# Designing Context-Based Marketing:
# Product Recommendations under Time Pressure*

Kohei Kawaguchi†        Kosuke Uetake‡        Yasutora Watanabe§

December 26, 2018

## 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 the recommendation menu influence attention and utility. We find that time pressure reduces consumer attention but increases utility in general. Time pressure moderates the effect of recommendations for 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 simulation, we find that the revenue-maximizing number of recommendations increases with time pressure. Optimizing the number of recommendations for each vending machine and for each time of day increases the total sales volume by 4.5% relative to the actual policy, 1.9% points more than traditional consumer-segment-based targeting.

**Keywords**: Consideration Set, Context-based marketing, Time pressure, Recommendations, Menu Effects

†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 more 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 like. 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. For example, Dhar and Nowlis (1999) find that consumers are more likely to avoid making any decisions when they are under time pressure in order 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. Our companion paper, Kawaguchi, Uetake, and Watanabe (Forthcoming) 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.

In this paper, we investigate how to optimize marketing interventions to contextual factors. More specifically, we study the design of the product recommendation system for beverage vending machines under 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, time pressure affects the effectiveness of product recommendations, 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. Estimating a consideration set model with these features allows us to measure how context factors and marketing interventions jointly influence consumer attention and utility, while taking into account the "menu" effects in order to design rich context-based marketing.

A number of unique features of our setup allow us to estimate the flexible consideration

---

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

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.[2] By utilizing the train schedule information, we can precisely measure time pressure at the minute level.

Second, our consumer purchase data come from a period when the company owning the vending machines executed an experiment on product recommendations. The vending machines we study are equipped with a recommendation system that can change recommendations according to customer attributes recognized by the camera installed at the top of the machines. 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 can 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. Based on the recent development of the decision theory literature on the consideration set model,[3] this variation in assortment (that is, the available set of products) gives us a useful source of identification for the consideration set model. Specifically, it allows 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)).[4] With the variation in

---

[2]In our previous research (Kawaguchi, Uetake, and Watanabe (Forthcoming)), we conducted a validation experiment to study if the time to the next train is strongly correlated with the time pressure felt by subjects, and we confirmed this relationship.

[3]See, e.g., Masatlioglu, Nakajima, and Ozbay (2012) and Manzini and Mariotti (2014)

[4]The marketing and economics literature has considered the effect of advertisements on attention and utility separately as an informational role of advertisements (Stigler (1961); Grossman and Shapiro (1984); Milgrom and Roberts (1986)) and as a persuasive role of advertisements (Becker and Murphy (1993)). The effect on utility sug-

product availability, advertising variables need not be excluded from the utility formation in our model.

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; and iv) menu characteristics significantly affect consumer attention-in particular, the number of total recommendations increases the attention level in general, but in a decreasing order. Moreover, we find significant heterogeneity in the effects of recommendations across customer segments.

To quantify the efficacy of product recommendations, we calculate the elasticities of recommendations to purchase incidence. We find that, compared to the baseline where no product is recommended, recommending a product increases its sales by 45% (*own* elasticity) on average and increases non-recommended products' sales by 6.7% through spillover effects (*cross* elasticity). Overall, the sales of a vending machine increase by 8% due to a recommendation. We then decompose these elasticities into the attention channel and the utility channel. We find that own elasticity is driven mainly by the utility channel, while overall purchase

gests the existence of the persuasive role, while the effect on attention suggests the existence of the informational role. Ackerberg (2001) proposes a reduced-form way to separately estimate the information effect and the prestige effect using the variation in consumer experiences. Ching and Ishihara (2010) and Ching and Ishihara (2012) develop discrete-choice models of the physician's brand choice, which is similar to the consideration set model, and study the informational and persuasive roles of detailing.

elasticity is driven by the attention channel. Moreover, we find that these elasticities vary by time pressure; the own elasticities of recommendations increase with time pressure, while the cross elasticities decrease with time pressure. Overall, the purchase elasticity decreases with time pressure.

Our main goal is to derive managerial implications for designing product recommendations under 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 product, it may not be optimal to recommend too many products because it dilutes consumer attention. If we fix the degree of time pressure at the actual level, we find that the optimal number of recommendations is approximately 10. As the degree of time pressure decreases, we find that the company should recommend more products.

Second, we examine the revenue-maximizing recommendation policy that adjusts the number of recommendations at the machine and time-of-day level. We find that this policy increases sales by 4.5% compared to the actual policy. By contrast, when the company sets the number of recommendations uniformly across vending machines and times, sales can increase only up to 2.4% compared to the actual policy. The traditional consumer-segment-based optimization of the number of recommendation increases only by 2.6%. These results indicate the potential impacts of context-based recommendations.

**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)).[5] The consideration set model is useful as it allows

---

[5]Some papers use direct information about consideration sets, such as survey data. 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 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

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)).

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 variation is available. Abaluck and Adams (2018) propose a new identification strategy for the consideration set model based on asymmetric demand responses to the change in product characteristics. Our paper adds to this literature by explicitly considering the effects of context 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) 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, our companion paper, Kawaguchi, Uetake, and Watanabe (Forthcoming), examines the effectiveness of product recommendations when consumers are under time pressure. The paper finds that time pressure weakens this effectiveness.

Finally, our identification strategy relies on the idea developed by the growing body of literature in decision theory on choice-theoretic axiomatization of consideration set models (e.g.,

---

there is a potential reporting bias due to the nature of surveys.

Masatlioglu, Nakajima, and Ozbay (2012); De Clippel and Rosen (2014); Manzini and Mariotti (2014)). These studies consider how consumers' utilities and consideration sets can be elicited using variation in product availability, and establish conditions for the consideration set model to be rationalized by the data. We exploit the variation in product availability based on this approach to identify our consideration set model, which allows us to include advertising in both consumer attention and utility. Without such variation, the effects of the advertisement variable cannot be separately identified and therefore must be excluded from utility, as in Van Nierop, Bronnenberg, Paap, Wedel, and Franses (2010) and Goeree (2008).

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 decision from vending machines placed at train stations in the Tokyo metropolitan area. For details on the setup, please see Kawaguchi, Uetake, and Watanabe (Forthcoming), 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.[6] 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).[7] The company controls the recom-

---

[6]Because of privacy concerns, the cameras do not record any information on consumer characteristics. Hence, neither we nor the company can use the information from the cameras except to change the recommendations. In the current system, all vending machines must follow the same policy at the same time.

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

mendation 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.



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 our companion paper, Kawaguchi, Uetake, and Watanabe (Forthcoming). 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 time of day for weekdays during the week of July 15 to 26, 2013, as shown in Table 1.[8]

The experiment creates exogenous variations in recommendations, which also creates variation in the number of recommended products among treatment groups.[9] Because available

---

recommended.

[8]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.

[9]Note that the experiment is not meant to create exogenous variations in product availability. Instead, it creates

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 (Forthcoming). We conduct 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, we show the product recommendations chosen by the company, but we do not use the data for these slots in our empirical analysis.

products are different across vending machines, the number of recommended products can 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 context 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 period of time. Because trains in Tokyo

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 that the endogeneity concern on product availability, however, could be limited. The company delegates the decision on fulfilling vending machines to local agents who are responsible for several machines or stations. Based on our discussion with the company, the company is worried about the poor performance of capturing local demand by the current system.

operate punctually and arrive frequently, passengers tend to be under the influence of time pressure.[10]

We obtain the time schedule of the Japan Railway East on weekdays during the experiment. 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.[11]

## 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. In addition, 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 does not allow a consumer to spend enough time considering all available products, and

---

[10]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.

[11]To examine the validity of these proxy variables, in Kawaguchi, Uetake, and Watanabe (Forthcoming), we 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 is available from the authors upon request.

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).

## 2.5 Data

Table 2: Summary Statistics

|  | variable | N | Mean | Sd | Min | Max |
|---|---|---|---|---|---|---|
| *Machine-date-time* | Sales units | 7309 | 4.53 | 3.6 | 0 | 37 |
|  | Sales value (JPY) | 7309 | 614.65 | 488.8 | 0 | 5100 |
|  | Machine recommendation | 7246 | 0.58 | 0.49 | 0.0 | 1.0 |
|  | Number of recommendations | 7309 | 2.59 | 2.56 | 0.0 | 10.0 |
|  | Number of unique products | 7309 | 29.63 | 2.12 | 22.0 | 34.0 |
|  | Degrees Celsius | 7246 | 26.02 | 1.39 | 20.5 | 28.5 |
| *Time Pressure* | Minutes to next train | 32464 | 2.35 | 11.73 | 0 | 298.62 |
|  | Minutes between following trains | 32464 | 3.02 | 8.77 | 1 | 325.00 |
| *Product* | Price | 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 |
|  | Slots per product | 209801 | 1.15 | 0.37 | 1 | 5 |
| *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 is only of point club members. Junior is no greater than 30 and old is above 50. The availability of a product is the proportion of machine-time in which the product is available.

The main data set is directly obtained from the company's point-of-sales database. The data contain 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 tor 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.

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.[12]

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. Thus, endogeneity in price is not a major concern.

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.

---

[12]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
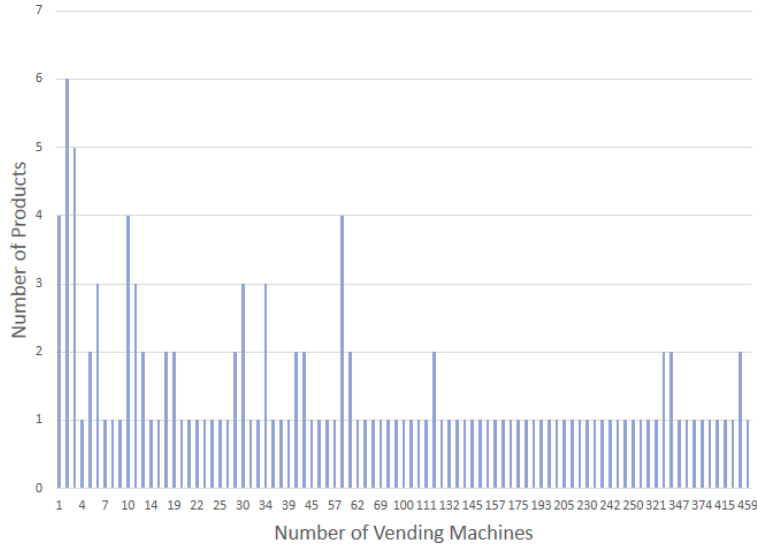
Figure 2: Product Availability: The horizontal axis is the number of vending machines that a particular product is available. The vertical axis is the number of products that corresponds to the number of machines that displays the product.

## 2.6 Key Findings from Kawaguchi et al (2018)

Because the findings of Kawaguchi, Uetake, and Watanabe (Forthcoming) motivate us to estimate the consideration set model that incorporates time pressure in the framework, we briefly discuss the key findings of our companion paper.

1. The first key finding is that recommendations increase not only the sales of recommended products, but also the sales of non-recommended products. A possible explanation of this effect would be the spillover effect of consumer attention to non-recommended products.

2. The second key finding is that the time to the next train, which captures the degree of time pressure, affects both machine-level and product-level sales. This finding implies that time pressure is likely to affect consumers' intention to buy a product and consumers' choices.

12

3. Finally, we find that the effect of recommendations is moderated as time pressure increases (the time to the next train shortens). We suspect that consumers respond to recommendations differently when they are under time pressure. This finding motivates us to explore whether such an effect results from attention or from preference.[13]

## 3  Model

To examine the effect of time pressure and recommendation on consumer attention and preference, we estimate a consideration set model in which 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 incorporate such behavioral effects into a choice model framework in a parsimonious way.

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

---

[13]On the other hand, crowd pressure, proxied by the number of passengers at the station, had only weak and not very robust impact on the effectiveness of product recommendations. Therefore, we drop that variable from the the current paper.

**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:

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

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 $\mathcal{C}_{it} \equiv \{j \in \mathcal{J}_t | C_{ijt} = 1\}$. The level of attention $V^*_{ijt}$ depends on the consumer and the product characteristics as follows:

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

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 $t$. $\alpha_{ij}$ is a consumer-level random effect in attention to product $j$, and $\zeta_j$ is a product-specific shock common across consumers, and $\varepsilon_{ijt}$ 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.[14] 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 co-

---

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

efficients 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. $\boldsymbol{\nu}_i$ follows an i.i.d. standard normal distribution. We assume that the off-diagonal terms of $\boldsymbol{\Sigma}$ are zero.

**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:

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

where we include a similar set of variables, $\boldsymbol{M}_t^U$ and $\boldsymbol{X}_{jt}^U$. $P_j$ is the logarithm price of product $j$, which affects only the utility and is excluded from the attention.[15] $\beta_{ij}$ is a consumer-level random effect in the utility of product $j$, and $\xi_j$ is a product-specific shock that is common across consumers. We assume that $\eta_{ijt}$ 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. $\boldsymbol{v}_i$ follows an i.i.d. standard normal distribution. We assume that the off-diagonal terms of $\boldsymbol{\Omega}$ are zero.

---

[15]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 attention to consider a product.

15

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

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

We assume that the error terms in $V_{ijt}^*$ and ones in $U_{ijt}^*$ are independent. This assumption is standard in the consideration set literature, such as Goeree (2008) and Van Nierop, Bronnenberg, Paap, Wedel, and Franses (2010).[16]

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 $\times$ 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 $\times$ Temp | Product category dummy $\times$ 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 $\times$ Volume | Interaction between container type and volume. |
| $A_{ijt}$ | PR | Indicator for whether product j is recommended. |
| | PR $\times$ T1 | Interaction between PR and T1 |
| $P_j$ | | Price of product $j$ |

**Empirical Specification**  Table 3 lists the variables we include in the estimation model. As discussed, the main departure of our consideration set model from the standard one comes

---

[16]We acknowledge that this assumption may be restrictive in some situations. Notable exceptions that do not rely on this assumption include Bronnenberg and Huang (forthcoming), Dehmany and Otter (2014), and Gaynor, Propper, and Seiler (2016). These papers consider specific form of dependence in attention.

from the menu- and context-related variables, $\boldsymbol{M}_t^V$, and $\boldsymbol{M}_t^U$. Note that these variables affect both attention and utility of all products. $M^V$ includes MR, NM, NR, T1, T2, while $M^U$ includes T1, T2, and Temp. 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. 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 and negative impacts on consumer 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, our companion paper shows 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 include T1 and T2 (and their interaction terms) in both attention and utility.

For the consumer utility, we interact Temp with product category dummies to see how different weather conditions affect the category of drinks consumers choose. We also include T1 and T2 in the utility because time pressure is likely to affect purchase probability, conditional on attention. We do *not* include MR, NM, and NR in the utility because we do not find any theoretical reason why they might shift consumer preference at the product level.

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. Cat, Shape, and Volume are product characteristics and are included for

both attention and utility.

Finally, 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, advertising may affect both consumer attention and preference through information provision and persuasion effects, respectively.

## 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 why could be either that the product is not preferred or that it was not considered. A typical approach to identification in the literature is to use ad hoc exclusion restrictions (e.g., Goeree (2008); Van Nierop, Bronnenberg, Paap, Wedel, and Franses (2010)).[17] In this approach, advertising, $A_{ijt}$, is usually excluded from utility.

We take an alternative approach because we are interested in the effects of recommendations on both attention and utility. A specification in which advertisement affects both attention and utility allows us to explore both informational persuasive roles of advertisement, as in Ching and Ishihara (2010) and Ching and Ishihara (2012). Hence, the identification of our model requires another source of identifying variation.

In this paper, we take advantage of the variation in product availability for identification. Intuitively speaking, whether a product is available or not affects consumer attention (because consumers cannot pay attention to unavailable products), whereas it does not directly shift consumers' utility of that product. In other words, the product (un)availability works as a sort of excluded variable that affects only attention without influencing utility.

The idea of using product availability to identify consumer attention and utility is based

---

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

on the recent decision theory literature on the consideration set models, such as Masatlioglu, Nakajima, and Ozbay (2012) and Manzini and Mariotti (2014). These papers show how changes in feasible sets and associated changes in choices help one to identify consumer attention and utility separately.[18] We provide some intuitive examples and more formal discussions in the online appendix. Although the examples are rather simple, the key insight of this literature is that changes in assortment and corresponding changes in choices (choice probabilities) provide us with useful restrictions to identify the consideration set model. With this insight, we have no need to make ad hoc exclusion restrictions on advertising, given the rich variation in product availability in our setup.

Another key feature of our consideration set model is that it includes menu variables and context variables, such as the number of products sold and the proxy variable for time pressure. In general, it is not easy to obtain the variables that capture context variables outside of the laboratory, and it is not straightforward to calculate relevant menu-related variables as in our vending machine setup. Menu-related variables in our model also serve as excluded variables that help identification because these variables arguably do not affect utility. 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 shifted by the menu variables, may not play an important role in our setup.

One may be concerned that the variation in product availability is already used to identify the non-linear parameters of utility. However, this is not the case. Identification of the random coefficients requires rich variation in the market- or consumer-choice-level characteristics (Berry and Haile (2011); Fox, Kim, Ryan, and Bajari (2012)), but not necessarily the variation in product availability per se. As long as there are enough variables altering the market shares,

---

[18]For example, suppose that a customer purchases beverage A but she switches to beverage B when another beverage C is not available. For simplicity, suppose for now that there are no stochastic shocks. Then, this *choice reversal* is caused by the change in her consideration set, and we can infer that beverage C is in her consideration set. This is because if she is not aware of beverage C in the first case, she faces exactly the same choice situation and must choose the same beverage or the choice reversal will not happen. Once we know that she is aware of beverage C, we can infer that she prefers beverage A to C simply because she selects beverage A over C.

the non-linear parameters are identified without the variation in the product availability. Of course, the over-identifying restrictions from the product availability still helps identification of the non-linear parameters of utility.

The effect of advertising 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 is made exogenously, there is no explicit reason why advertising is correlated with unobserved consumer heterogeneity.[19]

Lastly, we assume that attention probabilities are independent, *conditional on the observables.* 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 exist data that contain the set of all possible combinations of goods, we can identify the consideration set formation process completely nonparametrically. We provide a more formal discussion in the online appendix.

## 4.2   Simulated Minimum Distance Estimator

We estimate the consideration set model discussed in Section 3 with a simulated minimum distance estimator, which matches the empirical market share of each product to the associated simulated market share. In this subsection, we describe the estimation procedure. We demonstrate the validity of our estimation strategy using Monte Carlo simulations in the online appendix.

Given the realization of random effects, $\{\zeta_j, \xi_j\}_{j \in \mathscr{J}}$, $\{\alpha_{ij}, \beta_{ij}\}_{i \in \mathscr{I}, j \in \mathscr{J}}$, $\{\nu_i, \upsilon_i\}_{i \in \mathscr{I}}$ 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 fol-

---

[19]In Kawaguchi, Uetake, and Watanabe (Forthcoming), we confirm that the allocation is actually not correlated with key variables such as demographics and sales patterns.

lows:

$$p_{ijt} \equiv \mathbb{P}\{D_{ijt} = 1\}$$

$$= \sum_{\mathscr{C}_{it} \subset \mathscr{J}_t} \mathbb{P}\{U^*_{ijt} \geq \max_{k \in \mathscr{C}_{it}}\{U^*_{ikt}\}\} \prod_{l \in \mathscr{C}_{it}} \mathbb{P}\{C_{ilt} = 1\} \prod_{m \notin \mathscr{C}_{it}} \mathbb{P}\{C_{imt} = 0\} \quad (4.1)$$

$$\equiv \sum_{\mathscr{C}_{it} \subset \mathscr{J}_t} \pi_{ijt}(\mathscr{C}_{it}) \prod_{l \in \mathscr{C}_{it}} \gamma_{ilt} \prod_{m \notin \mathscr{C}_{it}} (1 - \gamma_{imt}).$$

Because we assume that $\varepsilon_{ijt}$ and $\eta_{ijt}$ in equations (3.2) and (3.4) follow a Type 1 extremum value distribution, respectively, we have

$$\pi_{ijt}(\mathscr{C}_{it}) = \frac{\exp(U_{ijt})}{1 + \sum_{k \in \mathscr{C}_{it}} \exp(U_{ikt})}, \text{ and } \gamma_{ijt} = \frac{\exp(V_{ijt})}{1 + \exp(V_{ijt})}.$$

Then, the expected choice probability is $\int \boldsymbol{p} \, dF(\{\zeta_j, \xi_j\}_{j \in \mathscr{J}}, \{\alpha_{ij}, \beta_{ij}\}_{i \in \mathscr{I}, j \in \mathscr{J}}, \{\boldsymbol{v}_i, \boldsymbol{v}_i\}_{i \in \mathscr{I}})$, where $\boldsymbol{p} = \{p_{ijt}\}_{j \in \mathscr{J}, i \in \mathscr{I}, t \in \mathscr{T}}$, which is approximated by simulations with the following steps. First, we simulate the consumer- and product-level shocks in equations (3.2)–(3.5) for $N_1$ times. We draw $\boldsymbol{v}_i$ and $\boldsymbol{v}_i$ from a standard normal distribution $N_1$ times for each consumer, and $\alpha_{ij}$ and $\beta_{ij}$ from a standard normal distribution $N_1$ times for each consumer and product, and $\zeta_j$ and $\xi_j$ from a standard normal distribution $N_1$ time s for each product.

Second, we draw $\varepsilon_{ijt}$ from Type I extreme value distribution $N_2$ times for each consumer, product, and purchase occasion. Using these simulated draws, we construct $N_2$ simulated consideration sets (each of which is denoted by $n_2$) for each simulated random coefficient (denoted by $n_1$), $\mathscr{C}_{it}^{(n_1, n_2)}$ by $\mathscr{C}_{it}^{(n_1, n_2)} = \{j \in \mathscr{J}_t | C_{ijt}^{(n_1, n_2)} = 1\}$, using equations 3.2 and 3.3.

Third, given the simulated consideration sets, we calculate the conditional choice probabilities for each consumer, product, purchase occasion, and simulated shocks by

$$\pi_{ijt}(\mathscr{C}_{it}^{(n_1, n_2)}) = \frac{\exp(U_{ijt}^{(n_1)})}{1 + \sum_{k \in \mathscr{C}_{it}^{(n_1, n_2)}} \exp(U_{ikt}^{(n_1)})}, \quad (4.2)$$

where $U_{ijt}^{n_1}$ is calculated based on equations (3.4) and (3.5).

21

Fourth, using equation (4.2), we derive the choice probabilities for each consumer, product, purchase occasion, and the same individual-level shocks by

$$p_{ijt}^{(n_1)} = \frac{1}{N_2} \sum_{n_2=1}^{N_2} \sum_{\mathscr{C}_{it} \subset \mathscr{I}_t} \pi_{ijt}(\mathscr{C}_{it}^{(n_1,n_2)}),$$  (4.3)

and the expected choice probabilities for each consumer, product, and purchase occasion is

$$s_{ijt} = \frac{1}{N_1} \sum_{n_1=1}^{N_1} p_{ijt}^{(n_1)}.$$

Finally, we match the simulated expected choice probabilities with the actual choice data.[20] Because we do not observe the consumers who visited the vending machine but did not make a purchase, we assume that the number of consumers who visited the vending machines in a station during a time period is proportional to the number of passengers in the station during the time period, which we call the market size of purchase occasion $t$ denoted by $S_t$.[21] Let $\mathscr{I}_t$ be the set of consumers who purchased a product in purchase occasion $t$ and $N_t$ be its size. Then, we construct the following distance measure in terms of the inside market share and the choice probability of the outside option:

$$\sum_{t \in \mathscr{T}} \sum_{i \in \mathscr{I}_t} \sum_{j \in \mathscr{I}_t} [D_{ijt} - s_{ijt}/(1 - s_{i0t})]^2 + \sum_{t \in \mathscr{T}} [(S_t - N_t)/S_t - s_{0t}]^2,$$

where $s_{0t}$ is the predicted choice probability of the outside option evaluated at the average covariates in the purchase occasion $t$. The first term matches the actual and simulated inside choice probabilities, and the second term matches the choice probabilities of the outside

---

[20]We set $N_1 = 16$ and $N_2 = 100$. In the Monte Carlo simulation details provided in the online appendix, we report the results of the sensitivity analysis in which we change $N_1$ and $N_2$ and find that the results are barely affected.

[21]We estimate the number of passengers at the station where the vending machine is installed, which is calculated from the 2010 Metropolitan Transportation Census of the Ministry of Land, Infrastructure, and Tourism (MLIT). The details can be found in Kawaguchi, Uetake, and Watanabe (Forthcoming). The market sizes are used to adjust the difference in the consumer size across vending machines, so only the relative sizes are important. The mismeasurement in the absolute level of the market size is reflected only in the estimated size of the intercept parameter and does not affect the counterfactual prediction.

option. We find parameters that minimize the distance and obtain standard errors by boot-straping at the market level. We repeat sub-sampling of size 1000 at the purchase occasion level 100 times and use the average as the estimate and its 2.5 and 97.5 percentiles as the 95% confidence interval.

## 5   Estimation Results

### 5.1   Parameter Estimates

We report the estimated coefficients in Table 4. 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. The coefficient on recommendation spillover (*MR*) in attention is positive, which implies that when the machine recommends some products, consumers pay more attention to each product in the machine, including non-recommended ones. 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, again, the 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 attention, but negative for utility. Hence, consumers are more aware of

23

Table 4: The Estimation Result of Parameters

| Attention | Estimate | 95% C.I. | Utility | Estimate | 95% C.I. |
|---|---|---|---|---|---|
| PR | 0.106 | [0.104, 0.107] | PR | 0.089 | [0.085, 0.090] |
| MR | 0.083 | [0.080, 0.084] | NS | 0.100 | [-0.029, 0.136] |
| NR | 0.021 | [0.018, 0.023] | price | -1.670 | [-1.677, -1.665] |
| $NR^2$ | -0.216 | [-0.221, -0.213] | PR × T1 | 0.007 | [0.003, 0.044] |
| PR × T1 | 0.128 | [0.126, 0.131] | T1 | -0.037 | [-0.048, -0.004] |
| NR × T1 | 0.095 | [0.094, 0.098] | T2 | -0.020 | [-0.025, 0.000] |
| NS | 0.083 | [0.081, 0.086] | Green tea | 0.011 | [0.007, 0.015] |
| NM | -0.100 | [-0.102, -0.097] | Other tea | 0.035 | [0.033, 0.038] |
| T1 | 0.133 | [0.131, 0.136] | Water | 0.007 | [0.002, 0.011] |
| T2 | 0.098 | [0.097, 0.101] | Sport | 0.035 | [0.034, 0.038] |
| Green tea | 0.028 | [0.026, 0.031] | Canned coffee | 0.001 | [-0.004, 0.004] |
| Other tea | -0.006 | [-0.008, -0.003] | Bottled coffee | 0.014 | [0.008, 0.018] |
| Water | 0.038 | [0.037, 0.042] | Black tea | 0.006 | [-0.004, 0.010] |
| Sport | 0.028 | [0.027, 0.032] | Carbonated | 0.020 | [0.016, 0.024] |
| Canned coffee | 0.032 | [0.031, 0.035] | Fruit | -0.006 | [-0.007, -0.003] |
| Bottled coffee | 0.036 | [0.035, 0.040] | Healthy | 0.011 | [0.010, 0.015] |
| Black tea | 0.360 | [0.359, 0.363] | Other drink | 0.020 | [0.019, 0.023] |
| Carbonated | 0.031 | [0.030, 0.035] | Green tea × Temp | 0.027 | [0.025, 0.028] |
| Fruit | 0.026 | [0.025, 0.029] | Other tea × Temp | 0.026 | [0.025, 0.026] |
| Healthy | 0.012 | [0.011, 0.016] | Water × Temp | 0.029 | [0.027, 0.029] |
| Other drink | 0.033 | [0.031, 0.036] | Sport × Temp | 0.027 | [0.026, 0.028] |
| | | | Canned coffee × Temp | 0.023 | [0.021, 0.024] |
| | | | Bottled coffee × Temp | 0.037 | [0.035, 0.038] |
| | | | Black tea × Temp | 0.009 | [0.007, 0.009] |
| | | | Carbonated × Temp | 0.019 | [0.018, 0.020] |
| | | | Fruit × Temp | 0.021 | [0.020, 0.022] |
| | | | Healthy × Temp | 0.024 | [0.022, 0.024] |

*Note:* The estimates are the average of the sub-sampling estimates. The 95% confidence interval is the 2.5 and 97.5 percentiles of the sub-sampling estimates. The size of a sub-sample is 1000. The sub-sampling is repeated 100 times. We omit some estimates from the table to save the space.

Table 5: The Estimation Result of Demography Dependence

| | Gender | PR in Attention | | MR in Attention | | PR in Utility | |
|---|---|---|---|---|---|---|---|
| | | Estimate | 95% C.I. | Estimate | 95% C.I. | Estimate | 95% C.I. |
| *Junior* | Female | 0.190 | [0.185, 0.194] | 0.165 | [0.161, 0.169] | 0.177 | [0.172, 0.179] |
| | Male | 0.276 | [0.269, 0.281] | 0.217 | [0.210, 0.222] | 0.245 | [0.236, 0.249] |
| *Senior* | Female | 0.106 | [0.104, 0.107] | 0.083 | [0.080, 0.084] | 0.089 | [0.085, 0.090] |
| | Male | 0.192 | [0.188, 0.195] | 0.134 | [0.128, 0.136] | 0.157 | [0.150, 0.160] |
| *Old* | Female | 0.179 | [0.175, 0.183] | 0.171 | [0.166, 0.175] | 0.327 | [0.318, 0.330] |
| | Male | 0.265 | [0.259, 0.270] | 0.222 | [0.216, 0.227] | 0.395 | [0.382, 0.399] |

*Note:* The estimate is the average of the sub-sampling estimates. The 95% confidence interval is the 2.5 and 97.5 percentiles of the sub-sampling estimates. The size of a sub-sample is 1000. The sub-sampling is repeated 100 times. Junior is younger than 30, senior is between the ages of 30 and 50, and old is 50 or older. The values are mean coefficients at each demographic.

products that occupy more physical space in the vending machine, but they are less attracted to buying them.

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

More importantly, the effectiveness of product recommendations is also affected by time pressure. We find that coefficients on PR× T1 and PR× T2 are 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 our previous paper Kawaguchi, Uetake, and Watanabe (Forthcoming) and suggests that consumers do not want to be told what to buy when they are in hurry.

Table 5 reports the heterogeneity in the coefficients of product recommendations on attention and preference. Our model includes the gender and age of customers as categorical variables in the random coefficients and the table reports the mean coefficient for each cate-

gory of consumers. We find significant heterogeneity in the effects of product recommendations across different demographic groups. An interesting finding is that female consumers are less affected by recommendations in terms of both attention and utility across all age groups. Another interesting result is that old consumers tend to feel greater utility from purchasing recommended products. One of the interesting managerial questions that arises 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.

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 $s_{jt}^0$. 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 $s_{kt}^{1j}$. Based on these notations, we define the *own elasticity* of the choice probability by $\frac{s_{jt}^{1j} - s_{jt}^0}{s_{jt}^0}$ and the *cross elasticity* by $\frac{\sum_{k \neq j, 0}(s_{kt}^{1j} - s_{kt}^0)}{\sum_{k \neq j, 0} s_{kt}^0}$. Note that we sum up choice probabilities of non-recommended products $k$ here. Finally, we define the *total elasticity* by $\frac{\sum_{k \neq 0}(s_{kt}^{1j} - s_{kt}^0)}{\sum_{k \neq 0} s_{kt}^0}$, 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.[22]

**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

---

[22]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.

Table 6: Mean Elasticity of Choice Probabilities to Product Recommendation

|  | Total | Own | Cross |
|---|---|---|---|
| *Full* | 0.080 | 0.454 | 0.067 |
| *Attention channel only* | 0.070 | 0.181 | 0.066 |
| *Utility channel only* | 0.008 | 0.225 | 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.

machine sales by 8%, the sales of *recommended* products by 45.4%, and the sales of *non-recommended products* by 6.7%.[23]

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 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 0.07 when recommendations affect only attention and 0.008 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 0.181 when recommendations only affect attention and 0.225 when recommendations only affect utility. Thus, the effect of recommendations on *a recommended product* works mainly through the utility channel.[24]

**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

---

[23]These elasticities are not as high 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%.

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

Table 7: Total Elasticity of Choice Probabilities to Product Recommendation by Time Pressure

| | (a) Full | | | | (b) Attention channel only | | | | (c) Utility channel only | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *Time* | *0* | *5* | *10* | | *0* | *5* | *10* | | *0* | *5* | *10* |
| *all* | 0.087 | 0.06 | 0.025 | | *0.0777* | 0.0503 | 0.0147 | | *0.007* | 0.008 | 0.01 |
| *own* | 0.406 | 0.42 | 0.325 | | *0.1613* | 0.1335 | 0.0229 | | *0.206* | 0.249 | 0.294 |
| *cross* | 0.075 | 0.047 | 0.014 | | *0.0748* | 0.0473 | 0.0144 | | *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.

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, as time pressure weakens, total elasticity decreases, own elasticity increases, and cross elasticity decreases.

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 own elasticities, cross elasticities, and overall elasticities all increase as time pressure increases. We find opposite patterns in panel (c); both overall and own elasticities increase as time pressure declines due to a lack of a spillover effect.

# 6 Designing Revenue-Maximizing Recommendations

Using the estimates of the consideration set model, we conduct several counterfactual simulations to derive managerial implications. In particular, we consider how many products the company should recommend under time pressure. 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 revenue-maximizing number of recommen-

dations is an important managerial question. In this exercise, we study how machine-level sales change by the number of recommended products and their sensitivity to time pressure. Hereafter, we regard "revenue maximization" as "optimization".[25]

To avoid computationally intractable combinatorial optimization problems, we do not consider the optimal combination of recommended products. Instead, we randomly select which products to recommend, given how many products to recommend. More precisely, we first randomly order available products ($j = \{1, ..., J_t\}$)[26] for each occasion. 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.

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



(a)    (b)
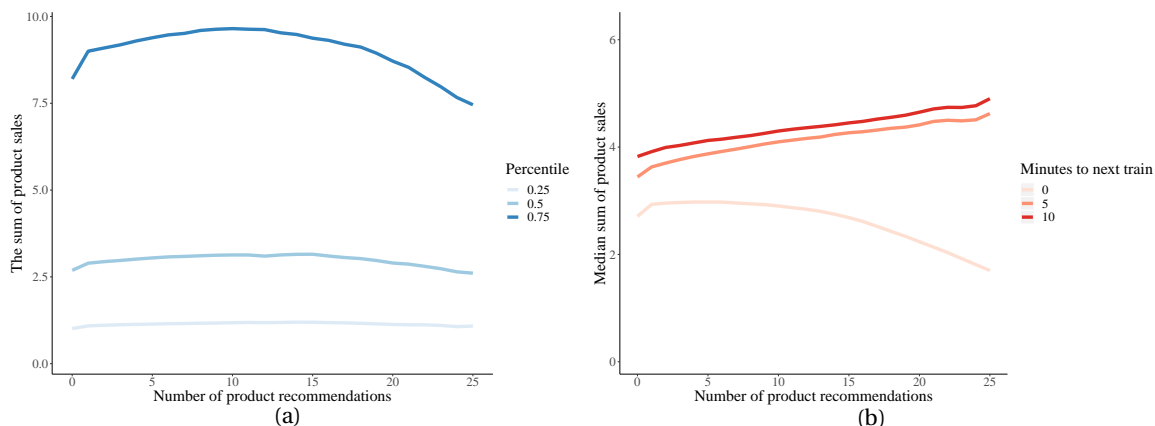
*Note:* For each purchase occasion, begin with a setting in which there is no product recommendation and then randomly pick up a product to recommend. Continue this process until all products in the vending machine are recommended. At each number of product recommendations, compute the machine-level 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 Pressures**    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

---

[25]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.

[26]$J_t$ is the number of available products in vending machine $k$ at occasion $t$.

Figure 4: Machine-level Sales by the Number of Recommendations: Attention Channel Only
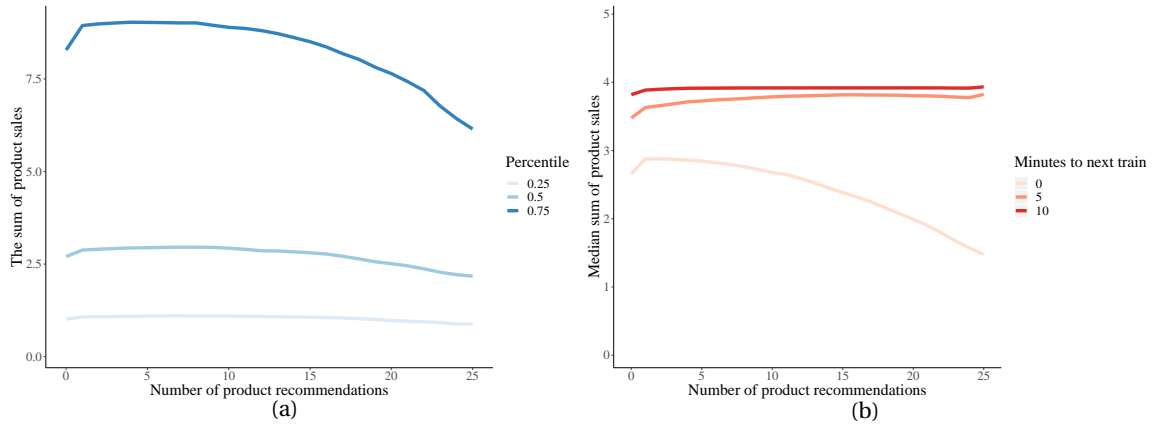


*Note:* The settings are the same as those in Figure 3 except that the utility channel is suppressed.

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 11 at the median.

**Optimal Recommendations Under Counterfactual Time Pressure**     Then, 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.

**Optimal Recommendations When Recommendation Only Affect Attention**     Although in Figure 3 we use the full model in which recommendations affect both attention and utility, it is unclear how attention and utility affect the result. To understand how the optimal number of recommendations changes when recommendations can affect only attention, in Figure 4, we show the relationship between the number of recommendations and sales at different quar-

tiles under the actual time pressure (panel (a)) and at different degrees of time pressures when recommendations do not affect utility.

Panel (a) shows that the optimal number of recommendations is smaller when recommendations have no impact on utility. Panel (b) shows that sales do not change much by the number of recommendations when there is little time pressure. Under high time pressure, however, it is optimal to recommend only two products, which is a smaller number than in the case in Figure 4.

**Optimality of Contextual Recommendation**  Finally, we consider how much context-based recommendations can improve the total 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 the context or attribute. Second, we consider the *optimal attribute-based targeting*, in which the number of recommendations is optimized for each consumer segment (gender × age class) level. The targeting policy is similar to traditional targeting based on observable characteristics. The third policy is the *optimal contextual policy*, which recommends the revenue-maximizing number of recommendations for each vending machine and time of day. 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.

Table 8 reports the results, which includes the sales units under each policy in the first column, the percenetage change in the sales in the second column, and the mean and standard deviation of the optimal number of recommendations in the third and fourth columns. Note that the sample for the counterfactual simulations consist of the customers with membership, not the entire customers.

31

We find that the uniform policy can improve sales by 2.4% relative to the current policy and it recommends 9 products. The targeting policy slightly increases the sales volume but only by an additional 0.2 percentage points relative to the uniform policy. Additionally, the targeting policy recommends slightly more products on average. Lastly, the contextual policy can achieve even higher performance with 4.5% of sales increase, or 2.1% more than the uniform policy. The contextual policy recommends 13.7 products on average. Therefore, we find that the contextual policy outperforms the traditional attribute-based recommendation policy in our setup.

Table 8: Optimal Number of Recommendations and Impacts on Sales Volume

|  | Sales volume | | NR | |
| --- | --- | --- | --- | --- |
|  | Units | Change | Mean | S.D. |
| *Actual* | 40838 | 1.000 | 4.48 | 1.70 |
| *Uniform* | 41823 | 1.024 | 9.00 | - |
| *Target* | 41893 | 1.026 | 9.67 | 1.63 |
| *Contextual* | 42666 | 1.045 | 13.69 | 7.13 |

Note that there is a limitation in our counterfactual simulation. In our analysis, we focus on the number of recommended products, but we do not optimize which products to recommend, because it is computationally infeasible to search across all possible combinations. The company can improve even more by carefully designing the optimal combination of products for recommendations. Our counterfactual simulation provides only the lower bound of the potential impacts of context-based marketing.

## 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 question, 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 a 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. We study how many products should be recommended depending on time pressure. Our results show that the revenue maximizing number of recommendations is approximately 10 if the distribution of time pressure is at the actual level, while as the degree of time pressure decreases, the revenue-maximizing number of recommendations increases. Finally, we find that the context-based recommendation system can outperform the traditional attribute-based recommendation system.

## 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.

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., 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.

DE CLIPPEL, G., AND K. ROSEN (2014): "Bounded Rationality and Limited Datasets," *mimeo.*

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.

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

GAYNOR, M., C. PROPPER, AND S. SEILER (2016): "Free to Choose? Reform, Choice and Consideration Sets in the English National Health Service," *American Economic Review*, 106, 3521–57.

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 (Forthcoming): "Effectiveness of Product Recommendation under Time and Social Pressures," *Marketing Science*.

LEWBEL, A. (2000): "Semiparametric Qualitative Response Model Estimation with Instrumental Variables and Unknown Heteroscedasticity," *Journal of Econometrics*, 97, 145–77.

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): "Stochastic Choice and Consideration Sets," *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.

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

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.

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.

# 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    Identification by Product Availability

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

## B.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 con-

| Case | 1 | 2 |
|---|---|---|
| Feasible Sets | 1, 2, 3 | 1, 2 |
| Choices | 1 | 2 |

Table 1: 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.

sideration 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 sets of products are available, as in Table 1. 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.[27] 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,

---

[27] 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.

sometimes preference or attention are only partially identified or are not identified.[28]

**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.

## B.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 $J' = \{1,2\}$.

---

[28] 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.

Then, the market share of $j = 1$ for feasible set $J$ is

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

where $\pi_1(\{1\})$ is the conditional choice probability of product 1 given consideration set $C = \{0, 1\}$, and $\gamma_1$ is the probability of having consideration set $C = \{0, 1\}$ given $J = \{1\}$.[29] Similarly, the market share of product 1 for feasible set $J'$ is written as

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

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 (B.1) and (B.2), we obtain the following:

$$\underbrace{\frac{s_1(J) - s_1(J')}{s_1(J)}}_{\text{\%-change in product 1's market share}} = \gamma_2 \left[ 1 - \frac{\pi_1(\{1, 2\})}{\pi_1(\{1\})} \right]$$

$$= \underbrace{\gamma_2}_{\text{Attention probability of product 2}} \tag{B.3}$$

$$\times \underbrace{\frac{\pi_1(\{1\}) - \pi_1(\{1, 2\})}{\pi_1(\{1\})}}_{\text{\%-change in product 1's conditional choice probability}}.$$

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.

---

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

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 $\gamma_1$ does not appear in the equation, as it is canceled out. In addition, note that $\gamma_2$ does not appear in the second term on the right-hand side. Hence, if there is sufficient variation in the variable that shifts only attention to product 2, one can *nonparametrically* identify $\gamma_2$. 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. The formal proposition is the following.

**Proposition 1.** *Suppose there exist feasible sets $J$ and $J_j = J \setminus \{j\}$ 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

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

Now, we can identify $\gamma_3$ from choice observations from $J_3 = \{1, 2\}$. We can write $s_1(J_3)$ as follows:

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

Combining Eqs. B.4 and B.5 yields

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

Note that $\gamma_1$ does not appear on the right-hand side of Eq. (B.6) and that $\gamma_3$ 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, $\gamma_3$, and %-change of product 1's conditional choice probability when product 3 is added. The only difference from Eq. (B.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. (B.6), driving $Z_3 \to +\infty$, we obtain $\frac{s_1(J)-s_1(J_3)}{s_1(J_3)} \to \gamma_3$.[30] This identifies $\gamma_3$ nonparametrically. Using $J_1$ and $J_2$ with $J$, similarly, we can nonparametrically identify $\gamma_1$ and $\gamma_2$, respectively. Thus, all attention parameters are identified.[31] Note that it is straightforward to identify parameters in $\gamma$s 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.

With more than two products in the feasible set, the decomposition in Eq. (B.3) becomes slightly more complicated, but similar intuition works. With $J = \{1,2,3\}$, it is possible to show

---

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

[31]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 $\gamma_k$ for $k \in J \setminus \{j\}$, 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 $\gamma_j$.

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 shifts only $gamma_3$, we can identify $\gamma_3$.

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 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 $\gamma$s. When we observe more variations in feasible sets, such as $J_{12}$, $J_{13}$, and $J_{23}$, we can have more restrictions to identify $\gamma_j$. 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.

## B.3 Identification with Exclusion Restrictions

In this section, we provide a proof of identifying the consideration set model with exclusion restrictions. Although not explicitly mentioned, the existing papers use the 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 $\pi_{jt}$, 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 Lewbel (2000), Berry and Haile (2011), and Fox, Kim, Ryan, and Bajari (2012). ∎

Observe that the exclusion restriction is crucial for this identification. If $A_{ijt}$ also impacts preference, as in our model, both $\pi_j$ and $\gamma_j$ 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 $A_{ijt} \to +\infty$.

45

Table 2: 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.

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, we use $A_{ijt}$ for the exclusion restriction, but it is not necessary. We can use other variables such as menu-related variables to the extent that these variables are excluded from the utility.

# Appendix C   Additional Results

**Elasticities on Consideration Probability**   Table 2 presents the elasticity of recommendations to the probability that a product is included in the consideration set instead of the purchase incidence. As in Table 6, we calculate own and cross elasticities. We find that recommendations increase the probability of being considered by 19.1% for recommended products and 6% for non-recommended products.

Table 3 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.

Table 3: 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 2 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.

# Appendix D    Monte Carlo Simulation

In the main text, we set the number of simulated shocks at $N_1 = 16$ and $N_2 = 100$ to compute the simulated choice probabilities. In this section, we study the sensitivity of the estimator to the number of shocks to simulate the choice probabilities. In our empirical application, there are about 30 products in each vending machine, and hence, the number of potential consideration sets could be very large. Because it is computationally infeasible to draw a large number of consideration sets for each step of the estimation routine, we estimate a version of the consideration set model we study in the main text under different number of simulation draws.

To do so, we consider the following version of the model. Because the number of consideration sets depends on the number of products in a vending machine, we set it similar to the actual number of 30. To make the computation feasible, we scale down the other dimensions of the problem: there are 100 products (including the outside option), 2 purchase occasions, and 100 consumers. The latent attentions and utilities are as follows:

$$V_{ijt}^* = \alpha_{PRi} PR_{jt} + \alpha_{MRi} MR_{jt} + \alpha_{T1} T1_t + \alpha_{PR,T1} PR_{jt} \times T1_t + \alpha_{ij} + \zeta_j + \epsilon_{ijt},$$

with:

$$\alpha_{PRi} = \alpha_{PR} + \alpha_{PR,Male} Male_i \sigma v_{PRi},$$

$$\alpha_{MRi} = \alpha_{MR} + \alpha_{MR,Male} Male_i + \sigma v_{MRi},$$

and:

$$U_{ijt}^* = \beta_{PRi} PR_{jt} + \beta_{T1} T1_t + \beta_{PR,T1} PR_{jt} \times T1_t + \beta_{ij} + \xi_j + \eta_{ijt},$$

with:

$$\beta_{PRi} = \beta_{PR} + \beta_{PR,Male} Male_i \sigma v_{PRi}.$$

We fix the parameters and exogenous variables. We then draw shocks 100 times to replicate the data. We then estimate the parameters using the same procedure as we discussed in Section 4.2 for simulated data with varying numbers of $N1$ and $N2$. Specifically, in addition to $(N1, N2) = (16, 100)$, we used $(N1, N2) = (16, 1000)$ and $(N1, N2) = (160, 100)$ to study the sensitivity of the estimator to the number of simulations. Finally, we computed the Monte Carlo mean squared errors for each pair of $(N1, N2)$ and decomposed them into the bias and variance terms.

The results are summarized in Table 4. We find that the parameter estimates do not change much even if we increase the number of simulations. The mean squared errors are small for most of the parameters and specifications. Although the mean squared errors are relatively large for some parameters such as the standard deviation of $\xi_{jt}$, the main parameters of interest such as PR and PR × T1 are precisely estimated for each specification. Hence, our estimates and conclusions are robust to the number of simulations; because there is a small probability that a large number of consideration sets will be realized, they do not contribute to the objective function much.

Table 4: Monte Carlo Mean Squared Errors of the Estimators

| | variable | MI | MC | True | Estimate | Bias | Variance | Mse |
|---|---|---|---|---|---|---|---|---|
| *Utility* | | | | | | | | |
| 1 | T1 | 16 | 100 | 0.17921 | 0.16100 | 0.00033 | 0.00442 | 0.00476 |
| 2 | T1 | 16 | 1000 | 0.17921 | 0.16877 | 0.00011 | 0.00633 | 0.00644 |
| 3 | T1 | 160 | 100 | 0.17921 | 0.15819 | 0.00044 | 0.00477 | 0.00521 |
| 4 | PR × T1 | 16 | 100 | 0.03636 | 0.06053 | 0.00058 | 0.00500 | 0.00558 |
| 5 | PR × T1 | 16 | 1000 | 0.03636 | 0.06647 | 0.00091 | 0.00747 | 0.00838 |
| 6 | PR × T1 | 160 | 100 | 0.03636 | 0.05895 | 0.00051 | 0.00637 | 0.00688 |
| 7 | PR × Male | 16 | 100 | 0.00250 | 0.02546 | 0.00053 | 0.00687 | 0.00740 |
| 8 | PR × Male | 16 | 1000 | 0.00250 | 0.06498 | 0.00390 | 0.02288 | 0.02679 |
| 9 | PR × Male | 160 | 100 | 0.00250 | 0.05100 | 0.00235 | 0.02090 | 0.02325 |
| 10 | PR | 16 | 100 | 0.06103 | 0.07116 | 0.00010 | 0.00306 | 0.00316 |
| 11 | PR | 16 | 1000 | 0.06103 | 0.07604 | 0.00023 | 0.01045 | 0.01068 |
| 12 | PR | 160 | 100 | 0.06103 | 0.08401 | 0.00053 | 0.01403 | 0.01456 |
| *Attention* | | | | | | | | |
| 13 | T1 | 16 | 100 | 0.01128 | 0.03304 | 0.00047 | 0.00285 | 0.00332 |
| 14 | T1 | 16 | 1000 | 0.01128 | 0.08184 | 0.00498 | 0.02891 | 0.03389 |
| 15 | T1 | 160 | 100 | 0.01128 | 0.05420 | 0.00184 | 0.01027 | 0.01211 |
| 16 | PR × T1 | 16 | 100 | 0.07733 | 0.07987 | 0.00001 | 0.00445 | 0.00446 |
| 17 | PR × T1 | 16 | 1000 | 0.07733 | 0.09400 | 0.00028 | 0.01873 | 0.01901 |
| 18 | PR × T1 | 160 | 100 | 0.07733 | 0.07523 | 0.00000 | 0.00667 | 0.00668 |
| 19 | PR × Male | 16 | 100 | 0.00655 | 0.01874 | 0.00015 | 0.00170 | 0.00185 |
| 20 | PR × Male | 16 | 1000 | 0.00655 | 0.03300 | 0.00070 | 0.00507 | 0.00577 |
| 21 | PR × Male | 160 | 100 | 0.00655 | 0.01701 | 0.00011 | 0.00100 | 0.00111 |
| 22 | PR | 16 | 100 | 0.11631 | 0.11801 | 0.00000 | 0.00251 | 0.00252 |
| 23 | PR | 16 | 1000 | 0.11631 | 0.10710 | 0.00008 | 0.00672 | 0.00680 |
| 24 | PR | 160 | 100 | 0.11631 | 0.11434 | 0.00000 | 0.00474 | 0.00474 |
| 25 | MR × Male | 16 | 100 | 0.01119 | 0.02045 | 0.00009 | 0.00084 | 0.00092 |
| 26 | MR × Male | 16 | 1000 | 0.01119 | 0.04502 | 0.00114 | 0.01072 | 0.01186 |
| 27 | MR × Male | 160 | 100 | 0.01119 | 0.02629 | 0.00023 | 0.00185 | 0.00208 |
| 28 | MR | 16 | 100 | 0.00993 | 0.01951 | 0.00009 | 0.00142 | 0.00151 |
| 29 | MR | 16 | 1000 | 0.00993 | 0.07876 | 0.00474 | 0.15595 | 0.16069 |
| 30 | MR | 160 | 100 | 0.00993 | 0.02385 | 0.00019 | 0.00170 | 0.00189 |
| *Standard deviations* | | | | | | | | |
| 31 | $\sigma$ | 16 | 100 | 0.10000 | 0.12759 | 0.00076 | 0.01118 | 0.01194 |
| 32 | $\sigma$ | 16 | 1000 | 0.10000 | 0.24060 | 0.01977 | 0.13501 | 0.15478 |
| 33 | $\sigma$ | 160 | 100 | 0.10000 | 0.18158 | 0.00666 | 0.03505 | 0.04171 |
| 34 | S.D. $\zeta_{jt}$ | 16 | 100 | 0.29220 | 0.29801 | 0.00003 | 0.00747 | 0.00751 |
| 35 | S.D. $\zeta_{jt}$ | 16 | 1000 | 0.29220 | 0.36155 | 0.00481 | 0.17298 | 0.17779 |
| 36 | S.D. $\zeta_{jt}$ | 160 | 100 | 0.29220 | 0.28807 | 0.00002 | 0.00661 | 0.00662 |
| 37 | S.D. $\xi_{jt}$ | 16 | 100 | 0.29659 | 0.28507 | 0.00013 | 0.02950 | 0.02963 |
| 38 | S.D. $\xi_{jt}$ | 16 | 1000 | 0.29659 | 0.29126 | 0.00003 | 0.08428 | 0.08431 |
| 39 | S.D. $\xi_{jt}$ | 160 | 100 | 0.29659 | 0.38092 | 0.00711 | 0.09547 | 0.10259 |
| 40 | S.D. $\beta_{ij}$ | 16 | 100 | 0.78317 | 0.75816 | 0.00063 | 0.05315 | 0.05378 |
| 41 | S.D. $\beta_{ij}$ | 16 | 1000 | 0.78317 | 0.81126 | 0.00079 | 0.14915 | 0.14994 |
| 42 | S.D. $\beta_{ij}$ | 160 | 100 | 0.78317 | 0.76206 | 0.00045 | 0.10159 | 0.10203 |
| 43 | S.D. $\alpha_{ij}$ | 16 | 100 | 1.46147 | 1.47538 | 0.00019 | 0.00567 | 0.00586 |
| 44 | S.D. $\alpha_{ij}$ | 16 | 1000 | 1.46147 | 1.60895 | 0.02175 | 1.30497 | 1.32672 |
| 45 | S.D. $\alpha_{ij}$ | 160 | 100 | 1.46147 | 1.48163 | 0.00041 | 0.00867 | 0.00908 |
| 46 | $\omega$ | 16 | 100 | 0.10000 | 0.07881 | 0.00045 | 0.00906 | 0.00950 |
| 47 | $\omega$ | 16 | 1000 | 0.10000 | 0.08781 | 0.00015 | 0.01219 | 0.01234 |
| 48 | $\omega$ | 160 | 100 | 0.10000 | 0.09378 | 0.00004 | 0.01613 | 0.01617 |

*Note:* The Monte Carlo simulation is repeated for 100 times. The estimate is the simulation average and bias, variance, and mean-squared-errors are due to the simulation distribution.