Rational Inattention as an Empirical Framework:

Application to the Welfare Effects of New-Product Introduction[*][†]

Joonhwi Joo[‡]

April 1, 2020

Abstract

In the conventional economic models of discrete choice where consumers are fully informed about the choice alternatives, consumer welfare increases whenever more alternatives are added to the choice set. However, when consumers face incomplete information about the set of available products, consumer welfare can either rise or fall when more products are added to the choice set. Incompletely informed consumers can choose an inferior, new alternative in the choice set, and thus the welfare implications of increasing variety becomes an empirical question. To account for the role of consumers' incomplete information about the choice alternatives, this paper develops a novel empirical framework of discrete-choice built upon a new interpretation of the Rational Inattention (RI) theory Matějka and McKay (2015); Fosgerau et al. (2019). The empirical framework is applied to a case of the consumer welfare effects from the introduction of the new Tide Pods product into the market in 2012. I find Tide Pods' entry reducted consumer welfare, on average, because its pricing was higher than the benefit provided to consumers. A counterfactual pricing experiment suggests a 6%-7% price cut of Tide Pods will make an average consumer just as well off as prior to Tide Pods' entry.

[*]I am grateful to my advisors, Ali Hortaçsu, Jean-Pierre Dubé, Pradeep K. Chintagunta, and Doron Ravid, for their guidance and support. I thank Filip Matějka, Kyeongbae Kim, Linda M. Schilling, Michael Grubb, Sanjog Misra, Inyoung Chae, Nanda Kumar, Bart J. Bronnenberg, Mingyu Joo, Ram C. Rao, Lawrence Y. Chung, seminar participants at the University of Chicago, UBC Sauder, UT Dallas Jindal, UCL School of Management, IIOC 2017, INFORMS Marketing Science 2018, ESAM 2019, and 2nd DFW Marketing Junior Symposium for suggestions and discussions. This study was supported in part by the facility and staff at the University of Chicago Social Sciences Computing Service. This paper includes figures calculated or derived based on data from The Nielsen Company (US), LLC, Copyright 2018, provided by the Kilts Center for Marketing at the University of Chicago Booth School of Business. The conclusions drawn from the Nielsen data are mine and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

[†]This paper subsumes the job market paper version "Quantity-Surcharged Larger Package Sales as Rationally Inattentive Consumers' Choice."

[‡]Naveen Jindal School of Management, University of Texas at Dallas. `joonhwi.joo@utdallas.edu`

1 Introduction

A typical Costco store carries around 4,000 SKUs, far less than a typical Walmart Supercenter, which carries around 120,000 SKUs. Costco has developed a reputation for high customer satisfaction through its superior product curation and the provision of preferred assortments of goods with low prices.[1] Costco's reputation suggests adding more product alternatives may not always increase consumer well-being per se. Whether the increasing variety of products actually benefits consumers is an empirical question, the answer to which depends on the composition of the choice set and how well consumers are informed about it. Recent evidence from lab and field experiments have confirmed how, in some contexts, adding more choice alternatives may not always be beneficial to consumers (e.g., Broniarczyk et al., 1998; Iyengar and Lepper, 2000; Iyengar et al., 2004; Bertrand et al., 2010). However, microeconomic models with fully informed consumers unambiguously predict the addition of more alternatives to the choice set increases consumer welfare in both the classic Walrasian demand framework and the discrete-choice random utility models (RUM henceforth, McFadden, 1974, 1978, 1981; McFadden and Richter, 1991; Small and Rosen, 1981; Herriges and Kling, 1999; Dagsvik and Karlström, 2005; Bhattacharya, 2015). A gap exists between a growing base of empirical evidence and what can be evaluated using existing full-information empirical demand frameworks.

This paper evaluates the consumer-welfare gains of Tide Pods laundry detergents' introduction into the market in 2012. To ensure consumer welfare is not mechanically predicted to increase with variety, I use a novel choice-modeling framework that accounts for consumers' incomplete information over the set of available product choices. Building on the rational inattention (RI) framework of Matějka and McKay (2015); Fosgerau et al. (2019), the model allows for both the endogenous acquisition of information and purchase choices. A sub-optimal purchase decision may occur when the acquired information about the set of available products is incomplete. Accordingly, the addition of more product alternatives to the choice set can potentially reduce consumer welfare.

On each choice occasion, a consumer confronts an exogenous choice set of products with deterministic utilities that are unknown prior to gathering information. The consumer then needs to "learn" about her/his alternative-specific consumption utilities before making the purchase decision. Some learning is costless, such as information generated through promotions or own purchase history. Such costless learning enables the consumer to form a subjective prior belief about each product's consumption utility, which is not necessarily unbiased. To capture the costly aspects of learning, I assume the information-gathering cost is proportional to the expected generalized entropy reduction between the prior and posterior belief distribution over each product's consumption utilities. The cost of information increases with the degree of precision. Thus, it is never optimal for a consumer to fully resolve the uncertainty over the alternative-specific consumption utilities, implying incom-

[1] OCED Glossary of Statistical Terms defines consumer welfare as "the individual benefits derived from the consumption of goods and services. In theory, individual welfare is defined by an individual's own assessment of his/her satisfaction, given prices and income. Exact measurement of consumer welfare therefore requires information about individual preferences." https://stats.oecd.org/glossary/detail.asp?ID=3177

plete information acquisition. After the costly information-acquisition stage, the consumer updates her/his beliefs using Bayes rule and makes a purchase decision based on the posterior belief over alternative-specific consumption utilities. Discrete-choice probabilities akin to those of the classic RUM models are derived as a result. The derived discrete-choice probabilities, however, reflect both the true alternative-specific consumption utilities and the possibly biased subjective prior belief over them.

The RI-based choice model developed herein allows for the possibility that consumer welfare decreases with the addition of new products. When the possibly biased subjective prior belief affects the choice, an inattentive consumer may not discover potentially superior alternatives in the choice set, even after the costly information acquisition. Consequently, the addition of inferior choice alternatives could lead to worse consumer outcomes if superior choice alternatives are no longer discovered during the information-acquisition stage.

By conducting a counterfactual consumer-welfare simulation on consumers' choice-set composition, I find the consumer surplus declines, on average, after the new product introduction in the case study of the launch of Tide Pods. I show the welfare declines arise from Tide charging higher prices for the pods than the value that an average consumer would place on them. Tide would need to cut the prices of Tide Pods by 6%-7%, all else equal, to make consumers, on average, just as well off as prior to the launch. This potential for welfare-reducing product introductions highlights the managerial importance of Costco's product-curation and pricing strategy. In fact, in digital e-commerce markets where shelf space is costless, the managerial relevance of product curation may be even more pronounced. The findings herein suggest excessive variety could potentially harm consumers if they are impeded from discovering superior product alternatives, a feature that is explicitly assumed away in conventional discrete-choice models. Under the RI-based demand framework proposed, a manager could potentially increase consumer welfare through product curation and a limited supply of variety.

To the best of my knowledge, this paper is the first in the literature to develop a general empirical framework and apply it to real-world data, in which an increased variety does not always improve consumer welfare. In addition to the nonmonotonicity of consumer welfare with respect to the number of alternatives, the RI-based choice model developed herein contributes to the literature on consumer consideration, using a similar two-stage "consideration-then-choice" formulation (e.g., product consideration by Allenby and Ginter 1995; Mehta et al. 2003 and price consideration by Ching et al. 2009, 2014, Forthcoming). I also use a similar empirical strategy to distinguish between consideration and choice by excluding demand shifters such as promotions and purchase history from the alternative-specific utilities (e.g., Allenby and Ginter, 1995; Mehta et al., 2003; Ching et al., 2009, 2014, Forthcoming; Terui et al., 2011; Murthi and Rao, 2012), including them only in the information-gathering or "consideration" stage. The RI-based choice model also contributes to the literature analyzing the role of the dispersion parameter in conventional RUM models. In the current framework, the dispersion parameter represents the unit information costs, which I parameterize as a function of the shopping environment and observed demographics. The coefficient estimates turn out to be broadly consistent with the search-cost story.[2]

Benchmarking the RI-based discrete-choice model against conventional RUM models, I find the proposed framework performs better both in and out of sample. The added flexibility of the consideration stage and the parametrization of the dispersion term may explain this superior performance.

The remainder of the paper is organized as follows. Section 2 illustrates the major features of the empirical framework developed in the paper and connects them to the related literature. Section 3 formally develops the RI-based empirical framework of discrete-choice and consumer-welfare evaluation. Section 4 applies the empirical framework to the Tide Pods' introduction to evaluate the associated consumer-welfare gains, and section 5 concludes.

2 Overview of the Framework and the Relation to Extant Literature

This section illustrates the choice contexts and welfare calculations considered throughout the present paper by simple examples, summarizes the main contributions, and connects them to the extant literature.

2.1 Rationalizing the Observed Brand Choices Using the RI Framework

I model consumers' brand choice as a process of information acquisition from a shelf by distinguishing and recognizing the prices and product attributes of different alternatives to learn about the own-consumption utility that each alternative would yield. The following example describes one such choice context that I consider throughout the present paper.

Example 2.1. A consumer wants to buy one laundry detergent item that maximizes the alternative-specific consumption utility, which is fixed and deterministic. However, all the laundry detergent items in the aisle are indistinguishable at this point to the consumer.

S/he is familiar with some laundry detergent items on the shelf, because s/he bought them in the past. When s/he walks into the laundry detergent aisle, s/he immediately perceives how many different items are on the shelf and what items are under promotion, such as in-store displays or feature advertisements. Previous purchases and promotions form the subjective prior belief over the consumption utilities of each alternative, but alternatives are otherwise indistinguishable at this point.

To distinguish each alternative and learn about the alternative-specific consumption utility, the consumer has to read labels and/or price tags. This information-acquisition process is costly because it takes time and effort. Although the specific information-acquisition process is not observable to

[2] By contrast, in the discrete-choice RUM framework, the dispersion parameter is usually normalized as 1 because interpreting that consumers with some demographics make more idiosyncratic choices than others is difficult.

the researcher, spending an infinite amount of time and effort to gather every bit of information on all alternatives on the shelf is never optimal.

After the costly information-acquisition process, the consumer updates the subjective prior beliefs using the newly acquired information about true alternative-specific consumption utilities to form the posterior beliefs. Based on the posterior beliefs on consumption utilities, the consumer chooses one alternative that s/he believes will yield the highest consumption utility among the available alternatives, including the outside option. The choice does not deterministically designate the alternative with the highest utility, because uncertainty about alternative-specific consumption utility was not fully resolved, due to incomplete information acquisition.

The stylized example above illustrates two key features that are common to any brand-choice context – at least *some* costly information acquisition over the shelf space is necessary, and, the acquired information on the alternatives is not likely to be complete.

The consumer-choice model that I develop in this paper, which is underpinned by the consumer's optimization over costly information acquisition, is built upon a new interpretation of recent theoretical works on rational inattention by Matějka and McKay (2015); Fosgerau et al. (2019). Matějka and McKay used Shannon (1948) entropy as the information-cost function to rationalize an extended logit form of discrete-choice probability, and Fosgerau et al. extended Matějka and McKay's result to any functional form of discrete-choice probability by employing a generalized entropy cost function. For empirical application and counterfactual, I depart from the conventional interpretation on consumers' prior distribution that it reflects the true distribution of the alternative-specific consumption utilities, and that it is fully known to the consumers. Instead, I allow for the gap between consumers' subjective prior belief and the true alternative-specific consumption utilities.

The choice context described in Example 2.1 has two important features that make our RI-based discrete-choice model more applicable to such contexts than the workhorse full-information discrete-choice RUM model: (i) The source of stochasticity in consumer choice is attributed to incomplete information acquisition over alternative-specific consumption utility, and (ii) the (posterior) choice-probability shifters have two separate sources, respectively, from the true consumption utilities and the possibly biased subjective prior beliefs over them. The subjective prior is affected by "free" information such as promotions or own purchase history.

(i). The marketing literature has provided ample evidence that consumers simply do not consider all the variants in the aisle of a store (e.g., Hoyer and Brown, 1990; Drèze et al., 1994; Wedel and Pieters, 2000). Studies have documented and investigated the importance of costly information acquisition and consumer learning about prices or other product attributes in the brand-choice context, incomplete learning of which can make the choices stochastic (e.g., Anderson and Simester (1998); Monroe and Lee (1999); Ching et al. (2009, 2014, Forthcoming)). In particular, Ching et al. (2009, 2014) specifically claim the limited attention of consumers as the source of stochastic category-choice decision in their price-consideration model. My modeling approach based on the RI framework is based on the same intuition: consumers have to acquire information about alternative-specific consumption utilities, which is costly. In the model of consumer choice developed in the present paper, consumer choices are stochastic because consumers are highly unlikely to make choices after acquiring all the information about every alternative in their choice set. By contrast, the workhorse full-information discrete-choice RUM framework assumes consumers have complete knowledge about the true alternative-specific consumption utilities including the "idiosyncratic shock" component that is not observable to a researcher. Choices in the full-information RUM framework are stochastic to a researcher due to this unobserved idiosyncratic shock component. Two other important strands of literature in economics and marketing, which emphasize the role and the cost of information friction, are consumer-search literature (e.g., Stigler, 1961; Diamond, 1971; Reinganum, 1979; Weitzman, 1979; Burdett and Judd, 1983; Carlson and McAfee, 1983; Stahl II, 1989; Chade and Smith, 2006; Kim et al., 2010; Ursu, 2018, Forthcoming) and consideration-set literature (e.g., Mehta et al., 2003; Kim et al., 2010; Honka, 2014; Palazzolo and Feinberg, 2015; Honka and Chintagunta, 2017; Ursu, 2018; Morozov, 2019; Honka et al., Forthcoming; Ursu, Forthcoming).

(ii). My modeling approach based on the RI framework can separately incorporate the informational choice-probability shifters such as promotions or purchase history in the choice probability, which are not a component of the true alternative-specific consumption utility. Those informational choice-probability shifters tweak the subjective prior belief over alternative-specific consumption utilities of the consumer. The resulting subjective prior can be biased; thus, it may not necessarily reflect the true alternative-specific consumption utility.

Several extant papers have addressed the separate role of informational choice-probability shifters in the marketing literature from different perspectives. Anderson and Simester (1998) provide the insight that promotional activities may draw consumers' attention to the category, emphasizing the informational role of the promotions. Anderson and Simester (2001) find sale signs are less effective when more items on the shelves have sale signs. This finding is consistent with my modeling approach in that promotions play an informational role by making the promoted items salient and thus easily distinguished from non-promoted items. Allenby and Ginter (1995); Terui et al. (2011); Murthi and Rao (2012) examine the role of advertisements, in-store displays, and features to find they play an important role in the consideration-set formation stage, but they do not directly affect the consumption utilities. Anderson and Simester (2013) find the effect of learning from consumers' own previous purchases on the advertising spillover effects to competitors. In a similar spirit, Ching et al. (2009, 2014, Forthcoming) use promotional activities as the exclusion restriction for the category consideration stage in their price-consideration model. The motivation is that those promotional activities may only affect the consumers' information, not the utility from consumption. The present paper closely follows the literature discussed above, to consider the in-store display/feature variables and own-purchase-history variables as informational choice-probability shifters, or equivalently, probabilistic consideration shifters. They are used as exclusion restrictions from the alternative-specific consumption utility.

The subjective prior belief over each choice alternative is initialized at each choice incidence in Example 2.1. In other words, it is de facto assuming consumers need to recover information again at each shopping instance over prices, product attributes, and specific items' shelf-space locations. I make this assumption, which follows the tradition in the marketing literature on brand choice dating back to at least Gudagni and Little (1983), not just because of the empirical tractability, but also because the information-acquisition process being modeled here could be the process to distinguish and recognize the prices and product attributes of different alternatives in the aisle. Note the experience of previous purchase is modeled in a way that can still affect the prior belief over consumption utilities, which may reflect the presence of choice inertia.

2.2 Non-monotonicity of Consumer Welfare in the Number of Choice Alternatives

The RUM-based standard consumer-welfare-evaluation framework by McFadden (1981); Small and Rosen (1981); Herriges and Kling (1999); Dagsvik and Karlström (2005); Bhattacharya (2015) is built upon the full-information assumption in consumer choice. The framework has been applied to various contexts including price changes, new-product introduction, excluding alternatives in the choice set, measuring the impact of brands in consumer welfare, and so on (see, e.g., Petrin, 2002; Hortaçsu and Syverson, 2004; Goldfarb et al., 2009; Morozov, 2019 among numerous others). Under the full-information assumption, a natural conclusion is that adding more alternatives to consumers' choice set must make consumers weakly better off, because consumers can simply avoid choosing the added alternative if it is a "bad" one when they have complete knowledge on the alternatives. Thus, in essence, "SKU rationalization" problem would not exist for a retailer when the full-information RUM-based consumer-welfare-evaluation framework is employed, because consumer value increases monotonically with additional variants. Only costs on the supply side such as shelf space would ever limit variety.

The possibility that more choice alternatives may harm consumers, however, is well known. Iyengar and Lepper (2000); Iyengar et al. (2004); Bertrand et al. (2010) document that more choice alternatives lead to less choice in the context of grocery shopping and financial-product choice. Diehl and Poynor (2010) report that a larger assortment may raise consumers' expectations, which leads to consumers being disappointed more by the choice they make. Broniarczyk et al. (1998); Boatwright and Nunes (2001) examine the effect of reducing assortments, suggesting evidence that consumers may not be harmed and sales may even increase.

This paper develops a general empirical framework to provide a systematic method to evaluate whether adding a choice alternative harms or benefits consumers when they have only incomplete information over the alternatives. The flexibility of the empirical framework whereby adding more alternatives may either benefit or harm consumer welfare is a useful property for studying SKU rationalization, assortment management, optimal product line, store choice, and so on.

Let us consider a toy example that includes two alternatives $\{1, 2\}$ with the deterministic part of the alternative-specific utility $u_1 = u_2 = 0$. Assume, for simplicity, the additive i.i.d. double-exponential "idiosyncratic" utility shock so that the resulting choice probability of each alternative is $\frac{\exp(0)}{\exp(0)+\exp(0)} = 0.5$, and the consumer welfare evaluated under RUM assumptions is 0.693.[3] When a third, "good" alternative is added to the choice set where $u_3 = 1$, the consumer welfare increases to 1.551. Next, when a third, "neutral" alternative is added to the choice set where $u_3 = 0$, consumer welfare increases to 1.097. Lastly, when a third, "bad" alternative is added to the choice set where $u_3 = -1$, consumer welfare still increases to 0.862. It indeed has to increase as long as $u_3 > -\infty$. This monotonicity property in the RUM framework is a result of attributing the choice of $u_3 = -1$ as it is somehow preferred more by the consumer than the other two alternatives. This property may be justified under the strong assumption that the consumers have complete knowledge over every alternative-specific utility in her/his choice set, and that the stochasticity of the discrete-choice model is attributed to the "idiosyncratic" utility component that is not observable to the researcher. The *RUM Welfare* column of Table 1 summarizes our discussion in this paragraph thus far.

If, on the contrary, consumers end up having only incomplete knowledge over the alternative-specific utility, the monotonicity of the consumer-welfare formula can be problematic; consumers may have chosen a bad alternative simply because they did not know a better alternative existed in their choice set. To illustrate, suppose the alternative-specific utility $u_1 = u_2 = 0$ contains no "idiosyncratic utility shock." Consumers, however, need to learn about the alternative-specific utility of each product, which is costly. If the information acquisition over alternative specific utility is incomplete, the choices are still stochastic; that is, the best alternative is not always chosen, due to incomplete information. Matějka and McKay (2015); Fosgerau et al. (2019) show that under the assumption of optimal information acquisition with a suitably chosen information-cost function, the choice probability of each alternative can still take the simple logit form: $\frac{\exp(0)}{\exp(0)+\exp(0)} = 0.5$.[4]

The goal of the present paper is to develop a consumer-welfare-evaluation framework such that the monotonicity in consumer welfare in the number of alternatives breaks down when the source of the stochasticity in discrete choices is the consumers' incomplete knowledge over the true alternative-specific consumption utilities. Adding a "bad" alternative may even lower consumer welfare, which is the expected utility from the choice problem. An intuitive reasoning behind this nonmonotonicity can be found in the formula of the consumer welfare from the perspective of the researcher, which is defined by "the (gross) benefit of choice net of the cost of information." Adding a bad alternative reduces the (gross) benefit of choice when the consumer has only incomplete information over choice alternatives, and the cost of information can either increase or decrease depending on the information structure. Although how the exact calculation can be carried out is relegated to section 3 after the choice problem and the consumer-welfare formula are formally introduced, the *RI Welfare* column

[3]The usual log-sum formula $\ln(\exp(0) + \exp(0))$ yields this number.

[4]This calculation is under the assumptions that (i) consumers have symmetric subjective prior belief distribution over alternative-specific utilities, (ii) the Shannon entropy information-cost function is employed, and (iii) the true utility is fixed and degenerate, whereas consumers do not know them. We relax (i) and (ii) in the subsequent sections.

of Table 1 illustrates what happens when $u_3 = 1$, $u_3 = 0$, and $u_3 = -1$ are added to the choice set, respectively.[5] In our example, when a bad alternative is added to the choice set, the RI-based consumer welfare decreases to –0.237 from 0. Notably, when a neutral alternative is added to the choice set, the RI-based consumer welfare is not changed from the baseline. Comparing the RI welfare with RUM welfare in the "Adding Neutral" case, it is immediate that the RUM-based welfare evaluation measure exaggerates the utility gains from adding more alternatives in the choice set, whereas the RI-based welfare evaluation measure does not. All else being equal, it would indeed be more sensible that the consumer welfare does not change when the alternative with $u_3 = 0$ is added to the choice set.

Table 1: A Toy Example: Comparison of Welfare Calculation in RI and RUM

Case	Alternative	u_j	Choice Prob.	RI Welfare	RUM Welfare
Baseline	1	0	0.5	0	0.693
	2	0	0.5		
Adding Good	1	0	0.212	0.453	1.551
	2	0	0.212		
	3	1	0.576		
Adding Neutral	1	0	0.333	0	1.097
	2	0	0.333		
	3	0	0.333		
Adding Bad	1	0	0.422	–0.237	0.862
	2	0	0.422		
	3	–1	0.155		

Note. The table illustrates the welfare changes associated with adding a "bad," "neutral," and "good" alternative, respectively, to the choice set. The RI welfare is calculated using equation (3.10) in section 3, and the RUM welfare is calculated using the usual log-sum formula.

2.3 Relations to Other Streams of the Literature in Marketing and Economics

The RI-based empirical framework of discrete choice developed in the present paper reconciles the early literature of stochastic brand choice with the RI/RUM discrete-choice framework via Dirichlet distribution and a generalized entropy information-cost function, allowing for the effect of prices and other marketing-mix variables in the choice of an alternative. Herniter (1973); Bass (1974)

[5] The calculation is as follows. For the Adding Good case, RI Welfare is

$$\begin{aligned} W_{RI}(0,0,1) = {} & 0.212 \times 0 + 0.212 \times 0 + 0.576 \times 1 \\ & - 1 \times \{0.212 \times (\ln(0.212) - \ln(0.333)) + \\ & 0.212 \times (\ln(0.212) - \ln(0.333)) + \\ & 0.576 \times (\ln(0.576) - \ln(0.333))\}. \end{aligned}$$

Calculation for the Adding Bad case is similar.

rationalize the observed market shares by the maximum-entropy principle, and Bass et al. (1976) show the distribution of choice probabilities in Herniter; Bass's framework can be modeled as a Dirichlet distribution, which is extensively used in the proof of Theorem 3.1 in the present paper. Herniter; Bass; Bass et al., however, attribute probabilistic choices solely to inherent stochasticity in consumer preferences, not allowing the effect of prices or other marketing-mix variables.

Also closely related to deriving the discrete-choice probabilities from imperfect perception and/or costly information are Block and Marschak (1960); Shugan (1980); de Palma et al. (1994). Block and Marschak; Shugan; de Palma et al. rationalize stochastic choice by attributing the stochasticity to the decision-maker's limited information-processing ability, and then derive probabilistic choice rules from minimizing the cost of comparing alternatives (Shugan, 1980), or errors stemming from incomplete information processing (de Palma et al., 1994). To the best of my knowledge, this paper is the first in the literature to develop a consumer-welfare-evaluation framework built upon a general discrete-choice framework microfounded on the decision-maker's limited information-processing ability.[6]

3 The Empirical Framework of Discrete Choice and Consumer-Welfare Evaluation Based on Rational Inattention

In this section, I develop an empirical framework of discrete choice and consumer-welfare evaluation based on a new interpretation of RI. Then, I state and prove the necessary theorem that allows us to estimate the discrete-choice RI model developed in the present section as if estimating usual discrete-choice RUM models. Finally, I derive the associated social-welfare function based on the proposed choice model and compare it with the social-welfare function derived from the RUM framework.

Overview of the Consumer's Choice Problem, Terminology, and Notation As an overview, I summarize the endowed information structure of the problem and procedure on consumers' choice of an alternative below. I assume a consumer needs to go through all of these stages at each choice incidence.

- *(Endowed Information Structure)* True alternative-specific consumption utility of a consumer is deterministic and fixed. However, the consumer perceives all the alternatives as indistinguishable at this point.
- *(Choice Stage 0: Formation of Consumer's Subjective Prior)* The consumer forms a subjective prior belief over alternative-specific consumption utilities based on

[6]In a recent work by Morozov (2019), consumers' consideration process is explicitly modeled and blended with the RUM framework to estimate the gains from the introduction of solid state drives in the storage market. Morozov finds the welfare gains from new-product introduction can be underestimated when consumers' consideration process is ignored.

- The number of alternatives presented,
- Past purchases s/he made, and
- Current promotions in the store.

- *(Choice Stage 1: Optimization over Information Strategies)* The consumer optimally chooses the following, taking the information-cost function and the subjective prior belief formed in the previous stage as given:
  - The strategy of costly attention allocation, that is, what and how much information to acquire, and
  - The action strategy, that is, what to buy given a particular posterior belief over alternative-specific consumption utilities.
- *(Choice Stage 2: Choice of an Alternative)* The consumer executes the information strategy, learns about the alternative-specific consumption utilities, updates the subjective prior belief in the Bayesian fashion to yield the posterior belief, and chooses an alternative according to the optimized outcome of the posterior belief.

(Endowed Information Structure) The consumer walks into the store. $\mathcal{J}_i = \{0, 1, ..., J_i\}$ denotes the set of alternatives available to consumer i. $\mathcal{J}_i$ contains the outside option, which is denoted by option 0. The true values of the alternative-specific consumption utilities are deterministic and fixed, represented by the degenerate distribution $Q^0(\cdot)$ that places the unit mass at the true alternative-specific consumption-utility vector $\mathbf{u}_i := \{u_{i,j}\}_{j \in \mathcal{J}_i}$. However, all the alternatives are indistinguishable to the consumer at this point. It can be represented as consumers being endowed with a full-support, subjective prior distribution over the ex-ante perceived alternative-specific utility $\mathbf{v}_i$ that is invariant to the permutations of entries of $\mathbf{v}_i$.

(Choice Stage 0: Formation of Consumer's Subjective Prior) The endowed symmetric subjective prior distribution of the consumer is tweaked by some "free" information that the consumer obtains: the consumer remembers which items and products s/he purchased in the past, and perceives in-store promotions such as feature advertisements. Let $\mathbf{D}_i := \{\mathbf{d}_{i,j}\}_{j \in \mathcal{J}_i}$ denote the observed informational shifters of the subjective prior belief on the alternative-specific consumption utility. $\mathbf{d}_{i,j}$ includes the dummies of the previous purchase history and current promotion status. This information, acquired without cost, results in a subjective prior belief over $\mathbf{v}_i$, which I denote $Q_i(\mathbf{v}_i) := Q(\mathbf{v}_i | \mathcal{J}_i, \mathbf{D}_i)$.

(Choice Stage 1: Optimization over Information Strategies) Taking the information-cost function as given, the consumer optimizes over what and how much information to acquire. The consumer also decides what to buy when a particular posterior belief on alternative-specific consumption utilities is given.

(Choice Stage 2: Choice of an Alternative) The consumer engages in the costly information acquisition to learn about the alternative-specific consumption utilities. What the consumer perceives

to learn about is $\mathbf{v}_i$. However, true utility $\mathbf{u}_i$ is the alternative-specific consumption utility that the consumer learns through the costly information-acquisition process. Therefore, the optimization procedure yields the conditional choice probability denoted by $\Pr_i(i \text{ Chooses } j|\mathbf{u}_i)$ that conditions on $\mathbf{u}_i$[7] as the solution from the perspective of the researcher.

Finally, let us define the unconditional choice probability $\pi_{i,j}$ by the expected conditional choice probability, expectation taken against the consumer's subjective prior distribution $Q_i(\mathbf{v}_i)$:

$$\begin{aligned}\pi_{i,j} &:= E_{Q_i}\left[\Pr_i(i \text{ Chooses } j|\mathbf{v}_i)\right] \\ &= \int \Pr_i(i \text{ Chooses } j|\mathbf{v}_i)\, Q_i(d\mathbf{v}_i)\,. \end{aligned} \tag{3.1}$$

Notice $\pi_{i,j}$ does not condition either on the perceived ex-ante alternative-specific consumption utility vector $\mathbf{v}_i$ nor the true alternative-specific consumption utility vector $\mathbf{u}_i$.

Table 2 summarizes the notations used throughout the paper. Matějka and McKay (2015) show the problem of optimization over information strategies described above, which is a high-dimensional problem, can be equivalently formulated as directly choosing the set of conditional choice probabilities when the Shannon entropy information-cost function is employed. This equivalent formulation reduces the dimensionality of *(Choice Stage 1)* and *(Choice Stage 2)* problem drastically, which is how Fosgerau et al. (2019) formulate the RI-choice problem using the generalized entropy cost function. In what follows, I follow Fosgerau et al. in formulating the consumers' choice problem that consumers directly optimize over the set of consumers' conditional choice probabilities $\{\Pr_i(i \text{ Chooses } j|\mathbf{v}_i)\}_{j\in\mathcal{J}_i}$ conditioning on perceived $\mathbf{v}_i$.

3.1 Choice of an Alternative: The Problem from the Perspective of the Consumer

In this subsection, I formally develop the consumer's choice problem of an alternative, and provide the form of the solution from the perspective of the consumer. The form of the choice probability is derived in this subsection, taking $Q_i(\mathbf{v}_i)$ as if the true distribution of consumption utilities. The discrepancy between $Q_i(\mathbf{v}_i)$ and $Q^0(\mathbf{u}_i)$ will be exploited later when I derive the likelihood from the perspective of the researcher.

Choice Stage 0: Formation of Consumer's Subjective Prior The symmetric and non-degenerate subjective prior distribution over the perceived consumption utilities are tweaked by $\mathbf{D}_i := \{\mathbf{d}_{i,j}\}_{j\in\mathcal{J}_i}$, the informational shifters of the subjective prior belief distribution. $Q_i(\mathbf{v}_i) := Q(\mathbf{v}_i|\mathcal{J}_i, \mathbf{D}_i)$, consumer i's subjective prior distribution over the perceived consumption utilities $\mathbf{v}_i$, is formed as a result.

[7] I denote $\Pr_i(i \text{ Chooses } j|\mathbf{u}_i) = \Pr(i \text{ Chooses } j|\mathbf{u}_i; \mathcal{J}_i, \mathbf{D}_i)$ with a slight abuse of notation.

Table 2: Notations Used throughout the Paper

Notation	Explanation
i	Individual consumer/decision-maker
j	Alternative, $j = 0$ is outside option
$\mathcal{J}_i := \{0, 1, ..., J_i\}$	Set of alternatives available to consumer i
$\mathbf{v}_i := \{v_{i,j}\}_{j \in \mathcal{J}_i}$	Consumer i's perceived alternative-specific consumption utility
$\mathbf{u}_i := \{u_{i,j}\}_{j \in \mathcal{J}_i}$	Vector of consumer i's true, deterministic alternative-specific consumption utility
$(p_j, \mathbf{x}_j)$	Observed shifters of $u_{i,j}$. p_j is price and $\mathbf{x}_j$ is product attributes
$\eta_{i,j}$	Unobserved (to the researcher) shifters of $u_{i,j}$, observable to the consumer
$\mathbf{D}_i := \{\mathbf{d}_{i,j}\}_{j \in \mathcal{J}_i}$	Observed informational shifters of the subjective prior belief distribution
$Q_i(\mathbf{v}_i) := Q(\mathbf{v}_i \vert \mathcal{J}_i, \mathbf{D}_i)$	i's subjective prior belief distribution over the perceived consumption utilities $\mathbf{v}_i$
$Q^0(\mathbf{u}_i)$	True distribution of true $\mathbf{u}_i$ that is degenerate, placing the unit mass at the true $\mathbf{u}_i$
$\Pr_i(i \text{ Chooses } j \vert \mathbf{u}_i)$	Conditional choice probability of i choosing j, conditioning on true $\mathbf{u}_i$
$\Pr_i(i \text{ Chooses } j \vert \mathbf{v}_i)$	Conditional choice probability of i choosing j, conditioning on perceived $\mathbf{v}_i$
$\pi_{i,j} = E_{Q_i}[\Pr_i(i \text{ Chooses } j \vert \mathbf{v}_i)]$	Unconditional choice probability of i choosing j, expectation taken against Q_i
$c_i(\cdot, \cdot)$	i's information cost function
μ_i	Reciprocal of i's unit information cost
$\mathbf{w}_i$	Observed shifters of the unit information cost
$W_{RUM}(\cdot, \cdot)$	RUM-based social-welfare Function
$W_{RI}(\cdot, \cdot)$	RI-based social-welfare Function
$\mathbf{H}(\cdot) := \nabla(\exp W_{RUM}(\cdot))'$	Transpose of exponentiated gradient of $W_{RUM}(\cdot)$
$H_j(\cdot)$	j'th element of $\mathbf{H}(\cdot)$

Note. This table summarizes the notations used throughout the paper and their explanations.

Choice Stage 1: Optimization over Conditional Choice Probability The consumer maximizes the expected gain from choice net of information cost with respect to her/his own subjective prior $Q_i(\mathbf{v}_i)$ over the vector of alternative-specific consumption utilities $\mathbf{v}_i$:

$$\max_{\{\Pr_i(i \text{ Chooses } j|\mathbf{v}_i)\}_{j \in \mathcal{J}_i}} E_{Q_i}\left[\sum_{j \in \mathcal{J}_i} v_{i,j} \Pr_i(i \text{ Chooses } j|\mathbf{v}_i) - \text{Information cost}\right], \tag{3.2}$$

where the conditional choice probabilities have to be nonnegative. Formally, the consumer's optimization problem is

$$\max_{\{\Pr_i(i \text{ Chooses } j|\mathbf{v}_i)\}_{j \in \mathcal{J}_i}} \int_{\mathbf{v}_i} \left\{\sum_{j \in \mathcal{J}_i} v_{i,j} \Pr_i(i \text{ Chooses } j|\mathbf{v}_i) - c_i\left(\boldsymbol{\pi}_i, \{\Pr_i(i \text{ Chooses } j|\mathbf{v}_i)\}_{j \in \mathcal{J}_i}\right)\right\} Q_i(d\mathbf{v}_i), \tag{3.3}$$

with the constraints that $\sum_{j \in \mathcal{J}_i} \Pr_i(i \text{ Chooses } j|\mathbf{v}_i) = 1$ and $\Pr_i(i \text{ Chooses } j|\mathbf{v}_i) \geq 0 \; \forall j \in \mathcal{J}_i, \forall \mathbf{v}_i \in \mathbb{R}^{|\mathcal{J}_i|}$.

Let μ_i be the reciprocal of individual i's unit information cost. Define a vector-valued function $\mathbf{H}(\mathbf{v}_i) := (H_1(\mathbf{v}_i), ..., H_{J_i}(\mathbf{v}_i))'$, which is the exponentiated gradient of McFadden's social-surplus function $W_{RUM}(\mathbf{v}_i)$ in the RUM context.[8] The information-cost function $c_i(\cdot,\cdot)$ in (3.3) has the following form:

$$\begin{aligned} & c_i\left(\boldsymbol{\pi}_i, \{\Pr_i(i \text{ Chooses } j|\mathbf{v}_i)\}_{j\in\mathcal{J}_i}\right) \\ =& \mu_i^{-1} \sum_{j\in\mathcal{J}_i} \Pr_i(i \text{ Chooses } j|\mathbf{v}_i) \left\{\ln H_j^{-1}\left(\{\Pr_i(i \text{ Chooses } j|\mathbf{v}_i)\}_{j\in\mathcal{J}_i}\right) - \ln H_j^{-1}(\boldsymbol{\pi}_i)\right\}. \end{aligned} \tag{3.4}$$

The information-cost function is proportional to the difference in the generalized entropy of unconditional and conditional choice probabilities.

The generalized entropy is defined in Appendix A. The shape of the solution to the optimization problem (3.3) will be inherited from the shape of the generalized entropy cost function. Note that if $\mathbf{H}(\mathbf{v}_i)$ is the identity function, the information-cost function reduces to the expected difference in Shannon entropy. The information-cost function is convex in $\{\Pr_i(i \text{ Chooses } j|\mathbf{v}_i)\}_{j\in\mathcal{J}_i}$,[9] which makes the information associated with more "precise" conditional choice probability more expensive.

Choice Stage 2: Choice of an Alternative and the Solution of the Consumer's Problem During Stage 2, the consumer chooses one alternative out of $\mathcal{J}_i$, according to the conditional choice probability $\Pr_i(i \text{ Chooses } j|\mathbf{v}_i)$ after costly information acquisition on the true alternative-specific consumption utility.

The solution for the problem (3.3) yields the conditional choice probability after costly information acquisition as

$$\Pr_i(i \text{ Chooses } j|\mathbf{v}_i) = \frac{H_j\left(\exp\left(\ln\left(\mathbf{H}^{-1}(\boldsymbol{\pi}_i)\right) + \mu_i\mathbf{v}_i\right)\right)}{\sum_{k\in\mathcal{J}_i} H_k\left(\exp\left(\ln\left(\mathbf{H}^{-1}(\boldsymbol{\pi}_i)\right) + \mu_i\mathbf{v}_i\right)\right)}, \tag{3.5}$$

which conditions on the "realized" $\mathbf{v}_i$. The solution (3.5), combined with the definition of unconditional choice probability (3.1), leads to the condition that the consumer's subjective prior belief distribution $Q_i(\cdot)$ must satisfy the following fixed-point equation:

$$\pi_{i,j} = \int_{\mathbf{v}_i} \frac{H_j\left(\exp\left(\ln\left(\mathbf{H}^{-1}(\boldsymbol{\pi}_i)\right) + \mu_i\mathbf{v}_i\right)\right)}{\sum_{k\in\mathcal{J}_i} H_k\left(\exp\left(\ln\left(\mathbf{H}^{-1}(\boldsymbol{\pi}_i)\right) + \mu_i\mathbf{v}_i\right)\right)} Q_i(d\mathbf{v}_i) \qquad \forall j \in \mathcal{J}_i \tag{3.6}$$

[8]

$$W_{RUM}(\mathbf{v}_i) := E\left[\max_{j\in\mathcal{J}_i}\{v_{i,j} + \epsilon_{i,j}\}\right]$$

with $\epsilon_{i,j}$ being appropriately distributed additive idiosyncratic preference shocks. See Appendix A for the precise definition, and Online Appendix G for examples of $\mathbf{H}(\cdot)$ functions.

[9]By Proposition 8 of Fosgerau et al. (2019).

for $\pi_{i,j}$ such that $\pi_{i,j} > 0$ (Proposition 6 of Fosgerau et al., Corollary 2 of Matějka and McKay).

Two remarks are in order. First, the fixed-point equation (3.6) places an important restriction on the relation between $Q_i(\cdot)$ and $\{\pi_{i,j}\}_{j \in \mathcal{J}_i}$, especially when μ_i and $\{\pi_{i,j}\}_{j \in \mathcal{J}_i}$ are identified from data, whereas $Q_i(\cdot)$ cannot be known to a researcher directly. I will explain further how this restriction can be satisfied in section 3.2. Second, the exact functional form of $\Pr_i(i \text{ Chooses } j|\mathbf{v}_i)$ depends crucially on the functional form of $\mathbf{H}(\cdot)$ that is inherited from the cost function (3.4). By employing the appropriate generalized entropy information-cost function introduced by Chiong et al. (2016); Fosgerau et al. (2019), (3.5) can yield the functional form of the choice probability resulting from any additive RUM choice models.

3.2 Choice of an Alternative: The Solution from the Perspective of the Researcher and the Likelihood

The consumer's problem is stated and solved in section 3.1 to derive the conditional-choice-probability expression (3.5), under the endowed information structure that the consumer takes her/his subjective prior $Q_i(\cdot)$ as the true distribution of alternative-specific consumption utility. However, a discrepancy exists between the consumer's subjective prior belief $Q_i(\cdot)$ and the true distribution $Q^0(\cdot)$ over the alternative-specific consumption utility. Because the realized alternative-specific consumption utility vector is $\mathbf{u}_i$ with probability one, the actual conditional choice probability of choosing an alternative after costly information acquisition would be the following, which conditions on the true $\mathbf{u}_i$:

$$\Pr_i(i \text{ Chooses } j|\mathbf{u}_i) = \frac{H_j\left(\exp\left(\ln\left(\mathbf{H}^{-1}(\boldsymbol{\pi}_i)\right) + \mu_i \mathbf{u}_i\right)\right)}{\sum_{k \in \mathcal{J}_i} H_k\left(\exp\left(\ln\left(\mathbf{H}^{-1}(\boldsymbol{\pi}_i)\right) + \mu_i \mathbf{u}_i\right)\right)}. \tag{3.7}$$

Three components contribute to the formation of (3.7): (i) The unit information cost μ_i^{-1} and the functional form of $\mathbf{H}(\cdot)$, which are inherited from the information-cost function (3.4); (ii) the true distribution $Q^0(\cdot)$ of the alternative-specific consumption utility that is degenerate and places unit mass on $\mathbf{u}_i$ pins $\mathbf{u}_i$ down; and (iii) the consumer's subjective prior belief $Q_i(\cdot)$ over the alternative-specific consumption utility, combined with the fixed-point equation that is described and stated below, pins $\boldsymbol{\pi}_i$ down.

The above expression of the conditional choice probability (3.7) will be taken as the building block of the likelihood in our empirical analysis; that is, the distribution of $\Pr_i(i \text{ Chooses } j|\mathbf{u}_i)$ is observed in the data. Combined with modest structural assumptions and parametrizations, $(\mathbf{u}_i, \mu_i, \boldsymbol{\pi}_i)$ will be taken as objects identified from the data. From a researcher's perspective, the subsequent question would be whether $(\mathbf{u}_i, \mu_i, \boldsymbol{\pi}_i)$ identified from the data would be consistent with a consumer's choice process that solves the problem (3.2) stated in section 3.1.

Because I assume the true $\mathbf{u}_i$ is deterministic and can differ from the perceived $\mathbf{v}_i$, $\mathbf{u}_i$ identified from the data in the conditional choice probability would be consistent with the procedures of a rationally inattentive consumer's choice that we have described thus far.[10] We are more concerned about $\boldsymbol{\pi}_i$, an argument of the conditional choice probability that is defined in close relation with $\{Q_i(\mathbf{v}_i), \mu_i\}$ via (3.6). For an estimate of the unconditional choice probabilities $\boldsymbol{\pi}_i (> 0)$ to be justified as consistent with the RI theory, a non-degenerate subjective prior belief distribution $Q_i(\cdot)$ over the perceived utilities that yields $\boldsymbol{\pi}_i$ must exist. In the RI theory literature, the subjective prior belief distribution $Q_i(\cdot)$ is fully specified, and then the unconditional choice probability $\{\pi_{i,j}\}_{j\in\mathcal{J}_i}$ is solved to satisfy the set of fixed-point equations (3.6).[11] The direction that should be taken here for the empirical application is the converse, because the goal is to estimate the utility and information parameters using the choice data. The following Theorem 3.1 provides the necessary result – when any $(\mathbf{u}_i, \mu_i, \boldsymbol{\pi}_i)$ is identified from the data, $Q_i(\cdot)$ exists that is consistent with (3.6).[12] Therefore, the unconditional choice probabilities $\boldsymbol{\pi}_i$ identified from the choice data can be justified as resulting from a rationally inattentive decision-maker's choice.

Theorem 3.1. *(Existence of Subjective Prior Belief Distribution) Fix the individual i. Let* $J_i (\in \mathbb{N}) \geq 2$ *and let* $\mathcal{J}_i = \{0, 1, ..., J_i\}$*. For each* $j \in \mathcal{J}_i$*, let* $\pi_{i,j} > 0$ *such that* $\sum_{k\in\mathcal{J}_i} \pi_{i,k} = 1$ *is given and fixed. Let* $\mu_i > 0$ *and* $\mathbf{H} : \mathbb{R}^{J_i} \to \mathbb{R}^{J_i}$ *also be given and fixed. Then, a probability measure* Q_i *over possible utility levels exists such that for each* $j (\in \mathcal{J}_i)$*, the following (i) and (ii) hold:*

(i) For each $j (\in \mathcal{J}_i)$*,*

$$\pi_{i,j} = \int \frac{H_j\left(\exp\left(\ln\left(\mathbf{H}^{-1}(\boldsymbol{\pi}_i)\right) + \mu_i \mathbf{v}_i\right)\right)}{\sum_{k\in\mathcal{J}_i} H_k\left(\exp\left(\ln\left(\mathbf{H}^{-1}(\boldsymbol{\pi}_i)\right) + \mu_i \mathbf{v}_i\right)\right)} Q_i(d\mathbf{v}_i). \tag{3.8}$$

(ii) For each $j \in \mathcal{J}_i$*,* $Q_i(\{\mathbf{v}_i : v_{i,j} > v_{i,k}\ \forall k \neq j\}) > 0$*.*

Proof. See appendix B. □

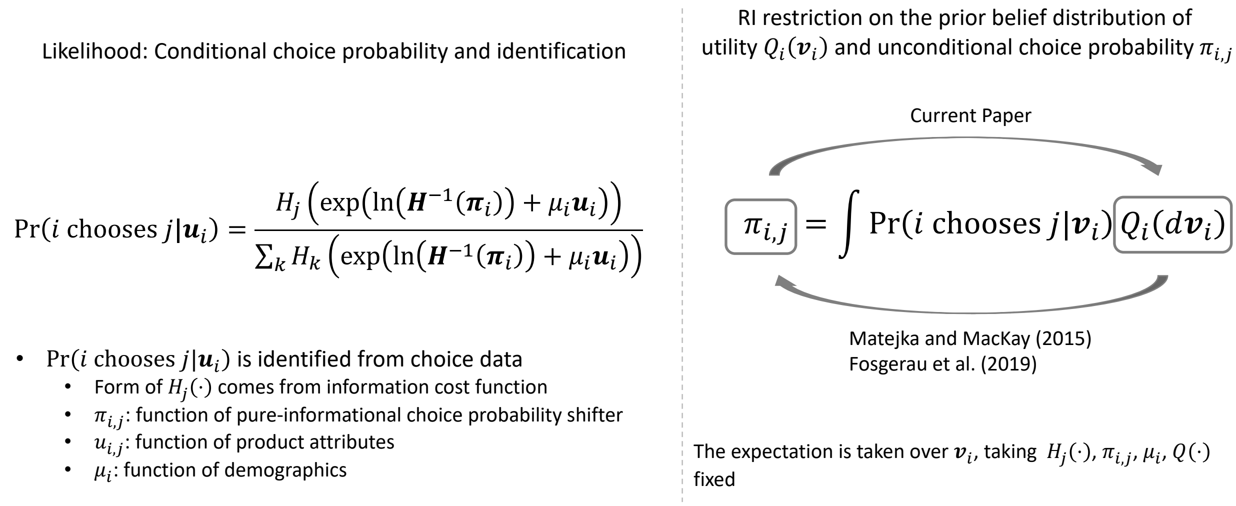
Figure 3.1 summarizes the discussion in this section thus far.

[10] The consumer thinks that s/he is discovering $\mathbf{v}_i$, but in fact s/he is discovering the true, fixed $\mathbf{u}_i$ from the costly information acquisition.

[11] In the extant RI theory literature, $Q_i(\cdot)$, which is nondegenerate, is usually taken as the true distribution of the alternative-specific consumption utilities, that is, $Q_i(\cdot) = Q^0(\cdot)$ using our notation. The extant RI theory literature further assumes the decision-maker knows $Q^0(\cdot)$ fully.

[12] A restriction placed here is that every alternative must have a positive unconditional choice probability, namely, $\pi_{i,j} > 0$, which is not the restriction necessarily imposed by the RI theory literature. For example, Caplin et al. (Forthcoming) allows for the zero unconditional choice probability and interprets the alternatives with zero unconditional choice probability as not included in the decision-maker's consideration set. By contrast, I restrict the unconditional choice probability of all possible alternatives to be positive, because (i) zero unconditional choice probability cannot be distinguished from infinitesimal unconditional choice probability with only a finite number of choice samples in hand, and (ii) in the aggregate demand-estimation context, market-share-equation inversion does not apply when an alternative has zero choice probability, leading to identification failure. I interpret higher $\pi_{i,j}$ as a higher probability of alternative j being considered in the alternative-choice stage.

Figure 3.1: Unconditional and Conditional Choice Probability, RI Restrictions, and Identification



3.3 Welfare Analysis: Social-Surplus Function in a Rational Inattention Framework from the Perspective of the Researcher

Subsections 3.1-3.2 show the RI and RUM frameworks can lead to similar choice-probability expressions under mild restrictions imposed on the RI framework. The welfare calculations, however, can be starkly different because the respective choice-probability expressions are based on fundamentally different microfoundations. In this subsection, I define the social-surplus functions within the RI-based choice model developed above, provide the corresponding compensating variation formula (CV henceforth), and compare it with the conventional RUM-based compensating variation formula.[13] Compensating variation, the hypothetical monetary compensation necessary to make consumers indifferent to changes in the choice-set composition and/or prices, has been the prominent apparatus for evaluating changes in consumer welfare in the literature. The concept of CV was first developed in Walrasian demand theory dating back to at least Hicks (1939). McFadden (1978, 1981); Small and Rosen (1981); Herriges and Kling (1999); Dagsvik and Karlström (2005); Bhattacharya (2015) later adopted the concept within the context of the discrete-choice RUM framework.

In the remainder of the paper, I assume both $v_{i,j}$ and $u_{i,j}$ are quasilinear in the income y_i of the consumer.[14] I further assume both the consumer and the researcher know the functional form of the

[13]The terms social-surplus function and social-welfare function are used interchangeably hereafter.

[14]This quasilinear specification follows the convention in the discrete-choice literature, which simplifies the welfare analysis substantially, because the income effects are canceled out (Small and Rosen (1981) and numerous applied welfare analyses using the CV formula therein). Recent advances of discrete-choice RUM-based welfare frameworks in the literature, including Herriges and Kling (1999); Dagsvik and Karlström (2005); Bhattacharya (2015), consider the nonlinear form of alternative-specific consumption utilities, thereby allowing for income effects. Developing a more flexible RI-based consumer-welfare-evaluation framework that can possibly accommodate income effects exceeds the scope of the present research.

utility. Specifically, $u_{i,j}$ has the following form:

$$u_{i,j} := (y_i - p_j)\,\beta_1^i + \chi_{i,j}, \tag{3.9}$$

where $\chi_{i,j}$ is the "quality index" function that can possibly take product attributes as its argument. y_is are canceled out in the respective conditional-choice-probability expressions (3.5) and (3.7), thereby implying the absence of an income effect both in the perceived and the actual solutions to the consumers' optimization problem.[15] In principle, β_1^i is allowed to vary across heterogeneous individuals i; thus, the proposed welfare-evaluation framework can possibly accommodate the random-coefficients specifications.

The social-surplus function $W_{RI}^i(\mathbf{u}_i, \mathcal{J}_i)$ for individual i with choice alternatives $\mathcal{J}_i$ is defined by the following:

$$\begin{aligned}
W_{RI}^i(\mathbf{u}_i, \mathcal{J}_i) &:= E_{Q^0}\left[\sum_{j\in\mathcal{J}_i} \Pr_i(i \text{ Chooses } j|\mathbf{u}_i)\, u_{i,j} - c_i\left(\{\pi_{i,j}\}_{j\in\mathcal{J}_i}, \{\Pr_i(i \text{ Chooses } j|\mathbf{u}_i)\}_{j\in\mathcal{J}_i}\right)\right] \\
&= \int \left\{\sum_{j\in\mathcal{J}_i} \Pr_i(i \text{ Chooses } j|\mathbf{u}_i)\, u_{i,j} - c_i\left(\{\pi_{i,j}\}_{j\in\mathcal{J}_i}, \{\Pr_i(i \text{ Chooses } j|\mathbf{u}_i)\}_{j\in\mathcal{J}_i}\right)\right\} Q^0(d\mathbf{u}_i) \\
&= \sum_{j\in\mathcal{J}_i} \Pr_i(i \text{ Chooses } j|\mathbf{u}_i)\, u_{i,j} \\
&\quad - \mu_i^{-1}\left\{\sum_{j\in\mathcal{J}_i} \Pr_i(i \text{ Chooses } j|\mathbf{u}_i)\left\{\ln H_j^{-1}\left(\{\Pr_i(i \text{ Chooses } j|\mathbf{u}_i)\}_{j\in\mathcal{J}_i}\right) - \ln H_j^{-1}\left(\{\pi_{i,j}\}_{j\in\mathcal{J}_i}\right)\right\}\right\}.
\end{aligned} \tag{3.10}$$

The last equality follows from the assumption that $Q^0(\cdot)$ is a degenerate distribution. The Gibbs inequality ensures the information-cost term is always positive regardless of the realized $(\boldsymbol{\pi}_i, \mathbf{u}_i, \mu_i)$. Let superscript 0 denote the state before the changes in prices and/or choice-set composition, and let superscript 1 denote the state after the changes, respectively. CV_{RI}^i is defined by

$$CV_{RI}^i := \frac{1}{\beta_1^i}\left\{W_{RI}^i\left(\mathbf{u}_i^1, \mathcal{J}_i^1\right) - W_{RI}^i\left(\mathbf{u}_i^0, \mathcal{J}_i^0\right)\right\}, \tag{3.11}$$

where $-\beta_1^i$ is the price coefficient in the alternative-specific utility (3.9). CV_{RI}, compensating

[15] The result of the zero-income-effect follows directly from the fact that $\mathbf{H}$ and $\mathbf{H}^{-1}$ are homogeneous of degree 1 (Proposition 8 of Fosgerau et al.), combined with the assumption that $u_{i,j}$ is quasilinear in income.

variation for the target population, can be obtained by aggregating CV_{RI}^i across the population.[16][17][18]

Four important remarks are in order. First, when evaluating the welfare changes associated with a policy change such as pricing or choice-set composition, the expectation should be taken against the true distribution of consumption utilities $Q^0(\cdot)$, not consumers' subjective prior $Q_i(\cdot)$. It is particularly pertinent when a researcher wants to evaluate the actual welfare changes associated with a policy change, not the welfare changes as perceived by consumers. Second, $W_{RI}(\mathbf{u}_i, \mathcal{J}_i)$ is no longer monotonic in the number of alternatives $|\mathcal{J}_i|$. That is, consumer welfare can even be strictly smaller when more options are available from which to choose. This important feature can be attributed mainly to the discrepancy between $Q^0(\cdot)$ and $Q_i(\cdot)$ – the true alternative-specific consumption utility $\mathbf{u}_i$ may well be different from consumers' perceptions of it; that is, the subjective prior $Q_i(\cdot)$ can be biased. Third, $W_{RI}(\mathbf{u}_i, \mathcal{J}_i)$ separates the consumption utility $\{u_{i,j}\}_{j\in\mathcal{J}_i}$ from the informational choice-probability shifters inside $\Pr_i(i \text{ Chooses } j|\mathbf{u}_i)$,[19] thereby allowing for the informational choice-probability shifters such as promotions or previous purchase history to be contained only in $\Pr_i(i \text{ Chooses } j|\mathbf{u}_i)$, not in $\{u_{i,j}\}_{j\in\mathcal{J}_i}$. Including consideration shifters in the utility term $u_{i,j}$ to calculate the social-surplus function will mislead the measure of consumer welfare, because they are not considered contributions to *consumption values*, but rather tweaks to the choice probabilities through information. Fourth, (3.10) is defined up to the level of consumers' income y_i under the assumed quasilinear functional form of alternative-specific consumption utility (3.9). During the calculation of consumer-welfare changes associated with a policy change such as compensating variation, one can normalize $y_i = 0$ under the maintained assumption (3.9), namely, the quasi-linearity

[16]Note $W_{RUM}^i(\mathbf{u}_i, \mathcal{J}_i)$ and CV_{RUM}^i defined within the RUM framework have the following form:

$$W_{RUM}^i(\mathbf{u}_i, \mathcal{J}_i) = E_{\epsilon_{i,j}}\left[\max_{j\in\mathcal{J}_i}\{u_{i,j} + \epsilon_{i,j}\}\right] \tag{3.12}$$

$$CV_{RUM}^i := \frac{1}{\beta_1^i}\left\{W_{RUM}\left(\mathbf{u}_i^1, \mathcal{J}_i^1\right) - W_{RUM}\left(\mathbf{u}_i^0, \mathcal{J}_i^0\right)\right\}. \tag{3.13}$$

When $\epsilon_{i,j}$s are distributed following the i.i.d. double exponential, the $E\max$ term of (3.12) reduces to the well-known log-sum formula. Otherwise, the $E\max$ terms are usually calculated using simulation integrals.

[17]Suppose $\mathbf{u}_i$ contains random coefficients or random effects, which can be interpreted as heterogeneous taste parameters or alternative-specific utility components that are observable only to the consumers and not to the researcher. A typical situation is that only the distribution of β_1^i is estimated or known to the researcher. Then, CV_{RI}^i and CV_{RUM}^i can be evaluated by the following formulas:

$$CV_{RI}^i = E\left[\frac{1}{\beta_1^i}\left\{W_{RI}^i\left(\mathbf{u}_i^1, \mathcal{J}_i^1\right) - W_{RI}^i\left(\mathbf{u}_i^0, \mathcal{J}_i^0\right)\right\}\right]$$

$$CV_{RUM}^i = E\left[\frac{1}{\beta_1^i}\left\{W_{RUM}^i\left(\mathbf{u}_i^1, \mathcal{J}_i^1\right) - W_{RUM}^i\left(\mathbf{u}_i^0, \mathcal{J}_i^0\right)\right\}\right],$$

where the expectation is taken against the distribution of random coefficients or random effects.

[18]$\frac{1}{\beta_1^i}$ is the factor that converts "utils" to a monetary term. If the alternative-specific consumption utility is defined by, for example,

$$u_{i,j} := (y_i - p_j)\,\beta_1^i\mu_i + \chi_{i,j}\mu_i,$$

the factor multiplied to $W_{RI}^i(\mathbf{u}_i, \mathcal{J}_i)$ should be $\frac{1}{\beta_1^i\mu_i}$.

[19]Recall (3.14).

of alternative-specific consumption utility in income. This feature is particularly useful because the individual consumer's income level is typically not observable a researcher.

In practice, when $(\boldsymbol{\pi}_i, \mathbf{u}_i, \mu_i)$ is estimated from the data, a consistent prediction $(\hat{\boldsymbol{\pi}}_i, \hat{\mathbf{u}}_i, \hat{\mu}_i)$ can be used in place of $(\boldsymbol{\pi}_i, \mathbf{u}_i, \mu_i)$. I provide some further details and discussions on the implementation of the proposed consumer-welfare-evaluation framework in appendix C.

3.4 Parametrization of the Choice Probabilities for Empirical Application

In this subsection, I explore further possibilities of parametrizations on the conditional-choice-probability expression (3.7).

Let us consider (3.7), which we take as the key building block for the likelihood:

$$\Pr_i (i \text{ Chooses } j|\mathbf{u}_i) = \frac{H_j \left(\exp\left(\ln\left(\mathbf{H}^{-1}(\boldsymbol{\pi}_i)\right) + \mu_i \mathbf{u}_i\right)\right)}{\sum_{k\in\mathcal{J}_i} H_k \left(\exp\left(\ln\left(\mathbf{H}^{-1}(\boldsymbol{\pi}_i)\right) + \mu_i \mathbf{u}_i\right)\right)}. \tag{3.14}$$

Without the term $\ln\left(\mathbf{H}^{-1}(\boldsymbol{\pi}_i)\right)$, the choice probability lines up exactly with that of RUM with an error-term distribution corresponding to the function $\mathbf{H}(\cdot)$. In other words, the choice-probability expression above is equivalent to the RUM choice probability with the utility vector shifted by $\ln\left(\mathbf{H}^{-1}(\boldsymbol{\pi}_i)\right)$, which is not directly related to the consumption utility vector $\mathbf{u}_i$. Furthermore, the positive partial derivatives of $\mathbf{H}$ and $\mathbf{H}^{-1}$ with respect to its arguments motivate us to define a mapping:

$$\tilde{\boldsymbol{\pi}}_i := \ln\left(\mathbf{H}^{-1}(\boldsymbol{\pi}_i)\right).$$

[20] (3.14) simplifies to

$$\Pr_i (i \text{ Chooses } j|\mathbf{u}_i) = \frac{H_j \left(\tilde{\boldsymbol{\pi}}_i \circ \exp(\mu_i \mathbf{u}_i)\right)}{\sum_{k\in\mathcal{J}_i} H_k \left(\tilde{\boldsymbol{\pi}}_i \circ \exp(\mu_i \mathbf{u}_i)\right)}, \tag{3.15}$$

where $\circ$ denotes the Hadamard product.

One remarkable feature of (3.15) is that $\{\pi_{i,j}\}_{j\in\mathcal{J}_i}$ or $\{\tilde{\pi}_{i,j}\}_{j\in\mathcal{J}_i}$ can shift the choice probabilities without affecting the consumption utility $\{u_{i,j}\}_{j\in\mathcal{J}_i}$. By contrast, all the choice-probability shifters must be contained in the "utility" term $u_{i,j}$ in the RUM framework, which is unrealistic, because choice-probability shifters such as advertisement or promotions only tweak consumers' information, not the consumption value. Formulating $\tilde{\boldsymbol{\pi}}_i$ as a function of consideration shifters would be reasonable considering the structure of consumers' optimization problem described in the beginning of the current subsection. Furthermore, the expression $\tilde{\boldsymbol{\pi}}_i$ in (3.15) allows for a wide range of flexibility in the choice of the functional form. The simplest parametrization of $\tilde{\pi}_{i,j}$ would be $\tilde{\pi}_{i,j} \propto \exp\left(\mathbf{d}'_{i,j}\gamma\right)$, where $\mathbf{d}_{i,j}$ is a vector that shifts the probability of an alternative j considered in the alternative-

[20] $\mathbf{H}$ and $\mathbf{H}^{-1}$ are homogeneous of degree 1 (Proposition 8 of Fosgerau et al.); therefore, one can rescale $\tilde{\boldsymbol{\pi}}_i$ so that $\tilde{\boldsymbol{\pi}}_i$ sums up to 1. Nonnegative partial derivatives of $\mathbf{H}$ follow from the convexity of McFadden's social-surplus function defined in the RUM context, which guarantees the monotonicity of $H_j(\cdot)$ and $H_j^{-1}(\cdot)$.

choice stage. Such a parametrization is well microfounded, and it also allows the estimation problem to be substantially simpler than that of the extant consideration-set literature that has to include all the possible permutations of the consideration sets in the model likelihood. The parametrization allows for applied researchers to estimate the RI models in the same way as if estimating various RUM models with additional utility shifters $\mathbf{d}_{i,j}$.[21]

4 Application: Introduction of Pods Laundry Detergent

Tide introduced its "pods" laundry detergents to the market early in 2012. In this section, I quantify and compare the consumer-welfare changes associated with the introduction of "pods" laundry detergents in 2012, respectively, using the consumer welfare formulas from RI and RUM frameworks.

4.1 Data

The data used are a combination of Nielsen-Kilts Homescan and Retail Measurement Services (RMS henceforth) data. Homescan data record all the consumer packaged goods (CPG henceforth) items purchased in the panel households, using the barcode scanner issued to each household, along with the identity of the store where purchased if available. RMS data record one third to one half of the entire U.S. CPG transactions. I matched the shopping trips of the Homescan data with the RMS weekly sales using the store code and week information to construct the estimation data, because Homescan does not provide the set of alternatives that the participating households were facing. Then, I dropped the stores that do not provide the display and feature information in RMS. The matching and cleaning process resulted in 170,968 choice observations with 17 million alternative observations throughout the sample period 2006-2016. Further details about data matching and the cleaning procedure are relegated to Online Appendix F.

4.2 Maximum Likelihood Estimation of Model Parameters Using Individual-level Choice Data

In this subsection, I describe the details about the utility specification, likelihood, identification, and inference for the maximum likelihood estimation of the RI-based empirical model of consumer choice introduced in the previous section, using the individual-level laundry detergent choice data.[22] Then, I examine the contribution of the consideration shifters in model fit to compare the performance of the RI-based empirical model of discrete-choice with the RUM-based discrete-choice model.

[21] The original model by Matějka and McKay (2015); Fosgerau et al. (2019) is much richer in the sense that any prior distribution over consumption utilities is allowed, from which interesting choice patterns may occur. The parametrization of $\tilde{\boldsymbol{\pi}}_i$ proposed here simplifies and restricts the prior distribution in a specific way that is initialized at the symmetric prior, and then dictated by the distribution of $\mathbf{d}_{i,j}$. The RI-theory model is also restricted here in a way that does not allow for the possibilities of zero unconditional choice probabilities.

[22] If only market-level data are available, the model can also be estimated using the market-share inversion *á la* Berry (1994).

Utility Specification Let $j\,(\in \mathcal{J}_i)$ denote an alternative; p_j, the per-package price; and $\mathbf{x}_j$, characteristics of item j, including brand and package-size dummies. Each alternative j denotes an item that is available on the shelf. $j'\,(\neq j)$ can denote the same product with j that has a different package size. Let $\chi_{i,j}$ of equation (3.9) in section 3.3 take the following form:

$$\chi_{i,j} = \beta_0 + \mathbf{x}_j'\boldsymbol{\beta}_2 + \eta_{i,j},$$

where $\mathbf{x}_j$ is product attributes observable to a researcher, and $\eta_{i,j}$ represents the utility from product attributes that are not observable to the researcher. Assume further that the utility parameter $\beta_1^i = \beta_1$ for all i. The alternative-specific utility of consumer i purchasing item j is then

$$\begin{aligned} u_{i,j} &:= \beta_0 - p_j\beta_1 + \mathbf{x}_j'\boldsymbol{\beta}_2 + \eta_{i,j} \qquad \text{for } j \neq 0 \\ u_{i,0} &:= 0. \end{aligned} \tag{4.1}$$

[23] The intercept term β_0, which is normalized as 1, can be interpreted as the mean relative utility of consuming any laundry detergent in comparison to the outside option.

For empirical tractability, I specify further structures on the unconditional choice probabilities $\{\pi_{i,j}\}_{j\in\mathcal{J}_i}$.[24] First, I model that the prior belief distribution $Q_i(\mathbf{u}_i) \equiv Q(\mathbf{u}_i; \mathcal{J}_i, \mathbf{D}_i)$ is affected only by the set of alternatives $\mathcal{J}_i$ and the dummies $\mathbf{D}_i$, which represent the previous purchase history and promotion status. To that end, I parameterize the unconditional choice probability $\pi_{i,j} \equiv \pi_j(\mathcal{J}_i, \mathbf{D}_i)$ as proportional to $\exp\left(\mathbf{d}_{i,j}'\boldsymbol{\gamma}\right)$, that is,

$$\pi_{i,j} \propto \exp\left(\mathbf{d}_{i,j}'\boldsymbol{\gamma}\right), \tag{4.2}$$

where $\mathbf{d}_{i,j}'\boldsymbol{\gamma}$ captures the informational choice-probability shifters. $\mathbf{d}_{i,j}$, the vector of previous purchase history and promotions, plays a role as a consideration shifter that does not affect the alternative-specific utility $u_{i,j}$ that consumer i gains from consuming an alternative j. Next, I model consumers' unit cost of information that depends on the demographics and the consumer's familiarity with the specific store in which the consumer was shopping. Specifically,

$$\mu_i := \exp\left(\mathbf{w}_i'\boldsymbol{\theta}\right), \tag{4.3}$$

where $\mathbf{w}_i$ is the vector of demographics that does not include the constant term.

Information Cost Function and Likelihood To operationalize the RI-based discrete-choice and consumer-welfare-evaluation framework developed in section 3, I use the Shannon entropy for

[23]Note that the common term $y_i\beta_1$ is canceled out from $u_{i,j}$ and $u_{i,0}$.

[24]$\tilde{\pi}_{i,j} = \pi_{i,j}$ when $\mathbf{H}$ is an identity function.

the information-cost function, in which case $\mathbf{H}$ is an identity function.[25] The resulting conditional choice probability takes the form of simple logit:

$$\Pr_i\left(i \text{ Chooses } j|\mathbf{u}_i\right) = \frac{\exp\left(\mathbf{d}'_{i,j}\boldsymbol{\gamma} + \mu_i\left[\beta_0 - p_j\beta_1 + \mathbf{x}'_j\boldsymbol{\beta}_2 + \eta_{i,j}\right]\right)}{1 + \sum_{j'\in\mathcal{J}_i\setminus 0}\exp\left(\mathbf{d}'_{i,j'}\boldsymbol{\gamma} + \mu_i\left[\beta_0 - p_{j'}\beta_1 + \mathbf{x}'_{j'}\boldsymbol{\beta}_2 + \eta_{i,j'}\right]\right)}. \tag{4.4}$$

Because $\eta_{i,j}$ terms are not observable to the researcher, I model the distribution of $\eta_{i,j}$ as $\eta_{i,j} \sim$ i.i.d. $\mathcal{N}\left(0, \sigma^2\right)$, take $\eta_{i,j}$ as random effects, and estimate σ along with other model parameters.

Note in the choice probability (4.4) that μ_i captures the observed heterogeneity in the unit cost of information acquisition. Larger μ_i implies the consumer reacts more sensitively to the price and attribute differences. The choice-probability expression inherently contains the components of individual coefficients, in that the effective utility coefficient is $(-\mu_i\beta_1, \mu_i\boldsymbol{\beta}_2)$, where μ_i captures the heterogeneity in the unit information cost of each consumer. If, for example, $\mu_i \to +\infty$, which implies consumer i's unit information cost is zero, the consumer will choose the alternative that yields the highest $u_{i,j}$ among her/his choice set with probability 1, and the consideration shifters $\mathbf{d}_{i,j}$ will not play any role in the conditional choice probability. On the other hand, if $\mu_i = 0$, the information is infinitely costly, and thus consumers will simply choose to use the information contained in the consideration shifters $\mathbf{d}_{i,j}$.

The fact that the parameter μ_i, which is the inverse of the unit information cost, is multiplied by $u_{i,j}$ in the conditional-probability specification would have an important implication on the effectiveness of the consideration shifters $\mathbf{d}_{i,j}$: the effect of consideration shifters would be stronger for the individuals with smaller μ_i. Suppose, for example, individual i has an infinitely large unit information cost so that $\mu_i = 0$. In such a case, the actual consumption utility $u_{i,j}$ does not affect the resulting choice probability; thus, $\mathbf{d}_{i,j}$ is the only factor that affects the resulting choice probability. Because $\mathbf{d}_{i,j}$ includes promotional activities, promotions would be more effective for individuals with a higher unit-information cost, that is, smaller μ_i.

One last factor to consider is the panel-projection weight ω_i provided by Nielsen, taken as sampling weights. ω_i makes our likelihood expression slightly nonstandard. The maximum-likelihood problem is given by

$$\max_{(\boldsymbol{\gamma},\boldsymbol{\beta},\boldsymbol{\theta},\sigma)} \prod_{i,j} E_{\boldsymbol{\eta}_i}\left[\Pr_i\left(i \text{ Chooses } j\right)\right]^{\omega_i \mathbf{1}(i \text{ Chooses } j)},$$

where the expectation is taken against the distribution of $\eta_{i,j}$s. I use simulation integrals to approximate for the expectation during implementation.

[25] Note again that different possibilities can exist in the information-cost function than in the Shannon entropy. A different functional form of $\mathbf{H}(\cdot)$ in the information-cost function will lead to a different functional form of the conditional choice probability, and therefore the likelihoods. Some possibilities of $\mathbf{H}(\cdot)$ are listed in Online Appendix G.

Identification The consideration-shifter parameter $\boldsymbol{\gamma}$ is identified by the variation in the promotional activities and the household's own purchase history, under the assumption that those information shifters do not directly affect the consumption utility. The utility parameters $(-\beta_1, \boldsymbol{\beta}_2)$ are identified from the cross-product variations in the prices and product attributes, as well as the variations in the choice set $\mathcal{J}_i$ in each shopping instance. The information-cost parameter $\boldsymbol{\theta}$ is identified using the cross-household variations in the demographics and familiarity with the specific store that the household was shopping in. Because identifying μ_i and α simultaneously is impossible, I normalize the utility of consuming one pack of detergent as fixed at $\beta_0 = 1$, which allows μ_i to be identified against β_0.

4.3 Estimation Results and Model Fit

Model-Parameter Estimation Results Table 3 summarizes the model-parameter estimation results. The effective magnitude of the utility parameters $\left(-\hat{\beta}_1, \hat{\boldsymbol{\beta}}_2, \hat{\sigma}\right)$ are $\left(-\hat{\beta}_1\mu_i, \hat{\boldsymbol{\beta}}_2\mu_i, \hat{\sigma}\mu_i\right)$, which is specific to each individual i. The implied *Mean Price Elasticity* is –2.133, which is of a reasonable magnitude.[26] All the major brand coefficients are positive and statistically and economically significant, implying a substantial brand premium exists in this market.

The $\hat{\boldsymbol{\gamma}}$ estimates indicate the consideration shifters have strong effects on consumers' choice. In particular, strong inertia effects on the choice of laundry detergents seem to exist, as captured by the coefficients of *Same UPC Purchased Within 1 Year* and *Same Product (Different UPC) Purchased Within 1 Year* variables.

The unit-information-cost parameter estimates $\hat{\boldsymbol{\theta}}$ relate to the shopping environment and demographics. Because μ_i is the inverse of the unit information cost, the positive $\hat{\boldsymbol{\theta}}$ component implies the corresponding variable affects the unit information cost negatively. For example, if the household has visited the specific store before and is therefore familiar with it, learning the locations of shelves will require less time and effort, as reflected in a positive coefficient estimate of *Visited the Same Store Within 1 Year* variable. Estimates imply households with higher income and a household head with a college degree exhibit a higher unit information cost, whereas the presence of a non-working spouse, living in an apartment, and not having children exhibit a lower unit information cost. Except for the fact that the household head being employed tends to have lower unit information cost, $\hat{\boldsymbol{\theta}}$ estimates are broadly consistent with the opportunity-cost-of-time story, and therefore with the search-cost story. Thus, we can see that the unit information cost parameterized in our RI framework provides face validity to the results.

[26] Under the functional form of the conditional choice probability (4.4), the own-price elasticity is given by

$$E_{\boldsymbol{\eta}_i}\left[\frac{\partial \Pr_i (i \text{ Chooses } j|\mathbf{u}_i)}{\partial p_j}\right]\frac{p_j}{E_{\boldsymbol{\eta}_i}\left[\Pr_i (i \text{ Chooses } j|\mathbf{u}_i)\right]} = -\hat{\beta}_1\mu_i p_j \frac{E_{\boldsymbol{\eta}_i}\left[\Pr_i (i \text{ Chooses } j|\mathbf{u}_i)\left(1-\Pr_i (i \text{ Chooses } j|\mathbf{u}_i)\right)\right]}{E_{\boldsymbol{\eta}_i}\left[\Pr_i (i \text{ Chooses } j|\mathbf{u}_i)\right]},$$

which uses the approximation of market shares s_j as $s_j \simeq E_{\boldsymbol{\eta}_i}\left[\Pr_i (i \text{ Chooses } j)\right]$. I also assume the regularity conditions for differentiation under the integral sign hold.

Table 3: Model-Parameter Estimates

Mean Price Elasticity		–2.133			
Utility Parameter $\left(-\hat{\beta}_1, \hat{\boldsymbol{\beta}}_2, \hat{\sigma}\right)$				Consideration Shifter Parameter $\hat{\boldsymbol{\gamma}}$	
Per-pack Price $\left(-\hat{\beta}_1\right)$	–0.133***	75oz. ≤ Pack ≤150oz.	–0.212***	In-store Display	0.564***
	(0.009)		(0.036)		(0.018)
Pack ≤75oz.	–0.522***	150oz. ≤ Pack ≤225oz.	0.027	Feature Ad	0.765***
	(0.026)		(0.050)		(0.015)
All	0.363***	Tide	0.576***	Same UPC	3.405***
	(0.028)		(0.041)	Purchased Within 1 Year	(0.019)
Arm&Hammer	0.330***	Wisk	0.434***	Same Product (Different UPC)	1.841***
	(0.026)		(0.034)	Purchased Within 1 Year	(0.020)
Gain	0.315)***	Xtra	0.200***		
	(0.026)		(0.021)	Unit Information Cost Parameter $\hat{\boldsymbol{\theta}}$	
Purex	0.249***			Visited the Same Store	0.214***
	(0.021)			Within 1 Year	(0.027)
Powder	0.056***	Oxi-Clean	0.074***	Income Ratio to FPL	–0.058***
	(0.011)	/ Baking Soda	(0.013)		(0.006)
Fabric Softener	–0.159***	Colorsafe	0.158***	Apartment	0.074**
	(0.032)		(0.017)		(0.033)
Febreze	0.004	Soft	0.066**	Non-working Spouse	0.176***
	(0.013)		(0.029)		(0.026)
All Temperature	0.060***	Stain Remover	0.136***	Household Size	0.072***
	(0.016)	/ Deep Clean	(0.015)		(0.012)
Bleach	–0.080***	Unscented /	0.008	Head Employment	0.253***
	(0.011)	Sensitive / Baby	(0.007)		(0.033)
Ultra	0.035***	Low Cl / S / P	0.024**	Head College Degree	–0.064***
	(0.007)		(0.010)		(0.023)
$n\times$Concentrated	0.044***	Pod / Tablet / Sheet	–0.058***	Married,	0.008
	(0.005)		(0.015)	Living Together	(0.026)
High Efficiency	0.097***	sd$(\eta_{i,j})$ $(\hat{\sigma})$	0.239***	No Child	0.295***
	(0.009)		(0.048)		(0.033)
Choice Obs.	170698	Sample Size	17M	$\bar{\mu} = \overline{\exp\left(\mathbf{w}_i'\hat{\boldsymbol{\theta}}\right)}$	1.829
AIC	1171000	BIC	1171403	Log-likelihood	–585460

*p<0.1; **p<0.05; ***p<0.01

Note. This table summarizes the model-parameter-estimation results from the maximum likelihood estimation. Mean elasticity and $\bar{\mu} = \overline{\exp\left(\mathbf{w}_i'\hat{\boldsymbol{\theta}}\right)}$ are calculated after weighting for the panel-projection factor. Due to the utility specification of the model, μ_i should be multiplied by the utility-parameter coefficients $\left(-\hat{\beta}_1, \hat{\boldsymbol{\beta}}_2\right)$ to find the effective magnitudes. Comparing the magnitudes of $\left(-\hat{\beta}_1, \hat{\boldsymbol{\beta}}_2\right)$ with the promotion parameter $\hat{\boldsymbol{\gamma}}$ estimates is only sensible after multiplying μ_i's by $\left(-\hat{\beta}_1, \hat{\boldsymbol{\beta}}_2\right)$. The data used for the estimation are the matched sample of laundry detergent purchases in the Nielsen-Kilts consumer panel data and scanner data for households during 2006-2016. Standard-error estimates are in parentheses. AIC, BIC, and log-likelihood are calculated after adjusting the panel-projection factor weights sum up to the number of choice observations.

In-Sample and Out-of-Sample Model Fits with Different Specifications on the Information Shifters Asking how our RI framework performs empirically relative to the RUM framework is worthwhile. To this end, I compare the in-sample and out-of-sample fits for alternative specifications. Table 4 compares the model-fit measures for alternative specifications on the information shifters. To compare the out-of-sample predictive model fit, I randomly subsampled the 10,000 holdout choices and estimated the model with the remaining 160,698 choices. Column (1) corresponds to our preferred specification, the discrete-choice RI model that incorporates the consideration shifters fully, and column (16) corresponds to the discrete-choice RUM model without any consideration shifters or parametrizing the dispersion parameter.[27] Although not reported here for brevity, the estimated model parameters of column (1) are virtually the same as what are reported in Table 3, and the utility-parameter estimates $\left(-\hat{\beta}_1, \hat{\boldsymbol{\beta}}_2, \hat{\sigma}\right)$ are similar across all the specifications (1)-(16).

Three in-sample fit measures (log-likelihood, AIC, and BIC) and two out-of-sample prediction measures (hit rate and average hit probability) are applied to the holdout sample and are reported in Table 4. Hit rate is defined by the percentage of times the choice alternative with the highest predicted choice probability is actually chosen out of 10,000 holdout samples. Average hit probability is defined by the average predicted choice probability of the chosen alternative. Because, on average, each choice instance has around 100 choice alternatives, hit rate and average hit probability would be around 0.01 if the specified model were simply choosing alternatives randomly. One can immediately note that incorporating the consideration shifters in the model improves model fit substantially by comparing all five model-fit measures across columns (1)-(16). A simple comparison between column (1) (RI-logit with all the consideration shifters) and column (16) (RUM-logit without any consideration shifters) shows a remarkable difference, in that the hit rate improves more than twice, and average hit probability more than four times. Among other consideration shifters included and compared in the model, own purchase history improves model fit the most. The results of model-fit comparison presented here reconfirm the importance of incorporating the consideration shifters in the likelihood, and the RI-based discrete-choice framework developed in section 3 provides a proper microfoundation for incorporating both the consideration shifters and utility shifters in the likelihood.

4.4 Counterfactual Welfare Simulation Associated with Tide Pods' Introduction

The goal of this subsection is to calculate the consumers' benefit from the entry of Tide Pods laundry detergent, using the RI-based welfare formula proposed in section 3.3, and to compare it with the welfare calculation from the conventional RUM-based welfare formula.

The CV with respect to the introduction of Tide Pods into the market is calculated as follows. Recall the consumer-choice model is estimated by pooling the pre- and post- Tide Pods' introduction data. The model-parameter estimates $\left(-\hat{\beta}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\theta}}, \hat{\sigma}\right)$ reported in Table 3 are taken as

[27] Or the unit information cost μ_i^{-1} in the RI context.

Table 4: Comparison of Model Fit across Different Specifications on the Consideration Shifters

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log-likelihood	–551251	–552190	–561205	–562306	–563028	–563959	–671663	–673312
AIC	1102504	1104382	1122413	1124614	1126058	1127920	1343327	1346627
BIC	1102514	1104392	1122423	1124624	1126068	1127930	1343337	1346637
Hit Rate	0.259	0.259	0.257	0.257	0.238	0.237	0.110	0.109
Average Hit Probability	0.102	0.101	0.097	0.096	0.087	0.086	0.024	0.024
Display / Feature	O	O	O	O	O	O	O	O
Own UPC Purchase History	O	O	O	O	X	X	X	X
Own Product Purchase History	O	O	X	X	O	O	X	X
Demographics as Unit Information Cost	O	X	O	X	O	X	O	X

	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Log-likelihood	–557638	–558641	–567575	–568737	–569729	–570755	–679598	–681322
AIC	1115277	1117283	1135151	1137475	1139461	1141512	1359198	1362645
BIC	1115287	1117293	1135161	1137485	1139471	1141522	1359208	1362655
Hit Rate	0.259	0.260	0.262	0.261	0.241	0.240	0.116	0.115
Average Hit Probability	0.099	0.098	0.095	0.093	0.083	0.082	0.021	0.021
Display / Feature	X	X	X	X	X	X	X	X
Own UPC Purchase History	O	O	O	O	X	X	X	X
Own Product Purchase History	O	O	X	X	O	O	X	X
Demographics as Unit Information Cost	O	X	O	X	O	X	O	X

Note. AIC, BIC, and log–likelihood are calculated after scaling the panel-projection factor weights so that they sum up to the number of choice observations. The estimation sample has 160,698 choices with a total of 15,989,188 alternatives. The holdout sample has 10,000 choices with a total of 994,378 alternatives.

fixed. Let $\hat{u}_{i,j}$, $\hat{\mu}_i$, $\hat{\pi}_{i,j}$, and $\hat{\Pr}_i\left(i \text{ Chooses } j|\hat{\mathbf{u}}_i\right)$ be their consistent predictions obtained by plugging the estimated parameters $\left(-\hat{\beta}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\theta}}, \hat{\sigma}\right)$ back into the respective functions (4.2), (4.3), and (4.4) in section 4.2. Using the formulas (3.10) and (3.12) provided in section 3.3, $W^i_{RI}\left(\hat{\mathbf{u}}^1_i, \mathcal{J}^1_i\right)$ and $W^i_{RUM}\left(\hat{\mathbf{u}}^1_i, \mathcal{J}^1_i\right)$ are evaluated using the post-Tide-Pods-introduction data. Tide Pods laundry detergent items are removed from the consumers' choice set to evaluate $W^i_{RI}\left(\hat{\mathbf{u}}^0_i, \mathcal{J}^0_i\right)$ and $W^i_{RUM}\left(\hat{\mathbf{u}}^0_i, \mathcal{J}^0_i\right)$ using the same set of formulas.[28] CV^i_{RI} and CV^i_{RUM} are then calculated following the formulas (3.11)

[28] Because we have assumed the Shannon-entropy cost function that yields the simple logit form of the conditional choice probabilities, the function $\mathbf{H}(\cdot)$ is an identity function in this case, which simplifies (3.10) in section 3.3 as the following:

$$W^i_{RI}\left(\hat{\mathbf{u}}_i, \mathcal{J}_i\right) = \sum_{j\in\mathcal{J}_i} \hat{\Pr}_i\left(i \text{ Chooses } j|\hat{\mathbf{u}}_i\right)\hat{u}_{i,j} - \hat{\mu}_i^{-1}\left\{\sum_{j\in\mathcal{J}_i} \hat{\Pr}_i\left(i \text{ Chooses } j|\hat{\mathbf{u}}_i\right)\left\{\ln \hat{\Pr}_i\left(i \text{ Chooses } j|\hat{\mathbf{u}}_i\right) - \ln \hat{\pi}_{i,j}\right\}\right\}. \quad (4.5)$$

and (3.13), respectively.

The welfare counterfactual analysis conducted in this subsection is *ceteris paribus* except for removing the Tide Pods laundry detergent from consumers' choice set from post-Tide-Pods-introduction data. Equilibrium price and promotion responses by removing Tide Pods laundry detergent from consumers' choice set are ignored for the following two reasons. First, many other laundry detergents are in the market where Tide Pods takes only around 5% of the market shares measured in dollars; therefore, the equilibrium responses from competitors by removing Tide Pods are likely to be small. Second, the supplier side, which is necessary to determine the equilibrium responses in prices and/or other marketing-mix variables from the competitors, is not explicitly modeled here.[29] The rest of the welfare function and CV calculation follows as explained in section 3.3.

Figure 4.1 depicts the distribution of the CV with respect to Tide Pods' entry into the market. Each observation corresponds to each shopping instance in the estimation data. Note that adding Tide Pods laundry detergent to consumers' choice set does not monotonically increase CV^i_{RI}, and only around one third of CV^i_{RI} are positive. By contrast, CV^i_{RUM} is constantly positive and tends to be much higher than CV^i_{RI}.

Average per Shopping Trip row of Table 5 summarizes the average CV per shopping trip. Note that average CV^i_{RI} per shopping trip is slightly negative, whereas average CV^i_{RUM} is positive. Comparing columns (1) and (2), CV^i_{RUM} overestimates the CV not just in terms of magnitude, but also in signs, compared to CV^i_{RI}. *Annually Projected* row of Table 5 presents the results after projecting the results to the entire U.S. population using the panel-projection factors. CV^i_{RI} is –\$8 million, CV^i_{RUM} is \$73 million, and the total annual Pods detergent sales equal \$346 million. CV^i_{RUM} amounts to more than one fifth of the total annual Pods detergents sales, which is unrealistic because consumers may easily switch to other laundry detergents that are close substitutes.[30]

Taking the product characteristics of Tide Pods as given, while taking the prices as something suppliers can adjust, the next question related to consumer welfare concerns pricing. To that end, I examine how much price cuts are necessary to make an average consumer just as well off as before the introduction of Tide Pods. Table 6 presents the results of counterfactual simulations on Tide Pods pricing. The table presents the counterfactual CV_{RI} aggregated and projected to the population when, *ceteris paribus*, Tide Pods laundry detergent has n% price discounts. Approximately 6%-7%

The model corresponding to RUM, which does not include consideration shifters or the parameterized unit-information-cost function, is also separately estimated to find the welfare changes calculated using the RUM-based log-sum formula. See Online Appendix D.2 for the model-parameter estimates of the RUM model that does not include any informational-choice-probability shifters in the likelihood.

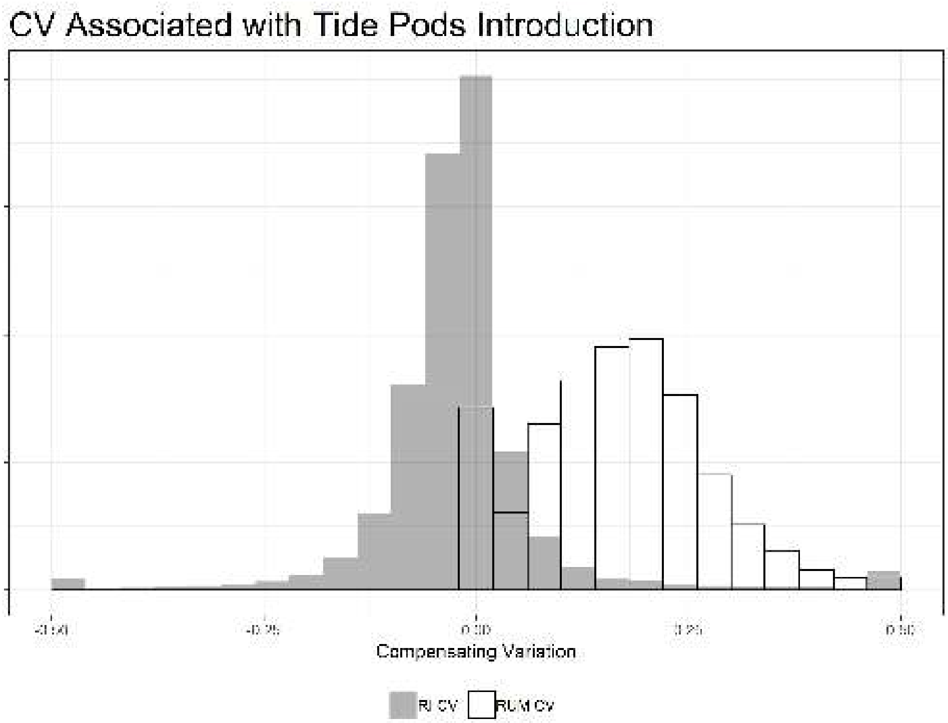
[29] The consumer-choice and consumer-welfare-evaluation framework developed in section 3 does not prevent incorporating equilibrium responses during the counterfactual consumer-welfare evaluation such as in Petrin (2002); Morozov (2019), as long as a researcher is willing to explicitly model the supplier side, including the mode of competition, equilibrium concept, and pricing strategy.

[30] Online Appendix D.2 reports the CV calculated using the formula based on full-information RUM after including all the informational-choice-probability shifters. It still overestimates the gains from the introduction of Tide Pods, thereby suggesting the result reported in Table 5 is not because of omitting the informational-choice-probability shifters in the likelihoods, but because of the differences in the welfare-evaluation formula itself.

price discounts can make an average consumer just as well off as before the introduction of Tide Pods. The results presented in Table 6 suggest the prices of Tide Pods were higher than the utility values they provided to an average consumer.

Online Appendix E reports robustness-check results using different empirical specifications. In all empirical specifications compared, the finding that the average consumer welfare gains from Tide Pods' introduction evaluated using the RI framework is slightly negative, whereas the full-information RUM framework substantially exaggerates the consumer-welfare gains, is robust. The robustness-check results suggest the differences between CV_{RI} and CV_{RUM} reported in this section are not due to a specific empirical specification of the choice model; they suggest the differences should be attributed to the differences in the welfare formula itself.

Figure 4.1: Compensating Variation Associated with Tide Pods' entry



Note. The histogram illustrates the simulated compensating variation (CV) associated with the introduction of Tide Pods to the market. The RI CV change is calculated using the formula (3.11), and the RUM CV change is calculated using the formula (3.12). The model parameter used to simulate the change in CV is from Table 3. The data used for CV simulation are the 2012-2016 subsample of the estimation data, which is after the introduction of Tide Pods in 2012. The frequency is weighted for the panel-projection factor.

Table 5: Average and Annually Projected Compensating Variation per Shopping Trip Associated with Tide Pods' Introduction

	(1) CV_{RI}	(2) CV_{RUM}
Average per Shopping Trip	–\$0.019	\$0.157
Annually Projected	–\$8,028,950	\$72,538,564
Pods Sales/Yr	\$346,294,884	
Detergent Sales/Yr	\$6,449,062,852	

Note. The *Annually Projected* row of CV_{RI} and CV_{RUM} columns are calculated using the subsample of the estimation data during 2012-2016, and then projecting to the entire U.S. population using the panel-projection factors. *Pods Sales/Yr* and *Detergent Sales/Yr* rows are calculated using the total Nielsen RMS laundry detergent sales of 2012-2016, multiplied by 2.5 and then annualized.

Table 6: Counterfactual Tide Pods Price Discount and CV_{RI}

Discount	0%	5%	6%	7%	8%	10%	15%
CV_{RI}	–\$8,028,950	–\$1,824,098	–\$519,362	\$807,681	\$2,157,505	\$4,927,456	\$12,286,636

Note. This table presents the hypothetical annually projected CV_{RI} of Tide Pods' introduction when the Tide Pods laundry detergents are uniformly discounted by n%.

5 Conclusion

This paper proposes an empirical framework of discrete choice and consumer-welfare evaluation built upon RI theory, where the source of stochasticity in consumer choice probability stems from the corresponding information-acquisition costs. Unlike the conventional RUM models assuming consumers' complete knowledge of their consumption utility on all alternatives in the choice set, the proposed RI model takes into account consumers' degree of (in)attention to the consumption utility and the corresponding information-acquisition costs. The framework allows consumer welfare to decrease when an inferior option is added to a choice set, primarily due to the information-acquisition cost that leads to consumers' incomplete information acquisition.

Based on the conventional knowledge of consumer welfare from the workhorse RUM models, managers may be misled to believe consumers may always benefit from a larger choice set, which is related to a firm's product-line decision. However, the application of the proposed framework in the laundry detergent category shows an introduction of a new product may harm consumers' welfare, depending on the consumption utility of the additional product to the product line and individual consumers' information-acquisition cost. We find the introduction of Tide Pods actually reduced consumer welfare, on average, which is not possible to capture under the conventional RUM framework. After the attributes of a new product are determined, the remaining managerial decision is pricing, which may either mitigate or aggravate the effects on consumer welfare. The counterfactual analysis on the pricing of Tide Pods suggests the introduction of Tide Pods detergents would improve average consumer welfare only if the manufacturer were to reduce prices by at least 6%-7%.

The empirical case study of Tide Pods' introduction demonstrates the RI-based consumer-welfare-evaluation framework developed in the present paper can be used as a rationale to suggest manufacturers' and retailers' decision-making in various layers, which may include, but is not limited to, composition of individual product attributes during a new-product development, product-line design, product curation of a category, and pricing. The framework can be useful in contexts where customer satisfaction is a key concern of a marketing manager, so the proper evaluation of consumer welfare becomes imperative.

As a final remark, the framework developed in this paper maintained the assumption that the alternative-specific utility is quasilinear in income; therefore, the income effect is zero. A useful extension of the RI-based discrete-choice and consumer-welfare-evaluation framework developed in this paper would be to allow for a non-quasilinear form of alternative-specific consumption utility, thereby allowing for possibly nonlinear income effects.

References

Allenby, G. M. and J. L. Ginter (1995): "The effects of in-store displays and feature advertising on consideration sets," *International Journal of Research in Marketing*, 12, 67–80.

Anderson, E. T. and D. Simester (2013): "Advertising in a competitive market: The role of product standards, customer learning, and switching costs," *Journal of Marketing Research*, 50, 489–504.

Anderson, E. T. and D. I. Simester (1998): "The role of sale signs," *Marketing Science*, 17, 139–155.

——— (2001): "Are sale signs less effective when more products have them?" *Marketing Science*, 20, 121–142.

Bass, F. M. (1974): "The theory of stochastic preference and brand switching," *Journal of Marketing Research*, 11, 1–20.

Bass, F. M., A. Jeuland, and G. P. Wright (1976): "Equilibrium stochastic choice and market penetration theories: Derivations and comparisons," *Management Science*, 22, 1051–1063.

Berry, S. (1994): "Estimating discrete-choice models of product differentiation," *RAND Journal of Economics*, 25, 242–262.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile prices in market equilibrium," *Econometrica*, 63, 841–890.

Bertrand, M., D. Karlan, S. Mullainathan, E. Shafir, and J. Zinman (2010): "What's advertising content worth? Evidence from a consumer credit marketing field experiment," *Quarterly Journal of Economics*, 125, 263–306.

Bhattacharya, D. (2015): "Nonparametric welfare analysis for discrete choice," *Econometrica*, 83, 617–649.

Block, H. D. and J. Marschak (1960): *Random orderings and stochastic theories of responses*, Boston, USA: D. Reidel Publishing Company, vol. 1 of *Economic Information, Decision, and Prediction*, chap. 7, 172–217.

Boatwright, P. and J. C. Nunes (2001): "Reducing assortment: An attribute-based approach," *Journal of Marketing*, 65, 50–63.

Brenkers, R. and F. Verboven (2006): "Liberalizing a distribution system: The European car market," *Journal of the European Economic Association*, 4, 216–251.

Bresnahan, T. F., S. Stern, and M. Trajtenberg (1997): "Market segmentation and the sources of rents from innovation: Personal computers in the late 1980s," *RAND Journal of Economics*, 28, S17–S44.

Broniarczyk, S. M., W. D. Hoyer, and L. Mcalister (1998): "Consumers' perceptions of the assortment offered in a grocery cagegory: The impact of item reduction," *Journal of Marketing Research*, 35, 166–176.

Burdett, K. and K. L. Judd (1983): "Equilibrium price dispersion," *Econometrica*, 51, 955–969.

Caplin, A., M. Dean, and J. Leahy (Forthcoming): "Rational inattention, optimal consideration sets and stochastic choice," *Review of Economic Studies*.

Carlson, J. A. and R. P. McAfee (1983): "Discrete equilibrium price dispersion," *Journal of Political Economy*, 91, 480–493.

Chade, H. and L. Smith (2006): "Simultaneous search," *Econometrica*, 74, 1293–1307.

Ching, A. T., T. Erdem, and M. P. Keane (2009): "The price consideration model of brand choice," *Journal of Applied Econometrics*, 24, 393–420.

——— (2014): "A simple method to estimate the roles of learning, inventories, and category consideration in consumer choice," *The Journal of Choice Modelling*, 13, 60–72.

——— (Forthcoming): "How much do consumers know about the quality of products? Evidence from the diaper market," *Japanese Economic Review*.

Chiong, K. X., A. Galichon, and M. Shum (2016): "Duality in dynamic discrete-choice models," *Quantitative Economics*, 7, 83–115.

Chu, C. (1989): *A paired combinatorial logit model for travel demand analysis*, Western Periodicals, vol. 4 of *Transport Policy, Management and Technology Towards 2001:Selected Proceedings of the Fifth World Conference on Transport Research*, 295–309.

Dagsvik, J. K. and A. Karlström (2005): "Compensating variation and Hicksian choice probabilities in random utility models that are nonlinear in income," *Review of Economic Studies*, 72, 57–76.

Davis, P. and P. Schiraldi (2014): "The flexible coefficient multinomial logit (FC-MNL) model of demand for differentiated products," *RAND Journal of Economics*, 45, 32–63.

de Palma, A., G. M. Myers, and Y. Y. Papageorgiou (1994): "Rational choice under an imperfect ability to choose," *American Economic Review*, 84, 419–440.

Diamond, P. A. (1971): "A model of price adjustment," *Journal of Economic Theory*, 3, 156–168.

Diehl, K. and C. Poynor (2010): "Great expectations?! Assortment size, expectations, and satisfication," *Journal of Marketing Research*, 47, 312–322.

Drèze, X., S. Hoch, and M. E. Purk (1994): "Shelf management and space elasticity," *Journal of Retailing*, 70, 301–326.

Fosgerau, M., E. Melo, A. d Palma, and M. Shum (2019): "Discrete choice and rational inattention: a general equivalence result," *Working Paper*.

Goldfarb, A., Q. Lu, and S. Moorthy (2009): "Measuring brand value in an equilibrium framework," *Marketing Science*, 28, 69–86.

Gudagni, P. M. and J. D. C. Little (1983): "A Logit model of brand choice calibrated on scanner data," *Marketing Science*, 2, 203–238.

Herniter, J. D. (1973): "An entropy model of brand purchase behavior," *Journal of Marketing Research*, 10, 361–375.

Herriges, J. A. and C. L. Kling (1999): "Nonlinear income effects in random utility models," *Review of Economics and Statistics*, 81, 62–72.

Hicks, J. R. (1939): *Value and Capital*, Oxford, 2nd ed.

Honka, E. (2014): "Quantifying search and switching costs in the US auto insurance industry," *RAND Journal of Economics*, 45, 847–884.

Honka, E. and P. K. Chintagunta (2017): "Simultaneous or sequential? Search strategies in the U.S. auto insurance industry," *Marketing Science*, 36, 21–42.

Honka, E., A. Hortaçsu, and M. A. Vitorino (Forthcoming): "Advertising, consumer awareness, and choice: Evidence from the U.S. banking industry," *RAND Journal of Economics*.

Hortaçsu, A. and C. Syverson (2004): "Product differentiation, search costs, and competition in the mutual fund industry: A case study of S&P 500 index funds," *Quarterly Journal of Economics*, 119, 403–456.

Hoyer, W. D. and S. P. Brown (1990): "Effects of brand awareness on choice for a common, repeat-purchase product," *Journal of Consumer Research*, 19, 141–148.

Iyengar, S. S., G. Huberman, and W. Jiang (2004): *How much choice is too much? Contributions to 401(k) retirement plans*, Oxford: Oxford University Press, 83–96, Pension Design and Structure: New Lessons from Behavioral Finance.

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

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

Koppelman, F. S. and C.-H. Wen (2000): "The paired combinatorial logit model: properties, estimation and application," *Transportation Research Part B*, 34, 75–89.

Matějka, F. and A. McKay (2015): "Rational inattention to discrete choices: A new foundation for the multinomial logit model," *American Economic Review*, 105, 272–298.

McFadden, D. (1974): *Conditional logit analysis of qualitative choice behavior*, Academic Press, chap. 4, 105–142.

——— (1978): *Modelling the choice of residential location*, North-Holland Publishing Company, vol. 3 of *Spatial interaction theory and planning models*, chap. 3, 75–96.

——— (1981): *Econometric models of probabilistic choice*, The MIT Press, chap. 5, 198–272, Structural analysis of discrete data with econometric applications.

McFadden, D. and M. K. Richter (1991): *Stochastic rationality and revealed stochastic preference*, Westview Press, 161–186.

Mehta, N., S. Rajiv, and K. Srinivasan (2003): "Price uncertainty and consumer search: A structural model of consideration set formulation," *Marketing Science*, 22, 58–84.

Monroe, K. B. and A. Y. Lee (1999): "Remembering versus knowing: Issues in buyers' processing of price information," *Journal of the Academy of Marketing Science*, 27, 207–225.

Morozov, I. (2019): "Measuring benefits from new products in markets with information frictions," *Working Paper*.

Murthi, B. and R. C. Rao (2012): "Price awareness and consumers' use of deals in brand choice," *Journal of Retailing*, 88, 34–46.

Nevo, A. (2001): "Measuring market power in the ready-to-eat cereal industry," *Econometrica*, 69, 307–342.

Palazzolo, M. and F. Feinberg (2015): "Modeling consideration set substitution," *Working Paper*.

Petrin, A. (2002): "Quantifying the benefits of new products: The case of the minivan," *Journal of Political Economy*, 110, 705–729.

Reinganum, J. F. (1979): "A simple model of equilibrium price dispersion," *Journal of Political Economy*, 87, 851–858.

Shannon, C. E. (1948): "A mathematical theory of communication," *The Bell System Technical Journal*, 27, 379–423, 623–656.

Shugan, S. M. (1980): "The cost of thinking," *Journal of Consumer Research*, 7, 99–111.

Small, K. A. and H. S. Rosen (1981): "Applied welfare economics with discrete choice models," *Econometrica*, 49, 105–130.

Stahl II, D. O. (1989): "Oligopolistic pricing with sequential consumer search," *American Economic Review*, 79, 700–712.

Stigler, J. J. (1961): "The economics of information," *Journal of Political Economy*, 69, 213–225.

Terui, N., M. Ban, and G. M. Allenby (2011): "The effect of media advertising on brand consideration and choice," *Marketing Science*, 30, 74–91.

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

Ursu, R. M. (Forthcoming): "The power or rankings: Quantifying the effect of rankings on online consumer search and purchase decisions," *Marketing Science*.

Wedel, M. and R. Pieters (2000): "Eye fixations on advertisements and memory for brands: A model and findings," *Marketing Science*, 19, 297–312.

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

Appendix

A The Generalized Entropy

The generalized entropy function is proposed by Chiong et al. (2016), and has been investigated and used extensively in Fosgerau et al. (2019) to generalize Matějka and McKay (2015)'s results to any discrete-choice probability.

Define a vector-valued function $\mathbf{H}(\mathbf{u}_i)$, which is the exponentiated gradient of McFadden's social-surplus function $W_{RUM}(\mathbf{u}_i, \mathcal{J}_i)$[31] in the RUM context. Let

$$W_{RUM}(\mathbf{u}_i) := E\left[\max_{j \in \mathcal{J}_i} \{u_{i,j} + \epsilon_{i,j}\}\right],$$

where the idiosyncratic utility shock $\boldsymbol{\epsilon}_i$ follows a full-support distribution with density. The vector-valued function is defined by $\mathbf{H}(\mathbf{u}_i) := (H_1(\mathbf{u}_i), ..., H_{J_i}(\mathbf{u}_i))' = \nabla(\exp W_{RUM}(\mathbf{u}_i))'$. Note the Williams-Daly-Zachary theorem yields

$$\Pr(i \text{ chooses } j|\mathbf{u}_i) = \frac{H_j(\mathbf{u}_i)}{\sum_{k \in \mathcal{J}_i} H_k(\mathbf{u}_i)}$$

in McFadden's RUM framework.[32]

The functional form of the generalized entropy information-cost function is derived from the exponentiated social-surplus function as follows. Fosgerau et al. (2019) show $\mathbf{H} : \mathbb{R}^{J_i} \to \mathbb{R}_+^{J_i}$ is globally invertible in its domain and homogeneous of degree 1, and thus its inverse $\mathbf{H}^{-1}$ is also homogeneous of degree 1.

Now let $W^*(\mathbf{y})$ denote the convex conjugate of $W_{RUM}(\mathbf{u})$, which solves

$$W_{RUM}(\mathbf{u}) = \max_{\mathbf{y} \in \bar{\mathcal{A}}} \left\{\mathbf{y}'\mathbf{u} - W^*(\mathbf{y})\right\},$$

where $\bar{\mathcal{A}}$ is the unit simplex in $\mathbb{R}^{J_i}$. Fosgerau et al. show

$$W^*(\mathbf{y}) = \begin{cases} \mathbf{y}' \ln \mathbf{H}^{-1}(\mathbf{y}) & \text{if } \mathbf{y} \in \bar{\mathcal{A}} \\ \infty & \text{otherwise.} \end{cases}$$

Define the negative convex conjugate $-W^*(\mathbf{y})$ as the generalized entropy, which is concave in its domain. When $\mathbf{H}$ is an identity function, the generalized entropy reduces to the well-known Shannon entropy.

[31] For simplicity in notation, $W_{RUM}(\mathbf{u}_i, \mathcal{J}_i)$ is abbreviated as $W_{RUM}(\mathbf{u}_i)$ in the remainder of this section.

[32] Some well-known $\mathbf{H}(\mathbf{u}_i)$ functions are listed in Appendix G.

B Proof of Theorem 3.1

Proof. The proof is by construction. Fix individual i and drop index i for the sake of notational convenience. Drop alternative 0 without losing generality, and let $\mathcal{J} = \{1, 2, ..., J\}$. Normalize $u_J = 0$ as the utility of the outside option. Let $\tilde{\mathbf{u}} := (u_1, u_2, ..., u_{J-1}, 0)$.[33]

For each $k \in \mathcal{J}$, let $\pi_k > 0$ be given. Let $\mu > 0$ be given. Define a mapping $D : \mathbb{R}^{J-1} \to \mathbb{R}^J$, where the jth element of the mapping is

$$\{D(\tilde{\mathbf{u}})\}_j = \frac{H_j\left(\exp\left(\ln\left(\mathbf{H}^{-1}(\boldsymbol{\pi})\right) + \mu\tilde{\mathbf{u}}\right)\right)}{\sum_{k\in\mathcal{J}} H_k\left(\exp\left(\ln\left(\mathbf{H}^{-1}(\boldsymbol{\pi})\right) + \mu\tilde{\mathbf{u}}\right)\right)} =: y_j. \tag{B.1}$$

Because the mapping $\mathbf{H} : \mathbb{R}^J \to \mathbb{R}^J_+$ is a bijection, it is globally invertible[34]; therefore, the mapping $D : \mathbb{R}^{J-1} \to \mathbb{R}^J$ defined in (B.1) is globally invertible with the range

$$\mathcal{A} := \left\{\mathbf{y} \in \mathbb{R}^J : 0 < y_1, y_2, ..., y_J < 1 \text{ and } \sum_{k=1}^{J} y_k = 1\right\}.$$

$\mathcal{A}$ is the interior of the unit simplex in $\mathbb{R}^J$, which is a $J-1$ dimensional object. Note $y_J = 1 - \sum_{k=1}^{J-1} y_k$ because of the normalization constraint, and $\pi_J = 1 - \sum_{k=1}^{J-1} \pi_k$ by construction because $\{\pi_k\}_{k\in\mathcal{J}}$ is a set of unconditional choice probabilities.

Take any $\chi > 0$ and define $\pi_k^\chi := \chi\pi_k$. Consider the density q^D of the Dirichlet distribution over $\mathcal{A}$ with concentration parameters $\left(\pi_1^\chi, \pi_2^\chi, ..., \pi_J^\chi\right)$, given by

$$q^D\left(y_1, y_2, ..., 1 - \sum_{k=1}^{J-1} y_k\right) = \frac{1}{B\left(\pi_1^\chi, \pi_2^\chi, ..., \pi_J^\chi\right)} \prod_{k=1}^{J} y_k^{\pi_k^\chi - 1},$$

where

$$B\left(\pi_1^\chi, \pi_2^\chi, ..., \pi_J^\chi\right) := \frac{\prod_{k=1}^{J} \Gamma\left(\pi_k^\chi\right)}{\Gamma\left(\sum_{k=1}^{J} \pi_k^\chi\right)}$$

and $\Gamma(\cdot)$ is the Gamma function. Let $\tilde{\mathbf{y}} := (y_1, y_2, ..., y_{J-1})$. It is known for Dirichlet distributions that, for $1, 2, ..., J-1$,

$$\int_{\mathcal{A}} y_j q^D(\tilde{\mathbf{y}})\, d\tilde{\mathbf{y}} = \frac{\pi_j^\chi}{\sum_{k\in\mathcal{J}} \pi_k^\chi}.$$

[33] This normalization is innocuous because H_j is homogeneous of degree 1.
[34] See Proposition 2 of Fosgerau et al. (2019) for the proof of global invertibility of $\mathbf{H}$.

Therefore,

$$\int_{\mathcal{A}} y_j q^D(\tilde{\mathbf{y}})\, d\tilde{\mathbf{y}} = \frac{\pi_j}{\sum_{k\in\mathcal{J}} \pi_k} = \pi_j \tag{B.2}$$

$$\int_{\mathcal{A}} q^D(\tilde{\mathbf{y}})\, d\tilde{\mathbf{y}} = 1. \tag{B.3}$$

Applying the change of variables $y_j = \{D(\tilde{\mathbf{u}})\}_j$, respectively, on (B.2) and (B.3) yields

$$\pi_j = \int_{\mathbb{R}^{J-1}} \{D(\tilde{\mathbf{u}})\}_j \left|\det(\nabla D(\tilde{\mathbf{u}}))\right| q^D(D(\tilde{\mathbf{u}}))\, d\tilde{\mathbf{u}} \quad \forall j \in \mathcal{J}$$

$$1 = \int_{\mathbb{R}^{J-1}} \left|\det(\nabla D(\tilde{\mathbf{u}}))\right| q^D(D(\tilde{\mathbf{u}}))\, d\tilde{\mathbf{u}}.$$

[35] Define

$$q(\tilde{\mathbf{u}}) := \left|\det(\nabla D(\tilde{\mathbf{u}}))\right| q^D(D(\tilde{\mathbf{u}})) \tag{B.4}$$

as the density of the distribution $Q(\cdot)$. Because $D : \mathbb{R}^{J-1} \to \mathcal{A}$ is globally invertible and q^D has full support over $\mathcal{A}$, q has a full support over $\mathbb{R}^{J-1}$, which is a sufficient condition for (ii). This concludes the proof. □

Theorem 3.1 (i) is the necessary condition given in Corollary 2 of Matějka and McKay (2015) as well as in Proposition 6 (i) of Fosgerau et al. (2019), and Theorem 3.1 (ii) is the condition that requires the constructed $Q_i(\cdot)$ is indeed coherent with $\pi_{i,j} > 0$ for all $j \in \mathcal{J}_i$. I show the existence by constructing a probability density function $q_i : \mathbb{R}^{J_i} \to (0, 1)$ corresponding to Q_i, of which the support is $\mathbb{R}^{J_i}$. Although the class of density functions I construct here is a mixture resulting from the Dirichlet distribution, other possibilities may exist. The necessary conditions for q^D over $\mathcal{A}$ are (B.3) holds and $E[y_j] = \pi_j \; \forall j \in \mathcal{J}$. The restrictions that the latter condition impose are only on the first moments. Hence, other possibilities could construct such probability density q^D, and, by using (B.4), any density q^D that satisfies these two conditions can be converted to the density $q(\tilde{\mathbf{u}})$ that we are looking for. q^D does not have to be a Dirichlet density. Finding the sufficient condition for q^D exceeds the scope of this research.

C Further Details and Possibilities for the Consumer-Welfare Evaluation

Section 3.3 provides the RI-based consumer-welfare formulas, and section 4.4 demonstrates how the RI-based consumer-welfare-evaluation framework can be applied to the case study of Tide Pods' introduction. In this section, I provide further possibilities and discussions on the consumer welfare

[35] This is a slight abuse of notation as we are fixing $u_J = 0$.

evaluation. Specifically, two major issues are discussed in this section: (i) stability of demand parameter estimates, and (ii) equilibrating responses from the supplier side.

(i): In section 4.4, only one set of parameters $(-\beta_1, \boldsymbol{\beta}_2, \boldsymbol{\gamma}, \boldsymbol{\theta}, \sigma)$ are estimated by pooling the data for pre- and post- Tide Pods introduction. This can be justified well when the span of the sample is relatively short, and thus, consumers' taste and information costs are relatively stable throughout the sample period. However, when it is more appropriate to assume that consumers' utility parameters are varying over time, the model parameters for pre-change $\left(-\beta_1^0, \boldsymbol{\beta}_2^0, \boldsymbol{\gamma}^0, \boldsymbol{\theta}^0, \sigma^0\right)$ and post-change $\left(-\beta_1^1, \boldsymbol{\beta}_2^1, \boldsymbol{\gamma}^1, \boldsymbol{\theta}^1, \sigma^1\right)$ can be estimated separately. Then, the proposed formula for the compensating variation can be adjusted accordingly, as follows:

$$CV_{RI}^i := \frac{1}{\beta_1^1} W_{RI}^i \left(\mathbf{u}_i^1, \mathcal{J}_i^1\right) - \frac{1}{\beta_1^0} W_{RI}^i \left(\mathbf{u}_i^0, \mathcal{J}_i^0\right).$$

[36] During an application, consistent estimates of the model parameters can be plugged in to yield the consistent prediction.

(ii): Equilibrium responses in prices and promotions are ignored in the Tide Pods entry case studied, mainly because the market shares of Tide Pods were small enough. However, in many instances, considering the equilibrium responses of the competitors on prices and promotional variables may be more appropriate when conducting consumer-welfare counterfactuals, which should depend highly on the circumstances of the individual market or industry being studied. To do so, explicitly modeling the supplier side is inevitable. In particular, specifying the modes of competition, such as Bertrand/Cournot/Stackelberg/differentiated-products monopoly, would be crucial in determining the counterfactual pricing. A classic example would be the static Bertrand-Nash assumption in Berry et al. (1995); Nevo (2001); Petrin (2002). For other possibly endogenous marketing variables including promotions, modeling the equilibrium responses of all those endogenous marketing variables can be sometimes infeasible or implausible. A researcher can conduct robustness checks in such cases.

[36] If heterogeneity in model parameters exists across individuals (i.e., random coefficients), it can be accommodated in the way described in section 3.3.

Appendix for Online Publication

D More on the Model, Likelihood, Estimation, and Counterfactuals

D.1 Implementation Details of the Baseline Specification

Maximum Simulated Likelihood The maximum-likelihood-estimation problem that I consider is

$$\max_{(\boldsymbol{\gamma},\boldsymbol{\beta},\boldsymbol{\theta},\zeta)} \ln\left(\prod_{i,j} E_{\boldsymbol{\eta}_i}\left[\Pr_i\left(i \text{ Chooses } j\right)\right]^{\omega_i \mathbf{1}(i \text{ Chooses } j)}\right), \tag{D.1}$$

where

$$\begin{aligned} E\left[\Pr_i\left(i \text{ Chooses } j\right)\right] &\approx \frac{1}{n_s}\sum_{s=1}^{n_s}\Pr_i\left(i \text{ Chooses } j;\boldsymbol{\eta}_{i,s}\right) \\ &:= \frac{1}{n_s}\sum_{s=1}^{n_s}\frac{\exp\left(\mathbf{d}'_{i,j}\boldsymbol{\gamma}+\mu_i\left[\alpha+p_j\beta_1+\mathbf{x}'_j\boldsymbol{\beta}_2+\eta_{i,j,s}\varsigma\right]\right)}{1+\sum_{j'\in\mathcal{J}_i\setminus 0}\exp\left(\mathbf{d}'_{i,j'}\boldsymbol{\gamma}+\mu_i\left[\alpha+p_{j'}\beta_1+\mathbf{x}'_{j'}\boldsymbol{\beta}_2+\eta_{i,j',s}\varsigma\right]\right)}. \end{aligned}$$

$n_s = 100$ is used to simulate the integrals.

I supply the exact derivatives of the objective function (D.1) to the nonlinear optimizer. The log-likelihood is

$$\begin{aligned} &\mathbf{1}\left(i \text{ Chooses } j\right)\omega_i \ln E\left[\Pr_i\left(i \text{ Chooses } j\right)\right] \\ =&\mathbf{1}\left(i \text{ Chooses } j\right)\omega_i \ln\left(\frac{1}{n_s}\sum_{s=1}^{n_s}\frac{\exp\left(\mathbf{d}'_{i,j}\boldsymbol{\gamma}+\mu_i\left[\alpha+p_j\beta_1+\mathbf{x}'_j\boldsymbol{\beta}_2+\eta_{i,j,s}\varsigma\right]\right)}{1+\sum_{j'\in\mathcal{J}_i\setminus 0}\exp\left(\mathbf{d}'_{i,j'}\boldsymbol{\gamma}+\mu_i\left[\alpha+p_{j'}\beta_1+\mathbf{x}'_{j'}\boldsymbol{\beta}_2+\eta_{i,j',s}\varsigma\right]\right)}\right), \end{aligned}$$

where s denotes the index for the simulated $\boldsymbol{\eta}_i$ normal random variable.

The corresponding gradients are given as follows:

$$\frac{\partial \ln \Pr\left(i \text{ Chooses } j\right)}{\partial \gamma^{(p)}} = \frac{1}{\sum_{s=1}^{n_s}\Pr\left(i \text{ Chooses } j;\boldsymbol{\eta}_{i,s}\right)}\sum_{s=1}^{n_s}\left\{d_{i,j}^{(p)} - \frac{\sum_{j'\in\mathcal{J}_i\setminus 0} d_{i,j'}^{(p)}\exp\left(\mathbf{d}'_{i,j}\boldsymbol{\gamma}+\mu_i u_{i,j',s}\right)}{1+\sum_{j'\in\mathcal{J}_i\setminus 0}\exp\left(\mathbf{d}'_{i,j'}\boldsymbol{\gamma}+\mu_i u_{i,j',s}\right)}\right\}\Pr\left(i \text{ Chooses } j;\boldsymbol{\eta}_{i,s}\right)$$

$$\frac{\partial \ln \Pr\left(i \text{ Chooses } j\right)}{\partial \beta^{(p)}} = \frac{\mu_i}{\sum_{s=1}^{n_s}\Pr\left(i \text{ Chooses } j;\boldsymbol{\eta}_{i,s}\right)}\sum_{s=1}^{n_s}\left\{x_{i,j}^{(p)} - \frac{\sum_{j'\in\mathcal{J}_i\setminus 0} x_{i,j'}^{(p)}\exp\left(\mathbf{d}'_{i,j}\boldsymbol{\gamma}+\mu_i u_{i,j',s}\right)}{1+\sum_{j'\in\mathcal{J}_i\setminus 0}\exp\left(\mathbf{d}'_{i,j}\boldsymbol{\gamma}+\mu_i u_{i,j',s}\right)}\right\}\Pr\left(i \text{ Chooses } j;\boldsymbol{\eta}_{i,s}\right)$$

$$\frac{\partial \ln \Pr\left(i \text{ Chooses } j\right)}{\partial \theta^{(p)}} = \frac{w_i^{(p)}\mu_i}{\sum_{s=1}^{n_s}\Pr\left(i \text{ Chooses } j;\boldsymbol{\eta}_{i,s}\right)}\sum_{s=1}^{n_s}\left\{u_{i,j,s} - \frac{\sum_{j'\in\mathcal{J}_i\setminus 0} u_{i,j',s}\exp\left(\mathbf{d}'_{i,j}\boldsymbol{\gamma}+\mu_i u_{i,j',s}\right)}{1+\sum_{j'\in\mathcal{J}_i\setminus 0}\exp\left(\mathbf{d}'_{i,j}\boldsymbol{\gamma}+\mu_i u_{i,j',s}\right)}\right\}\Pr\left(i \text{ Chooses } j;\boldsymbol{\eta}_{i,s}\right)$$

$$\frac{\partial \ln \Pr\left(i \text{ Chooses } j\right)}{\partial \zeta^{(p)}} = \frac{\mu_i}{\sum_{s=1}^{n_s}\Pr\left(i \text{ Chooses } j;\boldsymbol{\eta}_{i,s}\right)}\sum_{s=1}^{n_s}\left\{\eta_{i,j,s} - \frac{\sum_{j'\in\mathcal{J}_i\setminus 0} \eta_{i,j',s}\exp\left(\mathbf{d}'_{i,j}\boldsymbol{\gamma}+\mu_i u_{i,j',s}\right)}{1+\sum_{j'\in\mathcal{J}_i\setminus 0}\exp\left(\mathbf{d}'_{i,j'}\boldsymbol{\gamma}+\mu_i u_{i,j',s}\right)}\right\}\Pr\left(i \text{ Chooses } j;\boldsymbol{\eta}_{i,s}\right).$$

Inference The asymptotic covariance matrix is also slightly different from the usual maximum-likelihood estimator. Sampling weights must be considered when calculating the asymptotic covariance matrix. Let ∇ denote the gradient operator, and let

$$\nabla_{i,j} := \nabla \left[\omega_i E \left[\Pr_i \left(i \text{ Chooses } j \right) \right] \right]$$

denote the score function for each sample evaluated at the optimum. The asymptotic covariance matrix formula for the estimator is

$$\mathbf{V} = \mathbf{\Omega}^{-1} \mathbf{\Delta} \mathbf{\Omega}^{-1}, \tag{D.2}$$

where

$$\mathbf{\Omega} := -\sum_{i,j} \nabla \left(\nabla'_{i,j} \right)$$
$$\mathbf{\Delta} := \sum_{i,j} \left(\nabla'_{i,j} \nabla_{i,j} \right).$$

I use (D.2) in the following for the inference on model-parameter estimates.

D.2 Model-Parameter Estimates for the RUM Specification

Table 7 presents the estimation results from the RUM specification, without including any consideration shifters or parametrizing the information-cost function. As usual in estimating any discrete-choice RUM model, the dispersion parameter is normalized as 1. The utility-parameter estimates and the implied price elasticities are similar to the preferred specification presented in Table 3.

Table 7: Model-Parameter Estimates, Consideration Shifters Not Included (RUM Specification)

Mean Price Elasticity		–2.405	
Utility Parameter $\left(-\hat{\beta}_1, \hat{\boldsymbol{\beta}}_2\right)$			
Per-pack Price $\left(-\hat{\beta}_1\right)$	–0.272***	75oz. ≤ Pack ≤150oz.	0.519***
	(0.003)		(0.058)
Pack ≤75oz.	–0.061	150oz. ≤ Pack ≤225oz.	0.917***
	(0.053)		(0.062)
All	0.926***	Tide	1.417***
	(0.019)		(0.018)
Arm&Hammer	0.799***	Wisk	1.063***
	(0.018)		(0.025)
Gain	0.686***	Xtra	0.534***
	(0.023)		(0.023)
Purex	0.672***		
	(0.017)		
Powder	–0.171***	High Efficiency	0.147***
	(0.016)		(0.010)
Fabric Softener	–0.282***	Oxi-Clean	0.107***
	(0.053)	/ Baking Soda	(0.021)
Febreze	–0.108***	Colorsafe	0.341***
	(0.022)		(0.021)
All Temperature	0.058**	Soft	–0.024
	(0.028)		(0.051)
Bleach	–0.162***	Stain Remover	0.260***
	(0.018)	/ Deep Clean	(0.019)
Ultra	0.124***	Unscented /	–0.010
	(0.010)	Sensitive / Baby	(0.012)
$n\times$Concentrated	0.105***	Low Cl / S / P	0.050***
	0.006)		(0.017)
sd($\eta_{i,j}$) ($\hat{\sigma}$)	0.103*	Pod / Tablet / Sheet	–0.359***
	(0.062)		(0.026)
Choice Obs.	170698	Sample Size	17M
AIC	1447938	BIC	1448209

*p<0.1; **p<0.05; ***p<0.01

Note. This table summarizes the model-parameter-estimation results from the maximum-likelihood estimation. The data used for the estimation are the matched sample of laundry detergent purchases in the Nielsen-Kilts consumer panel data and scanner data for households during 2006-2016. Standard-error estimates are in parentheses. AIC and BIC are calculated after adjusting the panel-projection factor weights sum up to the number of choice observations.

D.3 Further Results on the Counterfactual Welfare Simulation Associated with Tide Pods' Introduction

Table 8 reports further results for the counterfactual welfare simulation. Columns (1) and (2) coincide with those of Table 5 in section 4.4. Columns (3) and (4) are benchmarks, calculated CV^*_{RI} and CV^*_{RUM}, respectively, for the "hybrid" models in which the information shifters are included in the alternative-specific consumption utilities. Although including the information shifters in the alternative-specific consumption-utility terms may lack the microfoundation, a meaningful exercise could be to examine whether the drastic difference in CV_{RI} and CV_{RUM} associated with Tide Pods' entry reported in Table 5 should be attributed to the welfare formula itself or to not including the information shifters in the conditional choice probability and alternative-specific utility term. Specifically, CV^*_{RI} and CV^*_{RUM} are calculated by using the following formula:

$$
\begin{aligned}
CV^*_{RI} &= E_{\boldsymbol{\eta}_i}\left[\frac{1}{\hat{\beta}_1\hat{\mu}_i}\left\{W^*_{RI}\left(\hat{\tilde{\mathbf{u}}}^1_i, \mathcal{J}^1_i\right) - W^*_{RI}\left(\hat{\tilde{\mathbf{u}}}^0_i, \mathcal{J}^0_i\right)\right\}\right] \\
CV^*_{RUM} &= E_{\boldsymbol{\eta}_i}\left[\frac{1}{\hat{\beta}_1\hat{\mu}_i}\left\{W^*_{RUM}\left(\hat{\tilde{\mathbf{u}}}^1_i, \mathcal{J}^1_i\right) - W^*_{RUM}\left(\hat{\tilde{u}}_i, \mathcal{J}^0_i\right)\right\}\right] \\
W^*_{RUM}\left(\hat{\tilde{\mathbf{u}}}_i, \mathcal{J}_i\right) &= \ln\left\{\sum_{j\in\mathcal{J}_i}\exp\left(\hat{\tilde{\mathbf{u}}}_i\right)\right\} \\
W^*_{RI}\left(\hat{\tilde{\mathbf{u}}}_i, \mathcal{J}_i\right) &= \sum_{j\in\mathcal{J}_i}\hat{\Pr}_i\left(i \text{ Chooses } j|\hat{\mathbf{u}}_i\right)\hat{\tilde{u}}_{i,j} \\
&\quad - \hat{\mu}_i^{-1}\left\{\sum_{j\in\mathcal{J}_i}\hat{\Pr}_i\left(i \text{ Chooses } j|\hat{\mathbf{u}}_i\right)\left\{\ln\hat{\Pr}_i\left(i \text{ Chooses } j|\hat{\mathbf{u}}_i\right) - \ln\hat{\pi}_{i,j}\right\}\right\},
\end{aligned}
$$

where

$$
\begin{aligned}
\hat{\tilde{u}}_{i,j} &:= \mathbf{d}'_{i,j}\hat{\boldsymbol{\gamma}} + \hat{\mu}_i\left(\hat{\beta}_0 - p_j\hat{\beta}_1 + \mathbf{x}'_j\hat{\boldsymbol{\beta}}_2 + \eta_{i,j}\right) \qquad \text{for } j \neq 0 \\
\hat{\tilde{u}}_{i,0} &:= 0.
\end{aligned} \tag{D.3}
$$

In Table 8, CV^*_{RUM} is still positive and unrealistically large in magnitude, whereas CV^*_{RI} is slightly negative and of a much more reasonable magnitude. It reconfirms the assertion that the RUM-based CV formula overestimates the consumer gains from the new,increased size of the choice set even after including all the information shifters in the alternative-specific utility term $\tilde{u}_{i,j}$. The $E\max$ form of the RUM-based formula systematically overestimates the gains from the increased size of the choice set, by attributing the source of the stochastic component during discrete choice to unobserved preference shock.

Table 8: Average and Annually Projected Compensating Variation per Shopping Trip Associated with Tide Pods' introduction

	(1) CV_{RI}	(2) CV_{RUM}	(3) CV^*_{RI}	(4) CV^*_{RUM}
Average per Shopping Trip	–\$0.019	\$0.157	–\$0.011	\$0.198
Annually Projected	–\$8,028,950	\$72,538,564	–\$1,178,488	\$88,566,797

Note. The *Annually Projected* row is calculated using the subsample of the estimation data during 2012-2016, and then projecting to the entire U.S. population using the panel-projection factors.

E Robustness Checks

In this section, I report the model-parameter estimates and counterfactual welfare-simulation results for two different empirical specifications as a robustness check.

The first specification is identical to the main specification in section 4, except that it does not include the random effects $\eta_{i,j}$ in the alternative-specific utility specification. The alternative-specific utility is

$$\begin{aligned} u_{i,j} &:= \beta_0 - p_j\beta_1 + \mathbf{x}_j'\boldsymbol{\beta}_2 \qquad \text{for } j \neq 0 \\ u_{i,0} &:= 0. \end{aligned} \tag{E.1}$$

The expression for the conditional choice probability for the first specification without random effects is as follows:

$$\Pr_i\left(i \text{ Chooses } j|\mathbf{u}_i\right) = \frac{\exp\left(\mathbf{d}_{i,j}'\boldsymbol{\gamma} + \mu_i\left[\beta_0 - p_j\beta_1 + \mathbf{x}_j'\boldsymbol{\beta}_2\right]\right)}{1 + \sum_{j'\in\mathcal{J}_i\setminus 0}\exp\left(\mathbf{d}_{i,j'}'\boldsymbol{\gamma} + \mu_i\left[\beta_0 - p_{j'}\beta_1 + \mathbf{x}_{j'}'\boldsymbol{\beta}_2\right]\right)}.$$

In the second specification, the inverse of the unit information cost μ_i is normalized to 1. Instead, the demographics are interacted with the prices, resulting in the following conditional choice probability expression as:

$$\Pr_i\left(i \text{ Chooses } j|\mathbf{u}_i\right) = \frac{\exp\left(\mathbf{d}_{i,j}'\boldsymbol{\gamma} + \left[\beta_0 - \tilde{\mathbf{p}}_j'\tilde{\boldsymbol{\beta}}_1 + \mathbf{x}_j'\boldsymbol{\beta}_2\right]\right)}{1 + \sum_{j'\in\mathcal{J}_i\setminus 0}\exp\left(\mathbf{d}_{i,j'}'\boldsymbol{\gamma} + \left[\beta_0 - \tilde{\mathbf{p}}_{j'}'\tilde{\boldsymbol{\beta}}_1 + \mathbf{x}_{j'}'\boldsymbol{\beta}_2\right]\right)},$$

where $\tilde{\mathbf{p}}_j$ is the price interacted with all the demographic variables $\mathbf{w}_i$.

Utility-parameter estimates from the first specification turn out to be similar to those of the specification's random-effects counterpart, namely, the main empirical specification in section 4. Utility-parameter estimates from the second specification are considerably different from those of the main empirical specification. However, the implications from counterfactual welfare analyses are surprisingly similar in the main empirical specification and two robustness-check specifications reported in the current section; that is, CV_{RI} is slightly negative, whereas CV_{RUM} is positive and unrealistically large. Sections E.1 and E.2 report the results from the first and second specifications, respectively.

E.1 Results from Specification 1 – Specification without Random Effects: Model-Parameter Estimates, Model-Fit Measures, and Counterfactual Simulation Results

Table 9: Model-Parameter Estimates, Consideration Shifters Included (RI Specification)

Mean Price Elasticity		–2.126			
Utility Parameter $\left(-\hat{\beta}_1, \hat{\boldsymbol{\beta}}_2\right)$				Consideration Shifter Parameter $\hat{\boldsymbol{\gamma}}$	
Per-pack Price $\left(-\hat{\beta}_1\right)$	–0.129***	75oz. ≤ Pack ≤150oz.	–0.195***	In-store Display	0.552***
	(0.008)		(0.036)		(0.017)
Pack ≤75oz.	–0.499***	150oz. ≤ Pack ≤225oz.	0.038	Feature Ad	0.752***
	(0.024)		(0.050)		(0.014)
All	0.352***	Tide	0.562***	Same UPC	3.362***
	(0.026)		(0.038)	Purchased within 1 Year	(0.012)
Arm&Hammer	0.319***	Wisk	0.422***	Same Product (Different UPC)	1.831***
	(0.024)		(0.031)	Purchased within 1 Year	(0.019)
Gain	0.306***	Xtra	0.193***		
	(0.024)		(0.020)	Unit Information Cost Parameter $\hat{\boldsymbol{\theta}}$	
Purex	0.240***			Visited the Same Store	0.223***
	(0.019)			Within 1 Year	(0.026)
Powder	0.056***	High Efficiency	0.094***	Income Ratio to FPL	–0.054***
	(0.010)		(0.008)		(0.006)
Fabric Softener	–0.157***	Oxi-Clean	0.072***	Apartment	0.077**
	(0.031)	/ Baking Soda	(0.013)		(0.031)
Febreze	0.004	Colorsafe	0.154***	Non-working Spouse	0.177***
	(0.012)		(0.016)		(0.025)
All Temperature	0.061***	Soft	0.067**	Household Size	0.068***
	(0.016)		(0.028)		(0.011)
Bleach	–0.078***	Stain Remover	0.131***	Head Employment	0.254***
	(0.011)	/ Deep Clean	(0.014)		(0.032)
Ultra	0.033***	Unscented /	0.011	Head College Degree	–0.051**
	(0.006)	Sensitive / Baby	(0.007)		(0.021)
$n\times$Concentrated	0.044***	Low Cl / S / P	0.024**	Married,	0.009
	(0.005)		(0.010)	Living Together	(0.025)
		Pod / Tablet / Sheet	–0.058***	No Child	0.296***
			(0.015)		(0.032)
Choice Obs.	170698	Sample Size	17M	$\bar{\mu} = \overline{\exp\left(\mathbf{w}_i'\hat{\boldsymbol{\theta}}\right)}$	1.868
AIC	1171086	BIC	1171478	Log-likelihood	–585504

*p<0.1; **p<0.05; ***p<0.01

Note. This table summarizes the model-parameter-estimation results from the maximum-likelihood estimation. Mean elasticity and $\bar{\mu} = \overline{\exp\left(\mathbf{w}_i'\hat{\boldsymbol{\theta}}\right)}$ are calculated after weighting for the panel-projection factor. Due to the utility specification of the model, μ_i should be multiplied by the utility-parameter coefficients $\left(-\hat{\beta}_1, \hat{\boldsymbol{\beta}}_2\right)$ to find the effective magnitudes. Comparing the magnitudes of $\left(-\hat{\beta}_1, \hat{\boldsymbol{\beta}}_2\right)$ with the promotion-parameter $\hat{\boldsymbol{\gamma}}$ estimates is only sensible after multiplying μ_i's by $\left(-\hat{\beta}_1, \hat{\boldsymbol{\beta}}_2\right)$. The data used for the estimation are the matched sample of laundry detergent purchases in the Nielsen-Kilts consumer panel data and scanner data for

Table 10: Model-Parameter Estimates, Consideration Shifters Not Included (RUM Specification)

Mean Price Elasticity		–2.572	
Utility Parameter $\left(-\hat{\beta}_1, \hat{\boldsymbol{\beta}}_2\right)$			
Per-pack Price $\left(-\hat{\beta}_1\right)$	–0.291***	75oz. ≤ Pack ≤150oz.	–1.065***
	(0.003)		(0.050)
Pack ≤75oz.	–1.718***	150oz. ≤ Pack ≤225oz.	–0.627***
	(0.052)		(0.052)
All	0.918***	Tide	1.445***
	(0.018)		(0.018)
Arm&Hammer	0.769***	Wisk	1.074***
	(0.017)		(0.025)
Gain	0.704***	Xtra	0.406***
	(0.023)		(0.023)
Purex	0.638***		
	(0.016)		
Powder	–0.204***	High Efficiency	0.137***
	(0.016)		(0.010)
Fabric Softener	–0.268***	Oxi-Clean	0.106***
	(0.053)	/ Baking Soda	(0.021)
Febreze	–0.108***	Colorsafe	0.357***
	(0.022)		(0.021)
All Temperature	0.057**	Soft	−0.045
	(0.029)		(0.051)
Bleach	–0.179***	Stain Remover	0.245***
	(0.018)	/ Deep Clean	(0.019)
Ultra	0.104***	Unscented /	–0.022*
	(0.010)	Sensitive / Baby	(0.012)
n×Concentrated	0.093***	Low Cl / S / P	0.035**
	(0.006)		(0.017)
		Pod / Tablet / Sheet	–0.369***
			(0.026)
Choice Obs.	170698	Sample Size	17M
AIC	1442265	BIC	1442276

*p<0.1; **p<0.05; ***p<0.01

Note. This table summarizes the model-parameter-estimation results from the maximum-likelihood estimation.. The data used for the estimation are the matched sample of laundry detergent purchases in the Nielsen-Kilts consumer panel data and scanner data for households during 2006-2016. Standard-error estimates are in parentheses. AIC and BIC are calculated after adjusting the panel-projection factor weights sum up to the number of choice observations.

Table 11: Comparison of Model Fit across Different Specifications on the Consideration Shifters

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log-likelihood	–551281	–552195	–561255	–562324	–563028	–563959	–671664	–673313
AIC	1102565	1104392	1122511	1124650	1126058	1127921	1343329	1346628
BIC	1102575	1104402	1122521	1124660	1126068	1127931	1343339	1346638
Hit Rate	0.264	0.263	0.263	0.261	0.243	0.242	0.113	0.111
Average Hit Probability	0.102	0.101	0.098	0.097	0.087	0.086	0.024	0.024
Display / Feature	O	O	O	O	O	O	O	O
Own UPC Purchase History	O	O	O	O	X	X	X	X
Own Product Purchase History	O	O	X	X	O	O	X	X
Demographics as Unit Information Cost	O	X	O	X	O	X	O	X
Demographics Interacted with Utility	X	X	X	X	X	X	X	X

	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Log-likelihood	–557638	–558641	–567576	–568737	–569730	–570755	–679599	–681322
AIC	1115279	1117284	1135153	1137476	1139461	1141513	1359200	1362647
BIC	1115289	1117294	1135163	1137486	1139471	1141523	1359210	1362657
Hit Rate	0.263	0.264	0.266	0.265	0.247	0.246	0.119	0.117
Average Hit Probability	0.099	0.098	0.095	0.093	0.083	0.082	0.021	0.021
Display / Feature	X	X	X	X	X	X	X	X
Own UPC Purchase History	O	O	O	O	X	X	X	X
Own Product Purchase History	O	O	X	X	O	O	X	X
Demographics as Unit Information Cost	O	X	O	X	O	X	O	X
Demographics Interacted with Utility	X	X	X	X	X	X	X	X

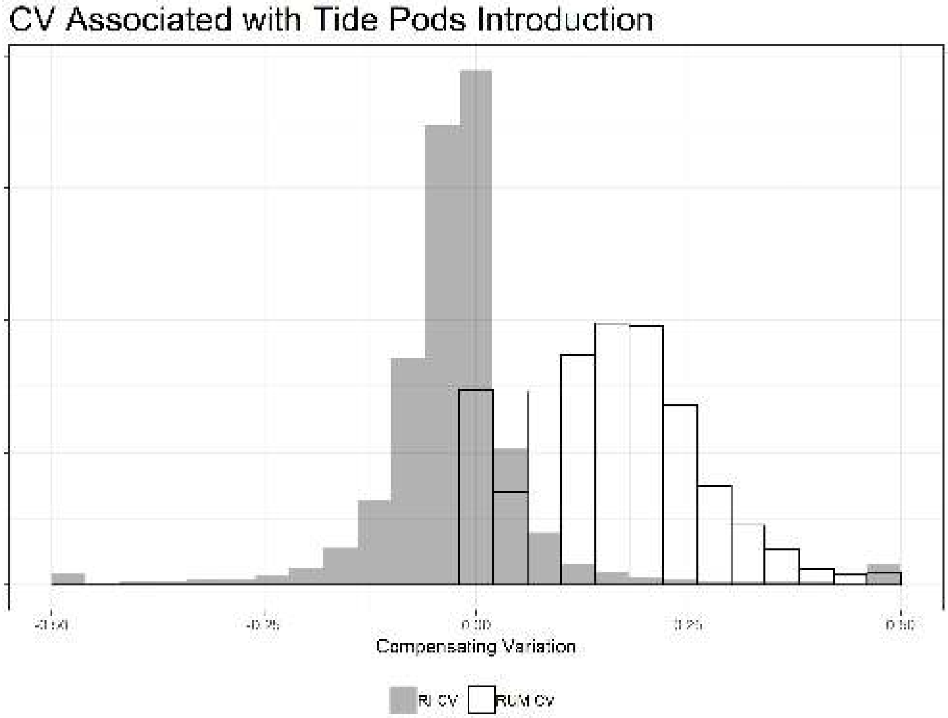
Note. AIC, BIC, and log–likelihood are calculated after scaling the panel-projection factor weights so that they sum up to the number of choice observations. The estimation sample has 160,698 choices with a total of 15,989,188 alternatives. The holdout sample has 10,000 choices with a total of 994,378 alternatives.

Table 12: Average and Annually Projected Compensating Variation per Shopping Trip Associated with Tide Pods' Introduction

	(1) CV_{RI}	(2) CV_{RUM}	(3) CV^*_{RI}	(4) CV^*_{RUM}
Average per Shopping Trip	–\$0.022	\$0.157	–\$0.016	\$0.197
Annually Projected	–\$9,083,832	\$68,992,703	–\$3,349,424	\$88,482,962

Note. The *Annually Projected* row is calculated using the subsample of the estimation data during 2012-2016, and then projecting to the entire U.S. population using the panel-projection factors.

Figure E.1: Compensating Variation Associated with Tide Pods' Entry



Note. The histogram illustrates the simulated compensating variation (CV) associated with the introduction of Tide Pods to the market. RI CV change is calculated using the formula (3.11), and RUM CV change is calculated using the formula (3.12). The model parameter used to simulate the change in CV is from Table 3. The data used for CV simulation are the 2012-2016 subsample of the estimation data, which is after the introduction of Tide Pods in 2012. The frequency is weighted for the panel-projection factor.

E.2 Results from Specification 2 – Interaction of Price Coefficients With Demographics: Model-Parameter Estimates, Model-Fit Measures, and Counterfactual Simulation Results

Table 13: Model-Parameter Estimates, Consideration Shifters Included (RI Specification)

Mean Price Elasticity	–2.378				
Utility Parameter				Consideration-Shifter Parameter	
$p_j \times$const	–0.269***	75oz. $\leq$ Pack $\leq$150oz.	0.044	In-store Display	0.555***
	(0.009)		(0.052)		(0.017)
Pack $\leq$75oz.	–0.497***	150oz. $\leq$ Pack $\leq$225oz.	0.474***	Feature Ad	0.751***
	(0.049)		(0.056)		(0.014)
All	0.680***	Tide	1.085***	Same UPC	3.357***
	(0.020)		(0.019)	Purchased within 1 Year	(0.012)
Arm&Hammer	0.635***	Wisk	0.812***	Same Product (Different UPC)	1.843***
	(0.018)		(0.026)	Purchased Within 1 Year	(0.019)
Gain	0.587***	Xtra	0.386***		
	(0.024)		(0.024)	Demographics Interacted with Price	
Purex	0.479***			$p_j \times$Visited the Same Store	–0.021***
	(0.018)			Within 1 Year	(0.003)
Powder	0.117***	High Efficiency	0.186***	$p_j \times$Income Ratio to FPL	0.012***
	(0.016)		(0.010)		(0.001)
Fabric Softener	–0.323***	Oxi-Clean	0.139***	$p_j \times$Apartment	–0.018***
	(0.054)	/ Baking Soda	(0.022)		(0.005)
Febreze	–0.010	Colorsafe	0.306***	$p_j \times$Non-working Spouse	–0.013***
	(0.023)		(0.022)		(0.004)
All Temperature	0.128***	Soft	0.134***	$p_j \times$Household Size	0.006***
	(0.028)		(0.051)		(0.002)
Bleach	–0.146***	Stain Remover	0.245***	$p_j \times$Head Employment	–0.020***
	(0.018)	/ Deep Clean	(0.020)		(0.005)
Ultra	0.061***	Unscented /	0.037***	$p_j \times$Head College Degree	0.020***
	(0.011)	Sensitive / Baby	(0.012)		(0.003)
$n\times$Concentrated	0.094***	Low Cl / S / P	0.052***	$p_j \times$Married,	–0.003
	(0.007)		(0.018)	Living Together	(0.004)
		Pod / Tablet / Sheet	–0.116***	$p_j \times$No Child	0.000
			(0.026)		(0.005)
Choice Obs.	170698	Sample Size	17M	$\bar{\mu} = \overline{\exp\left(\mathbf{w}_i'\hat{\boldsymbol{\theta}}\right)}$	1
AIC	1170752	BIC	1171144	Log-likelihood	–585337

*p<0.1; **p<0.05; ***p<0.01

Note. This table summarizes the model-parameter-estimation results from the maximum-likelihood estimation. The data used for the estimation are the matched sample of laundry detergent purchases in the Nielsen-Kilts consumer panel data and scanner data for households during 2006-2016. Standard-error estimates are in parentheses. AIC, BIC, and log-likelihood are calculated after adjusting the panel-projection factor weights sum up to the number of choice observations.

Table 14: Model-Parameter Estimates, Consideration Shifters Not Included (RUM Specification)

Mean Price Elasticity		–2.586			
Utility Parameter				Demographics Interacted with Price	
p_j×const	–0.293***	75oz. ≤ Pack ≤150oz.	0.584***	p_j×Visited the Same Store	–0.044***
	(0.009)		(0.057)	Within 1 Year	(0.004)
Pack ≤75oz.	–0.009	150oz. ≤ Pack ≤225oz.	0.973***	p_j×Income Ratio to FPL	0.014***
	(0.053)		(0.062)		(0.001)
All	0.940***	Tide	1.435***	p_j×Apartment	–0.013**
	(0.019)		(0.018)		(0.005)
Arm&Hammer	0.814***	Wisk	1.082***	p_j×Non-working Spouse	–0.016***
	(0.018)		(0.025)		(0.004)
Gain	0.716***	Xtra	0.534***	p_j×Household Size	0.003**
	(0.023)		(0.023)		(0.002)
Purex	0.681***			p_j×Head Employment	–0.028***
	(0.017)				(0.005)
Powder	–0.170***	High Efficiency	0.144***	p_j×Head College Degree	0.034***
	(0.016)		(0.010)		(0.003)
Fabric Softener	–0.281***	Oxi-Clean	0.113***	p_j×Married,	0.001
	(0.053)	/ Baking Soda	(0.021)	Living Together	(0.004)
Febreze	–0.104***	Colorsafe	0.345***	p_j×No Child	–0.005
	(0.022)		(0.021)		(0.005)
All Temperature	0.054**	Soft	–0.029		
	(0.028)		(0.051)		
Bleach	–0.165***	Stain Remover	0.262***		
	(0.018)	/ Deep Clean	(0.019)		
Ultra	0.126***	Unscented /	–0.008		
	(0.010)	Sensitive / Baby	(0.012)		
n×Concentrated	0.107***	Low Cl / S / P	0.051***		
	(0.006)		(0.017)		
		Pod / Tablet / Sheet	–0.356***		
			(0.026)		
Choice Obs.	170698	Sample Size	17M	$\bar{\mu} = \overline{\exp\left(\mathbf{w}_i'\hat{\boldsymbol{\theta}}\right)}$	1
AIC	1443408	BIC	1443418	Log-likelihood	–721703

*p<0.1; **p<0.05; ***p<0.01

Note. This table summarizes the model-parameter-estimation results from the maximum-likelihood estimation. The data used for the estimation are the matched sample of laundry detergent purchases in the Nielsen-Kilts consumer panel data and scanner data for households during 2006-2016. Standard-error estimates are in parentheses. AIC and BIC are calculated after adjusting the panel-projection factor weights sum up to the number of choice observations.

Table 15: Comparison of Model Fit across Different Specifications on the Consideration Shifters

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log-likelihood	–551094	–561210	–562635	–671377	–557458	–567545	–569322	–679253
AIC	1102190	1122422	1125272	1342757	1114917	1135092	1138645	1358508
BIC	1102200	1122432	1125282	1342767	1114927	1135102	1138655	1358518
Hit Rate	0.261	0.260	0.242	0.111	0.263	0.264	0.246	0.117
Average Hit Probability	0.102	0.097	0.086	0.024	0.098	0.094	0.082	0.021
Display / Feature	O	O	O	O	X	X	X	X
Own UPC Purchase History	O	O	X	X	O	O	X	X
Own Product Purchase History	O	X	O	X	O	X	O	X
Demographics as Unit Information Cost	X	X	X	X	X	X	X	X
Demographics Interacted with Utility	O	O	O	O	O	O	O	O

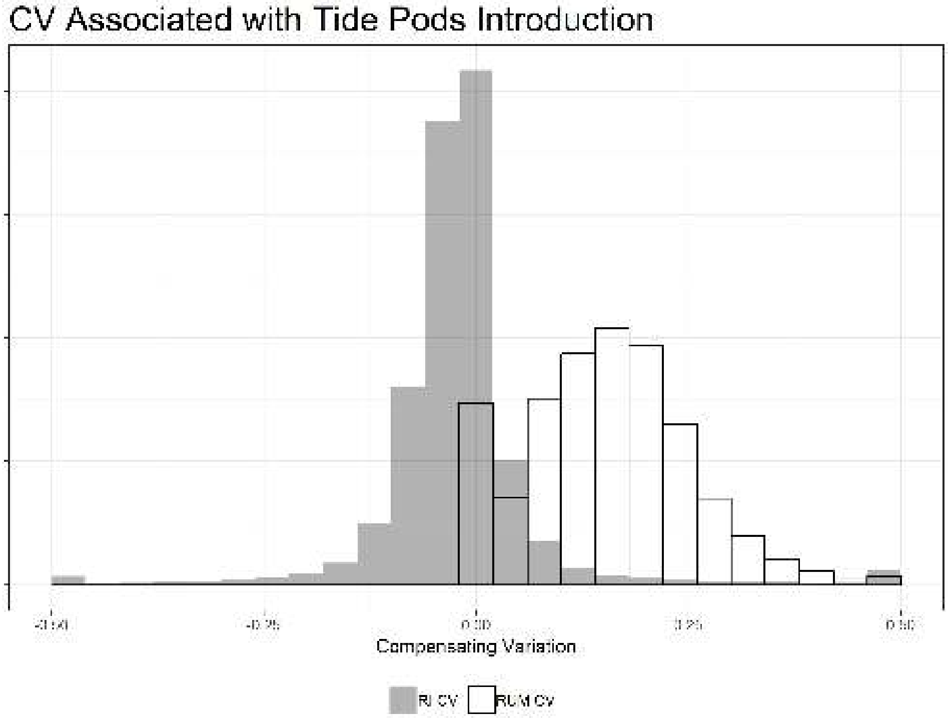
Note. AIC, BIC, and log-likelihood are calculated after scaling the panel-projection factor weights so that they sum up to the number of choice observations. The estimation sample has 160,698 choices with a total of 15,989,188 alternatives. The holdout sample has 10,000 choices with a total of 994,378 alternatives.

Table 16: Average and Annually Projected Compensating Variation per Shopping Trip Associated with Tide Pods' Introduction

	(1) CV_{RI}	(2) CV_{RUM}	(3) CV^*_{RI}	(4) CV^*_{RUM}
Average per Shopping Trip	–\$0.020	\$0.154	–\$0.030	\$0.194
Annually Projected	–\$7,932,798	\$67,098,056	–\$8,667,197	\$76,414,322

Note. The *Annually Projected* row is calculated using the subsample of the estimation data during 2012-2016, and then projecting to the entire U.S. population using the panel-projection factors.

Figure E.2: Compensating Variation Associated with Tide Pods' entry



Note. The histogram illustrates the simulated compensating variation (CV) associated with the introduction of Tide Pods to the market. RI CV change is calculated using the formula (3.11), and RUM CV change is calculated using the formula (3.12). The model parameter used to simulate the change in CV is from Table 3. The data used for CV simulation are the 2012-2016 subsample of the estimation data, which is after the introduction of Tide Pods in 2012. The frequency is weighted for the panel-projection factor.

F Data Appendix

The Nielsen-Kilts Homescan and RMS datasets are well known in the marketing and economics academic community, and the subscriptions are publicly available to academic researchers for a fee. I therefore relegate most of the details on the data to the data manuals provided by Nielsen-Kilts, and describe only some necessary details on the data-cleaning and processing procedure I have taken.

F.1 Matching Household Shopping-Trip Data with Scanner Data by Store Code

The choice set that the panel household faces is not recorded in the Homescan panel data. The only data source that I can obtain as the proxy for the choice set is the sales information of the scanner data. Therefore, I match shopping-trip data with the sales information of the scanner data, using the store-code and week variables. Because the store-code variable is missing for some shopping trips, I summarize how many shopping trips have store-code variables available in Table 17. I drop the samples with missing store codes.

Table 17: Fraction of Trips for Which Store Code Is Available

Year	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Matched Fraction	0.48	0.46	0.45	0.46	0.78	0.76	0.76	0.76	0.76	0.76	0.77
No. of Trips	178K	289K	275K	263K	248K	253K	233K	227K	219K	217K	233K

Note. This table presents the fraction of shopping trips for which the store code is available. Years 2004 and 2005 are omitted because the scanner data are available only for 2006-2016. The Matched Fraction row is the fraction of shopping trips for which the store code is available among the number of trips. The No. of Trips row is the number of shopping trips in which any laundry detergent is bought in the corresponding year, weighted by the household projection factor.

F.2 Product Attributes

Many product-attribute abbreviations in the raw data are not standard, and thus I manually coded the labels of 10,911 UPCs to identify unique products. I often searched and matched the UPCs with external databases to decode abbreviations.[37] I classify the functional product characteristics of laundry detergents as in Table 18. Each criterion listed in Table 18 has the same label in the same brand, but it might have different labeling across brands. Therefore, I manually classify the functional product characteristics of laundry detergents to make the labeling consistent.[38] In addition to differences in their formulas, laundry detergents have different scents. I separately code and classify the scents because consumers are likely to recognize the same laundry detergent formula

[37]Further details on the UPC handling and the external UPC databases used to identify the product characteristics can be found in Online Appendix F.3.

[38]For example, some brands refer to the oxygen-cleaning formula as "oxi-clean," whereas other brands refer to it as "active oxygen" or "oxifoam." The different names refer to essentially the same functional characteristic. Therefore, I classify "oxi-clean," "active oxygen," and "oxyfoam" as "oxi-clean." Similar classification occurs for most other product characteristics displayed in Table 18.

with different scents as different products. My data contain 262 scents across all brands. The product-attributes classification is sufficient to capture most product descriptions in the raw data.

Table 18: Summary of Observed Product Characteristics

Characteristics	Obs.	Characteristics	Obs.	Characteristics	Obs.
liquid	7849	high efficiency	2953	baking soda	46
powder	3066	oxi-clean	191	plant based	33
fabric softener	585	colorsafe	505	low sudsing	72
Febreze	206	soft	691	low Chlorine	4
all temperature	401	unscented	1153	low Sulfate	431
bleach	1695	sensitive skin	64	low Phosphorous	780
stain remover	159	baby	181	tablets	163
deep clean	128	pre-treater	2	sheet	31
ultra	4007	wrinkle reducer	4	refill	91
$n\times$concentrated	8963	enzyme	137		

Note. Observations are the counts of UPCs of the corresponding characteristics. The data contain 10,911 different UPCs.

F.3 Barcodes and External Barcode Databases

In this subsection, I describe how I processed the barcodes and matched them with the external databases. Nielsen uses 13-digit EAN-13 (International Article Number, previously referred to as European Article Number), which is a superset of UPC-A.[39] EAN-13 beginning with 0 coincides with UPC-A. For convenience, I refer to both EAN-13 and UPC-A as UPC unless otherwise noted.

Nielsen did not record the check digits that are required to search the external UPC database. Check digits are calculated as below. Let x_k be the kth digit of EAN-13. Then, x_{13}, the check digit, is defined to satisfy the following equation:

$$(x_1 + x_3 + ... + x_{11}) + 3\,(x_2 + x_4 + ... + x_{12}) + x_{13} = z \times 10,$$

for some nonnegative integer z.

The original product description of the Nielsen-Kilts data is highly abbreviated, and often no clue is available to help figure out which abbreviation means what sort of product characteristic. Therefore, I searched external UPC databases for the majority of the products to match with the product characteristics. I searched different UPC databases to identify the product characteristics, because a complete and comprehensive UPC database does not exist. Table 19 lists the databases that I use for search.

[39]Because neither database records check digits, the last 10 digits of the UPC and Nielsen EAN should coincide, provided they are from the same barcodes.

Table 19: UPC/EAN Databases

Database	Web Address
Amazon	www.amazon.com
Ebay	www.ebay.com
EANdata	www.eandata.com
UPC database	www.upcdatabase.org
UPC index	www.upcindex.com

G Some Well-Known GEV Class Social-Surplus Functions and Their Exponentiated Derivatives

Define $\mathcal{H}(\mathbf{u}) := \exp(W_{RUM}(\mathbf{u}))$ by the potential function of $\mathbf{H}(\mathbf{u})$, i.e., $H_j(\mathbf{u}) = \{\nabla\mathcal{H}(\mathbf{u})\}_j = \{\nabla \exp(W_{RUM}(\mathbf{u}))\}_j$, where $W(\cdot)$ is McFadden's social-surplus function.

G.1 Logit and Nested Logit

The $\mathbf{H}$ function for simple logit is

$$\mathcal{H}(\mathbf{u}) = \sum_{k\in\mathcal{J}} \exp(u_k)$$
$$H_j(\mathbf{u}) = \exp(u_j).$$

The $\mathbf{H}$ function for one-level nested logit is

$$\mathcal{H}(\mathbf{u}) = \sum_{K}\left(\sum_{k\in B_K}(\exp(u_k))^{\frac{1}{\lambda_K}}\right)^{\lambda_K}$$
$$H_j(\mathbf{u}) = \left(\sum_{k\in B_I}(\exp(u_k))^{\frac{1}{\lambda_J}}\right)^{\lambda_J-1}(\exp(u_j))^{\frac{1}{\lambda_J}-1},$$

where λ_K may differ by K. The potential function for the two-level nested-logit generating function is similarly defined as

$$\mathcal{H}(\mathbf{u}) = \sum_{\mathcal{K}}\left(\sum_{K\in\mathcal{K}}\left(\left(\sum_{k\in B_K}(\exp(u_k))^{\frac{1}{\lambda_K}}\right)^{\lambda_K}\right)^{\frac{1}{\lambda_{\mathcal{K}}}}\right)^{\lambda_{\mathcal{K}}}.$$

In these nested-logit models, the set of bins at each nesting level partitions $\mathcal{J}$. These models are developed by McFadden (1978); Brenkers and Verboven (2006).

G.2 Product-Differentiation Logit

The product-differentiation logit model generating function developed by Bresnahan et al. (1997) is

$$\mathcal{H}(\mathbf{u}) = \sum_g a_g \left(\sum_{K \in g} \left(\sum_{k \in B_g} (\exp(u_k))^{\frac{1}{\lambda_g}} \right)^{\lambda_g} \right) + (\exp(u_0))$$

$$H_j(\mathbf{u}) = a_g \left(\sum_{k \in B_g} (\exp(u_k))^{\frac{1}{\lambda_J}} \right)^{\lambda_J - 1} (\exp(u_j))^{\frac{1}{\lambda_J} - 1} \qquad \text{for } i \neq 0,$$

where for each g, $a_g \in (0,1]$ such that $\sum_g a_g = 1$, and $\lambda_g \in (0,1)$.

G.3 Paired Combinatorial Logit

The paired combinatorial logit-generating function by Chu (1989); Koppelman and Wen (2000) is

$$\mathcal{H}(\mathbf{r}) = \sum_{k=1}^{J-1} \sum_{l=k+1}^{J} \left((\exp(u_k))^{\frac{1}{\lambda_{kl}}} + (\exp(u_l))^{\frac{1}{\lambda_{kl}}} \right)^{\lambda_{kl}}$$

$$H_j(\mathbf{r}) = \sum_{k \neq j} (\exp(u_j))^{\frac{1}{\lambda_{kj}} - 1} \left((\exp(u_j))^{\frac{1}{\lambda_{jk}}} + (\exp(u_k))^{\frac{1}{\lambda_{jk}}} \right)^{\lambda_{jk} - 1},$$

where $\lambda_{jk} \in (0,1]$ for all distinct $k, l \in \mathcal{J}$.

G.4 Flexible-Coefficient Multinomial Logit

The flexible-coefficient multinomial logit-generating function by Davis and Schiraldi (2014) is

$$\mathcal{H}(\mathbf{r}) = \sum_{k \in \mathcal{J}} \sum_{k \neq j} a_{jk} \left(\frac{(\exp(u_j))^{\frac{1}{\lambda}} + (\exp(u_k))^{\frac{1}{\lambda}}}{2} \right)^{\tau\lambda} + \sum_{k \in \mathcal{J}} a_{kk,\mathcal{J}} (\exp(u_j))^{\tau}$$

$$H_j(\mathbf{r}) = \tau \sum_{k \neq j} b_{jk} \left(\frac{(\exp(u_j))^{\frac{1}{\sigma}} + (\exp(u_k))^{\frac{1}{\sigma}}}{2} \right)^{\tau\sigma - 1} \exp(u_j)^{\frac{1}{\sigma} - 1} + \tau b_{jj} (\exp(u_j))^{\tau - 1},$$

where $a_{jk} \geq 0$, $\tau > 0$, $\lambda > 0$, $\tau\lambda \leq 1$, and for each j, a good $l \in \mathcal{J}$ exists such that $b_{jl} > 0$.