Designing Context-Based Marketing:

Product Recommendations under Time Pressure*

Kohei Kawaguchi[†] Kosuke Uetake[‡] Yasutora Watanabe[§]

April 25, 2020

Abstract

We study how to design product recommendations when consumers' attention and utility are influenced by time pressure—a prominent example of the context effect—and menu characteristics, such as the number of recommended products in the assortment. Using unique data on consumer purchases from vending machines on train platforms in Tokyo, we develop and estimate a structural consideration set model in which time pressure and recommendations can influence attention and utility. We find that time pressure reduces consumer attention but increases utility. Time pressure moderates the effect of recommendations for the attention of both recommended and non-recommended products, and utility for recommended products. Moreover, the number of total recommendations increases consumer attention in general, but in a diminishing way. In our counterfactual simulations, we find that the revenue-maximizing number of recommendations decreases with time pressure, and that optimizing recommending products to accommodate time pressure by a greedy algorithm increases total sales volume by 3.7% relative to the actual policy, 0.6% points more than traditional consumer-segment-based targeting policy. This effect is larger than 10% price discounts, which increases the revenue only by 0.4% at the margin.

Keywords: Attention, Consideration set, Context-based marketing, Time pressure, Recommendations, Menu effects

*We thank Bart Bronnenberg, Andrew Ching, Jean-Pierre Dube, Yufeng Huang, Ahmed Khwaja, Vineet Kumar, Carl Mela, Aniko Oery, Thomas Otter, Stephan Seiler, Jiwoong Shin, Matt Shum, K. Sudhir, Elie Tamer, Raphael Thomadsen, and Nathan Yang for useful discussions and suggestions. We also thank seminar/conference participants at the 2017 UTD Forms Conference, the 2016 Marketing Science Conference, University of Tokyo, and Yale, and financial support from Japan Center for Economic Research. This paper is previously circulated as "Identifying Consumer Attention: A Product-Availability Approach." Identification argument in the previous draft is available in Online Appendix C. All errors are ours.

[†]Hong Kong University of Science and Technology. kkawaguchi@ust.hk
[‡]Yale School of Management. kosuke.uetake@yale.edu
[§]University of Tokyo. yasutora.watanabe@gmail.com

1 Introduction

Context-based marketing attracts increasing attention from marketing managers as more real-time consumer behavioral data become available. Foursquare, for example, sends messages to consumers when they are close to shops or restaurants that they are predicted to visit. In fact, a large number of behavioral studies in marketing find that consumers tend to behave differently under context factors such as time and social pressures. Dhar and Nowlis (1999) find that consumers are more likely to avoid making any decisions when they are under time pressure to save their cognitive resources. Using in-store shopper movement data, Hui, Bradlow, and Fader (2009) find that consumers' purchase decisions are affected by time pressure. More recently, Kawaguchi, Uetake, and Watanabe (2019) study the effects of product recommendations in a beverage vending machine under time and crowd pressures, and find that time pressure significantly affects the effectiveness of marketing interventions.

Despite its practical importance, how context-based marketing works, and how managers can optimize it are not well understood. This paper aims at filling the gap by studying the mechanism behind the effect of recommendations under contextual factors, and by investigating how to optimize marketing interventions to contextual factors. More specifically, we decompose the effect of recommendation to the one through consumer attention and the other through preference when consumers are under time pressure, and study the design of the recommendation system for beverage vending machines, taking into account the effects of time pressure.[1] To do so, we structurally estimate a consideration set model in which both time pressure and product recommendation can affect consumer attention and utility. Moreover, our model allows time pressure to affect the effectiveness of product recommendations and allows consumer attention to depend on "menu"-related variables such as the number of recommended products and the number of unique products in the assortment. Our consid-

[1] We focus on time pressure given that Kawaguchi, Uetake, and Watanabe (2019) show large effects of time pressure on the effectiveness of recommendation in the same context, while the effects of crowd pressure are small and not robust.

eration set model is motivated by a finding of Kawaguchi, Uetake, and Watanabe (2019) that recommendations have strong positive spillover effects to the sales of the non-recommended product on the menu.[2] Similar spillover effects of advertising are also found in online advertising (e.g., Sahni (2016)). As a recommendation to other products is unlikely to affect consumer's preference over the product, a natural interpretation of the spillover effect is that it affects consumer's attention to non-recommended products. Disentangling this mechanism through which recommendation affects preference and attention is essential to design the optimal recommendation, especially under contextual factors that may influence the effectiveness of recommendations on preference and attention in a different way.

We use the data on the consumer beverage purchases from vending machines in Japan, which is also used in Kawaguchi, Uetake, and Watanabe (2019). There are several unique features in the setup, which allow us to estimate our flexible consideration set model. First, a reliable proxy variable that captures time pressure is readily available in our setup, allowing us to study time pressure in a non-laboratory environment, in contrast to extant studies, which are mostly laboratory-based. The vending machines we study are located on the platforms of train stations in Tokyo. Hence, the time until the next train is a natural proxy for the time pressure consumers feel when purchasing a product.[3] By utilizing the train schedule information, we can precisely measure time pressure at the minute level.

Second, our consumer purchase data comes from a period when the company owning the vending machines executed an experiment on product recommendations. Since the vending machines we study are equipped with a recommendation system based on the installed camera, which recognizes consumer attributes, the vending machines can change recommenda-

[2] Moreover, Ching, Erdem, and Keane (2009), Ching and Hayashi (2010), Conlon and Mortimer (2013) (and other papers) also point out the importance of incorporating choice set heterogeneity to correctly infer the consumer preference. Ching, Erdem, and Keane (2009) estimate the price consideration model and show the crucial role of allowing flexible consideration sets. Ching and Hayashi (2010) use unique survey data that contain information about choice sets and study the effect of credit card reward programs on the consumer's payment method choice. Conlon and Mortimer (2013) use the information about stock-out to construct exact choice sets and show that the estimates of price elasticity are biased if stock-outs are not taken into account.

[3] In Kawaguchi, Uetake, and Watanabe (2019), they conducted a validation experiment to study if the time to the next train is strongly correlated with the time pressure felt by subjects, and they confirmed this relationship.

tions depending on consumer demographic information. Estimating the effect of recommendations is challenging in general due to the endogeneity bias resulting from the fact that popular products are likely to be recommended, and the popularity of the recommended products may be wrongly attributed to the effect of recommendations. With our exogenous experimental variation in product recommendations, we can identify the causal effect of product recommendations without much concern for the endogeneity of recommendations.

Third, product assortments vary greatly across vending machines. This variation in assortment (that is, the available set of products) allows us to create menu-related variables such as the number of unique products sold in the vending machine and the number of columns that each product occupies. These variables shift consumer attention and hence are important to consider optimal product recommendations. Moreover, the menu-related variables tend to be excluded from the consumer utility function, which helps us identify the consideration set model. These additional exclusion restrictions allow us to include advertising variables for both attention and utility. Typically, in existing papers that estimate consideration set models, advertisements are included only in consumer attention, but not in utility (see, e.g., Goeree (2008) and Van Nierop, Bronnenberg, Paap, Wedel, and Franses (2010)). Including advertising in the utility function allows us to (indirectly) test the signaling and persuasive role of advertising. [4]

[4] As Bagwell (2005) summarizes, the marketing and economics literature have theoretically considered the informative role and the persuasive role of advertisements (e.g., Stigler (1961); Grossman and Shapiro (1984); Milgrom and Roberts (1986), Becker and Murphy (1993)). Empirically, these roles are tested in various settings. For example, the literature of structural learning models, as surveyed by Ching, Erdem, and Keane (2013) and Ching, Erdem, and Keane (2017), studies how advertisements affect the learning process and consumer preference. In this literature, Ching and Ishihara (2010) estimate a dynamic learning model of detailing in the pharmaceutical industry, in which physicians face uncertainty about the effectiveness of drugs and learn it over time through prescription. Detailing can directly affect the physician's prescription decisions and also the physician's learning process by providing information. Also, Ching and Ishihara (2012) study the information and persuasive role of detailing in the pharmaceutical industry. Their identification strategy is based on the observation that the informative component of detailing is chemical specific while the persuasive component is brand specific. Focusing on markets where some drug companies use a comarketing agreement, they identify the informative role of detailing by looking at the market shares of chemicals and the total detailing efforts aggregated across brands using the same chemical. Papers on informative advertising in the literature of structural learning do not model attention or consideration set with only a few exceptions, such as Ching, Erdem, and Keane (2014) where the price consideration mechanism of Ching, Erdem, and Keane (2009) is incorporated into a structural learning model.

Outside the structural learning literature, Ackerberg (2001) proposes a reduced-form way to separately estimate

Lastly, the experiment on the product recommendations and the variations in the product assortment at the vending machine level generate variation in the "menu"-related variables for consumer attention such as the number of recommended products, the number of unique products, the number of slots for each product, etc. In our set up, the menu-related variables are well defined and have ample cross-machine variations. We demonstrate the importance of taking into account the effects of these "menu"-related variables, and our counterfactual simulations examine how many products the company should recommend under varying degrees of time pressure.

The estimation results show that i) as time pressure increases, consumers pay less attention to each product but more likely to purchase products; ii) product recommendations positively and significantly affect both attention and utility; iii) time pressure weakens the effectiveness of recommendations; iv) menu characteristics significantly affect consumer attention- in particular, the number of total recommendations increases the attention level in general, but in decreasing order (Table 4). Moreover, we find significant heterogeneity in the effects of recommendations across customer segments (Table 5).

To quantify the efficacy of product recommendations on attention and utility, we calculate the elasticities of recommendations to purchasing incidence. We find that, compared to the baseline where no product is recommended, recommending a product increases its sales by 18.7% (*own* elasticity) on average, which can be decomposed into the effect through attention (9%) and utility (8.6%) (Table 6). Moreover, we find that these elasticities vary by time pressure, and they are maximized under moderate time pressure (Table 7).

The findings above lead us to study the design of recommendations that depend on time pressure through a series of counterfactual analyses. We first investigate how the revenue-maximizing *number* of recommendations varies by the degree of time pressure. Although each recommendation may increase the attention and choice probability of the recommended

the information effect and the prestige effect using the variation in consumer experiences. Anand and Shachar (2011) study the informational role of advertising in matching consumers with products when they are uncertain about product attributes.

product, it may not be optimal to recommend too many products because it dilutes consumer attention.

To do so, we compare the performance of three different recommendation strategies with the actual recommendation. Our first recommendation system maximizes the total sales by choosing the same number of recommendations for all vending machines (*uniform policy*). Our second recommendation system maximizes the total sales by consumer segment, which is basically similar to the traditional targeting strategy based only on the consumer demographic variables (*segment-based policy*). Finally, the third recommendation system, the context-based recommendation strategy, adjusts the number of recommended products by time to the next train for each vending machine. We find that the uniform policy and the consumer-segment-based policy can increase the sales by 0.82% and 0.88% compared to the actual policy, whereas the context-based policy increases sales by 1.79% (Table 8). Hence, our results show that context-based marketing can increase sales more than traditional marketing in our setting.

Lastly, we quantify the impacts of optimizing which products to recommend (not just the number of recommendations as in the previous case) at the consumer-segment-level, and at the context-level. Because finding the optimal products to recommend is computationally infeasible, we use a greedy algorithm to find semi-optimal products to recommend. Again, the results show that the context-based policy can outperform the existing strategies: uniform policy and consumer-segment-based policy increases the revenue by 3.13% and 3.15%, whereas the context-based policy increases sales by 3.71% (Table 9). Thus, the context-based recommendations can further increase sales by optimizing the set of recommended products together. These effects are comparable to the effects of price discounts. Moreover, using product recommendations and maximizing its benefits by making them context-based can be more attractive for a manager than price discounts. By simulation, we find that 5% price discount of all products boosts sales volume by 5.5%. However, because the price decreases, the revenue increases only by 0.2% ($1.055 \times 0.95 = 1.00225$). Cutting the price of all products by 10% increases the sales volume by 11.6%. This results in 0.4% increases in the revenue ($1.116 \times 0.9 = 1.0044$).

Related Literature This paper builds on the literature on the consideration set models that have been studied both in marketing and economics for a long time (see, e.g., Manski (1977), Roberts and Lattin (1991), Allenby and Ginter (1995), Mehta, Rajiv, and Srinivasan (2003), Ching, Erdem, and Keane (2009), Masatlioglu, Nakajima, and Ozbay (2012), and Manzini and Mariotti (2014)).[5] The consideration set model is useful as it allows one to study the effect of advertising on consumer attention (e.g., Van Nierop, Bronnenberg, Paap, Wedel, and Franses (2010)) or the consequence of ignoring limited attention to biased estimates of price elasticities (e.g., Goeree (2008) and Conlon and Mortimer (2013)).[6]

Some recent papers expand the extant literature. For example, Bronnenberg and Huang (forthcoming) and Dehmany and Otter (2014) propose an alternative approach to model consideration sets, that exploits the variations both in quantity purchased and in purchased products. Although there is no sufficient variation in quantity choice in our empirical setting (i.e., almost all customers purchase only one product on a purchase occasion), this is a useful approach when such variations exist. Abaluck and Adams (2018) propose a new identification strategy for the consideration set model based on asymmetric demand responses to changes in product characteristics. Our paper adds to this literature by explicitly considering the effects of contextual factors in the consumer purchase funnel as well as exploiting the variation in product availability.

Our paper is also related to the literature on time pressure. Although it is beyond the scope of our paper to list all papers related to time pressure in the psychology and consumer behavior literature, let us mention a few papers that are highly relevant to ours. Dhar and Nowlis (1999)

[5] Some papers use direct information about consideration sets, such as survey data. Ching and Hayashi (2010), Draganska and Klapper (2011), Honka, Hortacsu, and Vitorino (2017), and Palazzolo and Feinberg (2015), for example, employ survey data in which each consumer is asked which products they consider when making a purchase decision. This type of data directly identifies consideration sets. Due to the increased availability of detailed consumer search data, information about consideration sets would be more available in some markets such as online retailers. However, it could sometimes be costly to obtain such data (e.g., the financial cost of running a large-scale consumer survey) and there is a potential reporting bias due to the nature of surveys.

[6] Goeree (2008) studies the role of advertising for consumer information and choices in the US personal computer market. She pays special attention to the endogeneity of advertisement, and constructs approximated optimal instruments to deal with it.

find the choice deferral effect under time pressure, which indicates that consumers are less likely to make a purchase decision under time pressure. Hui, Bradlow, and Fader (2009) test the choice deferral effect in a supermarket purchase environment using consumer movement data. Reutskaja, Nagel, Camerer, and Rangel (2011) study the search process of subjects under time pressure in a laboratory setting using eye-tracking and find that choices are affected by time pressure. Finally, Kawaguchi, Uetake, and Watanabe (2019) examine the effectiveness of product recommendations when consumers are under time pressure. The paper finds that time pressure weakens this effectiveness, and we discuss the relationship to this paper below.

Lastly, our paper is related to the literature using the data from vending machines. Anupindi, Dada, and Gupta (1998), Conlon and Mortimer (2013), and Conlon and Mortimer (2019) study a vending machine data to see the effects of stock-outs on consumer demand. Anupindi, Dada, and Gupta (1998) estimate a reduced-form Poisson regression, while Conlon and Mortimer (2013) estimate a structural discrete-choice demand model, taking into account stock-outs. Finally, Conlon and Mortimer (2019) study the effects of vertical rebates in the vending machine industry by using the data from a field experiment.

Relationship to Kawaguchi, Uetake, and Watanabe (2019) Our consideration set model is motivated by one of the findings of Kawaguchi, Uetake, and Watanabe (2019), which use the same data set like ours, that recommendations have strong positive spillover effects to the sales of non-recommended products on the menu. As recommendations to other products are not likely to affect the preference for non-recommended products, this finding suggests that attention plays a critical role in understanding how recommendations affect consumer choices, in particular under contextual effects. Although Kawaguchi, Uetake, and Watanabe (2019) presents that recommendation and contextual factors interact by a series of reduced-form regression analysis, it is silent about the mechanism through which recommendations affect separately through preference and attention and how contextual factors interact with them. This limits the managerial implications of the study because a detailed understanding of the mechanism is essential in designing recommendations. The current paper adopts the consideration set model that separately identifies the effects of recommendation to preference and attention under contextual factors so that one can study the design of recommendations under contextual factors.[7]

The rest of the paper is organized as follows. Section 2 presents the background and the data for our empirical setup. Section 3 presents the consideration set model, and Section 4 discusses our identification and estimation strategies. Section 5 reports the estimation results, and Section 6 reports the counterfactual simulations. Finally, Section 7 concludes.

2 Background and Data

2.1 Background

We study consumers' beverage purchase decisions from vending machines placed at train stations in the Tokyo metropolitan area. For details on the setup, please see Kawaguchi, Uetake, and Watanabe (2019), which use the data set from the same setup.

In this study, we use approximately 460 vending machines that have a recommendation system. The machine recognizes the age and gender of each consumer with a camera attached to the top of the front panel and then recommends a different set of products depending on the consumer characteristics according to a pre-specified rule.[8] The recommendations are displayed on the front panel of the vending machine with colorful and flashing pop-ups and are hence easily recognizable by consumers (see Figure 1).[9] The company controls the recommendation policy through a centralized system but can vary it only by the time of day (morning: before 10 am, daytime: between 10 am and 6 pm, and nighttime: after 6 pm), and cannot do so at the machine level or hour level.

[7]Moreover, although it is possible to estimate a fully flexible discrete choice model that relates recommendations for each product to a sales of each product with sufficient data, this problem is high-dimensional, and hence we need a structural consideration set model to capture spill-overs in a lower-dimensional space.

[8]Because of privacy concerns, the cameras do not record any information on consumer characteristics. Hence, neither the company nor we can use the information from the cameras except to change the recommendations.

[9]Note that consumers can see all of the available products on display regardless of whether or not they are recommended.

Figure 1: An image of the touch-panel and product recommendations: The product recommendations are the flashing red bubble signs with the word "Recommended." Image supplied by the company.

2.2 Field Experiment on Product Recommendations

The company conducted an experiment with us to measure the impact of product recommendations using these vending machines. We briefly describe the experimental design, and the details can be found in Kawaguchi, Uetake, and Watanabe (2019). In the experiment, the company created the treatment condition, in which a set of products was recommended, and the control condition, in which no product was recommended. The set of recommended products in the experiment is chosen exogenously. The company then randomly allocated the treatment and control conditions at three different times of day for weekdays during the week of July 15 to 26, 2013, as shown in Table 1.[10]

The experiment creates exogenous variations in recommendations, which also creates variation in the number of recommended products among treatment groups.[11] Because available products are different across vending machines, the number of recommended products can

[10]In Table 1, the sign "-" indicates that the company ran its regular recommendations, for which the company (not us) chose which products to recommend. We do not use these observations because of endogeneity concerns: the set of recommended products is likely to include more popular products; hence, estimates could be biased upwards.

[11]Note that the experiment is not meant to create exogenous variations in product availability. Instead, it creates exogenous variations in product recommendations, which are typically endogenously determined by the firm. Although we cannot completely rule out the concern for endogeneity in product availability, we think the endogeneity concern could be limited, because the company delegates the decision on fulfilling vending machines to local agents who are responsible for several machines or stations, and the company is worried about the poor performance of capturing local demand by the current system, based on our discussion with the company.

Table 1: Experimental Design

	15 Mon	16 Tue	17 Wed	18 Thu	19 Fri	22 Mon	23 Tue	24 Wed	25 Thu	26 Fri
Morning	-	T	-	T	C	C	-	T	C	-
Daytime	T	-	T	C	-	-	T	C	-	T
Night	-	T	C	-	T	T	C	-	T	C

Note: This table is the same as Table 1 in Kawaguchi, Uetake, and Watanabe (2019). The company conducted the experiment using treatment T and control C. The number at the top is the day in July 2013, and the second line represents the day of the week. No product is recommended for control C. In the slot with a bar, the vending machine shows the product recommendations chosen by the company, but we do not use the data for these slots in our empirical analysis.

also be different across vending machines under the treatment condition. From the experimental data, we construct the following variables: MR, PR, and NR. MR is a machine-time level indicator variable showing that the machine at the time is under treatment, PR is a machine-product-time level indicator variable for whether a particular product is recommended, and NR is the number of recommended products in the vending machine at the time.

2.3 Time Pressure

One of our main interests is to examine the effects of contextual factors in designing recommendation systems. We focus on time pressure, a prominent example of the context effects (see, e.g., Dhar and Nowlis (1999)). Time pressure, however, is not usually measurable in a non-laboratory environment. Thus, most extant studies on time pressure are conducted in laboratory settings. In our case, we exploit a naturally occurring exogenous variation, *train schedule*. The idea is that consumers feel more time pressure when the next train is approaching because they make a purchase decision within a limited time. Because trains in Tokyo operate punctually and arrive frequently, passengers tend to be under the influence of time pressure.[12]

We obtain the schedule of the Japan Railway East on weekdays during the experiment.

[12] Since electronic bulletin boards at the ticket gate, concourse, and platform of each station display the departure times of the next train and the one after, consumers can easily tell how soon the next train will arrive.

The data cover all of the trains and stations that the railway company operates. Using the train schedule data, we calculate the time until the next train, measured in minutes, as the primary proxy variable for time pressure (denoted as $T1$). One may wonder if the consumer may not feel time pressure if the train after the next one arrives shortly, even though the next train arrives shortly. Hence, we also create $T2$, which is the time until the train after the next one, to address this possibility as well.[13]

2.4 Menu Effects

In addition to studying time pressure, we examine the menu effects, which are the effects of the characteristics of the menu, such as the number of products in the assortment and the number of recommended products. Although existing empirical works on the consideration set model focus mostly on the effects of product attributes on consumer attention and utility, the design of assortment can also have impacts on consumer attention (see, e.g., Chandon, Wesley, Bradlow, and Young (2009)). Our vending machine setup is unique in that the choice menu is relatively simple compared to therein the other settings such as a grocery store, and the menu-related variables are well defined.

We consider the effects of the number of available unique products in a machine and the number of slots assigned to each product in a machine. Also, we consider the number of recommended products, which we introduced in the previous section, as a menu-related variable. These variables may influence consumer attention: too many products in an assortment do not allow a consumer to spend enough time considering all available products, and a product occupying more slots allows a consumer to pay more attention to it (see, e.g., the top row of Figure 1, where multiple products occupy more than one slot, and one recommended product occupies three slots).

[13]To examine the validity of these proxy variables, in Kawaguchi, Uetake, and Watanabe (2019), they run a field test with about 100 undergraduate subjects. The results indicate that consumers feel more pressure as the next train approaches. The details of the validation test are available from the authors upon request.

2.5 Data

Table 2: Summary Statistics

Type of Data	*Variable*	*N*	*Mean*	*Sd*	*Min*	*Max*
Machine/date/time/product						
	Recommendation	209801	0.09	0.28	0	1
	Slots per product	209801	1.15	0.37	1	5
Machine/date/time						
	Sales units	7309	4.53	3.6	0	37
	Sales value (JPY)	7309	614.65	488.8	0	5100
	Recommendation	7246	0.58	0.49	0.0	1.0
	# of recommendations	7309	2.59	2.56	0	10
	# of unique products	7309	29.63	2.12	22	34
	Degrees Celsius	7246	26.02	1.39	20.5	28.5
Purchase						
	Mins to next train	32464	2.35	11.73	0	299
	Mins btwn following train	32464	3.02	8.77	1	325
Product						
	Price (JPY)	95	134.63	18.38	100	200
	Volume (ml)	95	330.68	132.12	100	600
	Availability	95	0.30	0.29	0	1
	Plastic bottle	95	0.66	0.48	0	1
	Can	95	0.27	0.45	0	1
	Glass bottle	95	0.06	0.24	0	1
Consumer						
	Male	32464	0.70	0.46	0	1
	Junior	32464	0.16	0.36	0	1
	Old	32464	0.17	0.38	0	1

Notes: The data includes only point club members. Junior is no greater than 30 and old is above 50. Recommendation in the first panel is a dummy variable to indicate whether a product is recommended at a machine on a certain purchase timing, whereas recommendation in the second panel is a dummy variable to indicate whether a machine recommends products or not. The availability of a product is the proportion of machine-time in which the product is available.

The primary data set is directly obtained from the company's point-of-sale database. The data contain time and product of purchase as well as 460 vending machines that are equipped with the recommendation system. We focus on the customers who registered with the company's membership program as we can observe their demographic information. We use their demographics to study heterogeneous effects of recommendations. These sets of customers use an electric card, which also works as a commuter card, to purchase products. The purchase time is recorded at the second level, which enables the precise measurement of the time to the next train. In the data, there are, unfortunately, not many repeat customers during the period we study.

Table 2 reports the summary statistics of the variables we use in the estimation. The average sales by the sample customers per machine in a time period (i.e., morning, daytime, and nighttime) are about 4.5. An average vending machine sells 29.6 products, among which 2.6 products are recommended. We find large variations in these variables as well. The average temperature is 26.6 degrees in Celsius, which is typical for the summer season in Tokyo.[14]

We calculate the minutes to the next train (T1) and the minutes between the next train and the one after (T2) from the train schedule data. The average number of minutes to the next train is 2.4 minutes, and the interval is 3.0 minutes. Hence, trains arrive at stations quite frequently and create time pressure for passengers.

There are approximately 100 distinct beverage products across all vending machines, and the average price is 134 Japanese Yen (about $1.3). Beverages are sold in three types of packages: plastic bottles, cans, and glass bottles, and the average volume is 330 ml. There is no price variation across locations and time within a product. At the time of the experiment, there was also little price variation across products, controlling for the volume and the package.

The bottom part of Table 2 reports the customers' demographic information. In our estimation, we use information about customers whose characteristics are recorded, which is about 32,000 customers. About 69% of them are male, and about 16% are categorized as junior (no older than 30), and 17% are old (older than 50) according to the company's categorization.

To show the variation in product availability, Figure 2 presents how many products are available at how many vending machines. Among about 100 products that the company sells, more than 50% are available at less than 30 locations out of 460. This figure implies that product availability has large cross-sectional (that is, cross-machine) variations.

[14]The information about temperature is obtained from the Japan Meteorological Agency, and the locations are matched to each station. https://www.data.jma.go.jp/gmd/risk/obsdl/index.php

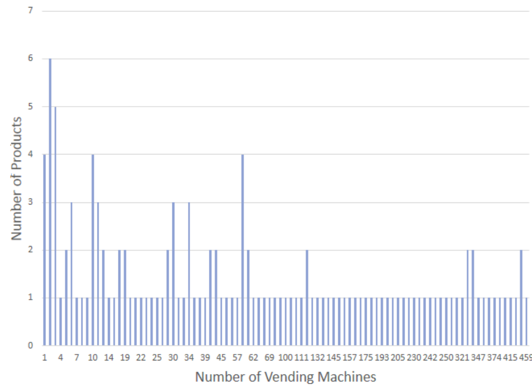


Figure 2: Product Availability: The horizontal axis is the number of vending machines that a particular product is available. The vertical axis is the number of products that corresponds to the number of machines that displays the product.

Finally, we provide a balance check for covariates between treatment and control in the online appendix B. We confirm that there is no systematic difference between the two groups.

3 Model

As we discussed above, one of the key findings of Kawaguchi, Uetake, and Watanabe (2019) is that product recommendations can spill over to other products that are not recommended. Motivated by this finding, we consider a consideration set model in which recommendations can spill over via consumer attention, and time pressure can affect attention and preference. In a consideration set model, the consumer first forms her consideration set and then chooses the product with the highest utility from the consideration set. Hence, the consumer may not be aware of some of the available products and may not necessarily choose her utility-maximizing product if she is not aware of it. Using the consideration-set model allows us to decompose the effectiveness of product recommendations into one from the attention channel and another from the preference channel, which we cannot do this with the usual multinomial logistic regression model.

The model is a discrete choice model wherein consumers do not necessarily know or do not consider all of the available products. The set $\mathscr{J}$ consists of all goods, regardless of their availability. The set of goods available in each choice occasion is a subset of $\mathscr{J}$, and we denote the set of available products in purchase occasion t as $\mathscr{J}_t \subseteq \mathscr{J}$. Consumers can always choose the outside option and buy nothing, $j = 0 \in \mathscr{J}_t$. We call the set of available goods in purchase occasion t (and the outside option of not buying) as the *feasible set*, denoted by $\mathscr{J}_t$, while we call the set of products that consumer i actually considers as the *consideration set*, denoted by $\mathscr{C}_{it}$. In the first stage, the consideration set for consumer i (that is, $\mathscr{C}_{it}$) is determined, and in the second stage, consumer i chooses a product from those in $\mathscr{C}_{it}$ that maximizes her utility. We explain each stage in order.

Stage 1: Consideration Set Formation The first stage concerns how the consideration set is formed. Whether a good is included in the consideration set is determined by the level of attention that a consumer pays to it, which is denoted by V^*_{ijt}. To be precise, good j is included in the consideration set if the following condition is satisfied:

$$C_{ijt} = \mathbb{1}\{V^*_{ijt} > 0\}, \quad j \in \mathscr{J}_t, \tag{3.1}$$

where $C_{ijt} \in \{0,1\}$ indicates whether product j is considered ($C_{ijt} = 1$) or not ($C_{ijt} = 0$). We normalize the threshold at 0 without loss of generality. Then, consumer i's consideration set is written as $\mathscr{C}_{it} \equiv \{j \in \mathscr{J}_t | C_{ijt} = 1\}$. The level of attention V^*_{ijt} depends on the consumer and the product characteristics as follows:

$$V^*_{ijt} = \begin{cases} \alpha_{0i} A_{ijt} + \boldsymbol{\alpha}'_{1i} \boldsymbol{M}^V_t + \boldsymbol{\alpha}'_{2i} \boldsymbol{X}^V_{jt} + \alpha_{ij} + \zeta_j + \varepsilon_{ijt} \equiv V_{ijt} + \varepsilon_{ijt} & j \in \mathscr{J}_t \setminus \{0\} \\ \infty & j = 0, \end{cases} \tag{3.2}$$

where A_{ijt} is a dummy for product recommendation of product j for customer i at time t (advertisement in general), $\boldsymbol{M}^V_t$ is a vector of context factors and menu characteristics, which we will explain below. The vector $\boldsymbol{X}^V_{jt}$ contains a set of product-specific characteristics other than the price, some of which may vary by choice occasion t. α_{ij} is a consumer-level random effect in attention to product j (following a normal distribution with mean 0 and a standard error "S.D. α_{ij}"), and ζ_j is a product-specific shock common across consumers, and ε_{ijt} is a consumer-product-occasion-level i.i.d. idiosyncratic shock. Thus, the model allows for correlation in attention due to the product attributes and the consumer-level random effects, but attention is independent conditional on them.[15] This assumption follows the literature on the consideration set models (Goeree (2008); Van Nierop, Bronnenberg, Paap, Wedel, and Franses (2010)). The level of attention for good 0 (outside option) is positive infinity as the outside option is always included in the consideration set. We denote a vector of random coefficients by $\boldsymbol{\alpha}_i \equiv (\alpha_{0i}, \boldsymbol{\alpha}'_{1i}, \boldsymbol{\alpha}'_{2i})'$, which is a function of consumer characteristics $\boldsymbol{Z_i}$ as follows:

$$\boldsymbol{\alpha}_i = \boldsymbol{\alpha} + \boldsymbol{\Pi}^\alpha \boldsymbol{Z}_i + \boldsymbol{\Sigma}\boldsymbol{\nu}_i, \tag{3.3}$$

where $\boldsymbol{\Pi}^\alpha$ and $\boldsymbol{\Sigma}$ are the coefficients associated with the consumer characteristics and error terms. We assume that $\boldsymbol{\nu}_i$ follows an i.i.d. standard normal distribution and that the off-diagonal terms of $\boldsymbol{\Sigma}$ are zero to reduce dimensionality.

Stage 2: Product Choice In the second stage, given the consideration set $\mathscr{C}_{it}$ formed in the first stage, customer i chooses a product that maximizes his/her utility. Let us first introduce the utility from product j, U^*_{ijt}, which is given by:

$$U^*_{ijt} = \begin{cases} \beta_{0i}A_{ijt} - \beta_{1i}P_j + \boldsymbol{\beta}'_{2i}\boldsymbol{M}^U_t + \boldsymbol{\beta}'_{3i}\boldsymbol{X}^U_{jt} + \beta_{ij} + \xi_j + \eta_{ijt} \equiv U_{ijt} + \eta_{ijt} & j \in \mathscr{J}_t \setminus \{0\}, \\ \varepsilon_{i0t} & j = 0, \end{cases} \tag{3.4}$$

where we include a similar set of variables, $\boldsymbol{M}^U_t$ and $\boldsymbol{X}^U_{jt}$. P_j is the logarithm price of product j, which affects only the utility and is excluded from the attention.[16] β_{ij} is a consumer-level

[15]In our data, there are not many repeat customers. Therefore, it is not possible to include individual fixed effects in the consideration set formation.

[16]Note that the price is only included in utility, but not in attention. This formulation follows the model used in the literature such as Van Nierop, Bronnenberg, Paap, Wedel, and Franses (2010) and Ching, Erdem, and Keane (2009). Ching, Erdem, and Keane (2009) motivates this assumption by the fact that many advertisements do not contain price information (as in our case) and hence advertisements work as a trigger for consumers to pay atten-

random effect in the utility of product j (following a Normal distribution with mean 0 and a standard error "S.D. β_{ij}"), and ξ_j is a product-specific shock that is common across consumers. We assume that η_{ijt} follows an i.i.d. Type-I extreme value random distribution.

$\boldsymbol{\beta}_i \equiv (\beta_{0i}, \beta_{1i}, \boldsymbol{\beta}'_{2i}, \boldsymbol{\beta}'_{3i})'$ is a vector of random coefficients, which is a function of consumer characteristics $\boldsymbol{Z_i}$ as follows:

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\Pi}^{\beta} \boldsymbol{Z}_i + \boldsymbol{\Omega} \boldsymbol{v}_i, \tag{3.5}$$

where $\boldsymbol{\Pi}^{\beta}$ and $\boldsymbol{\Omega}$ are the coefficients associated with the consumer characteristics and error terms. We assume that $\boldsymbol{v}_i$ follows an i.i.d. standard normal distribution and that the off-diagonal terms of $\boldsymbol{\Omega}$ are zero to reduce dimensionality.

Finally, we describe a consumer's decision problem in the second stage. Let D_{ijt} be an indicator variable that takes a value of 1 when the good is chosen and 0 otherwise. Given the consideration set $\mathscr{C}_{it} = \{j \in \mathscr{J}_t | C_{ijt} = 1\}$ and the utility level $\{U^*_{ijt}\}_{j \in J_t}$, consumer i's choice can be described as

$$D_{ijt} = \mathbb{1}\{U^*_{ijt} \geq \max_{k \in \mathscr{C}_{it}} \{U^*_{ikt}\}\}, j \in \mathscr{C}_{it}. \tag{3.6}$$

We assume that the error terms in V^*_{ijt} and ones in U^*_{ijt} are independent. This assumption is not without loss of generality, but is standard in the consideration set literature.[17]

Empirical Specification Table 3 lists the variables we include in the estimated model. As discussed above, a fundamental departure of our consideration set model from the standard one comes from the menu- and context-related variables, $\boldsymbol{M}^V_t$ (in attention), and $\boldsymbol{M}^U_t$ (in preference). $\boldsymbol{M}^V_t$ includes MR, NM, NR, T1, T2, while $\boldsymbol{M}^U_t$ includes T1, T2, and Temp. Hence, MR, NM, and NR are included only in attention, while T1 and T2 are included in both attention and preference. Thus, some of the product, menu, and context-related variables are excluded

tion to a product.

[17] We acknowledge that this assumption may be restrictive in our setup. It could be the case that a product with a large attention shock may have a greater taste shock. Since we do not have enough data to identify the correlation between the two and also the computational cost of allowing the correlation is significantly high, we cannot incorporate that feature unfortunately. Hence, our estimates would be biased to the extent that the error terms in the attention and those in the utility function are correlated.

Table 3: Variable Definitions

Variable type	Variable Label	Description
M_t	MR	1 if there are any recommended products at occasion t
	NM	# of different products
	NR	# of total recommended products
	NR^2	The square of NR
	NR × T1	Interaction between NR and T1
	T1	Time to the next train
	T2	Time between next and the one after that
	Temp	Temperature
X_{jt}	NS	# of slots product j occupies
	Cat1-10	Product category dummy
	Cat1-10 × Temp	Product category dummy × temperature
	Shape1-3	Product container type: 1 for plastic bottle, 2 for aluminum can, and 3 for glass bottle
	Volume	Product size in ml
	Shape1-3 × Volume	Interaction between container type and volume.
A_{ijt}	PR	Indicator for whether product j is recommended.
	PR × T1	Interaction between PR and T1
	PR × Package	Interaction between PR and Package (can, bottle)
P_j		Price of product j

either from attention or utility. We do so for reducing the number of parameters to estimate based on a priori speculation about the decision process, but not mainly for identification.

We include MR in consumer attention because eye-catching recommendations may attract attention not only to recommended products but also to all products (including non-recommended products) in the assortment as Kawaguchi, Uetake, and Watanabe (2019) find. NM can affect attention either way; to the extent that consumers appreciate product variety, they may pay more attention to all products, but consumers may pay less attention to each product if there are too many products in the feasible set because paying attention to many options could be mentally costly. NR may also have positive or negative impacts on attention; more recommendations increase consumer attention to the point where consumers feel that the recommendations are too aggressive (see, e.g., Chae, Bruno, and Feinberg (forthcoming)). Another important set of variables we examine is the variables related to time pressure. Note that as T1 becomes smaller, the time pressure increases. Because time pressure might impact consumer attention negatively, a decrease in T1 might decrease attention. Although it is not theoretically clear how time pressure affects recommendation effectiveness, Kawaguchi, Uetake, and Watanabe (2019) show that time pressure lowers the effectiveness of a product recommendation. This effect will be captured by interacting T1 with advertisement variables such as PR and NR. We also include T1 and T2 in the utility because several studies in consumer psychology have found that consumer preference would vary under time pressure, such as Payne, Bettman, and Johnson (1988), Edland and Svenson (1993), and Nowlis (1995). For the consumer utility, we interact Temp with product category dummies to see how different weather conditions affect the category of drinks consumers choose.

Variables in X_{jt} vary by product (and market and time). NS is the number of slots that a product occupies at a vending machine. Because consumers might pay more attention to the products that occupy more space in a vending machine, we include this variable in the attention function.[18] Cat, Shape, and Volume are product characteristics and are included for both attention and utility.

Finally, A_{ijt} contains three types of variables. PR is a dummy variable indicating whether product j is recommended to customer i in market t, which affects both consumer attention and preference. As the theory of advertising shows, advertisements may affect consumer preference not only based on the "persuasive role" but also based on the "informative role," signaling through advertising (e.g., Nelson (1974) and Milgrom and Roberts (1986)).[19] We are also interested in how its effectiveness varies by time pressure (PR ×T1) and by product package.

[18]To the extent that NS is correlated with the local demand at each vending machine, the coefficient might be potentially biased.

[19]We cannot separately identify the two effects, because the classification of informational/persuasive role of advertising does not necessarily correspond to the classification of effects through utility/attention.

4 Identification and Estimation

4.1 Identification

In general, it is challenging to identify consideration set models because researchers typically observe only consumers' choices — not their consideration sets or their utilities. Hence, when a product is not chosen, the reason could be either that the product is not preferred or that it is not considered.

One approach to separately identify attention and utility is to use exclusion restrictions (e.g., Goeree (2008); Van Nierop, Bronnenberg, Paap, Wedel, and Franses (2010)).[20] In this approach, advertising, A_{ijt}, is usually excluded from the utility. A specification in which advertisement affects both attention and utility, however, allows us to explore both informational and persuasive roles of advertisement, as in Ching and Ishihara (2010) and Ching and Ishihara (2012). In our setup, recommendations A_{ijt} is included in both attention and utility, but we have menu-variables such as the total number of products sold and whether the vending machine is in the treatment[21], which serve as exclusion restrictions. Those menu-related variables are excluded from the utility function because these variables do not directly affect the utility from chosen products once the products are in the consideration set.

In our setup, we also include context variables, i.e., the proxy variables for time pressure. In general, it is not easy to obtain the variables that capture context variables outside of the laboratory. As we discuss in Section 2.3, we use the data on the train schedule to measure time pressure, T1 and T2. We treat these variables exogenous as it is unlikely that the train schedule is set depending on the consumer taste for beverages.[22]

Advertising in observational data is often correlated with unobserved heterogeneity. There-

[20] See also the discussion in Abaluck and Adams (2018), who also make a similar point about the existing literature.

[21] Product differentiation is mostly horizontal, and hence, compromising effects (see, e.g., Simonson (1989)) or decoy effects (see, e.g., Huber, Payne, and Puto (1982)), which indicate that consumer preference may be shifted by the menu variables, may not play an important role in our setup.

[22] Kawaguchi, Uetake, and Watanabe (2019) checks if there is any selection of customers by the degree of time pressure by a series of robustness checks. They confirm that the selection effect is limited in our context.

fore, the effect of advertising is identified with the variation in instrumental variables in the literature (see, e.g., Goeree (2008)).[23] In our case, it is identified from the exogenous variation in product recommendations that the company created in the experiment. Since the allocation of treatment conditions across different days and times of the day is made exogenously, there is no reason to think that advertisement is correlated with unobserved consumer heterogeneity in our setup.[24]

The identification of the price coefficient depends on a limited variation of data. In our data set, prices conditional on the product's package and size do not vary much,[25] which may limit the degree to which the price coefficient is recovered. Conditional on package and size, additional variations in the price mainly come from the package of size 280ml. Because this size of the package was relatively new, the company experimentally set different prices across products.

In order to identify the random coefficients, we exploit the variation in the available products and that in product recommendations. Also, the variation in the price may have some identification power, although very limited. [26] In our data set, however, the variation in prices, advertisements, and product availability may not be rich enough to allow nonparametric identification. Therefore, the identification of the random coefficients relies on the parametric assumptions we make. In particular, we do not allow any correlation in random coefficients conditional on observables, i.e., the off-diagonal cells of Σ and Ω are assumed to be zero. This assumption reduces the number of parameters to be estimated.

Lastly, we assume that attention probabilities are independent, *conditional on the observ-*

[23]Goeree (2008) uses attributes, product cost shifters, and advertising cost shifters of all other products to construct optimal instruments. Also, Goeree (2008) tests the exogeneity of exposure to advertisements using the cost of access to various media, and finds that this endogeneity is not an issue.

[24]In Kawaguchi, Uetake, and Watanabe (2019), they confirm that the allocation is not correlated with key variables such as demographics and sales patterns.

[25]When we regress prices on package dummies and size dummies, the model explains more than 90% of the variation in prices.

[26]To identify the random coefficients, it is necessary to have variation in the consumer-choice-level characteristics (see, e.g., Berry and Haile (2011); Fox, Kim, Ryan, and Bajari (2012)). For example, we have such variations when the price and advertisements for a particular product vary by time and the vending machine or when product availability varies by time and the vending machine.

ables.[27] Similarly, we assume that the random shocks for attention and the ones for preference are independent of each other. Although this assumption is maintained in the existing papers, this is indeed a strong assumption. Our data do not have rich variations to identify it.

4.2 Simulated Maximum Likelihood Estimation

We estimate the consideration set model discussed in Section 3 with a simulated maximum likelihood estimator, which uses simulated purchase probabilities to calculate the likelihood of individual purchase decisions.

Given the realization of random effects, $\{\zeta_j, \xi_j\}_{j\in\mathscr{J}}$, $\{\alpha_{ij}, \beta_{ij}\}_{i\in\mathscr{I}, j\in\mathscr{J}}$, $\{\boldsymbol{\nu}_i, \boldsymbol{v}_i\}_{i\in\mathscr{I}}$ in the model discussed in Section 3, the choice probability of products can be decomposed into the attention probabilities and the choice probabilities conditional on consideration sets, as follows:

$$\begin{aligned} p_{ijt} &\equiv \mathbb{P}\{D_{ijt}=1\} \\ &= \sum_{\mathscr{C}_{it}\subset\mathscr{J}_t} \mathbb{P}\{U^*_{ijt} \geq \max_{k\in\mathscr{C}_{it}}\{U^*_{ikt}\}\} \prod_{l\in\mathscr{C}_{it}} \mathbb{P}\{C_{ilt}=1\} \prod_{m\notin\mathscr{C}_{it}} \mathbb{P}\{C_{imt}=0\} \\ &\equiv \sum_{\mathscr{C}_{it}\subset\mathscr{J}_t} \pi_{ijt}(\mathscr{C}_{it}) \prod_{l\in\mathscr{C}_{it}} \gamma_{ilt} \prod_{m\notin\mathscr{C}_{it}} (1-\gamma_{imt}), \end{aligned} \tag{4.1}$$

where $\gamma_{ijt} = \mathbb{P}\{C_{ijt}=1\}$ and $\pi_{ijt}(\mathscr{C}_{it}) = \mathbb{P}\{U^*_{ijt} \geq \max_{k\in\mathscr{C}_{it}}\{U^*_{ikt}\}\}$.

Then, the likelihood function is

$$\int \prod_i \prod_j \prod_t p_{ijt}^{D_{ijt}} \, dF(\{\zeta_j, \xi_j\}_{j\in\mathscr{J}}, \{\alpha_{ij}, \beta_{ij}\}_{i\in\mathscr{I}, j\in\mathscr{J}}, \{\boldsymbol{\nu}_i, \boldsymbol{v}_i\}_{i\in\mathscr{I}}). \tag{4.2}$$

Note that equation (4.1) involves a large number of terms as it sums up $2^{|\mathscr{J}_t|}-1$ terms, which is the number of potential consideration sets. It is not computationally straightforward

[27] Although we acknowledge that the independence assumption helps identification with the actual variation in the data, the assumption is not necessary in theory. If there exists data on feasible sets that contain the set of all possible combinations of goods, we can potentially identify the correlation of attention probabilities flexibly. We provide a more formal discussion in the online appendix.

to accurately approximate this choice probability with a computationally feasible number of simulation draws of consideration sets.[28] Instead, we compute the choice probability with a method recently proposed by Lee (2019), which does not require one to simulate consideration sets. Lee (2019) theoretically shows that the choice probabilities based on his approximation algorithm converge to the true choice probabilities much faster than the simulation-based strategy.[29]

In particular, Lee (2019) shows that the choice probability can be *analytically* written as follows:

$$p_{ijt} = \gamma_{ijt} \int \prod_{l \in \mathscr{J}_t \setminus \{j\}} [1 - \gamma_{ilt} + \gamma_{ilt} P(U^* | U_{ilt})] p(U^* | U_{ijt}) dU^*, \tag{4.3}$$

where $P(|U_{ijt})$ is the conditional probability distribution of U^*_{ijt} given U_{ijt} and $p(U^*|U_{ijt}) = \frac{dP(U^*|U_{ijt})}{dU^*}$. Note that γ_{ijt}, γ_{ilt}, $P(U^*|U_{ilt})$, and $p(U^*|U_{ijt})$ can be evaluated as a function of U^* given parameters. A benefit of this transformation is that the choice probability, p_{ijt}, can be described as a univariate integration problem, which is much more accurately done than simulating a large number of consideration sets. Another benefit is that the integrand involves only $|\mathscr{J}_t|$ terms, which is computationally faster than simulating $2^{|\mathscr{J}_t|}$ terms.

The intuition behind this equation is straightforward. Suppose that there are only two alternatives in the choice set, j, l. Note that alternative j is chosen over alternative l if 1) alternative j is considered, and 2) alternative l is not considered or 3) alternative l is also considered and alternative j's utility is higher than alternative l's utility. The probability event 1 occurs is γ_j. The probability event 2 occurs is $1-\gamma_l$. Suppose that the utility of alternative j is u and the distribution of alternative l's utility is $P_l(\cdot)$. Then, the probability event 3 happens is $\gamma_l P_l(u)$. In summary, the probability of (event 1 and (event 2 or event 3)) occurs is $\gamma_j[1-\gamma_l+\gamma_l P_l(u)]$. This

[28]Existing papers either assume that certain products are in the consideration set based on, for example, past purchase or usage, and sum over all possible consideration sets or simulate choice sets using importance sampler as in Goeree (2008), which has some well-defined properties.

[29]Although Lee (2019) does not show the simulated maximum likelihood estimator based on his approximation of the choice probabilities is consistent and asymptotically normal, it is straightforward to show the asymptotic properties of the simulated maximum likelihood estimator based on Lee (2019)'s method by applying the standard proof of consistency and asymptotic normality of the simulated maximum likelihood estimator such as Lee (1995).

expression can be evaluated at any u as a function. Thus, by integrating this function with respect to u with its probability density, $p_j(\cdot)$, we can obtain the choice probability of alternative j over alternative l. Thus, the evaluation of choice probability can be achieved by integrating a scalar variable u. In Lee (2019), this integration is done by quadrature. The intuition is extended to the case with arbitrary many alternatives in the choice set.

We use the algorithm proposed by Lee (2019) to approximate equation (4.3). We then compute the likelihood function (equation (4.2)) by simulating random coefficients, $\{\zeta_j, \xi_j\}_{j\in\mathcal{J}}$, $\{\alpha_{ij}, \beta_{ij}\}_{i\in\mathcal{I}, j\in\mathcal{J}}$, $\{\boldsymbol{\nu}_i, \boldsymbol{v}_i\}_{i\in\mathcal{I}}$.[30] We will explain the estimation algorithm in the online appendix in detail. Lastly, since our estimation strategy does not have to simulate consideration sets, we do not have to correct standard errors due to high variance of simulated consideration sets. the

5 Estimation Results

5.1 Parameter Estimates

We report the estimated coefficients of the consideration set model in Table 4.[31] The first column shows the estimated coefficients of the attention function, and the second column shows those of the utility function. The heterogeneity in the effectiveness of product recommendations across consumer segments is summarized in Table 5.

The coefficient on product recommendation (*PR*) is positive for both attention and utility. Hence, recommendations not only increase the attention to the recommended product but also increase the utility, which implies that the product recommendation in our setup has both informational and persuasive roles, which should be combinations of effects through utility and attention. The coefficient on recommendation spillover (*MR*) in attention is positive, and

[30]For $\{\zeta_j, \xi_j\}_{j\in\mathcal{J}}$ and $\{\boldsymbol{\nu}_i, \boldsymbol{v}_i\}_{i\in\mathcal{I}}$, we take 100 random draws for each product. For $\{\alpha_{ij}, \beta_{ij}\}_{i\in\mathcal{I}, j\in\mathcal{J}}$, we take 100 random draws for each product and customer. We also try 500 random draws but the results barely change.

[31]In the online appendix, we also report the fit of the consideration set model. Also, we report the parameter estimates and the fit of a random-coefficient multinomial logit model to compare the results. We find that the consideration set model fits better to the data. We thank Associate Editor for suggesting this analysis.

Table 4: The Parameter Estimates

Attention	*Estimate*	*95% C.I.*	*Utility*	*Estimate*	*95% C.I.*
PR	0.093	[0.090, 0.104]	PR	0.047	[0.045, 0.054]
MR	0.053	[0.052, 0.060]	NS	0.202	[0.176, 0.242]
NR	0.192	[0.189, 0.195]	price	-1.072	[-1.110, -1.068]
NR^2	-0.321	[-0.322, -0.319]	PR × T1	0.019	[0.017, 0.023]
PR × T1	0.106	[0.105, 0.108]	T1	-0.063	[-0.068, -0.060]
NR × T1	0.055	[0.049, 0.058]	T2	-0.034	[-0.040, -0.032]
PR × NR	-0.123	[-0.127, -0.121]	Green tea	0.096	[0.090, 0.098]
NS	0.061	[0.058, 0.063]	Other tea	0.097	[0.090, 0.100]
NM	-0.005	[-0.006, -0.004]	Water	0.023	[0.018, 0.032]
T1	0.131	[0.065, 0.173]	Sport	0.099	[0.090, 0.103]
T2	0.006	[0.005, 0.010]	Canned coffee	0.080	[0.074, 0.083]
PR × Can	0.016	[0.015, 0.029]	Bottled coffee	0.035	[0.033, 0.037]
PR × Glass bottle	0.017	[0.015, 0.030]	Black tea	0.127	[0.125, 0.131]
MR × Can	0.021	[0.015, 0.028]	Carbonated	-0.003	[-0.010, 0.011]
MR × Glass bottle	0.008	[0.005, 0.021]	Fruit	0.106	[0.098, 0.109]
Green tea	-0.046	[-0.047, -0.044]	Healthy	0.040	[0.038, 0.047]
Other tea	0.027	[0.026, 0.029]	Other drink	0.137	[0.135, 0.150]
Water	-0.084	[-0.083, -0.076]	Green tea × Temp	-0.088	[-0.091, -0.087]
Sport	0.049	[0.045, 0.052]	Other tea × Temp	-0.093	[-0.095, -0.092]
Canned coffee	-0.145	[-0.146, -0.143]	Water × Temp	-0.084	[-0.086, -0.074]
Bottled coffee	0.155	[0.157, 0.164]	Sport × Temp	-0.085	[-0.109, -0.082]
Black tea	0.061	[0.059, 0.069]	Canned coffee × Temp	-0.097	[-0.100, -0.095]
Carbonated	0.109	[0.107, 0.120]	Bottled coffee × Temp	-0.095	[-0.098, -0.093]
Fruit	0.050	[0.047, 0.052]	Black tea × Temp	-0.115	[-0.118, -0.114]
Healthy	0.100	[0.097, 0.107]	Carbonated × Temp	-0.107	[-0.109, -0.097]
Other drink	0.055	[0.052, 0.068]	Fruit × Temp	-0.091	[-0.094, -0.089]
Can	-0.050	[-0.052, -0.048]	Healthy × Temp	-0.100	[-0.101, -0.099]
Glass bottle	0.136	[0.138, 0.152]	Other drink × Temp	-0.128	[-0.132, -0.124]
Plastic bottle × Volume	-0.023	[-0.028, -0.020]	Can	0.160	[0.153, 0.163]
Can × Volume	0.614	[0.612, 0.619]	Glass bottle	0.023	[0.020, 0.038]
Glass bottle × Volume	0.236	[0.236, 0.251]	Plastic bottle × Volume	0.278	[0.269, 0.281]
σ	0.671	[0.668, 0.684]	Can × Volume	0.085	[0.084, 0.089]
S.D. α_{ij}	0.025	[0.025, 0.039]	Glass bottle × Volume	-0.275	[-0.278, -0.262]
S.D. ζ_{jt}	0.465	[0.462, 0.479]	ω	0.031	[0.026, 0.042]
			S.D. β_{ij}	-0.044	[-0.065, -0.01]
			S.D. ξ_{jt}	0.188	[0.181, 0.19]

Note: Confidence intervals are constructed based on 100 bootstrapping samples. We omit some estimates from the table to save the space.

hence when the machine recommends some products, consumers pay more attention to each product in the machine, including non-recommended ones. The effect of PR and MR is also heterogeneous depending on the package as different packages attract different attention. The total number of recommended products, *NR*, has a positive first-order effect on attention, but the second-order effect is negative. Thus, consumers pay less attention to each product if there are too many recommended products. This finding may reflect the consumer's limited resource for attention.

The coefficient on the number of products in the assortment (*NM*) is negative for attention, which suggests that consumers pay less attention to each product if there are more products in the vending machine. Thus, cognitive constraint seems to dominate the love-for-variety effect with attention. The effect of the number of columns that each product occupies (*NS*) is positive for both attention and utility. Hence, consumers become more aware of products that occupy more physical space in the vending machine, and they are more attracted to buying them.

Regarding time pressure, we find that the coefficients on time pressure (*T1* and *T2*) are positive. Hence, consumers have lower attention under higher time pressure in general. In contrast, the coefficients on *T1* and *T2* are negative in utility, i.e., consumers are *more* likely to buy some products because they feel more time pressure.[32] A possible interpretation of these coefficients is as follows: time pressure deprives a consumer of her cognitive resources, leading her not to pay enough attention to the products, to not carefully examine the products, and to make an impulse to buy.

More importantly, the effectiveness of product recommendations is also affected by time pressure. We find that the coefficient on *PR*× *T1* is positive for both attention and utility. Hence, consumers pay less attention to recommended products when they are under time pressure, and they are less likely to buy the recommended products even when they are considered. This finding is consistent with Kawaguchi, Uetake, and Watanabe (2019) and suggests

[32]Although the signs are opposite, the overall effect of time pressure is positive. This overall positive effect is also shown in Kawaguchi, Uetake, and Watanabe (2019)

Table 5: The Parameter Estimates for Customer Heterogeneity

	Gender	*PR in Attention*		*MR in Attention*		*PR in Utility*	
		Estimate	*95% C.I.*	*Estimate*	*95% C.I.*	*Estimate*	*95% C.I.*
Junior							
	Female	0.237	[0.234, 0.259]	0.114	[0.112, 0.123]	0.126	[0.123, 0.137]
	Male	0.250	[0.245, 0.275]	0.120	[0.115, 0.137]	0.127	[0.122, 0.147]
Senior							
	Female	0.093	[0.090, 0.104]	0.053	[0.052, 0.060]	0.047	[0.045, 0.054]
	Male	0.106	[0.101, 0.120]	0.059	[0.055, 0.074]	0.048	[0.044, 0.065]
Old							
	Female	0.167	[0.162, 0.185]	0.129	[0.127, 0.138]	0.122	[0.118, 0.133]
	Male	0.180	[0.173, 0.201]	0.135	[0.130, 0.152]	0.123	[0.118, 0.142]

Note: Confidence intervals are constructed by 100 bootstrapping samples. We omit some estimates from the table to save the space. Junior is a dummy variable for customers whose age is younger than 30, senior is between the ages of 30 and 50, and old is 50 or older.

that consumers do not want to be told what to buy when they are in a hurry.

Table 5 reports the heterogeneity in the coefficients of product recommendations on attention and preference ($\mathbf{\Pi}^{\alpha}$ and $\mathbf{\Pi}^{\beta}$ in the random coefficients). Our model includes the gender and age of customers as categorical variables in the random coefficients. We find significant heterogeneity in the effects of product recommendations across different demographic groups. An interesting finding is that senior consumers (age 30-50) tend to pay less attention to recommended products and feel smaller utility from those products. We also find female customers pay less attention to recommended products and find less utility. One of the interesting managerial questions that arise here is which of the traditional attribute-based targeting and context-based recommendations has a higher impact and through which channel (attention vs. utility). We address this question in the subsequent section.

5.2 Advertisement Elasticities

To quantify the effects of recommendations on attention and utility under time pressure, we calculate advertising elasticities on attention, utility, and overall purchase incidence under different levels of time pressure.

Table 6: Mean Elasticity of Choice Probability to Recommendation

	Total	*Own*	*Cross*
Full	0.032	0.187	0.027
Attention channel only	0.029	0.090	0.027
Utility channel only	0.003	0.086	0.000

Note: The covariates are set at actual values except for variables related to product recommendation. We dropped the top 2.5% own elasticities because they are unstable.

Because in our setting advertising, PR, is a binary variable, we define *elasticities* as follows. For each purchase occasion t, we set $PR = NR = MR = 0$ in both attention and utility for all products $j \in J_t$ and calculate the choice probability of product j as s^0_{jt}. We then set $PR_{jt} = NR_t = MR_t = 1$ for each product j. We denote the choice probability of product k where only product j is recommended by s^{1j}_{kt}. Based on these notations, we define the *own elasticity* of the choice probability by $\frac{s^{1j}_{jt} - s^0_{jt}}{s^0_{jt}}$ and the *cross elasticity* by $\frac{\sum_{k \neq j,0}(s^{1j}_{kt} - s^0_{kt})}{\sum_{k \neq j,0} s^0_{kt}}$. Note that we sum up choice probabilities of non-recommended products k here. Finally, we define the *total elasticity* by $\frac{\sum_{k \neq 0}(s^{1j}_{kt} - s^0_{kt})}{\sum_{k \neq 0} s^0_{kt}}$, where the summation is taken over all products J_t. We compute these elasticities for each purchase occasion and for each product and report the mean in Table 6.[33]

Under Actual Time Pressure In the first row of Table 6, we present the mean of own, cross, and total elasticities as defined above using the full model *given the level of time pressure shown in the data*. We find that, on average, recommending a product increases the total vending machine sales by 3.2%, the sales of *recommended* products by 18.7%, and the sales of *non-recommended products* by 2.7%.[34]

We then decompose the elasticities in the first row into the effect through attention and the one through utility. In the second row, we calculate three types of elasticities when the coefficients on PR, NR, and MR in the utility are set to 0, i.e., recommendations affect only

[33] Remember that the elasticities we calculate here can be considered as the upper bounds because it includes the effect of changing NR and MR from 0 to 1.

[34] These elasticities are moderate compared to what some existing papers have found. For example, Breugelmans and Campo (2011) find that in-store displays can increase the sales of displayed products by up to 100%.

Table 7: Mean Elasticity of Choice Probability to Recommendation by Time Pressure

(a) Full

Time	*0*	*5*	*10*
all	0.029	0.037	0.02
own	0.156	0.183	0.196
cross	0.024	0.032	0.014

(b) Attention channel only

0	*5*	*10*
0.0266	0.0324	0.0127
0.0867	0.0548	0.0141
0.0244	0.0316	0.0138

(c) Utility channel only

0	*5*	*10*
0.002	0.004	0.007
0.063	0.123	0.205
0	0	0

Note: The setting is the same as those in Table 6 except that the time to next train (minutes) is set at 0, 5, and 10 minutes from the time of a purchase. Because some cases show extremely high elasticity due to the small baseline choice probabilities, we dropped cases with top 2.5% own elasticities.

attention. Similarly, in the third row, we calculate three elasticities when the coefficients on PR, NR, and MR in the attention are set to be 0, i.e., recommendations affect only utility.

The results show that the mean of the total elasticity is 0.03 when recommendations affect only attention and 0.003 when recommendations affect only utility. Hence, the recommendations affect *total vending machine sales* more through the attention channel than through the utility channel. This is because NR and MR are included in attention but not in utility. By contrast, own elasticity is 0.09 when recommendations only affect attention and 0.086 when recommendations only affect utility. Thus, the effect of recommendations on *a recommended product* works slightly more through the attention channel.[35]

Under varying degrees of time pressure We also calculate the three types of elasticities under varying degrees of time pressure, i.e., the time to the next train is 0 minutes (high time pressure), 5 minutes (medium time pressure), and 10 minutes (low time pressure). Note that it is a priori unclear whether elasticities increase or decrease with time pressure because both the numerator and denominator of the elasticities vary by time pressure.

In Panel (a) of Table 7, we present the mean of three elasticities under three different conditions of time pressure based on the full model. We find that total elasticity, own elasticity increases, and cross elasticity have an inverse-U relationship.

[35]In the online appendix, we also report the effects of recommendations on the probability that a product is included in the consideration set.

Panels (b) and (c) present the mean of elasticities when recommendations do not affect utility (Panel (b)) and do not affect attention (Panel (c)), respectively. In Panel (b), we find that total elasticities and cross elasticities have inverted-U shape, while own elasticities increase as time pressure increases. We find different patterns in Panel (c); both overall and own elasticities decrease as time pressure increases due to a lack of a spillover effect. These findings also motivate our counterfactual simulations to optimize recommendations based on the degree of time pressure.

6 Optimizing Recommendations under Time Pressure

6.1 Optimizing the Number of Recommendations under Time Pressure

Using the estimates of the consideration set model, we conduct several counterfactual simulations to derive managerial implications. In particular, we begin by considering *how many* products the company should recommend under time pressure.[36] Although recommendations in general increase the attention and utility of recommended products, recommending too many products can backfire because customers may start paying less attention to each product and when customers feel time pressure, they pay less attention to each product, and the effect of recommendations weakens. Hence, finding the optimal number of recommendations is an empirical question that we cannot tell the effect ex-ante.[37] Moreover, a reduced-form analysis by Kawaguchi, Uetake, and Watanabe (2019) cannot answer this counterfactual question. In this exercise, we study how machine-level sales change by the number of recommended products and their sensitivity to time pressure.

To avoid a computationally intractable combinatorial optimization problem, we do not consider the optimal combination of recommended products here. Instead, we randomly

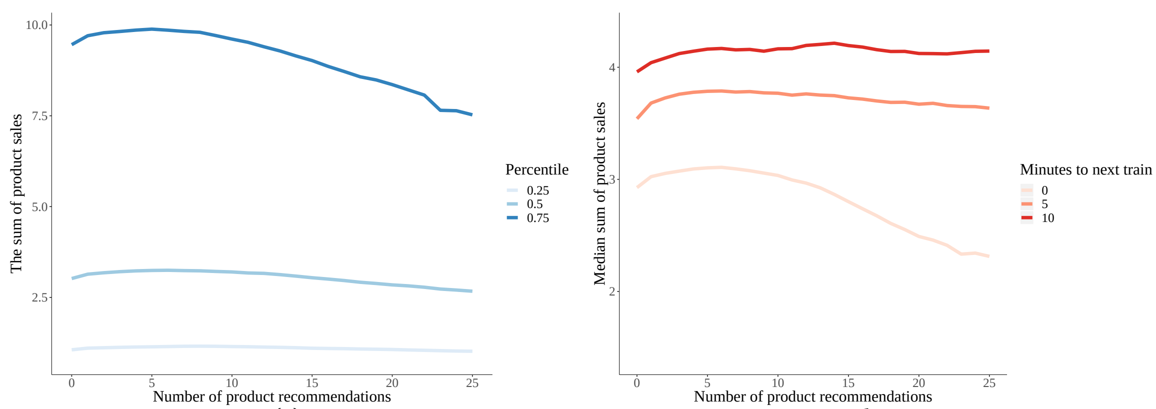
[36]In the online appendix, we show the results of the counterfactual simulations based on the multinomial logit model as a comparison. We find that the counterfactual recommendation systems based on the logit model tend to recommend too many products.

[37]According to the company, the gross margins are more or less the same across products. Therefore, the company sets the number of sales units, not the sales value or profit, as its objective variable.

select which products to recommend, given how many products to recommend. More precisely, we first randomly order available products ($j = \{1, ..., J_t\}$) for each occasion, where J_t is the number of available products in vending machine k at occasion t. We then recommend products from 1 to K, where K is the number of products to recommend. We repeat this procedure from $K = 1$ to $K = J_t$ for each vending machine and occasion. Later, we consider which products to recommend in a certain simplified way.

We first present Figure 3 to present the relationship between the sales and the number of recommendations. We present this figure as their relationship and how the time pressure affects it is not straightforward and it depend on the parameter estimates.

Figure 3: Machine-level Sales by the Number of Recommendations



(a) (b)

Note: For each purchase occasion, begin with a setting in which there is no product recommendation and then randomly pick up a product to recommend. Continue this process until all products in the vending machine are recommended. At each number of product recommendations, compute the machine-level sales units of products. Then, compute percentiles of the sum of inside product share for each number of recommendations (NR) across purchase occasions.

Optimal Recommendations Under Actual Time Pressure First, we show that machine-level sales vary by the number of recommendations *given the actual level of time pressure.* In Panel (a) of Figure 3, we plot the sales units per machine on the vertical axis against the number of recommendations per machine on the horizontal axis. The top curve shows the 75th percentile of the distribution of sales among vending machines, the middle 50th percentile, and the bottom 25th percentile. The figure shows that the optimal number of recommendations is about 7 at the median and sales decrease if there are too many recommended products.

Optimal Recommendations Under Counterfactual Time Pressure Next, we examine how time pressure affects the optimal number of recommendations by changing the degree of time pressure from what we observe in the data to counterfactual levels. Panel (b) of Figure 3 plots the median per-machine sales against the number of recommendations under three different levels of time pressure. We find that it is optimal to recommend more when there is less time pressure.

Optimality of Contextual Recommendation Now, we consider how much context-based recommendations can improve sales for the company. We do so by comparing the following three recommendation policies with the actual one. The first policy is the *optimal uniform policy*, which uniformly recommends the same *number* of products for all vending machines to maximize the sum of the sales from all vending machines during the time period. Hence, the uniform policy does not change recommendations based on context or attribute. Second, we consider the *optimal segment-based targeting policy*, in which the number of recommendations is chosen so as to maximize sales for *each consumer segment (gender* × *age class)* level. The segment-based targeting policy is similar to a traditional targeting strategy that is based only on observable customer characteristics. The third policy is the *contextual policy*, which recommends the optimal number of products for each vending machine and for each point of time. Note that the optimal contextual policy adjusts how many products to recommend depending on the degree of time pressure at each purchase occasion. Because there is no recommendation shown in the control group, the comparison is conducted based on the purchase occasions in the treatment group. Our baseline simulations randomly choose which products to recommend from the set of available products at each vending machine given a number of recommended products.

Table 8 reports the results, which include the percentage change in the sales in the first column, and the mean and standard deviation of the optimal number of recommendations in the second and third columns.

We find that the uniform policy can improve sales by 0.82% relative to the current policy, and it recommends 7 products. The segment-based targeting policy slightly increases the sales volume but only by an additional 0.06 percentage points relative to the uniform policy. Also, the targeting policy recommends slightly more products on average. Lastly, the contextual policy can achieve even higher performance with 1.79% of sales increase, and the contextual policy recommends 11 products on average. Therefore, we find that the contextual policy outperforms the traditional attribute-based recommendation policy in our setup.

Table 8: Sales and Optimal Recommendations

	Sales Change	*NR*	
		Mean	*S.D.*
Actual	-	4.44	1.74
Uniform	0.82%	7.00	-
Target	0.88%	7.50	1.22
Contextual	1.79%	11.13	7.09

Note: The table reports the sales change under the optimal number of recommendations for each algorithm against the actual recommendations in the first column, and the optimal number of recommendations in the second column.

6.2 Optimizing Products to Recommend under Time Pressure

In the previous analysis, we focus on the *number* of recommended products, but we do not consider *which* products to recommend, because it is computationally infeasible to search across all possible combinations. We now consider how much sales improve by taking that into account. Since it is computationally hard to compare all possible assortment combinations, we do so by implementing the following heuristic *greedy algorithm*. Hence, one can interpret our results as a lower bound of more sophisticated assortment optimization algorithm.[38] For

[38] In Operations Research, there are several algorithms to optimize assortment given a demand system such as multinomial logit and nested logit models (see, e.g., Davis, Gallego, and Topaloglu (2014) and Feldman and

each vending machine t, let $\tilde{J}_t$ denote the set of recommended products and $\Pi_t(J)$ denote the sales of vending machine t when the set of products $J \subseteq J_t$ is being recommended. Then, the greedy algorithm works as follows.

1. Set $\tilde{J}_t = \emptyset$.

2. Update $\tilde{J}_t = \tilde{J}_t \cup \{j\}$ such that $j = \arg\max_{j \in J_t \setminus \tilde{J}_t} \Pi_t(\tilde{J}_t \cup \{j\})$.

3. Stop if $\Pi_t(\tilde{J}_t \cup \{j\}) < \Pi_t(\tilde{J}_t)$ for all $j \in J_t \setminus \tilde{J}_t$.

Simply speaking, the greedy algorithm keeps adding a product that increases the total sales of the vending machine given the set of recommended products in the previous round. We implement the greedy algorithm for all vending machines in the data and for all four cases (Actual, Uniform, Target, and Contextual) and report the results in Table 9.

Table 9: Sales and Optimal Recommendation by Greedy Algorithm

	Sales	*NR*	
	Change	*Mean*	*S.D.*
Actual	-	4.44	1.74
Uniform	3.13%	8.00	-
Target	3.15%	8.50	0.84
Contextual	3.71%	11.05	5.63

Note: The table reports the sales change under the optimal number of recommendations for each algorithm against the actual recommendations in the first column, and the optimal number of recommendations in the second column.

We find that the greedy algorithm generates higher sales than our baseline case, but the results are qualitatively the same. Uniform and Target recommendations increase the sales by a similar percentage, while Contextual recommendations increase total sales by about 3.71%. Our counterfactual simulations provide only the lower bound of the potential impacts of context-

Topaloglu (2018)). It is well-known that the assortment optimization problem with a random coefficient multinomial logit model is in general NP-hard. So far, there is no algorithm for optimizing assortment with a consideration set model.

based marketing. If the company can optimize recommendations, total sales would increase even further.

To compare the size of these effects with other marketing tools, we conduct simulation with price discounts. By this simulation, we find that a 5% price discount of all products boosts sales volume by 5.5%. However, because the price decreases, the revenue increases only by 0.2% ($1.055 \times 0.95 = 1.00225$). Cutting the price of all products by 10% increases the sales volume by 11.6%. This results in 0.4% increases in the revenue ($1.116 \times 0.9 = 1.0044$). Therefore, using product recommendations and maximizing its benefits by making them context-based can be more attractive for a manager than price discounts.

7 Conclusion

This paper studies the effect of time pressure on consumer attention and utility and examines the optimization of product recommendation systems that adopt contextual factors. To answer the research questions, we take advantage of our unique setup of consumer beverage purchases from the vending machines on train station platforms, which allow us to measure the degree of time pressure that consumers feel from the train schedule information.

We build a structural model of the consideration set formation, in which time pressure and product recommendations can affect consumer attention and utility, and consumer attention can depend on "menu"-related variables such as the number of recommended products and the number of unique products in the assortment. The estimation results reveal several findings. First, we find that time pressure negatively affects consumer attention but positively affects utility. Second, product recommendations increase both attention and utility, but time pressure moderates the effectiveness of recommendations. Third, the number of total recommendations increases the attention level in general, but in decreasing order. Finally, there is significant heterogeneity in the effects of recommendations across customer segments.

Using the estimates, we conduct a series of counterfactual simulations to investigate the optimal design of the context-based recommendation system. In particular, we compare the context-based recommendation, which changes the number of recommended products based on the degree of time pressure for each machine with the customer attribute-based targeting as in a traditional targeting strategy. We find that the context-based recommendation outperforms the attribute-based targeting in our context. Moreover, the context-based recommendation increases the revenue more than a price cut at the margin.

Lastly, there are some interesting and important avenues for future research. First, we do not incorporate any dynamics into the model and are agnostic about the evolution of the consideration set over time. It is important to study how advertisements have a long-run effect on consumer preference and attention. Second, assortment design is very attractive research. In this paper, we do not investigate the optimal combination of products for recommendations, but it would be an interesting strategy to consider. We leave those topics for potential future research.

References

ABALUCK, J., AND A. ADAMS (2018): "What Do Consumers Consider Before They Choose? Identification from Asymmetric Demand Responses," *mimeo.*

ACKERBERG, D. (2001): "Empirically Distinguishing Informative and Prestige Effects of Advertising," *RAND Journal of Economics*, 32, 316–33.

ALLENBY, G., AND J. GINTER (1995): "The Effects of in-store displays and feature advertising on consideration sets," *International Journal of Research in Marketing*, 12, 67–80.

ANAND, B., AND R. SHACHAR (2011): "Advertising, the matchmaker," *RAND Journal of Economics*, 42(2), 205–45.

ANUPINDI, R., M. DADA, AND S. GUPTA (1998): "Estimation of Consumer Demand with Stock-Out Based Substitution: An Application to Vending Machine Products," *Marketing Science*, 17(4), 406–423.

BAGWELL, K. (2005): *Handbook of Industrial Organization*chap. The Economic Analysis of Advertising, pp. 1701–1844. North-Holland, Amsterdam.

BARROSO, A., AND G. LLOBET (2012): "Advertising and Consumer Awareness of New, Differentiated Products," *Journal of Marketing Research*, 49(6), 773–792.

BECKER, G., AND K. MURPHY (1993): "A Simple Theory of Advertising as a Good or Bad," *Quarterly Journal of Economics*, 108, 942–64.

BERRY, S., AND P. HAILE (2011): "Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers," .

BREUGELMANS, E., AND K. CAMPO (2011): "Effectiveness of In-Store Displays in a Virtual Store Environment," *Journal of Retailing*, 87(1), 75–89.

BRIESCH, R., P. CHINTAGUNTA, AND R. MATZKIN (2010): "Nonparametric Discrete Choice Models with Unobserved Heterogeneity," *Journal of Business and Economic Statistics*, 28, 291–307.

BRONNENBERG, B., AND Y. HUANG (forthcoming): "Pennies for Your Thoughts: Costly Product Consideration and Purchase Quantity Thresholds," *Marketing Science*.

CHAE, I., H. BRUNO, AND F. FEINBERG (forthcoming): "Wearout or Weariness? Measuring Potential Negative Consequences of Online Ad Volume and Placement," *Journal of Marketing Research*.

CHANDON, P., H. WESLEY, E. BRADLOW, AND S. YOUNG (2009): "Does in-store marketing work? Effects of the number and position of shelf facings on brand attention and evaluation at the point of purchase," *Journal of Marketing*, 73, 1–17.

CHING, A., T. ERDEM, AND M. KEANE (2009): "The Price Consideration Model of Brand Choice," *Journal of Applied Econometrics*, 24, 393–420.

CHING, A., T. ERDEM, AND M. KEANE (2017): *Handbook of Marketing Decision Models*chap. Empirical Models of Learning Dynamics: A Survey of Recent Developments, pp. 223–257. Springer.

CHING, A., AND F. HAYASHI (2010): “Payment card rewards programs and consumer payment choice,” *Journal of Banking and Finance*, 34(8), 1773–87.

CHING, A., AND M. ISHIHARA (2010): “The effects of detailing on prescribing decisions under quality uncertainty,” *Quantitative Marketing and Economics*, 8, 123–165.

——— (2012): “Measuring the Informative and Persuasive Roles of Detailing on Prescribing Decisions,” *Management Science*, 58, 1374–1387.

CHING, A. T., T. ERDEM, AND M. P. KEANE (2013): “Learning Models: An Assessment of Progress, Challenges, and New Developments,” *Marketing Science*, 32(6), 913–938.

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

CONLON, C., AND J. MORTIMER (2019): “Efficiency and Foreclosure Effects of Vertical Rebates: Empirical Evidence,” *mimeo*.

CONLON, C. T., AND J. H. MORTIMER (2013): “Demand Estimation under Incomplete Product Availability,” *American Economic Journal: Microeconomics*, 5, 1–30.

DAVIS, J. M., G. GALLEGO, AND H. TOPALOGLU (2014): “Assortment Optimization Under Variants of the Nested Logit Model,” *Operations Research*, 62(2), 250–273.

DEHMANY, K., AND T. OTTER (2014): “Utility and Attention – A Structural Model of Consideration,” *mimeo*.

DHAR, R., AND S. M. NOWLIS (1999): “The Effect of Time Pressure on Consumer Choice Deferral,” *Journal of Consumer Research*, 25, 369–384.

DRAGANSKA, M., AND D. KLAPPER (2011): “Choice Set Heterogeneity and the Role of Advertising: An Analysis with Micro and Macro Data,” *Journal of Marketing Research*, 48, 653–69.

EDLAND, A., AND O. SVENSON (1993): *Judgment and Decision Making Under Time Pressure*pp. 27–40. Springer US, Boston, MA.

FELDMAN, J., AND H. TOPALOGLU (2018): “Capacitated Assortment Optimization Under the Multinomial Logit Model with Nested Consideration Sets,” *Operations Research*, 66(2), 380–391.

FOX, J., K. KIM, S. RYAN, AND P. BAJARI (2012): “The Random Coefficients Logit Model Is Identified,” *Journal of Econometrics*, 166, 204–12.

GOEREE, M. (2008): “Limited Information and Advertising in the US Personal Computer Industry,” *Econometrica*, 76, 1017–74.

GROSSMAN, G., AND C. SHAPIRO (1984): “Informative Advertising with Differentiated Products,” *Review of Economic Studies*, 51, 63–81.

HONKA, E., A. HORTACSU, AND M. A. VITORINO (2017): “Advertising, Consumer Awareness, and Choice: Evidence from the U.S. Banking Industry,” *RAND Journal of Economics*, 48, 611–46.

HUBER, J., J. PAYNE, AND C. PUTO (1982): “Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis,” *Journal of Consumer Research*, 9, 90–98.

HUI, S., E. BRADLOW, AND P. FADER (2009): “Testing Behavioral Hypotheses Using an Integrated Model of Grocery Store Shopping Path and Purchase Behavior,” *Journal of Consumer Research*, 36, 478–93.

KAWAGUCHI, K., K. UETAKE, AND Y. WATANABE (2019): “Effectiveness of Product Recommendation under Time and Social Pressures,” *Marketing Science*, 38(2), 253–273.

LEE, L.-F. (1995): “Asymptotic Bias in Simulated Maximum Likelihood Estimation of Discrete Choice Models,” *Econometric Theory*, 11(3), 437–483.

LEE, Y. (2019): "Fast computation algorithm for the random consideration set model," *Economics Letters*, 179, 38 – 41.

MANSKI, C. (1977): "The Structure of Random Utility Models," *Theory and Decision*, 8, 229–54.

MANZINI, P., AND M. MARIOTTI (2014): "Stochastic Choice and Consideration Sets," *Econometrica*, 82, 1153–76.

MASATLIOGLU, Y., D. NAKAJIMA, AND E. OZBAY (2012): "Revealed Attention," *American Economic Review*, 102, 2183–2205.

MATZKIN, R. (1992): "Nonparametric and Distribution-Free Estimation of the Binary Threshold Crossing and the Binary Choice Models," *Econometrica*, pp. 239–70.

MEHTA, N., S. RAJIV, AND K. SRINIVASAN (2003): "Price Uncertainty and Consumer Search: A Structural Model of Consideration Set Formation," *Marketing Science*, 22, 58–84.

MILGROM, P., AND J. ROBERTS (1986): "Price and Advertising Signals of Product Quality," *Journal of Political Economy*, 94, 796–821.

NELSON, P. (1974): "Advertising as Information," *Journal of Political Economy*, 82(4), 729–754.

NOWLIS, S. M. (1995): "The effect of time pressure on the choice between brands that differ in quality, price, and product features," *Marketing Letters*, 6(4), 287–295.

PALAZZOLO, M., AND F. FEINBERG (2015): "Modeling Consideration Set Substitution," *mimeo.*

PAYNE, J., J. BETTMAN, AND E. JOHNSON (1988): "Adaptive Strategy Selection in Decision Making," *Journal of Experimental Psychology*, 14, 534–552.

REUTSKAJA, E., R. NAGEL, C. F. CAMERER, AND A. RANGEL (2011): "Search Dynamics in Consumer Choice under Time Pressure: An Eye-Tracking Study," *American Economic Review*, 101, 900–926.

Roberts, J. H., and J. M. Lattin (1991): "Development and Testing of a Model of Consideration Set Composition," *Journal of Marketing Research*, 28, 429–440.

Sahni, N. (2016): "Advertising Spillovers: Evidence from Online Field Experiments and Implications for Returns on Advertising," *Journal of Marketing Research*, 53, 459–78.

Simonson, I. (1989): "Choice Based on Reasons: The Case of Attraction and Compromise Effects," *Journal of Consumer Research*, 16, 158–174.

Stigler, G. L. (1961): "The Economics of Information," *Journal of Political Economy*, 71, 213–25.

Van Nierop, E., B. Bronnenberg, R. Paap, M. Wedel, and P. H. Franses (2010): "Retrieving Unobserved Consideration Sets from Household Panel Data," *Journal of Marketing Research*, 47, 63–74.

Online Appendix (Not for Publication)

Appendix A Picture of the Vending Machine

The figure shows a vending machine used for the experiment on a train platform. There is a digital touch panel in the front and a camera at the top of the machine recognizes the customer's age and gender. The vending machine makes recommendations based on the observed consumer characteristics. The image is supplied by the company.

Appendix B Balance Check

Our empirical analyses rely on the assumption that beverage demand is not correlated with the treatment condition conditional on observed characteristics. Although it is not possible to perfectly prove the exogeneity assumption, we can show how observable characteristics vary by the condition.

As Table 10 shows, no systematic difference in observable characteristics exists between the treatment and control conditions.[39]

[39]The difference might be statistically significant for some variables, but it is small in magnitude and does not affect our estimates.

Table 10: Balance Check

	Control		Treatment	
	Mean	Std Dev	Mean	Std Dev
Precipitation	3.853	8.91	3.807	9.03
Temperature	26.26	1.514	26.168	1.517
crowdedness_hour	337.2	767.4	308.7	703.5
time to next	3.436	5.865	3.187	5.163
time to after next	3.481	3.809	3.373	3.73
Female	0.3126	0.4636	0.3227	0.4675
Age 10	0.0148	0.1208	0.0155	0.1235
Age 20	0.1481	0.3552	0.1546	0.3615
Age 30	0.3447	0.4753	0.3469	0.476
Age 40	0.3218	0.4702	0.3191	0.4661
Age 50	0.1623	0.369	0.1639	0.3702

Notes: In the table, we compare some machine-level characteristics for the control group and the treatment group.

Appendix C Identification by Product Availability

In this section, we provide some discussion about identification of the consideration set model.

C.1 Intuition of the Product Availability Approach

We begin with the intuition of the identification based on product availability. The idea is based on the recently growing literature in decision theory, particularly Masatlioglu, Nakajima, and Ozbay (2012) and Manzini and Mariotti (2014). These studies exploit variations in product availability to infer consumer attention and preferences using only choice data. Masatlioglu, Nakajima, and Ozbay (2012) consider a case in which both preference and attention are deterministic, whereas Manzini and Mariotti (2014) consider a case in which only attention is stochastic. These papers provide us with a useful starting point, although the empirical consideration set models have stochastic error terms in both preference and attention.

Deterministic Preference and Attention Suppose there are three products (1, 2, and 3), and we observe two different choice occasions of the same decision maker in which different

Case	1	2
Feasible Sets	$1,2,3$	$1,2$
Choices	1	2

Table 11: Example: An example of feasible sets and choices from Masatlioglu et al. (2012). Product 3 must have been considered in Case 1 because choices are different in Cases 1 and 2. Removing 3 from the choice set cannot affect consumer choice if 3 has not been considered.

sets of products are available, as in Table 11. In Case 1, all of 1, 2, and 3 are available, and the consumer chooses product 1. In Case 2, only products 1 and 2 are available, and the consumer chooses 2. Note that the consumer may not be aware of some of the products, even if they are available.

Masatlioglu, Nakajima, and Ozbay (2012) consider the case with no stochastic shock in either preference or attention and argue that product 3 must have been considered in Case 1, because choices cannot be flipped from Case 1 to Case 2 if product 3 was not included in the consideration set in Case 1.[40] Note that without more observations, we cannot pin down preference and attention from this example. This can be because either product 1 is not included in the consideration set in Case 2 and $1 \succ 2$ in Case 1 or product 2 is not included in the consideration set in Case 1 and $2 \succ 1$ in Case 2. More data may allow us to infer preference and attention further, or more data may reject the rational choice model under limited attention. In fact, Masatlioglu, Nakajima, and Ozbay (2012) do not discuss what kind of data are necessary to identify preference and attention completely. Instead, they show that it is possible to test if consumers use rational choice with limited attention *given the data at hand* if and only if a weaker version of WARP (Weak Axiom of Revealed Preference) is satisfied. Hence, sometimes preference or attention are only partially identified or are not identified.[41]

[40]Formally, Masatlioglu, Nakajima, and Ozbay (2012) require that if an alternative does not attract attention from the decision maker, his/her consideration set does not change when such an item becomes unavailable.

[41]Note that full identification of attention in this example requires attention to be dependent on feasible sets. Requiring independence between attention and feasible sets in the deterministic model does not make much sense because it implies that a product is either always included or always excluded. Then, observed chosen products are always included in any consideration set and are unlikely to be rationalized by choice with limited attention. We thank Stephan Seiler for making this point.

Deterministic Preference and Stochastic Attention Extending this idea to the case with stochastic consideration set formation but a deterministic preference, in which the choice probability for each product can be observed, Manzini and Mariotti (2014) argue that product 3 must have been considered with *positive probability* in $\{1,2,3\}$ if the choice probability of product 1 and that of product 2 differ between these two cases. The reason for this argument is similar to that for the previous case: removing product 3 from the feasible set cannot affect consumer choice if product 3 has not been considered at all. This argument is the basic intuition behind the identification of attention. To identify preference, Manzini and Mariotti (2014) further argue that removing product 3 from the feasible set affects the choice probability of 1 only if $3 \succ 1$, because the inclusion of lower-ranked alternatives does not influence the choice when preference is deterministic. As such, variations in choice sets provide information about both attention and preference.

C.2 Identification Result

When both preference and attention are stochastic, then the change in choice behavior is not as discrete as in the deterministic preference case in Manzini and Mariotti (2014). Although the fully nonparametric identification of the model becomes more difficult, the basic intuition as outlined above carries over to the stochastic preference case, with some additional restrictions. The identification discussion in this section is nonparametric and therefore requires strong assumptions that are not necessary to identify the empirical model in Section 3.

Let us begin with a simple case in which there are two feasible sets $J = \{1\}$ and $J' = \{1,2\}$. Then, the market share of $j = 1$ for feasible set J is

$$s_1(J) = \pi_1(\{1\})\gamma_1, \tag{C.1}$$

where $\pi_1(\{1\})$ is the conditional choice probability of product 1 given consideration set $C = \{0,1\}$, and γ_1 is the probability of having consideration set $C = \{0,1\}$ given $J = \{1\}$.[42] Similarly,

[42] Because product 0 is the outside option and it is always included in the consideration set, we omit it in $\pi_j(\cdot)$.

the market share of product 1 for feasible set J' is written as

$$s_1(J') = \pi_1(\{1\})\gamma_1(1-\gamma_2) + \pi_1(\{1,2\})\gamma_1\gamma_2. \tag{C.2}$$

The first term on the right-hand side is the probability of choosing product 1 when the consideration set is $\{0,1\}$, and the second term is the probability of choosing product 1 when the consideration set is $C = \{0,1,2\}$. Furthermore, the probability of $C = \{0,1\}$ given J' is simply $\gamma_1(1-\gamma_2)$. Similarly, the probability of $C = \{0,1,2\}$ given J' is $\gamma_1\gamma_2$.

Now, using equations (C.1) and (C.2), we obtain the following:

$$\begin{aligned}
\underbrace{\frac{s_1(J) - s_1(J')}{s_1(J)}}_{\text{%-change in product 1's market share}} &= \gamma_2\left[1 - \frac{\pi_1(\{1,2\})}{\pi_1(\{1\})}\right] \\
&= \underbrace{\gamma_2}_{\text{Attention probability of product 2}} \\
&\quad \times \underbrace{\frac{\pi_1(\{1\}) - \pi_1(\{1,2\})}{\pi_1(\{1\})}}_{\text{%-change in product 1's conditional choice probability}} .
\end{aligned} \tag{C.3}$$

This relationship is the basis of our identification argument. Intuitively, the left-hand side, $\frac{s_1(J)-s_1(J')}{s_1(J)}$, is the percentage change if product 1's market share when product 2 is added to the feasible set, which is observable. The right-hand side of the equation decomposes it into two parts: i) the probability that product 2 is considered and ii) the percentage change in product 1's conditional choice probability when product 2 is added to the consideration set. This decomposition allows us to separate the attention probability for product 2 from other parts. If consumers are aware of all available products, it is easy to observe that $\frac{s_1(J)-s_1(J')}{s_1(J)} = \frac{\pi_1(\{1\})-\pi_1(\{1,2\})}{\pi_1(\{1\})}$.

Now, note that γ_1 does not appear in the equation, as it is canceled out. In addition, note that γ_2 does not appear in the second term on the right-hand side. Hence, if there is sufficient variation in the variable that only decreases the utility from product 2, one can *nonparametri-*

cally identify γ_2. Once these parameters are identified, we can apply the method established in the discrete choice literature to identify the remaining parts of the model. The key part of the identification is that the attention to product 2 is separated from preference terms, which allows us to identify the attention model nonparametrically. In other words, if we consider the effect of advertisement on product 1, the effect on attention for product 1, if any, does not change the ratio $\frac{s_1(J)-s_1(J')}{s_1(J)}$, and the effect of advertisements on product 1's choice probability depends on its effect on preference $1-\frac{\pi_1(\{1,2\})}{\pi_1(\{1\})}$ but not its effect on attention.

With more than two products in the feasible set, the decomposition in Eq. (C.3) becomes slightly more complicated, but similar intuition works. With $J=\{1,2,3\}$, it is possible to show that $\frac{s_1(J)-s_1(J_3)}{s_1(J_3)}=\gamma_3\left[1-\frac{\pi_1(\{1,3\})(1-\gamma_2)+\pi_1(\{1,2,3\})\gamma_2}{\pi_1(\{1\})(1-\gamma_2)+\pi_1(\{1,2\})\gamma_2.}\right]$, where the attention probability of the dropped product (i.e., product 3) is separated from the percentage change in product 1's conditional choice probability in the square bracket. Hence, given a large variation in the variable that only decreases the utility from product 3, we can identify γ_3.

The formal proposition is the following.

Proposition 1. *Suppose there exist feasible sets J and $J_j = J\backslash\{j\}$ for all $j\in J$. Then, the consideration set formation model, $\{\gamma_j\}_{j\in J}$, is identified.*

Proof: We can prove the proposition with the case of $J=\{1,2,3\}$ without loss of generality. In this case, we obtain

$$\begin{aligned} s_1(J) &= \pi_1(\{1\})\gamma_1(1-\gamma_2)(1-\gamma_3)+\pi_1(\{1,2\})\gamma_1\gamma_2(1-\gamma_3) \\ &\quad +\pi_1(\{1,3\})\gamma_1(1-\gamma_2)\gamma_3+\pi_1(\{1,2,3\})\gamma_1\gamma_2\gamma_3. \end{aligned} \tag{C.4}$$

Now, we can identify γ_3 from choice observations from $J_3=\{1,2\}$. We can write $s_1(J_3)$ as follows:

$$s_1(J_3)=\pi_1(\{1\})\gamma_1(1-\gamma_2)+\pi_1(\{1,2\})\gamma_1\gamma_2. \tag{C.5}$$

Combining Eqs. C.4 and C.5 yields

$$\frac{s_1(J)-s_1(J_3)}{s_1(J_3)}=\gamma_3\left[1-\frac{\pi_1(\{1,3\})(1-\gamma_2)+\pi_1(\{1,2,3\})\gamma_2}{\pi_1(\{1\})(1-\gamma_2)+\pi_1(\{1,2\})\gamma_2.}\right] \tag{C.6}$$

Note that γ_1 does not appear on the right-hand side of Eq. (C.6) and that γ_3 does not appear in the object in the square brackets. $\frac{s_1(J)-s_1(J_3)}{s_1(J_3)}$ is the percentage change in product 1's market share when product 3 is added to the feasible set (i.e., change from J_3 to J), which can be decomposed into attention probability for product 3, γ_3, and %-change of product 1's conditional choice probability when product 3 is added. The only difference from Eq. (C.3) is that product 2 might also be included in the consideration set. Now, we adopt the standard identification-at-infinity argument (see, e.g.,Matzkin (1992); Briesch, Chintagunta, and Matzkin (2010); Berry and Haile (2011)). In Eq. (C.6), driving $Z_3 \to +\infty$, we obtain $\frac{s_1(J)-s_1(J_3)}{s_1(J_3)} \to \gamma_3$.[43] This identifies γ_3 nonparametrically. Using J_1 and J_2 with J, similarly, we can nonparametrically identify γ_1 and γ_2, respectively. Thus, all attention parameters are identified.[44] Note that it is straightforward to identify parameters in γs in Section 2 by using a simple panel linear regression argument, as we have a panel structure in the data. ■

We omit consumer heterogeneity (i.e., random coefficients) in this discussion, but this is only for expositional simplicity. As we discussed in the main text, the identification of random coefficients relies not only on the variation in product availability but also on the variation in product/market characteristics across markets as shown in Berry and Haile (2011) and Fox, Kim, Ryan, and Bajari (2012), which show nonparametric (semiparametric) identification of random coefficient multinomial discrete-choice models.

Moreover, the proposition is derived using feasible sets J and $J_j = J\backslash\{j\}$, but it is possible to identify the model using different types of variation in feasible sets, for example, when there

[43]This procedure is possible because of the exclusion restriction, that is, Z_3 does not affect the attention of product 3.

[44]In a general case, we can always write $\frac{s_1(J)-s_1(J_j)}{s_1(J_j)}=\gamma_j F(\pi_1,\gamma_1,...,\gamma_J)$, where $F(\cdot)$ is a function of $\pi_1(\widetilde{J})$ for $\widetilde{J}\subseteq J$ such that $1\in\widetilde{J}$, and γ_k for $k\in J\backslash\{j\}$, and $F(\cdot)\to 1$ as $Z_j\to+\infty$. Hence, $\frac{s_1(J)-s_1(J_j)}{s_1(J_j)}\to\gamma_j$ when $Z_j\to+\infty$. This identifies γ_j.

are two products added to (or dropped from) the feasible set as $J = \{1,2,3\}$ and $J_{23} = J \backslash \{2,3\}$. Then, we can show that the ratio $\frac{s_1(J_{23})-s_1(J)}{s_1(J)}$ and variations in Z_2 and Z_3 allow us to identify $\gamma_2 + \gamma_3 - \gamma_2\gamma_3$. More formally, $\frac{s_1(J_{23})-s_1(J)}{s_1(J)} = \gamma_2 + \gamma_3 - \gamma_2\gamma_3 - \frac{\pi_1(\{1,2\})\gamma_2(1-\gamma_3)+\pi_1(\{1,3\})(1-\gamma_2)\gamma_3+\pi_1(\{1,2,3\})\gamma_2\gamma_3}{\pi_1(\{1\})}$. Since $\pi_1(J) \to 0$ as $Z_2, Z_3 \to +\infty$ for any J, we obtain the results in the main text, which imposes another over-identifying restriction over γs. When we observe more variations in feasible sets, such as J_{12}, J_{13}, and J_{23}, we can have more restrictions to identify γ_j. In fact, identification can be achieved under more general patterns of feasible sets. A key implication of this proposition is that product availability provides a source of identification in general.

Finally, note that we assume attention to be independent across products, following the literature on the consideration set models ((Goeree 2008); Van Nierop, Bronnenberg, Paap, Wedel, and Franses (2010); Barroso and Llobet (2012)). Although we acknowledge that this assumption helps computation significantly, in principle, we can relax the independence assumption because it is possible to use the overidentifying restrictions created by many types of feasible sets. In an extreme case, one can nonparametrically identify arbitrary correlation across errors in the attention function if there is a set of feasible sets that includes all possible combinations of products.

C.3 Identification with Exclusion Restrictions on Utility Formation

In this section, we provide a proof of identifying the consideration set model with exclusion restrictions on utility formation. Although not explicitly mentioned, the existing papers use this exclusion restriction to identify the consideration set model. Two main conditions for the identification are i) Advertisement A_{ijt} is excluded from utility but included for attention, that is, $\beta_{0i} = 0$, ii) the support of A_{ijt} is $(-\infty, +\infty)$ for all j, and iii) the attention level for product j is increasing in A_{ijt}, that is, $\alpha_{0i} > 0$.

Proposition 2. *Under the exclusion restriction and the large support conditions, the consideration set model is identified.*

Proof: When advertisement A_{ijt} is excluded from the choice probability conditional on consideration set π_{jt}, the choice probability of product j in market t is written as

$$s_{ijt}(J_t) = \prod_{C \subseteq J_t} \pi_j(\{X_{ikt}, Z_{kt}\}_{k \in C} | C) \prod_{k \in C} \gamma(X_{ikt}, A_{ikt}) \prod_{l \notin C} (1 - \gamma(X_{ilt}, A_{ilt}))$$

First, notice that $\lim_{A_{ikt} \to +\infty} \gamma_j(X_{ikt}, A_{ikt}) = 1$ without affecting $\gamma(X_{ijt}, A_{ijt})$ for any other j. This occurs because A_{ikt} affects only the consideration probability of product k and because the large support condition allows us to drive $A_{ijt} \to +\infty$ for all j. In other words, we can consider the situation in which all products are considered by driving $A_{ijt} \to +\infty$. Under such a situation, the model is exactly the same as the regular discrete-choice demand model. Hence, we can use the results from this stream of literature, such as Berry and Haile (2011), and Fox, Kim, Ryan, and Bajari (2012). ■

Observe that the exclusion restriction on utility formation is crucial for this identification. If A_{ijt} also impacts preference, as in our model, both π_j and γ_j move together, and it is not possible to determine where the expression converges. Note also that not only the exclusion restriction, but also the large support condition is crucial for identification. If A_{ijt} is not excluded, $\pi_j(\cdot)$ also move as $A_{ijt} \to +\infty$. Hence, one cannot tell if the consideration set model converges to a regular discrete-choice model. If A_{ijt} does not satisfy the large support condition, one cannot drive A_{ijt} to positive infinity to eliminate the effect of unobservable consideration sets. The product availability approach does not require these conditions for identification. Moreover, the independence assumption is also important, because it might be possible that more advertisement for product j leads to lower attention for product j', i.e., $\gamma_{j'} \to +0$ if $A_{ijt} \to +\infty$. Then, one cannot construct a situation in which all products are included in the consideration set. Therefore, this approach also relies on the independence assumption. Lastly, A_{ijt} can be other than advertising. We can use other variables such as menu-related variables to the extent that these variables are excluded from the utility.

Appendix D Estimation Algorithm

We explain the algorithm proposed by Lee (2019) to calculate the choice probabilities in 4.3. Since his algorithm does not consider random coefficients in attention and utility functions, we need to integrate choice probabilities out. Below, the algorithm runs given a fixed set of draws of random coefficients. Let $G(u) = \gamma_{ijt} \prod_{l \in \mathscr{J}_t \setminus \{j\}} [1 - \gamma_{ilt} + \gamma_{ilt} P(u|U_{ijt})] p(u|U_{ijt})$ and take any small tolerance value ϵ. Remember that $P(u|U)$ is the probability distribution of the idiosyncratic shock in the utility function.

1. Set $\overline{u} = \max\{u \in \mathbf{R} | u = P^{-1}(1 - \epsilon/2|\alpha_l^i), l \in \mathscr{J}_t\}$ and $\underline{u} = \min\{u \in \mathbf{R} | u = P^{-1}(\epsilon/2|\alpha_l^i), l \in \mathscr{J}_t\}$

2. Define $u_n = \underline{u} + (\overline{u} - \underline{u})\frac{n-1}{N}$ for $n = 1, ..., N+1$.

3. Let $\overline{G}(u_n) = \frac{G(u_n) + G(u_{n+1})}{2}$ for $n = 1, ..., N$. Then,

$$\hat{p}_{ijt} = \sum_{n=1}^{N} \overline{G}(u_n)(u_{n+1} - u_n). \tag{D.1}$$

Note that $\hat{p}_{ijt}$ is computed given a set of simulated draws of $\{\zeta_j, \xi_j\}_{j \in \mathscr{J}}, \{\alpha_{ij}, \beta_{ij}\}_{i \in \mathscr{I}, j \in \mathscr{J}}, \{\boldsymbol{\nu}_i, \boldsymbol{\upsilon}_i\}_{i \in \mathscr{I}}$.

Finally, the simulated likelihood can be written by

$$\frac{1}{N_s} \sum_{s=1}^{N_s} \prod_i \prod_j \prod_t \hat{p}_{ijt}^s,$$

where N_s is the number of simulations (which we set $N_s = 100$) and p_{ijt}^s is the choice probability given a set of simulated draws calculated based on equation (D.1). We draw $\{\zeta_j, \xi_j\}_{j \in \mathscr{J}}$, $\{\alpha_{ij}, \beta_{ij}\}_{i \in \mathscr{I}, j \in \mathscr{J}}, \{\boldsymbol{\nu}_i, \boldsymbol{\upsilon}_i\}_{i \in \mathscr{I}}$ from an i.i.d. standard Normal distribution, respectively. Lastly, the confidence intervals of the parameters are computed based on 100 bootstrapped samples.

Table 12: The Elasticity of Attention Probabilities to Product Recommendation

Elasticity	*Mean*	*Sd*	*Min*	*Max*	*Q1*	*Q2*	*Q3*
own	0.198	0.044	0	0.279	0.181	0.204	0.225
cross	0.065	0.017	0	0.122	0.056	0.068	0.076

Note: First, we compute the attention probability of a product in a purchase occasion for a consumer and then integrate consumer heterogeneity to obtain the attention probability at the product and purchase occasion level when there is no product recommendation, the product is recommended, and other product is recommended. Second, compute the changes in the attention probabilities. The summary statistics above are across product and purchase occasions. Because some cases show extremely high elasticity due to the small baseline choice probabilities, we dropped cases with the top 2.5% own elasticities.

Appendix E Additional Results

Elasticities on Consideration Probability Table 12 presents the elasticity of recommendations to the probability that a product is included in the consideration set instead of the purchase incidence. As in Section 5.2, we calculate own and cross elasticities. We find that recommendations increase the probability of being considered by 19.8% for recommended products and 6.5% for non-recommended products.

Table 13 presents the elasticity of attention probability to product recommendations according to different degrees of time pressure. We find that own elasticities of attention probability first increase as time pressure weakens and then decrease, while cross elasticities monotonically decrease, because the baseline attention level under $T1 = 10$ is already high, and hence, the increase measured as a percentage would be smaller.

Logit Results In this section, we compare the consideration set model with a random-coefficient multinomial logistic regression model. We estimate the following multinomial logistic regression model.

$$U^*_{ijt} = \begin{cases} \beta_{0i} A_{ijt} - \beta_{1i} P_j + \boldsymbol{\beta}'_{2i} \boldsymbol{M}_t + \boldsymbol{\beta}'_{3i} \boldsymbol{X}_{jt} + \beta_{ij} + \xi_j + \eta_{ijt} \equiv U_{ijt} + \eta_{ijt} & j \in \mathcal{J}_t \setminus \{0\}, \\ \varepsilon_{i0t} & j = 0, \end{cases} \tag{E.1}$$

Table 13: Elasticity of Attention Probabilities to Product Recommendation by Time Pressure

Time	*0*	*5*	*10*
own	0.152	0.250	0.189
cross	0.069	0.060	0.044

Note: The setting is the same as that in Table 12 except that the time to the next train (minutes) is set at 0, 5, and 10 minutes from the time of a purchase. Because some cases show extremely high elasticity due to the small baseline choice probabilities, we dropped cases with the top 2.5% own elasticities.

where the model includes all variables listed in Table 3. Hence, the difference between the full consideration model and the model above is the fact that this multinomial logistic regression model does not take into account the fact that the customer may not be aware of all available products in reality.

First, we show the parameter estimates from the multinomial logistic regression model. Although the estimated coefficients are mostly consistent with the estimates of the consideration model reported in Table 5, there are a few key differences. For example, we find that the coefficient on T1 is positive in the multinomial logistic regression, but the one in the preference part of the consideration set model is negative. In other words, the logistic regression model cannot separately estimate the differential impact of time pressure on attention and preference. This fact may have potentially a large managerial implications because the company may want to decrease time pressure for the product with greater taste.

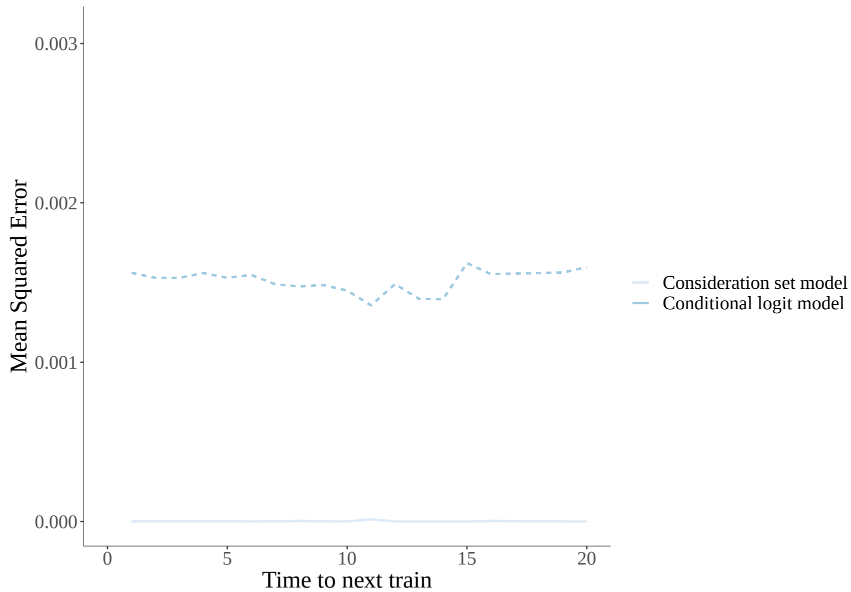
We then compare the fit of the two models. In Figure E.1, we plot the mean square error of the logit model and the consideration set model by T1. We find that the consideration set model fits to the data better overall. In particular, the consideration set model performs better than the logistic regression model when time pressure is very high or very low. In Table 15, we also report the log-likelihood and AIC of the multinomial logistic regression model and the consideration set model. The table shows that the consideration set model outperforms the logistic regression model in terms of both likelihood and AIC.

Table 14: The Estimation Result of Parameters: Logit Model

Utility	*Estimate*	*95% C.I.*
PR	0.091	[0.074, 0.127]
MR	0.078	[0.062, 0.087]
NR	0.093	[0.077, 0.128]
NR^2	-0.036	[-0.039, -0.026]
NS	0.076	[0.063, 0.101]
price	-1.335	[-1.383, -1.300]
PR × T1	0.053	[0.021, 0.075]
NR × T1	0.070	[0.047, 0.086]
PR× NR	-0.029	[-0.054, -0.022]
NS	0.076	[0.063, 0.101]
menu	-0.036	[-0.040, -0.027]
PR × Can	0.057	[0.010, 0.079]
PR × Glass bottle	0.071	[0.027, 0.181]
MR × Can	0.067	[0.039, 0.082]
MR × Glass bottle	0.063	[0.009, 0.239]
T1	-0.099	[-0.206, -0.027]
T2	-0.133	[-0.259, -0.026]
Green tea	0.050	[0.006, 0.066]
Other tea	0.120	[0.057, 0.137]
Water	0.045	[-0.044, 0.062]
Sport	0.116	[0.071, 0.188]
Canned coffee	0.060	[0.021, 0.074]
Bottled coffee	0.080	[0.022, 0.259]
Black tea	0.054	[-0.025, 0.072]
Carbonated	0.055	[-0.004, 0.229]
Fruit	0.050	[0.003, 0.063]
Healthy	0.058	[-0.013, 0.227]
Other drink	0.075	[-0.001, 0.263]
Green tea × Temp	-0.022	[-0.024, -0.018]
Other tea × Temp	-0.043	[-0.046, -0.039]
Water × Temp	-0.024	[-0.031, -0.020]
Sport × Temp	-0.029	[-0.033, -0.025]
Canned coffee × Temp	-0.035	[-0.040, -0.032]
Bottled coffee × Temp	0.003	[-0.002, 0.008]
Black tea × Temp	-0.007	[-0.011, -0.002]
Carbonated × Temp	-0.055	[-0.058, -0.051]
Fruit × Temp	-0.031	[-0.035, -0.027]
Healthy × Temp	-0.026	[-0.030, -0.021]
Other drink × Temp	0.010	[-0.024, 0.037]
Can	0.063	[-0.016, 0.091]
Glass bottle	0.064	[0.002, 0.240]
Plastic bottle × Volume	0.058	[-0.021, 0.079]
Can × Volume	0.091	[0.023, 0.279]
Glass bottle × Volume	0.076	[0.054, 0.091]
ω	0.076	[0.026, 0.176]

Note: Standard errors are estimated by 100 bootstrapping samples. We omit some estimates from the table to save the space.

Figure E.1: Mean Square Error by Time Pressure



Note: The mean square error is calculated by $\frac{1}{NJT}\sum(d_{ijt}-D_{ijt})^2$, where d_{ijt} is the observed choice by consumer i and D_{ijt} is the predicted choice probability of product j by consumer i.

Table 15: Fit of the Model

	Logit	Consideration Set
Log Likelihood	189.7	102.7
AIC	479.4	363.4

Lastly, we conduct the counterfactual simulation to find the optimal number of product recommendations using the estimated multinomial logit model. Table 16 shows that the optimal number of recommendations is surprisingly large (more than 20 products) and the resulted sales volume increase seem unreasonably high (greater than 20%). These results may imply that the consideration set model would fit to the data better and give us reasonable counterfactual simulation results.

Table 16: Counterfactual Simulation: Logistic Regression

	Sales Change	*NR*	
		Mean	*S.D.*
Actual	-	4.44	1.74
Uniform	22.17%	24.00	-
Target	22.30%	24.67	0.52
Contextual	24.87%	25.99	3.92