Online Appendix for: Swap and Save: How a Button Shapes Lasting Online Grocery Shopping Behavior

Rubén Gonzálvez
Complutense University of Madrid, rubenmgo@ucm.es

Xiaolin Li
The Hong Kong Polytechnic University, xiao-lin.li@polyu.edu.hk

Sarit Weisburd
The Hebrew University, sarit.weisburd@mail.huji.ac.il

Xinrong Zhu
Imperial College London, xzhu2@ic.ac.uk

Appendix

A. Examples of Swaps on the Platform

Figure A1 A Shopping Journey on the Platform

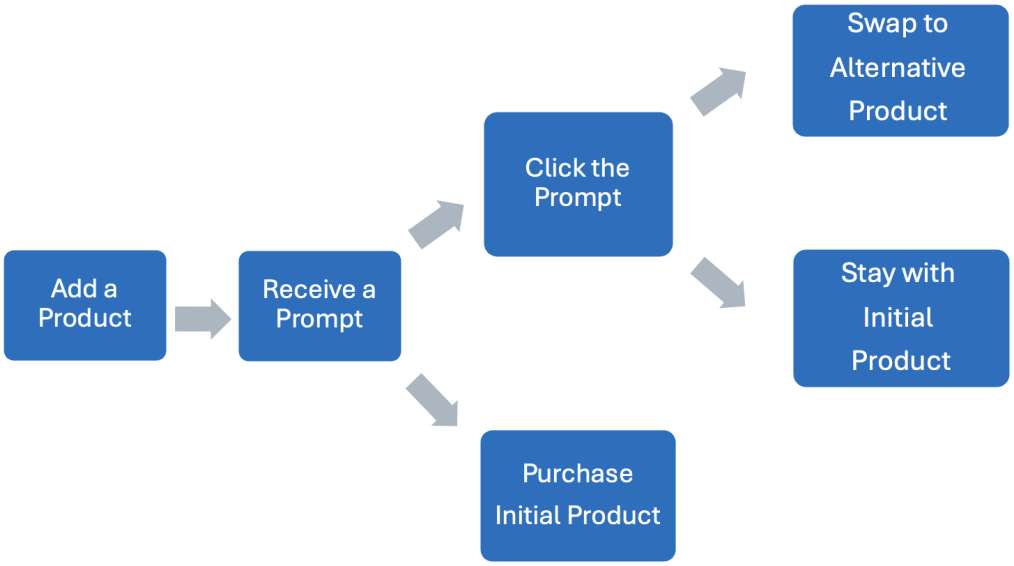
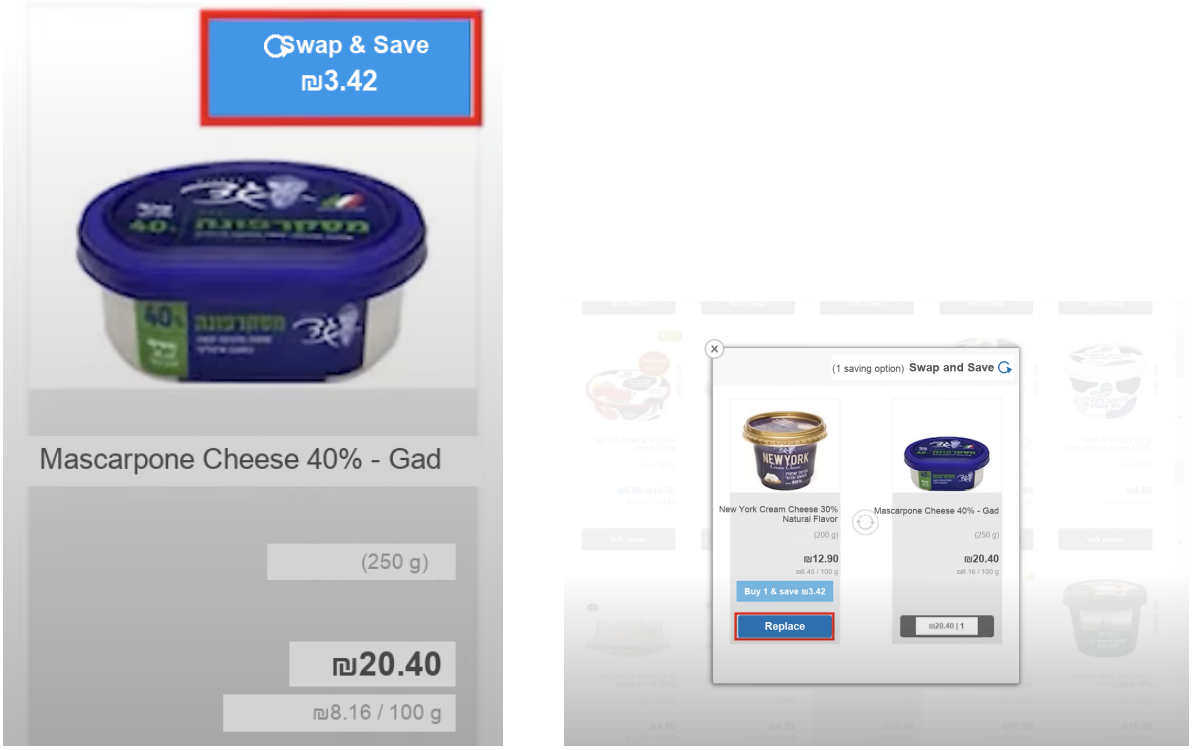


Figure A2 Example of a Swap & Save Button and Prompt (English Translation)



(a) Swap & Save Button (b) Prompt of Cheaper Alternatives

Notes: Panel (a) displays an example of the *swap & save* button, and Panel (b) shows the prompt of a cheaper alternative if the shopper clicks the button. The platform interface is in Hebrew; we translated it to English in this figure.

B. Sample Construction Details

This appendix provides step-by-step details on the construction of the swap, IV, and cross-category analysis samples.

B1. Raw Datasets

Analysis databases are created from two raw datasets from an Israeli price comparison platform that aggregates real-time information from consumers using the platform to shop at five online supermarkets:

- **Swap & Save database:** Records all interactions with product substitution prompts (2,365,840 observations; 19,449 shoppers).
- **Purchase database:** Item-level purchases during 2015 (6,178,044 observations; 19,965 shoppers).

B2. Swap Sample Construction

We construct the swap sample used in the CS DiD model following these steps.

- **Step 1: Sample restriction.** We merge the swap and purchase databases on basket and product identifiers, then restrict the sample to the mineral water category. This yields 71,433 purchases from 11,337 shoppers across 38 products.
- **Step 2: Swap events.** We identify swap events and record the initial product ($J1$) and alternative product ($J2$).
- **Step 3: Control group.** We define control shoppers as shoppers who were plausibly *not exposed* to the swap button. They must have at least one observation and satisfy the criteria introduced in Subsection 1.3, which yields 70,335 observations from 11,127 shoppers, including 1,420 treated and 5,699 control shoppers.
- **Step 4: Matching.** Control observations are matched to treated swap events for the same initial product and supermarket within a $\pm$30-day window using nearest-date matching. Unmatched observations are dropped. The matched sample contains 36,915 observations from 3,182 shoppers (1,420 treated; 1,762 control).
- **Step 5: Panel restrictions.** We drop singleton observations and shoppers with no pre-period purchases, leaving 36,551 observations from 3,105 shoppers (1,343 treated; 1,762 control).
- **Step 6: Outcomes and event windows.** We retain one observation per trip and define $\text{BuyAlternative}_{irt} = 1$ if the shopper purchases the alternative product, and 0 otherwise. We construct three event-study windows across the same 3,105 shoppers: $[-5, +5]$ (19,987 observations), $[-10, +10]$ (25,588 observations), and $[-15, +15]$ (27,923 observations).

B3. IV Sample Construction

The IV sample expands the CS DiD sample to include shoppers who were *exposed* to the swap button but did not click or swap. This enables instrumental variable estimation using button receipt as an instrument for actual swapping behavior.

- **Step 7:** Starting from the same cleaned purchase data and including only shoppers who were dropped from the swap sample, we define two additional levels of treatment exposure:

1. Stayers: shoppers who clicked the swap button to view alternatives but chose to keep their initial product.

2. Exposed non-clickers: shoppers who purchased a product for which a swap event existed at the same supermarket on the same date, meaning the button was likely displayed even if they did not interact with it. The control group is defined with the same criteria as in the swap sample.

- **Step 8: Matching and variable definitions.**

— Control purchases are matched to exposed shoppers' event dates using the same ±30-day window and closest-date rule as in the swap sample.

— For stayers and exposed non-clickers, we assign a counterfactual alternative product ($J2$). Since these shoppers did not actually swap, we cannot observe which product they would have been offered. We assign the most commonly swapped-to product for their initial product at the same supermarket and date. If no swap occurred at that supermarket-date, we use the most common swapped-to product for that initial product across all dates at the same supermarket.

— The final database after dropping singletons and shoppers not observed in the pre-period includes 57,336 observations from 5,984 shoppers:

∗ Exposed shoppers (swappers + stayers + exposed non-clickers): 4,079 shoppers, including 1,343 treated shoppers from the swap sample and an additional 2,736 stayer + exposed non-clicker shoppers.

∗ Unexposed shoppers: 1,905 shoppers, including 1,762 control shoppers from the swap sample and an additional 143 control shoppers matched to stayers and non-clicker shoppers.

Three analysis datasets are constructed by varying the event window: $[-5, +5]$ (34,409 observations), $[-10, +10]$ (42,191 observations), and $[-15, +15]$ (45,298 observations), each covering the same 5,984 shoppers.

B4. Cross-Category Sample Construction

The cross-category analysis extends the swap pipeline to 55 product categories beyond mineral water. The construction is identical to the swap sample (Steps 1–6) and is replicated independently for each category.

- **Step 1:** Using the full purchase database, we identify all product categories with at least 40,000 total purchases during the sample period. This threshold yields 55 categories, ranging from "Candies and chewing gum" (the smallest, with 4,965 final observations) to "Vegetables" (the largest, with 65,344 final observations).
- **Step 2:** For each of the 55 categories, we independently execute the same six steps described in constructing the swap sample. Each category produces its own final analysis sample, containing one observation per shopper-trip in the $[-5, +5]$ window.

C. Sample Details

Table C1 Summary Statistics of Prices and Promotions

	Mean	SD	P50	P25	P75
Price per liter ($)	0.553	0.481	0.297	0.234	0.918
Promotion Frequency (Percentage)	8.224	10.979	5.424	1.930	9.408
Promotion Duration (Days)	9.933	10.833	4.690	2	17.793
Promotion Depth ($)	-0.335	0.348	-0.221	-0.559	-0.060
Promotion Depth (Percentage)	-21.871	9.079	-23.412	-25.060	-15.353

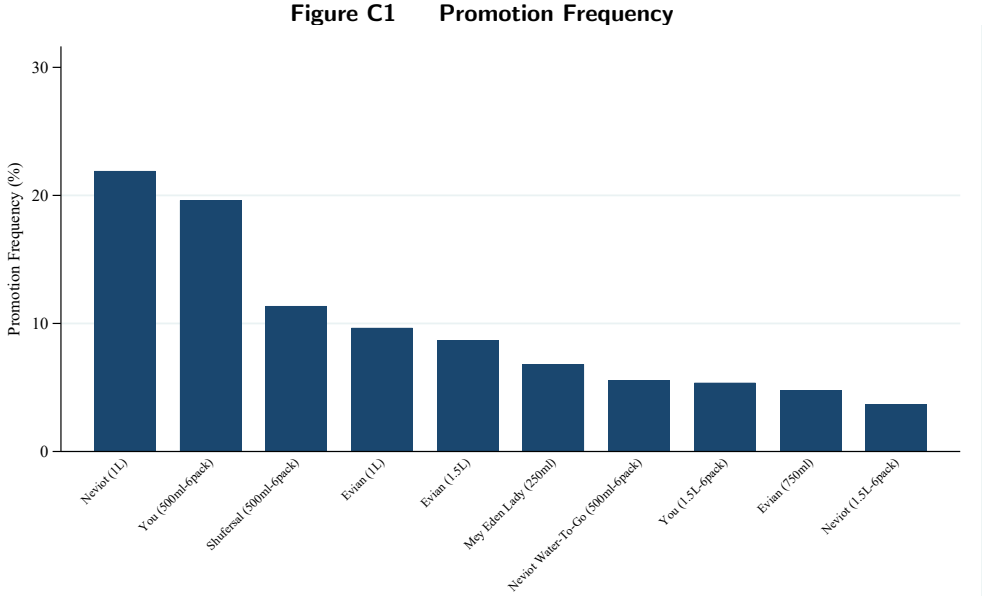
Notes: Price per liter is summarized at the product-supermarket-date level, and the other variables are summarized at the product-supermarket-promotion level. Promotion-related variables are conditional on the occurrence of a promotion.

Table C2 Summary Statistics of Main Products

Product	Brand	Quantity Share	Promotion Frequency	Price/L ($)	Size
Ein Gedi (6 pack)	Ein Gedi	37.11%	1.77%	0.23	2 liter
Neviot (6 pack)	Neviot	20.72%	3.70%	0.28	1.5 liter
Aqua Nova (6 pack)	Aqua Nova	16.01%	0.17%	0.20	2 liter
Mey Eden (6 pack)	Mey Eden	13.27%	0.80%	0.31	1.5 liter
Mey Eden Sport (6 pack)	Mey Eden	5.73%	1.20%	1.03	500ml
Mey Eden Sport (6 pack)	Mey Eden	1.93%	1.85%	0.98	750 ml
San Benedetto (6 pack)	Sun Bendo	1.44%	3.35%	0.46	1.5 liter
Neviot Water-To-Go (6 pack)	Neviot	1.02%	5.55%	1.24	500ml
Shufersal (6 pack)	Shufersal	0.83%	0.00%	0.30	1.5 liter
Neviot Water-To-Go (6 pack)	Neviot	0.74%	0.00%	1.05	755ml
You (6 pack)	Yo	0.27%	5.37%	0.30	1.5 liter
Mey Eden Baby (6 pack)	Mey Eden	0.25%	0.73%	2.61	250ml
San Benedetto (6 pack)	Sun Bendo	0.23%	0.00%	1.11	500ml
Shufersal (6 pack)	Shufersal	0.17%	11.36%	1.01	500ml
Neviot Baby (6 pack)	Neviot	0.14%	2.04%	2.00	330ml
Neviot	Neviot	0.08%	21.91%	0.78	1 liter
You (6 pack)	Yo	0.02%	19.61%	0.98	500ml
Spark (12 pack)	Spark	0.01%	0.00%	0.58	500ml
You	Yo	0.01%	0.00%	0.29	1.5 liter
Opal S.Y International (12 pack)	Opal	0.01%	0.00%	0.45	500ml

Notes: This table presents summary statistics on the top 20 products based on quantity share ranking. Promotion frequency is calculated as the number of days when a product is on sale at a given supermarket divided by the total number of days it is sold in the sample period.

Figure C1 Promotion Frequency



Notes: This figure shows promotion frequency for 10 products with the highest promotion frequency across all supermarkets in our sample. Promotion frequency is calculated as the number of days a product is on sale at a supermarket divided by the total number of days sold in the sample period. Averages across supermarkets are plotted.

D. Additional Descriptive Evidence

Table D1 Product Pairs of Top 10 Most Common Swaps and their Prices at the Swap Events

Initial Product	Initial Price	Alternative Product	Alternative Price	Price Difference	# of Swaps
Ein Gedi 2 liter (6 pack)	0.227	Aqua Nova 2 liter (6 pack)	0.192	-0.035	471
Mey Eden 1.5 liter (6 pack)	0.305	Neviot 1.5 liter (6 pack)	0.244	-0.061	398
Mey Eden 1.5 liter (6 pack)	0.310	Ein Gedi 2 liter (6 pack)	0.221	-0.088	207
Neviot 1.5 liter (6 pack)	0.266	Aqua Nova 2 liter (6 pack)	0.190	-0.076	195
Neviot 1.5 liter (6 pack)	0.277	Ein Gedi 2 liter (6 pack)	0.218	-0.059	187
Neviot Water-To-Go 500 ml (6 pack)	1.281	Mey Eden Sport 500 ml (6 pack)	0.994	-0.287	164
Ein Gedi 2 liter (6 pack)	0.235	Neviot 1.5 liter (6 pack)	0.225	-0.009	113
Mey Eden Sport 500 ml (6 pack)	1.088	Mey Eden Sport 750 ml (6 pack)	0.887	-0.201	110
Mey Eden 1.5 liter (6 pack)	0.307	Aqua Nova 2 liter (6 pack)	0.196	-0.111	92
Mey Eden Sport 500 ml (6 pack)	1.158	Neviot Water-To-Go 500 ml (6 pack)	0.917	-0.241	87

Notes: This table presents the top 10 most common product swaps across supermarkets, ranked by number of swaps. Price is measured as price per liter in dollars. Both the initial price and alternative price are prices at the swap events. "# of swaps" represents the total number of swaps between the initial and the alternative products.

Figure D1 Price Differences Between Swap Product Pairs: Supermarket 1

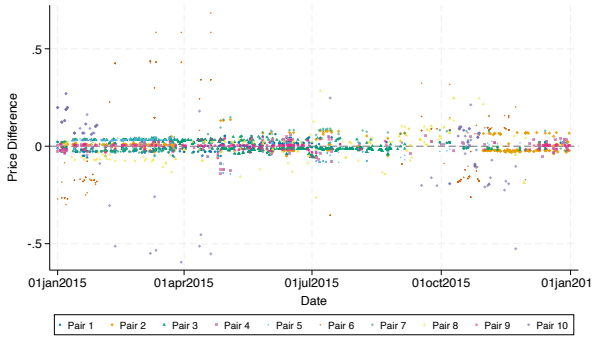


Figure D2 Price Differences Between Swap Product Pairs: Supermarket 2

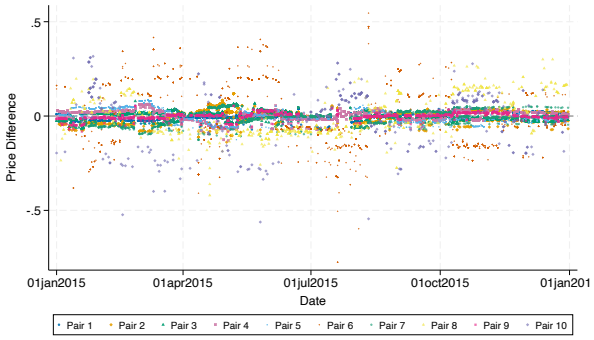


Figure D3 Price Differences Between Swap Product Pairs: Supermarket 3

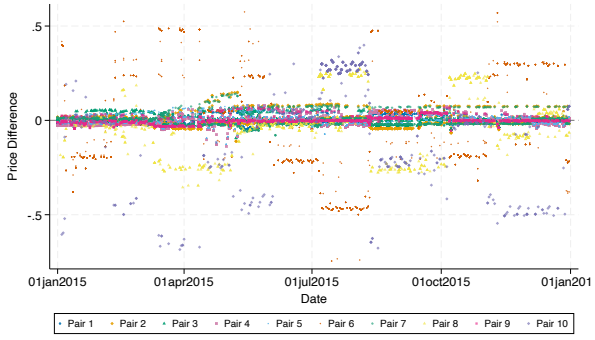


Figure D4 Price Differences Between Swap Product Pairs: Supermarket 4

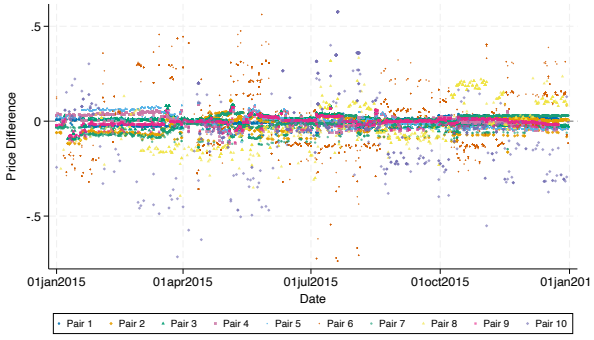
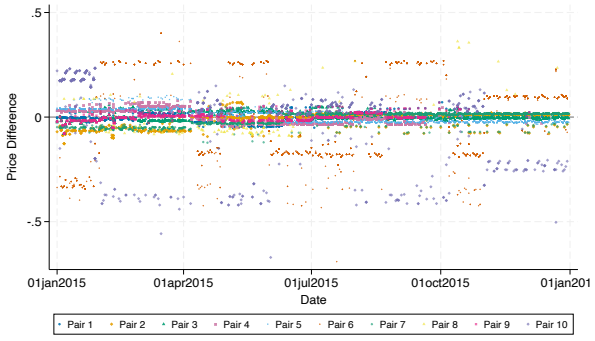


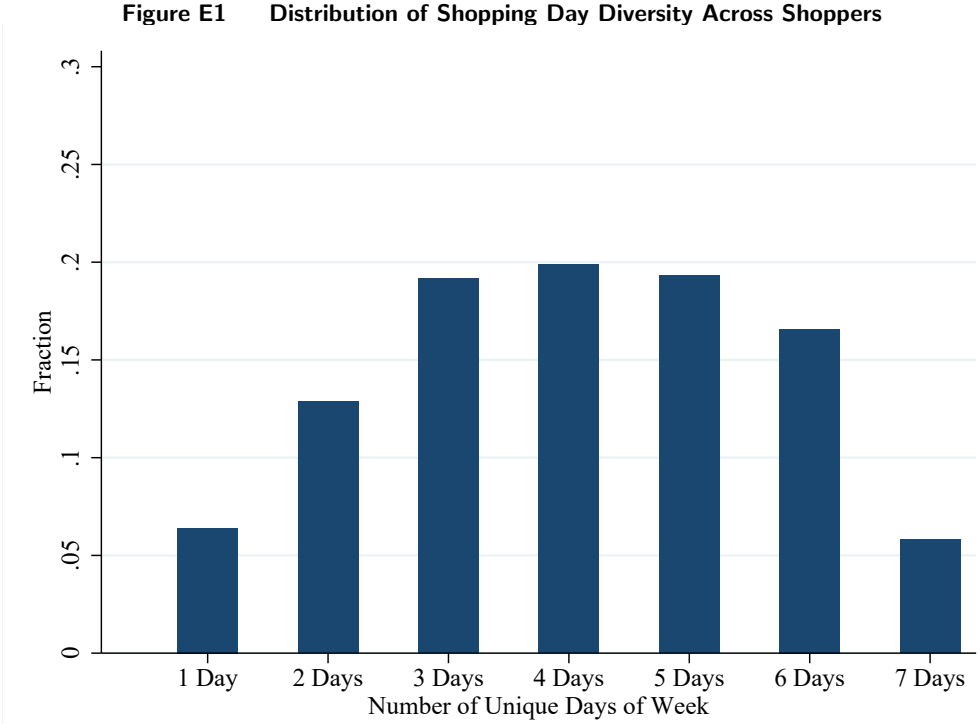
Figure D5 Price Differences Between Swap Product Pairs: Supermarket 5



Notes: Figures D1–D5 plot the price difference (alternative minus initial, in dollars per liter) for the ten most commonly swapped product pairs across calendar dates at each supermarket. Price differences fluctuate randomly around zero throughout the year, with no predictable seasonal pattern that shoppers could exploit through strategic purchase timing.

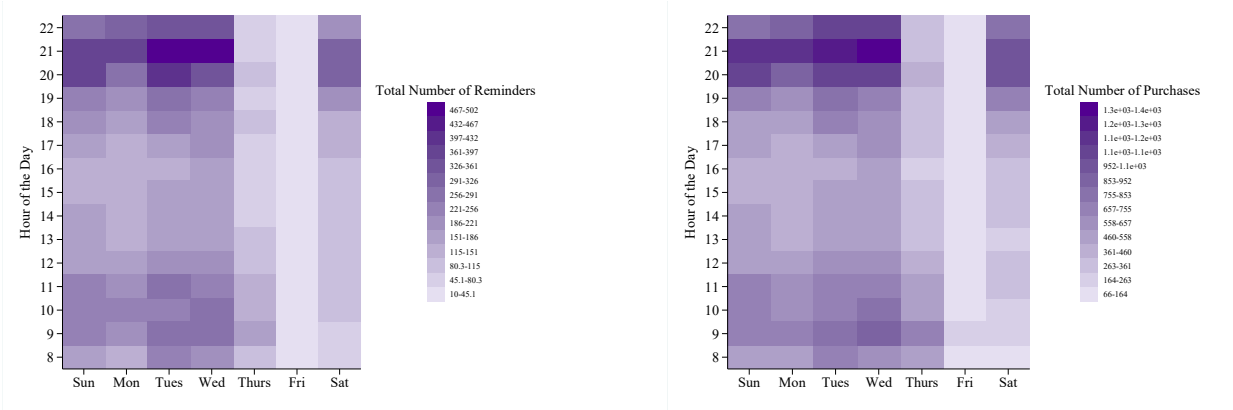
E. Shopper Purchasing Patterns

Figure E1 Distribution of Shopping Day Diversity Across Shoppers



Notes: The figure shows the distribution of the number of unique days of the week on which shoppers make purchases.

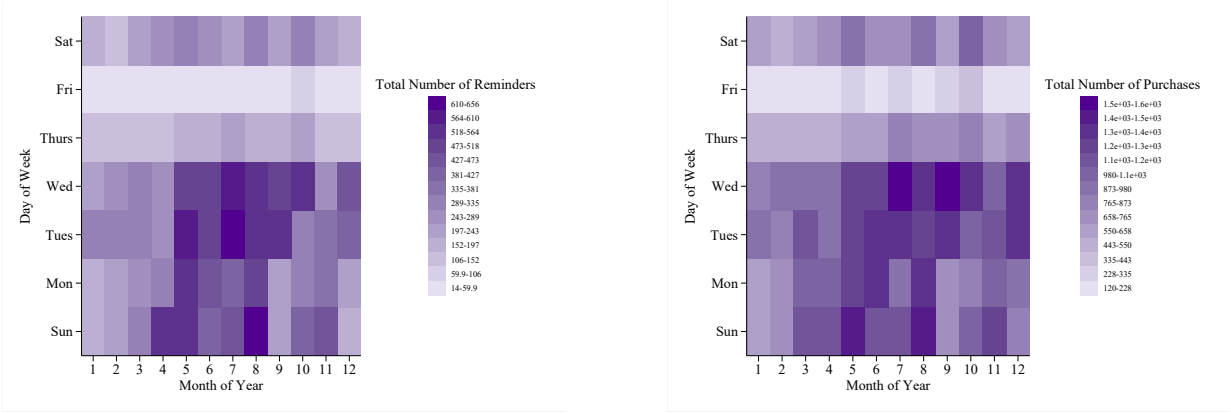
Figure E2 Number of Prompts and Purchases across Days and Hours



(a) Total Number of Prompts (b) Total Number of Purchases

Notes: These heatmaps show the total number of prompts (panel (a)) and the total number of purchases (panel (b)) by hour of the day and day of the week. Darker colors indicate higher counts.

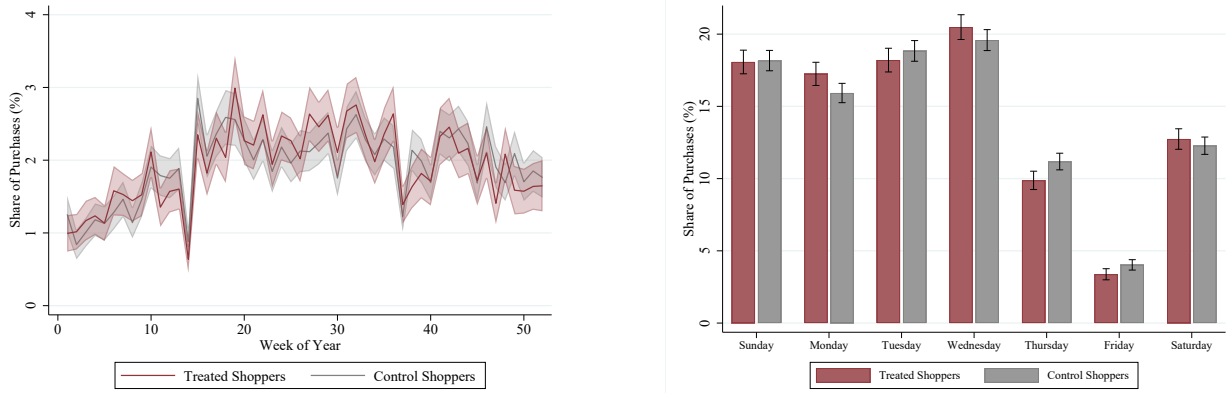
Figure E3 Number of Prompts and Purchases across Days and Months



(a) Total Number of Prompts (b) Total Number of Purchases

Notes: The heatmaps show the total number of prompts (panel (a)) and the total number of purchases (panel (b)) by day of the week and month of the year. Darker colors indicate higher counts.

Figure E4 Comparable Purchasing Patterns between Treated and Control Shoppers



(a) Share of Purchases by Weeks of the Year and Shopper Groups

(b) Share of Purchases by Day and Shopper Groups

Notes: The share of purchases by week is calculated as the number of purchases made by each shopper in each week, relative to their respective total purchases of the year. The share of purchases by day is calculated as the number of purchases made by each shopper on each day of the week, relative to their respective total purchases of the entire week. These calculations use all purchases made during the sample period by both treated and control shoppers. Shares are reported rather than counts because the treated and control groups differ in size, and shares facilitate comparison of seasonal and day-of-week patterns across groups.

F. IV Analysis

F1. IV Construction and Identifying Assumptions

Sample Construction. Shoppers who received the swap & save button (*exposed shoppers*) include: (i) shoppers who clicked the "swap & save" button to view cheaper alternatives but ultimately retained their initial product. (ii) Shoppers who purchased the same initial product on the same day and at the same supermarket as those who clicked the "swap & save" button, but who did not click the button themselves (and therefore do not appear in the *swap database*). For these shoppers, the alternative product ($J2$) is defined as the one most frequently chosen by shoppers who received a "swap & save" button after initially adding product ($J1$) at supermarket (r) on swap event day (e).[1] Among these 4,079 *exposed shoppers*, 66% clicked on the "swap & save" button, and 33% of them swapped to an alternative water product.

For each purchase exposed to the swap & save button ("exposed purchase"), we identify matching purchases by 1,905 unexposed shoppers who (i) shopped at the same supermarket within a 60-day window (30 days before and after the exposed purchase), (ii) did not receive the button, and (iii) purchased identical initial products. Then, we incorporate these exposed and unexposed shoppers into our main sample (for CS DiD) alongside the existing treated and control groups. Finally, we construct a normalized "purchase time" variable across all shoppers, centered on the event time and set to 0 at the first prompt or swap occurrence. For exposed shoppers, their event time is the first time they receive the button. For unexposed shoppers, we assign event timing based on their matched exposed counterpart.[2]

Identifying Assumptions. The causal effect can be identified if our IV meets the following conditions: (i) the receipt of the "swap & save" button is correlated with product swaps (relevance); (ii) the receipt of the button is essentially randomly assigned (independence); and (iii) the receipt of the button affects the outcome solely through its first-stage impact on swaps (exclusion restriction). We discuss how our IV satisfies these conditions below.

(i) Relevance: this is ensured by the institutional fact that receiving the "swap & save" button is necessary for a product swap. In Appendix F2, we also show the first-stage results and F statistic of the IV from the first-stage regression to verify this point.

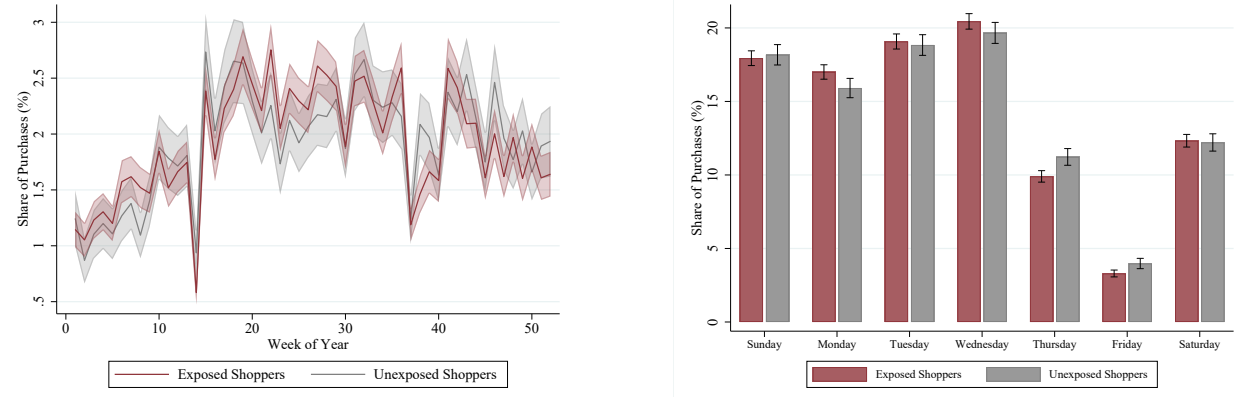
(ii) Independence: In IV DiD models, the independence assumption requires exposure to the "swap & save" button to be quasi-random. We show in Figure 1 that prices of the initial and alternative products fluctuate randomly and frequently across days of the week. Given this high degree of price variability, the "swap & save" buttons are as good as random to each individual shopper. Additionally, Figure F1 indicates that the shopping patterns (in terms of days and weeks of shopping) are also very similar between the exposed and unexposed shopper groups. Lastly, in Figure F2, we plot the coefficients from a TWFE intention-to-treat (ITT) event study of the probability of purchasing the alternative product. The figure shows that prior to exposure to the "swap & save" button, differences in demand between exposed and unexposed

[1] The average purchase rate of the most frequently chosen $J2$ conditional on $J1$ is 94.54%.

[2] When shoppers made multiple purchases on the same day at the same supermarket, we consolidate these into a single trip given the minimal time gap. We set the outcome variable to 1 if any alternative product purchase occurred that day.

shoppers remained stable over time, consistent with parallel pre-trends. This pattern indicates the absence of differential demand dynamics prior to exposure and suggests that any differences between groups are time-invariant. Furthermore, any residual unobserved heterogeneity at the shopper, retailer, shopping trip, or day-of-week level is absorbed by the IV DiD model's fixed effects.

Figure F1 Comparable Purchasing Patterns between Exposed and Unexposed Shoppers

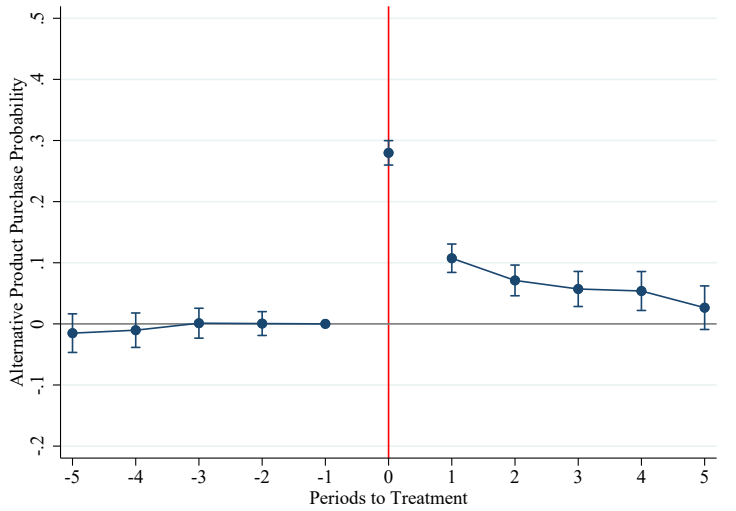


(a) Share of Purchases by Weeks of the Year and Shopper Groups

(b) Share of Purchases by Day and Shopper Groups

Notes: The share of purchases by week is calculated as the number of purchases made by each shopper in each week, relative to their respective total purchases of the year. The share of purchases by day is calculated as the number of purchases made by each shopper on each day of the week, relative to their respective total purchases of the entire week. These calculations use all purchases made during the sample period by both exposed and unexposed shoppers. Shares are reported rather than counts because the exposed and unexposed groups differ in size, and shares facilitate comparison of seasonal and day-of-week patterns across groups.

Figure F2 Effect of Prompt on Demand for Alternative Product

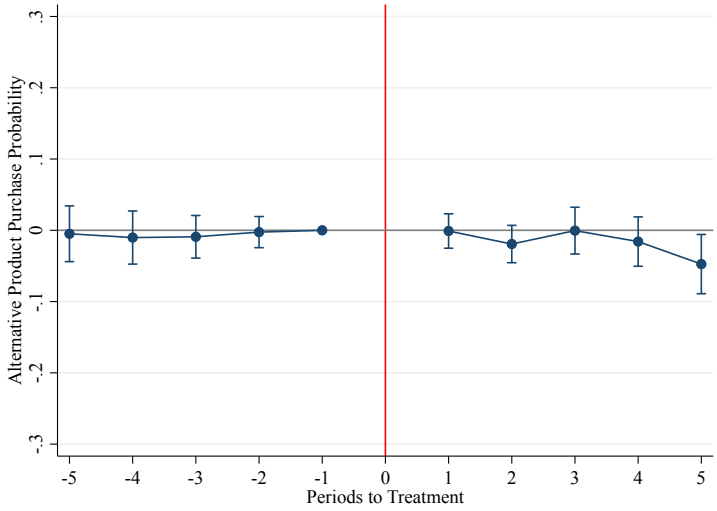


Notes: The figure plots δ_t and their 95% confidence intervals from a TWFE intention-to-treat (ITT) event study of the probability of purchasing the alternative product on indicators for periods relative to prompt receipt, controlling for retailer, day-of-week, shopping trip, and shopper fixed effects.

(iii) Exclusion Restriction: The instrument should ideally affect the outcome only through its effect on the treatment (a product swap). A potential concern is that exposure to the "swap & save" button may generate a persistent advertising or attention effect on the alternative product, even among shoppers who do not swap.

To examine this, we compare exposed non-swappers to matched control shoppers. If exposure directly increased future purchases of the alternative product, we might expect divergence in post-period trends between these groups, even though neither group swapped. We estimate a TWFE DiD event study with exposure as the treatment and the alternative-product purchase indicator as the outcome. The average post-exposure coefficient is not statistically significant ($\beta = -0.0103$, $p > 0.1$), and the event-study coefficients in Figure F3 show no systematic increase in purchases of the alternative product among exposed non-swappers. This suggests that exposure, absent any swapping, does not generate a persistent advertising or top-of-mind effect on demand for the alternative product.

Figure F3 Purchase Probability of Alternative Product: Exposed Non-Swappers vs Control Shoppers



Notes: The figure plots δ_t and their 95% confidence intervals estimated from Equation 2 when including only exposed non-swappers (shoppers exposed to the "swap & save" button who did not swap products) and unexposed control shoppers, with prompt receipt as the treatment. The vertical line denotes the time of exposure.

F2. IV Estimation Results

Table F1 reports the first-stage and second-stage results from estimating Equations 4 and 5. The first-stage estimate $\hat{\beta}_1$ implies that 36.9% of shoppers who received a "swap & save" button switched products (see Panel A). The instrument passes the Kleibergen-Paap weak instrument test, with an F-statistic of 889.6, which indicates that the instrument is highly correlated with the endogenous variable ($\text{PostSwap}_{i,t}$). The second-stage estimate ($\hat{\pi}_1$) indicates that responding to the button to switch products increased the probability of purchasing the alternative product in future shopping trips by 21.4 percentage points.

We also estimate an alternative IV specification following Lewbel (2012), which constructs internal instruments based on heteroskedasticity rather than relying solely on the prompt as an external instrument.[3] In this specification, the first stage includes both $PostPrompt_{irt}$ and the Lewbel-type generated instrument, and the second stage uses the fitted $PostSwap_{irt}$ as the regressor of interest. The Breusch-Pagan test rejects homoskedasticity in the first-stage residuals ($F = 82.24$, $p < 0.001$), supporting the relevance of the Lewbel instrument. The results, reported in Column (2) of Table F1, are consistent with and very close in magnitude to our baseline IV DiD estimates.

The Lewbel specification also allows us to assess whether the prompt has a direct effect on future purchases of the alternative by including $PostPrompt_{irt}$ in the second stage regression, rather than as an excluded instrument. Identification in this specification relies solely on the Lewbel IV. The estimated coefficient on $PostPrompt_{irt}$ is small and negative (-0.036, SE $= 0.013$). Thus, although the primary exclusion concern is

[3]This method generates instruments by interacting demeaned exogenous variables with first-stage residuals, replacing traditional exclusion restrictions with assumptions about the covariance structure of error terms.

that exposure to the prompt may directly increase future purchases of the alternative, we find no evidence supporting a positive direct effect.

Table F1 IV DiD First and Second Stages

	(1) IV DiD	(2) Lewbel IV
Panel A: First Stage (DV: PostSwap_{irt})		
PostPrompt_{irt}	0.3688***	0.3535***
	(0.0124)	(0.0120)
Lewbel IV_{irt}		1.9710***
		(0.1301)
Panel B: Second-Stage (DV: $\text{BuyAlternative}_{irt}$)		
$\widehat{\text{PostSwap}}_{irt}$	0.2144***	0.2647***
	(0.0271)	(0.0253)
F-Stat (Kleibergen-Paap)	889.640	482.87
F-Stat (Breusch-Pagan)		82.24
Number of Observations: 27,682		

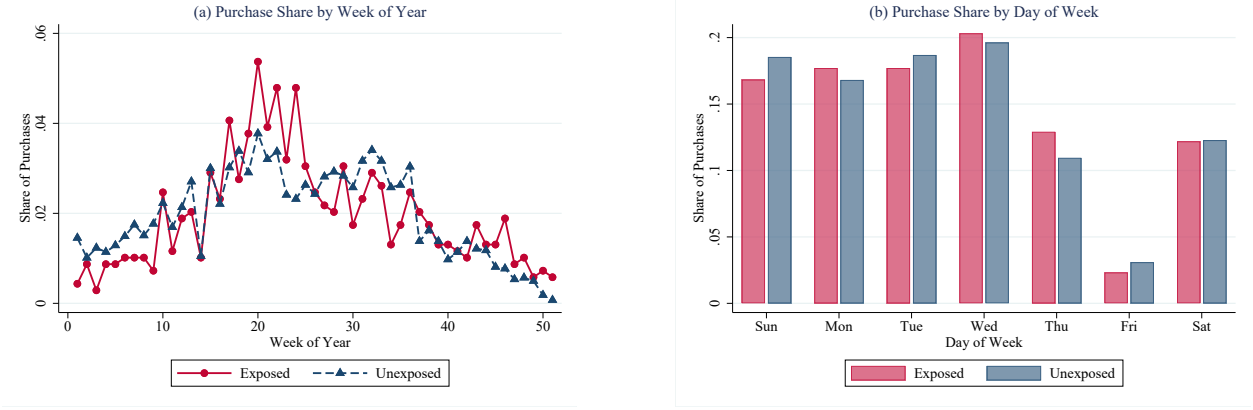
Notes: Each column represents a different IV estimation method. To avoid artificially inflating the treatment effect from swaps, the regressions exclude the treatment event. Standard errors, clustered at the shopper level, are shown in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

F3. Alternative IV: Stock-outs

We also developed an alternative instrument based on product stock-outs of the shopper's regularly purchased item, which provides a compelling source of exogenous variation in swap decisions. In this setting, the "swap & save" button suggests in-stock alternatives that are similar to the initially selected product, regardless of price.

Why stock-outs are a good quasi-random source of variation. Stock-outs of regularly purchased products are driven by supply chain disruptions, warehouse logistics, and replenishment timing — factors that are plausibly orthogonal to any individual consumer's price sensitivity or product preferences. Consistent with this, exposed shoppers (shopping when their regularly purchased product is out-of-stock) and unexposed shoppers (shopping for this same product when it is in-stock at the same supermarket within a 60 day window) exhibit very similar shopping patterns. To show this, Figure F4 plots the share of purchases occurring in each week of the year and each day of the week for both groups. The distributions of purchases are highly comparable between exposed and unexposed shoppers. Next, we also estimate a TWFE intention-to-treat (ITT) event study of the probability of purchasing the alternative product and plot the coefficients in Figure F5. The figure shows that prior to exposure to the stock-outs, differences in demand between exposed and unexposed shoppers remained stable over time, consistent with parallel pre-trends. This supports the assumption that stock-out exposure is quasi-random: absent the stock-out, exposed shoppers would likely have followed similar demand paths as unexposed shoppers. Moreover, temporary stock-outs affect future demand for the alternative product only by inducing trial of a new product, rather than through persistent price or informational effects.

Figure F4 Shopping Patterns of Exposed and Unexposed Shoppers (Stock-Out Sample)

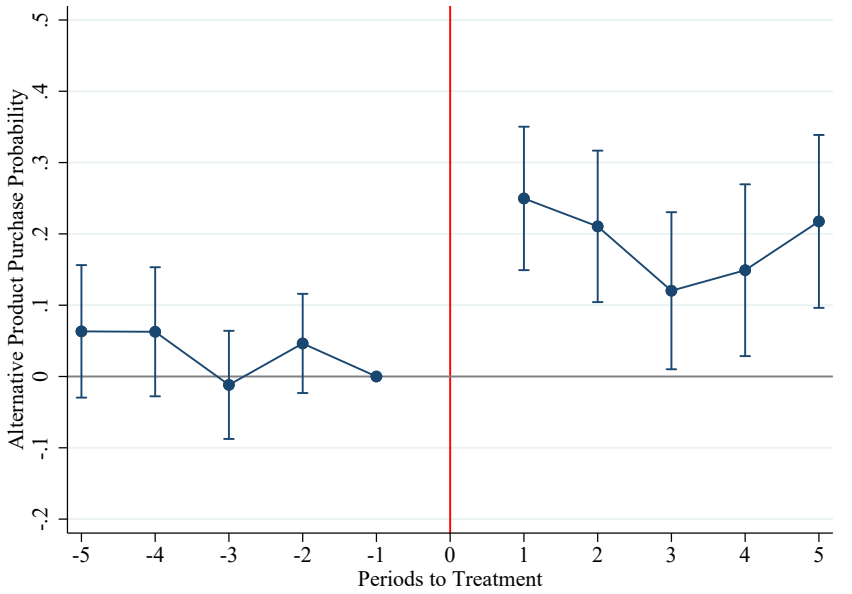


(a) Share of Purchases by Weeks of the Year and Shopper Groups

(b) Share of Purchases by Day and Shopper Groups

Notes: The share of purchases by week is calculated as the number of purchases made by each shopper in each week, relative to their respective total purchases of the year. The share of purchases by day is calculated as the number of purchases made by each shopper on each day of the week, relative to their respective total purchases of the entire week. Both panels use only pre-treatment observations (five trips before the stock-out event).

Figure F5 Effect of a Stock-Out Event on Demand for Alternative Product



Notes: The figure plots δ_t and their 95% confidence intervals estimated from the following equation: $\text{BuyAlternative}_{irt} = \alpha_0 + \sum_{\substack{t=-5 \\ t\neq 0}}^{5} \delta_t \, \text{PostStockout}_{irt} + Z_i + S_r + K_{d(t)} + D_t + \epsilon_{irt}$. The vertical line denotes the stock-out event time.

Sample Construction. To implement this analysis, we redefine event-time as the date on which a shopper's regularly purchased water product is out of stock. *Exposed shoppers* are those who arrived at the platform and found their regular product out-of-stock.[4] *Unexposed shoppers* are those who purchased the same regular product at the same supermarket—but while it was in-stock—within a 60-day window around the exposed shopper's visit.

Specifically, exposed shoppers include: (i) shoppers who clicked the "swap & save" button to view suggested in-stock alternatives and chose one of these alternatives. (ii) shoppers who clicked the "swap & save" button to view suggested in-stock alternatives but ultimately chose a different product. (iii) shoppers who selected an alternative without clicking the prompt. For exposed shoppers who did not actually swap, the "alternative product" is assigned as the product most frequently chosen by other shoppers facing the same out-of-stock initial product at the same supermarket.

For each purchase exposed to an out-of-stock event ("exposed purchase"), we identify matching purchases by unexposed shoppers who (i) shopped at the same supermarket within a 60-day window (30 days before and after the exposed purchase), (ii) had a history of purchasing the same out-of-stock product, and (iii) the product was not out-of-stock on the day they shopped. We construct a normalized "purchase time" variable across all shoppers, centered on the event time and set to 0 at the first out-of-stock occurrence. For exposed shoppers, their event time is the first time they encountered an out-of-stock event. For unexposed shoppers, we assign event timing based on their matched exposed purchase.

Estimation Results. We estimate the following two-stage model. Let $F_{it} = \text{PostSwap}_{irt}$, then the first stage predicts this interaction term as a function of a similar interaction term formed by the instrument, $\text{PostStockout}_{irt}$, which equals one if shopper i in supermarket r encountered an out-of-stock event by shopping trip t.

$$F_{irt} = \alpha_0 + \beta_1 \,\text{PostStockout}_{irt} + Z_i + S_r + K_{d(t)} + D_t + \nu_{irt} \tag{1}$$

$$\text{BuyAlternative}_{irt} = \pi_0 + \pi_1 F_{irt} + Z_i + S_r + K_{d(t)} + D_t + \mu_{irt} \tag{2}$$

Table F2 reports the first-stage and second-stage results from estimating Equations 1 and 2. The first-stage estimate $\hat{\beta}_1$ suggests that 53.5% of shoppers swap to the most popular alternative product when their regularly purchased product is out-of-stock (see Panel A). The instrument passes the Kleibergen-Paap weak instrument test, with an F-statistic of 135.6, which indicates that the instrument is highly correlated with the endogenous variable (PostSwap_{irt}). The second-stage estimate ($\hat{\pi}_1$) indicates that switching to the alternative product raised the probability of purchasing it in future shopping trips by 31.8 percentage points.

[4] 9.5% of the shoppers were exposed to stock-outs, and 51% of them swapped to an alternative product.

Table F2 Stock-out IV DiD First and Second Stages

Panel A: First Stage (DV: PostSwap_{irt})	
$\text{PostStockout}_{irt}$	0.5352***
	(0.0460)
Panel B: Second-Stage (DV: $\text{BuyAlternative}_{irt}$)	
$\widehat{\text{PostSwap}}_{irt}$	0.3183***
	(0.0701)
F-Stat (Kleibergen-Paap)	135.644
Number of Observations: 9,745	

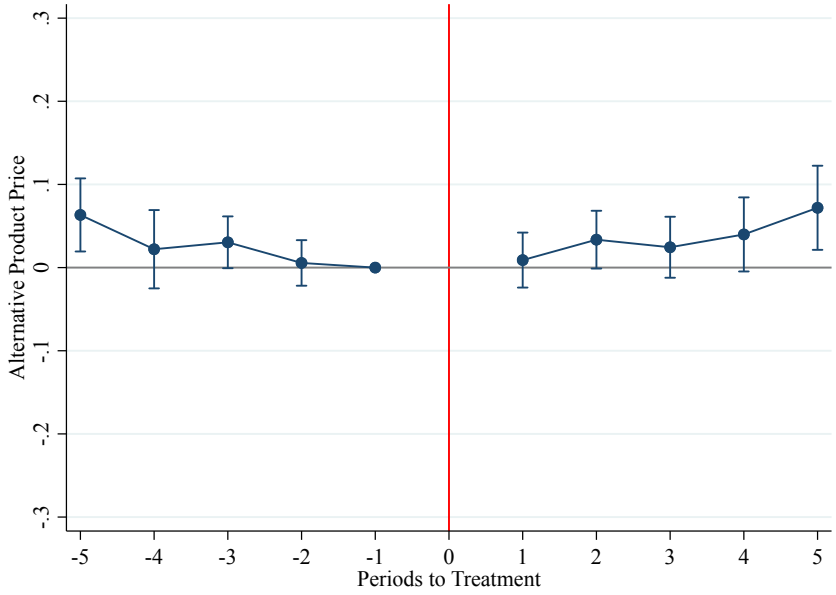
G. Additional Robustness and Placebo Tests

Table G1 Robustness Tests

	(1) Not-yet-treated as control	(2) Less frequent shoppers	(3) Treated cohort defined by month	(4) Additional fixed effects
PostSwap_{irt}	0.2706***	0.2379***	0.2081***	0.2149***
	(0.0323)	(0.0232)	(0.0241)	(0.0167)
Number of Observations	4,813	11,160	10,248	16,598

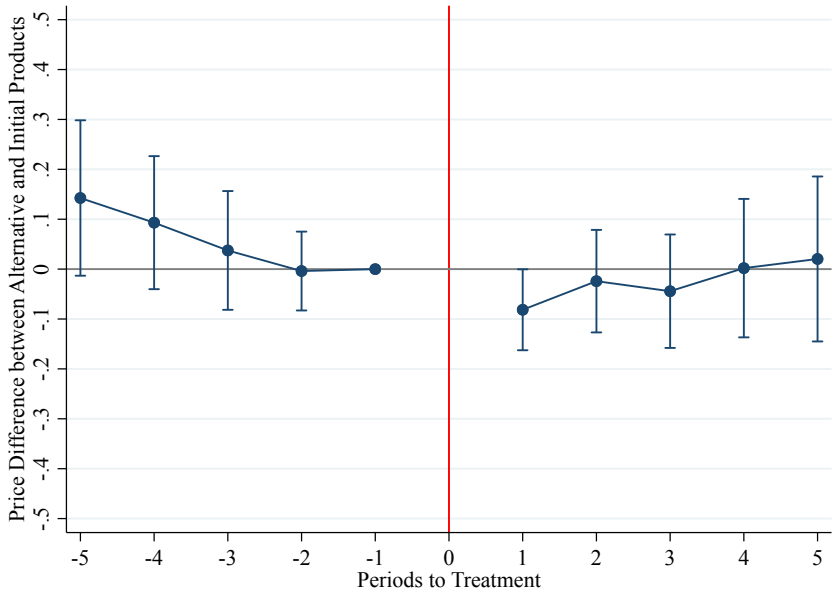
Notes: Each column represents a robustness check using the CS DiD method. Column (1) restricts the control group to not-yet-treated shoppers. Column (2) uses a subsample of below-mean shopping frequency shoppers. Column (3) applies a calendar-month cohort definition, collapsing the sample to monthly observations and coding $\text{BuyAlternative}_{irt} = 1$ if the alternative was purchased at least once that month. Column (4) augments the CS DiD model with retailer × day fixed effects. Standard errors, clustered at the shopper level, are shown in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Figure G1 Price Evolution of Alternative Product – TWFE DiD



Notes: The figure plots δ_t and their 95% confidence intervals estimated from a TWFE DiD event-study of the alternative product's price, controlling for retailer, day-of-week, shopping trip, and shopper fixed effects.

Figure G2 Price Difference of Alternative Product versus Initial Product



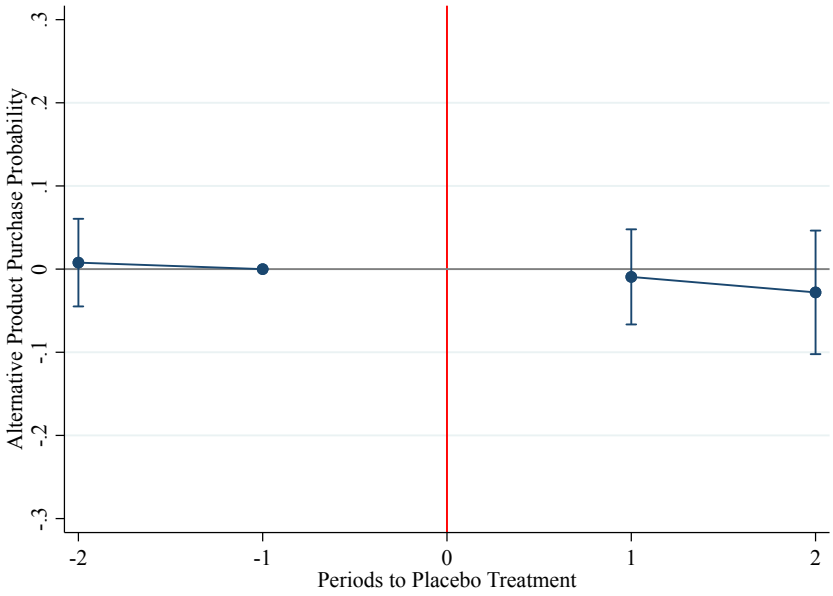
Notes: The figure plots δ_t and their 95% confidence intervals estimated from Equation 2. The outcome variable is the price difference between the alternative product and the initial product. The vertical line denotes the swap event time.

Table G2 Robustness to Control Group Matching Window

	(1)	(2)	(3)	(4)
	±2 days	±3 days	±5 days	±15 days
$\text{PostSwap}_{i,t}$	0.2305***	0.2332***	0.2388***	0.2390***
	(0.0206)	(0.0193)	(0.0188)	(0.0183)
Number of Observations	13,008	14,080	15,122	16,209

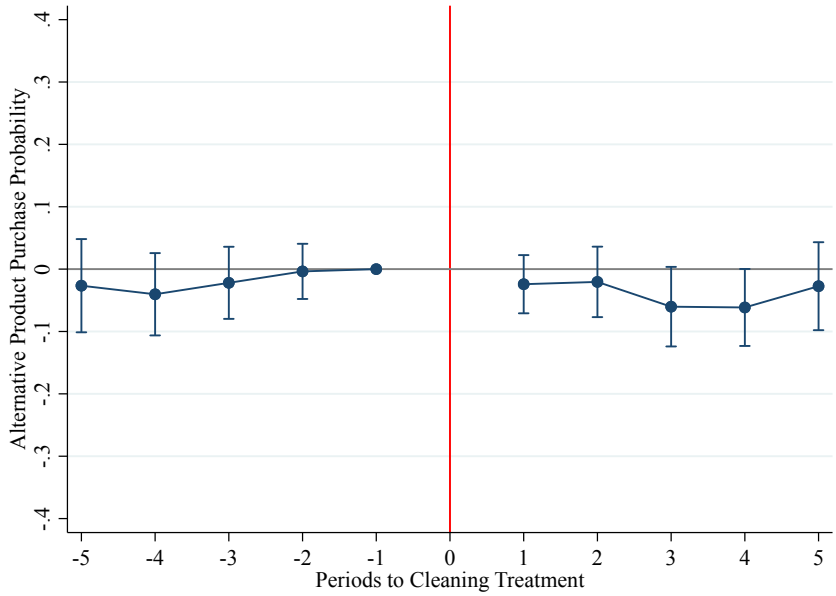
Notes: Each column re-estimates the CS DiD model using a different control group matching window. Controls are shoppers who purchased the same initial product at the same retailer within the specified number of days of the treated shopper's swap event. Standard errors, clustered at the shopper level, are shown in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Figure G3 Placebo Test – Treatment Time Defined at Pre-period



Notes: This figure presents estimates from a test for anticipation effects, examining whether soon-to-be-treated shoppers adjust their purchasing behavior prior to the swap. Specifically, the figure plots the coefficients δ_t and their 95% confidence intervals from Equation 2, estimated using a CS DiD specification with a placebo event time at $t-3$ (three trips before the actual swap), using only pre-swap observations. The vertical line denotes the placebo event time, corresponding to three trips before the swap.

Figure G4 Placebo Test – Treatment Defined by Cleaning Product Swap



Notes: This figure presents estimates from a placebo test, examining whether shoppers adjust their purchasing probability of the alternative water product, following a cleaning product swap. Specifically, the figure plots the coefficients δ_t and their 95% confidence intervals from Equation 2, estimated using a CS DiD specification with a placebo event time of the cleaning product swap (denoted by the vertical line).

References

Lewbel A (2012) Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business & Economic Statistics* 30(1):67–80.