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*

October 10, 2018

Abstract

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. In a survey administered during the taste test, 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. Surprisingly however, an overwhelming majority systematically chooses the private label in the blinded test. During the week after the intervention, the tested private label product market shares increase by 15 share points, 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, 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. The findings are consistent with a treatment effect of information on demand where the memory for this information decays slowly over time. Alternative explanations to the information treatment are ruled out.

JEL: L11, L15, M31, M37

*We are grateful to Bob Mariano and Roundy's Supermarkets for providing the data for this project. We benefitted from the comments and suggestions of Bryan Bollinger, Guido Imbens and Moshen Bayati, along with seminar participants at McGill University, the 2016 Data Science Academy at AC Nielsen, the 2016 Kilts Center Marketing Insights Conference and the 2016 Marketing Science Conference in Shanghai. Bronnenberg acknowledges research support from Stanford University as the 2017-2018 GSB Trust Faculty Fellow; Dubé acknowledges research support from the Kilts Center for Marketing and the Charles E. Merrill faculty research fund; and Sanders acknowledges the Marketing Science Institute for research support. E-mail addresses: bart.bronnenberg@tilburguniversity.edu, jdube@chicagobooth.edu, rsanders@rady.ucsd.edu.

1 INTRODUCTION

Private label brands in the consumer packaged goods (CPG) industry are still relatively underdeveloped 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, Dubé, Gentzkow, and Shapiro, 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.

[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.

It is not possible to design an in-store sampling promotion that randomly assigns subjects to information treatment conditions. Hence, participation in the blind taste test could be self-selected on unobserved aspects of consumer preferences for private labels. We propose a difference-in-differences (DID) approach to obtain a consistent estimate of the 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 fixed-effects specification that allows us to relax the parallel trends condition and allow for richer patterns both across time and across consumers.

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, 73% 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 and, and more in line with the initial self-reported quality beliefs. Finally, after the identity of the chosen product was revealed, 83% 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 qualitatively unchanged when we re-run our analysis at the weekly level using a matrix completion estimator (Athey, Bayati, Doudchenko, and Imbens, 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, as documented in Bronnenberg, Dhar, and Dubé (2009) and Bronnenberg, Dubé, and Gentzkow (2012).

Our findings add to the extant literature on the role of consumer information. 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, Dubé, Gentzkow, and Shapiro, 2015). However, it is unclear whether policies that directly communicate objective product information to consumers have a material impact on their brand choices. Bollinger, Leslie, and Sorensen (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, Coney, and Ruppe, 1983; Carrera and Villas-Boas, 2015). Similarly, in the automobile industry, Allcott and Knittel (2017) 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, Dubé, and Gentzkow, 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, Rajiv, and Srinivasan, 2004). These findings are consistent with a small theoretical literature on free sampling that also allows for forgetting (e.g., Heiman, McWilliams, Shen, and Zilberman, 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, Farley, and Lehmann, 1984; Dubé, Hitsch, and Manchanda, 2005; Sahni, 2012); although our blind taste test generates a much larger effect on demand than impressions from traditional advertising media.

Our findings also 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, Little, and Klein, 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, and comprise 1,329,900 unique customers and over 29 unique Mariano's stores. We retain the UPC-level purchase information for our 3 categories of interest: ice cream, yogurt and shelf-stable packaged cookies. 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 through 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 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.[2] 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] In the store with only 3 trials, the sampling booth only ran for 20 minutes and was shut down early.

Breyers (Ice Cream). At this point, 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?" 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 member participating in the same in-store test) of the first test (in case of exposure to 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. 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. The decision to use 157 days post treatment is due to the fact that our transaction data end at that time. In total, we retain 16,680 unique customers who make 211,790 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 private-label and branded Oreo style Cookies, 156 sampled branded and unbranded 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 categories. In total, we have 16,240 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. Appendix B discusses the criteria used to combine UPCs into a single SKU.

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.

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 36% of sales. Roundy's private label Greek yogurt is the lowest-share brand with only 5.6% of the market share. We observe some self-selection into the test sample. Our test panelists are more likely to buy Roundy's (6.7% versus 5.5% in control) and Noosa (24.0% versus 13.1% in control), but they are less likely to buy Chobani (26.7% versus 36.2% 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 those 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]?"[3] The

[3] We used Roundy's O's, Roundy's Select and Roundy's for the cookies, ice cream and Greek yogurt categories

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.[4] It is even more striking that 72% of the sample responded affirmatively to question three (Q3): “Did you prefer Roundy’s O’s Cookie or Nabisco Oreo Cookie?” The actual choice rate in Q3 was 27 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 12 percentage points higher than the choice rate in Q3, and 40 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 conditioning the sample on those subjects that purchased private label in the chain prior to the intervention.

The sharp increase in stated future private label purchase intention relative to predicted purchase is consistent with an information effect, but could also reflect a salience effect from the in-store taste test. 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.

---

respectively. We also used the national brands Nabisco Oreo, Chobani and Breyers for the cookies, Greek yogurt and ice cream categories respectively.

[4]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.

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 ($N = 371$), 63% 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 ($N = 293$), 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 constant the pre-test stated preference for the private label brand at $Q_2 = 1$, 68% ($N = 53$) of consumers who updated negatively ($Q_3 = 0$) indicated the intention to buy the private label next time. However, among those who remained positive ($Q_3 = 1$; $N = 240$), a strongly contrasting 95% indicated they would buy the private label next time. Holding constant, in rows 3 and 4, the pre-test stated preference for the private label brand at $Q_2 = 0$, among all who thought initially they would not prefer 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 ($Q_3 = 1$; $N = 235$), 97% 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 panel data described in section 2.4 to estimate the causal effect of the blind taste test on panelists' private label choices. Our approach consists of using a two-way fixed effects 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. In a subsequent section, we verify the robustness of our difference-in-differences (DID) estimator to some of its key underlying identifying assumption.

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., $\tau_{hct} \equiv \mathbb{I}_{\{treatment \text{ group},t\geq 0\}}$. 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 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:

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

As with most settings, the challenge is that we do not observe $(Y_{hct}(0) \mid \tau_{hct} = 1)$.

To resolve this problem, our key identifying assumption is that treated and untreated households follow *parallel trends*. More formally, we assume (e.g., Abadie, 2005):

$$\mathbb{E}\left(Y_{hct}(0) - Y_{hct'}(0) \mid \tau_{hct} = 1\right) = \mathbb{E}\left(Y_{hct}(0) - Y_{hct'}(0) \mid \tau_{hct} = 0\right), \text{where } t \neq t'. \tag{2}$$

This assumption imposes that any changes in Y_{hct} over time are independent of whether or not a household participated in the blind taste test.

We use the parallel trends assumption to form a standard two-way fixed-effects estimator of ATT_c based on the difference-in-differences 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:

$$\begin{aligned} Y_{hct} = \alpha_{hc} + \gamma_{ct} &+ \beta_{SR} \cdot \tau_{hct} \cdot \mathbb{I}_{\{t \in (0,6)\}} \\ &+ \beta_{MR} \cdot \tau_{hct} \cdot \mathbb{I}_{\{t \in (7,27)\}} \\ &+ \beta_{LR} \cdot \tau_{hct} \cdot \mathbb{I}_{\{t > 27\}} + \varepsilon_{hct}. \end{aligned} \tag{3}$$

The parameter α_{hc} is a panelist-category fixed effect, γ_{ct} is a category-week fixed effect,[5] and the indicator variables $\mathbb{I}_{\{t \in \mathscr{T}\}}$ denote whether trip date t falls in the time interval $\mathscr{T}$, 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 treatment effect 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 (3) that splits the long run effect, β_{LR}, into separate 4-week treatment effects, $(\beta_{\text{4-8 weeks}}, \beta_{\text{9-12 weeks}}, \beta_{\text{13-16 weeks}}, \beta_{\text{>16 weeks}})$.

4.2 DID and Identification

The heart of our identification strategy relies on the panel structure of our transaction data. To address the potential bias associated with self-selection into the blind taste, we include panelist×category-specific fixed effects in equation (3) to control for the persistent differences between panelists by category. In addition, we use the parallel trends assumption. While we cannot directly test this assumption, we can follow Angrist and Krueger (1999, p. 1299) to exploit the long time series

[5] 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.

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

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

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 7.9% 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 predicted mean difference of only 0.045 percentage points. However, the results are imprecise and we cannot rule out that the treatment group trend is as much as 0.21 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%.[6] In section 4.4, we will check the robustness of our ATT estimates to an estimator that uses a weaker assumption than parallel trends.[7]

–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. Nevertheless, below in section 4.4, we explore the robustness of our results to an estimator that relaxes the parallel trends assumption.

4.3 Treatment Effects

We now focus on the DID estimates using the linear probabilities model in equation (3). Table 9 presents the estimates for the DID regressions. We observe that the base share of the private label

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

[7]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.

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 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 off the pre-treatment data, controls for a constant trend in the pre- and post periods. The results in Tables 15 and 16 confirm the estimates of the DID analysis above. The larger effect sizes, 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 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 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. The figure shows that once we drop households who buy 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). Still, Figure 1 suggests a "day of the test" effect which might be capturing the salience effect of the sampling booth on choice. 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 private label versus national brand and present the results from using the purchases in the categories (Cookies, Ice Cream & Frozen Yogurt, and Yogurt) versus only the subcategories (Oreo-style Cookies, Ice Cream, and Greek Yogurt) used in the blind taste test. Figure 2 compares the results for each of the two choice set 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 long run under both definitions.

–insert Figure 2 here–

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.

–insert Figure 3 here–

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.

4.4 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, Bayati, Doudchenko, and Imbens (2017). As explained earlier, the key identification assumption underlying our DID estimates in section 4.3 is that treatment and control consumers' purchases follow parallel trends. Suppose, for instance, that some of the untreated consumers would never consider purchasing the private label. In that case, we would not observe any trend in their propensity to purchase private label and the mean trend in the control group would not likely be the same as that in the treated group. The interactive fixed-effects model no longer requires the same stable time paths in outcomes for treated and untreated households.

As before, we face the problem that we do not observe $(Y_{hct}(0)|\tau_{hct}=1)$. To impute these missing values, Athey, Bayati, Doudchenko, and Imbens (2017) model the outcomes in the following matrix form:

$$Y_{H\times T} = L_{H\times T} + \varepsilon_{H\times T} \tag{5}$$

where L is an $H\times T$ matrix to be estimated, and ε_{ht} is measurement error with $\mathbb{E}\left(\varepsilon|L\right)=0$. 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 DID 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 richer heterogeneity beyond unit fixed effects and common time-shocks. Unlike a standard factor model, Athey, Bayati, Doudchenko, and Imbens (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, Bayati, Doudchenko, and Imbens (2017):

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

where $\Theta = (\gamma_1,...,\gamma_N,\delta_1,...,\delta_T)$ are parameters to be estimated and the notation $\mathscr{O}$ denotes the set of observations, indexed by (h,t), for which the outcome is observed. Regularization depends on the penalty term, λ,[8] 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, Bayati, Doudchenko, and Imbens (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}

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

allow for time-varying unobservable heterogeneity.

Our estimate of the ATT is then:

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

For comparison purposes, we also use the following reformulation of our DID estimator to be comparable to the MC-NNM estimator above. Again, if we let $R = 2$ and $\gamma_{h2} = 1$ and $\delta_{t1} = 1$ (i.e. $L = 0$), our data-generating process simplifies to the two-way fixed-effects model again:

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

We can obtain OLS estimates $\hat{\gamma}_h$ and $\hat{\delta}_t$ with the special-case of $\lambda = 0$. Accordingly, a DID analog of the ATT can be obtained as follows:

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

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 households for which we do not observed any pre-treatment-period observations. [9] We define our dependent variable, Y_{ht},as the share of purchases that are private label and use this share as our dependent variable: $Y_{ht} = \frac{\#\ PL \text{ purchases by } h \text{ in week } t}{total\ \#\ purchases \text{ by } h \text{ in week } t}$. [10]

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

[9] Although not reported herein, the ATT^{DID} estimates are quite similar to those based on equation (3), reported above in section 4.3, suggesting that the deletion of households does not alter our key findings. Results are available from the authors upon request.

[10] 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.

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.

4.5 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 (3) with a single average (over time) treatment effect and an interaction between the treatment effect and the valence of the information:

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

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 concentrated 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 -10.3 percentage points or a positive effect as large as 0.9 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 -1.5 percentage points or as large as 4.1 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

Continuing our analysis, we now estimate a structural model 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.

An additional 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.

Finally, we use our demand estimates to simulate a counterfactual wherein all consumers are part of the blind taste test. Extrapolating from the average treatment effect on the treated consumers to the full population of consumers visiting the store requires additional structure and assumptions, which we discuss below.

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.

5.1 Model and Econometric Specification

We denote consumer loyalty card panelists by $h = 1,...,H$. On a trip during date t, a panelist chooses amongst the $j = 1,...,J$ products in a category or chooses $j = 0$, an outside option (i.e. "no purchase"). 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 $t = 0$. We use $j = PL$ and $j = NB$ to 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: $u^h_{jt} = v_j\left(p_{jt}, s^h_t; \Theta^h\right) + \varepsilon^h_{jt}$, where $v_j\left(p_{jt}, s^h_t; \Theta^h\right)$ represents the panelist's deterministic (conditional on the parameters Θ^h) utility ε^h_{jt} is an i.i.d. Type I Extreme Value distributed random utility term. The indirect utilities are as follows:

$$
u_j\left(p_{jt}, s_t^h; \Theta^h\right) = \begin{cases} \alpha_j^h + \eta^h p_{jt} + \gamma^h \mathbb{I}\{s_t^h = j\} + \varepsilon_{jt}^h, & j \notin \{PL,\ NB\} \\ \alpha_j^h + \beta_j^h \mathbb{I}_{\{h \in T, t \geq \tau\}} + \eta^h p_{jt} + \gamma^h \mathbb{I}\{s_t^h = j\} + \varepsilon_{jt}^h, & j \in \{PL,\ NB\} \\ \varepsilon_{jt}^h, & j = 0 \end{cases} \tag{9}
$$

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:

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

The conditional probability that panelist h chooses alternative j on trip t is

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

To allow for persistent heterogeneity in tastes, we assume panelist tastes, $\Theta^h = \left(\alpha_1^h, \ldots, \alpha_J^h, \beta_{PL}^h, \beta_{NB}^h, \eta^h, \gamma^h\right)$, are drawn from a population distribution $N\left(\bar{\Theta}, \Sigma\right)$ with mean $\bar{\Theta}$ and covariance matrix Σ. We estimate the model using MCMC with a chain consisting of 100,000 posterior draws, dropping the first 10,000 draws as a burn-in period.[11]

We now discuss the key assumptions for the identification of the blind taste test treatment effects, $\{\beta_{PL}, \beta_{NB}\}$. First, we assume that all panelist preferences are drawn from the common population distribution, $F(\Theta)$. The tastes for $\{\alpha_1, ..., \alpha_J, \eta, \gamma\}$ are identified by pooling all the panelists during the pre-test period, $t < 0$. The covariances between the stable taste parameters, $\{\alpha_1, ..., \alpha_J, \eta, \gamma\}$, and the treatment effects, $\{\beta_{SR}, \beta_{MR}, \beta_{LR}\}$, are identified off the test panelists' choices after the blind taste test. We assume that these covariances are common across test and

[11] We implemented the MCMC algorithm using the Bayesm package in R, and refer the reader to Rossi, Allenby, and McCulloch (2005) for a more thorough discussion of the approach.

control panelists and, hence, there is no self-selection on the ability to learn from the blind taste test. This assumption is not critical for estimation. But, it is critical for our counterfactual that requires us to make inferences about the treatment effect on untreated panelists.

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, Hitsch, and Rossi, 2010).

5.2 Structural Estimates

We fit the demand model to the transaction data from the Greek yogurt category. We first compare the fit of four 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 ($\gamma = 0$), and (4) demand with treatment effects and loyalty. For each specification, we compare results with versus without unobserved heterogeneity. 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.[12] 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. The best-fitting specification excludes loyalty but includes the treatment effect of the blind taste test. This finding also suggests that purchase reinforcement is not contributing to the persistent effect of the blind taste test.

–insert Table 13 here–

For posterior inference, we retain every 5th draw from the chain, leaving us with 8,000 posterior

[12]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.

draws. We report the posterior mean and 95% posterior credibility intervals for each of the hyper-parameters in Table 14. 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 14 here–

Consistent with the DID analysis presented previously, the blind taste test has a causal effect on utility. Interestingly, we find an effect on both the brand component of utility for 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 gap between the two brands shrinks by over 80%. From 7 days to 30 days after the test (Medium Run), the expected brand utility for Chobani decreases even more. Interestingly, we cannot rule out zero effect on the expected Roundy's utility. The reduction in the expected gap is now only 38%, relative to the no-treatment case. Finally, between 30 and 150 days after the test (Long Run), we observe a small and negative expected change in utility for both brands. We cannot rule out that the gap is now unchanged relative to the no-treatment case. At least in the short and medium runs, 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.

5.3 The Impact of Information on Demand

Of interest is whether treating the chain's entire consumer population with the blind taste test treatment would fundamentally alter the relative demand of the private label and national brand. 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.

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 demand exceeds the baseline demand. At each point, there is at least a 93% posterior probability that the medium run demand exceeds baseline demand. Recall that our long-run estimate of the treatment effect on Roundy's was found to be small and imprecise. Any tested differences between baseline demand and long-run demand are inconclusive. In sum, we can see that the blind taste test shifts out expected demand for Roundy's in the short run and medium run. Demand appears to shift back close to baseline 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 13.5 percentage points in the short run, and 6.7 percentage points in the medium run. These magnitudes are comparable to the treatment effects found with our DID estimator in section 4.3, even though we now control for prices, brand choice and purchase incidence. In contrast with our DID estimates, we now find a small long-term shift in demand and fail to rule out that it could be zero under 95% posterior credibility.

–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 long-run. At the average posted price of $1.23, the expected short-run demand declines by 14 percentage points, and medium-run demand by 18 percentage points. The expected long-run demand fall by 2.5 percentage points; but we fail to reject it is 0 under 95% posterior credibility.

–insert Figure 6 here–

–insert Figure 7 here–

Thus, we conclude that the direct and the competitive effects of informing consumers on brand shares are statistically and substantively important. These findings survive our controls for prices, brand choice and purchase incidence. Moreover, we find that the blind taste test not only influences perceived utility from the private label, it has a negative influence on the perceived utility of the tested national brand, Chobani. However, as in our DID regressions, the short run effects are much larger than the long run effects.

6 Conclusions

Our findings add to the growing literature studying the implications of consumer misinformation. For the categories studied, the private label alternative typically has a much 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. However, 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 effect on demand for the private label, even five months after the intervention. 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.

In a structural exercise, we estimate a demand system that controls for the causal factors at the point of purchase, such as prices and availability. We find that the blind taste test increases the preferences for the store brand, but decreases the preferences for the competing national brand, both relative to the outside good. Accordingly, the increase in demand for the private label stems from both a direct and an indirect informational effect. We use our estimates to simulate the total effect of assigning the entire day's store population to the blind taste test. We again find the large initial effect, followed by a gradual decay. Our findings are consistent with consumers learning and then forgetting gradually about the quality information from the blind taste test. This depreciation could reflect forgetting (e.g., Mehta, Rajiv, and Srinivasan, 2004) although we do not rule out alternative potential explanations (e.g. on-going marketing by the national brands re-establishes brand capital). Therefore, while our findings suggest we can change consumer demand for private labels, a single usage experience may be insufficient to generate a lasting effect. The information effects reported by us 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, Tellis, and Briesch, 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.

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

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

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

ALLCOTT, H., AND C. KNITTEL (2017): "Are Consumers Poorly-Informed about Fuel Economy? Evidence from Two Experiments," *NBER Working Paper No. 23076*.

ANGRIST, J. D., AND A. B. KRUEGER (1999): "Chapter 23 - Empirical Strategies in Labor Economics," vol. 3, Part A of *Handbook of Labor Economics*, pp. 1277 – 1366. Elsevier.

ASSMUS, G., J. U. FARLEY, AND D. R. LEHMANN (1984): "How Advertising Affects Sales: Meta-Analysis of Econometric Results," *Journal of Marketing Research*, 21(1), 65–74.

ATHEY, S., M. BAYATI, N. DOUDCHENKO, AND G. IMBENS (2017): "Matrix Completion Methods for Causal Panel Data Models," *Stanford GSB Working Paper*.

BAI, J. (2009): "Panel data models with interactive fixed effects," *Econometrica*, 77(4), 1229–1279.

BAIN, J. S. (1956): *Barriers to New Competition*. Cambridge: Harvard University Press.

BAWA, K., AND R. SHOEMAKER (2004): "The Effects of Free Sample Promotions on Incremental Brand Sales," *Marketing Science*, 23(3), 345–363.

BOLLINGER, B., P. LESLIE, AND A. SORENSEN (2011): "Calorie Posting in Chain Restaurants," *The American Economic Journal: Economic Policy*, 3(1), 91–128.

BRONNENBERG, B. J., S. K. DHAR, AND J.-P. DUBÉ (2009): "Brand History, Geography, and the Persistence of Brand Shares," *Journal of Political Economy*, 117, 87–115.

BRONNENBERG, B. J., J.-P. DUBÉ, AND M. GENTZKOW (2012): "The Evolution of Brand Preferences: Evidence from Consumer Migration," *American Economic Review*, 102(6), 2472–2508.

BRONNENBERG, B. J., J.-P. DUBÉ, M. GENTZKOW, AND J. M. SHAPIRO (2015): "Do Pharmacists Buy Bayer? Sophisticated Shoppers and the Brand Premium," *Quarterly Journal of Economics*, 130.

CARRERA, M., AND S. B. VILLAS-BOAS (2015): "Generic aversion and observational learning in the over-the-counter drug market," *Working Paper*.

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

COX, S. R., K. A. CONEY, AND P. F. RUPPE (1983): "The impact of comparative product ingredient information," *Journal of Public Policy & Marketing*, 2, 57–69.

DUBÉ, J.-P., G. J. HITSCH, AND P. MANCHANDA (2005): "An Empirical Model of Advertising Dynamics," *Quantitative Marketing and Economics*, 3, 107–144.

DUBÉ, J.-P., G. J. HITSCH, AND P. E. ROSSI (2010): "State dependence and alternative explanations for consumer inertia," *RAND Journal of Economics*, 41(3), 417–445.

ERDEM, T., AND M. P. KEANE (1996): "Decision-Making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets," *Marketing Science*, 15(1), 1–20.

GEDENK, K., AND S. A. NESLIN (1999): "The Role of Retail Promotion in Determining Future Brand Loyalty: Its Effect on Purchase Event Feedback," *Journal of Retailing*, 75(4), 433–459.

GIVON, M., AND D. HORSKY (1990): "Untangling the Effects of Purchase Reinforcement and Advertising Carryover," *Marketing Science*, 9(2), 171–187.

HEIMAN, A., B. MCWILLIAMS, Z. SHEN, AND D. ZILBERMAN (2001): "Learning and Forgetting: Modeling Optimal Product Sampling Over Time," *Management Science*, 47(4), 532–546.

IRONS, K. W., J. D. C. LITTLE, AND R. L. KLEIN (1983): *Proc. 1983 ORSA/TIMMS Marketing Sci. Conf*chap. Determinants of coupon effectiveness, pp. 157–164.

JIN, G. Z., AND P. LESLIE (2003): "The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards," *The Quarterly Journal of Economics*, 118(2), 409–451.

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

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

MEHTA, N., S. RAJIV, AND K. SRINIVASAN (2004): "Role of Forgetting in Memory-Based Choice Decisions: A Structural Model," *Quantitative Marketing and Economics*, 2, 107–140.

NEWTON, M. A., AND A. E. RAFTERY (1994): "Approximate Bayesian Inference with the Weighted Likelihood Bootstrap," *Journal of the Royal Statistical Society, Series B*, 56(1), 3–48.

ROSSI, P., G. ALLENBY, AND R. MCCULLOCH (2005): *Bayesian Statistics and Marketing*. John Wiley & Sons.

SAHNI, N. S. (2012): "Effect of Temporal Spacing between Advertising Exposures: Evidence from Online Field Experiments," *manuscript*.

SCHMALENSEE, R. (1982): "Product Differentiation Advantages of Pioneering Brands," *The American Economic Review*, 72, 349–365.

SETHURAMAN, R., G. J. TELLIS, AND R. A. BRIESCH (2011): "How Well Does Advertising Work? Generalizations from Meta-Analysis of Brand Advertising Elasticities.," *Journal of Marketing Research.*, 48(3), 457–471.

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

	Cookies				Ice Cream		Greek Yogurt	
Store Number and Location	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
	number of panelists	16240	4266	7529	6793
	number of transactions	208000	12420	49560	146000
control	transactions per household	12.806	2.911	6.582	21.500
panelists	(standard deviation)	(24.906)	(3.693)	(9.506)	(33.290)
	private label share	0.081	0.057	0.184	0.048
	(standard deviation)	(0.199)	(0.204)	(0.292)	(0.154)
	number of panelists	440	99	185	156
	number of transactions	3790	291	1052	2447
test	transactions per household	8.614	2.939	5.686	15.686
panelists	(standard deviation)	(17.107)	(3.063)	(7.695)	(25.913)
	private label share	0.125	0.131	0.226	0.080
	(standard deviation)	(0.254)	(0.286)	(0.312)	(0.205)

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		0.806	0.396	664
Do you think you will prefer PL?	X		0.441	0.497	664
Did you prefer PL?		X	0.715	0.452	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	235	371
		36.66%	63.34%	100.00%
yes		53	240	293
		18.09%	81.91%	100.00%
total		189	475	664
		28.46%	71.54%	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: Updating of beliefs

Valence of updating during taste test		After taste test: Will buy PL next time? no	yes	total
	negative update	17	36	53
	($Q_2 = 1,\ Q_3 = 0$)	32.08%	67.92%	100.00%
	positive neutral	12	228	240
	($Q_2 = 1,\ Q_3 = 1$)	5.00%	95.00%	100.00%
	negative neutral	67	69	136
	($Q_2 = 0,\ Q_3 = 0$)	49.26%	50.74%	100.00%
	positive update	8	227	235
	($Q_2 = 0,\ Q_3 = 1$)	3.40%	96.60%	100.00%
	total	104	560	664
		15.66%	84.34%	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)
hh effects	X	X	X	X
N	107200	6063	25350	75760
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)
hh×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	211800	12710	50610	148500
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-$\beta_{\text{4-8 weeks}}$	0.0527	0.0119	-0.0012	0.0833
	(0.0131)	(0.0366)	(0.0372)	(0.0132)
56-83 days-$\beta_{\text{9-12 weeks}}$	0.0132	0.2901	-0.0188	0.0157
	(0.0136)	(0.0623)	(0.0386)	(0.0133)
84-111 days-$\beta_{\text{12-16 weeks}}$	0.0037	0.0671	0.0358	-0.0128
	(0.0115)	(0.0311)	(0.0346)	(0.0114)
112-157 days-$\beta_{\text{>16 weeks}}$	0.0162	0.0566	0.0336	0.0090
	(0.0121)	(0.0339)	(0.0361)	(0.0119)
hh$\times$cat fixed effects $-\alpha_{hc}$	X	X	X	X
week$\times$cat fixed effects $-\gamma_{tc}$	X	X	X	X
N	211800	12710	50610	148500
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	0.51	0.51	0.28	0.31	0.24	0.22	0.29	0.30
(0-6 days)	(1.1e-03 ,1.00)	(-2e-04 ,0.99)	(0.06 ,0.50)	(0.10 ,0.52)	(0.08 ,0.39)	([0.09 ,0.43)	(0.18 ,0.41)	(0.16 ,0.41)
medium run	0.17	0.17	0.09	0.11	0.06	0.05	0.08	0.08
(7-27 days)	(-0.02 ,0.66)	(-0.02 ,0.54)	(-0.02 ,0.24)	(-0.01 ,0.25)	(-0.02 ,0.15)	(-0.03 ,0.14)	(0.01 ,0.15)	(0.02 ,0.16)
long run	0.08	0.07	5e-04	0.01	0.03	0.03	0.02	0.02
(28-157)	(-0.01 ,0.20)	(-3.3e-03 ,0.18)	(-0.07 ,0.08)	(-0.07 ,0.10)	(-0.02 ,0.09)	(-0.03 ,0.08)	(-0.02 ,0.07)	(-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. $ATT^{OLSfull}$ refers to the estimator used on the sample that includes treated units for whom we do not observe a pre-treatment period.

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 ($\beta^{\text{neg}\to\text{pos}}$)		0.067
		(0.012)
treatment after negative update ($\beta^{\text{pos}\to\text{neg}}$)		-0.051
		(0.028)
treatment after positive confirmation ($\beta^{\text{pos}\to\text{pos}}$)		0.054
		(0.014)
treatment after negative confirmation ($\beta^{\text{neg}\to\text{neg}}$)		0.009
		(0.014)
panelist$\times$category fixed effects–α_{hc}	X	X
week$\times$category fixed effects$-\gamma_{tc}$	X	X
N	211800	211800
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.

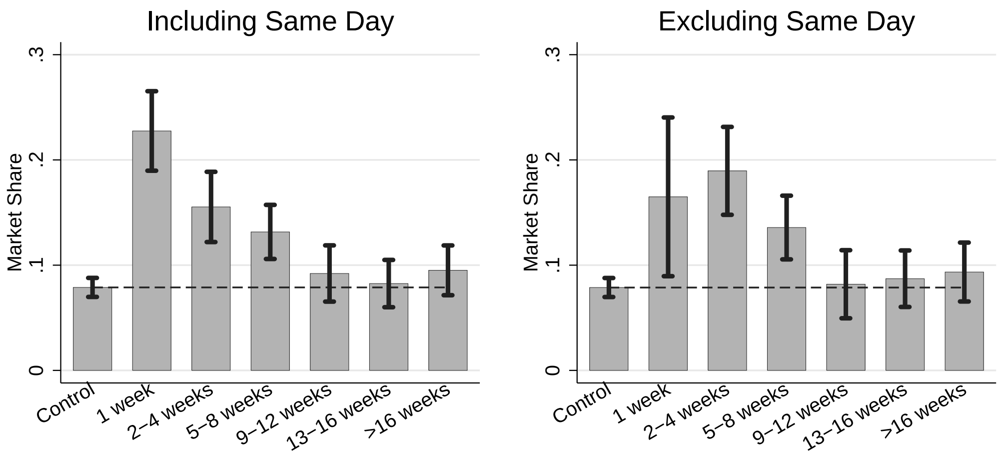
Table 13: Posterior Likelihood (Greek Yogurt)

	posterior log likelihood			
	no loyalty	loyalty	treatment & no loyalty	treatment & loyalty
homogeneous	-284,738.0	-245,380.5	-284,709.8	-245,369.4
random coefficients	-176,640.8	-177,023.8	-176,530.8	-176,826.1

Table 14: Hyper-parameter estimates (Greek Yogurt)

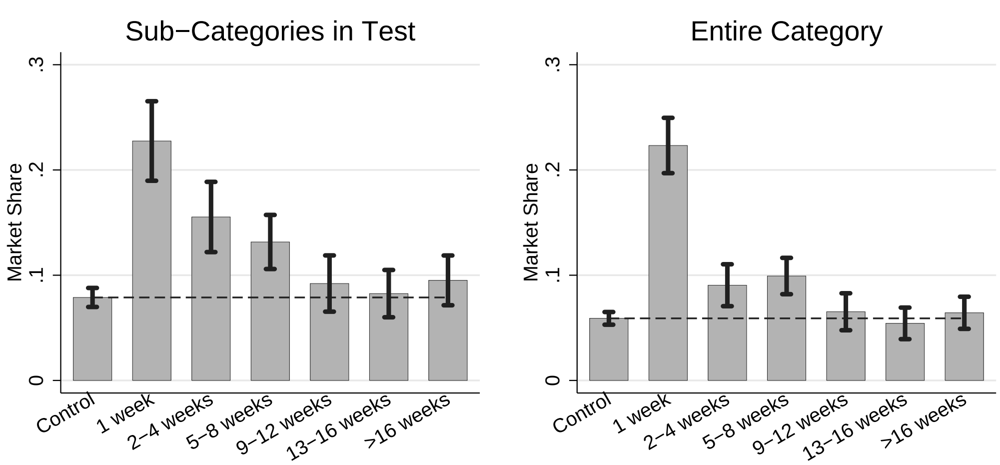
Coefficient	Population mean			Population st. dev.		
	50th	2.5th	97.5th	50th	2.5th	97.5th
Chobani $(\alpha_{Chobani})$	-2.18	-2.42	-1.93	4.92	4.74	5.11
Dannon Oikos (α_{Dannon})	-3.93	-4.22	-3.66	5.73	5.54	5.93
Fage (α_{Fage})	-3.75	-4.03	-3.49	5.06	4.87	5.26
Roundy's $(\alpha_{Roundy's})$	-6.00	-6.29	-5.71	5.20	5.01	5.40
Noosa (α_{Noosa})	-4.57	-4.97	-4.16	7.32	7.03	7.61
Yoplait Greek $(\alpha_{Yoplait})$	-4.74	-5.00	-4.48	5.17	4.98	5.36
price (η)	-2.86	-3.06	-2.66	3.77	3.62	3.92
(SR treat)*Chobani $\left(\beta^{SR}_{Chobani}\right)$	-0.94	-1.73	-0.44	1.77	1.15	2.36
(SR treat)*Roundy's $\left(\beta^{SR}_{Roundy's}\right)$	2.13	1.58	2.56	1.99	0.96	2.64
(MR treat)*Chobani $\left(\beta^{MR}_{Chobani}\right)$	-1.12	-1.67	-0.56	1.45	0.95	2.01
(MR treat)*Roundy's $\left(\beta^{MR}_{Roundy's}\right)$	0.31	-0.11	0.73	1.49	1.17	1.78
(LR treat)*Chobani $\left(\beta^{LR}_{Chobani}\right)$	-0.38	-0.92	0.03	1.41	1.00	1.77
(LR treat)*Roundy's $\left(\beta^{LR}_{Roundy's}\right)$	-0.78	-1.39	-0.17	1.76	1.20	2.36

Figure 1: Robustness – Panelists not Buying on the Treatment Day



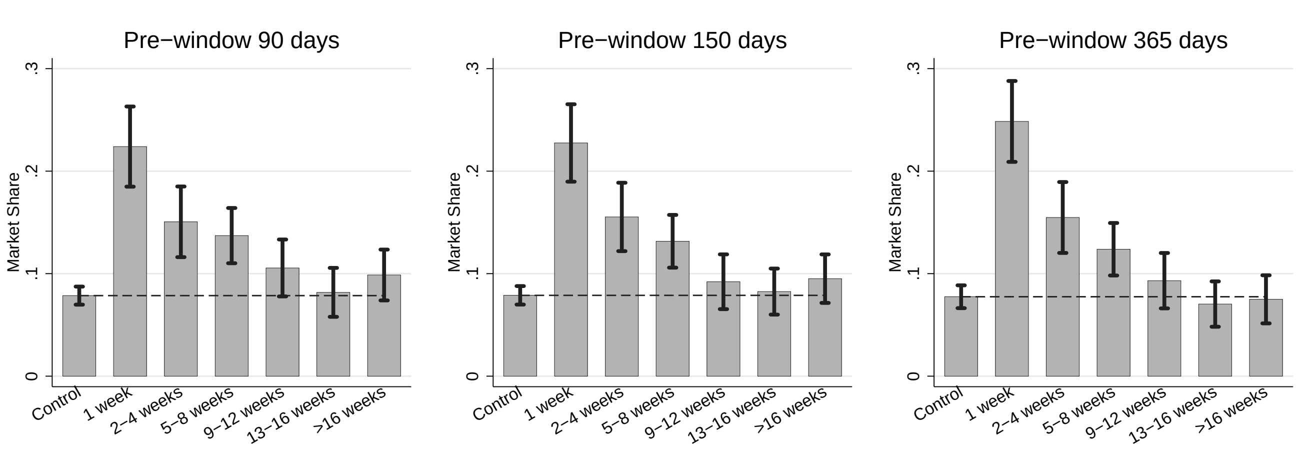
Note: The graph on the left reproduces the first column in table (10). The graph on the right represents the case where the difference-in-differences regression is estimated on the sub-sample of treated panelists who do not buy from the category on the treatment-day.

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

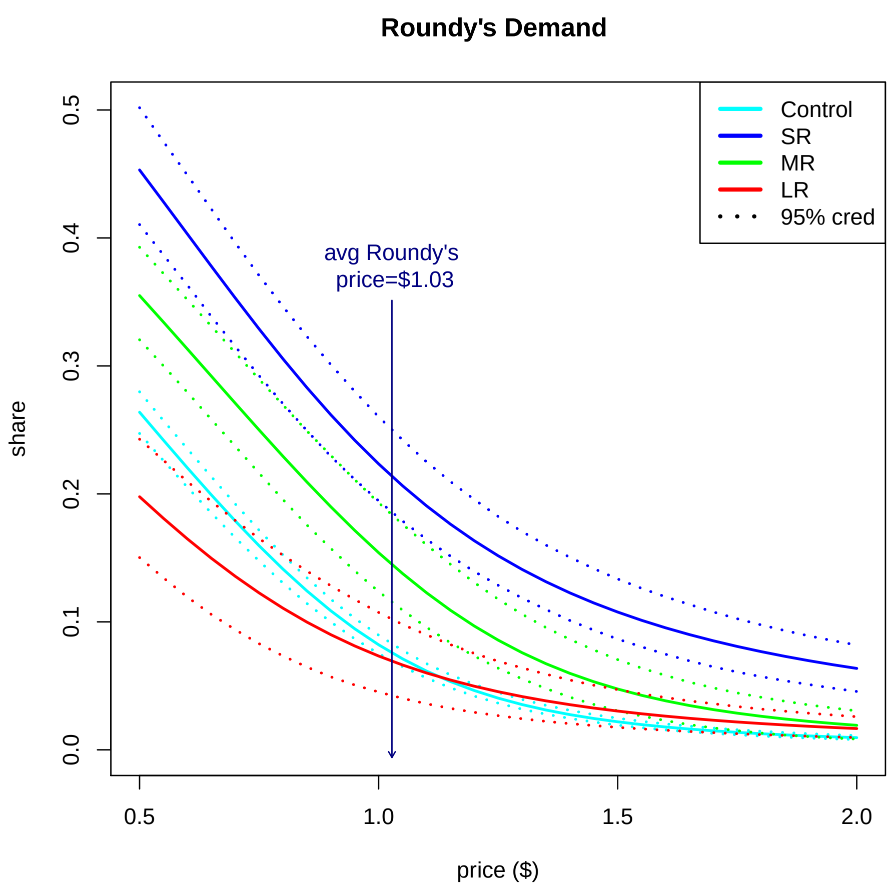


Figure 5: Posterior Expected Demand Shifts for Roundy's

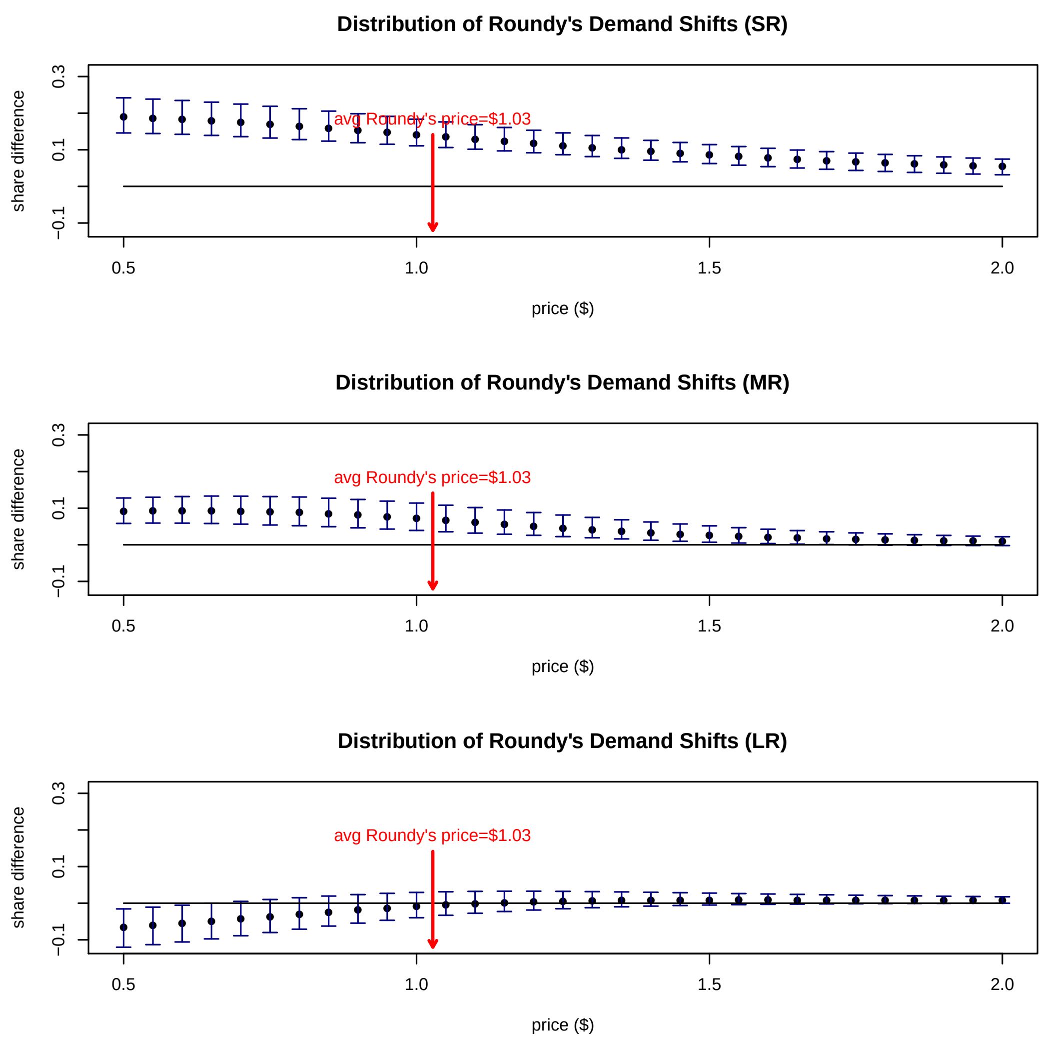


Figure 6: Posterior Expected Demand for Chobani

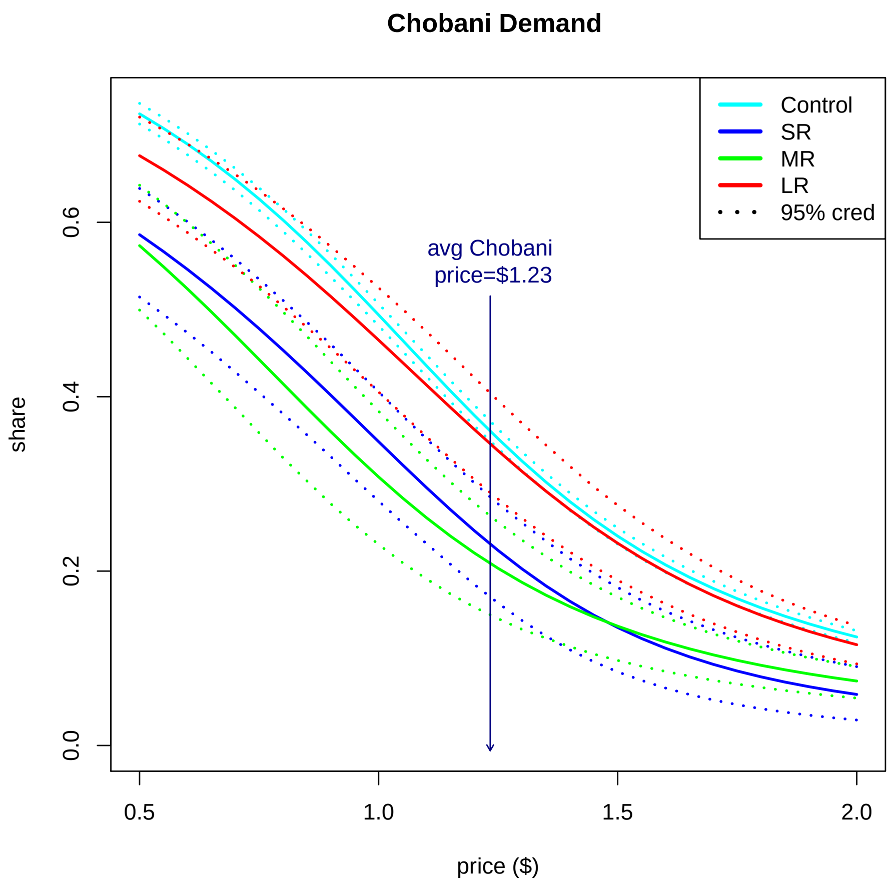
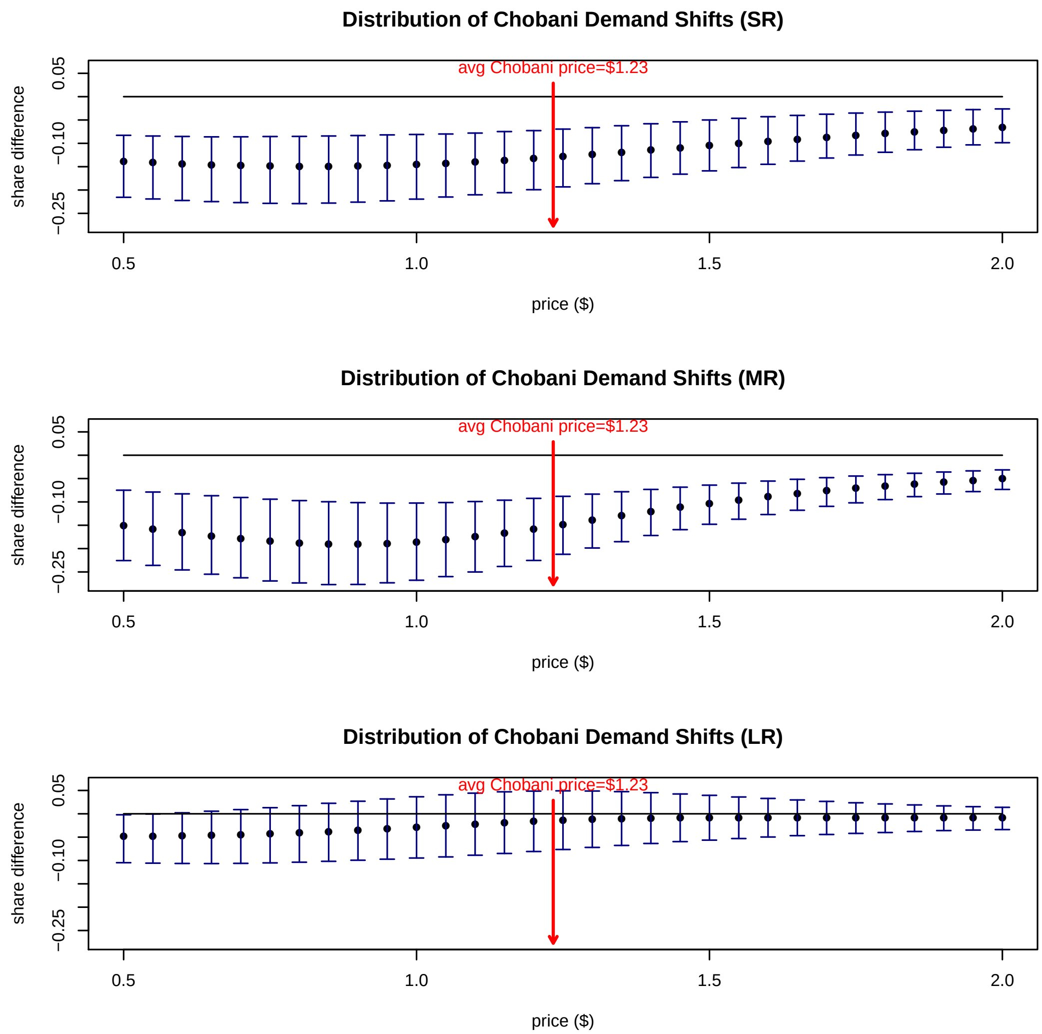


Figure 7: Posterior Expected Demand Shifts for Chobani



A Difference Regressions within Treatment Group

Table 15: 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)
hh$\times$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 16: 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)
hh$\times$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 Construction of Structural Estimation Sample

B.1 Misc. Corrections to the Data.

B.1.1 Incorrect Pack Sizes and Labeling:

Some of the originally-coded SKUs were incorrectly or imprecisely labeled.

- In the Roundy's file containing SKU information, some SKUs incorrectly list the equivalent units (EU) per pack, rather than the total EU associated with purchasing that particular SKU.
- This occurs for multi-packs in particular. To correct the item size, we search for inconsistency between the pack size as described in the item description and the total observed servings. For example, for yogurt, we search for items with substrings indicating a pack of four (e.g. "4P", "4PK", "4 PK") but also contain fewer than 7 ounces.

- We also manually check to make sure pack sizes are not mislabeled. In the raw data, 'CHOBANI GRK YGT RSPBRRY W CHOC', 'CHOBANI GRK YGT FIG BITE', 'CHOBANI GRK YGT PINEAPPLE BITE','DANNON OIKOS TRAD STRAWBERRY', 'YOPLAIT GRK BLUEBERRY GRAN PARFAIT are mislabeled and 1/4th their actual size. To correct these pack sizes, we multiply the EU of these units by 4.
- Similarly, 6 pack and 8 pack items were corrected. The products 'CHOBANI STRAWBERRY TUBES', 'CHOBANI BLUEBERRY TUBES', 'CHOBANI CHERRY TUBES', 'CHOBANI STRW BANA TUBES' were actually 8 packs although their substrings did not contain this obviously, and they were manually corrected.
- We also ensure SKUs are grouped together and properly classified as "greek yogurt." For example, "DANNON LIGHT & FIT" (a greek yogurt) was missed as part of the Dannon Oikos greek yogurt brand yet contains no substrings indicating it is in this class, so we manually collapse it into "DANNON OIKOS."

B.2 Cleaning Data:

- To determine the market leaders, we use purchase data from entire set of all Mariano's yogurt consumers.
- The top 11 packsize-brand combinations has median weekly market share of 76.5%. We drop the remaining packsize-brand combinations. This results in 6 unique brands.
- We focus only on single servings of these brands, and we capture 67% of the greek yogurt market share in EU.
- To construct SKU prices, we straight average of UPC prices.
- We retain only yogurt treatment and yogurt control customers, as created in the diff-in-diff estimation.
- We drop trips involving purchases of multiple brands.
- We run the analysis on a 150 (difference-in-difference) and 365-day (structural) pre-window to the end of our data, March 21, 2015.

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. This format conforms with the set-up used in the descriptive DID analysis in section 4. The posterior likelihood estimates in Table 17 confirm our earlier findings. Unobserved heterogeneity in tastes improves fit substantially, as does the conditioning on the treatment-related variables. In addition, we select the model without loyalty relative to the model controlling for loyalty based on posterior fit.

Similarly, our results regarding preferences in Table 18 also confirm our earlier findings. Once again, we find that the short and medium run treatment effect of the blind taste test on consumers'

Table 17: Posterior Likelihood (Greek Yogurt)

	posterior log likelihood			
	no loyalty	loyalty	treatment & no loyalty	treatment & loyalty
homogeneous	-195941.97	-167603.24	-195907.31	-167590.39
random coefficient	-117027.22	-117743.42	-116996.50	-117683.79

Table 18: Hyper-parameter estimates (Greek Yogurt)

Coefficient	Population mean			Population st. deviation		
	50th	2.5th	97.5th	50th	2.5th	97.5th
Chobani	-2.54	-2.94	-2.16	5.32	5.08	5.55
Dannon Oikos	-4.80	-5.27	-4.35	5.96	5.71	6.22
Fage	-4.28	-4.75	-3.88	5.39	5.13	5.66
Roundys	-7.02	-7.44	-6.64	5.61	5.37	5.85
Noosa	-5.18	-5.77	-4.62	7.63	7.26	8.02
Yoplait Greek	-5.16	-5.52	-4.78	5.47	5.26	5.70
price	-2.65	-2.95	-2.32	3.94	3.75	4.12
(SR treat)*Chobani	0.24	0.02	0.52	1.38	1.07	1.62
(SR treat)*Roundys	2.37	2.05	2.60	1.56	0.82	2.37
(MR treat)*Chobani	-1.15	-1.36	-1.00	1.02	0.82	1.27
(MR treat)*Roundys	1.09	0.82	1.46	*1.58*	*1.34*	*1.78*
(LR treat)*Chobani	0.06	-0.13	0.32	1.29	1.04	1.53
(LR treat)*Roundys	0.10	-0.22	0.34	1.33	1.08	1.66

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.