, if a consumer switches to another previously inspected product, we assume checkout abandonment is entirely driven by preference discovery, as she already has information on both products at the time of checkout and abandonment.

4 Estimation

This section presents our estimation strategy. We implement a simulated maximum likelihood approach, a method commonly used in the literature for estimating sequential search models.

4.1 Likelihood Contributions

We begin by constructing the likelihood function of the model. Specifically, we first derive the conditional likelihood of sequential search within each inspection round, given β_{ir}, c_{ijr}, and $u_{ir}^{outside}$, and then incorporate the Bayesian updating of perceived preferences across rounds to establish the unconditional likelihood.

The conditional likelihood is defined as the joint probability that the observed sequence of inspections and checkout decisions is subject to the optimal search rules described in Section 3.2. However, directly computing this probability leads to challenges in computation. To address this, we follow Proposition 1 in Zhang (2025) and recast the within-round sequential search as an equivalent partial ranking over action values. We have the following proposition:

Proposition 1. *Define:*

$$y_{ir} = \begin{cases} \min\{u_{i,h_r,r}, z_{i,J_r,r}\}, & \text{if } r = 1; \\ \min\{u_{i,h_r,r}, z_{i,J_r,r}\} \cdot I(J_r > J_{r-1}) + u_{i,h_r,r} \cdot I(J_r = J_{r-1}), & \text{if } r > 1. \end{cases}$$

Denote $J_0 = 0$ for symbolic simplification. Weitzman's Optimal Search Rules hold if and only if the following conditions are fulfilled:

1. *Distribution Condition: $u_{i,h_r,r} \leq z_{i,J_r,r}$ if $h_r < J_r$ and $J_r > J_{r-1}$ for all r ;*
2. *Rank Condition: $z_{i,J_{r-1}+1,r} \geq z_{i,J_{r-1}+2,r} \geq \ldots \geq z_{i,J_r,r}$ for all r ;*
3. *Inspection Choice Condition: $z_{ikr} \leq y_{ir}$ for all $k > J_r$;*
4. *Purchase Choice Condition: $u_{ijr} \leq y_{ir}$ for all $j \leq J_r$ and $j \neq h_r$.*

Zhang (2025) demonstrates that any optimal sequential search process satisfying the Weitzman assumptions can be recast as a single-stage partial ranking model, whose features are fully characterized by the conditions stated in Proposition 1 and yield the same probability as the original sequential search process. Based on this result, we express the conditional joint probability of the search decisions within inspection round r as follows:

$$\begin{aligned}
&\text{Prob}(\text{Decisions in round } r \mid u_{i,h_r,r}, \Theta_1, \beta_{ir}, \vec{p}_i, \vec{x}) \\
= &\underbrace{\Pr(z_{i,J_r,r}) \geq u_{i,h_r,r} \mid u_{i,h_r,r}, \Theta_1, \beta_{ir}, p_{i,J_r}, x_{i,J_r})^{I(J_r > h_r) \cdot I(J_r > J_{r-1})}}_{\text{Distribution Condition}} \\
&\cdot \underbrace{\prod_{j=J_{r-1}+1}^{J_r - 1} \Pr(z_{ijr} \geq z_{i,j+1,r} \mid z_{i,j+1,r}, \beta_{ir}, \Theta_1, p_{ij}, x_j)}_{\text{Rank Condition}} \\
&\cdot \underbrace{\prod_{j=1, j \neq h_r}^{J_r} \Pr(u_{ijr} \leq y_{ir} \mid y_{ir}, \beta_{ir}, \Theta_1, p_{ij}, x_j)}_{\text{Purchase Choice Condition}} \cdot \underbrace{\prod_{k=J_r+1}^{|\mathcal{M}_i|} \Pr(z_{ikr} \leq y_{ir} \mid y_{ir}, \beta_{ir}, \Theta_1, p_{ik}, x_k)}_{\text{Inspection Choice Condition}}
\end{aligned}$$

Here $\Theta_1 = \{\mu^{outside}, \xi^{outside}, \gamma, \bar{c}, \sigma_c, \sigma_\varepsilon\}$ represent the parameters that determine consumers' search and checkout decisions in an inspection round apart from the perceived price sensitivity.

In the above expression, inspection and purchase decisions within each inspection round are transformed into direct or indirect ordinal relationships between perceived utilities or reservation values and $u_{i,h_r,r}$. Since the checkout product may differ across rounds, we exclude the perceived utilities of products checked out in other rounds from the ordering conditions in each round. Due to the conditional independence inherent in partial ranking structures, this exclusion does not affect the relative ordering among the remaining values. Define the set of perceived price sensitivity and match values that determines checkout products' perceived utilities in all rounds by $\Psi_i = \{\varepsilon_{i,h_1}, \cdots \varepsilon_{i,h_R}, \beta_{i1}, \cdots, \beta_{iR}\}$, the likelihood of consumer i's inspection and checkout decisions within all inspection rounds is given by:

$$\begin{aligned}
\mathcal{L}_{i,search} = & \prod_{r=1}^{R} \Pr(m_\varepsilon(c_{i,J_r,r}) \geq u_{i,h_r,r} - v_{i,J_r,r} \mid \Psi_i, \Theta_1, \vec{p}_i, \vec{x})^{I(J_r > h_r) \cdot I(J_r > J_{r-1})} \\
& \cdot \prod_{r=1}^{R} \prod_{j=J_{r-1}+1}^{J_r - 1} \Pr(m_\varepsilon(c_{ijr}) \geq z_{i,j+1,r} - v_{ijr} \mid z_{i,j+1,r}, \Psi_i, \Theta_1, \vec{p}_i, \vec{x}) \\
& \cdot \prod_{j=1, j \notin \{h_1, h_2, \cdots, h_R\}}^{J_R} \Pr(\varepsilon_{ij} \leq \min_{r \in \{t_j, t_j+1, \cdots R\}} y_{ir} - v_{ijr} \mid y_{ir}, \Psi_i, \Theta_1, \vec{p}_i, \vec{x}) \\
& \cdot \prod_{r=1}^{R} \prod_{k=J_r+1}^{|\mathcal{M}_i|} \Pr(m_\varepsilon(c_{ijr}) \leq y_{ir} - v_{ikr} \mid y_{ir}, \Psi_i, \Theta_1, \vec{p}_i, \vec{x})
\end{aligned}$$

where t_j indicates the inspection round in which product j is inspected.

Consumer i's checkout decisions reveal the relative perceived utilities of the checked-out products and the associated shifts in perceived price sensitivity. A product may be preferred to another under the prior but inferior under the posterior. The observed checkout sequence requires that each product checked out must exhibit higher perceived utility than all previously inspected and checked-out alternatives. Define $\mathcal{I}_i$ as the family of Ψ_i that ensures $ui, h_r, r \geq u_{i,h_{r'},r}$ for all r and $r' \in 1, 2, \ldots, t_{h_r}$, the probability that consumer i sequentially checks out $\{h_1, h_2, \ldots, h_R\}$ across inspection rounds $\{1, 2, \ldots, R\}$ is given by:

$$\mathcal{L}_{i,discovery} = \Pr(\Psi_i \in \mathcal{I}_i \mid p_{i,h_1}, \cdots, p_{i,h_R}, x_{h_1}, \cdots, x_{h_R}, \Theta_2)$$

where $\Theta_2 = \{\bar{\beta}, \delta, \tau_1, \sigma_{\bar{\beta}}, \sigma_s\}$ are the determinant parameters for preference discovery. The overall parameter set to be determined is $\Gamma = \{\mu^{outside}, \xi^{outside}, \gamma, \bar{c}, \sigma_c, \sigma_\varepsilon, \bar{\beta}, \delta, \tau_1, \sigma_{\bar{\beta}}, \sigma_s\}$.

The overall likelihood of consumer i's search process is therefore:

$$\mathcal{L}_i(\Gamma; \vec{p}_{ij}, \vec{x}_j) = \mathcal{L}_{i,search} \cdot \mathcal{L}_{i,discovery}$$

It is important to note that $\mathcal{L}_{i,discovery}$ may be subject to a zero-probability issue. To ensure that $\mathcal{L}_{i,discovery} > 0$, it may be necessary to impose additional restrictions on the domain of β_{ir}. For instance, consider the case where consumer i undergoes two inspection rounds, 1 and 2, and checks out two distinct products, h_1 and h_2, both inspected in round 1. In this scenario, the perceived utilities of the checkout products must satisfy the following conditions:

$$\begin{cases} u_{i,h_1,1} \geq u_{i,h_2,1} \Rightarrow \varepsilon_{ih_2} - \varepsilon_{ih_1} \leq x_{h_1}\gamma + p_{i1}\beta_{i1} - x_{h_2}\gamma - p_{i2}\beta_{i1} \\ u_{i,h_2,2} \geq u_{i,h_1,2} \Rightarrow \varepsilon_{ih_2} - \varepsilon_{ih_1} \geq x_{h_1}\gamma + p_{i1}\beta_{i2} - x_{h_2}\gamma - p_{i2}\beta_{i2} \end{cases}$$

The inequality conditions above hold only if $(\beta_{i1} - \beta_{i2})(p_{ih_1} - p_{ih_2}) > 0$. The intuition is that, if a consumer, knowing two products, abandons her checkout with the more expensive one and turns to the cheaper, she must be more price sensitive after preference discovery at the checkout.

These domain constraints are common in our implementation and impose nontrivial limitations on the model. Our assumption that preference discovery occurs only in perceived price sensitivity across inspection rounds contributes to the model's traceability. However, the setup of a unique source of variation suffers from the curse of dimensionality. In an extreme case where consumer i undergoes R inspection rounds and checks out R distinct products all inspected in the first round, the implied sequence generates at least $(R-1) \cdot R$ inequality constraints on perceived utilities. By contrast, Ψ_i contains only R match values (ε_{i,h_r}) and R price sensitivity parameters (β_{ir}), yielding just $2 \cdot R$ degrees of freedom. As R increases, the number of constraints exceeds the dimensionality of Ψ_i, resulting in excessive restrictions. Nevertheless, for $R \leq 3$, all relevant constraints can be exhaustively enumerated, covering over 99.3% of inspections and 97.6% of checkouts in the data. Appendix D lists the full set of inequality constraints for a consumer's first three checkout decisions and details the implementation of $\mathcal{L}_{i,discovery}$ in each case.

Inspections do not directly constrain the range of perceived price sensitivity. Although abandoning a more expensive checkout to inspect a cheaper product may suggest increased price sensitivity, such behavior may also reflect a lower search cost drawn in the subsequent round. This highlights the importance of allowing search costs to vary across rounds. If perceived utility and reservation values depended solely on price, all products inspected in a new round would need to be uniformly more expensive or cheaper than the abandoned checkout. This pattern is inconsistent with the data, indicating the need for an additional source of variation to allow products in different price levels to be inspected following a checkout abandonment.

4.2 Identification

Our model's identification strategy consists of two parts. We first identify search parameters. The round-specific price sensitivity is identified with choices in each inspection round, while search cost and parameters for other list-page attributes are identified with inspection and checkout decisions throughout the search process. Then, parameters for the normal Bayesian preference discovery are identified through the variation of price sensitivity across rounds.

Prior research on sequential search often employs illustrative explanations to clarify how usual moment conditions in the data are related to model parameters. We adopt a similar informal approach in the main text to maintain accessibility. For readers seeking a more rigorous treatment, Appendix F presents a formal identification analysis, addressing both sequential search and preference discovery. While studies such as Morozov et al. (2021) and Ursu et al. (2023) develop identification arguments based on subsets of decisions, our formal approach leverages the full set of decision-making in consumers' within-round search processes. We believe the informal discussion in the main text offers sufficient clarity for applied purposes, while the formal analysis provides theoretical insights into identification of a general class of models.

4.2.1 Search Parameters Identified Within Inspection Rounds

Consider the perceived utilities and reservation values of product j to consumer i:

$$
\begin{aligned}
z_{ijr} &= x_j\gamma + p_{ij}\beta_{ir} + m_{\varepsilon}\left(c_{ijr}\right) \\
u_{ijr} &= x_j\gamma + p_{ij}\beta_{ir} + \varepsilon_{ij}
\end{aligned}
$$

As pointed out by Ursu et al. (2023) and Chung et al. (2024), the identification of σ_ε is theoretically feasible but empirically fragile.[13] To simplify estimation, we follow the existing literature and normalize $\sigma_\varepsilon = 1$.

It is important to emphasize that this assumption is not innocuous. Since the function $m_\varepsilon(c_{ijr})$ is not invariant to the scale of ε_{ij}, the assumed distribution of ε_{ij} has a substantial impact on the estimation of search costs (Zhang, 2025). Therefore, all search cost estimates are conditional on the assumed distribution of ε_{ij}, and the estimates should not be used directly for monetization. Nevertheless, this assumption does not affect the scale of preference and preference discovery parameters (see Chung et al., 2024; Greminger, 2024; Zhang, 2025). Hence, it remains valid for studying the process of preference discovery.

Under the optimal search rules and Proposition 1, products inspected earlier are likely to have higher perceived reservation values, while those with higher perceived utilities are more likely to be checked out. Accordingly, the position of a product in the search sequence and its checkout frequency help identify the attribute preference parameter γ and the average perceived price sensitivity $\bar{\beta}_{ir}$. The heterogeneity in β_{ir} is identified through variation in consumers' search paths. For instance, consider two consumers who inspect the same set of products and purchase the same one, but one begins with a more expensive option while the other starts with a cheaper one. This pattern suggests that the former is more likely to exhibit lower price sensitivity.

The log mean of the search cost $\bar{c}$ is identified by the length of inspection rounds. An inspection round ends when the inspected product with the highest perceived utility exceeds the perceived reservation values of uninspected products. The decreasing monotonicity of $m_\varepsilon(c)$ implies that lower search costs result in higher perceived reservation values, making consumers more inclined to inspect unknown products and delay the checkout decision.

4.2.2 Discovery Parameters Identified Across Inspection Rounds

Now, consider the parameters for across inspection rounds. These include the parameters governing the preference discovery process and those explaining variation in consumer decision-

[13] This difficulty arises from the properties of $m_\varepsilon(c_{ijr})$. The three parameters $\bar{c}$, σ_c, and σ_ε jointly determine its distribution, including both the mean and the variance. Although σ_c can be identified from variation across rounds, allowing for the theoretical identification of the remaining two parameters, the distributional shape of $m_\varepsilon(c_{ijr})$ also depends nonlinearly on these parameters. As a result, empirical identification is difficult to achieve. **?** proposes a Bayesian MCMC estimation approach that may help address this issue.

making independent of price disutility effects. Their identification relies on parameters identified within inspection rounds.

The standard deviation of search costs (σ_c) is identified by the frequency with which consumers abandon the checkout product and continue searching. When $\sigma_c = 0$, search costs have no variation, meaning all inspections in rounds $r > 1$ are driven purely by changes in price sensitivity. A positive σ_c introduces variability in perceived reservation values, leading to inspections independent of preference discovery. As σ_c increases, consumers are more likely to abandon checkout and extend their search process.

The mean true price sensitivity ($\bar{\beta}$), the uniform deviation in perceived sensitivity (δ), and the initial convergence rate (τ_1) are identified through the intercept, slope, and dispersion in prior means. A larger $|\bar{\beta}|$ indicates stronger price sensitivity throughout. The path of convergence is shaped jointly by δ and τ_1: both high $|\delta|$ and high τ_1 generate notable inter-round differences. However, while a high τ_1 implies rapid adjustment after the first checkout, leaving little change in later rounds, a low τ_1 implies gradual convergence. Consequently, differences between rounds 1 and 2 are more sensitive to τ_1, whereas large $|\delta|$ preserves cross-round variation regardless of learning speed.

The standard deviation of true sensitivities (σ_β), the cross-consumer dispersion in prior means (σ_0), and the standard deviation in preference signals (σ_s) are identified by the intercept, slope, and variation in prior variances. A larger σ_β increases prior variance uniformly across rounds, as preference learning does not eliminate true heterogeneity. In contrast, both a higher σ_0 and a lower σ_s steepen the decline in variance. A low σ_s indicates stronger belief updating in round 2, while a high σ_0 reflects more diffuse initial beliefs and thus slower convergence.

We show how the prior preferences that consumers act on in each inspection round in Figure 5. We can also see how the variation in preference discovery parameters δ, τ, σ_0 and σ_s influences the convergence curves of the preference belief. In principle, parameters in our preference discovery model are identified with the sequence data of three or more inspection rounds.

4.3 Monte Carlo Simulation

We report the Monte Carlo simulation results in Appendix Table E.2, which validate the proposed estimation method. A pseudo-sample of 50,000 consumers is generated, where each

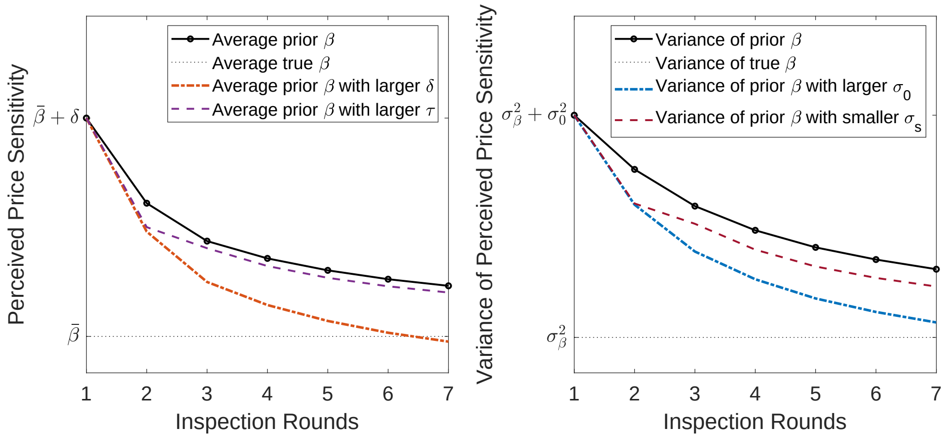

FIGURE 5 – Convergence of Perceived Price Sensitivity across Inspection Rounds

Notes: The two figures show the convergence of consumers' perceived price sensitivity in both mean and variance levels. We show the identification strategy with minor variable changes in the convergence route. In these changes, we keep $\bar{\beta}+\delta$ ($\sigma_\beta^2+\sigma_0^2$) in the left (right) subfigure constant and matched the changes in δ and τ (σ_0 and σ_s) to make their variation in the second inspection rounds similar. We see different convergence routes in the later inspection rounds.

product is characterized by three binary list-page attributes. These attributes yield eight distinct combinations, corresponding to eight different products. Including an outside option available from the second inspection round onward, consumers make search, checkout, and purchase decisions among nine alternatives. We implement the estimator as described in Appendix D. The estimates in Column 2 of Appendix Table E.2 closely match the true parameter values in Column 1, confirming the validity of our identification strategy and estimation procedure.

4.4 Estimation Sample

To reduce computational burden, we further restrict the estimation sample using the dataset described in Section 2.2. We aggregate cellphone SKUs across colors and specifications, yielding 275 distinct models. Among these, the 20 best-selling models (73% of total market sales) are treated as individual products. For each model, list-page attributes are defined based on the highest-selling specification. The remaining 255 models are grouped into three composite products based on price tiers. Markets are segmented by purchase week. In each week, the consumer's choice set includes all products viewed at least once. Weekly prices for each model are computed as the average of all recorded views in that week. Models unavailable in a given week are excluded from that week's choice set. As a result, consumers purchasing in different weeks face menus of 21 to 23 alternatives with week-specific pricing.

We consider only first views as inspections. Our model assumes that consumers are unaware of product-page information before inspection and fully informed when making checkout decisions. Thus, we treat all revisits to inspected products as free recalls of already-known information. Consequently, abandoning a checkout to revisit an inspected product is not counted as abandonment, as it is viewed as a recall of information. We record a checkout as abandonment only if a subsequent checkout to an inspected product is observed.

Since we focus on consumers' preference variation before and after their checkouts to investigate the preference discovery, we exclude consumers who only view products but do not check out or purchase [14]. The final sample includes 595,968 consumers. Table 2 describes the main statistics and price patterns around the checkouts.

TABLE 2 – Descriptive Statistics of the Model Estimation Sample

Panel A: Search Behavior (Sequence Level)

	All	Buyers	Non-buyers
Number of consumers	595,968	377,644 (63%)	218,324 (37 %)
Average number of products inspected	3.584	3.473	3.777
Average number of checkouts	1.301	1.322	1.265

Panel B: Search Behavior (Round Level)

	Round	All	Buyers	Non-buyers
No. of inspections in round	1	2.923	3.041	2.720
	2	0.570	0.367	0.923
	3	0.314	0.212	0.535
	>3	0.346	0.234	0.569
No. of consumers ending search in round	2	458,396	283,827	174,569
	3	106,226	72,912	33,314
	4	22,972	15,377	7,595

Panel C: Price Trends Across Inspection Rounds

	Round	All	Buyers	Non-buyers
Price of product inspected	1	390.1	390.9	388.5
	2	401.2	384.3	412.8
	3	389.1	373.2	402.6
Price of checkout products	1	379.2	375.6	385.5
	2	373.4	374.9	362.6
	3	359.7	360.7	352.6
	4	346.5	346.0	349.8

We first note that, conditional on checking out at least once, the non-purchase rate is 36.6%, compared to 77.3% in the sample used in Section 2.2. This indicates that consumers in our

[14] In other words, all consumers in our sample experience at least two inspection rounds

estimation sample are generally more purchase-oriented rather than engaging in mere browsing.

Turning to consumers' inspection and checkout behavior (see Panel A), consumers inspect, on average, only 3.6 models in the choice set of the 18-20 individual products and 3 synthetic products. Consumers check out multiple times: conditional on checking out at least once, each consumer has an average of 1.3 checkouts.

Panel B presents patterns in consumers' inspection behavior. Most inspections occur in the first round. After the first checkout, consumers who exit the market inspect more products than those who proceed to purchase. The checkout abandonment rate remains high across rounds: only 283,827 consumers (47.6% of all) purchase after their first checkout. Additionally, 137,572 consumers (17.8%) pursue at least two checkouts, a behavior that cannot be attributed solely to the additional information available at checkout. The ratio of purchasers to leavers remains relatively stable across the first three checkouts.

Finally, Panel C presents statistics on the prices of products inspected and checked out. Compared to the mixed price trend of inspected products, we observe a downward trend in the prices of products taken to checkout for both buyers and non-buyers, as seen in Section 2.2. This trend suggests that consumers display increasing price sensitivity in checkout decisions as the number of checkouts grows, while inspection decisions cannot be explained solely by preference changes.

Due to the large data scale, estimating our model on the entire sample is computationally demanding and unnecessary. Therefore, we randomly select a subsample of 50,000 consumers for estimation. As detailed in Section 4.1, we use the first three inspection rounds for estimation, using subsequent rounds as a reference for prediction accuracy. This limitation has minimal impact on the estimates, as it affects only 5.2% of the sample, and most of the inspections of the affected consumers occur within the first three rounds.

5 Estimation Results

Our estimation results of the model can be found in Table 3. Columns 1 and 2 are the main estimation results of our model. The estimated parameters include constant attribute-level preferences, discovery parameters for the changing price sensitivity, and parameters for search costs.

TABLE 3 – Parameter Estimates

	(1) Pref Discovery		(2) Weitzman	
Utility Parameters				
γ:				
RAM (1 GB)	0.0796	(0.0009)	0.0413	(0.0019)
Storage (100 GB)	0.0569	(0.0012)	0.0460	(0.0057)
Released after 2018	0.6289	(0.0049)	0.5925	(0.0087)
Apple	1.5337	(0.0077)	1.2425	(0.0165)
Samsung	0.1829	(0.0021)	0.1593	(0.0038)
Low-price synthetic	1.1822	(0.0067)	1.0409	(0.0132)
Mid-price synthetic	1.6120	(0.0074)	1.3495	(0.0146)
High-price synthetic	2.3655	(0.0096)	1.9184	(0.0197)
Discovery Parameters				
$\bar{\beta}$: Mean of true price sensitivity (100$)	-0.3353	(0.0014)	-0.1757	(0.0027)
δ: Prior deviation from true price sensitivity	0.0983	(0.0015)	-	
$\sigma_{\bar{\beta}}$: SD of true price sensitivity	0.0645	(0.0006)	0.1084	(0.0028)
σ_0: SD of initial belief means	0.1374	(0.0008)	-	
τ_1: Initial learning speed	0.5219	(0.0057)	-	
σ_s: SD of the preference signal	0.3752	(0.0061)	-	
Outside Option Value Parameters				
$u^{outside}$	-		1.1358	(0.0118)
$\mu_0^{outside}$	-0.4363	(0.0223)	-	
$\xi^{outside}$	3.6471	(0.1008)	-	
Search Cost Parameters				
$\bar{c}$: Mean of log search cost	-0.5696	(0.0051)	-1.6618	(0.0051)
σ_c: SD of log search cost	1.3364	(0.0051)	1.3364	(fixed)
N	50000		50000	

Notes: The estimation results are based on the subsample drawn from a cleaned-up sample of 595,968 consumers in Table 2. The standard deviation is calculated from the numerical information matrix approximation and is shown in the parenthesis. For each consumer, we draw 3,500 groups of errors in the preference discovery model estimation and 2,500 groups in the Weitzman model estimation.

As a reference, we also estimate a Weitzman-style model using inspections and the last checkout of the same sample, in which consumers search with their preferences drawn from the normal distribution $\mathcal{N}(\bar{\beta}, \sigma_\beta)$. The behavior of consumers checking out products in the middle of the search process is considered random, independent, uninformative, and costless, leading to no preference discovery. The Weitzman model estimates are stated in Columns 3 and 4. As Chung et al. (2024) specified, σ_c is difficult to identify in a Weitzman-style search and much relies on econometricians' assumption. Hence, we use the estimate from the preference discovery model for Weitzman model estimation.

All preference parameters toward product attributes are positive. Notice that these attributes can be obtained on the list page, influencing both consumers' inspection and checkout decisions. Our results are significant in all estimates partly because we have a large consumer sample, each making multiple choices in a wide market. Consumers have a positive preference for larger cellphone memory and storage, which coincides with intuition. Compared to other brands (Xiaomi, Huawei, and Oppo), Apple and Samsung provide significantly larger brand effects, consistent with their market shares. The estimates from the Weitzman model differ from the discovery model but have similar scale and relative size, which confirms our model estimates.

The estimation results of our model suggest clear evidence of preference changing across inspection rounds. The estimates of $\bar{\beta}$ and δ show that consumers enter the market with a significant positive bias compared to their true price sensitivity. For the rest of the estimates, we show the estimated and predicted perceived price sensitivity curve in the left side graph of Figure 6, assuming no stoppage at checkouts. We can see that consumers' mean of the perceived price sensitivity is around -0.24 against a mean of the true price sensitivity of -0.33. The graph indicates an overall underestimation of the price sensitivity at the beginning of search. As estimated from the first and second checkouts, consumers engage in search with an average initial deviation of 29.3%. They become more price-sensitive as the search continues and shrink the underestimation to 19% in the third inspection round, which displays an evident preference discovery effect. We predict that preference discovery will shrink the mean deviation to 12.4% after the fifth checkout. Specifically, if we assume that consumers in our population purchase in the inspection round where they are recorded in the data, the average price sensitivity increases by 10.6% compared to when consumers start searching.

The right side graph in Figure 6 illustrates the heterogeneity in the distribution and evolution of price sensitivity. We show both the interquartile range and the interval between the 2.5th and 97.5th percentiles of perceived price sensitivity, highlighting a broad dispersion in consumers' price sensitivity. Notably, 5.9% of simulated consumers start their search with a positive prior preference for higher prices. This finding aligns with the Weitzman model estimates in Table 3 and is consistent with those consumers seeking more advanced product alternatives, given that we do not control for product fixed effects. We also observe that the 2.5th percentile of true price sensitivity lies above the corresponding perceived sensitivity, suggesting that around 20%

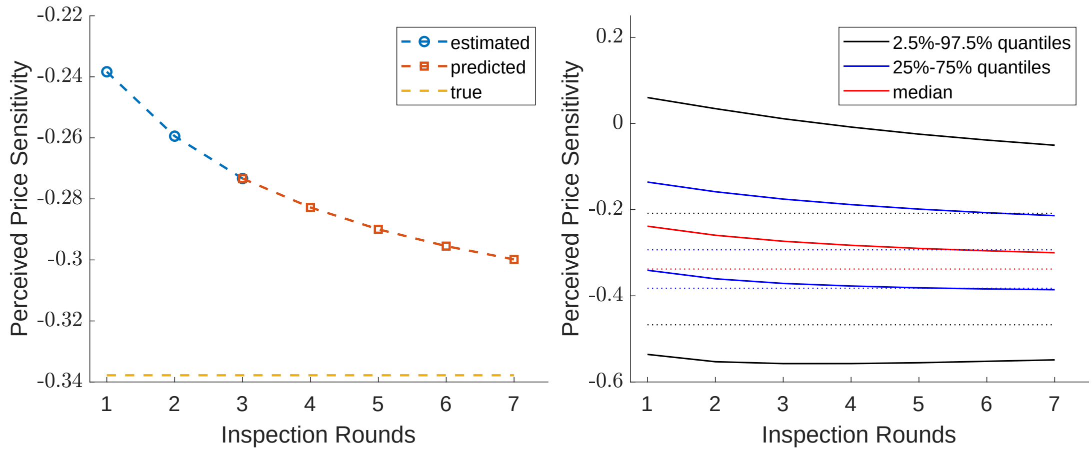

FIGURE 6 – Price Preference Evolution: Curve and Simulated Deviation

of consumers would become less price-sensitive if they fully learned their preferences through the search process.

Lastly, we report the mean and standard deviation estimates of log search cost in the bottom panel of Columns 1 and 2 in Table 3. In Weitzman models, estimating the degree of random variation in perceived reservation values relative to perceived utilities is often challenging due to weak identification issues. Our model, by allowing search cost draws across inspection rounds and assuming $\sigma_\varepsilon = 1$, enables us to estimate the stochasticity in reservation values. We find that σ_c is significantly large, with the standard deviation of $m_\varepsilon(c_{ijr})$ reaching approximately 2.9 when calculated with the given search cost parameters. As a comparison, Jiang et al. (2021) estimated heterogeneity at around 0.5 with a variance of 1 for ε using grid search. Such a big difference likely stems from three aspects. First, Jiang et al. (2021) focused on a narrow market with homogeneous goods (iPad mini 16G WiFi), examining reservation value variations among sellers of the same product, while our model considers a very diversified cellphone market. Second, Jiang et al. (2021) accounted for fixed effects across sellers, which our model does not include. Lastly, our results capture reservation value stochasticity both across consumers and across inspection rounds of the same consumer, the latter of which is absent in the Weitzman model, where reservation values are considered invariant for each consumer. This internal variation in our model explains a significant amount of checkout abandonment.

The results support the findings in the descriptive statistics in Section 4.4. Though preference changes were less apparent for the price trend of inspected products, they are pronounced in the checkout and purchase stages, indicated by the trend toward cheaper products at checkout.

Our model provides good predictions for the market share. After taking the weighted average across weekly markets, we show the market share data and predictions from the preference discovery model and the Weitzman model in Figure 8. The market share of each product is the ratio to the number of consumers who purchase, while the market share of the outside option is the ratio to the sample size.

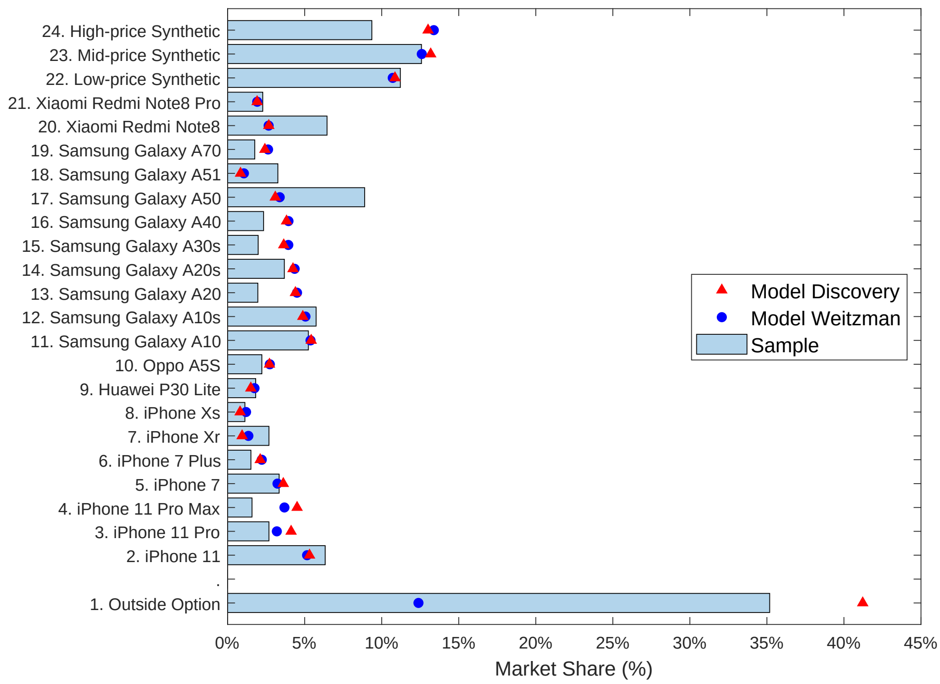

FIGURE 7 – Market Share: Data and Predictions: In Products

Notes: The market share of the outside option is calculated over the whole sample, while the market share of each product is calculated conditional on purchase.

When conditional on a purchase, our model predicts market share similarly to the Weitzman-style model. Although a predicted purchase in our model needs two checkouts in different inspection rounds under both prior and posterior preferences, we do not observe a significant decline in prediction quality. Both models perform poorly in some products because of the correlation in product values. For example, the Samsung Galaxy A50 and A40 are released closely, and their best-seller specifications share identical RAM size and storage. However, the product-page attributes of A50 are a much better choice than the A40, while in both models, their product-page values are assumed to be drawn independently. We show the market share across different brands in the following graph. The performance of market share predictions across brands is much closer to what we see in the data.

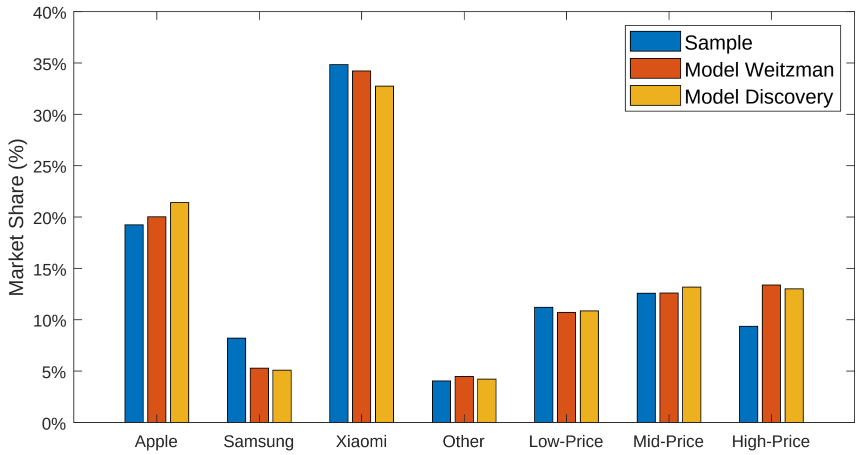

FIGURE 8 – Market Share: Data and Predictions: In Brand

Notes: The market share of each brand is calculated conditional on purchase.

Our model provides better predictions on the share of the leavers than the Weitzman model. Specifically, the preference discovery estimate for the outside option value after the second checkout is 1.1000, and after the third checkout, it is 1.6793. On the contrary, the Weitzman estimate reports 1.1358 between the two values. Our estimates suggest that the outside option becomes more attractive as the search goes on, showing that consumers leave the market not only driven by a time-invariant outside value but also process-related factors.

To further show the model fit, we illustrate the predicted leave-continue ratio after each checkout in Table 4. Our nonlinear specification for the outside option value provides a good approximation to the data, capturing the leave-continue ratios not only for the first two check-outs used in the estimation but also for subsequent rounds. For comparison, we estimated a preference discovery model with an additional restriction that the outside option value remains constant across rounds. In this case, the predicted leave-continue ratios deviate significantly from the actual data. This supports the notion that preference discovery alone cannot fully explain the transitions from checking out a product to leaving the market. External disruptions likely increase the probability of market exit as the search process extends over time. Additionally, a Weitzman-style model, which assumes invariant perceived utilities for all products, including the outside option, cannot capture variations in the outside option value.

Next, we report consumers' predicted response to product-level price changes. The summary statistics of the own-price elasticity are reported in Table 5, and we compare it with the own-price elasticity from the Weitzman model. Our model predicts a significantly more sensitive own-price elasticity in response to a 1% price change, which is slightly below -2 across all three markets, while the Weitzman model shows an elasticity around -1.5.

The distribution of the own-price elasticity of our model is shown in the left column of Figure 9. In general, higher-priced products exhibit greater elasticity in later rounds, while cheaper products become less elastic as the search process progresses. For cheaper products, a one-percent price change results in a smaller increase in price disutility, which cannot offset the utility gap created by product attributes and idiosyncratic preference signals. Conversely, for more expensive products, price increases do not elicit a significant negative market response due to the presence of consumers whose price insensitive or prefer luxury items. Price-insensitive consumers are not largely affected by price increases, while luxury chasers are attracted by more expensive prices. Consequently, the higher price does not drive many consumers away, while customers of other high-priced products are attracted by the higher price, mitigating the potential decrease in demand. In contrast, products in the mid-price range show highly elastic demand in response to price changes.

The right column of Figure 9 illustrates the own-price elasticity for checkouts across differ-

TABLE 4 – Simulated Leave-Continue Choice Table

Leave After:	1st Checkout	2nd Checkout	3rd Checkout	4th Checkout
Population	29.29%	24.21%	24.22%	25.53%
Sample	29.38%	25.09%	24.24%	27.93%
21-model market prediction	28.91%	24.54%	23.49%	23.79%
22-model market prediction	27.36%	23.83%	22.62%	22.99%
23-model market prediction	26.75%	23.48%	22.74%	22.98%
23-model, constant outside value	33.62%	7.40%	6.02%	5.19%

Notes: This table simulates the rate of consumers who choose to leave the market after each checkout using the estimates in Table 3 and the estimates to the same model only assuming the outside option value is round-invariant. The population corresponds to the data described in Table 2 and the sample corresponds to the 50,000 consumers used for estimation in 3.

TABLE 5 – Summary Statistics of Own-Price Elasticity

Market		Obs.	Mean	SD	Min	50%	Max
21-Model	Discovery	21	-2.0828	0.8388	-3.3075	-2.2314	-0.3909
	Weitzman	21	-1.4509	0.5721	-2.3291	-1.4630	-0.4389
22-Model	Discovery	22	-2.0369	0.7830	-3.2746	-2.1234	-0.5010
	Weitzman	22	-1.3960	0.5266	-2.2608	-1.3774	-0.4973
23-Model	Discovery	23	-2.0770	0.7552	-3.2485	-1.9520	-0.5916
	Weitzman	23	-1.4270	0.5106	-2.2067	-1.2865	-0.5461

Notes: The elasticities are derived by simulating how checkouts and demand change following a one-percent increase in a product's price.

ent inspection rounds. Generally, higher-priced products become more elastic in later rounds, whereas cheaper products become less elastic as the search process continues. Figure 6 reveals that consumers tend to be more price-sensitive in the later inspection rounds, and the proportion of consumers with a positive price preference diminishes. The loss of luxury-preferring consumers results in a more pronounced negative response to price increases, thereby increasing the price elasticity of high-priced products.

Our estimates suggest that the Weitzman model significantly underestimates price elasticity without accounting for the checkout abandonment and preference discovery. This occurs because the Weitzman model overlooks selections made during the search process, instead deferring them to the actual purchase observations. This leads to an overestimation of the search process duration and an underestimation of both consumers' true price sensitivity and the search cost. In reality, consumers already make selection decisions when they take a product to checkout; however, preference discovery, behavioral shocks, or higher outside value may prompt them to abandon the checkout product, leading to an unpredicted extension of the search process. While products that were not inspected before checkout may make use of the abandonment and be inspected in the following rounds, this does not necessarily increase their likelihood of being purchased, as consumers make checkout decisions with a more price-sensitive preference in later rounds. In short, cheap checkouts are more likely to be checked out in the next rounds, while expensive checkouts are not.

Corresponding to this, we conduct a counterfactual to show that a recommendation that reduces the search cost will not be as lucrative as predicted. Such recommendations are widely applied on online market platforms to increase the probability of being inspected. We consider

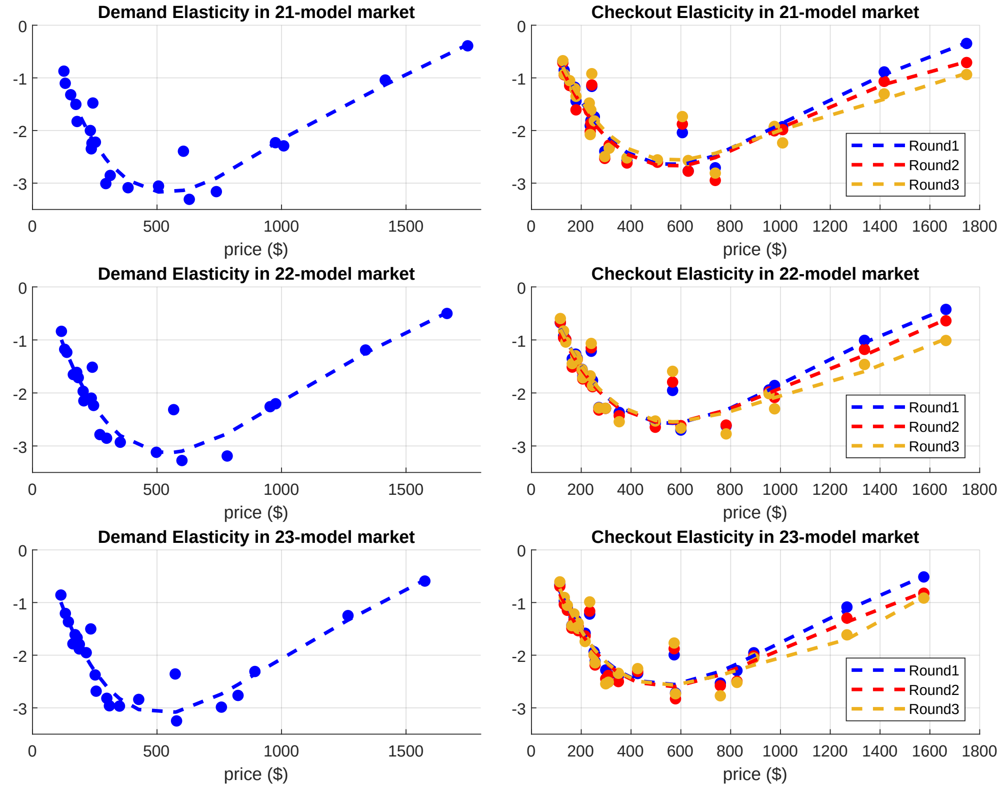

FIGURE 9 – Own-Price Elasticity: Demand and Checkout

an extreme case in which the search cost is reduced to 0 such that the recommended product is always inspected. We see the recommendation increase the product's revenue in both models, while in the Weitzman model, it is much larger than that in the preference discovery model. We show the comparative difference between the predicted revenue of the two models in Figure 10.

Our results indicate that the predicted benefits of recommendation are relatively similar between the two models for low-priced products. This is because these products are not significantly impacted by preference changes; their low prices inherently limit the extent to which perceived utility varies due to preference discovery. Consequently, the primary benefit of low-priced products from the recommendation is from consumers' increased inspections, which yields similar gains in both models. However, for higher-priced products, the prediction difference between the models becomes more pronounced. Although these products benefit from being included in the choice set without requiring inspections in both models, they are more affected by preference changes in our model, which puts them at a disadvantage once a checkout abandonment happens. In subsequent checkout decisions, the high price makes these products

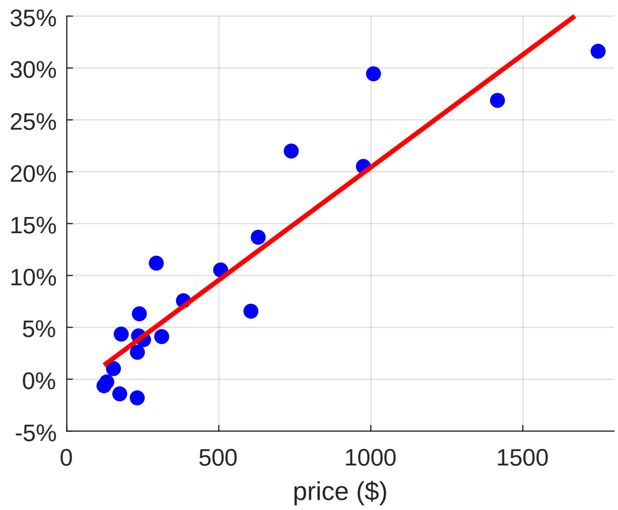

FIGURE 10 – Predicted Revenue Difference in Weitzman and Preference Discovery Models

Notes: The graph shows how much larger the revenue prediction in the Weitzman Model than in the Preference Discovery Model. We illustrate with a market of 21 models and simulate with 50000 consumers 50 times using the estimation obtained in Table 3.

less likely to be selected by consumers with increased price sensitivity. This is reflected in our counterfactual analysis, where the Weitzman model overestimates the benefits of recommending higher-priced products.

TABLE 6 – Welfare Changes of a One-click Buying Design (Percentage)

Market	21-model		22-model		23-model	
One-click Purchase Probability	50%	100%	50%	100%	50%	100%
Average Utility of Purchasers	-5.5	-8.8	-5.2	-8.3	-4.9	-7.9
Total Utility of All Consumers	2.7	4.7	2.8	4.6	2.8	4.8
Average Price of Purchased Products	-0.3	-0.5	-0.3	-0.5	-0.3	-0.6
Total Revenues of Sellers	8.4	14.2	8.0	13.4	7.9	13.1
Total Search Costs	-26.4	-44.5	-27.0	-45.6	-27.6	-46.5
Average Search Costs	-32.3	-51.6	-32.6	-52.3	-33.0	-53.0

Notes: This table examines the percentage changes in consumer utility and average utility, seller revenue and average product price, as well as consumer search costs and average search costs when introducing a one-click purchase mechanism that is triggered with a 50% or 100% probability during each checkout, compared to the baseline scenario where the mechanism does not exist. We conduct 50 simulations with $N = 50000$ simulated consumers using the parameter estimates in Table 3 and report the averages.

We conclude by reporting the welfare effect of consumer preference discovery. Specifically, we discuss how introducing a one-click buying design affects the true utility consumers derive from purchased products, the transaction revenue, and the search costs incurred. We assume that the one-click buying is randomly triggered with a fixed probability, allowing consumers to purchase a product without entering the checkout or engaging in preference discovery. The results are presented in Table 6. Notably, in our model, the outside option value changes across inspection rounds. To avoid considering the welfare effect from the outside option, we assign a utility value of 0 to consumers who ultimately leave the market, taking that they derive no utility from the current market.

Introducing a one-click buying design significantly increases the total utility consumers obtain from the market. However, the utility for each buyer decreases under this design. One-click buying can cause consumers to prematurely conclude their search process and make decisions before well discovering their preferences. On the other hand, the design reduces the likelihood of consumers leaving the market after checkouts, thereby increasing the total number of purchases. The average purchase price remains stable, as premature buying leads consumers who would still buy a product without the design to buy more expensive products, while others who would have exited the market instead will be retained to buy cheaper products. The total seller revenue increases significantly due to the higher amount of buyers.

We also observe that the total search costs incurred by consumers drop substantially under the one-click purchase design as they inspect fewer products. Due to the reasons described in Section 4.2, we cannot directly compare search costs with utility, so the overall welfare effect depends on the potential heterogeneity of private product information. When the information obtained through inspection has a better chance to significantly enhance consumer utility, the welfare gains from reduced search costs cannot offset the losses from reduced search and preference discovery. Otherwise, reducing unnecessary search and preference discovery may increase consumer surplus.

6 Conclusion

Inferring consumer preferences from choice data plays a central role in analyzing market structure, informing platform strategy, and shaping regulatory policy. However, this approach relies on a key assumption: that consumers fully know their preferences when making decisions. While widely adopted, this assumption has been increasingly questioned in recent studies, and its empirical implications remain underexplored due to the lack of suitable field contexts.

Using click-stream data from an e-commerce platform that records smartphone search and purchase behavior, we empirically examine the role of preference discovery in consumer search. We focus on the checkout stage, where no new information is introduced, allowing us to isolate behavioral adjustments captured by checkout abandonment decisions. Evidence from product prices and attribute convergence before and after checkout attempts points clearly to the presence of preference discovery. These patterns suggest that consumers do not act on fixed price sensitivity, but instead gradually form judgments about product value relative to price as they progress through the search process.

Motivated by this evidence, we develop a structural model that incorporates preference discovery and estimates the key parameters that govern its evolution. Capturing both checkout abandonments and preference discovery is essential for accurately understanding and modeling consumer decisions. Our results show that our model predicts significantly larger price elasticities compared to standard search models. This is because consumers often initiate tentative checkout attempts during the search process and exhibit greater price sensitivity after abandonment, which amplifies price competition. We also examine mechanisms such as one-click purchases that encourage early choices to become final decisions. While such designs may raise platform conversion rate and total revenue, they also risk limiting preference discovery, potentially reducing consumer decision quality and surplus.

Our findings underscore the significance and ubiquity of preference discovery in real-world consumer search. Rather than serving solely to acquire information, pre-purchase search also facilitates self-reflection, helping consumers better understand their preferences. As a result, the value of search lies not only in addressing information asymmetry but also in supporting preference discovery. For market designers, these insights suggest the importance of tailoring information disclosure, recommendation strategies, and search tool design to different stages of the search journey. For regulators, the conclusions emphasize the need to guard against platform mechanisms that prompt consumers to make premature decisions before their preferences have been adequately established. Future work can build on these findings to further examine the implications of preference discovery for market design and policy.

References

Ackerberg, D. A. (2003). Advertising, learning, and consumer choice in experience good markets: an empirical examination. *International Economic Review 44*(3), 1007–1040.

Ariely, D., G. Loewenstein, and D. Prelec (2003). "coherent arbitrariness": Stable demand curves without stable preferences. *The Quarterly Journal of Economics 118*(1), 73–106.

Balachander, S. and S. Ghose (2003). Reciprocal spillover effects: A strategic benefit of brand extensions. *Journal of Marketing 67*(1), 4–13.

Baymard Institute (2024). Baymard Institute: 49 cart abandonment rate statistics (2024). https://baymard.com/lists/cart-abandonment-rate. Last accessed on 18.10.2024.

Berry, S. T. and P. A. Haile (2014). Identification in differentiated products markets using market level data. *Econometrica 82*(5), 1749–1797.

Branco, F., M. Sun, and J. M. Villas-Boas (2012). Optimal search for product information. *Management Science 58*(11), 2037–2056.

Bronnenberg, B. J., J. B. Kim, and C. F. Mela (2016). Zooming in on choice: How do consumers search for cameras online? *Marketing Science 35*(5), 693–712.

Cao, X. and J. Zhang (2021). Preference learning and demand forecast. *Marketing Science 40*(1), 62–79.

Cerreia-Vioglio, S., F. Maccheroni, M. Marinacci, and A. Rustichini (2023). Multinomial logit processes and preference discovery: inside and outside the black box. *The Review of Economic Studies 90*(3), 1155–1194.

Chen, Y. and S. Yao (2017). Sequential search with refinement: Model and application with click-stream data. *Management Science 63*(12), 4345–4365.

Chick, S. E. and P. Frazier (2012). Sequential sampling with economics of selection procedures. *Management Science 58*(3), 550–569.

Ching, A. T. (2010). Consumer learning and heterogeneity: Dynamics of demand for prescription drugs after patent expiration. *International Journal of Industrial Organization 28*(6), 619–638.

Chung, J. H., P. Chintagunta, and S. Misra (2024). Simulated maximum likelihood estimation of the sequential search model. *Quantitative Marketing and Economics*, 1–60.

Cooke, K. (2017). Preference discovery and experimentation. *Theoretical Economics 12*(3), 1307–1348.

Crawford, G. S. and M. Shum (2005). Uncertainty and learning in pharmaceutical demand. *Econometrica 73*(4), 1137–1173.

De Los Santos, B., A. Hortaçsu, and M. R. Wildenbeest (2017). Search with learning for differentiated products: Evidence from e-commerce. *Journal of Business & Economic Statistics 35*(4), 626–641.

DeGroot, M. H. (1970). *Optimal Statistical Decisions*. John Wiley & Sons.

Delaney, J., S. Jacobson, and T. Moenig (2020). Preference discovery. *Experimental Economics 23*(3), 694–715.

Dzyabura, D. and J. R. Hauser (2019). Recommending products when consumers learn their preference weights. *Marketing Science 38*(3), 417–441.

Erdem, T. (1998). An empirical analysis of umbrella branding. *Journal of Marketing Research 35*(3), 339–351.

Erdem, T. and M. P. Keane (1996). Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing Science 15*(1), 1–20.

Gardete, P. M. and M. Hunter (2024). Multiattribute search: Empirical evidence and information design. *Marketing Science*.

Greminger, R. P. (2024). Heterogeneous position effects and the power of rankings. *arXiv preprint arXiv:2210.16408*.

Grenet, J., Y. He, and D. Kübler (2022). Preference discovery in university admissions: The case for dynamic multioffer mechanisms. *Journal of Political Economy 130*(6), 1427–1476.

Huang, G.-H., N. Korfiatis, and C.-T. Chang (2018). Mobile shopping cart abandonment: The roles of conflicts, ambivalence, and hesitation. *Journal of Business Research 85*, 165–174.

Jiang, Z., T. Chan, H. Che, and Y. Wang (2021). Consumer search and purchase: An empirical investigation of retargeting based on consumer online behaviors. *Marketing Science 40*(2), 219–240.

Jindal, P. and A. Aribarg (2021). The importance of price beliefs in consumer search. *Journal of Marketing Research 58*(2), 321–342.

Jolivet, G. and H. Turon (2019). Consumer search costs and preferences on the internet. *The Review of Economic Studies 86*(3), 1258–1300.

Kahneman, D. and A. Tversky (1979). Prospect theory: An analysis of decision under risk. *Econometrica 47*(2), 263–292.

Kahneman, D., P. P. Wakker, and R. Sarin (1997). Back to bentham? explorations of experienced utility. *The Quarterly Journal of Economics 112*(2), 375–406.

Keller, G. and S. Rady (1999). Optimal experimentation in a changing environment. *The Review of Economic Studies 66*(3), 475–507.

Kim, J. B., P. Albuquerque, and B. J. Bronnenberg (2010). Online demand under limited consumer search. *Marketing Science 29*(6), 1001–1023.

Koulayev, S. (2014). Search for differentiated products: identification and estimation. *The RAND Journal of Economics 45*(3), 553–575.

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

Moraga-González, J. L., Z. Sándor, and M. R. Wildenbeest (2023). Consumer search and prices in the automobile market. *The Review of Economic Studies 90*(3), 1394–1440.

Morozov, I., S. Seiler, X. Dong, and L. Hou (2021). Estimation of preference heterogeneity in markets with costly search. *Marketing Science 40*(5), 871–899.

Narita, Y. (2018). Match or mismatch? learning and inertia in school choice. *Available at SSRN: 3198417*.

Plott, C. R. (1996). Rational individual behavior in markets and social choice processes: the discovered preference hypothesis. In K. J. Arrow, E. Colombatto, M. Perlman, and C. Schmidt (Eds.), *Rational foundations of economic behavior*, Chapter 10, pp. 225–250. Macmillan Press.

Plott, C. R. and K. Zeiler (2005). The willingness to pay–willingness to accept gap, the "endowment effect," subject misconceptions, and experimental procedures for eliciting valuations. *American Economic Review 95*(3), 530–545.

Prelec, D. and G. Loewenstein (1998). The red and the black: Mental accounting of savings and debt. *Marketing Science 17*(1), 4–28.

Saleh, K. and A. Shukairy (2010). *Conversion optimization: The art and science of converting prospects to customers*. " O'Reilly Media, Inc.".

Szymanowski, M. and E. Gijsbrechts (2012). Consumption-based cross-brand learning: are private labels really private? *Journal of Marketing Research 49*(2), 231–246.

Thaler, R. (1985). Mental accounting and consumer choice. *Marketing Science 4*(3), 199–214.

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

Ursu, R., S. Seiler, and E. Honka (2023). The sequential search model: A framework for empirical research.

Ursu, R. M. (2018). The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions. *Marketing Science 37*(4), 530–552.

Ursu, R. M., Q. Wang, and P. K. Chintagunta (2020). Search duration. *Marketing Science 39*(5), 849–871.

van Gelder, K. (2023). Share of shoppers who usually compare prices before buying online in the United States as of August 2022. https://www.statista.com/statistics/1375931/price-comparison-online-shopping-habits-us/. Last accessed on 24.06.2024.

Weitzman, M. L. (1979). Optimal search for the best alternative. *Econometrica 47*(3), 641–654.

Wu, X., M. S. Lewis, and F. A. Wolak (2024). Search with learning in the retail gasoline market. *The RAND Journal of Economics*.

Zellermayer, O. (1996). *The pain of paying*. Carnegie Mellon University.

Zhang, T. (2025). Estimating sequential search models based on a partial ranking representation. *arXiv preprint arXiv:2501.07514*.

Appendix

A Data Cleaning Process

Our raw clickstream data includes consumers' view, checkout, and purchase clicks on the website. Each click is recorded with the product SKU ID, consumer ID, timestamp, product categories (not used because of incompleteness and errors), product price, and consumer session. The timestamps are in discrete intervals of 131,072 milliseconds, and for clicks occurring simultaneously in the data, we rank them following the dataset's original order. The cleaning process is as follows:

1. Match product SKU IDs to the attribute dataset containing information of 1,708 cellphone product pages on the website, discarding unmatched clicks.

2. Remove all observations with a price value of 0.

3. For repeated actions (i.e., same product id, time, and event type with a nearby click by the same user), retain only the first occurrence. This resulted in an initial search dataset with 84,711,440 observations, 5,453,475 consumers, and 1,699 product pages.

4. Defined steps: Each step comprises consecutive actions by the same consumer on the same SKU ID (e.g., checkout and purchase, or checkout, view, and checkout again). Since no other product pages are involved, we treated these actions as a single step, retaining only the highest-level event (purchase > checkout > view). This yielded 58,774,389 steps.

5. Exclude all steps occurring after a consumer's first purchase, leaving 49,577,080 steps.

6. Remove consumers with only one step, resulting in a dataset with 48,223,442 steps, 4,099,837 consumers, and 1,690 product pages.

7. Expand the sample for reduced-form analysis by creating "view" records for all checkout and purchase steps[1]: For steps containing a checkout, we created two records (view and checkout). For steps containing a purchase, we create three records (view, checkout, and purchase), all timestamped as occurring simultaneously. This produced a total of 49,722,578 data points.

[1] This expansion was not applied when defining steps, as some checkouts and purchases in the raw data were recorded without a view.

8. Sample selection: To control careless buyers or forgetting, we exclude sequences where the purchase or the last view before exit occurs between 5 minutes and two weeks after the first view click.

This results in the sample used in Section 2.2.

B More Tables and Figures for the Reduced-form Evidence

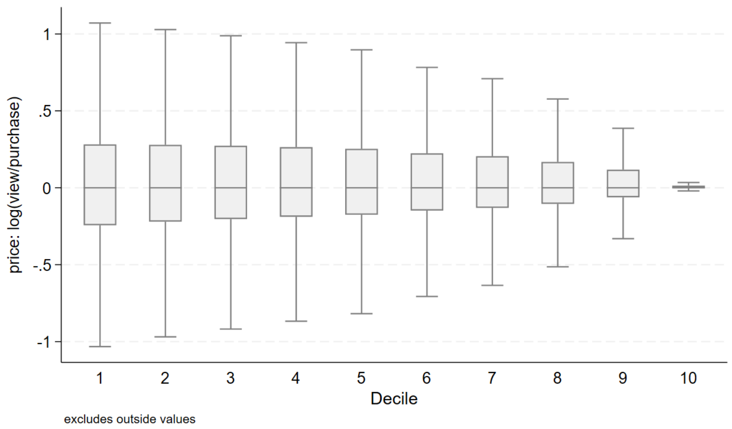

FIGURE B.1 – Convergence to Prices of Purchased Product

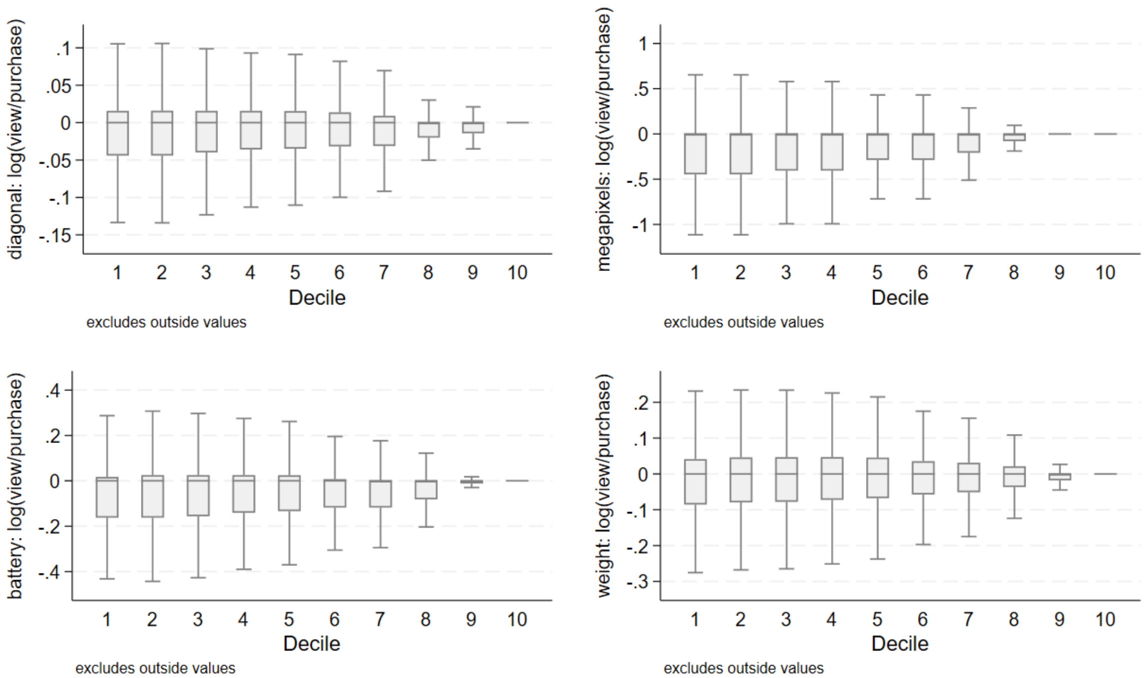

FIGURE B.2 – Convergence to Attributes of Purchased Product

	Full Sample			Subsample with 3 Checkouts			Subsample with 4 Checkouts					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
1 Checkout	-0.052***	-0.074***	-0.020***	-0.047***	-0.068***	-0.023***	-0.037***	-0.050***	-0.026***	-0.011**	-0.017**	-0.025***
	(0.001)	(0.002)	(0.002)	(0.003)	(0.004)	(0.004)	(0.004)	(0.006)	(0.006)	(0.004)	(0.006)	(0.006)
2 Checkouts	-0.065***	-0.093***	-0.037***	-0.111***	-0.168***	-0.042***	-0.055***	-0.073***	-0.049***	-0.013**	-0.022**	-0.047***
	(0.002)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.005)	(0.008)	(0.007)	(0.004)	(0.007)	(0.007)
3 Checkouts	-0.073***	-0.103***	-0.051***	0.000	0.000	0.000	-0.115***	-0.171***	-0.064***	-0.035***	-0.071***	-0.062***
	(0.003)	(0.004)	(0.005)	(.)	(.)	(.)	(0.005)	(0.007)	(0.008)	(0.004)	(0.006)	(0.008)
4 Checkouts	-0.075***	-0.106***	-0.046***	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.004)	(0.006)	(0.007)	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)
Step 2	-0.047***	-0.058***	0.013***	0.010***	0.025***	0.018***	0.009	0.020**	0.020***	0.002	0.009	0.017***
	(0.001)	(0.001)	(0.001)	(0.003)	(0.004)	(0.002)	(0.005)	(0.007)	(0.004)	(0.004)	(0.006)	(0.004)
Step 3	-0.039***	-0.042***	0.026***	-0.004	0.008	0.031***	0.005	0.017*	0.029***	-0.002	0.006	0.028***
	(0.001)	(0.001)	(0.001)	(0.003)	(0.004)	(0.003)	(0.005)	(0.007)	(0.004)	(0.004)	(0.006)	(0.004)
Step 4	-0.033***	-0.030***	0.037***	-0.000	0.014**	0.041***	0.004	0.016*	0.045***	-0.003	0.007	0.043***
	(0.001)	(0.001)	(0.001)	(0.003)	(0.005)	(0.003)	(0.005)	(0.008)	(0.005)	(0.004)	(0.007)	(0.005)
Step 5	-0.029***	-0.023***	0.047***	0.002	0.019***	0.047***	-0.005	0.014	0.050***	-0.012**	0.003	0.046***
	(0.001)	(0.001)	(0.001)	(0.003)	(0.005)	(0.003)	(0.005)	(0.008)	(0.005)	(0.004)	(0.007)	(0.005)
Step 6	-0.026***	-0.016***	0.055***	0.003	0.020***	0.053***	-0.003	0.008	0.055***	-0.010*	0.000	0.051***
	(0.001)	(0.001)	(0.001)	(0.003)	(0.005)	(0.003)	(0.005)	(0.008)	(0.005)	(0.004)	(0.007)	(0.006)
Step 7	-0.024***	-0.012***	0.062***	0.002	0.019***	0.060***	-0.001	0.020*	0.060***	-0.011*	0.009	0.054***
	(0.001)	(0.002)	(0.001)	(0.003)	(0.005)	(0.004)	(0.006)	(0.008)	(0.006)	(0.005)	(0.008)	(0.006)
Step 8	-0.023***	-0.007***	0.069***	0.004	0.033***	0.065***	-0.002	0.015	0.064***	-0.013**	0.001	0.058***
	(0.001)	(0.002)	(0.001)	(0.004)	(0.005)	(0.004)	(0.006)	(0.009)	(0.006)	(0.005)	(0.008)	(0.006)
Unobs Attri Dev	No	No	No	No	No	No	No	No	No	Yes	Yes	Yes
Adjusted R^2	0.006	0.005	0.013	0.017	0.016	0.013	0.016	0.013	0.016	0.322	0.198	0.055
Observations	3467864	3467864	3467864	377045	377045	377045	170171	170171	170171	165216	165216	165216

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

Notes: Standard errors in parentheses, clustered at the individual level. Restrict sample to sequence with a purchase. iOS is a binary variable

TABLE B.1 – Estimates of Absolute Log Deviations of Other Attributes

C Computation of the (Perceived) Reservation Value

The definition of reservation value is a possessed value that equalizes the expected gain from inspecting a product j and the search cost of product j:

$$
\begin{aligned}
c_{ijr} &= \int_{\varepsilon_{ij} > \bar{v} - v_{ijr}} (\varepsilon_{ij} - (\bar{v} - v_{ijr}))\, dF(\varepsilon_{ij}) \\
&= \left(1 - F\left(\frac{\bar{v} - v_{ijr}}{\sigma_\varepsilon}\right)\right) \int_{\varepsilon_{ij} > (\bar{v} - v_{ijr})}^{\infty} (\varepsilon_{ij} - (\bar{v} - v_{ijr})) \frac{f(\varepsilon_{ij})}{1 - F\left(\frac{\bar{v} - v_{ijr}}{\sigma_\varepsilon}\right)} d\varepsilon_{ij} \\
&= \left(1 - F\left(\frac{\bar{v} - v_{ijr}}{\sigma_\varepsilon}\right)\right) \cdot \mathrm{E}(\varepsilon_{ij} - (\bar{v} - v_{ijr}) \mid \varepsilon_{ij} > (\bar{v} - v_{ijr})) \\
&= \left(1 - \Phi\left(\frac{\bar{v} - v_{ijr}}{\sigma_\varepsilon}\right)\right) \cdot \left[\sigma_\varepsilon \frac{\phi\left(\frac{\bar{v} - v_{ijr}}{\sigma_\varepsilon}\right) - 0}{1 - \Phi\left(\frac{\bar{v} - v_{ijr}}{\sigma_\varepsilon}\right)} - (\bar{v} - v_{ijr})\right] \\
&= \sigma_\varepsilon \cdot \left[\phi\left(\eta_{ijr}\right) - \left(1 - \Phi\left(\eta_{ijr}\right)\right) \eta_{ijr}\right]
\end{aligned}
$$

where $\eta_{ijr} = \frac{\bar{v} - v_{ijr}}{\sigma_\varepsilon}$. We apply the assumption ε_{ij} following a normal distribution in the proof above. $\Phi(\cdot)$ and $\phi(\cdot)$ denote cdf and pdf of the Gaussian distribution respectively. The first-order condition with respect to $\bar{v}$ of the right-hand side is:

$$
\begin{aligned}
&\frac{\partial \sigma_\varepsilon \cdot \left[\phi\left(\eta_{ijr}\right) - \left(1 - \Phi\left(\eta_{ijr}\right)\right) \eta_{ijr}\right]}{\partial \bar{v}} \\
&= \sigma_\varepsilon \cdot \frac{1}{\sigma_\varepsilon} \cdot \left[\phi'(\eta_{ijr}) - (1 - \Phi(\eta_{ijr})) + \phi(\eta_{ijr})\eta_{ijr}\right] \\
&= -\eta_{ijr}\phi(\eta_{ijr}) - (1 - \Phi(\eta_{ijr})) + \phi(\eta_{ijr})\eta_{ijr} \\
&= -(1 - \Phi(\eta_{ijr}))
\end{aligned}
$$

which is always negative with a finite η_{ijr}. Notice that the left-hand side has a positive derivative, it implies a bijection between $\bar{v}$ and c_{ijr}. Therefore, we have a unique solution of $\bar{v}$, denoted by z_{ijr}. Define $m(x) = [(1 - \Phi(x))x + \phi(x)]^{-1}$, we obtain the linear specification in Section 3.2.

D Likelihood Function Construction

In Appendix D, we first compute the $\mathcal{L}_{i,discovery}$ in detail:

$$\mathcal{L}_{i,discovery} = \Pr(I_i \in \mathcal{I}_i \mid p_{i,h_1}, p_{i,h_2}, \cdots, p_{i,h_r}, x_{h_1}, x_{h_2}, \cdots, x_{h_r}, \bar{\beta}, \delta, \tau_1, \sigma_{\bar{\beta}}, \sigma_s)$$

where $I_i = \{\varepsilon_{i,h_1}, \cdots \varepsilon_{i,h_R}, \beta_{i1}, \cdots, \beta_{iR}\}$. Then, we show how to implement a Geweke-Hajivassilou-Keane (GHK) style simulator to the likelihood function.

The checkout products in all inspection rounds $\{h_1, h_2, \cdots, h_R\}$ together with the last inspected products $\{J_1, J_2, \cdots J_R\}$ are observed from the data. In the classic Weitzman model, selection decisions reflect stable utility relationships in the model. In our multi-round model, however, the perceived utilities of checkout products are compared in different inspection rounds, with observed changes in checkout outcomes reflecting variations in consumer preferences. To be more specific than Section 4.1, let $\Xi_i = \{\varepsilon_{i,h_1} \cdots \varepsilon_{i,h_R}\}$ and $B_i = \{\beta_{i1}, \cdots \beta_{iR}\}$ are the two subsets of Ψ_i. It requires constraints on B_i such that Ξ_i is non-empty. The curse of dimensionality limits the specification of these constraints in long search processes because the number of inequality conditions exceeds the elements in Ψ_i as the number of rounds increases. Therefore, we focus on the first three rounds of consumer search. In the following analysis, first, for each case of the observed checkout sequence, we outline the inequality conditions on Ψ_i. Then, we partition these conditions into preference constraints on B_i that guarantee the existence of a solution, and value constraints on Ξ_i conditional on the preference constraints being satisfied.

Notice that there are no inequality conditions when there is only one inspection round in the search sequence. We restate the inequalities and preference constraints of two-round search processes in D.1, and extend to three-round search processes in D.2 and D.3. We introduce the conditional distribution of the posterior mean in the preference discovery to facilitate the implementation of the preference constraints in D.4. The implementation of our simulated likelihood estimator will be introduced in D.5.

D.1 Two-round Search Process

Suppose we observe consumer i's search process consists of two inspection rounds, with $\Psi_i = \{\varepsilon_{i,h_1}, \varepsilon_{i,h_2}, \beta_{i1}, \beta_{i2}\}$. When $h_1 = h_2$, the consumer checks out the same product in two inspection rounds. Since we consider a two-round search sequence, it must be that the consumer purchases the first checkout product. In this case, there are no inequality conditions between perceived checkout product utilities. When $h_1 \neq h_2$, we consider two subcases below:

D.1.1 $J_1 < h_2 \leq J_2$

In this subcase, consumer i inspects product h_2 after she abandons h_1. She does not know the perceived utility of h_2 when checking out h_1 but knows product h_1 when checking out h_2. The perceived utilities of h_1 and h_2 are compared only in Round 2, leading to one inequality condition:

$$u_{i,h_2,2} \geq u_{i,h_1,2} \Rightarrow \varepsilon_{i,h_1} - \varepsilon_{i,h_2} < v_{i,h_2,2} - v_{i,h_1,2} = x_{h_2}\gamma + \beta_{i1} p_{ih_2} - (x_{h_1}\gamma + \beta_{i1} p_{ih_1})$$

We can always find $\{\varepsilon_{i,h_1}, \varepsilon_{i,h_2}\}$ with any preferences $\{\beta_{i,h_1}, \beta_{i,h_2}\}$ given, so there is no preference constraint. In reality, this subcase would not happen because the consumer does not stop searching after checking out product h_2 (no purchase or leaving concludes the search). We raise this subcase for completeness and to facilitate the understanding of more complicated cases.

D.1.2 $h_2 \leq J_1$

This subcase is the example given in Section 4.1, consumer i inspects both h_1 and h_2 in the first inspection round. We discussed the subcase $h_2 = h_1$ and now focus on the subcase of two different checkout products, i.e., $h_2 \neq h_1$. There are two inequality conditions that Ξ_i need to fulfill with preferences given:

$$\begin{cases} u_{i,h_1,1} \geq u_{i,h_2,1} \Rightarrow \varepsilon_{i,h_1} - \varepsilon_{i,h_2} > v_{i,h_2,1} - v_{i,h_1,1} = x_{h_1}\gamma + \beta_{i1} p_{ih_1} - (x_{h_2}\gamma + \beta_{i1} p_{ih_2}) \\ u_{i,h_2,2} \geq u_{i,h_1,2} \Rightarrow \varepsilon_{i,h_1} - \varepsilon_{i,h_2} < v_{i,h_2,2} - v_{i,h_1,2} = x_{h_1}\gamma + \beta_{i2} p_{ih_1} - (x_{h_2}\gamma + \beta_{i2} p_{ih_2}) \end{cases}$$

We are not able to find Ξ_i that satisfies both inequalities unless the following condition is imposed on β_{i1}, β_{i2}:

$$(\beta_{i1} - \beta_{i2})(p_{ih_1} - p_{ih_2}) > 0$$

D.2 Three-round Search Process: Three checkout products

When we consider the first three inspection rounds in search processes with at least three rounds, the inequality conditions on Ψ_i become more complicated. We first discuss all subcases in which consumer i checks out three different products.

D.2.1 $h_2 \leq J_1, h_3 \leq J_1$

In this subcase, all checkout products are inspected in the first inspection round, and compared with each other in all three rounds. We observe such a checkout sequence only when the following six inequality conditions on Ψ_i hold simultaneously:

$$\begin{cases} u_{i,h_1,1} \geq u_{i,h_2,1} \Rightarrow \varepsilon_{i,h_1} - \varepsilon_{i,h_2} \geq v_{i,h_2,1} - v_{i,h_1,1} \\ u_{i,h_1,1} \geq u_{i,h_3,1} \Rightarrow \varepsilon_{i,h_1} - \varepsilon_{i,h_3} \geq v_{i,h_3,1} - v_{i,h_1,1} \\ u_{i,h_2,2} \geq u_{i,h_1,2} \Rightarrow \varepsilon_{i,h_1} - \varepsilon_{i,h_2} \leq v_{i,h_2,2} - v_{i,h_1,2} \\ u_{i,h_2,2} \geq u_{i,h_3,2} \Rightarrow \varepsilon_{i,h_2} - \varepsilon_{i,h_3} \geq v_{i,h_3,2} - v_{i,h_2,2} \\ u_{i,h_3,3} \geq u_{i,h_1,3} \Rightarrow \varepsilon_{i,h_1} - \varepsilon_{i,h_3} \leq v_{i,h_3,3} - v_{i,h_1,3} \\ u_{i,h_3,3} \geq u_{i,h_2,3} \Rightarrow \varepsilon_{i,h_2} - \varepsilon_{i,h_3} \leq v_{i,h_3,3} - v_{i,h_2,3} \end{cases}$$

Ξ_i that satisfies these conditions is non-empty only when B_i fulfill the following preference constraints:

$$\begin{cases} (\beta_{i1} - \beta_{i2})(p_{i,h_1} - p_{i,h_2}) \geq 0 & \text{(D.2.1.a)} \\ (\beta_{i1} - \beta_{i3})(p_{i,h_1} - p_{i,h_3}) \geq 0 & \text{(D.2.1.b)} \\ (\beta_{i2} - \beta_{i3})(p_{i,h_2} - p_{i,h_3}) \geq 0 & \text{(D.2.1.c)} \end{cases}$$

Notice when the two preference constraints involving the medium-price checkout product are fulfilled, the rest one is naturally satisfied. For example, given $p_{i,h_1} > p_{i,h_2} > p_{i,h_3}$, if Condition (D.2.1.a) and Condition (D.2.1.c) hold, Condition (D.2.1.b) would also hold without any further condition. We use the subscript *high*, *mid*, and *low* to replace subscript r in the price preference β_{ir} and the checkout product h_r in inspection rounds with the highest, medium, and lowest checkout product price correspondingly. The formal restrictions imposed on $\{\beta_{i1}, \beta_{i2}, \beta_{i3}\}$ are:

$$\begin{cases} (\beta_{i,high} - \beta_{i,mid})(p_{i,h_{high}} - p_{i,h_{mid}}) \geq 0 \\ (\beta_{i,mid} - \beta_{i,low})(p_{i,h_{mid}} - p_{i,h_{low}}) \geq 0 \end{cases}$$

D.2.2 $J_1 < h_2 < J_2, h_3 < J_1$

When h_2 is not inspected in the first round, it is not compared to h_1 with preference β_{i1}. The inequality conditions on Ψ_i are given as follow:

$$\begin{cases} u_{i,h_1,1} \geq u_{i,h_3,1} \Rightarrow \varepsilon_{i,h_1} - \varepsilon_{i,h_3} \geq v_{i,h_3,1} - v_{i,h_1,1} & \text{(D.2.2.a)} \\ u_{i,h_2,2} \geq u_{i,h_1,2} \Rightarrow \varepsilon_{i,h_1} - \varepsilon_{i,h_2} \leq v_{i,h_2,2} - v_{i,h_1,2} & \text{(D.2.2.b)} \\ u_{i,h_2,2} \geq u_{i,h_3,2} \Rightarrow \varepsilon_{i,h_2} - \varepsilon_{i,h_3} \geq v_{i,h_3,2} - v_{i,h_2,2} & \text{(D.2.2.c)} \\ u_{i,h_3,3} \geq u_{i,h_1,3} \Rightarrow \varepsilon_{i,h_1} - \varepsilon_{i,h_3} \leq v_{i,h_3,3} - v_{i,h_1,3} & \text{(D.2.2.d)} \\ u_{i,h_3,3} \geq u_{i,h_2,3} \Rightarrow \varepsilon_{i,h_2} - \varepsilon_{i,h_3} \leq v_{i,h_3,3} - v_{i,h_2,3} & \text{(D.2.2.e)} \end{cases}$$

Notice that h_3 is compared to h_1 in Rounds 1 and 2, and compared and h_2 in Rounds 2 and 3. Hence, if $h_{mid} = h_3$, combining inequalities (D.2.2.a), (D.2.2.d), (D.2.2.c) and (D.2.2.e), we have the same preference restrictions as in section D.2.1. Inequality (D.2.2.b) becomes a redundant constraint.

If $h_{mid} \neq h_2$, we consider the rest two subcases separately.

1. Suppose $h_{mid} = h_1$. Combining Inequalities (D.2.2.a) and (D.2.2.d) leads to $v_{i,h_1,3} - v_{i,h_3,3} \leq \varepsilon_{i,h_3} - \varepsilon_{i,h_1} \leq v_{i,h_1,1} - v_{i,h_3,1}$. $\varepsilon_{i,h_3} - \varepsilon_{i,h_1}$ has a positive interval only when Restriction 1 is satisfied: $(\beta_{i1} - \beta_{i3})(p_{i,h_1} - p_{i,h_3}) \geq 0$.

   If Inequality (D.2.2.b) also holds, we must have the following:

$$\varepsilon_{i,h_3} - \varepsilon_{i,h_2} \leq v_{i,h_1,1} - v_{i,h_3,1} + v_{i,h_2,2} - v_{i,h_1,2} \qquad \text{(D.2.2.f)}$$

On the other hand, Inequalities (D.2.2.c) and (D.2.2.e) require:

$$v_{i,h_2,3} - v_{i,h_3,3} \leq \varepsilon_{i,h_3} - \varepsilon_{i,h_2} \leq v_{i,h_2,2} - v_{i,h_3,2} \tag{D.2.2.g}$$

We want to determine preference constraints such that (D.2.2.f) and (D.2.2.g) can fulfill simultaneously. The upper bound of (D.2.2.g) is naturally satisfied when $v_{i,h_1,1} - v_{i,h_3,1} + v_{i,h_2,2} - v_{i,h_1,2} \leq v_{i,h_2,2} - v_{i,h_3,2}$, which is equivalent to $(p_{i,h_1} - p_{i,h_3})(\beta_{i2} - \beta_{i1}) \geq 0$ (Restriction 2).

If Restriction 2 holds, Inequality (D.2.2.f) is binding. A positive interval for $\varepsilon_{i,h_3} - \varepsilon_{i,h_2}$ only exists when $v_{i,h_1,1} - v_{i,h_3,1} + v_{i,h_2,2} - v_{i,h_1,2} > v_{i,h_2,3} - v_{i,h_3,3}$, equivalent to:

$$(p_{i,h_2} - p_{i,h_3})\beta_{i3} \leq (p_{i,h_1} - p_{i,h_3})\beta_{i1} + (p_{i,h_2} - p_{i,h_1})\beta_{i2}$$

This existence condition is redundant when both Restrictions 1 and 2 hold. Combine Restrictions 1 and 2, we have:

$$\begin{cases} \beta_{i3} \leq \beta_{i1} \leq \beta_{i2} & \text{if } h_3 = h_{low} \\ \beta_{i3} \geq \beta_{i1} \geq \beta_{i2} & \text{if } h_3 = h_{high} \end{cases}$$

If Restriction 2 does not hold, Inequality (D.2.2.f) is redundant to (D.2.2.g), The existence condition for $\varepsilon_{i,h_3} - \varepsilon_{i,h_2}$ becomes $(p_{i,h_2} - p_{i,h_3})(\beta_{i2} - \beta_{i3}) \geq 0$ (Restriction 3). Combining Restrictions 1 and 3 leads to:

$$\begin{cases} \beta_{i3} \leq \beta_{i2} \leq \beta_{i1} & \text{if } h_3 = h_{low} \\ \beta_{i3} \geq \beta_{i2} \geq \beta_{i1} & \text{if } h_3 = h_{high} \end{cases}$$

Therefore, the subcase has a positive probability only when β_{i3} fulfill the following preference constraints:

$$\begin{cases} \beta_{i3} \leq \min\{\beta_{i1}, \beta_{i2}\} & \text{if } h_3 = h_{low} \\ \beta_{i3} \geq \max\{\beta_{i1}, \beta_{i2}\} & \text{if } h_3 = h_{high} \end{cases}$$

2. Suppose $h_{mid} = h_2$. Combining Inequalities (D.2.2.c) and (D.2.2.e) leads to $v_{i,h_2,3} - v_{i,h_3,3} \leq \varepsilon_{i,h_3} - \varepsilon_{i,h_2} \leq v_{i,h_2,2} - v_{i,h_3,2}$. The interval is non-zero when the following Re-

striction 1 is satisfied: $(\beta_{i2} - \beta_{i3})(p_{i,h_2} - p_{i,h_3}) \geq 0$.

If Inequality (D.2.2.b) also hold, we have the following:

$$\varepsilon_{i,h_3} - \varepsilon_{i,h_1} \geq v_{i,h_2,3} - v_{i,h_3,3} + v_{i,h_1,2} - v_{i,h_2,2} \tag{D.2.1.2.h}$$

At the same time, Inequalities (D.2.2.a) and (D.2.2.d) requires:

$$v_{i,h_1,3} - v_{i,h_3,3} \leq \varepsilon_{i,h_3} - \varepsilon_{i,h_1} \leq v_{i,h_1,1} - v_{i,h_3,1} \tag{D.2.2.i}$$

We want to set restrictions such that (D.2.1.2.h) and (D.2.2.i) can hold simultaneously. Notice that (D.2.1.2.h) is binding only when $v_{i,h_2,3} - v_{i,h_3,3} + v_{i,h_1,2} - v_{i,h_2,2} \geq v_{i,h_1,3} - v_{i,h_3,3}$, which is equivalent to $(p_{i,h_1} - p_{i,h_2})(\beta_{i2} - \beta_{i3}) \geq 0$. This condition is naturally satisfied with Restriction 1 and $h_{mid} = h_2$ hold. Hence, (D.2.1.2.h) determines the lower bound, and $\varepsilon_{i,h_3} - \varepsilon_{i,h_1}$ exists only if $v_{i,h_1,1} - v_{i,h_3,1} \geq v_{i,h_2,3} - v_{i,h_3,3} + v_{i,h_1,2} - v_{i,h_2,2}$, which is equivalent to Restriction 2:

$$(p_{i,h_2} - p_{i,h_3})\beta_{i3} \leq (p_{i,h_1} - p_{i,h_3})\beta_{i1} + (p_{i,h_2} - p_{i,h_1})\beta_{i2}$$

Combining Restrictions 1 and 2, we have the following preference constraints:

$$\begin{cases} \beta_{i3} \leq \min\left\{\beta_{i1}, \dfrac{(p_{i,h_1} - p_{i,h_3})\beta_{i1} + (p_{i,h_2} - p_{i,h_1})\beta_{i2}}{p_{i,h_2} - p_{i,h_3}}\right\} & \text{if } h_3 = h_{low} \\ \beta_{i3} \geq \max\left\{\beta_{i1}, \dfrac{(p_{i,h_1} - p_{i,h_3})\beta_{i1} + (p_{i,h_2} - p_{i,h_1})\beta_{i2}}{p_{i,h_2} - p_{i,h_3}}\right\} & \text{if } h_3 = h_{high} \end{cases}$$

D.2.3 $h_2 < J_1, J_1 < h_3 < J_2$

When h_3 is not inspected in the first round, it is not compared to h_1 with preference β_{1r}. $\{I_i\}$ needs to satisfy the following conditions:

$$\begin{cases} u_{i,h_1,1} \geq u_{i,h_2,1} \Rightarrow \varepsilon_{i,h_1} - \varepsilon_{i,h_2} \geq v_{i,h_2,1} - v_{i,h_1,1} \\ u_{i,h_2,2} \geq u_{i,h_1,2} \Rightarrow \varepsilon_{i,h_1} - \varepsilon_{i,h_2} \leq v_{i,h_2,2} - v_{i,h_1,2} \\ u_{i,h_2,2} \geq u_{i,h_3,2} \Rightarrow \varepsilon_{i,h_2} - \varepsilon_{i,h_3} \geq v_{i,h_3,2} - v_{i,h_2,2} \\ u_{i,h_3,3} \geq u_{i,h_1,3} \Rightarrow \varepsilon_{i,h_1} - \varepsilon_{i,h_3} \leq v_{i,h_3,3} - v_{i,h_1,3} \\ u_{i,h_3,3} \geq u_{i,h_2,3} \Rightarrow \varepsilon_{i,h_2} - \varepsilon_{i,h_3} \leq v_{i,h_3,3} - v_{i,h_2,3} \end{cases}$$

Notice that now, h_2 is compared to h_1 in rounds 1 and 2 and h_3 in rounds 2 and 3, while h_1 and h_2 are only compared in round 2. This subcase is a completely mirrored situation of Section D.2.2. We can verify by going through the process in D.2.2 and obtaining similar preference constraints, with only simple changes of exchanging β_{i2} and β_{i3} and h_2 and h_3. Eventually, the preference constraints are as follows: when $h_{mid} = h_2$, it follows the subcase in Section D.2.1 with the same constraints. When $h_{mid} = h_1$:

$$\begin{cases} \beta_{i2} \leq \min\{\beta_{i1}, \beta_{i3}\} & \text{if } h_3 = h_{low} \\ \beta_{i2} \geq \max\{\beta_{i1}, \beta_{i3}\} & \text{if } h_3 = h_{high} \end{cases}$$

when $h_{mid} = h_3$:

$$\begin{cases} \beta_{i2} \leq \min\left\{\beta_{i1}, \dfrac{(p_{i,h_1} - p_{i,h_2})\beta_{i1} + (p_{i,h_3} - p_{i,h_1})\beta_{i3}}{p_{i,h_3} - p_{i,h_2}}\right\} & \text{if } h_2 = h_{low} \\ \beta_{i2} \geq \max\left\{\beta_{i1}, \dfrac{(p_{i,h_1} - p_{i,h_2})\beta_{i1} + (p_{i,h_3} - p_{i,h_1})\beta_{i3}}{p_{i,h_3} - p_{i,h_2}}\right\} & \text{if } h_2 = h_{high} \end{cases}$$

For the convenience of implementation, we rewrite the preference constraints of the two subcases: when $h_{mid} = h_1$:

$$\begin{cases} (\beta_{i2} - \beta_{i1})(p_{i,h_2} - p_{i,h_1}) \geq 0 \\ (\beta_{i3} - \beta_{i2})(p_{i,h_3} - p_{i,h_2}) \geq 0 \end{cases}$$

when $h_{mid} = h_3$:

$$\begin{cases} (\beta_{i2} - \beta_{i1})(p_{i,h_2} - p_{i,h_1}) \geq 0 \\ (p_{i,h_1} - p_{i,h_3})\beta_{i3} \leq (p_{i,h_2} - p_{i,h_3})\beta_{i2} + (p_{i,h_1} - p_{i,h_2})\beta_{i1} \end{cases}$$

Hence, the constraints on β_{i2} are just related to β_{i1}.

D.2.4 $J_1 < h_2 < J_2, J_1 < h_3 < J_2$

In this subcase, both h_2 and h_3 are not compared to h_1 in the first inspection round. We eliminate two of the inequalities in Section D.2.1:

$$\begin{cases} u_{i,h_2,2} \geq u_{i,h_1,2} \Rightarrow \varepsilon_{i,h_1} - \varepsilon_{i,h_2} \leq v_{i,h_2,2} - v_{i,h_1,2} & \text{(D.2.4.a)} \\ u_{i,h_2,2} \geq u_{i,h_3,2} \Rightarrow \varepsilon_{i,h_2} - \varepsilon_{i,h_3} \geq v_{i,h_3,2} - v_{i,h_2,2} & \text{(D.2.4.b)} \\ u_{i,h_3,3} \geq u_{i,h_1,3} \Rightarrow \varepsilon_{i,h_1} - \varepsilon_{i,h_3} \leq v_{i,h_3,3} - v_{i,h_1,3} & \text{(D.2.4.c)} \\ u_{i,h_3,3} \geq u_{i,h_2,3} \Rightarrow \varepsilon_{i,h_2} - \varepsilon_{i,h_3} \leq v_{i,h_3,3} - v_{i,h_2,3} & \text{(D.2.4.d)} \end{cases}$$

Combining inequalities (D.2.2.b), (D.2.2.d), we get:

$$(\beta_{i2} - \beta_{i3})(p_{i,h_2} - p_{i,h_3}) > 0.$$

This is the preference constriant that preserves a non-zero domain for Ξ_i.

D.2.5 $h_2 < J_1, h_3 > J_2$

In this subcase, only h_1 and h_2 are compared in the first two rounds. Similar to the previous subcase, we have four inequalities for I_i:

$$\begin{cases} u_{i,h_1,1} \geq u_{i,h_2,1} \Rightarrow \varepsilon_{i,h_1} - \varepsilon_{i,h_2} \geq v_{i,h_2,1} - v_{i,h_1,1} & \text{(D.2.5.a)} \\ u_{i,h_2,2} \geq u_{i,h_1,2} \Rightarrow \varepsilon_{i,h_1} - \varepsilon_{i,h_2} \leq v_{i,h_2,2} - v_{i,h_1,2} & \text{(D.2.5.b)} \\ u_{i,h_3,3} \geq u_{i,h_1,3} \Rightarrow \varepsilon_{i,h_1} - \varepsilon_{i,h_3} \leq v_{i,h_3,3} - v_{i,h_1,3} & \text{(D.2.5.c)} \\ u_{i,h_3,3} \geq u_{i,h_2,3} \Rightarrow \varepsilon_{i,h_2} - \varepsilon_{i,h_3} \leq v_{i,h_3,3} - v_{i,h_2,3} & \text{(D.2.5.d)} \end{cases}$$

Inequalities (D.2.2.a) and (D.2.2.b) can hold simultaneously only when:

$$(\beta_{i1}-\beta_{i2})(p_{i,h_1}-p_{i,h_2})>0.$$

The other two inequalities can always be satisfied by proper-valued $\{\varepsilon_{i1}\}$ for any price sensitivities.

D.2.6 $h_2 > J_1, h_3 > J_2$

In this subcase, I_i only fulfills three inequalities:

$$\begin{cases} u_{i,h_1,1} \geq u_{i,h_2,1} \Rightarrow \varepsilon_{i,h_1}-\varepsilon_{i,h_2} \geq v_{i,h_2,1}-v_{i,h_1,1} \\ u_{i,h_3,3} \geq u_{i,h_1,3} \Rightarrow \varepsilon_{i,h_1}-\varepsilon_{i,h_3} \leq v_{i,h_3,3}-v_{i,h_1,3} \\ u_{i,h_3,3} \geq u_{i,h_2,3} \Rightarrow \varepsilon_{i,h_2}-\varepsilon_{i,h_3} \leq v_{i,h_3,3}-v_{i,h_2,3} \end{cases}$$

Notice that constraints on the preferences are not required. There always exists Ξ_i that fulfills the inequality conditions with any preferences given.

D.3 Three-round Search Process: Two or One checkout products

Consumer i may check out a product multiple times in her search process. If the checkout product is the same in two or more inspection rounds, we eliminate the inequality conditions on the perceived checkout product utilities across these rounds.

D.3.1 $h_1, h_2, h_3 \leq J_1$ with Two out of Three Identical

Denote the two inspection rounds with the same checkout product by s_1 and s_2 and the other inspection round by o. We eliminate two conditions in Section D.2.1 because they degenerate into two equalities. The remaining four inequalities must still be satisfied simultaneously.

$$\begin{cases} u_{i,h_{s_1},s_1} \geq u_{i,h_o,s_1} \Rightarrow \varepsilon_{i,h_o}-\varepsilon_{i,h_{s_1}} \leq v_{i,h_{s_1},s_1}-v_{i,h_o,s_1} \\ u_{i,h_{s_2},s_2} \geq u_{i,h_o,s_2} \Rightarrow \varepsilon_{i,h_o}-\varepsilon_{i,h_{s_2}} \leq v_{i,h_{s_2},s_2}-v_{i,h_o,s_2} \\ u_{i,h_o,o} \geq u_{i,h_{s_1},o} \Rightarrow \varepsilon_{i,h_o}-\varepsilon_{i,h_{s_1}} \geq v_{i,h_{s_1},o}-v_{i,h_o,o} \\ u_{i,h_o,o} \geq u_{i,h_o,s_2} \Rightarrow \varepsilon_{i,h_o}-\varepsilon_{i,h_{s_2}} \leq v_{i,h_{s_2},o}-v_{i,h_o,o} \end{cases}$$

To ensure the existence of $\varepsilon_{i,h_o} - \varepsilon_{i,h_{s_1}}$, the following preference restrictions must hold:

$$\begin{cases} (\beta_{i,s_1} - \beta_{i,o})(p_{i,h_{s_1}} - p_{i,h_o}) \geq 0 \\ (\beta_{i,s_2} - \beta_{i,o})(p_{i,h_{s_2}} - p_{i,h_o}) \geq 0 \end{cases}$$

Since $p_{i,h_{s_1}} = p_{i,h_{s_2}}$, the above conditions imply that β_{i,s_1} and β_{i,s_2} should be simultaneously greater than or less than $\beta_{i,o}$.

D.3.2 $h_1 = h_2 \leq J_1, J_1 < h_3 \leq J_2$

The comparison between h_3 and $h_1 = h_2$ occurs in the second and third inspection rounds. The inequalities for the perceived utilities are:

$$\begin{cases} u_{i,h_2,2} \geq u_{i,h_3,2} \Rightarrow \varepsilon_{i,h_2} - \varepsilon_{i,h_3} \geq v_{i,h_3,2} - v_{i,h_2,2} \\ u_{i,h_3,3} \geq u_{i,h_2,3} \Rightarrow \varepsilon_{i,h_2} - \varepsilon_{i,h_3} \leq v_{i,h_3,3} - v_{i,h_2,3} \end{cases}$$

The two conditions can hold simultaneously only when:

$$(\beta_{i2} - \beta_{i3})(p_{i,h_2} - p_{i,h_3}) > 0.$$

with $p_{i,h_1} = p_{i,h_2}$.

D.3.3 $h_1 = h_3 \leq J_1, J_1 < h_2 \leq J_2$

Product h_2 and product $h_1 = h_3$ are compared under the preferences β_{i2} and β_{i3}. The conditions for the perceived utilities are:

$$\begin{cases} u_{i,h_2,2} \geq u_{i,h_1,2} \Rightarrow \varepsilon_{i,h_2} - \varepsilon_{i,h_1} \geq v_{i,h_1,2} - v_{i,h_2,2} \\ u_{i,h_3,3} \geq u_{i,h_2,3} \Rightarrow \varepsilon_{i,h_2} - \varepsilon_{i,h_3} \leq v_{i,h_3,3} - v_{i,h_2,3} \end{cases}$$

$\varepsilon_{i,h_2} - \varepsilon_{i,h_3}$ can satisfy both conditions only when:

$$(\beta_{i2} - \beta_{i3})(p_{i,h_2} - p_{i,h_3}) > 0.$$

with $p_{i,h_1} = p_{i,h_3}$.

D.3.4 $J_1 < h_2 = h_3 \leq J_2$

Product h_2 is not compared to h_1 in the first inspection round. The perceived utility inequalities reduce to:

$$\begin{cases} u_{i,h_2,2} \geq u_{i,h_1,2} \Rightarrow \varepsilon_{i,h_2} - \varepsilon_{i,h_1} \geq v_{i,h_1,2} - v_{i,h_2,2} \\ u_{i,h_3,3} \geq u_{i,h_1,3} \Rightarrow \varepsilon_{i,h_3} - \varepsilon_{i,h_1} \geq v_{i,h_1,3} - v_{i,h_3,3} \end{cases}$$

with $h_2 = h_3$. The two inequalities can be satisfied simultaneously as long as $\varepsilon_{i,h_2} - \varepsilon_{i,h_1}$ is large enough. We do not need to impose additional preference constraints.

D.3.5 $h_1 = h_2 \leq J_1, h_3 > J_2$

The only equality in this subcase is that product h_3 surpasses h_1 in the last inspection round:

$$u_{i,h_3,3} \geq u_{i,h_1,3} \Rightarrow \varepsilon_{i,h_3} - \varepsilon_{i,h_1} \geq v_{i,h_1,3} - v_{i,h_3,3}$$

with $h_1 = h_2$. The condition can hold with any price preferences.

D.3.6 One checkout product

In this last case, we have $h_1 = h_2 = h_3$. There are no other checkout products to be compared with, and no inequality conditions on perceived checkout product utilities.

D.4 Preference Discovery Distribution

In this section, we show how the preference restrictions are imposed in preference discovery. The conditional distribution of the preference posterior mean $\beta_{i,r+1}$ given the prior mean $\beta_{i,r}$, the true preference β_i, the prior variance σ_r^2 and the signal variance σ_s^2 is given by:

$$\beta_{i,r+1} = \beta_{i,r} + \frac{\tau_r^2}{1+\tau_r^2}(\beta_i^s - \beta_{i,r}) = \frac{1}{1+\tau_r^2}\beta_{i,r} + \frac{\tau_r^2}{1+\tau_r^2}\beta_i^s$$

Hence, the conditional expectation and variance of of $\beta_{i,r+1}$ is given by:

$$\begin{aligned} \mathrm{E}(\beta_{i,r+1}|\beta_i) &= \frac{1}{1+\tau_r^2}\beta_{i,r} + \frac{\tau_r^2}{1+\tau_r^2}\beta_i \equiv \tilde{\beta}_{i,r} \\ \mathrm{Var}(\beta_{i,r+1}|\beta_i) &= \left(\frac{\tau_r^2}{1+\tau_r^2}\sigma_s\right)^2 \equiv \tilde{\sigma}_{s,r}^2 \end{aligned}$$

The conditional distribution of $\beta_{i,r+1}$ is:

$$\beta_{i,r+1}|\beta_{i,r} \sim \mathcal{N}\left(\tilde{\beta}_{i,r},\ \tilde{\sigma}^2_{s,r}\right)$$

The preference discovery process is equivalent to recursively drawing the posterior mean from the conditional distribution above. When preference restrictions are imposed, the underlying distribution becomes a truncated normal distribution with the same parameters of $\{\tilde{\beta}_{i,r},\ \tilde{\sigma}^2_{s,r}\}$

D.5 Simulated Likelihood Construction

Following the restrictions and inequalities in section D, we implement our estimator with a Geweke-Hajivassilou-Keane (GHK)-style simulated likelihood. A GHK-style simulator recursively makes random draws while enforcing inequality conditions on the randomnesses imposed by the previous draws. Jiang et al. (2021) and Chung et al. (2024) introduced the GHK simulator of a Weitzman-style search model. Compared to their models, the selected product in our model is replaced with the checkout product in all inspection rounds, and the preference discovery process constrains the perceived utilities of checkout products. The construction of the simulated likelihood can be summarized as follows:

1. Make n_d draws of β_{i1}, denoted by β_{i1}^d.

2. If there are at least two inspection rounds, draw β_{i2} conditional on β_{i1}^d following the distribution in D.3.4. Impose the following truncation conditions:

   - In default cases, the upper bound $\beta_{i2}^{u,d} = +\infty$, the lower bound $\beta_{i2}^{\ell,d} = -\infty$.
   - In cases D.1.2, D.2.1, D.2.2 with $h_{mid} = h_3$, D.2.3 and D.3.1 with $h_1 \neq h_2$: If $p_{i,h_2} > p_{i,h_1}$, set the lower bound $\beta_{i2}^{\ell,d} = \beta_{i1}^d$: if $p_{i,h_2} < p_{i,h_1}$, set the upper bound $\beta_{i2}^{u,d} = \beta_{i1}^d$

   Obtain β_{i2}^d. Compute $\Pr(\beta_{i2}|\beta_{i1}^d)$, the probability that β_{i2}^d falls in the interval $\{\beta_{i2}^{\ell,d}, \beta_{i2}^{u,d}\}$.

3. If there are at least three inspection rounds, draw β_{i3} for each draw conditional on β_{i2}^d and β_{i1}^d following the distribution in D.4. Similarly to the previous step, impose the following truncation conditions on the distribution by defining an upper and lower truncation

boundary. The default boundaries are $\beta_{i3}^{u,d} = +\infty$ and $\beta_{i3}^{\ell,d} = -\infty$. Adopt the corresponding restrictions in each case in D.2 and D.3 to replace the boundary conditions, and draw β_{i3} from the truncated conditional distribution.

Obtain β_{i3}^d. Compute $\Pr(\beta_{i3}|\beta_{i2}^d, \beta_{i1}^d)$, the probability that β_{i3}^d falls in the interval $\{\beta_{i3}^{\ell,d}, \beta_{i3}^{u,d}\}$.

4. Denote the ordered set of preferences drawn in steps 1, 2, and 3 by B_i^d, $B_i^d = \{\beta_{i1}^d\}$ (one inspection round), $\{\beta_{i1}^d, \beta_{i2}^d\}$ (two inspection rounds) or $\{\beta_{i1}^d, \beta_{i2}^d, \beta_{i3}^d\}$ (three inspection rounds). Compute $\Pr(B_i) = \Pr(\beta_{i2}|\beta_{i1}^d)$ for two-round sequences, and $\Pr(B_i) = \Pr(\beta_{i2}|\beta_{i1}^d)\Pr(\beta_{i3}|\beta_{i2}^d, \beta_{i1}^d)$ for three-round sequences. $\Pr(B_i) = 1$ if there is only one inspection round.

5. For each draw d, given drawn preferences B_i^d, make random draws of $\Xi_i = \{\varepsilon_{i,h_1}\}$ (one inspection round), $\{\varepsilon_{i,h_1}, \varepsilon_{i,h_2}\}$ (two inspection rounds) or $\{\varepsilon_{i,h_1}, \varepsilon_{i,h_2}, \varepsilon_{i,h_3}\}$ (three inspection rounds) with respect to the inequality conditions on I_i in each specific case in D.1 and D.2. One can realize it by drawing some ε first, and draw other εs conditional on the previous draws. For example, in D.1.2, we can first draw ε_{i,h_1}^d without any restrictions, then draw ε_{i,h_2}^d conditional on ε_{i,h_1}^d. The orders for drawing perceived utilities in three-round search sequences are listed below.

   - In D.2.1, draw $\varepsilon_{i,h_{mid}}^d$ first, then draw $\varepsilon_{i,h_{high}}^d$ and $\varepsilon_{i,h_{low}}^d$ conditional on $\varepsilon_{i,h_{mid}}^d$. As discussed, when $\varepsilon_{i,h_{mid}}^d - \varepsilon_{i,h_{low}}^d$ and $\varepsilon_{i,h_{high}}^d - \varepsilon_{i,h_{mid}}^d$ satisfy the inequality conditions, $\varepsilon_{i,h_{high}}^d - \varepsilon_{i,h_{low}}^d$ will also satisfy the condition with preference constraints held.
   - In D.2.2 and D.2.3. There are only binding constraints on $\varepsilon_{i,h_3} - \varepsilon_{i,h_1}$ and $\varepsilon_{i,h_3} - \varepsilon_{i,h_2}$. We draw ε_{i,h_3} first, then draw ε_{i,h_1} and ε_{i,h_2} conditional on ε_{i,h_3}^d.
   - In D.2.4 and D.2.5, draw ε for the products inspected in the same round following D.3.1. ε of the product inspected in another round is drawn lastly.
   - In D.2.6, draw $\varepsilon_{i,h_1}, \varepsilon_{i,h_2}, \varepsilon_{i,h_3}$ sequentially.
   - In D.3, draw ε for the checkout products in any order.

Obtain Ξ_i^d for each draw. Compute $\Pr(\Xi_i|B_i^d)$ the probability that Ξ_i satisfy the inequality conditions by taking the product of conditional probabilities of ε_{ir}. Combining Ξ_i^d and

B_i^d, we get the draw of the information set I_i^d.

6. With I_i^d, compute $u_{i,h_r,r}^d$ for all inspection rounds r and $v_{i,j,r}$ for all products j in the market and all inspection rounds r. Compute $\text{Prob}_{i,disc}^d = \Pr(B_i)\Pr(\Xi_i|B_i^d)$

7. For each d, draw one c_{i,J_r}^d for each inspection round r:

   - If $h_r < J_r$, draw $c_{i,J_r,r}^d$ with truncation condition $c_{i,J_r,r}^d \leq m^{-1}(u_{i,h_r,r}^d - v_{i,J_r,r}^d)$. Compute $\Pr(c_{i,J_r,r}^d \leq m_\varepsilon^{-1}(u_{i,h_r,r}^d - v_{i,J_r,r}^d))$.
   - If $h_r = J_r$, randomly draw c_{i,J_r}^d. Set $\Pr(c_{i,J_r,r}^d \leq m_\varepsilon^{-1}(u_{i,h_r,r}^d - v_{i,J_r,r}^d)) = 1$

   In each inspection round r, compute $y_{ir}^d = \min\{u_{i,h_r,r}, z_{i,J_r,r}\}$. Compute $\text{Prob}_{i,check}^d = \prod_r \Pr(c_{i,J_r,r}^d \leq m_\varepsilon^{-1}(u_{i,h_r,r}^d - v_{i,J_r,r}^d))$ for all inspection rounds r.

8. For each inspection round r with $J_{r+1} > J_r + 1$, draw $c_{i,J_{r+1}-1,r}^d, \cdots, c_{i,J_r+1,r}^d$ recursively such that $c_{i,j+1,r}^d \leq m_\varepsilon^{-1}(v_{i,j,r}^d + m_\varepsilon(c_{i,j,r}^d) - v_{i,j+1,r}^d) = m_\varepsilon^{-1}(z_{i,j,r}^d - v_{i,j+1,r}^d)$. The condition ensures that the perceived reservation values are ranked in descending order of the search process within each inspection round. Compute $\text{Prob}_{i,rank}^d = \prod_j \Pr(c_{i,j+1,r}^d \leq m_\varepsilon^{-1}(z_{i,j,r}^d - v_{i,j+1,r}^d))$ for each $j \in \cup^r \{J_r + 1, \cdots, J_{r+1} - 1\}$. If $J_{r+1} \leq J_r + 1$ for all inspection round r, set $\text{Prob}_{i,rank}^d = 1$.

9. For each inspection round r, compute $\text{Prob}_{i,inspect}^d = \prod_j \Pr(\varepsilon_{i,j} \leq \min_r y_{ir}^d - v_{i,j,r}^d)$ for all j inspected but not checked out in any inspection rounds r. If no product is inspected but not checked out, set $\text{Prob}_{i,inspect}^d = 1$.

10. For each inspection round r, compute $\text{Prob}_{i,stop}^d = \prod_r \prod_{k_r} \Pr(c_{i,k_r,r} \geq m_\varepsilon^{-1}(y_{ir}^d - v_{i,k_r,r}^d)$ for all inspection rounds r and all product k_r not inspected at the end of the inspection round r. If $J_1 = |\mathcal{M}_i|$, set $\text{Prob}_{i,stop}^d = 1$.

11. Take the product of all probabilities above for each draw:

$$\mathcal{L}_i^d = \text{Prob}_{i,disc}^d \cdot \text{Prob}_{i,check}^d \cdot \text{Prob}_{i,rank}^d \cdot \text{Prob}_{i,inspect}^d \cdot \text{Prob}_{i,stop}^d$$

   and calculate the mean across all draws as the simulated likelihood of consumer i.

E Monte Carlo Simulation Results

TABLE E.2 – Monte Carlo simulation results

	True Value	Estimates
γ_1	0.5	0.5002
γ_2	0.2	0.2018
γ_3	-0.2	-0.1992
$\bar{\beta}$	-1.2	-1.1974
δ	0.6	0.5975
$\sigma_{\bar{\beta}}$	0.2	0.2035
σ_0	0.25	0.2523
τ_1	0.8	0.7974
σ_s	0.4	0.3783
$\mu_0^{outside}$	-1.6	-1.6039
$\xi^{outside}$	1.0	1.0265
$\bar{c}$:	-0.75	-0.7504
σ_c:	0.25	0.2527
D	2500	2500
Log-Likelihood	-258975	-258969
N	50000	50000

Notes: The Monte-Carlo simulation assumes a market of 8 products and an outside option. Products are assigned with individual-specific prices.

F Proof of Identification

Our model identifies preference and search cost parameters similar to those in the literature with Weitzman-style search models. The difference is that the preferences are identified within each inspection round, which is used to identify the parameters in the preference discovery process. The individual-level revealed preference and parameters can be identified from a consumer's observed inspection and checkout decisions. Given that these individual-level parameters have been identified, the distributions of the parameters are known to us, and the parametric preference discovery process can be identified. The parametric functional form is assumed prerogatively because consumers in our model may terminate their search process due to within-round checkout decisions, which exogenously determines the truncation of the preference discovery.

Therefore, a nonparametric identification is not available but a parametric form preference discovery process can still be identified.

F.1 Identification within Inspection Rounds

The proof of within-round identification closely follows **?**. Consider the perceived reservation value of product j to consumer i:

$$z_{ijr} = x_j\gamma + p_{ij}\beta_{ir} + m_\varepsilon\left(c_{ijr}\right)$$

Here, c_{ijr} follows a log-normal distribution with variance known from the assumption, and $m_\varepsilon(.)$ monotonically decreases with a known function form. Therefore, we can identify the mean and the variance of c_{ijr} as long as $m_\varepsilon(c_{ijr})$ is identified. We denote $\zeta = m_\varepsilon(c_{ijr})$ and rewrite the perceived reservation value as follows:

$$z_{ijr} = x_j\gamma + p_{ij}\beta_{ir} + \zeta_{ijr} \equiv v_{ijr} + \zeta_{ijr}$$

First, we consider the case where the checkout product h_r is not the last product inspected in the current inspection round. We represent perceived reservation values of all products and perceived utilities of inspected products in stacked vectorized forms as follows:

$$\begin{gathered}
\boldsymbol{z}_{ir}^{k} = \left(z_{i,J_r,r}, z_{i,J_r-1,r}, \cdots, z_{i,J_{r-1}+1,r}\right)^\top, \quad \boldsymbol{z}_{ir}^{u} = \left(z_{i,J_r+1}, \cdots, z_{i,|\mathcal{M}|,r}\right)^\top, \\
\boldsymbol{u}_{ir}^{k'} = \left(u_{i,1,r}, \cdots, u_{i,h_r-1,r}, u_{i,h_r+1,r}, \cdots u_{i,J_r,r}\right)^\top \\
\boldsymbol{z}_{ir}^{k} = \boldsymbol{v}_{ir}^{k} + \boldsymbol{\zeta}_{ir}^{k}, \quad \boldsymbol{z}_{i}^{u} = \boldsymbol{v}_{ir}^{u} + \boldsymbol{\zeta}_{ir}^{u}, \quad \boldsymbol{u}_{i}^{k'} = \boldsymbol{v}_{ir}^{k} + \boldsymbol{\varepsilon}_{ir}^{k}
\end{gathered}$$

If no product is inspected in round r, $\boldsymbol{z}_{ir}^{k}$ is an empty vector. Conditional on products $1, 2, \cdots, J_{r-1}$ are inspected before round r, the probability of observing consumer i inspect

$J_{r-1}+1,\cdots,J_r$ and checkout product h_r is given by:

$$\text{Prob}(\text{round } r) = \Pr\left(\underbrace{\hat{D}}_{(J_r-J_{r-1}+|\mathcal{M}_i|-1)\times(J_r-J_{r-1}+|\mathcal{M}_i|)} \begin{pmatrix} u_{i,h_r,r} \\ \boldsymbol{z}_{ir}^k \\ \boldsymbol{z}_{ir}^u \\ \boldsymbol{u}_{ir}^{k'} \end{pmatrix}_{(J_r-J_{r-1}+|\mathcal{M}|)\times 1} \leq 0 \right)$$

$$= \Pr\left(\hat{D} \begin{pmatrix} \varepsilon_{i,h_r} \\ \zeta_{ir}^k \\ \zeta_{ir}^u \\ \varepsilon_{ir}^k \end{pmatrix} \leq -\hat{D} \begin{pmatrix} \zeta_{ir}^k \\ \boldsymbol{v}_{ir}^k \\ \boldsymbol{v}_{ir}^u \\ \boldsymbol{v}_{ir}^{k'} \end{pmatrix} \right), \quad \text{where } \hat{D} = \begin{pmatrix} \hat{D}_1 & \hat{D}_2 \\ \hat{D}_3 & \hat{D}_4 \end{pmatrix}$$

Consisting four parts, the difference matrix D is of rank $J_r - J_{r-1} + |\mathcal{M}_i| - 1$:

$$\hat{D}_1 = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -1 & 0 \\ 0 & 0 & \cdots & 0 & 1 & -1 \end{pmatrix}_{(J_r-J_{r-1})\times(J_r-J_{r-1}+1)}, \quad \hat{D}_2 = \{0\}_{(J_r-J_{r-1})\times(|\mathcal{M}_i|-1)}$$

$$\hat{D}_3 = \begin{pmatrix} -1 & 0 & \cdots & 0 \\ -1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & 0 & \cdots & 0 \end{pmatrix}_{(|\mathcal{M}_i|-1)\times(J_r-J_{r-1}+1)}, \quad \hat{D}_4 = I_{(|\mathcal{M}_i|-1)\times(|\mathcal{M}_i|-1)}.$$

Following Train (2009), since we know the distribution of ε_{ij} as a scale normalization, identification for a within-round search process holds with a location normalization as Berry and Haile (2014) specified in a perfect information environment.

Then, we consider the case where $h_r = J_r$ and $J_r > J - r - 1$. The only difference is that we now take out $z_{i,J_r,r}$ because we cannot identify the relative size between $z_{i,J_r,r}$ and $u_{i,J_r,r}$. However, we know that both values are larger than the perceived utilities of inspected products and perceived reservation values of uninspected products. Hence, we follow all vectorized expressions and express the probability of observing consumer i's search and checkout decisions in round r as:

$$\text{Prob}(\text{round } r) = \Pr\left(\underbrace{\tilde{D}}_{(J_r - J_{r-1} + 2\cdot|\mathcal{M}_i| - 3)\times(J_r - J_{r-1} + |\mathcal{M}_i|)} \begin{pmatrix} u_{i,h_r,r} \\ \boldsymbol{z}_{ir}^k \\ \boldsymbol{z}_{ir}^u \\ \boldsymbol{u}_{ir}^{k'} \end{pmatrix}_{(J_r - J_{r-1} + |\mathcal{M}|)\times 1} \leq 0 \right)$$

$$= \Pr\left(\tilde{D} \begin{pmatrix} \varepsilon_{i,h_r} \\ \boldsymbol{\zeta}_{ir}^k \\ \boldsymbol{\zeta}_{ir}^u \\ \boldsymbol{\varepsilon}_{ir}^k \end{pmatrix} \leq -\tilde{D} \begin{pmatrix} \boldsymbol{\zeta}_{ir}^k \\ \boldsymbol{v}_{ir}^k \\ \boldsymbol{v}_{ir}^u \\ \boldsymbol{v}_{ir}^{k'} \end{pmatrix} \right), \quad \text{where } D = \begin{pmatrix} \tilde{D}_1 & \tilde{D}_2 \\ \tilde{D}_3 & \tilde{D}_4 \\ \tilde{D}_5 & \tilde{D}_6 \end{pmatrix}$$

The difference matrix $\tilde{D}$ consists of six parts:

$$\tilde{D}_1 = \begin{pmatrix} 0 & 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 & -1 \end{pmatrix}_{(J-1)\times(J+1)}, \quad D_2 = \{0\}_{(J-1)\times(|\mathcal{M}|-1)},$$

$$\tilde{D}_3 = \begin{pmatrix} -1 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & 0 & 0 & \cdots & 0 \end{pmatrix}_{(|\mathcal{M}|-1)\times(J+1)}, \quad \tilde{D}_4 = I_{(|\mathcal{M}|-1)\times(|\mathcal{M}|-1)}$$

$$\tilde{D}_5 = \begin{pmatrix} 0 & -1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & -1 & 0 & \cdots & 0 \end{pmatrix}_{(|\mathcal{M}|-1)\times(J+1)}, \quad \tilde{D}_6 = I_{(|\mathcal{M}|-1)\times(|\mathcal{M}|-1)}$$

$\tilde{D}$ is also of rank $J_r - J_{r-1} + |\mathcal{M}_i| - 1$, one smaller than the number of errors and equal to the number of error differences. The model remains identified as in the previous case.

The way of computing the joint probability is very similar. Instead of imposing the inequality $\varepsilon_{ij} \leq \min\{u_{i,h_r,r}, z_{i,J_r,r}\} - v_{ijr}$, for all product j inspected but never taken to the checkout, we impose $\varepsilon_{ij} \leq \min_r\{\min\{u_{i,h_r,r}, z_{i,J_r,r}\} - v_{ijr}\}$ for all r in which product j has been inspected. By doing this, only one distinctive restriction on ε_{ij} is binding, and the model stays identified.

F.2 Identification across Inspection Rounds

From the within-round estimation, we identify consumers' price preferences individually. Because we only have limited observations on three inspection rounds, nonparametric identification of the preference discovery process is unavailable. However, if we focus on a normal Bayesian learning process, the learning parameters can be just identified with the distribution parameters of prior preferences. Remember the distributions of the real preference, the prior belief, the prior mean, and the preference signal:

$$\text{Real preferences: } \beta_i \sim \mathcal{N}(\bar{\beta}, \sigma_\beta^2)$$
$$\text{Prior belief: } \mathcal{N}(\beta_{i1}, (\tau_1\sigma_s)^2)$$
$$\text{Prior mean: } \beta_{i1} \sim \mathcal{N}(\beta_i + \delta, \sigma_0^2)$$
$$\text{Preference signals } \beta_{ir}^s \sim \mathcal{N}(\beta_i, \sigma_s^2)$$

For the real preference, the prior mean and the preference signal, the difference between the realized draw and the mean of the distribution is denoted by $\Delta_{i\beta}$, Δ_{i0} and Δ_{ir} for the signal received at the end of inspection round r. Consumer i's prior mean, the mean of prior means across consumers, and the variance of prior means across consumers in round 1 are as follows:

$$\beta_{i1} = \bar{\beta} + \Delta_{i\beta} + \Delta_{i0} + \delta$$
$$\bar{\beta}_1 = \bar{\beta} + \delta$$
$$\sigma_{\beta_1}^2 = \sigma_\beta^2 + \sigma_0^2$$

Notice that the prior mean is the preference that consumer i acts on in round 1, and the mean of prior means is the observed mean of the preference the sample acts on in round 1.

After entering the checkout page, consumers receive a preference signal. Following DeGroot (1970), consumer i's prior mean, the mean of prior means across consumers, and the variance of prior means across consumers in round 2 are as follows:

$$\begin{aligned}\beta_{i2} &= \frac{1}{\tau_1^2+1}\beta_{i1} + \frac{\tau_1^2}{\tau_1^2+1}\beta_{i1}^s \\ &= \frac{1}{\tau_1^2+1}\left(\bar{\beta}+\Delta_{i\beta}+\Delta_{i0}+\delta\right) + \frac{\tau_1^2}{\tau_1^2+1}\left(\bar{\beta}+\Delta_{i\beta}+\Delta_{i1}\right) \\ &= \bar{\beta}+\Delta_{i\beta}+\frac{1}{\tau_1^2+1}\left(\Delta_{i0}+\delta\right)+\frac{\tau_1^2}{\tau_1^2+1}\Delta_{i1} \\ \bar{\beta}_2 &= \bar{\beta}+\frac{1}{\tau_1^2+1}\cdot\delta \\ \sigma_{\beta_2}^2 &= \sigma_\beta^2+\left(\frac{1}{\tau_1^2+1}\right)^2\sigma_0^2+\left(\frac{\tau_1^2}{\tau_1^2+1}\right)^2\sigma_s^2\end{aligned}$$

In the second inspection round, some consumers purchase products taken to the checkout in the first round. Because these consumers do not perform a third round, the preference distribution we observed in the third inspection round is censored. We recover the distribution of the preferences via simulation, and represent the recovered mean and variance of the sample preference similar to the previous rounds.

$$\begin{aligned}\beta_{i3} &= \frac{1}{2\cdot\tau_1^2+1}\beta_{i1} + \frac{\tau_1^2}{2\cdot\tau_1^2+1}\beta_{i1}^s + \frac{\tau_1^2}{2\cdot\tau_1^2+1}\beta_{i2}^s \\ &= \frac{1}{2\cdot\tau_1^2+1}\left(\bar{\beta}+\Delta_{i\beta}+\Delta_{i0}+\delta\right) + \frac{\tau_1^2}{2\cdot\tau_1^2+1}\left(\bar{\beta}+\Delta_{i\beta}+\Delta_{i1}\right) + \frac{\tau_1^2}{2\cdot\tau_1^2+1}\left(\bar{\beta}+\Delta_{i\beta}+\Delta_{i2}\right) \\ &= \bar{\beta}+\Delta_{i\beta}+\frac{1}{2\cdot\tau_1^2+1}\left(\Delta_{i0}+\delta\right)+\frac{\tau_1^2}{2\cdot\tau_1^2+1}\Delta_{i1}+\frac{\tau_1^2}{2\cdot\tau_1^2+1}\Delta_{i2} \\ \bar{\beta}_2 &= \bar{\beta}+\frac{1}{2\cdot\tau_1^2+1}\cdot\delta \\ \sigma_{\beta_2}^2 &= \sigma_\beta^2+\left(\frac{1}{2\cdot\tau_1^2+1}\right)^2\sigma_0^2+2\cdot\left(\frac{\tau_1^2}{2\cdot\tau_1^2+1}\right)^2\sigma_s^2\end{aligned}$$

With means and variances in three inspection rounds observed and the preference in each round individually identified, the six parameters that describe the preference discovery process are just identified under the given functional form.

Notice that the functional form of the preference discovery depends on the researchers' choice. Setting a more flexible functional form for the Bayesian learning process is also possible. However, a more flexible learning process usually relates to more parameters. To identify the additional parameters, one may need more moment conditions other than the mean and variances of the preferences in the first three inspection rounds, which could be difficult to obtain in a preference discovery process.