Rational Inattention as an Empirical Framework:
Application to the Welfare Effects of New-Product Introduction[*,†]

Joonhwi Joo[‡]

September 17, 2019

Abstract

Consumer-welfare, as explicated by the conventional economic models of consumer-choice, increases whenever more alternatives are added to consumers' choice set. However, whether consumer-welfare is enhanced or reduced with more choice alternatives is an empirical question that must take into account not only the expanded choice set but also the incomplete information under which consumer choice occurs. To answer the question properly, this paper evaluates consumer welfare effects by developing a novel empirical framework of discrete-choice based on the Rational Inattention (RI) theory by Matějka and McKay (2015); Fosgerau et al. (2019). The resulting consumer-welfare-evaluation framework is general in that it allows consumer-welfare to either improve or reduce by adding alternatives to the choice set. This empirical framework is applied to a case of new-product introduction, namely, Tide Pods' entry into the market in 2012, to assess its effects on consumer-welfare. The empirical study finds that the average net welfare gain from Tide Pods' entry is negative because the pricing is too high relative to the benefit they provide to consumers. A counterfactual pricing experiment suggests that 7%-8% price cut of Tide Pods will make consumers just as well-off as prior to Tide Pods' entry.

[*] I am grateful to my advisors, Ali Hortaçsu, Jean-Pierre Dubé, Pradeep 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, 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 is known for its effort on better product curations of more preferred goods in a product category with low prices, leading to higher customer satisfaction, which is closely related to consumer-welfare.[1] The above example leads to the question: Are more alternatives always better for consumers? The answer to this empirical question should depend on whether the corresponding changes in the set of alternatives actually benefit consumers. Recent evidence from field experiments in the context of grocery shopping and financial product choice, including Iyengar and Lepper (2000); Iyengar et al. (2004); Bertrand et al. (2010), suggest the possibilities of more choice alternatives may harm consumers by inducing them to step away from choice. However, consumers must be better off whenever more alternatives are added to the choice set in the Walrasian demand framework and discrete-choice random utility maximization (RUM henceforth) framework developed by McFadden (1974, 1978, 1981); McFadden and Richter (1991); Small and Rosen (1981). Thus, a gap exists between what is empirically known or conjectured and what can be evaluated using existing empirical demand frameworks.

This paper evaluates the consumer-welfare gains of Tide Pods laundry detergents' introduction to the market in 2012. To allow the flexibility of consumer-welfare either being improved or reduced by the introduction of a new-product, I develop an empirical framework of consumer-choice and consumer-welfare-evaluation based on Matějka and McKay (2015); Fosgerau et al. (2019)'s rational inattention (RI henceforth) theory on discrete-choices. It turns out that the average consumer-welfare gains associated with Tide Pods' entry is negative. The counterfactual simulation on the pricing of Tide Pods suggests their prices were higher than the value consumers placed on them. Cutting the prices of Tide Pods by 7%-8% would make consumers just as well-off as prior to Tide Pods' introduction.

The flexibility of consumer-welfare gains being either positive or negative by adding more alternatives to consumers' choice set stems from the possibility that consumers may be inattentive to potentially superior alternatives in their choice set. Adding more "bad" alternatives to their choice set may increase the chance that those bad alternatives are chosen simply because consumers are inattentive to better alternatives in their choice set. This possibility is pertinent in almost every choice context where acquiring costly information about choice alternatives is necessary, no matter how small the amount of time and effort invested. Brand choice in a grocery shopping context is a good example. At each shopping instance, consumers need to spend time and effort distinguishing between and learning about the consumption utilities of each alternative on a shelf by, for example, reading price tags or product labels.

[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

Brand choice in grocery shopping can be modeled using Matějka and McKay (2015)'s RI-based discrete-choice framework in the following way. For each consumer's choice instance, an exogenously given finite choice set exists, from which the consumer intends to choose one alternative that will yield the highest consumption utility. The consumer may gather some "free" information, such as promotions or own purchase history, to form the prior belief distribution over the consumption utilities. Acquiring information about the true alternative-specific consumption utility is costly, where the information-gathering cost is proportional to the expected Shannon entropy reduction between the prior and posterior belief distribution over alternative-specific consumption utilities. The information is assumed to become more and more expensive as it becomes more precise. Thus, it is never optimal for a consumer to fully resolve the uncertainty over the alternative-specific utilities, implying imperfect information acquisition. After the costly information acquisition, the consumer updates the prior belief in a Bayesian fashion, yielding the posterior belief over consumption utilities. The choice based on the posterior belief is not deterministic because of the imperfect information acquisition. The choice probability after the costly information acquisition turns out to have an extended-logit form, and Fosgerau et al. (2019) extends Matějka and McKay's results to rationalize any discrete-choice probability as stemming from the RI model, by employing a generalized-entropy information-cost function.

In addition to the nonmonotonicity of consumer-welfare with respect to the number of alternatives, the RI-based empirical framework of discrete-choice proposed in this paper has two prominent features in rationalizing discrete-choices compared to the workhorse discrete-choice RUM framework. First, the RI-based empirical framework of discrete-choice inherently incorporates consumers' consideration process prior to the choice stage. The RI-based empirical framework of discrete-choice developed in this paper is similar to the consideration-set models (e.g., Allenby and Ginter 1995; Mehta et al. 2003) or the price-consideration models (Ching et al. 2009, 2014, Forthcoming) in that it formulates the choice problem as a two-stage process – consideration and choice. I follow the approaches and findings of Allenby and Ginter (1995); Mehta et al. (2003); Ching et al. (2009, 2014, Forthcoming); Terui et al. (2011); Murthi and Rao (2012) that informational choice-probability shifters such as promotions or purchase history should be excluded from alternative-specific consumption utility, whereas they should be included during consumers' consideration process. Separately incorporating such informational choice-probability shifters and consumption utilities in the choice probabilities is particularly pertinent in the context of consumer-welfare-evaluation because promotions or purchase history is not what is consumed; meanwhile, they still shift the choice probabilities. Second, the RI-based derivation of the discrete-choice probabilities provides a new interpretation of what corresponds to the dispersion parameter in RUM-based derivation: the unit information cost. I parameterize the unit information cost as a function of shopping environments and observed demographics and estimate the coefficients. The coefficient estimates turn out to be broadly consistent with the search-cost story.[2] Because the

[2]By contrast, in the discrete-choice RUM framework, the dispersion parameter is usually normalized as 1 because

RI-based discrete-choice model can incorporate these two features in the resulting choice probabilities, I compare in-sample and out-of-sample model fits to the workhorse RUM. I find the RI-based discrete-choice model fits the data substantially better than the RUM-based model, thus providing one metric of validation.

The empirical finding of the present paper that the average consumer-welfare gains from the introduction of Tide Pods were actually negative sheds new light on the implications of new-product introduction: it may not be always beneficial for consumers. More broadly, the flexible consumer-welfare-evaluation framework developed in the present paper would have direct managerial implications regarding customer satisfaction, especially when consumers' information acquisition about choice alternatives is costly, and thus curation of choice alternatives becomes a key managerial problem. Costco's product-curation strategy mentioned above may well be understood as deriving from the perspective that simply adding more alternatives will not lead to customer satisfaction. Another example would be e-commerce contexts where the cost of shelf space is not a concern of a manager. A manager may be tempted to add as many choice alternatives as possible because doing so would benefit both the consumers and the sellers under the conventional consumer-welfare-evaluation frameworks, which is not true when consumers' information acquisition is costly. Further, the RI-based consumer-welfare-evaluation framework developed in the present paper may be used to provide a rationale for what desirable curation of choice alternatives and/or what desirable category pricing scheme would lead to customer satisfaction.

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

The RI choice model that I appeal to in this paper, which is underpinned by the consumer's optimization over costly information acquisition, is based on recent theoretical works 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

interpreting that consumers with some demographics make more idiosyncratic choices than others is difficult.

Fosgerau et al. extended Matějka and McKay's result to any functional form of discrete-choice probability by employing a generalized entropy cost function. The RI-based empirical framework of discrete-choice developed in this paper can be widely applied to the choice contexts where at least *some* information acquisition, of which the specific information-acquisition process is unobservable to a researcher, has to be involved in order to distinguish different alternatives. The following example illustrates one such choice context that is considered throughout the present paper.

Example 2.1. A consumer wants to buy one laundry-detergent item. She 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 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 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 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 prior beliefs using the newly acquired information about 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 choice context described in the above example has two important features that make our RI-based discrete-choice model more applicable to such contexts than the workhorse 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 – alternative-specific consumption utility and informational choice-probability shifters such as promotions or own purchase history.

(i). The marketing literature has 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. The RI framework is based on the same intuition: consumers have to acquire information about alternative-specific consumption utilities, which is costly. In the RI framework, 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 RUM framework assumes consumers have complete knowledge about alternative-specific consumption utilities including the "idiosyncratic shock" component that is not observable to a researcher. Choices in the 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 of information friction and cost of information acquisition, 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) and consideration-set literature (e.g., Mehta et al. 2003; Honka 2014; Palazzolo and Feinberg 2015; Honka and Chintagunta 2017; Honka et al. Forthcoming).

(ii). 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 alternative-specific consumption utility. Several extant papers have addressed this issue 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 that sale signs are less effective when more items on the shelves have sale signs. This finding is consistent with the RI theory 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 purchase 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.

Note that the prior belief over each choice alternative is initialized at each choice incidence in Example 2.1. In other words, it is de facto assuming that learning over specific items' shelf-space locations or consumers' own reference price point on each alternative does not occur. We assume this feature, which follows the tradition in the marketing literature on brand choice dating back to at least Gudagni and Little (1983), mainly for the tractability of the framework because it allows our decision problem to be a static problem as opposed to a dynamic optimization problem. However, the experience of previous purchase may still affect the prior belief over consumption utilities in the empirical model developed in the following sections, which may reflect the presence of choice inertia.

2.2 Non-monotonicity of consumer-welfare in the Number of Choice Alternatives in the RI Framework

The discrete-choice RUM-based consumer-welfare-evaluation has been a standard consumer-welfare-evaluation apparatus since McFadden (1981); Small and Rosen (1981), and has been applied to various contexts including 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 among numerous others). The extant consumer-welfare-evaluation framework based on RUM exhibits the monotonicity in the number of alternatives. That is, *ceteris paribus*, adding one more alternative to consumers' choice set always increases the expected utility of choice, no matter how "bad" the added choice alternative is. However, the possibility that more choice alternatives may harm consumers 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 consumers to be disappointed more by the choice they end up making. This paper develops a general empirical framework to provide a systematic method to evaluate whether adding a choice alternative harms or benefits consumers.

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. Now, when a third, "bad" alternative is added to the choice set where $u_3 = -1$, the 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 occurs as choosing alternative 3 with $u_3 = -1$ is because the consumer somehow liked it better than the other two alternatives. This property may well be justified under the strong assumption that the consumers have a 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 have incomplete knowledge over the alternative-specific utility, the monotonicity of the consumer-welfare formula can be problematic; consumers may have

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

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 alternative, 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]

Matějka and McKay (2015) document the possibility of an "irregularity" in the discrete-choice RI model whereby adding an additional alternative to the choice set may increase the probability of existing alternatives being chosen, which cannot happen in the discrete-choice RUM model. The goal of the present paper is to go one step further: if the source of the stochasticity in discrete-choices is incomplete knowledge of the consumers over the alternative-specific utilities, the monotonicity in consumer-welfare in the number of alternatives breaks down. 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 consumers' objective function of choice, 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, 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 of Table 1 illustrates what happens when $u_3 = 1$ and $u_3 = -1$ are added to the choice set, respectively.[5] Notably, when a bad alternative is added to the choice set, the RI-based consumer welfare decreases to –0.237 from 0.

[4]This calculation is under the assumptions that consumers have no prior information over the alternatives and the Shannon entropy information-cost function is employed. We relax these assumptions in the subsequent sections.

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

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

Calculation for the Adding Bad case is similar.

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 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" and "good" alternative, respectively, to the choice set. The RI welfare is calculated using equation (3.9) 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 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) rationalizes the observed market shares by the maximum entropy principle, and Bass et al. (1976) shows 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 information cost are Shugan (1980); de Palma et al. (1994). Shugan; de Palma et al. formulate the limited information-processing ability of the decision maker, and then derive probabilistic choice rules from minimizing the cost of comparing alternatives (Shugan, 1980), or errors stemming from imperfect information processing (de Palma et al., 1994).

3 The Empirical Framework of discrete-choice Based on Rational Inattention

In this section, I develop an empirical framework on the choice of an alternative based on RI, where a consumer intends to choose one alternative that would maximize the alternative-specific utility. Then, I state and prove the necessary theorem that allows us to estimate the discrete-choice RI models as if estimating usual discrete-choice RUM models. Finally, I derive the associated social-welfare function based on RI and compare it with the social-welfare function derived from the

RUM framework.

Overview of the Choice Problem, Terminology, and Notation As an overview, I summarize the procedure on the choice of an alternative below.

- (Choice Stage 0 – Prior Formation) The consumer forms a prior belief over alternative-specific utilities based on
  - 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 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 utilities.
- (Choice Stage 2 – Choice of an Alternative) The consumer executes the information strategy, updates the prior belief in the Bayesian fashion to yield the posterior belief, and chooses an alternative according to the optimized outcome of the posterior belief.

The terms decision maker and consumer are used interchangeably hereafter. $\mathcal{J}_i = \{0, 1, ..., J_i\}$ denotes the set of alternatives available to consumer i. $\mathcal{J}_i$ contains the outside option, which is denoted by option 0. $\mathbf{u}_i := \{u_{i,j}\}_{j \in \mathcal{J}_i}$ is the vector of alternative-specific utility for consumer i. The choice structure of the consumer is as follows. The consumer, who remembers which items and products s/he purchased in the past, walks into the store. S/he immediately perceives the choice set $\mathcal{J}_i$ and the promotion status without incurring cost. Let $\mathbf{D}_i := \{\mathbf{d}_{i,j}\}_{j \in \mathcal{J}_i}$ denote the observed informational shifters of the prior belief on the alternative-specific consumption utility $\mathbf{u}_i$. $\mathbf{d}_{i,j}$ includes the dummies of the previous purchase history and current promotion status. This information, acquired without cost, results in a prior belief over $\mathbf{u}_i$, which I denote $Q_i(\mathbf{u}_i) := Q(\mathbf{u}_i | \mathcal{J}_i, \mathbf{D}_i)$.[6] Then, the consumer optimizes over the information strategy as described above, and engages in the costly information acquisition to learn about the consumption utilities. The optimization procedure yields the conditional choice probability denoted by $\Pr_i(i \text{ Chooses } j | \mathbf{u}_i)$ as the solution.

[6]Analogously, 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.

Finally, let us define the unconditional choice probability $\pi_{i,j}$ by the expected conditional choice probability,

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

which does not condition on the true consumption utility $\mathbf{u}_i$. Heuristically, $\pi_{i,j}$ may have an interpretation of the probability of consumer i choosing j if the costly information acquisition about the true consumption utility $\mathbf{u}_i$ does not occur. Table 2 summarizes the notations used throughout the paper.

Matějka and McKay (2015) show that 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 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 conditional choice probabilities $\left\{\Pr_i\left(i \text{ Chooses } j|\mathbf{u}_i\right)\right\}_{j\in\mathcal{J}_i}$.

3.1 Choice of an Alternative with Costly Information Acquisition

In this subsection, I formally develop a model on the choice of an alternative and prove the existence theorem of the prior belief distribution based on the setup of Matějka and McKay (2015); Fosgerau et al. (2019). In the following, I describe the choice step backwards from the Stage 2.

Choice Stage 2: Choice of an Alternative During the second stage, the consumer chooses one alternative out of $\mathcal{J}_i$, according to the conditional choice probability $\Pr_i\left(i \text{ Chooses } j|\mathbf{u}_i\right)$ after costly information acquisition.

Choice Stage 1: Optimization over Conditional Choice Probability Acquiring information on the true choice-specific consumption utility $u_{i,j}$ that each alternative j provides is costly. The RI framework does not specify consumers' specific process of information acquisition such as mode of search, abstracting from what and how much information consumers acquire. The main behavioral restriction it imposes is that the information-cost function is proportional to the expected generalized entropy difference between the unconditional (before information acquisition) and the conditional (after information acquisition) choice probabilities. The convexity of the information-cost function makes the information more expensive when the conditional choice probability is more "precise" in the sense that it accurately designates the alternative that yields the highest

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
$u_{i,j}$	Alternative-specific consumption utility of consumer i consuming j
$\mathbf{u}_i := \left\{u_{i,j}\right\}_{j\in\mathcal{J}_i}$	Vector of consumer i's alternative specific consumption utility
$\left(p_j, \mathbf{x}_j\right)$	Observed shifters of $u_{i,j}$, where p_j is price and $\mathbf{x}_j$ is product attributes
$\mathbf{d}_{i,j}$	Observed informational shifters of the prior belief on the alternative utility $u_{i,j}$
$\mathbf{D}_i := \left\{\mathbf{d}_{i,j}\right\}_{j\in\mathcal{J}_i}$	Observed informational shifters of the prior belief on the alternative utility $\mathbf{u}_i$
$Q_i(\mathbf{u}_i) := Q(\mathbf{u}_i \vert \mathcal{J}_i, \mathbf{D}_i)$	i's subjective prior belief distribution over the consumption utilities $\mathbf{u}_i$
$\Pr_i(i \text{ Chooses } j \vert \mathbf{u}_i)$	Conditional choice probability of i choosing j, conditioning on true $\mathbf{u}_i$
$\pi_{i,j} = E_{Q_i}\left[\Pr_i(i \text{ Chooses } j \vert \mathbf{u}_i)\right]$	Unconditional choice probability of i choosing j, not conditioning on true $\mathbf{u}_i$
$c_i(\cdot, \cdot)$	i's information cost function
μ_i	Reciprocal of i's unit information cost
$\mathbf{w}_i$	Observed shifters of the unit information cost
$W_{RUM}(\mathbf{u}_i)$	RUM-based Social Welfare Function
$W_{RI}(\mathbf{u}_i)$	RI-based Social Welfare Function
$\mathbf{H}(\mathbf{u}_i) := \nabla\left(\exp W_{RUM}(\mathbf{u}_i)\right)'$	Transpose of exponentiated gradient of $W_{RUM}(\mathbf{u}_i)$
$H_j(\mathbf{u}_i)$	j'th element of $\mathbf{H}(\mathbf{u}_i)$

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

utility to the consumer.

The consumer maximizes the expected gain from choice net of information cost:

$$\max_{\{\Pr_i(i \text{ Chooses } j|\mathbf{u}_i)\}_{j\in\mathcal{J}_i}} E_{Q_i}\left[\sum_{j\in\mathcal{J}_i} u_{i,j}\Pr_i(i \text{ Chooses } j|\mathbf{u}_i) - \text{Information cost}\right], \tag{3.1}$$

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

$$\max_{\{\Pr_i(i \text{ Chooses } j|\mathbf{u}_i)\}_{j\in\mathcal{J}_i}} \int_{\mathbf{u}_i}\left\{\sum_{j\in\mathcal{J}_i} u_{i,j}\Pr_i(i \text{ Chooses } j|\mathbf{u}_i)\right\} Q_i(d\mathbf{u}_i) - c_i\left(\boldsymbol{\pi}_i, \{\Pr_i(i \text{ Chooses } j|\mathbf{u}_i)\}_{j\in\mathcal{J}_i}\right), \tag{3.2}$$

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

Let μ_i be the reciprocal of individual i's unit information cost. Define a vector-valued function $\mathbf{H}(\mathbf{u}_i) := \left(H_1(\mathbf{u}_i), ..., H_{J_i}(\mathbf{u}_i)\right)'$, which is the exponentiated gradient of McFadden's social surplus function $W_{RUM}(\mathbf{u}_i)$ in the RUM context.[7] The information-cost function $c_i(\cdot,\cdot)$ in (3.2) has the following form:

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

The information-cost function is proportional to the expected difference in the generalized entropy of unconditional and conditional choice probabilities, which is defined in Appendix A. The shape of the solution to the optimization problem (3.2) will be inherited from the shape of the generalized entropy cost function. Note that if $\mathbf{H}(\mathbf{u}_i)$ is the identity function, the information-cost function reduces to the expected difference in Shannon entropy. The information-cost function is convex in $\{\Pr_i(i \text{ Chooses } j|\mathbf{u}_i)\}_{j\in\mathcal{J}_i}$,[8] which makes the information associated with more "precise" conditional choice probability more expensive.

Solution of the Problem Problem (3.2) can yield any fully supported additive RUM choice probability as its solution by employing the appropriate generalized entropy information-cost function introduced by Chiong et al. (2016); Fosgerau et al. (2019). The solution for the problem (3.2) yields the conditional choice probability after costly information acquisition as

$$
\Pr_i(i \text{ Chooses } j|\mathbf{u}_i) = \frac{H_j\left(\exp\left(\ln\left(\mathbf{H}^{-1}(\boldsymbol{\pi}_i)\right) + \mu_i\mathbf{u}_i\right)\right)}{\sum_{k\in\mathcal{J}} H_k\left(\exp\left(\ln\left(\mathbf{H}^{-1}(\boldsymbol{\pi}_i)\right) + \mu_i\mathbf{u}_i\right)\right)} \quad \text{a.s.} \tag{3.3}
$$

(3.3) will be taken as the building block for the likelihood of alternative-choice.

Existence Theorem of the Prior Belief Distribution For $\left\{\pi_{i,j}\right\}_{j\in\mathcal{J}_i}$ such that $\pi_{i,j} > 0$, the first-order conditions of (3.2) lead to the condition that the prior belief distribution $Q_i(\cdot)$ must satisfy the following fixed-point equation:

$$
\pi_{i,j} = \int_{\mathbf{u}_i} \frac{H_j\left(\exp\left(\ln\left(\mathbf{H}^{-1}(\boldsymbol{\pi}_i)\right) + \mu_i\mathbf{u}_i\right)\right)}{\sum_{k\in\mathcal{J}} H_k\left(\exp\left(\ln\left(\mathbf{H}^{-1}(\boldsymbol{\pi}_i)\right) + \mu_i\mathbf{u}_i\right)\right)} dQ_i(\mathbf{u}_i) \tag{3.4}
$$

(Proposition 6 of Fosgerau et al., Corollary 2 of Matějka and McKay). For an estimate of the unconditional choice probabilities $\left\{\pi_{i,j}\right\}_{j\in\mathcal{J}_i}$ to be justified as consistent with the RI theory, a prior belief

[7]

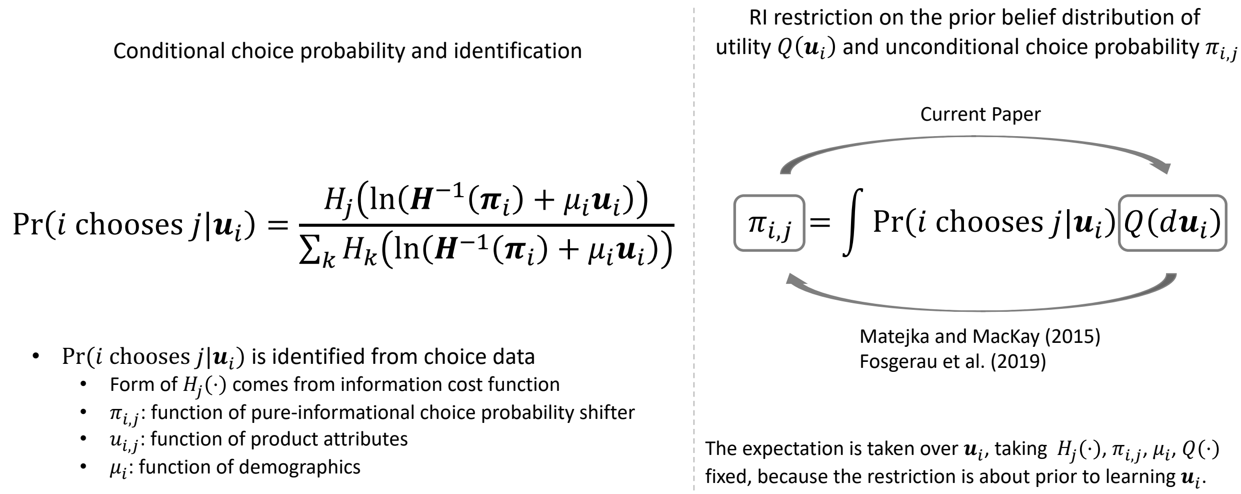
$$
W_{RUM}(\mathbf{u}_i) := E\left[\max_{j\in\mathcal{J}_i}\left\{u_{i,j} + \epsilon_{i,j}\right\}\right]
$$

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

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

distribution $Q_i(\cdot)$ over the utilities that yields $\{\pi_{i,j}\}_{j\in\mathcal{J}_i}$ must exist. In the RI theory literature, the 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 so as to satisfy the set of fixed-point equations (3.4). 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. Figure 3.1 summarizes the discussion in this subsection thus far.

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



The following theorem justifies that any combination of $\left\{\mu_i, \{\pi_{i,j}\}_{j\in\mathcal{J}_i}\right\}$ backed out from choice data can be rationalized as a rationally inattentive decision maker's choice. It asserts that for any given $\mu_i > 0$ and positive unconditional choice probabilities $\{\pi_{i,j}\}_{j\in\mathcal{J}_i}$, a prior belief distribution $Q_i(\cdot)$ exists that yields $\{\pi_{i,j}\}_{j\in\mathcal{J}_i}$.[9] Therefore, the unconditional choice probabilities backed out directly from the choice data can be justified as resulting from a rationally inattentive decision maker's choice.

Theorem 3.1. *(Existence of Prior Belief Distribution) Fix 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 $\sum_{k\in\mathcal{J}} \pi_{i,k} = 1$ be given. Let $\mu_i > 0$ also be given. Then, a probability measure Q_i exists over $\mathbb{R}^{J_i}$ such that for each $j\ (\in \mathcal{J}_i) \neq 0$, the following (i) and (ii) hold:*

[9] A restriction 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 theory literature of RI. 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.

(i) For each $j\ (\in \mathcal{J}_i) \neq 0$,

$$\pi_{i,j} = \int \frac{H_j\left(\exp\left(\ln\left(\mathbf{H}^{-1}\left(\boldsymbol{\pi}_i\right)\right) + \mu_i \mathbf{u}_i\right)\right)}{\sum_{k\in\mathcal{J}} H_k\left(\exp\left(\ln\left(\mathbf{H}^{-1}\left(\boldsymbol{\pi}_i\right)\right) + \mu_i \mathbf{u}_i\right)\right)} Q_i\left(d\mathbf{u}_i\right). \tag{3.5}$$

(ii) For each $j \in \mathcal{J}_i$, $Q_i\left(\left\{\mathbf{u}_i \in \mathbb{R}^{J_i} : u_{i,j} > u_{i,k}\ \forall k \neq j\right\}\right) > 0$.

Proof. See Appendix B. □

3.2 Parametrization of the Choice Probabilities for Empirical Application

Theorem 3.1 shows that the existence of the prior belief distribution consistent with any unconditional choice probabilities identified from the data. It implies that any discrete-choice probability can be rationalized as resulting from the choice of a rationally inattentive decision maker. In this subsection, I provide further possibilities of parameterizations on the conditional choice probability expression (3.3) that would facilitate the interpretations of the discrete-choice probabilities as resulting from the RI-choice problem.

Let us consider (3.3), which we take as the key building block for the likelihood:

$$\Pr_i\left(i \text{ Chooses } j|\mathbf{u}_i\right) = \frac{H_j\left(\exp\left(\ln\left(\mathbf{H}^{-1}\left(\boldsymbol{\pi}_i\right)\right) + \mu_i \mathbf{u}_i\right)\right)}{\sum_{k\in\mathcal{J}} H_k\left(\exp\left(\ln\left(\mathbf{H}^{-1}\left(\boldsymbol{\pi}_i\right)\right) + \mu_i \mathbf{u}_i\right)\right)}. \tag{3.6}$$

Without the term $\ln\left(\mathbf{H}^{-1}\left(\boldsymbol{\pi}_i\right)\right)$, 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 $\ln\left(\mathbf{H}^{-1}\left(\boldsymbol{\pi}_i\right)\right)$, 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:

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

[10] (3.6) simplifies to

$$\Pr_i\left(i \text{ Chooses } j|\mathbf{u}_i\right) = \frac{H_j\left(\tilde{\boldsymbol{\pi}}_i \circ \exp\left(\mu_i \mathbf{u}_i\right)\right)}{\sum_{k\in\mathcal{J}} H_k\left(\tilde{\boldsymbol{\pi}}_i \circ \exp\left(\mu_i \mathbf{u}_i\right)\right)}, \tag{3.7}$$

where $\circ$ denotes the Hadamard product.

One remarkable feature of (3.7) is that $\left\{\pi_{i,j}\right\}_{j\in\mathcal{J}_i}$ or $\left\{\tilde{\pi}_{i,j}\right\}_{j\in\mathcal{J}_i}$ can shift the choice probabilities without affecting the consumption utility $\left\{u_{i,j}\right\}_{j\in\mathcal{J}_i}$. 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' informa-

[10] $\mathbf{H}$ 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 $\mathbf{H}$ follows from the convexity of McFadden's social-surplus function defined in the RUM context, which guarantees the monotonicity of $H_j(\cdot)$ and $H_j^{-1}(\cdot)$.

tion. 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 section. Furthermore, the expression $\tilde{\boldsymbol{\pi}}_i$ in (3.7) 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-choice stage. Such a parametrization is well microfounded, in the meantime allowing 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 exactly as if estimating various RUM models with additional utility shifters $\mathbf{d}_{i,j}$.[11]

Example: Simple Logit A prominent example of the parametrization would be the simple multinomial logit. Suppose the information-cost function employed is the Shannon entropy; therefore, $\mathbf{H}$ is an identity function with $\tilde{\pi}_{i,j} = \pi_{i,j}$. Let $\mathbf{w}_i$ be shifters of unit information cost, let p_j be price, and let $\mathbf{x}_j$ be observed product characteristics. Let us parametrize $\pi_{i,j} \propto \exp\left(\mathbf{d}_{i,j}'\gamma\right)$, $\mu_i = \exp\left(\mathbf{w}_i'\boldsymbol{\theta}\right)$, and $u_{i,j} := \alpha - p_j\beta_1 + \mathbf{x}_j'\boldsymbol{\beta}_2$ with the normalization $u_{i,0} = 0$. Then, (3.7) reduces to

$$\begin{aligned} \Pr_i\left(i \text{ Chooses } j|\mathbf{u}_i\right) &= \frac{\exp\left(\mathbf{d}_{i,j}'\gamma + \mu_i u_{i,j}\right)}{1 + \sum_{j' \in \mathcal{J}_i} \exp\left(\mathbf{d}_{i,j'}'\gamma + \mu_i u_{i,j'}\right)} \\ &= \frac{\exp\left(\mathbf{d}_{i,j}'\gamma + \mu_i\left(\alpha - p_j\beta_1 + \mathbf{x}_j'\boldsymbol{\beta}_2\right)\right)}{1 + \sum_{j' \in \mathcal{J}_i} \exp\left(\mathbf{d}_{i,j'}'\gamma + \mu_i\left(\alpha - p_{j'}\beta_1 + \mathbf{x}_{j'}'\boldsymbol{\beta}_2\right)\right)}, \end{aligned} \tag{3.8}$$

which is the simple multinomial logit choice probability but more flexible in the sense that the consideration-shifter term $\mathbf{d}_{i,j}'\gamma$ and unit information cost μ_i are included.

Note in the choice probability (3.8) that μ_i captures the observed heterogeneity in the unit-information-acquisition cost of each household. 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 coefficients are $(-\mu_i\beta_1, \mu_i\boldsymbol{\beta}_2)$, 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 using the

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

information contained in the consideration shifters $\mathbf{d}_{i,j}$.[12]

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 at all; 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 smaller μ_i.

3.3 Welfare Analysis: Social-Surplus Function in Rational Inattention Framework

Section 3.1 shows 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 substantially different because the respective choice-probability expressions are based on starkly different microfoundations. In this subsection, I define the social-surplus functions within the RI framework, provide the corresponding compensating variation formula, and compare it with the RUM-based compensating variation formula.[13]

RI-based Social-Surplus Function Assume the exact value of $(\boldsymbol{\pi}_i, \mathbf{u}_i, \mu_i)$ is known to a researcher, whereas consumer i's subjective prior belief over the consumption utilities $Q_i(\mathbf{u}_i)$ is not. This situation is typical of what an empirical researcher would face where only $(\boldsymbol{\pi}_i, \mathbf{u}_i, \mu_i)$ is identified from choice data. Our goal is to evaluate the expected utility of choice net of information cost under this situation. A natural choice in the perspective of a researcher during the consumer-welfare predictions and counterfactuals would be to treat $Q_i(\mathbf{u}_i)$ as if it were a degenerate distribution with point mass at the realized value of $\mathbf{u}_i$, because the exact value of $\mathbf{u}_i$ is identified from data and already known to the researcher. Thus, the proposed social-surplus function $W_{RI}(\mathbf{u}_i)$ for

[12]Also note that $u_{i,j}$ can contain either the individual effects/coefficients, such that those individual effects/coefficients are not observable to a researcher but are observable to the consumer and affects the consumer's choice. The individual effects/coefficients can be treated as if they were fixed/random effects/coefficients during the estimation. We do not pursue this direction in the present paper because it either intricates the identification problem of the random effects/coefficients, or incurs a curse-or-dimensionality problem by having to introduce $\geq$10,000 indicator variables to treat them as fixed effects in a nonlinear model.

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

individual i with choice alternatives $\mathcal{J}_i$ is defined by the following:

$$\begin{aligned} W_{RI}\left(\mathbf{u}_i\right) &:= \sum_{j\in\mathcal{J}_i} \Pr_i\left(i \text{ Chooses } j|\mathbf{u}_i\right) u_{i,j} - \text{Information cost} \\ &= \sum_{j\in\mathcal{J}_i} \Pr_i\left(i \text{ Chooses } j|\mathbf{u}_i\right) u_{i,j} \\ &\quad - \mu_i^{-1}\left\{\sum_{j\in\mathcal{J}_i} \Pr_i\left(i \text{ Chooses } j|\mathbf{u}_i\right)\left\{\ln H_j^{-1}\left(\left\{\Pr_i\left(i \text{ Chooses } j|\mathbf{u}_i\right)\right\}_{j\in\mathcal{J}_i}\right) - \ln H_j^{-1}\left(\boldsymbol{\pi}_i\right)\right\}\right\}. \end{aligned} \tag{3.9}$$

The Gibbs' inequality ensures the information-cost term is always positive regardless of the realized $(\boldsymbol{\pi}_i, \mathbf{u}_i, \mu_i)$. In practice, when $(\boldsymbol{\pi}_i, \mathbf{u}_i, \mu_i)$ is estimated from data, a consistent prediction $(\hat{\boldsymbol{\pi}}_i, \hat{\mathbf{u}}_i, \hat{\mu}_i)$ can be used in place of $(\boldsymbol{\pi}_i, \mathbf{u}_i, \mu_i)$.

Two important remarks are in order. First, $W_{RI}\left(\mathbf{u}_i\right)$ 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 is due to the choice-specific utility $u_{i,j}$ not containing the idiosyncratic-utility-shock term in (3.9), as opposed to the social-welfare function $E\left[\max_{j\in\mathcal{J}_i}\left\{u_{i,j} + \epsilon_{i,j}\right\}\right]$ defined in the RUM context. Second, $W_{RI}\left(\mathbf{u}_i\right)$ separates the consumption utility $\left\{u_{i,j}\right\}_{j\in\mathcal{J}_i}$ from the informational choice-probability shifters inside $\Pr_i\left(i \text{ Chooses } j|\mathbf{u}_i\right)$,[14] thereby allowing for the informational choice-probability shifters such as promotions or previous purchase history to be contained only in $\Pr_i\left(i \text{ Chooses } j|\mathbf{u}_i\right)$, not in $\left\{u_{i,j}\right\}_{j\in\mathcal{J}_i}$. 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 subject to consumption.

Compensating Variation Calculation in RI and RUM Compensating variation (CV henceforth), 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. The concept of CV was first developed in Walrasian demand theory, and later McFadden (1978, 1981); Small and Rosen (1981) adopted the concept within the context of the discrete-choice RUM framework. Let superscript 0 denote before the changes in prices and/or choice-set composition, and let superscript 1 denote after the changes, respectively. CV_{RI} and CV_{RUM} are defined by

$$CV_{RI} := \frac{1}{\beta_1}\left\{W_{RI}\left(\mathbf{u}_i^1\right) - W_{RI}\left(\mathbf{u}_i^0\right)\right\} \tag{3.10}$$

$$CV_{RUM} := \frac{1}{\beta_1}\left\{E\left[\max_{j\in\mathcal{J}_i^1}\left\{u_{i,j}^1 + \epsilon_{i,j}\right\}\right] - E\left[\max_{j\in\mathcal{J}_i^0}\left\{u_{i,j}^0 + \epsilon_{i,j}\right\}\right]\right\}, \tag{3.11}$$

[14] Recall (3.6).

where $-\beta_1$ is the price coefficient in the alternative-specific utility term. When $\epsilon_{i,j}$s are distributed following the i.i.d. double exponential, $E\max$ terms of (3.11) reduce to the well-known log-sum formula. Otherwise, the $E\max$ terms are usually calculated using simulation integrals.

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 E.

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.[15] 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.

Utility Specification and Likelihood 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

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

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. Define the per-ounce utility of consumer i purchasing item j as

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

The intercept term α, which is to be 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 $\{\pi_{i,j}\}_{j \in \mathcal{J}_i}$. 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 $\exp\left(\mathbf{d}_{i,j}' \gamma\right)$, that is,

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

where $\mathbf{d}_{i,j}' \gamma$ 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 per-ounce 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.

I use the Shannon entropy for the information-cost function, which yields the simple logit form of the conditional choice probability. The probability of choosing alternative j becomes

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

which is taken as the conditional likelihood of i choosing j.

Identification The consideration-shifter parameter γ is identified by the variation in the promotional activities and 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 $\alpha = 1$, which allows μ_i to be identified against α.

Inference 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, given by

$$\prod_{i,j} \Pr_i \left(i \text{ Chooses } j | \mathbf{u}_i \right)^{\omega_i \mathbf{1}(i \text{ Chooses } j)} .$$

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

$$\nabla_{i,j} := \nabla \left[\omega_i \ln \Pr_i \left(i \text{ Chooses } j | \mathbf{u}_i \right) \right]$$

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

$$\mathbf{V} = \mathbf{\Omega}^{-1} \mathbf{\Delta} \mathbf{\Omega}^{-1}, \tag{4.5}$$

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 (4.5) in the following for the inference on model-parameter estimates.

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 $\left(-\hat{\beta}_1, \hat{\boldsymbol{\beta}}_2 \right)$ are $\left(-\hat{\beta}_1 \mu_i, \hat{\boldsymbol{\beta}}_2 \mu_i \right)$, which is specific to each individual i. The implied price elasticity of an alternative j is calculated by the following formula:

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

which uses the approximation of market shares s_j as $s_j \simeq \Pr_i(i \text{ Chooses } j|\mathbf{u}_i)$. The implied *Mean Price Elasticity* is -2.232, which is of a reasonable magnitude. All the major brand coefficients are positive and statistically and economically significant, implying a substantial brand premium exists in this market.

The $\hat{\gamma}$ 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 the store, learning the locations of shelves will require less time and effort. It is 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 lower unit information cost. Except 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 information cost parameterized in our RI framework provides face validity to the results.

In-sample and Out-of-sample Model Fits with Different Specifications on the Information Shifters
It is worth asking how our RI framework performs empirically relative to the RUM framework. 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 choice samples 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 parameterizing the dispersion parameter.[16] 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 $\left(-\hat{\beta}_1, \hat{\boldsymbol{\beta}}_2\right)$ 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 are applied to the holdout sample (Hit Rate and Average Hit Probability) reported in Table 4. Hit Rate is defined by the proportion out of 10,000 holdout samples that the choice alternative with the highest predicted choice probability is actually chosen, and Average Hit Probability is defined by the average predicted choice probability of the chosen alternative. Because,

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

Table 3: Model-Parameter Estimates from Maximum Likelihood Estimation

Mean Price Elasticity		-2.232			
Utility Parameter $\left(-\hat{\beta}_1, \hat{\boldsymbol{\beta}}_2\right)$				Consideration Shifter Parameter $\hat{\gamma}$	
Per-pack Price $\left(-\hat{\beta}_1\right)$	-0.129^{***}	75oz. $\leq$ Pack $\leq$150oz.	-0.195^{***}	In-store Display	0.552^{***}
	(0.008)		(0.036)		(0.017)
Pack $\leq$75oz.	-0.499^{***}	150oz. $\leq$ Pack $\leq$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 $\hat{\gamma}$ 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.

on average, each choice instance has 100 choice alternatives, Hit Rate and Average Hit Probability would be around 0.01 if the specified model is 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.

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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log-likelihood	–551276	–584913	–561266	–595326	–563005	–597734	–671885	–713029
AIC	1102631	1169904	1122608	1190729	1126087	1195545	1343845	1426133
BIC	1103021	1170296	1122987	1191111	1126467	1195927	1344215	1426505
Hit Rate	0.267	0.265	0.269	0.268	0.245	0.246	0.112	0.115
Average Hit Probability	0.102	0.102	0.098	0.098	0.086	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	–557641	–591384	–567593	–601720	–569729	–604642	–679874	–721131
AIC	1115360	1182846	1135263	1203516	1139534	1209361	1359822	1442337
BIC	1115750	1183238	1135643	1203898	1139914	1209743	1360191	1442709
Hit Rate	0.261	0.261	0.265	0.264	0.245	0.245	0.118	0.118
Average Hit Probability	0.098	0.098	0.094	0.095	0.082	0.082	0.021	0.022
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. Estimation sample has 160,698 choices with total of 15,992,553 alternatives. Holdout sample has 10,000 choices with total of 991,013 alternatives.

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 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. As in usual practice, I ignore equilibrium price and promotion responses by removing Tide Pods laundry detergent from consumers' choice set, because the supplier side is not explicitly modeled and many other laundry detergents are in the market where Tide Pods takes only around 0.5% of the market share.[17]

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 $\left(-\hat{\beta}_1, \hat{\boldsymbol{\beta}}_2, \hat{\gamma}, \hat{\boldsymbol{\theta}}\right)$ back into the respective functions (4.2), (4.3), and (4.4) in section 4.2. 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 (3.9) in section 3.3 as the following:

$$\begin{aligned} W_{RI}\left(\hat{\mathbf{u}}_i\right) = & \sum_{j\in\mathcal{J}_i} \hat{\Pr}_i\left(i \text{ Chooses } j|\hat{\mathbf{u}}_i\right)\hat{u}_{i,j} \\ & - \hat{\mu}_i^{-1}\left\{\sum_{j\in\mathcal{J}_i}\hat{\Pr}_i\left(i \text{ Chooses } j|\hat{\mathbf{u}}_i\right)\left\{\ln\hat{\Pr}_i\left(i \text{ Chooses } j|\hat{\mathbf{u}}_i\right) - \ln\hat{\pi}_{i,j}\right\}\right\}. \end{aligned} \tag{4.6}$$

The model corresponding to RUM, which does not include consideration shifters or parameterized unit-information-cost function, is also separately estimated to find the welfare changes calculated using the RUM-based formula.[18] The rest of the welfare function and compensating variation calculation follows as explained in section 3.3.

Figure 4.1 depicts the distribution of the compensating variation 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_{RI}, and only around one third of CV_{RI} are positive. By contrast, CV_{RUM} is constantly positive and tends to be much higher than CV_{RI}.

Average per Shopping Trip row of Table 5 summarizes the average compensating variation per shopping trip, in which the average is calculated after weighting for the panel projection factors ω_i. CV_{RI} and CV_{RUM} are calculated according to the formulas (3.10) and (3.11), respectively. Note that average CV_{RI} per shopping trip is slightly negative, whereas CV_{RUM} is positive. Comparing columns (1) and (2), CV_{RUM} overestimates the compensating variation not just in terms of magnitude, but also in signs, compared to CV_{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_{RI} is –\$9.2 million, CV_{RUM} is \$69 million, and the total annual Pods detergent sales equal \$346 million.

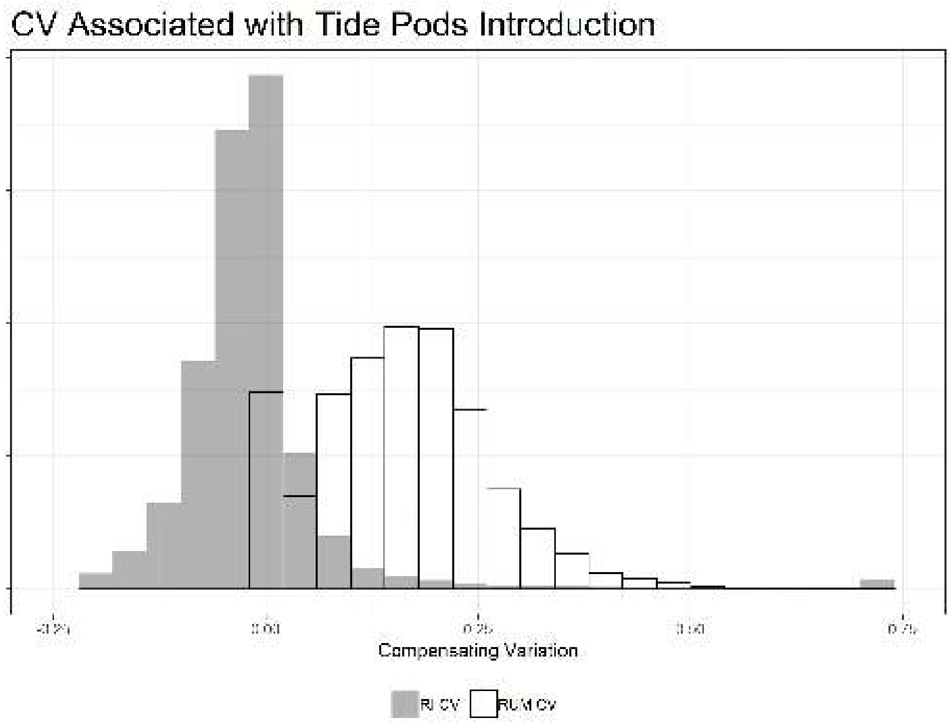
[17] My dataset contains 357 different laundry-detergent brands, with an average of 100 SKUs on the shelves.

[18] See Online Appendix D for the model-parameter estimates of RUM model.

CV_{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.

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 is about pricing. To that end, I examine how necessary price cuts are to making consumers 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} when, *ceteris paribus*, Tide Pods laundry detergent has n% price discounts. Approximately 7%-8% price discounts can make the consumers 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 consumers.

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. RI CV change is calculated using the formula (3.10), and RUM CV change is calculated using the formula (3.11). 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	–\$0.021	\$0.157
Annually Projected	–\$9,275,187	\$68,993,055
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%	7%	8%	10%	15%
CV_{RI}	–\$9,275,187	–\$3,043,713	–\$400,704	\$954,861	\$3,736,554	\$11,126,600

Note. This table presents the hypothetical average 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 based on RI theory, where the source of stochasticity in consumer choice probability stems from corresponding information-acquisition costs. Unlike the conventional RUM models assuming consumers' complete knowledge of their consumption utility on all alternatives in the observed choice set, the proposed RI model considers consumers' degree of (in)attention to the consumption utility and corresponding costs to acquire more information on the product-specific utility. In doing so, the framework allows consumer-welfare to decrease when an inferior option is added to a choice set, primarily due to the cost of information acquisition that leads to consumers' imperfect information acquisition.

Based on the conventional knowledge of consumer-welfare from the workhorse RUM models, managers may be misled to believe that 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 category of laundry detergent shows that 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 7%-8%.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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

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.

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 $W_{RUM}(\mathbf{u}_i)$ 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.[19]

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

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

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

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

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.

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

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 $\tilde{\mathbf{u}} := (u_1, u_2, ..., u_{J-1}, 0)$.[20]

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,[21] and 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 $1, 2, ..., J-1$,

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

[20]This normalization is innocuous because H_j is homogeneous of degree 1.

[21]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

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

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

[22] Define

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

as the density of the distribution $Q(\cdot)$. Because $D : \mathbb{R}^{J-1} \to \mathcal{A}$ is globally invertible and q^D has a 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 $E\left[y_j\right] = \pi_j \ \forall j \in \mathcal{J}$. The restrictions 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.

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

Appendix for Online Publication

C Log-Likelihood and Its Derivatives for the Model-Parameter Estimation

The maximum likelihood estimation problem that I consider is

$$\max_{(\gamma,\boldsymbol{\beta},\boldsymbol{\theta})} \ln \left(\prod_{i,j} \Pr\left(i \text{ Chooses } j\right)^{\omega_i \mathbf{1}(i \text{ Chooses } j)} \right),$$

for which I supply the exact derivatives of the nonlinear optimizer.

The log-likelihood of the first block is

$$\mathbf{1}\left(i \text{ Chooses } j\right) \omega_i \ln \Pr\left(i \text{ Chooses } j\right),$$

where

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

The corresponding gradients are

$$\begin{aligned}
\frac{\partial \ln \Pr\left(i \text{ Chooses } j\right)}{\partial \gamma^{(p)}} &= d^{(p)}_{i,j} - \frac{\sum_{j' \in \mathcal{J}_i \setminus 0} d^{(p)}_{i,j'} \exp\left(\mathbf{d}'_{i,j'}\boldsymbol{\gamma} + \mu_i u_{i,j'}\right)}{1 + \sum_{j' \in \mathcal{J}_i \setminus 0} \exp\left(\mathbf{d}'_{i,j'}\boldsymbol{\gamma} + \mu_i u_{i,j'}\right)} \\
\frac{\partial \ln \Pr\left(i \text{ Chooses } j\right)}{\partial \beta^{(p)}} &= x^{(p)}_{i,j}\mu_i - \frac{\sum_{j' \in \mathcal{J}_i \setminus 0} x^{(p)}_{i,j'}\mu_i \exp\left(\mathbf{d}'_{i,j'}\boldsymbol{\gamma} + \mu_i u_{i,j'}\right)}{1 + \sum_{j' \in \mathcal{J}_i \setminus 0} \exp\left(\mathbf{d}'_{i,j'}\boldsymbol{\gamma} + \mu_i u_{i,j'}\right)} \\
\frac{\partial \ln \Pr\left(i \text{ Chooses } j\right)}{\partial \theta^{(p)}} &= w^{(p)}_i \mu_i u_{i,j} - \frac{\sum_{j' \in \mathcal{J}_i \setminus 0} w^{(p)}_i \mu_i u_{i,j'} \exp\left(\mathbf{d}'_{i,j'}\boldsymbol{\gamma} + \mu_i u_{i,j'}\right)}{1 + \sum_{j' \in \mathcal{J}_i \setminus 0} \exp\left(\mathbf{d}'_{i,j'}\boldsymbol{\gamma} + \mu_i u_{i,j'}\right)}.
\end{aligned}$$

D Estimation Results for the RUM Specification

Table 7 presents the estimation results from the RUM specification, without including any consideration shifters nor parametrizing the information-cost function. As usual in estimating any discrete-choice RUM models, 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.[23]

[23] Because the dispersion parameter is normalized to 1 in the specification of Table 7, the average value of $\hat{\mu}_i$ has to be multiplied by the utility parameter estimates $\left(-\hat{\beta}_1, \hat{\boldsymbol{\beta}}_2\right)$ presented in Table 3 to directly compare the magnitude of the

Table 7: Model Parameter Estimates from Maximum Likelihood Estimation

Mean Price Elasticity		-2.679	
Utility Parameter $\left(-\hat{\beta}_1, \hat{\boldsymbol{\beta}}_2\right)$			
Per-pack Price $\left(-\hat{\beta}_1\right)$	-0.291^{***}	75oz. $\leq$ Pack $\leq$150oz.	-1.065^{***}
	(0.003)		(0.050)
Pack $\leq$75oz.	-1.718^{***}	150oz. $\leq$ Pack $\leq$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)
$n\times$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. 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 the parentheses. AIC and BIC are calculated after adjusting the panel-projection factor weights sum up to the number of choice observations.

E 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.

E.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 8. I drop the samples with missing store codes.

Table 8: 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.

E.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.[24] I classify the functional product characteristics of laundry detergents as in Table 9. Each criterion listed in Table 9 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.[25] In ad-

utility parameter estimates presented in Table 7.

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

[25]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 9.

dition 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 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 9: 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. 10,911 different UPCs are observed in the data.

E.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.[26] 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 10 lists the databases that I use for search.

[26]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 10: 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

F 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.

F.1 Logit and Nested Logit

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

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

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

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

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).

F.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 $\lambda_g \in (0, 1)$.

F.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 $k, l \in \mathcal{J}$.

F.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 $b_{jl} > 0$.