Consumer Misinformation and the Brand Premium: A Private Label Blind Taste Test*

Bart J. Bronnenberg, *CentER, Tilburg University, and CEPR*
Jean-Pierre Dubé, *University of Chicago Booth School of Business, and NBER*
Robert E. Sanders, *University of California, San Diego Rady School of Management*

September 16, 2019

Abstract

To study consumer brand misinformation, we run in-store blind taste tests with a retailer's private label food brands and the leading national brand counterparts in three large CPG categories. Subjects self-report very high expectations about the quality of the private labels relative to national brands. However, they predict a relatively low probability of choosing them in a blind taste test. An overwhelming majority systematically chooses the private label in the blinded test. Using program evaluation methods, we find that the causal effect of this intervention on treated consumers increases their market share for the tested private label product by 15 share points during the week after the intervention, on top of a base share of 8 share points. However, the effect diminishes to 8 share points during the second to fourth week after the test and to 2 share points during the second to fifth month after the test. Using a structural model of demand that controls for the self-selected participation and allows for heterogeneous treatment effects, we show these effects survive controls for point-of-purchase prices, purchase incidence, and the feedback effects of brand loyalty. We also find that the intervention increases the preference for the private label brands, and that it decreases the preference for the national brands, relative to the outside good. Interpreting the intervention as an information treatment about the product, we find evidence consistent with an economically large informational barrier on demand for the private label product relative to an established national brand.

keywords: private label, consumer information, brands and branding, market structure, self-selection

JEL: L11, L15, M31, M37

1 INTRODUCTION

Private label brands in the consumer packaged goods (CPG) industry are still relatively under-developed in the United States (US) relative to other western economies. According to a 2014 global Nielsen survey, private labels accounted for only 18% of US CPG sales, which is comparable to the weighted global average of 16.5% but much smaller than shares exceeding 40% in European countries like the United Kingdom, Switzerland and Spain.[1] In Europe, private labels represent $1 of every $3 spent on CPG. In spite of the spending gap between the US and Europe, survey evidence suggests that there is no gap in private label quality perceptions: 75% of US respondents agreed with the statement "Private Labels are a good alternative to name brands." In Europe, the rate is comparable at 69%. In spite of perceptions, US consumers routinely pay a large price premium for national brands. Recent research finds that US consumers could save $44 billion annually by switching to a comparable store brand when available (Bronnenberg et al., 2015).

We investigate the extent to which US consumers' willingness-to-pay a CPG brand price premium relative to private label alternatives is driven by an information barrier. In cooperation with Mariano's, a large, mid-western supermarket chain, we conduct a series of in-store blind taste tests that match the chain's private label products against the leading national brand competitors in several CPG categories. Mariano's carries a high-quality line of private label products using the private label, "Roundy's," along with a premium line of private label products under the private label "Roundy's Select." Prior to sampling the products, subjects are asked several questions regarding their private label beliefs. After the blind taste test, when the identities of the sampled products are revealed, subjects self-report their future purchase intentions for the private label. Each participant sampled products from only one category: Cookies (Roundy's O's versus Oreos), Greek Yogurt (Roundy's Greek versus Chobani) or Ice Cream (Roundy's Select versus Breyers). These product categories exhibit a substantial national brand price premium (36.6% in Cookies, 19.7% in Greek Yogurt and 24.8% in Ice Cream). To measure a treatment effect of the blind taste test, the survey responses are matched to each subject's loyalty card account so that national brand and private label purchases can be tracked, within-consumer, before and after the in-store intervention.

Since participation in the in-store blind taste test and survey is voluntary, our treatment sample

---

[1]"The State of Private Label Around the World: where it's growing, where it's not, and what the future holds," Nielsen, November 2014.

could be self-selected on unobserved aspects of consumer brand preferences and brand information. To resolve the self-selection, we exploit the panel structure of our data and apply a difference-in-differences (DID) approach to obtain a consistent estimate of the average causal effect of the blind taste test on the treated consumers. We use time stamps on consumers' transactions to construct a "control group" of consumers who shopped in the store-day of the blind taste tests. We then compare within-household changes in private label purchasing behavior on trips 150 days before and 157 days after the date of the blind taste test for our test and control consumers.

Our key identifying assumption consists of the usual parallel trends condition. We validate our test and control design using the pre-test purchase panel data, finding little systematic differences in preferences for test and control consumers and no evidence of non-parallel trends. We also show that the estimated treatment effect of the blind taste test is robust to an alternative matrix completion estimator that relaxes the parallel trends assumption and allows for a broad class of heterogeneous unobserved trends.

We begin with an analysis of the survey data. Across the three categories, 81% of participants agreed that overall, Mariano's "Roundy's" private label is as good as the national brands. Surprisingly, only 44% of participants predicted they would pick the private label over the national brand in the blind taste test. However, 72% of participants preferred the Roundy's private label immediately after the blind taste test (but before revealing the identities of brands), which is much higher than pure chance. Finally, after the identity of the chosen product was revealed, 84% of participants predicted that they would buy the Roundy's private label next time they shopped in the category they sampled.

Our DID estimates indicate a large initial impact of the blind taste test on purchases. During the week after the blind taste test, the pooled private label share for test consumers increased by 15 percentage points on a base of 8 percentage points. This effect is much larger than the usual advertising effects from traditional media like television. The effect size varies considerably across the three categories: 48 share points in Cookies, 22 share points in Ice Cream and 10 share points in Greek Yogurt. After the first week, the treatment effect declines. During the period spanning 1 to 4 weeks after the test, the pooled treatment effect across categories falls to 8 share points. During the period spanning 1 to 5 months after the test, the pooled treatment effect across categories falls to 2 share points. The latter still represents a quarter of the initial private label share. These findings are

3

qualitatively unchanged when we re-run our analysis at the weekly level using a matrix completion estimator (Athey et al., 2017) that relaxes the parallel trends assumption.

An analysis of the valence effects suggest that the information conveyed by the blind taste test is not merely generating a salience effect, in contrast with traditional promotional tools like in-aisle displays. In particular, the treatment effect is much larger for subjects that derived a positive signal from the taste test, suggesting an informative role of the blind taste test. The persistence in the effect of the intervention indicates that the blind taste test is generating more than an instantaneous promotional effect. Finally, even if we exclude the day of the intervention, we still find a large treatment effect during the first week suggesting a carry-over effect of the intervention into subsequent trips.

A limitation of the DID estimates is that they do not allow for heterogeneity in the treatment effect and they do not control for prices, purchase incidence, the presence of other substitute brands in the category that were not sampled in the blind taste test, and the potentially confounding role of purchase reinforcement through brand loyalty (e.g., Givon and Horsky, 1990). We estimate a random coefficients choice model to control for these various factors in the Greek Yogurt category, which we selected due to its relatively high purchase incidence. Our random coefficients analysis focuses on the largest brands in the category.

We find that, once we control for heterogeneity, we reject a model with brand loyalty (i.e., inertia) in favor of one without. We therefore conclude that any persistence in the effect of the blind taste test is not merely picking up the indirect feedback effect of brand loyalty. Our main finding regarding the short, medium and long-run treatment effects of the blind taste test are robust to the various controls. We also find that the blind taste test increases the consumer preference for the private label and decreases the utility for the tested national brand.

To assess the role of our informational intervention, we use the structural estimates to simulate the counterfactual scenario in which all consumers visiting the store participate in the blind taste test. We predict such a policy would generate a large initial outward shift in private label demand, ceteribus paribus. But, over time (1 to 5 months post test), the demand shift would weaken, converging back towards the initial pre-treatment levels. These findings suggest that the one-time information treatment may not be sufficient to overwhelm the persistent effects of brand capital documented in Bronnenberg et al. (2009) and Bronnenberg et al. (2012). However, the effect is

more than 7 times larger than the current trend in the growth of the CPG private label share in the US (e.g., Dubé et al., 2018), suggesting that established brands create substantial information barriers to entry.

Our findings add to the extant literature on the role of consumer information and consumer willingness-to-pay for branded goods (see the literature survey in Bronnenberg et al., 2019). Earlier work has found that product knowledge and domain expertise are associated with private label purchases. Pharmacists are considerably more likely to buy private label headache medicines, and chefs are considerably more likely to buy private label pantry staples (Bronnenberg et al., 2015). However, it is unclear whether policies that directly communicate objective product information to consumers have a material impact on their brand choices. Bollinger et al. (2011) find that calorie posting nudges consumers towards healthier product choices and Jin and Leslie (2003) find that restaurant hygiene report cards lead more demand for clean restaurants and a larger supply of high-hygiene restaurants. In contrast, in the CPG industry, consumers still tend to pick a higher-priced national CPG brand even when they are told that the private label is objectively comparable in quality (Cox et al., 1983; Carrera and Villas-Boas, 2015). Similarly, in the automobile industry, Allcott and Knittel (2019) find that providing consumers with fuel economy information does not affect consumer car purchases. We find that providing consumers with their own, subjective CPG food information has a long lasting, yet largely transient, effect on their private label purchases.

Our findings also add to the broader literature on branding as a barrier to entry in consumer goods markets (Bain, 1956; Schmalensee, 1982; Bronnenberg et al., 2012). The survey results indicate that consumers underestimate their likelihood of choosing the private label in a blind taste test, in spite of their stated belief that the private label brands are as good as the national brands. In contrast with most of the structural learning literature (e.g., Erdem and Keane, 1996; Ackerberg, 2003), we find that the information effect erodes over time and purchase behavior reverts back towards the pre-test purchase rates, possibly due to forgetting (e.g., Mehta et al., 2004). These findings are consistent with a small theoretical literature on free sampling that also allows for forgetting (e.g., Heiman et al., 2001). The decline is also consistent with the empirical advertising literature in which advertising is found to have a persistent effect on demand that decays slowly over time (e.g., Clarke, 1976; Assmus et al., 1984; Dubé et al., 2005; Sahni, 2012); although our blind taste test generates a much larger effect on demand than impressions from traditional

5

advertising media.

Finally, our findings contribute to the managerial literature on free samples and non-price promotions. Price promotions, like coupons, are typically only found to have short-term direct effects on consumer purchases (e.g., Klein, 1981; Irons et al., 1983), with any longer-term effects typically arising through purchase feedback (e.g., Gedenk and Neslin, 1999). The findings on non-price promotions are more mixed. Gedenk and Neslin (1999) find either no effect or purely a purchase feedback effect from non-price promotions like feature ads and sampling. We explicitly test for and reject purchase feedback, finding a direct long-term effect from the blind taste tests, similar to the long-term effects in Bawa and Shoemaker (2004)'s study of free sampling campaigns. Unlike past studies of free samples, we explicitly treat our consumer subjects with information about their subjective taste for branded versus private label goods. The information treatment changed the beliefs of many consumers regarding the perceived quality gap between national brand and private labels.

This paper is organized as follows. Section 2 discusses the blind taste test and the data used in this study. Section 3 reports on survey findings. Next, section 4 discusses the results of our DID analysis, and section 5 explains the structural model and analysis. Finally, section 6 concludes.

2   Data

The data originate from a partnership with Roundy's, the parent company of several midwestern supermarket chains. We use data from the "Mariano's" chain, located in Cook County, Illinois. The database comprises three sources: (1) transaction-level data collected through shoppers' loyalty cards, (2) SKU-level data tracking daily price and product availability data by store, and (3) the in-store beliefs and preference survey conducted during our blind taste test. We now describe each data set in detail.

2.1   Loyalty Card Data

First, we collected transaction-level data through Mariano's loyalty card database. The data span the 57-month period from July, 2010 (the opening of the first Mariano's store in Arlington Heights, IL) to March, 2015. They include 70,635,896 transactions, time-stamped by day, hour, and minute,

6

and comprise 1,329,900 unique customers and 29 unique Mariano's stores. We retain the UPC-level purchase information for our 3 categories of interest: (shelf-stable) cookies, ice cream, and yogurt. The purchase data include the panelist's unique loyalty card number, the quantity of each UPC purchased, the price paid net of discounts, the trip date and the unique store number.

2.2 Store-Level Data

We also obtained a store-level database tracking the weekly shelf prices and product availability of each of the UPCs in our 3 categories of interest: (shelf-stable) cookies, ice cream, and yogurt. The data span 29 stores, 3,301 UPCs and the time period between July, 2010 and April, 2015.

2.3 Blind Taste Test Survey Data

The blind taste tests were conducted at the end of October, 2014. For each of the three categories, between 10 and 16 separate blind taste test sessions were conducted. In a given store-day, there was never more than one tested category. In total, the blind taste tests spanned 5 stores[2] across 8 days. Table 1 summarizes the date and location of each of the 36 sessions. A given session typically lasted 6 hours and resulted in 3 to 82 respondents.[3] During a session, a "free samples" table was manned by a trained sales associate near the main entrance of the store. The associate used a pre-programmed tablet device to administer the survey and to record responses. Mariano's management trained each sales associate regarding how to use the tablet device and how to administer the survey. Each participant was first asked to swipe her loyalty card. The participant was then asked to answer *yes*/*no* to the question: "Do you think Roundy's branded food products are at least as good as their national brand counter-parts?" After answering, the sales associate explained which two products the participant would sample. Before sampling, the participant was asked to answer a second *yes*/*no* question: "Do you think you will prefer [own brand being tested] or [national brand being tested]?" Depending on the session, the participant compared either Roundy's O's and Oreos (Cookies), Roundy's Greek and Chobani (Greek Yogurt), or Roundy's Select and

---

[2]Bob Mariano (the CEO of Roundy's Corporation at the time of the study) and his team determined exactly when and where the interventions would take place. We have no reason to believe the specific stores and/or the consumer population we analyzed was self-selected to deliver specific information effects (e.g., the test did not target stores that were systematically under-performing on private label sales).

[3]In the store with only 3 trials, the sampling booth only ran for 20 minutes and was shut down early.

Breyers (Ice Cream). The Mariano's employee offered the participant samples of the private label and branded good with the product identities masked. The participant was then asked to answer a third question: "Which product did you prefer?" At this point, the identity of the selected product was revealed to the participant. Finally, the participant was asked a fourth *yes/no* question: "Next time you shop for (product category being tested), will you buy (own brand being tested)?"

– insert Table 1 here –

In total, we collected 1,228 responses from 1,119 unique card holders who participated in one of the 36 blind taste test sessions listed in Table 1. For each unique card holder, we retained the first response (in case of multiple household members participating in the same in-store test) or a randomly determined response (in case of an exact minute tie) of the first test (in case of participation in multiple in-store tests) for our analysis. Our final treatment sample therefore contains 1,119 responses and unique card holders.

2.4 Estimation Sample

For our analysis, we focus on the population of shoppers who shopped in one of our 5 test stores on a date during which we ran one of the blind taste test sessions listed in Table 1. We then define our sample period as the 150 days prior to the blind taste test and the 157 days following the blind taste test that we have at our disposal. We observe panelist shopping data as early as 2010, when the first Mariano's stores opened in Chicago. However, we only use the 150 days prior to the taste test to mitigate the potentially confounding effects of changing tastes over time. In total, we retain 16,684 unique customers who make 211,810 unique transactions within the three categories studied across all 29 stores during the time period of interest. We match households' blind-taste test responses with pre- and post-intervention purchase data using their unique loyalty card number. Among the 1,119 consumers who participated in a blind taste test, 440 (39%) can be matched to purchase data of the varieties they tasted (e.g., Greek yogurt). Of these 440 participants, 99 sampled Oreo style Cookies, 156 sampled Greek Yogurt and 185 sampled Ice Cream. These panelists represent the test sample for our analysis.

For our control sample, we use all panelists who (a) shopped on the same day and location as the 36 blind taste test sessions, (b) did not participate in the test itself, and (c) purchase in the tested

8

categories.[4] In total, we have 16,244 unique control panelists.

Table 2 provides descriptive statistics of our final estimation sample. For each of the control and test groups, we report the unique number of panelists, the total number of transactions and the average number of transactions per panelist. We also report the average private label share, computed as the average share of purchases by volume in the time period.

– insert Table 2 here –

2.5 Estimation Sample for Structural Analysis

For the structural analysis, we assemble a choice panel from the Greek Yogurt category that allows us to control for prices, product availability, and other point-of-purchase factors. We focus on the Greek Yogurt category because of the relatively high purchase rate.

We define the Greek Yogurt market as the top-selling stock-keeping-units (SKUs) of Greek-labeled yogurt sold at Mariano's stores during our sample period. We construct these SKUs using the store level data described above. A SKU consists of the combination of UPCs with the same brand and pack size and with coordinated pricing.

We then retain the 6 top-selling SKUs, all of which are single-serving sized and which represent 60.3% of the total volume of Greek yogurt sold during our sample period: Chobani, Dannon, Fage, Roundy's, Noosa and Yoplait. A SKU's price in a given store-day consists of the average price across the UPCs available for that SKU on the same store-day. Table 3 lists each of our SKUs used for the structural analysis. For each SKU, we list the brand name and each of the underlying UPCs included. We also report the average shelf price and the average share of Greek yogurt volume sold across all stores and days during the entire estimation sample period.

In contrast with the DID results, we use a longer pre-treatment window of one year (365 days), to improve the precision of our estimates of heterogeneity. In Appendix C, we report estimates for a 150-day pre-treatment window that conforms with the DID analysis. Our findings do not change qualitatively with the shorter window.

---

[4]We also considered alternative definitions of the control group for robustness. After our primary definition, we redefined the control group as the set of customers who shopped at the blind taste test location during the hours of the day that the blind taste test was running. Alternatively, we defined it as the set of customers who shopped at that location outside of the window of the blind taste test (and thus did not have the opportunity to self-select into the test). In both instances, our results remain substantively unchanged from those reported here.

We merge the SKU data with our transaction data by store and date. We only retain those panelists that shopped in one of our test stores on a day during which one of the Greek yogurt blind taste tests was fielded, and those weeks corresponding to our estimation sample period (365 days prior to the test and 149 after the test). A retained panelist must purchase Greek yogurt at least once during the sample period. In total, we observe 259 test panelists that make 21,869 trips across the 29 stores during the sample period. We also observe 10,837 control panelists that make 827,995 trips across the 29 stores during the sample period. For each observed trip, we track the chosen item (if one of our Greek Yogurt SKUs was purchased) along with the entire choice set including available SKUs on that trip and their respective prices. If none of the 6 SKUs is chosen, we classify the trip as a "no purchase" occasion. Effectively, we have defined an "outside good." This outside good is chosen on 93.7% of the trips during the entire sample period.

Greek Yogurt purchases appear to satisfy the "discrete choice" assumption since 91.3% of the purchase trips result in the purchase of a single SKU. We dropped the remaining trips that involved purchasing multiple brands.

– insert Table 3 here –

Table 4 summarizes the unconditional and conditional (on purchase) shares for the test and control groups. Chobani is the top-selling product in the category, with 35% of sales. Roundy's private label Greek yogurt is the lowest-share brand with only 4.5% of the market share. We observe some self-selection into the test sample. Our test panelists are more likely to buy Roundy's (7.8% versus 4.4% in control) and Noosa (28.8% versus 14.0% in control), but they are less likely to buy Chobani (28.5% versus 34.8% in control).

– insert Table 4 here –

3   Survey Findings

We now summarize the main descriptive results of the blind taste test survey. The survey consisted of four questions regarding panelists' beliefs about the private label brand in the three categories tested (Oreo-style Cookies, Greek Yogurt, Ice Cream), with two questions asked before and two

after the blind taste test. Table 5 provides summary statistics for the responses from the 664 subjects that make at least one purchase from the tested categories during the sample period.

A total of 81% of our sample responded affirmatively to question one regarding private label quality in general (Q1): "Do you think Roundy's branded food products are as good as their national brand counterpart?" However, only 44% of our sample responded affirmatively to question two (Q2): "Do you think you will prefer Roundy's [product] or [competitor brand product]?"[5] The large difference in responses between Q1 and Q2 is surprising since it seems to suggest subjects do not trust their own beliefs about private label quality in general.[6] It is even more striking that 72% of the sample responded affirmatively to question three (Q3): "Did you prefer Roundy's [product] or [competitor brand product]?" The actual choice rate in Q3 was 28 percentage points higher than the predicted choice rate in Q2. After the identity of the chosen brand was revealed to subjects, 84% of the sample responded affirmatively to question four (Q4): "Next time you shop for Cookies, will you buy Roundy's private label O's Cookie?" The predicted purchase intention is 13 percentage points higher than the choice rate in Q3, and 41 percentage points higher than the predicted choice rate in Q2.

– insert Table 5 here –

The results are robust to a more micro analysis of the three individual categories. Although not reported in further detail, the category-specific survey responses mimic the findings in Table 5 qualitatively. The findings are also robust to using only the subjects who buy from the tested varieties or subcategories (i.e., the sample of the DID analysis in the next section containing 440 subjects) and to using the full sample of 1,119 subjects.

The high response to Q1 could represent misinformation, but could also reflect an acquiescence bias whereby respondents simply provided the affirmative answer to please the surveyor. Also, the sharp increase in stated future private label purchase intention in Q4 relative to the predicted purchase in Q3 is consistent with an information effect, but could also reflect a salience effect or a choice-supportive bias from subjects who had just selected the private label in the blind taste test

---

[5]We used Roundy's O's, Roundy's Select and Roundy's for the cookies, ice cream and Greek yogurt categories respectively. We also used the national brands Nabisco Oreo, Chobani and Breyers for the cookies, Greek yogurt and ice cream categories respectively.

[6]It is possible that a large group of participants is indifferent between private labels and national brands but choose the national brand as a tie-breaker.

component.

As preliminary evidence of an information effect, we now study the impact of the valence of the information conveyed by the blind taste test. The response to Q2 reveals aspects of a subject's prior belief about her relative preference for Roundy's versus the corresponding national brand. The comparison of a subject's response to Q2 and Q3 indicates the sign of the information signal provided to the subject from the blind taste test. For instance, a subject who predicted choosing the national brand on Q2 and who chose the private label on Q3 received a positive signal about the private label brand. We now analyze whether the impact of the blind taste test on choice is moderated by the sign of the information signal. A limitation of this exercise is that we cannot randomly assign subjects to information treatments. Therefore the sign of the information signal is self-selected on a panelist's tastes and our findings should be interpreted purely as correlational.

Table 6 gives the results. From the table, we observe that among all shoppers who thought that they would not prefer the private label brand before the taste test ($N = 375$), 64% chose the private label in the blind taste test and, accordingly, received a positive signal about the private label. In contrast, among those who did think that they would prefer the private label brand ($N = 289$), only 18% chose the national brand in the blind taste test and, accordingly, received a negative signal about the private label. Taken at face value, our results are suggestive that the probability of deriving a positive signal conditional on having a negative prior is much higher than the probability of deriving a negative signal conditional on having a positive prior.

– insert Table 6 here –

Table 7 shows the association between the signal derived from the taste test and the stated intention to purchase private label on the next purchase occasion. The table is structured to facilitate contrasts between households who hold the same pre-test beliefs but who differ in their post-test beliefs. In particular, looking at the first two rows, holding the pre-test stated preference for the private label brand constant at a high level ($Q_2 = 1$), 69% ($N = 52$) of consumers who updated negatively ($Q_3 = 0$) indicated the intention to buy the private label next time. However, among those who remained positive ($Q_3 = 1$; $N = 237$), a strongly contrasting 95% indicated they would buy the private label next time. Holding, in rows 3 and 4, the pre-test stated preference for the private label brand constant at a lower level ($Q_2 = 0$), among all who thought initially they would not pre-

fer the private label brand and continued to do so ($Q_3 = 0$; $N = 136$), 51% claim to buy the private label brand on a next occasion, whereas, among those who updated positively ($Q_3 = 1$; $N = 239$), 96% indicated the intention to buy the private label next time, almost doubling the intent to buy the private label. In sum, holding constant initial beliefs, the taste test resulted in updated beliefs that are strongly correlated to the intent to buy private label in the future.

– insert Table 7 here –

In the next section, we use subjects' actual purchase behavior after the date of the blind taste test to analyze whether the blind taste test had a persistent effect on their beliefs and shopping behavior.

4 Difference-in-Differences (DID) Analysis

4.1 Method

We use the shopping data described in section 2.4 to estimate the causal effect of the blind taste test on panelists' private label choices. Our identification strategy relies on the panel structure of our transaction data. We use a two-way fixed effects (TWFE) model to compare the differences in shopping behavior, before versus after the test dates, between our participants in the taste tests and all other shoppers in the test store on the same date. This *difference-in-differences* (DID) approach controls for time-invariant confounders with the treatment using consumer fixed effects, and common time confounders using time fixed effects.

Let $h = 1, ..., H$ denote panelists, each with a unique individual loyalty card, and let $c = 1, ..., C$ denote the product categories. We index the panelist's trip dates by $t = -T_{hb}, ..., 0, ..., T_{he}$ where $t = 0$ indicates the date of the blind taste test, T_{hb} is the total number of days between the first observed trip for h and the date of the test, and T_{he} is the total number of days elapsed between the date of the test and the last observed trip. Let $\tau_{hct} \in \{0, 1\}$ indicate whether panelist h was "treated" in category c prior to or on date t, meaning that she participated in the blind taste test, i.e., $\tau_{hct} \equiv \mathbb{I}_{\{\text{treatment group}, t \geq 0\}}$. Also, let τ_{hc} (without a time subscript) indicate whether panelist h is ever "treated" in the duration of our panel. Let $Y_{hct} \in \{0, 1\}$ indicate whether panelist h buys the private label in category c, conditional on making a category purchase on date t. Using the familiar

13

potential outcomes framework, for each individual h and time period t, we are interested in the potential outcomes $Y_{hct}(\tau_{hct})$.

Our empirical goal consists of obtaining an estimate of the average treatment effect of the blind taste test for a given category c on the treated units and time periods:

$$
\begin{aligned}
ATT_c &= \mathbb{E}\left(Y_{hct}(1) - Y_{hct}(0)\,|\,\tau_{hct} = 1\right) \\
&= \mathbb{E}\left(Y_{hct}(1)\,|\,\tau_{hct} = 1\right) - \mathbb{E}\left(Y_{hct}(0)\,|\,\tau_{hct} = 1\right).
\end{aligned}
\tag{1}
$$

As with most settings, the challenge is that we do not observe $(Y_{hct}(0)\,|\,\tau_{hct} = 1)$. In addition, the non-random assignment of treatment means that some consumers may self-select into the treatment condition, in part, based on their heterogeneous treatment effects.

We use a standard two-way fixed-effects estimator of ATT_c based on the DID between treated and untreated households. In particular, we use a linear probability model to predict a panelist's choice between the private label and the tested national brand in a category c on trip date t, conditional on purchase:

$$
\begin{aligned}
Y_{hct} = \alpha_{hc} + \gamma_{tc} &+ \beta_{SR,c} \cdot \tau_{htc} \cdot \mathbb{I}_{\{t \in (0,6)\}} \\
&+ \beta_{MR,c} \cdot \tau_{htc} \cdot \mathbb{I}_{\{t \in (7,27)\}} \\
&+ \beta_{LR,c} \cdot \tau_{htc} \cdot \mathbb{I}_{\{t > 27\}} + \varepsilon_{htc}.
\end{aligned}
\tag{2}
$$

The parameter α_{hc} is a panelist-category fixed effect, γ_{tc} is a category-week fixed effect,[7] and the indicator variables $\mathbb{I}_{\{t \in \mathcal{T}\}}$ denote whether trip date t falls in the time interval $\mathcal{T}$, measured in days. The parameters $\{\beta_{SR}, \beta_{MR}, \beta_{LR}\}$ capture the average treatment effect of the blind taste test on the propensity to buy a private label versus a national brand, conditional on purchase. We allow the ATT to vary with the duration of time elapsed since the date of the in-store taste test: a short run effect of the taste test during the first 7 days after the test (β_{SR}), a medium run effect of the taste test during the time interval between 7 days and 27 days after the test (β_{MR}), and a long run effect of the taste test during the time interval between 28 days and 157 days after the test (β_{LR}). We also report a version of equation ((2)) that splits the long run effect, β_{LR}, into separate 4-week treatment

---

[7]To define category-week fixed effects, we use the 7-day periods relative to the date that the blind taste test took place in a store.

effects, $(\beta_{\text{4-8 weeks}}, \beta_{\text{9-12 weeks}}, \beta_{\text{13-16 weeks}}, \beta_{\text{>16 weeks}})$.

The consistency of our DID estimator of $\{\beta_{SR}, \beta_{MR}, \beta_{LR}\}$ relies on *exogeneity*: $\mathbb{E}(\varepsilon_{hc}|\alpha_{hc}, \gamma_{tc}, \tau_{htc}) = 0$. The consumer fixed-effects control for self-selection on unobserved, time-invariant factors. Second, the inclusion of time fixed-effects that are common across consumers implicitly assumes that treated and untreated households follow *parallel trends*. In the potential outcomes framework, this implicit assumption can be stated as follows (e.g., Abadie, 2005):

$$\mathbb{E}\left(Y_{htc}(0) - Y_{ht'c}(0)|\tau_{hc} = 1\right) = \mathbb{E}\left(Y_{htc}(0) - Y_{ht'c}(0)|\tau_{hc} = 0\right), \text{where } t \neq t'. \tag{3}$$

Condition (3) assumes that any changes in Y_{htc} over time are independent of whether or not a household participated in the blind taste test.

A recent literature has demonstrated that, in general, the TWFE can exhibit bias and may not estimate the ATT when the timing of treatment is staggered across consumers (e.g., Athey and Imbens, 2018; de Chaisemartin and D'Haultfoeuille, 2019; Goodman-Bacon, 2018; Imai and Kim, 2019). These biases are unlikely to be large in our setting where all of the treated panelists received treatment within an eight-day window and the treatment was administered once only in a given category.[8] Therefore, the DID will generate the ATT of interest in our setting.

While we cannot directly test the parallel trends assumption (3), we can follow Angrist and Krueger (1999, p. 1299) to exploit the long time series in our panel data and test for parallel trends during the pre-treatment period. First, we conduct a direct pre-treatment test for parallel trends by estimating

$$Y_{htc} = \alpha_{hc} + \delta_c t + \omega_c t \mathbb{I}_{\{\text{treatment group}\}} + \varepsilon_{htc}, \ t < 0 \tag{4}$$

using all panelists during the pre-treatment period. The parameter δ_c is the common trend and ω_c is a deviation from the common trend for the treatment group. Results are displayed in Table 8 for a pre-treatment window of 150 days, i.e., $T_{hb} \leq 150$. The common trend is small, relative to the 7.8% baseline private label share, and statistically insignificant. Of interest is the null hypothesis of parallel trends: $H_0: \omega_c = 0$. We fail to reject the null of parallel trends, with an economically small

---

[8]Goodman-Bacon (2018) develops several results characterizing these biases. Following Goodman-Bacon (2018) theorem 1, the weights on the terms comparing treated and "not-yet-treated" consumers will be extremely small since they represent a tiny fraction of the total number of observations. Since the variance in time for which consumers in the treatment group are treated is similar across stores, the weight on the terms comparing treated and control consumers will be close to one.

predicted mean difference of only 0.04 percentage points. However, the results are imprecise and we cannot rule out with a high degree of confidence that the treatment group trend is as much as 0.20 percentage points lower. This finding is robust to the use of alternative pre-treatment window lengths as long as one year. With a one-year window length, we do reject the null of equal trends at the 5% significance level; but the difference in the weekly trend size remains small at 0.091%.[9] In section 4.3, we will check the robustness of our *ATT* estimates to an estimator that uses a weaker assumption than parallel trends.[10]

–insert Table 8 here–

Our tests appear to support the assumption of parallel trends. This evidence in conjunction with the panelist-specific fixed effects should ensure the consistency of our DID estimates of the treatment effect from the blind taste tests.

4.2   Estimated Treatment Effects

We now focus on the DID estimates using the linear probabilities model in equation (2). Table 9 presents the estimates for the DID regressions. We observe that the base share of the private label across the three categories ranges from 4.0 (Greek Yogurt) to 19.1 percent (Ice Cream). Pooled across categories, panelists, and purchase occasions, the baseline probability of buying the private label brand is 7.9 percent.

–insert Table 9 here–

In the first column of Table 9, we pool the three categories and allow for a common treatment effect. We find that during the week after the blind taste test, the short run purchase probability of the private label brand increases by 15.1 percentage points to 23.0 percent; although we cannot rule out an increase as small as 11.3 percentage points at the 5% significance level. The blind taste test thus tripled the market share of the private label in the short run. We also find a large

---

[9]The significance of the difference reflects the large sample size of $N = 206,532$ panelist-trips for the 365-day pre-window.

[10]Although not reported herein, we also experimented with placebo tests that assigned a treatment date arbitrarily during the pre-treatment period. In this specification, we also fail reject the assumption of parallel trends. Still, we cannot reject moderate-sized differences between treatment and control consumers once we account for the statistical uncertainty.

and significant short-run effect of the category-specific treatment effects reported in columns two, three and four.

Pooling across categories, we find that between 7 and 27 days after the blind taste test, the medium run purchase probability is still 7.8 percentage points larger than before the test, doubling the baseline purchase probability of 7.9 percent; although we cannot rule out an increase as small as 4.4 percentage points at the 5% significance level. Therefore, the pooled treatment effect declines relative to the short-run effect. Pooling across categories, the difference between the short run and medium run effects is large enough that we can reject the null hypothesis that they are the same. The numerical difference between the short run and medium run effect is large for each individual category. However, the difference is not always significant due to the limited sample sizes in specific categories.

During the four-month period starting 28 days after the test, the pooled (across categories) private label share is still 2.3 percentage points higher than during the pre-treatment period, although we cannot rule out that it is as small as 0.8 percentage points at the 5% significance level. Hence, the treatment effect of the in-store blind taste test declines further over time, albeit more slowly over a 5-month period. Still, compared to the baseline purchase probability of 7.9%, the taste test increases the relative share of the private label by 29% in the long run. In the pooled analysis, we reject the null hypothesis of no difference between the medium run and long run effects. These differences are also significant in the category level analysis for Greek Yogurt.

Table 10 decomposes the long-term treatment effect by allowing for separate consecutive 4 week effects. Column one indicates that the pooled treatment effect across categories decays from over time. The effect becomes statistically insignificant after 8 weeks. Even though after 16 weeks the treatment effect is insignificant in all three categories, we cannot rule out effects as large as several percentage points at the 5% significance level.

–insert Table 10 here–

As a robustness check, Appendix A reports the treatment effects using only the difference (before versus after treatment) within the treatment group. A time trend, identified from the pre-treatment data, controls for a constant trend in the pre- and post periods. The results in Tables 16 and 17 confirm the estimates of the DID analysis above. The larger effect sizes in these robustness

17

checks, especially for the long run, highlight the importance of the two-way fixed effect specification used above. The DID approach uses the untreated group to identify post-treatment time effects that might otherwise generate spurious treatment effects.

We conduct three additional robustness checks. For brevity these are reported for the pooled data only, corresponding to the first column of table 10. First, we check the sensitivity of our results to the definition of the short-run time interval. A concern is that, in addition to communicating product information, the in-store free samples booth may simply create a short-run salience effect, similar to standard in-store promotions like in-aisle or end-of-aisle displays. We re-estimate the DID regression dropping treated panelists who bought from the test-category on the same day as the taste test. Figure 1 compares the predicted private label share level when households who buy from the category of interest at the test day are retained (as before) versus excluded. Figure 1 indeed suggests a "day of the test" effect which might be capturing the salience effect of the sampling booth on choice. The figure shows that once we drop households who buy from the category on the test day, our point estimate for the short-run effect falls by several percentage points and we fail to reject the hypothesis that short-run and medium-run effects are the same (the lower sample size from sub-setting the data also generates a noisier estimate of the short-run effect). Importantly, however, we continue to find a persistent treatment effect after the date of the test even among those who do not buy on the test day.

–insert Figure 1 here–

As a second robustness check, we check the sensitivity of our results to the definition of the three categories and present the results from using the subcategories (Oreo-style Cookies, Ice Cream, and Greek Yogurt) as in Table 10 along with the results when using the purchases in the broader categories (Cookies, Ice Cream & Frozen Yogurt, and Yogurt) used in the blind taste test. Figure 2 compares the results for each of the two category definitions. The short-run effect appears to be robust across the two definitions. However, our point estimate for the medium run effect is several percentage points smaller when we use the entire category (this difference is statistically significant). This finding is not surprising since the broader category definition includes product varieties that were not explicitly tested even though they use the same brand names as those in the test. Most important, we continue to find a statistically significant effect in the medium run and

18

long run under both definitions.

As a final robustness check, we check the sensitivity of our results to the length of the pre-treatment window. As shown in Figure 3, we find that the magnitude and significance of our estimated treatment effects appear to be robust to the different window lengths.

We conclude that the effects of the blind taste test between the tested private label food products and their corresponding leading national-brand competitors are both large and persistent over a period of several months, although the effects decline over time.

4.3   Relaxing the Parallel Trends Assumption

In this section, we briefly explore the robustness of our treatment effects estimates by specifying a more general interactive fixed-effects model (e.g., Bai, 2009) and using the matrix completion estimator proposed by Athey et al. (2017). As explained earlier, one of the key identification assumption underlying our DID estimates in section 4.2 is that treatment and control consumers' purchases follow *parallel trends*. The interactive fixed-effects model allows for heterogeneity in trends across consumers and, thus, does not require the same stable time paths in outcomes for treated and untreated households.

As before, we face the problem that we do not observe $(Y_{htc}(0)|\tau_{htc} = 1)$. To simplify the remainder of this section, we remove the category index, c. To impute these missing values and calculate ATT as in equation 1, Athey et al. (2017) model the untreated potentialoutcomes in the following matrix form:

$$Y(0)_{H \times T} = L_{H \times T} + \varepsilon_{H \times T} \tag{5}$$

where L is an $H \times T$ matrix to be estimated, and ε_{ht} is measurement error. We modify the strict exogeneity assumption as follows: $\mathbb{E}(\varepsilon|L) = 0$. Standard factor models assume that L has a lower-rank approximation (of rank R) that can be expressed as the product of common time-factors, $V_{T \times R}$, and heterogeneous cross-sectional factor loadings, $U_{H \times R}$. That is, L can be decomposed as $L = U_{H \times R}V'_{T \times R}$ and a missing outcome can be approximated using $L_{ht} \approx \sum_{r=1}^{R} U_{hr}V_{tr}$. The factor

model, therefore, nests the TWFE approach as a special case when $R = 2$, $U_h = \begin{bmatrix} \gamma_h & 1 \end{bmatrix}$ and $V_t = \begin{bmatrix} 1 \\ \delta_t \end{bmatrix}$. Larger-rank R allows for a more general factor model than simply unit fixed effects and common time-shocks. Furthermore, unlike a standard factor model, Athey et al. (2017) allow R to grow with H and T, allowing for richer patterns of unobserved heterogeneity.

For estimation, we use the following regularized regression proposed by Athey et al. (2017):

$$\left(\hat{L}, \hat{\Theta} \right) = \operatorname*{arg\,min}_{\{L_{ht}\},\{\gamma_h\},\{\delta_t\}} \left[\sum_{(h,t) \in \mathcal{O}} \frac{(Y_{ht}(0) - L_{ht} - \gamma_h - \delta_t)^2}{|\mathcal{O}|} + \lambda \, \|L\|_1 \right] \tag{6}$$

where $\Theta = (\gamma_1, ..., \gamma_N, \delta_1, ..., \delta_T)$ are parameters to be estimated and the notation $\mathcal{O}$ denotes the set of observations, indexed by (h,t), for which the untreatedoutcome is observed. Regularization depends on the penalty term, λ,[11] and the Nuclear norm, $\|L\|_1 = \sum_{i=1}^{min(H,T)} \sigma_i(L)$, where $\{\sigma_i(L)\}_{i=1}^{min(H,T)}$ are the singular values of L. Accordingly, this problem is termed matrix completion with nuclear norm minimization (MC-NNM) because the objective consists of estimating the components of the matrices U and V above. We refer the interested reader to Athey et al. (2017) for technical details and note that, like them, we do not regularize the fixed effects to ensure we control for time-invariant heterogeneity and common time-shocks. The additional factors L_{ht} allow for time-varying unobservable heterogeneity.

Our estimate of the ATT is then:

$$\begin{aligned} A\hat{T}T^{MC-NNM} &= \frac{\sum_{(h,t)} \tau_{ht} \left(Y_{ht}(1) - \hat{Y}_{it}(0) \right)}{\sum_{(h,t)} \tau_{ht}} \\ &= \frac{\sum_{(h,t)} \tau_{ht} \left(Y_{ht}(1) - \hat{L}_{ht} - \hat{\gamma}_h - \hat{\delta}_h \right)}{\sum_{(h,t)} \tau_{ht}}. \end{aligned} \tag{7}$$

Again, if we let $R = 2$ a and $L = 0$, our data-generating process simplifies to the two-way fixed-effects model again:

$$Y_{ht}(0) = \gamma_h + \delta_t + \varepsilon_{ht}.$$

---

[11]The parameter λ is selected using 5-fold cross-validation and out-of-sample RMSE.

We can construct the MC-NNM analog of our DID estimates from section 4 as follows:

$$A\hat{T}T^{DID} \quad = \quad \frac{\sum_{(h,t)} \tau_{ht} \left(Y_{ht}(1) - \hat{\gamma}_h - \hat{\delta}_t \right)}{\sum_{(h,t)} \tau_{ht}} \tag{8}$$

which estimates $\hat{\gamma}_h$ and $\hat{\delta}_t$ for the special-case of $L_{ht} = 0 \; \forall \, h, t$.[12]

To manage dimensionality, we collapse the estimation sample to a weekly frequency for each household, instead of daily. In addition, we need to drop those treated households for which we do not observe any pre-treatment-period observations because their fixed effects are not identified.[13] We define our dependent variable, Y_{ht}, as the share of purchases that are private label and use this share as our dependent variable: $Y_{ht} = \frac{\#\,PL \text{ purchases by } h \text{ in week } t}{total \, \# \, purchases \text{ by } h \text{ in week } t}$. [14]

Table 11 reports the point estimates and bootstrapped (at the household-level) 95% confidence intervals for ATT^{MC-NNM} and ATT^{DID}. As before, we allow the treatment effect to differ over the short run of one week after the intervention (SR), the medium run from the end of the first week to end of four weeks after the intervention (MR), and the long run from the end of the fourth week to the end of our sample (LR).

The differences between the point estimates of ATT^{MC-NNM} and ATT^{DID} in Table 11 are negligible and alleviate concerns that our main results are driven by a violation of the parallel trends assumption. Indeed, the same general findings emerge as those in our primary specification in Section 4. We find a significant and large effect of the blind taste test on private label purchase likelihood which then decays over time to a smaller, but economically meaningful, long run point estimate. The robustness of these findings to the methods in this section suggest our initial results were not merely spurious, as would have been the case had they been identified off un-modeled deviations in the time trends between the control and treatment group.

---

[12] Both our OLS and MC estimators generate consistent estimates of ATT^{DID}. But, the MC estimator is less efficient because it excludes the post-treatement observations for the treatment group when estimating the time and unit fixed effects.

[13] Although not reported herein, the ATT^{DID} estimates are quite similar to those based on equation (2), reported above in section 4.2, suggesting that the deletion of households does not alter our key findings. Results are available from the authors upon request.

[14] Again, although not reported herein, ATT^{DID} computed with the daily versus weekly outcomes are almost identical, suggesting that the time-aggregation also does not alter our key findings. Results are available from the authors upon request.

4.4 Valence of Information

We now explore the potential information content of the treatment effect. A novel aspect of our data is that we surveyed panel members about their preference for the private label relative to the national brand before and after the blind taste test. Within panelist, this format allows us to measure the moderating effect of the valence of the subjective information from the blind taste test on the treatment effect. Panelist fixed effects allow us to control for the potential self-selection into a positive versus negative initial state.

We use the stated preference, elicited immediately before and after the blind taste test to measure the sign of the information signal conveyed by the free samples. If a panelist predicted she would prefer the national brand and then stated she preferred the private label after the taste test, we classify the information as a positive update. If the panelist predicted she would prefer the private label and still does so after the test, we classify the information as positive confirmation. Similarly, we define a negative update as a predicted preference for the private label and a preference for the national brand after the taste test. Finally, if the panelist predicted she would prefer the national brand and still does so after the taste test, we classify this information as negative confirmation.

We re-run the DID regressions in (2) with a single average (over time) treatment effect and an interaction between the treatment effect and the valence of the information:

$$
\begin{aligned}
Y_{hct} = \alpha_{hc} + \gamma_{ct} &+ \beta^{\text{neg}\to\text{pos}} \cdot \tau_{hc} \cdot \mathbb{I}_{\{t \geq 0\}} \\
&+ \beta^{\text{pos}\to\text{neg}} \cdot \tau_{hc} \cdot \mathbb{I}_{\{t \geq 0\}} \\
&+ \beta^{\text{pos}\to\text{pos}} \cdot \tau_{hc} \cdot \mathbb{I}_{\{t \geq 0\}} \\
&+ \beta^{\text{neg}\to\text{neg}} \cdot \tau_{hc} \cdot \mathbb{I}_{\{t \geq 0\}} + \varepsilon_{hct}.
\end{aligned}
$$

Table 12 presents the results against the baseline of the average treatment effect over the entire post-test time window. In the column labeled "Baseline" we confirm that the blind taste test has a large positive impact on the probability to purchase a private label over a branded product in the three categories tested. The propensity to buy private label increases from from 7.6% to 12.0%, a larger than 50% increase.

–insert Table 12 here–

Moving to the column labeled "Valence", we find that the positive treatment effect is concen-

trated among those who have positive evaluations of the private label after the taste test: those panelists that update their preference for the private label positively and those that confirm their positive prior disposition. For those who update negatively on the private label, the point estimate for the effect of the blind taste test is negative yet insignificant. We cannot rule out a negative treatment effect on the private label choice propensity as large as -10.7 percentage points or a positive effect as large as 0.5 percentage points at the 5% significance level. Finally, for those who remain negative about the private label brand we fail to reject a null effect, although we cannot rule out an effect as small as -2.1 percentage points or as large as 3.9 percentage points at the 5% level. These findings suggest that the blind taste test impacts mostly those participants who derive a positive signal from the experience although we cannot rule out a negative treatment effect for those participants who derived a negative signal.

5 Structural Analysis

5.1 Overview

Using program evaluation methods above, we documented strong and persistent treatment effects of the blind taste tests on the treated consumers. These findings indicate that subjective product information can change a consumer's reliance on brand to make purchase decisions. Ultimately, we would like to quantify the extent to which such information affects the overall market structure of the category. Thus, we would like to predict the counterfactual effect of treating all consumers with the blind taste test, which we will interpret as information.

The policy simulation is complicated by the fact that participation in the blind taste test is voluntary and may be self-selected on unobserved aspects of tastes. Self-selection was handled using a *fixed-effects* approach in our DID and MC estimators. However, in this section we develop a structural choice model in which we estimate the distribution of heterogeneity, including for the treatment effects of the blind taste test. Because correlation of these components with the treatment assignment can create an endogeneity bias, our structural model includes the selection decision. We use the estimates to predict the treatment effect on the untreated consumers and evaluate a counterfactual exercise in which all consumers become informed (i.e., as if all consumers

participate in the blind taste test).

In addition to measuring the impact of information on market structure, a structural choice model also allows us to test for a number of potentially confounding factors that could not easily be addressed by the DID analysis. The DID approach in section 4 focused on the binary outcome of private label choice conditional on purchase. It did not consider the specific brand choice alternatives or the no-purchase option. In addition, the DID estimator did not control for several potential confounding factors including variation across trips in the set of available brands in the category and/or the products' prices. Moreover, while the DID approach is robust to heterogeneity, it does not provide a characterization of the heterogeneity in the treatment effect. Since we expect consumers to have different degrees of experience and information with the various brands and the private label, we anticipate heterogeneity in the amount of information conveyed by the taste test and therefore in its effect size. To account for the role of causal factors like the choice set, prices and the role of consumer heterogeneity, we fit a random coefficients logit demand system to the transaction panel data.

A final concern is that any persistence in the treatment effect of the blind taste test could be identified spuriously from omitted state dependence in demand, or "purchase feedback" effects. We also estimate a version of the model that allows for structural state dependence in choices to control for such feedback effects, allowing for a more robust test of the causal effect of the blind taste test.

5.2 Model and Econometric Specification

We denote consumer loyalty card panelists by $h = 1, ..., H$. On a trip during time t, a panelist chooses amongst the $j = 1, ..., J$ products in a category or chooses $j = 0$, an outside option (i.e., "no purchase of any of the J products"). As before, we assume $t = -T_{hb}, ..., 0, ..., T_{he}$ where the blind taste test occurs at date $t = 0$. We assume the timing of trips is exogenous to demand in the given category. A self-selected subset of the panelists, $\mathbb{T}$, participates in a blind taste test on date $t = 0$, indicated by τ^h where $\tau^h = 1$ if $h \in \mathbb{T}$, and 0 otherwise. We index the subset of tested brands using $\mathbb{J} = \{PL, NB\}$ where $j = PL$ and $j = NB$ denote the private label and national brand alternatives that were sampled during the blind taste test.

On trip date t, panelist h derives the following conditional indirect utility from choosing alternative j: $u_{jt}^h = v_j\left(p_{jt}, s_t^h, \tau^h; \Theta^h\right) + \varepsilon_{jt}^h$, where $v_j\left(p_{jt}, s_t^h, \tau^h; \Theta^h\right)$ represents the panelist's deterministic (conditional on the parameters Θ^h) utility, and ε_{jt}^h is an i.i.d. Type I Extreme Value distributed random utility term. The deterministic utilities are as follows:

$$v_j\left(p_{jt}, s_t^h, \tau^h; \Theta^h\right) = \begin{cases} \alpha_j^h + \beta_j^h \mathbb{I}_{\{\tau^h = 1, t \geq 0, j \in \mathbb{J}\}} + \eta^h p_{jt} + \gamma^h \mathbb{I}_{\{s_t^h = j\}}, & j = 1, \ldots, J \\ 0, & j = 0 \end{cases} \tag{9}$$

where p_{jt} is the price of product j. The state variable $s_t^h \in \{1, \ldots, J\}$ indicates the previous product purchased by panelist h such that repeat-buying the same product generates a marginal utility of γ^h. The coefficients $\{\beta_{PL}, \beta_{NB}\}$ allow for the possibility that a panelist's brand preferences for the private label or tested national brand change in response to the information from the blind taste test. We let the treatment effects $\{\beta_{PL}, \beta_{NB}\}$ vary over time as follows:

$$\beta_j = \begin{cases} \beta_j^{SR}, & t \in (0, 6) \\ \beta_j^{MR}, & t \in (7, 29) \\ \beta_j^{LR}, & t \in (30, 149) \end{cases}, \ j \in \{PL, \ NB\}.$$

Let $y_{jt}^h = 1$ indicate if consumer h purchases brand j on trip t, and 0 otherwise. The conditional probability that panelist h chooses alternative j on trip t is

$$\Pr\left\{y_{jt}^h = 1 | p_t, s_t^h, \tau^h, \Theta^h\right\} = \frac{\exp\left(v_j\left(p_{jt}, s_t^h, \tau^h; \Theta^h\right)\right)}{1 + \sum_{k=1}^J \exp\left(v_k\left(p_{kt}, s_t^h, \tau^h; \Theta^h\right)\right)}. \tag{10}$$

To complete the choice model, we specify a distribution for the persistent tastes across consumers, $\Theta^h = \left(\alpha_1^h, \ldots, \alpha_J^h, \beta_{PL}^h, \beta_{NB}^h, \eta^h, \gamma^h\right)$. We use a standard hierarchical structure where

$$\Theta^h = \bar{\Theta} + v^h, \ v^h \sim N(0, \Sigma)$$

$$\bar{\Theta}|\Sigma \sim N\left(\bar{\bar{\Theta}}, a^{-1}\Sigma\right) \tag{11}$$

$$\Sigma \sim IW(v, V)$$

25

where $\bar{\bar{\Theta}}$, a, v and V are prior parameters set by the researcher.

5.3 Self-Selection

Consumers' voluntary selection into the blind taste test, $\tau^h = 1$, is unlikely to be random. In the context of our formal choice model, $\mathbb{I}_{\{\tau^h=1, t \geq \tau, j \in \mathbb{J}\}}$ will be endogenous if the probability of being in the treatment group, $\Pr\{\tau^h = 1\}$, depends on unobserved aspects of consumer tastes, Θ^h. Participationrequires a consumer to incur the hassle costs associated with time spent swiping her card, responding to the beliefs survey, and sampling the tested products. So, the decision to participatemay reflect benefits, many of which may be independent of unobserved preferences for the tested brands. For instance, consumers who consolidate a high proportion of their shopping in the chain may perceive a higher long-term benefit from the information treatment. Alternatively, consumers who are frequent shoppers may already be informed about the Roundy's products and thus perceive little potential for learning. At the same time, participation could also select on consumers' heterogeneous preferences. Someone with a high utility for either of the tested brands may value the immediate gratification of eating a "free sample." Alternatively, someone with a high potential information benefit (treatment effect) may also be more likely to participate.

To capture these sources of self-selection, we assume that on the date of the intervention, a consumer h will voluntarily participate as long as the incremental benefits of doing so are positive:

$$\Delta U^h \equiv w^{h\prime}\lambda_w + \Theta^{h\prime}\lambda_\Theta + \xi^h > 0$$

ΔU^h are the incremental benefits from participation, w^h is a vector of observed measures of consumer $h's$ shopping intensity at the Mariano's chain, Θ^h are consumer $h's$ preferences for the tested product category as in the previous section, $\Lambda = \left(\lambda'_w, \lambda'_\Theta\right)'$ are participation tastes, and ξ^h is an i.i.d. draw from a logistic distribution representing additional random net benefits from participation. The conditional probability that consumer h selects into the blind taste test is then

$$\Pr\left\{\tau^h = 1 | w^h, \Theta^h, \Lambda\right\} = \frac{1}{1 + \exp\left(-w^{h\prime}\lambda_w - \Theta^{h\prime}\lambda_\Theta\right)}. \tag{12}$$

Our complete model consists of the following conditional distributions

$$y_t^h | p_{jt}, s_t^h, \tau^h, \Theta^h, \quad \{\text{Brand Choice}\} \tag{13}$$

and

$$\tau^h | w^h, \Theta^h, \Lambda, \quad \{\text{Self-Selection}\} \ . \tag{14}$$

Equation (14) affects our inferences about Θ^h because the set of consumers receiving treatment may be self-selected on these tastes. Ignoring this selection equation, (14), and conducting inference using only equation (13) could lead to asymptotic bias in Θ^h because the selection component of the likelihood contains information about Θ^h. The likelihood of the parameters under the full model is

$$\ell\left(\Theta^h, \Lambda\right) = \prod_t \prod_j \Pr\{y_{jt}^h = 1 | p_t, s_t^h, \tau^h, \Theta^h\}^{y_{jt}^h} \prod_{i=0}^{1} \Pr\left\{\tau^h = i | w^h, \Theta^h, \Lambda\right\}^{\mathbb{I}\{\tau^h = i\}} . \tag{15}$$

This formulation is similar to the targeted marketing problem (e.g., Manchanda et al., 2004) and to the initial conditions bias in state-dependent choice models (e.g., Simonov et al., 2019).

Below we estimate both the conditional choice model alone (equation 13) with the hierarchy described in equation 11, and the complete, selection-corrected specification with both the choice and selection models (equations 13 and 14) combined with the hierarchy in equation 11. We conduct inference using a blocked random-walk Metropolis algorithm (see Appendix B for details).

Conceptually, all the taste parameters, $\{\alpha_1, ..., \alpha_J, \eta, \gamma\}$, could be identified by pooling all the panelists during the pre-test period, $t < 0$. The selection concerns apply primarily to the treatment effects, $\left\{\beta_j^{SR}, \beta_j^{MR}, \beta_j^{LR}\right\}_{j \in \{PL, NB\}}$. The selection parameters rely on the excluded variables w^h to provide a source of independent variation. For our counterfactual that assigns treatment to the entire consumer population, we rely on the assumption that all consumers, treated and untreated, have the same covariance structure in their tastes, Σ. Accordingly, we can conduct inference on the posterior treatment effects for the untreated set of consumers.

An additional concern is that persistence in the effect of the blind taste test could be identified spuriously from omitted brand loyalty. Suppose the blind taste test only has an immediate direct effect on demand, causing a consumer to switch to the private label at the time of the intervention.

The demand inertia associated with loyalty could create an indirect long-term effect of the blind taste test through purchase reinforcement (e.g., Givon and Horsky, 1990). The inclusion of the loyalty term, $\mathbb{I}\{s_t^h = j\}$, controls for such purchase reinforcement (e.g., Keane, 1997; Dubé et al., 2010).

5.4   Structural Estimates

We fit the demand model to the transaction data from the Greek yogurt category. We first compare the fit of six different specifications in total: (1) baseline demand ($\beta_{PL} = \beta_{NB} = \gamma = 0$), (2) baseline demand with loyalty ($\beta_{PL} = \beta_{NB} = 0$), (3) demand with treatment effects ($\gamma = 0$), (4) demand with treatment effects and loyalty, (5) demand with treatment effects and self-selected participation ($\gamma = 0$), and (6) demand with treatment effects, loyalty and self-selected participation. For each specification, we compare results with versus without unobserved heterogeneity. In estimation, we use our MCMC algorithm to take 40,000 posterior draws per specification. Of these, we drop the first 10,000 draws as a burn-in period and next retain every 5th draw from the chain. This leaves us with 8,000 draws for posterior inference. We assess posterior model fit using the Newton and Raftery (1994) approximation of the posterior likelihood.

Table 13 reports the posterior likelihood of each model. As expected, unobserved heterogeneity improves model fit substantially. We also find that after controlling for heterogeneity, the inclusion of loyalty worsens posterior fit.[15] The inclusion of treatment effects improves fit, but the magnitude of the improvement is small due to the fact that our test panelists represent only a small fraction of the sample. Finally, we also find an additional small improvement from the correction for self-selection (i.e., relaxing the restriction $\lambda_\Theta = 0$). The best-fitting specification excludes loyalty, but includes the treatment effect of the blind taste test and allows for self-selection. The exclusion of loyalty suggests that purchase reinforcement is not contributing to the persistent effect of the blind taste test.

–insert Table 13 here–

In Table 14, we report the the posterior means and 95% credibility intervals for each of the selection parameters, $\mathbb{E}(\Lambda|\mathbf{D})$, for our best-fitting specification (5). We find that the shopping

---

[15]The posterior likelihood has a built-in control for over-fitting as can be seen by its asymptotic approximation, the Schwarz criterion, which penalizes models with more parameters.

statistics, average basket size in dollars and average basket size in total items both have a significant impact on the probability of participating in the blind taste test. The low precision of our estimated effects of the taste parameters is perhaps unsurprising given the relatively small proportion of compliers in the Yogurt category coupled with the fact that individual-level taste parameters are estimated with statistical error. However, our credibility intervals reveal that we cannot rule out relatively large effects.

–insert Table 14 here–

We report the posterior mean and 95% posterior credibility intervals for each of the hyper-parameters, $\mathbb{E}\left(\bar{\Theta}|\mathbf{D}\right)$ and the diagonal elements of $\mathbb{E}\left(\Sigma|\mathbf{D}\right)$ in Table 15. As expected, Chobani has, overall, the highest mean brand preference, $\alpha_{Chobani}$, and Roundy's has the lowest overall brand preference, $\alpha_{Roundy's}$. Also as expected, the posterior probability that the mean price effect is negative is 100%.

–insert Table 15 here–

Consistent with the DID analysis presented previously, the blind taste test has a causal effect on utility. Interestingly, we find an effect on the marginal utilities consumers assign to both Roundy's and Chobani. In the week after the test (Short Run), the expected brand taste for Roundy's increases, while the expected brand taste for Chobani decreases. The expected utility gap between the two brands shrinks by almost 80%. From 7 days to 30 days after the test (Medium Run), the expected brand utility for Chobani decreases even more, although the increase for Roundy's has diminished substantially and we cannot reject an effect of zero. At the same time, we also cannot reject that the medium-term effect on Roundy's is positive and of the same magnitude as the long-term effect. The expected utility gap between Chobani and Roundy's is nevertheless still 41% smaller than during the pre-test period. Finally, between 30 and 150 days after the test (Long Run), the impact on Chobani is very small and we cannot reject that it's expected utility is unchanged from the pre period. In contrast, we find a strong positive effect on Roundy's leading to an expected utility gap that is still 22% lower.

Table 15 also reveals considerable heterogeneity in preferences, including in the estimated treatment effect of the blind taste test. In particular, our findings are consistent with heterogeneity in the sign of the treatment effect, with a non-trivial minority of consumers having a potential

negative treatment effect on their propensity to buy the private label in the long-run. In the long-run, almost 70% of the mass of the treatment effect on Roundy's is positive: $\Pr\left(\beta_{PL}^{LR} > 0 | \mathbf{D}\right) = 0.7$. This finding is consistent with our valence analysis above whereby half our respondents derived a positive signal, and almost 20% derived either a negative signal or a neutral signal confirming their predicted preference for Chobani. We will allow for this heterogeneity in the treatment effects when we conduct our information counterfactual below.

In sum, our findings of a treatment effect of the blind taste test on preferences survives our controls for prices, brand choice and the purchase incidence choice. Moreover, it does not appear to be moderated by purchase feedback through a brand loyalty mechanism.

5.5   The Impact of Information on Demand

In addition to controlling for point-of-purchase demand shifters, our key objective with the structural model is to quantify the magnitude of the information effect on the market structure for the Greek Yogurt category. We use our demand estimates from the previous section to simulate the impact of this counter-factual information treatment scenario on the entire consumer population. We compare the posterior mean baseline demand with no information treatment to the short-run, medium-run and long-run demand that would prevail if all consumers received the blind taste test treatment.

A critical assumption for our analysis is that the estimated treatment effects on brand preferences reported in section 5.4 reflect information about consumers' objective tastes for the brands. Our analysis of valence in section 4.4 is suggestive that receiving a positive signal from the intervention causes a larger treatment effect on brand choice. However, we cannot directly test this assumption. In Figure 4, we plot the posterior expected demand for Roundy's Greek yogurt holding its competitors' prices fixed at their mean levels during the sample period. The demand curve is conditional on purchase so we can compare with our DID results. The plot compares the posterior mean and 95% credibility interval for expected untreated demand (control) as well as the short run, medium run and long run demands along a wide range of prices. As a reference, we indicate expected demand at Roundy's average in-sample posted price of $1.03. At each point along the grid of prices, there is close to a 100% posterior probability that the short-run, medium-run and long-

run demand exceeds the baseline demand. Consistent with our discussion of the hyper-parameters, we cannot distinguish statistically between the medium and long-run effect sizes. However, it is clear that the blind taste test significantly shifts out demand for the Roundy's private label brand even in the long run.

–insert Figure 4 here–

To quantify the economic magnitude of the demand effects, the three panels in Figure 5 plot the distribution of the posterior expected shift in Roundy's demand across a range of prices for the SR, MR and LR respectively. Once again, we hold all the competitors' prices fixed at their average in-sample levels. At each price point, we report the expected magnitude of the shift in the demand relative to the untreated (pre) level (in share points) as well as the the 95% credibility interval, indicated by the whiskers. As a reference, we once more indicate the expected demand shift at Roundy's average in-sample posted price of $1.03. We can see that the outward shifts in demand are significant, even in the long-run. Although not reported in the figure, the posterior mean market share (conditional on purchase) for Roundy's at the average posted price of $1.03 increases relative to the baseline by 22.1 percentage points in the short run, 6.0 percentage points in the medium run, and 3.2 percentage points in the long run. These magnitudes are slightly larger (especially in the short-run) than the treatment effects found with our DID estimator in section 4.2. Recall that the DID estimator reported the average treatment effect on the treated whereas we are now using our model to measure the average treatment effect for the consumer population, having controlled for prices, brand choice, purchase incidence and taste heterogeneity. Moreover, our analysis herein focuses on share of total Greek Yogurt sales, not just the share of sales to the private label and leading national brand. In sum, if we interpret our treatment effects as information, then even though the treatment effect depreciates over time, private label would still command an additional 3 percentage points of market shares five months after the intervention. As a comparison, Dubé et al. (2018) document a long-term trend of 0.45 percentage points per year in the growth of CPG private labels in the US.

–insert Figure 5 here–

We also plot the analogous conditional demands and distributions of demand shifts for Chobani in Figures 6 and 7, respectively. We predict a large decline in share for Chobani in the short run and,

even more, in the medium run. In the long run, we do not expect to observe a permanent change in Chobani's share of Greek Yogurt sales. However, the differences between baseline and long-run Chobani demand are measured imprecisely and, using a 95% posterior credibility interval, we cannot rule out a long-run decline in share as large as 3 percentage points, which would account for almost all of the long-term gains to the private label.

–insert Figure 6 here–

–insert Figure 7 here–

If we interpret our long-run estimate as the permanent effect of the information treatment, then the private label would acquire an additional 3.2 percentage points of sales share (assuming prices do not re-adjust to their new equilibrium levels). Over half these gains come from Chobani, which loses 1.8 percentage points of sales share. But a sizeable shift, nearly 1 percentage point, comes from Noosa, which was not one of the tested brands.

Interestingly, the shift would benefit the retailer who obtains a higher margin on the sale of the private label ($0.31) than on Chobani ($0.27), in spite of the lower price charged for private label. Some of these gains derive from an expected 3.1% increase in the number of buyers. Overall, the long-run effect of information increases the retailer's expected total variable category profits by 3.7%, and its expected private label profits by 38%.

Therefore, we conclude that the direct and the competitive effects of counterfactually treating all consumers to the blind taste test intervention are statistically and substantively important and, holding prices fixed, would lead to substantial substitution to the private label even over a longer horizon of half a year. These findings survive our controls for prices, brand choice and purchase incidence. If we interpret the treatment effect as information, these findings add to our knowledge of information and the brand premium. Whereas Bronnenberg et al. (e.g., 2015) observed a role for objective information (e.g., efficacy of headache medecine), we now document a role for subjective information (e.g., the tastiness of food products).

6 Conclusions

Our findings add to the growing literature studying the implications of consumer misinformation. For the categories studied, the private label alternative has a lower market share than the leading national brands. The majority of participants in our blind taste tests self-reported a high perception of the quality of private labels, describing them as at least as good as the leading national brand. While this response could reflect acquiescence bias during the survey, it is nevertheless surprising that a much smaller proportion of these same respondents predicted they would pick the private label over the top national brand in a blind taste test.

Using three blind taste tests, we find that the majority of test participants chose the private label over the leading national brand. Using a difference-in-differences approach, we find that participation in the blind taste test has a persistent positive treatment effect on the treated consumers' demand for the private label, even five months after the intervention. This finding is robust to a large-dimensional factor model structure that relaxes the usual parallel trends assumption required in the DID specification. An exploratory analysis shows a strong association between the valence of the information conveyed by the taste test (i.e. positive versus negative signal) and the impact on future purchase behavior.

We use a structural choice model to measure the effect of the information treatment on Greek Yogurt demand, while controlling more thoroughly for the causal factors at the point of sale. We then use the estimates to measure the impact of the blind taste test on the category market structure. If we interpret the taste test estimates as an information treatment, then we predict that more informed consumers would be more likely to purchase Greek Yogurt, and would be more likely to pick the private label. Moreover, total retail category profits would be higher, mostly due to consumer substitution towards the higher-margin private label product.

These information effects are qualitatively different from the usual in-store advertising effects, like a display, in terms of their duration. Indeed, past research has not documented such direct long-term effects from in-store advertising. The information effects are also much larger than the typical estimates of traditional, e.g., television, advertising effects on demand (see, e.g., Sethuraman et al., 2011).

Our findings also contribute to the literature studying the barriers to entry created by established

brands. Even though the majority of our consumers pick the private label in the blind taste test, only a minority predicted they would. This inconsistency illustrates the obstacles facing the launch and growth of new brands, including private labels that may provide comparable value to consumers at a lower price. In the case of private labels, the information effect on shares after 5 months is 7 times larger than the long-term trend in recent growth in CPG private label shares in the US.

Finally, our findings add to the established wisdom on free-sampling campaigns. The finding of a long lasting effect suggests an investment benefit from our informative, non-price promotion, in contrast with what has been detected in past work regarding price promotions. The depreciation of the effect is also consistent with past theoretical work allowing for learning and forgetting from sampling campaigns. Our results suggest that on-going repetitions of the information treatment may be necessary to generate a more permanent benefit to private labels. In future work, it would be interesting to explore whether such on-going non-price promotions would be cost effective as a long-term strategy. More broadly, it would be interesting to investigate whether repeated information treatments could be sufficient to overwhelm the barriers created by brand capital for the leading national brands.

References

A. Abadie. 2005. Semiparametric difference-in-differences estimators. *Review of Economic Studies* 72, 1 (2005), 1–19.

Daniel A. Ackerberg. 2003. Advertising, Learning, and Consumer Choice in Experience Good Markets: An Empirical Examination. *International Economic Review* 44, 3 (2003), 1007–1040.

Hunt Allcott and Christopher Knittel. 2019. Are Consumers Poorly-Informed about Fuel Economy? Evidence from Two Experiments. *American Economic Journal: Economic Policy* 11, 1 (2019), 1–37.

Joshua D. Angrist and Alan B. Krueger. 1999. Chapter 23 - Empirical Strategies in Labor Economics. Handbook of Labor Economics, Vol. 3, Part A. Elsevier, 1277 – 1366. https://doi.org/10.1016/S1573-4463(99)03004-7

Gert Assmus, John U. Farley, and Donald R. Lehmann. 1984. How Advertising Affects Sales: Meta-Analysis of Econometric Results. *Journal of Marketing Research* 21, 1 (1984), 65–74.

Susan Athey, Mohsen Bayati, Nikolay Doudchenko, and Guido Imbens. 2017. Matrix Completion Methods for Causal Panel Data Models. *Stanford GSB Working Paper* (2017).

Susan Athey and Guido Imbens. 2018. Design-based Analysis in Difference-In-Differences Settings with Staggered Adoption. *Working Paper* (2018).

Jushan Bai. 2009. Panel data models with interactive fixed effects. *Econometrica* 77, 4 (2009), 1229–1279.

Joe S. Bain. 1956. *Barriers to New Competition*. Cambridge: Harvard University Press.

Kapil Bawa and Robert Shoemaker. 2004. The Effects of Free Sample Promotions on Incremental Brand Sales. *Marketing Science* 23, 3 (2004), 345–363.

Bryan Bollinger, Philip Leslie, and Alan Sorensen. 2011. Calorie Posting in Chain Restaurants. *The American Economic Journal: Economic Policy* 3(1) (2011), 91–128.

Bart J. Bronnenberg, Sanjay K. Dhar, and Jean-Pierre Dubé. 2009. Brand History, Geography, and the Persistence of Brand Shares. *Journal of Political Economy* 117 (2009), 87–115.

Bart J. Bronnenberg, Jean-Pierre Dubé, and Matthew Gentzkow. 2012. The Evolution of Brand Preferences: Evidence from Consumer Migration. *American Economic Review* 102, 6 (2012), 2472–2508.

Bart J. Bronnenberg, Jean-Pierre Dubé, Matthew Gentzkow, and Jesse M. Shapiro. 2015. Do Pharmacists Buy Bayer? Sophisticated Shoppers and the Brand Premium. *Quarterly Journal of Economics* 130 (2015).

Bart J. Bronnenberg, Jean-Pierre Dube, and Sridhar Moorthy. 2019. *The Handbook of the Economics of Marketing*. Vol. 1. Elsevier, Chapter The Economics of Brands and Branding.

Mariana Carrera and Sofia Berto Villas-Boas. 2015. Generic aversion and observational learning in the over-the-counter drug market. *Working Paper* (2015).

Darral G. Clarke. 1976. Econometric Measurement of the Duration of Advertising Effect on Sales. *Journal of Marketing Research* 13, 4 (1976), 345–357.

Steven R. Cox, Kenneth A. Coney, and Peter F. Ruppe. 1983. The impact of comparative product ingredient information. *Journal of Public Policy & Marketing* 2 (1983), 57–69.

Clement de Chaisemartin and Xavier D'Haultfoeuille. 2019. Two-way fixed effects estimators

with heterogeneous treatment effects. *Working Paper* (2019).

Jean-Pierre Dubé, Guenter Hitsch, and Peter Rossi. 2018. Income and Wealth Effects on Private Label Demand: Evidence From the Great Recession. *Marketing Science* 37, 1 (2018), 22–53.

Jean-Pierre Dubé, Günter J. Hitsch, and Puneet Manchanda. 2005. An Empirical Model of Advertising Dynamics. *Quantitative Marketing and Economics* 3 (2005), 107–144.

Jean-Pierre Dubé, Günter J. Hitsch, and Peter E. Rossi. 2010. State dependence and alternative explanations for consumer inertia. *RAND Journal of Economics* 41, 3 (2010), 417–445.

Tülin Erdem and Michael P. Keane. 1996. Decision-Making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets. *Marketing Science* 15, 1 (1996), 1–20.

Karen Gedenk and Scott A. Neslin. 1999. The Role of Retail Promotion in Determining Future Brand Loyalty: Its Effect on Purchase Event Feedback. *Journal of Retailing* 75, 4 (1999), 433–459.

Moshe Givon and Dan Horsky. 1990. Untangling the Effects of Purchase Reinforcement and Advertising Carryover. *Marketing Science* 9, 2 (1990), 171–187.

Andrew Goodman-Bacon. 2018. Difference-in-Differences with variation in Treatment Timing. *NBER Working Paper 25018* (2018).

Amir Heiman, Bruce McWilliams, Zhihua Shen, and David Zilberman. 2001. Learning and Forgetting: Modeling Optimal Product Sampling Over Time. *Management Science* 47, 4 (2001), 532–546.

Kosuke Imai and In Song Kim. 2019. On the Use of Two-way Fixed Effects Regression Models for Causal Inference with Panel Data. *Working Paper* (2019).

Karl W. Irons, John D. C. Little, and Robert L. Klein. 1983. *Proc. 1983 ORSA/TIMMS Marketing Sci. Conf.* Chapter Determinants of coupon effectiveness, 157–164.

Ginger Z. Jin and Philip Leslie. 2003. The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards. *The Quarterly Journal of Economics* 118(2) (2003), 409–451.

Michael P. Keane. 1997. Modeling Heterogeneity and State Dependence in Consumer Choice Behavior. *Journal of Business & Economic Statistics* 15, 3 (1997), 310–327.

Robert L. Klein. 1981. *Marketing: Measurement and Analysis. Proc. Third ORSA/TIMS Special Interest Conf. on Market Measurement and Analysis.* Institute of Management Sciences, Chapter Using supermarket scanner panels to measure the effectiveness of coupon promotions, 118–124.

Puneet Manchanda, Peter E. Rossi, and Pradeep Chintagunta. 2004. Response Modeling with Nonrandom Marketing-Mix Variables. *Journal of Marketing Research* 41, 4 (2004), 467–478.

Nitin Mehta, Surendra Rajiv, and Kannan Srinivasan. 2004. Role of Forgetting in Memory-Based Choice Decisions: A Structural Model. *Quantitative Marketing and Economics* 2 (2004), 107–140.

Michael A. Newton and Adrian E. Raftery. 1994. Approximate Bayesian Inference with the Weighted Likelihood Bootstrap. *Journal of the Royal Statistical Society, Series B* 56, 1 (1994), 3–48.

Peter Rossi, Greg Allenby, and Rob McCulloch. 2005. *Bayesian Statistics and Marketing.* John Wiley & Sons.

Navdeep S. Sahni. 2012. Effect of Temporal Spacing between Advertising Exposures: Evidence from Online Field Experiments. *manuscript* (2012).

Richard Schmalensee. 1982. Product Differentiation Advantages of Pioneering Brands. *The American Economic Review* 72 (1982), 349–365.

Raj Sethuraman, Gerard J. Tellis, and Richard A. Briesch. 2011. How Well Does Advertising Work? Generalizations from Meta-Analysis of Brand Advertising Elasticities. *Journal of Marketing Research.* 48, 3 (2011), 457–471.

Andrey Simonov, Jean-Pierre Dubé, Guenter Hitsch, and Peter E. Rossi. 2019. State-Dependent Demand Estimation with Initial Conditions' Correction. *Chicago Booth Working Paper* (2019).

Table 1: Number of participants in each blind taste test session

Store Number and Location	Cookies				Ice Cream		Greek Yogurt	
	10/15	10/20	10/23	10/26	10/22	10/25	10/21	10/24
8502 (Vernon Hills)		22	37	20	46	76	47	29
8509 (Frankfort)		17	11	65	35	41	46	51
8515 (Chicago)		12	63	25	17	27	18	14
8516 (Chicago)		5	5	30	15	34	4	3
8529 (Western Springs)	24	30	27	55	34	75	25	34

Table 2: Descriptive Statistics of Estimation Sample

sub-sample	variable	pooled	Cookies	Ice Cream	Greek Yogurt
control panelists	total number of panelists	16244	4266	7529	6793
	total number of transactions	208020	12418	49555	146047
	transactions per household	12.806	2.911	6.582	21.500
	(standard deviation)	(24.906)	(3.693)	(9.506)	(33.290)
	private label share of transactions	0.120	0.062	0.198	0.065
	(standard deviation)	(0.271)	(0.224)	(0.331)	(0.198)
test panelists	total number of panelists	440	99	185	156
	total number of transactions	3790	291	1052	2447
	transactions per household	8.614	2.939	5.686	15.686
	(standard deviation)	(17.107)	(3.063)	(7.695)	(25.913)
	private label share of transactions	0.186	0.169	0.248	0.124
	(standard deviation)	(0.341)	(0.345)	(0.374)	(0.282)

Note: Transactions and private label share of transactions are per household. Standard deviations are computed across households.

Table 3: SKUs, the underlying UPCs, the average shelf price, and the average share of Greek Yogurt volume sold across all stores and days

Brand	UPCs	Average Price	Average volume share of Greek Yogurt
Chobani	39	1.21	0.21
Dannon Oikos	38	1.19	0.08
Fage	21	1.34	0.14
Noosa	11	1.82	0.09
Roundy's	18	1.00	0.03
Yoplait Greek	33	1.16	0.05

Table 4: Choice shares for greek yogurt

	Chobani	Dannon	Fage	Roundy's	Noosa	Yoplait	No Purchase
Unconditional Shares							
Control	0.021	0.008	0.014	0.003	0.009	0.007	0.939
Test	0.011	0.005	0.005	0.003	0.012	0.004	0.960
All	0.021	0.008	0.014	0.003	0.009	0.007	0.939
Conditional Shares							
	Chobani	Dannon	Fage	Roundy's	Noosa	Yoplait	
Control	0.348	0.125	0.230	0.044	0.140	0.112	
Test	0.285	0.118	0.129	0.078	0.288	0.101	
All	0.347	0.125	0.228	0.045	0.142	0.112	

Table 5: Descriptive statistics of the survey questions

	before taste test	after taste test	mean	st. dev.	N
PL as good as national brand?	X		0.809	0.394	664
Do you think you will prefer PL?	X		0.435	0.496	664
Did you prefer PL?		X	0.717	0.451	664
Next time, buy PL?		X	0.843	0.364	664

Note: The total number of respondents (N) is equal to the number of survey respondents who have a purchase history in the category surveyed.

Table 6: Updating of beliefs

Before taste test: Will prefer PL?	After taste test: Did prefer PL?		
	no	yes	total
no	136	239	375
	36.27%	63.73%	100.00%
yes	52	237	289
	17.99%	82.01%	100.00%
total	188	476	664
	28.31%	71.69%	100.00%

Note: The total number of respondents (N) is equal to the number of survey respondents who have a purchase history in the category surveyed.

Table 7: Valence of beliefs update and intention to purchase

| | | After taste test: Will buy PL next time? | | |
		no	yes	total
Valence of updating during taste test	negative update $(Q_2 = 1, Q_3 = 0)$	16 30.77%	36 69.23%	52 100.00%
	positive neutral $(Q_2 = 1, Q_3 = 1)$	12 5.06%	225 94.94%	237 100.00%
	negative neutral $(Q_2 = 0, Q_3 = 0)$	67 49.26%	69 50.74%	136 100.00%
	positive update $(Q_2 = 0, Q_3 = 1)$	9 3.77%	230 96.23%	239 100.00%
	total	104 15.66%	560 84.34%	664 100.00%

Note: The total number of respondents (N) is equal to the number of survey respondents who have a purchase history in the category surveyed. Q_2- Do you think you will prefer the Private Label brand? (pre-test) Q_3- Did you prefer the Private Label brand?

Table 8: Parallel trends

	Pooled	Cookies	Ice cream	Yogurt
constant	0.0784	0.0644	0.1771	0.0439
	(0.0034)	(0.0130)	(0.0111)	(0.0030)
trend (δ)	0.0000	-0.0002	0.0004	0.0000
	(0.0001)	(0.0004)	(0.0004)	(0.0001)
trend $\times$ treatment ($\omega\tau_h$)	-0.0004	0.0013	-0.0007	-0.0005
	(0.0008)	(0.0028)	(0.0024)	(0.0008)
panelist effects	X	X	X	X
N	107172	6063	25352	75757
R^2	0.6028	0.8477	0.6349	0.5884

Note: Regressions use a pre-treatment period of 150 days prior to a panelist's store visit on the day of a blind taste test. Standard errors in parentheses. All specifications include panelist fixed effects. The trend variables are in weeks.

Table 9: Difference in difference regressions

	Pooled	Cookies	Ice cream	Yogurt
constant	0.0788	0.0704	0.1913	0.0397
	(0.0046)	(0.0174)	(0.0143)	(0.0045)
0-6 days -β_{SR}	0.1513	0.4785	0.2167	0.0988
	(0.0193)	(0.0546)	(0.0602)	(0.0187)
7-27 days-β_{MR}	0.0775	0.1596	0.1248	0.0403
	(0.0170)	(0.0535)	(0.0417)	(0.0184)
28-157 days-β_{LR}	0.0232	0.0626	0.0270	0.0201
	(0.0079)	(0.0239)	(0.0237)	(0.0078)
panelist×cat fixed effects $-\alpha_{hc}$	X	X	X	X
week×cat fixed effects $-\gamma_{tc}$	X	X	X	X
$H_0 : \beta^{SR} = \beta^{MR}$	reject	reject		
$H_0 : \beta^{MR} = \beta^{LR}$	reject			reject
N	211810	12709	50607	148494
R^2	0.5754	0.7726	0.5695	0.5186

Note: Standard errors in parentheses. The regressions show the short, medium, and long run treatment effects of the blind taste tests on the probability of choosing the private label. The regressions account for fixed effects for each combination of panelist and category. The pre taste-test window is 150 days.

Table 10: Difference in difference regressions

	Pooled	Cookies	Ice cream	Yogurt
constant	0.0789	0.0701	0.1912	0.0396
	(0.0046)	(0.0174)	(0.0143)	(0.0045)
0-6 days -β_{SR}	0.1487	0.4738	0.2120	0.0963
	(0.0193)	(0.0548)	(0.0603)	(0.0187)
7-27 days-β_{MR}	0.0765	0.1600	0.1188	0.0404
	(0.0170)	(0.0535)	(0.0416)	(0.0184)
28-55 days-$\beta_{4\text{-}8\text{ weeks}}$	0.0527	0.0119	-0.0012	0.0833
	(0.0131)	(0.0366)	(0.0372)	(0.0132)
56-83 days-$\beta_{9\text{-}12\text{ weeks}}$	0.0132	0.2901	-0.0188	0.0157
	(0.0136)	(0.0623)	(0.0386)	(0.0133)
84-111 days-$\beta_{12\text{-}16\text{ weeks}}$	0.0037	0.0671	0.0358	-0.0128
	(0.0115)	(0.0311)	(0.0346)	(0.0114)
112-157 days-$\beta_{>16\text{ weeks}}$	0.0162	0.0566	0.0336	0.0090
	(0.0121)	(0.0339)	(0.0361)	(0.0119)
panelist×cat fixed effects $-\alpha_{hc}$	X	X	X	X
week×cat fixed effects $-\gamma_{tc}$	X	X	X	X
N	211810	12709	50607	148494
R^2	0.5755	0.7731	0.5696	0.5188

Note: Standard errors in parentheses. The regressions show the short, medium, and long run treatment effects of the blind taste tests on the probability of choosing the private label. The regression account for fixed effects for each combination of customer and category. The pre taste-test window is 150 days.

Table 11: Average Treatment Effect on the Treated over Time

	Cookies		Ice Cream		Yogurt		Pooled	
	ATT^{MC-NNM}	ATT^{DID}	ATT^{MC-NNM}	ATT^{DID}	ATT^{MC-NNM}	ATT^{DID}	ATT^{MC-NNM}	ATT^{DID}
short run (0-6 days)	0.51 (1.1e-03 ,1.00)	0.51 (-2e-04 ,0.99)	0.28 (0.06 ,0.50)	0.31 (0.10 ,0.52)	0.24 (0.08 ,0.39)	0.22 (0.09 ,0.43)	0.29 (0.18 ,0.41)	0.30 (0.16 ,0.41)
medium run (7-27 days)	0.17 (-0.02 ,0.66)	0.17 (-0.02 ,0.54)	0.09 (-0.02 ,0.24)	0.11 (-0.01 ,0.25)	0.06 (-0.02 ,0.15)	0.05 (-0.03 ,0.14)	0.08 (0.01 ,0.15)	0.08 (0.02 ,0.16)
long run (28-157)	0.08 (-0.01 ,0.20)	0.07 (-3.3e-03 ,0.18)	5e-04 (-0.07 ,0.08)	0.01 (-0.07 ,0.10)	0.03 (-0.02 ,0.09)	0.03 (-0.03 ,0.08)	0.02 (-0.02 ,0.07)	0.02 (-0.01 ,0.07)

Note: Point estimates represent means of the estimators across the $B = 200$ bootstrap replications that draw individual customers with replacement. Confidence intervals in square brackets represent the 2.5^{th} and 97.5^{th} quantiles of the estimators across the $B = 200$ bootstrap replications.

Table 12: Valence of information

	Baseline	Valence
constant	-0.032	0.069
	(0.050)	(0.034)
average treatment	0.044	
	(0.008)	
treatment after positive update ($\beta^{\text{neg}\rightarrow\text{pos}}$)		0.067
		(0.012)
treatment after negative update ($\beta^{\text{pos}\rightarrow\text{neg}}$)		-0.051
		(0.028)
treatment after positive confirmation ($\beta^{\text{pos}\rightarrow\text{pos}}$)		0.054
		(0.014)
treatment after negative confirmation ($\beta^{\text{neg}\rightarrow\text{neg}}$)		0.009
		(0.015)
panelist$\times$category fixed effects$-\alpha_{hc}$	X	X
week$\times$category fixed effects$-\gamma_{tc}$	X	X
N	211810	211810
R^2	0.573	0.575

Note: Standard errors in parentheses. Regressions account for fixed effects for each combination of panelist and category and are pooled across categories.

Table 13: Posterior Likelihood (Greek Yogurt)

	posterior log likelihood					
	no loyalty	loyalty	treatment & no loyalty	treatment & loyalty	treatment, no loyalty, & self-selection	treatment, loyalty, & self-selection
homogeneous	-285,943.8	-246,586.1	-285,915.4	-246,574.3	-	-
random coefficients	-177,802.0	-178,230.6	-177,699.9	-178,182.9	-177,588.3	-177,658.6

Note: Posterior log-likelihoods account for selection on observables. Cases with "self-selection" also account for selection on persistent, unobserved tastes.

Table 14: Self-Selected Participation into theBlind Taste Test (Greek Yogurt)

| Coefficient	mean	(2.5th, 97.5th)
intercept (λ)	-1.945	(-2.977,-0.917)
Avg. \$ basket size ($\lambda_{\$}$)	-0.037	(-0.053,-0.02)
Avg. items per basket (λ_{items})	0.076	(0.024,0.123)
$\alpha^h_{Chobani}$ ($\lambda_{Chobani}$)	0.127	(-0.063,0.307)
$\alpha^h_{Roundy's}$ ($\lambda_{Roundy's}$)	-0.034	(-0.193,0.097)
$\beta^{SR}_{Chobani}$ ($\lambda^{SR}_{Chobani}$)	-0.018	(-0.483,0.71)
$\beta^{SR}_{Roundy's}$ ($\lambda^{SR}_{Roundy's}$)	-0.334	(-0.838,0.159)
$\beta^{MR}_{Chobani}$ ($\lambda^{MR}_{Chobani}$)	0.393	(-0.059,0.874)
$\beta^{MR}_{Roundy's}$ ($\lambda^{MR}_{Roundy's}$)	-0.497	(-1.014,0.134)
$\beta^{LR}_{Chobani}$ ($\lambda^{LR}_{Chobani}$)	0.268	(-0.084,0.648)
$\beta^{LR}_{Roundy's}$ ($\lambda^{LR}_{Roundy's}$)	-0.151	(-0.552,0.282)

Note: Population parameters estimated using model with self-selected participation into theblind taste test. In terms of notation, the row with label $\alpha^h_{Chobani}$ ($\lambda_{Chobani}$) reports the estimate for $\lambda_{Chobani}$, the impact of the preference component $\alpha^h_{Chobani}$ on participation.

Table 15: Hyper-parameter estimates (Greek Yogurt)

| Coefficient	Population mean		Population st. dev.	
	mean	(2.5th, 97.5th)	mean	(2.5th, 97.5th)
Chobani ($\alpha_{Chobani}$)	-2.159	(-2.386,-1.934)	4.931	(4.761,5.105)
Dannon Oikos (α_{Dannon})	-3.910	(-4.168,-3.652)	5.728	(5.551,5.905)
Fage (α_{Fage})	-3.719	(-3.938,-3.498)	5.061	(4.887,5.254)
Roundy's ($\alpha_{Roundy's}$)	-5.961	(-6.237,-5.685)	5.196	(5.021,5.389)
Noosa (α_{Noosa})	-4.556	(-4.998,-4.155)	7.332	(7.075,7.599)
Yoplait Greek ($\alpha_{Yoplait}$)	-4.732	(-4.998,-4.483)	5.179	(5.011,5.353)
price (η)	-2.875	(-3.053,-2.693)	3.780	(3.641,3.917)
(SR treat)*Chobani ($\beta^{SR}_{Chobani}$)	-0.262	(-0.423,-0.11)	1.016	(0.761,1.338)
(SR treat)*Roundy's ($\beta^{SR}_{Roundy's}$)	2.719	(2.241,3.031)	1.345	(1.13,1.996)
(MR treat)*Chobani ($\beta^{MR}_{Chobani}$)	-1.406	(-1.637,-1.162)	1.076	(0.917,1.199)
(MR treat)*Roundy's ($\beta^{MR}_{Roundy's}$)	0.151	(-0.165,0.464)	1.532	(1.286,1.815)
(LR treat)*Chobani ($\beta^{LR}_{Chobani}$)	-0.085	(-0.430,0.338)	1.377	(1.181,1.568)
(LR treat)*Roundy's ($\beta^{LR}_{Roundy's}$)	0.756	(0.282,1.175)	1.653	(1.48,1.900)

Note: Population parameters estimated using model with self-selected participation into the blind taste test (only diagonal elements of Σ are reported)

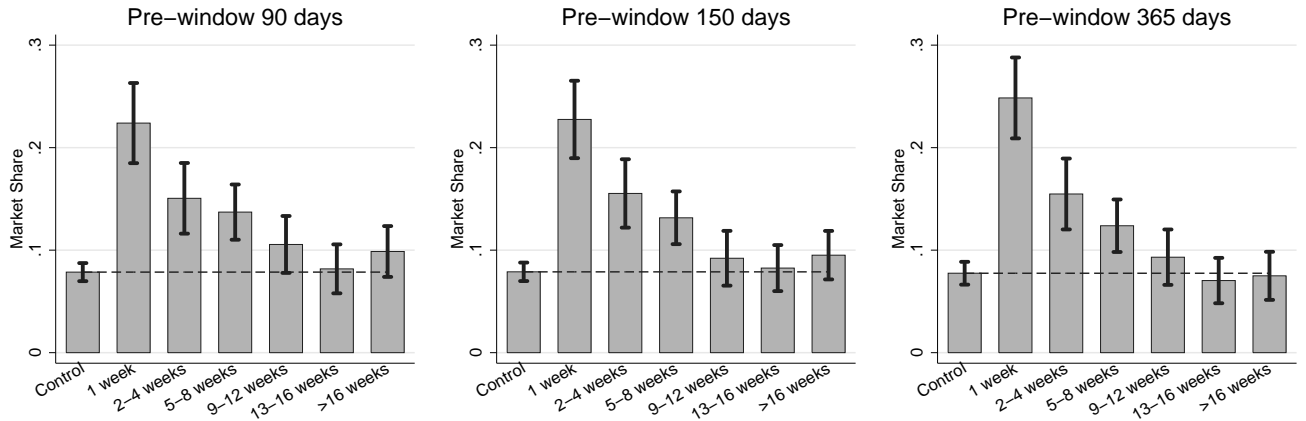## Figure 1: Robustness – Panelists not Buying on the Treatment Day
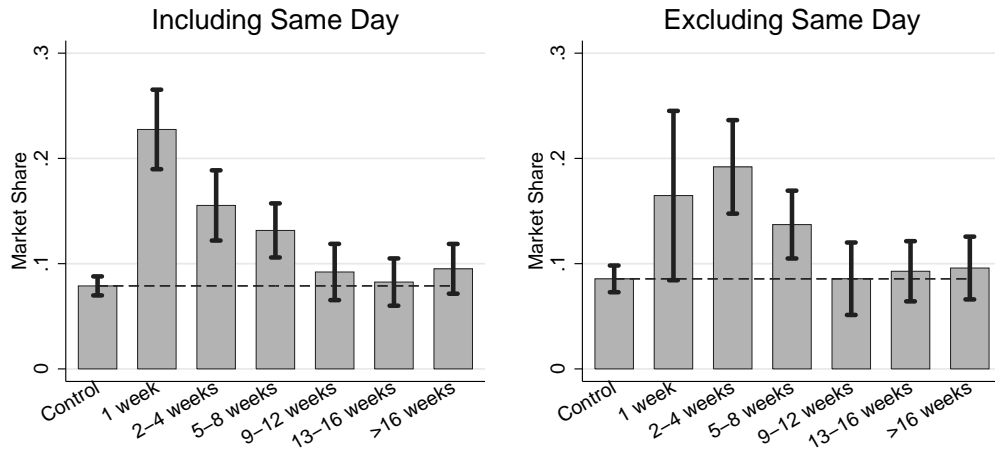
Note: The graph on the left reproduces the first column in table (10). The graph on the right represents the case where the difference-in-differences regression is estimated on the sub-sample of treated panelists who do not buy from the category on the treatment-day.

Figure 2: Robustness – Different Scope of Category
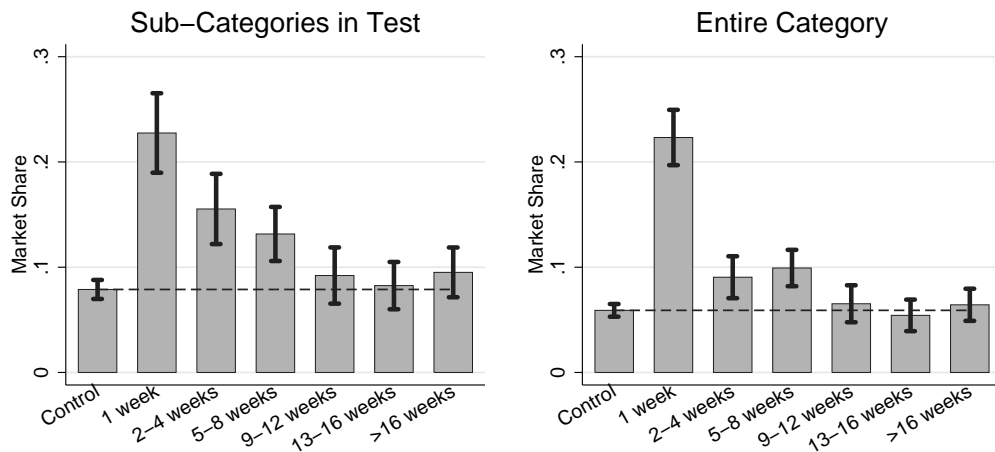
Note: The graph on the left reproduces the first column in table (10). The graph on the right represents the case where the difference-in-differences regression is estimated using all items in the three categories.

Figure 3: Robustness –Different Pre-Treatment Windows

Note: The graph in the middle reproduces the first column in table (10). The graph on the left and right represent the case where the difference-in-differences regression is estimated using shorter and longer pre-treatment windows, respectively.
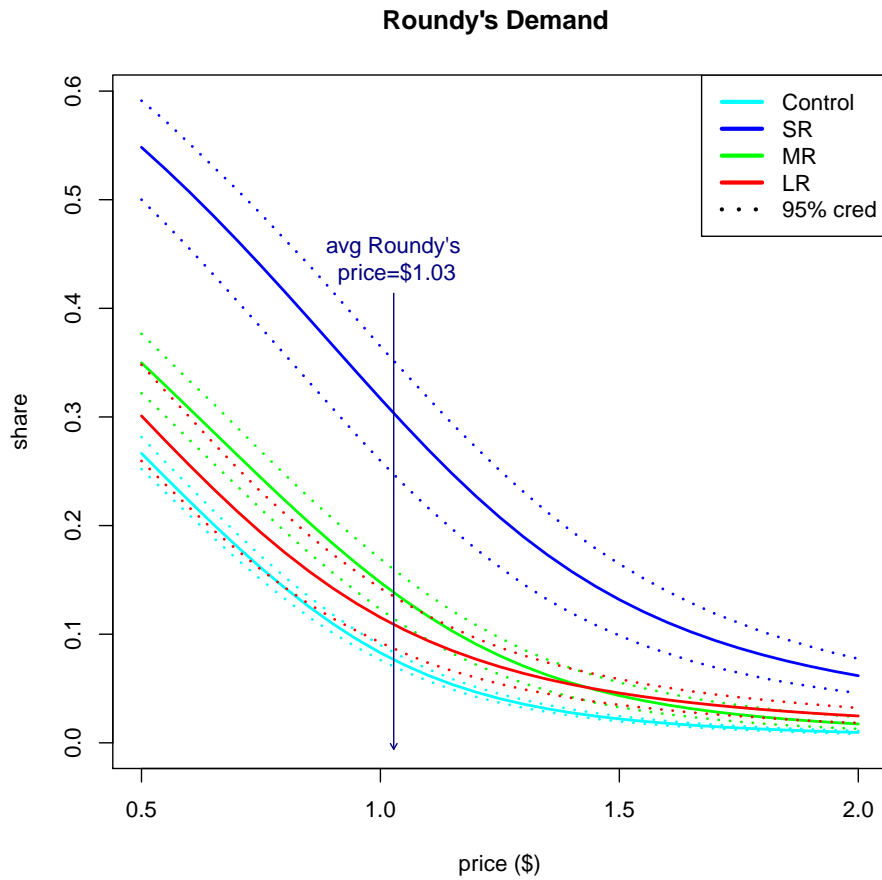
Figure 4: Posterior Expected Demand for Roundy's

Figure 5: Posterior Expected Demand Shifts for Roundy's

Distribution of Roundy's Demand Shifts (SR)

Distribution of Roundy's Demand Shifts (MR)

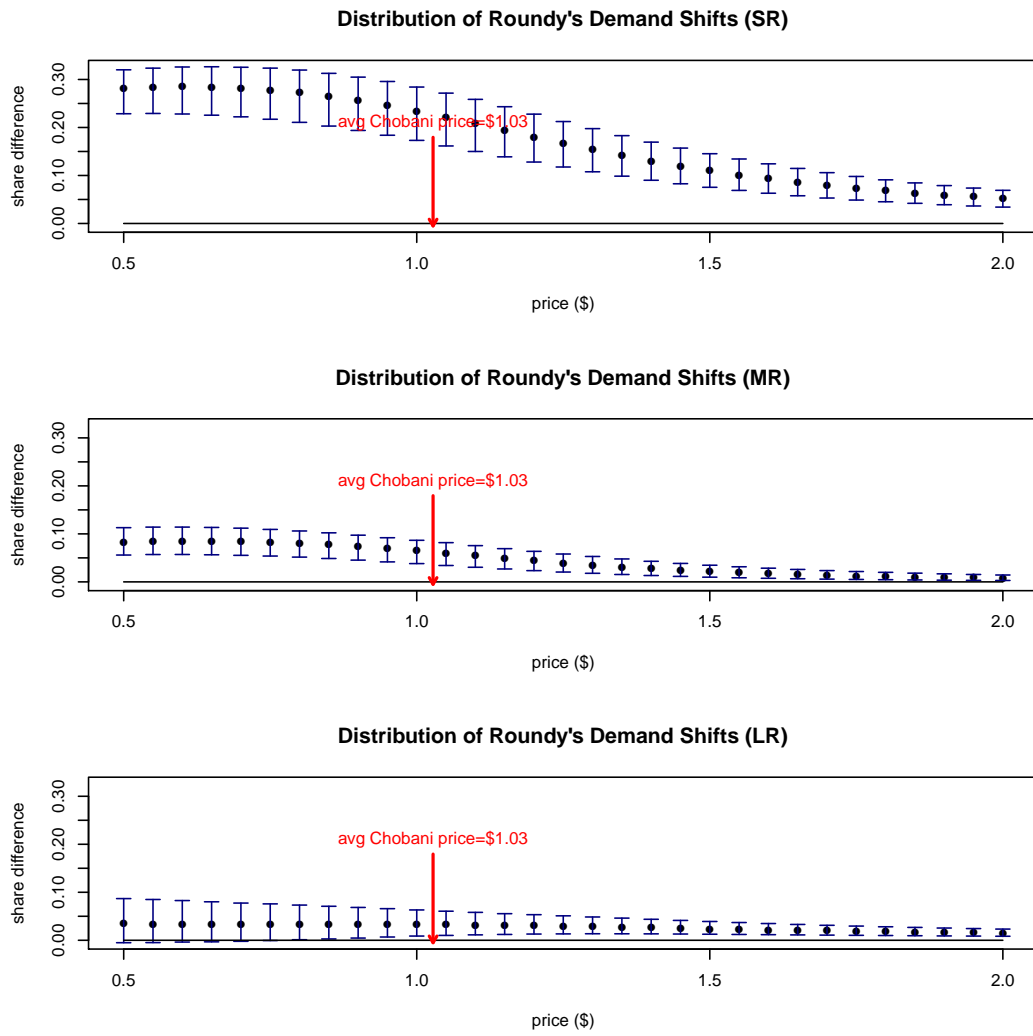**Distribution of Roundy's Demand Shifts (LR)**

Figure 6: Posterior Expected Demand for Chobani
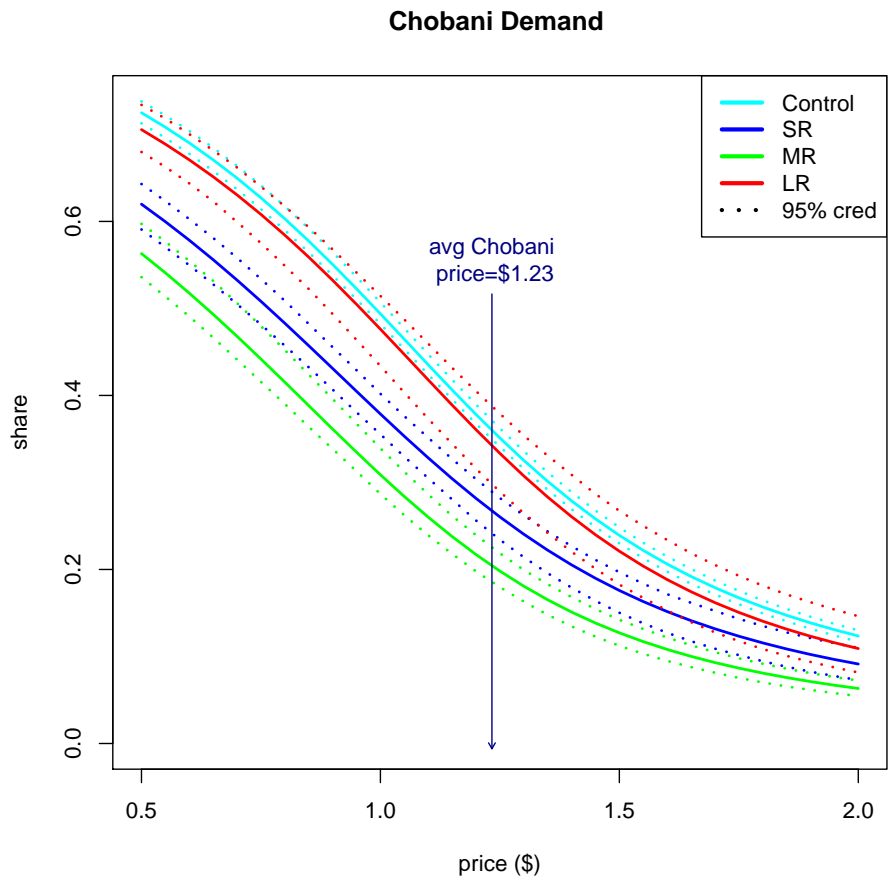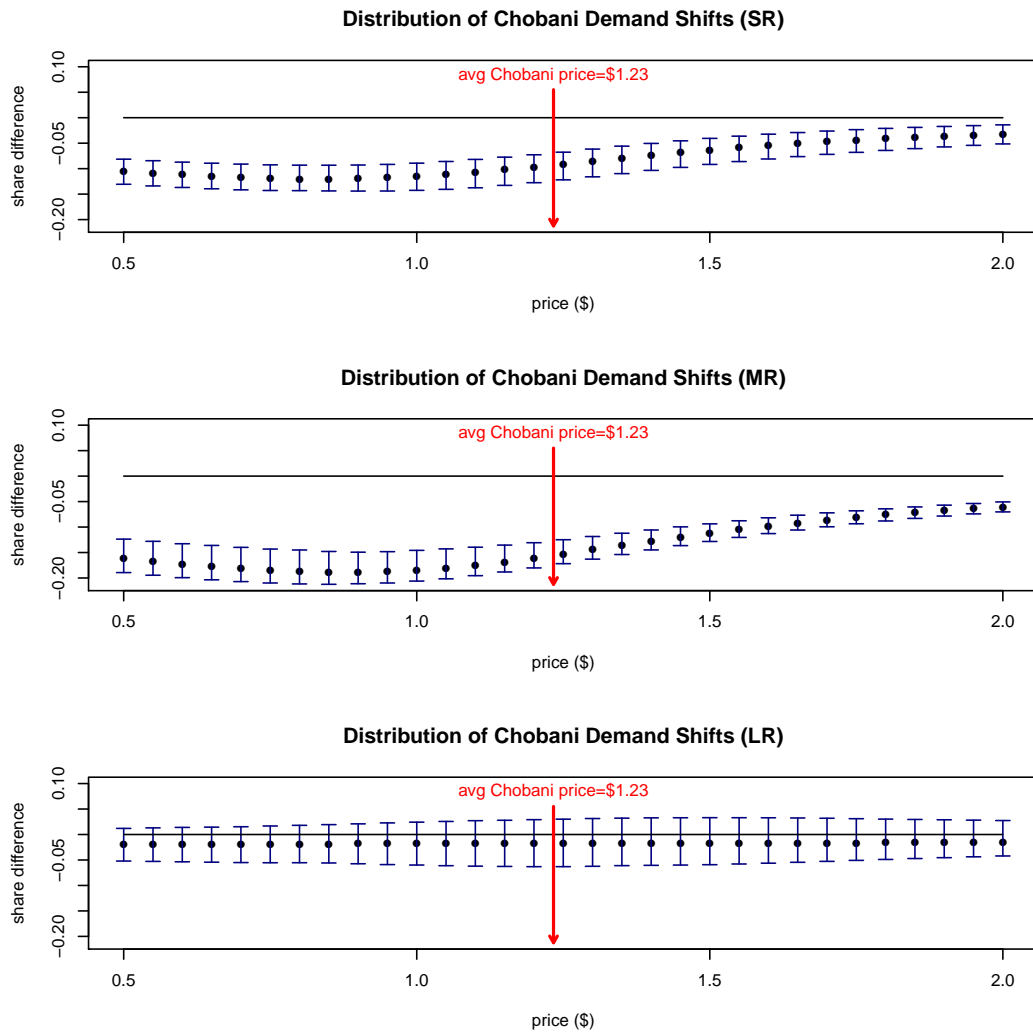

Chobani Demand

Figure 7: Posterior Expected Demand Shifts for Chobani

A Difference Regressions within Treatment Group

Table 16: Difference regressions

	Pooled	Cookies	Ice cream	Yogurt
constant	0.0765	0.0725	0.2026	0.0244
	(0.0102)	(0.0386)	(0.0280)	(0.0100)
0-6 days -β_{SR}	0.1832	0.4751	0.2425	0.1380
	(0.0242)	(0.0855)	(0.0688)	(0.0238)
7-27 days-β_{MR}	0.1271	0.1502	0.1433	0.1024
	(0.0226)	(0.0867)	(0.0535)	(0.0240)
28-157 days-β_{LR}	0.0821	0.0544	0.0122	0.1113
	(0.0205)	(0.0798)	(0.0547)	(0.0204)
linear time trend - γ	-0.00032	0.00014	0.00009	-0.00051
	(0.00011)	(0.00046)	(0.00029)	(0.00011)
panelist×cat fixed effects $-\alpha_{hc}$	X	X	X	X
N	3790	291	1052	2447
R^2	0.5991	0.7617	0.5662	0.5774

Note: Standard errors in parentheses. The regressions show the short, medium, and long run treatment effects of the blind taste tests on the probability of choosing the private label. The regressions account for fixed effects for each combination of panelist and category. The pre taste-test window is 150 days.

Table 17: Difference regressions

	Pooled	Cookies	Ice cream	Yogurt
constant	0.0938	0.0526	0.2179	0.0423
	(0.0104)	(0.0433)	(0.0276)	(0.0103)
0-6 days -β_{SR}	0.1644	0.4848	0.2292	0.1162
	(0.0244)	(0.0877)	(0.0691)	(0.0239)
7-27 days-β_{MR}	0.1068	0.1651	0.1262	0.0809
	(0.0227)	(0.0897)	(0.0531)	(0.0241)
28-55 days-$\beta_{\text{4-8 weeks}}$	0.0765	0.0106	-0.0231	0.1303
	(0.0207)	(0.0800)	(0.0535)	(0.0208)
56-83 days-$\beta_{\text{9-12 weeks}}$	0.0358	0.3047	-0.0516	0.0643
	(0.0234)	(0.1175)	(0.0595)	(0.0233)
84-111 days-$\beta_{\text{12-16 weeks}}$	0.0277	0.0994	-0.0097	0.0384
	(0.0242)	(0.1027)	(0.0639)	(0.0240)
112-157 days-$\beta_{\text{>16 weeks}}$	0.0459	0.0692	-0.0060	0.0722
	(0.0281)	(0.1126)	(0.0729)	(0.0283)
linear time trend- γ	-0.00012	-0.00001	0.00022	-0.00028
	(0.00012)	(0.00055)	(0.00030)	(0.00012)
panelist×cat fixed effects $-\alpha_{hc}$	X	X	X	X
N	3790	291	1052	2447
R^2	0.5991	0.7732	0.5668	0.5809

Note: Standard errors in parentheses. The regressions show the short, medium, and long run treatment effects of the blind taste tests on the probability of choosing the private label. The regression account for fixed effects for each combination of customer and category. The pre taste-test window is 150 days.

B   Inference

We base our inferences on the hierarchical model described in section 5. We use an MCMC algorithm to simulate the posterior distribution of the model parameters. At each stage r of the chain, we cycle through the following set of conditional draws:

1. Consumer tastes $\{\Theta^h\}$

Recall that $\{\Theta^h\} \sim N(\bar{\Theta}, \Sigma)$. Given draws for $\Lambda^{r-1}, \bar{\Theta}^{r-1}$ and Σ_Θ^{r-1}, we use a random-walk Metropolis-Hastings algorithm that samples a candidate draw $\Theta^{h,c} = \Theta^{h,r-1} + \tilde{\Theta}^h$ where $\tilde{\Theta}^h$ is drawn from a multivariate normal proposal density, $N\left(0, s^2\tilde{\Sigma}^h\right)$ with $\tilde{\Sigma}^h = \left(H^h + \left(\Sigma_\Theta^{r-1}\right)^{-1}\right)^{-1}$ and H^h is the Hessian of consumer $h's$ choice likelihood (ignoring selection) evaluated at the MLE for the fractional likelihood (see Rossi et al. (2005)). We then use the following acceptance rate for each

candidate draw

$$\Pr\left(\text{accept } \Theta^{h,c}\right) = \min\left\{\frac{\ell\left(\Theta^{h,c},\Lambda^{r-1}\right)\phi\left(\Theta^{h,c}|\bar{\Theta}^{r-1},\Sigma_{\Theta}^{r-1}\right)}{\ell\left(\Theta^{h,r-1},\Lambda^{r-1}\right)\phi\left(\Theta^{h,r-1}|\bar{\Theta}^{r-1},\Sigma_{\Theta}^{r-1}\right)},1\right\}$$

where $\phi\left(\Theta^{h,c}|\bar{\Theta}^{r-1},\Sigma_{\Theta}^{r-1}\right)$ is the density of a Normal distribution with mean and variance $\left(\bar{\Theta}^{r-1},\Sigma_{\Theta}^{r-1}\right)$ evaluated at $\Theta^{h,c}$.

2. The selection parameters Λ

Given draws for $\left\{\Theta^{h,r}\right\}$, we use a random-walk Metropolis-Hastings algorithm that samples a candidate draw $\Lambda^{c} = \Lambda^{r-1} + \tilde{\Lambda}$ where $\tilde{\Lambda}$ is drawn from a multivariate Normal proposal density, $N\left(0,(H_{\Lambda}+A_{\Lambda})^{-1}\right)$ where H_{Λ} is the Hessian of the maximized selection likelihood and A_{Λ} is a matrix of prior parameters set by the researcher. We then use the following acceptance rate for each candidate draw

$$\Pr\left(\text{accept } \Lambda^{c}\right) = \min\left\{\frac{\prod_{h}\Pr\left\{\tau^{h}=1|w^{h},\Theta^{h},\Lambda^{c}\right\}\phi\left(\Lambda^{h,c}|0,A_{\Lambda}\right)}{\prod_{h}\Pr\left\{\tau^{h}=1|w^{h},\Theta^{h},\Lambda^{r-1}\right\}\phi\left(\Lambda^{r-1}|0,A_{\Lambda}\right)},1\right\}.$$

3. The Population means $\bar{\Theta}$

Given draws of $\left\{\Theta^{h}\right\}$, standard conjugate theory can be used to generate draws of the hyper-parameters $\bar{\Theta}$ and Σ, where $\bar{\Theta}|\Sigma \sim N\left(\bar{\bar{\Theta}},a^{-1}\Sigma\right)$ and $\Sigma \sim IW(v,V)$.

C Robustness of Structural Estimates

To assess robustness of our estimates, we re-estimate the multinomial choice demand system for the Greek Yogurt data using a shorter pre-treatment time window of 150 days, to correspond to the sample used for the DID analysis in section 4. We compare the estimates for our best-fitting specification from section 5.4 using this shorter pre-treatment window in Table 18. Once again, we find that the short and medium run treatment effect of the blind taste test on consumers' mean brand preference for Roundy's are large and have a close to 100% posterior probability of being positive. The long run effect is small and has only a 73% posterior probability of being positive. Unexpectedly, the short run treatment effect on the mean brand preference for Chobani, though small, has a 99% posterior probability of being positive. The medium run treatment effect is, as before, large and has a close to 100% posterior probability of being negative. Also, as before, the long run effect is very small and has only a 58% probability of being negative. In summary, most of our qualitative findings from before are robust to the shorter pre-treatment time window. However, we lose precision in this shorter sample.

Table 18: Hyper-parameter estimates (Greek Yogurt) using 150-day pre-test window (as in DID analysis)

Coefficient	Population mean		Population st. dev.	
	mean	(2.5th, 97.5th)	mean	(2.5th, 97.5th)
Chobani $(\alpha_{Chobani})$	-2.553	(-2.838,-2.257)	5.341	(5.092,5.610)
Dannon Oikos (α_{Dannon})	-4.859	(-5.202,-4.515)	6.000	(5.763,6.284)
Fage (α_{Fage})	-4.305	(-4.631,-3.959)	5.409	(5.14,5.683)
Roundy's $(\alpha_{Roundy's})$	-7.068	(-7.470,-6.645)	5.631	(5.329,5.882)
Noosa (α_{Noosa})	-5.201	(-5.774,-4.633)	7.654	(7.284,8.065)
Yoplait Greek $(\alpha_{Yoplait})$	-5.161	(-5.472,-4.847)	5.496	(5.261,5.760)
price (η)	-2.635	(-2.869,-2.408)	3.957	(3.745,4.157)
(SR treat)*Chobani $(\beta_{Chobani}^{SR})$	-1.037	(-1.287,-0.570)	1.199	(0.965,1.371)
(SR treat)*Roundy's $(\beta_{Roundy's}^{SR})$	1.559	(1.226,2.168)	1.670	(1.116,2.411)
(MR treat)*Chobani $(\beta_{Chobani}^{MR})$	-1.027	(-1.412,-0.765)	1.298	(1.060,1.725)
(MR treat)*Roundy's $(\beta_{Roundy's}^{MR})$	0.382	(0.082,0.730)	1.386	(1.209,1.646)
(LR treat)*Chobani $(\beta_{Chobani}^{LR})$	1.20e-04	(-0.244,0.199)	1.300	(1.048,1.532)
(LR treat)*Roundy's $(\beta_{Roundy's}^{LR})$	-0.346	(-0.649,0.044)	1.520	(1.075,2.064)

Population parameters estimated using model with self-selected participation into theblind taste test.