A ONLINE APPENDIX

A.1 Analysis of the Online Channel Using Weekly Aggregation

For this analysis, we use a more granular level of aggregation that is at the week level. In total, we use 68 weeks for the weekly aggregation: 34 weeks before and 34 weeks after. Note that we drop 26 DMAs from the analysis due to missing observations. Of the 184 remaining DMAs, we assign 125 in the treatment group and 59 in the control group for Model-1. In Model-2, we match the experimental units through propensity score matching and include 59 DMAs in the treatment and 59 DMAs in the control group. Observing the results in Model-1 and Model-2, we find that our findings remain consistent.

Table 14: DID Analysis of the Online Channel Using Weekly Aggregation

Variables	**ONLINE PURCHASES Model-1**		**ONLINE PURCHASES Model-2**	
TREAT*AFTER	-0.03^{**}	(0.011)	-0.04^{***}	(0.012)
TIME-VARYING CONTROLS	YES		YES	
YEAR-WEEK FE	YES		YES	
DMA FE	YES		YES	
PROP. SCORE MATCHING	NO		YES	
N	12,512		8,024	
R^2	0.60		0.63	

Note: Standard errors (in parentheses) are heteroskedasticity robust and clustered at the DMA level. FE denotes fixed effects.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We also verify that the treatment and control groups follow parallel trends in the pre-BOPS period. The coefficient for TREAT $*$ TREND is not statistically significant in Model-1 ($p > 0.1$, $R^2 = 0.57$) nor in Model-2 ($p > 0.1$, $R^2 = 0.60$).

A.2 Analysis of the BM Channel Using Weekly Aggregation

Just as we did in §A.1, we use weekly aggregation for the BM channel. In total, we use 68 weeks: 34 weeks before and 34 weeks after. Note that we drop 1 DS-A store from the analysis due to missing observations. Of the 277 DS-A stores, we assign 152 in the treatment group and 125 in the control group for Model-1. In Model-2, we match the experimental units through using a distance approach and include 125 stores in the treatment and 125 stores in the control group. Observing the results in Model-1 and Model-2, we find that our findings remain consistent.

We also verify that the treatment and control groups follow parallel trends in the pre-BOPS period. The coefficient for TREAT $*$ TREND is not statistically significant in Model-1 ($p > 0.1$, $R^2 = 0.88$) nor in Model-2 ($p > 0.05$, $R^2 = 0.88$).

Table 15: DID Analysis of the BM Channel Using Weekly Aggregation

Variables	STORE PURCHASES Model-1	STORE PURCHASES Model-2
TREAT*AFTER	-0.02^{***} (0.006)	-0.02^{***} (0.006)
TIME-VARYING CONTROLS	YES	YES
YEAR-WEEK FE	YES	YES
STORE FE	YES	YES
REDUCED MODEL	NO	YES
N	18,836	17,000
R^2	0.87	0.87

Note: Standard errors (in parentheses) are heteroskedasticity robust and clustered at the store level. FE denotes fixed effects. REDUCED MODEL refers to employing equal number of units in the treatment and control groups based on distance approach.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

A.3 Analysis of the Online Channel After Removing Extreme Distances

In this section, we replicate our analysis of the online channel after removing DMAs with extreme median customer distances. In Figure 1, we first present the distribution of median customer distances for 206 DMAs that we use in our analyses. Note that in the last bin (> 200 miles), we have 7 DMAs and their median customer distances are as follows: 237.88, 238.96, 261.24, 225.75, 302.94, 1366.16, and 1446.13 miles. The last two distances (1366.16 and 1446.13 miles) belong to market areas located in Alaska, where there is no DS-B store.

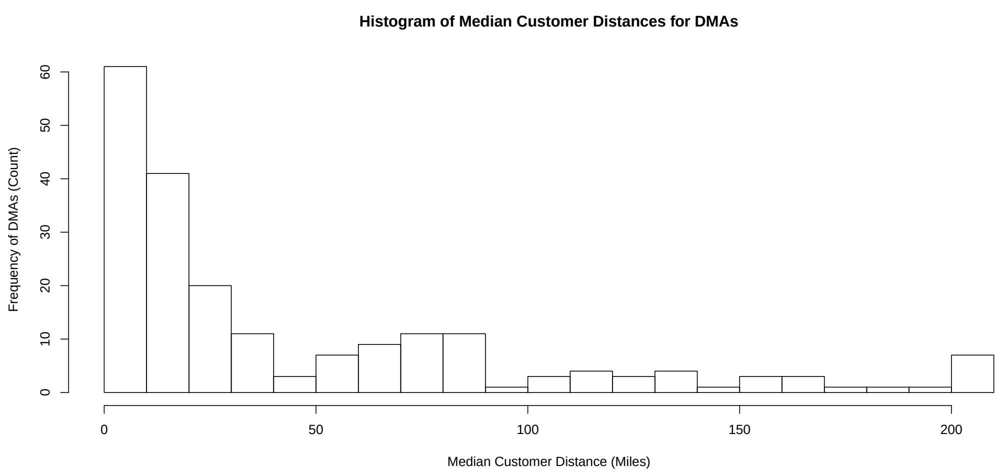

Figure 1: Distribution of Median Customer Distance across DMAs (before)

To identify extreme observations, we focus on the 5th and 95th percentiles, which gives us 4.6 miles and 171.9 miles, respectively. Hence, we eliminate all DMAs with a median customer distance below 4.6 miles and above 171.9 miles. Doing so gives us 185 DMAs for the analysis, 125 of which are in the treatment group and the remaining 60 are in the control group. Figure 2 presents the distribution after we remove the DMAs with extreme median customer distances.

Next, we report our findings in Table 16. Observing the results, we find that our findings remain

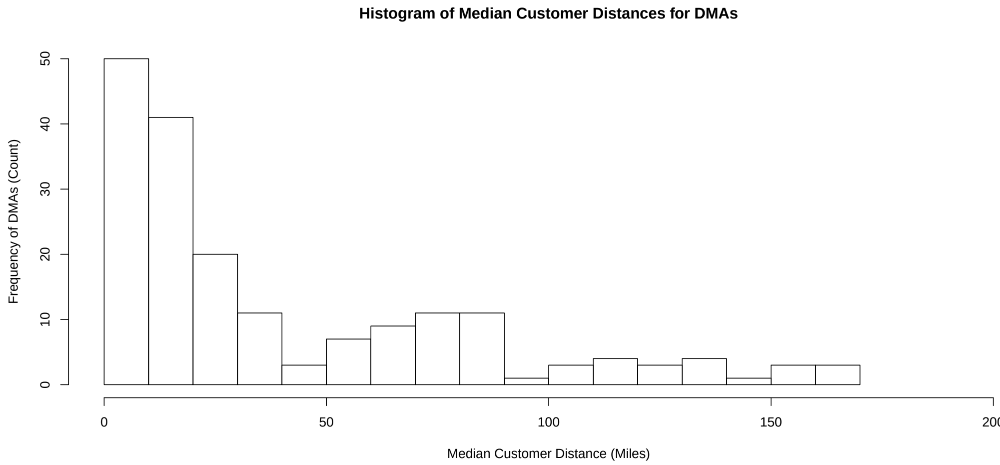

Figure 2: Distribution of Median Customer Distance across DMAs (after)

consistent.

Table 16: DID Analysis of the Online Channel After Removing Extreme Distances

Variables	ONLINE PURCHASES Model-1		ONLINE PURCHASES Model-2	
TREAT*AFTER	-0.05^{**}	(0.017)	-0.05^{**}	(0.019)
TIME-VARYING CONTROLS	YES		YES	
YEAR-MONTH FE	YES		YES	
DMA FE	YES		YES	
PROP. SCORE MATCHING	NO		YES	
N	2,960		1,920	
R^2	0.55		0.52	

Note: Standard errors (in parentheses) are heteroskedasticity robust and clustered at the DMA level. FE denotes fixed effects.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We also verify that the treatment and control groups follow parallel trends in the pre-BOPS period. The coefficient for TREAT $*$ TREND is not statistically significant in Model-1 ($p > 0.1$, $R^2 = 0.37$) and Model-2 ($p > 0.1$, $R^2 = 0.35$). Note that we also replicate the same analyses (removing DMAs with extreme median customer distances) by using weekly aggregation and verify that our findings remain consistent with what we report in Table 16. Furthermore, conducting the same analyses by removing extreme observations only in the right tail, once again we verify that our results are consistent.

A.4 Analysis of the BM Channel After Removing Extreme Distances

In this section, we replicate our analysis of the BM channel after removing stores with extremely low and high distances. In Figure 3, we first present the distribution of distances for 278 DS-A stores that we use in our analyses. Note that for distance >= 150 miles, we have 14 stores and their distances to closest DS-B store range between 151.6 miles and 266.2 miles.

To identify extreme distances, we focus on 5 and 95 percentiles, which gives us 0 miles and 146.47 miles,

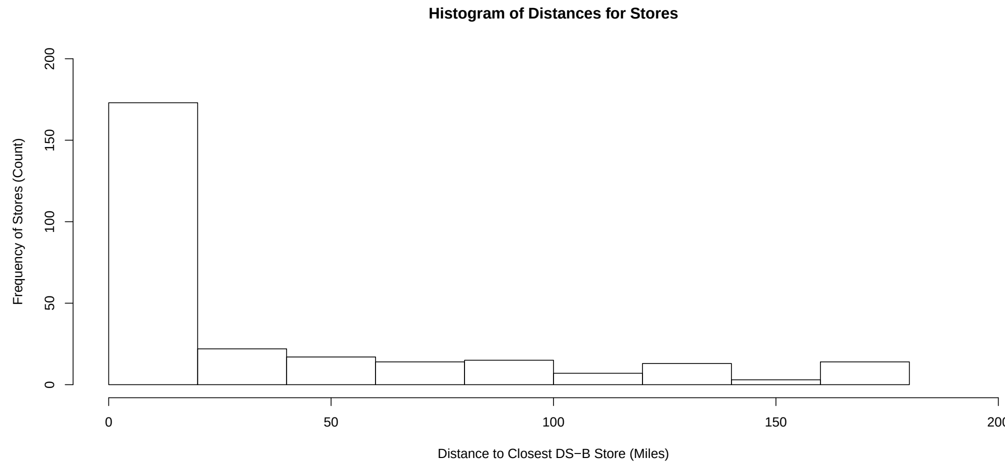

Figure 3: Distribution of Distances across Stores (before)

respectively. Hence, we eliminate all DS-A stores with a distance to the closest DS-B store outside the range of 0 to 146.47 miles. Doing so gives us 244 stores for the analysis, 133 of which are in the treatment group and the remaining 111 are in the control group. Figure 4 presents the distribution after we remove the stores with extreme distances.

Figure 4: Distribution of Distances across Stores (after)

We report the findings in Table 17. Observing the results, we find that our findings remain consistent.

We also verify that the treatment and control groups follow parallel trends in the pre-BOPS period. The coefficient for TREAT $*$ TREND is not statistically significant in Model-1 ($p > 0.1$, $R^2 = 0.85$) and Model-2 ($p > 0.05$, $R^2 = 0.85$). Note that we also replicate the same analyses (removing DS-A stores with extreme distances) by using weekly aggregation and verify that our findings remain consistent with what we report in Table 17. Furthermore, conducting the same analyses by removing extreme observations only in the right tail, once again we verify that our results are consistent.

Table 17: DID Analysis of the BM Channel After Removing Extreme Distances

Variables	STORE PURCHASES Model-1	STORE PURCHASES Model-2
TREAT*AFTER	-0.02^{***} (0.006)	-0.02^{**} (0.006)
TIME-VARYING CONTROLS	YES	YES
YEAR-MONTH FE	YES	YES
STORE FE	YES	YES
REDUCED MODEL	NO	YES
N	3,904	3,552
R^2	0.85	0.85

Note: Standard errors (in parentheses) are heteroskedasticity robust and clustered at the store level. FE denotes fixed effects. REDUCED MODEL refers to employing equal number of units in the treatment and control groups based on distance approach.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

A.5 Placebo Tests for the Online Channel

For the online channel placebo tests, we conduct two separate analyses. First, we adopt a conservative approach by distancing our placebo analyses from the original BOPS launch date. Second, we vary the artificial (fake) implementation date in the original pre-intervention period and conduct our analyses accordingly. We will start with the first one.

Using observations from the pre-intervention period, we pick the artificial implementation date in the middle and replicate our original analyses. We report our results in Table 18. The second column reports findings for the monthly aggregation while the third column reports findings for the weekly aggregation. Observing the results across both models, we see that the coefficient for the variable of interest, TREAT $*$ AFTER, is not significant in any models.

Table 18: Placebo Tests for the Online Channel

Variables	ONLINE PURCHASES (monthly)	ONLINE PURCHASES (weekly)
TREAT*AFTER	0.02 (0.020)	0.02 (0.016)
TIME-VARYING CONTROLS	YES	YES
YEAR-WEEK/MONTH FE	YES	YES
DMA FE	YES	YES
N	1,648	6,494
R^2	0.35	0.53

Note: Standard errors (in parentheses) are heteroskedasticity robust and clustered at the DMA level. FE denotes fixed effects.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We also verify that the treatment and control groups follow parallel trends during the artificial pre-intervention period. The coefficient for TREAT $*$ TREND is not statistically significant in the first model (monthly) ($p > 0.1$, $R^2 = 0.49$) and the second model (weekly) ($p > 0.1$, $R^2 = 0.56$).

Next, we vary the artificial BOPS implementation time during the real pre-intervention period. For this analysis, we use the weekly aggregation approach since it provides us with more flexibility compared to the monthly aggregation. Note that we have 34 weeks in the original pre-intervention period. To have at least 10 weeks before and 10 weeks after, we replicate our analyses choosing the artificial implementation date between week 10 and week 24. We present the results of this placebo test in Figure 5.

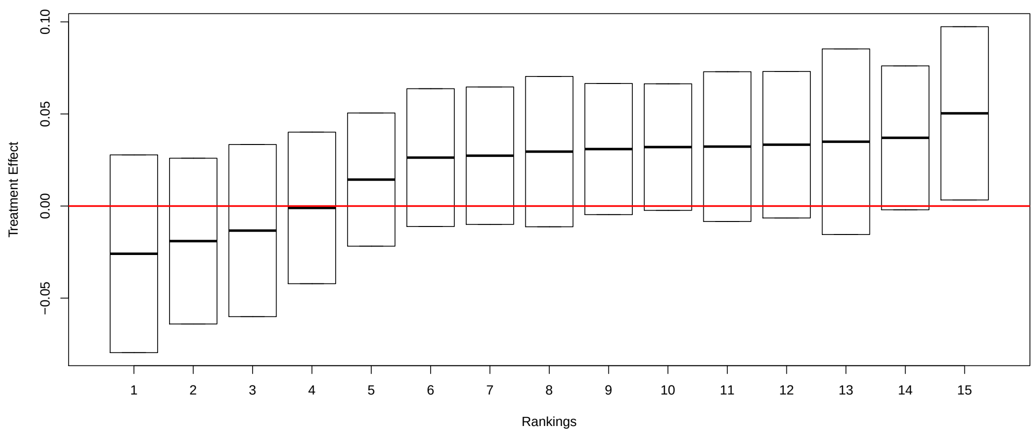

Figure 5: Placebo Test for the Online Channel

Notes: For each trial, we present the estimate for the coefficient and its 95% confidence interval using a box plot. Note that each box plot corresponds to an estimation of equation (1) using DMA-week as the unit of analysis. The confidence interval for our true estimate is [-0.054, -0.008].

Of the 15 trials, we find that only one trial produces a coefficient that is statistically significant at the 5% level but its significance is weaker than our original estimate. We also observe that there are only four trials that generate a negative coefficient for the interaction term, but none of these results are statistically significant ($p > 0.1$). Hence, these findings indicate that the negative impact of BOPS on DS-A's online channel operations does not arise from the structure of the data set.

A.6 Placebo Tests for the BM Channel

Just as we did in §A.5, we conduct two separate placebo tests for the BM channel. First, we adopt a conservative approach by distancing our placebo analyses from the original BOPS launch date. Second, we vary the artificial (fake) implementation date in the original pre-intervention period and conduct our analyses accordingly. We will start with the first one.

Using observations from the pre-intervention period, we pick the artificial implementation date in the middle and replicate our original analyses. We report our results in Table 19. The second column reports findings for the monthly aggregation while the third column reports findings for the weekly aggregation. Observing the results across both models, we see that the coefficient for the variable of interest, TREAT $*$ AFTER, is not significant in any models.

We also verify that the treatment and control groups follow parallel trends during the artificial pre-

Table 19: Placebo Tests for the BM Channel

Variables	**STORE PURCHASES (monthly)**		**STORE PURCHASES (weekly)**	
TREAT*AFTER	−0.01	(0.009)	−0.00	(0.009)
TIME-VARYING CONTROLS	YES		YES	
YEAR-MONTH/WEEK FE	YES		YES	
STORE FE	YES		YES	
N	2,224		9,418	
R^2	0.81		0.88	

Note: Standard errors (in parentheses) are heteroskedasticity robust and clustered at the store level. FE denotes fixed effects.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

intervention period. The coefficient for TREAT $*$ TREND is not statistically significant in the first model (monthly) ($p > 0.1$, $R^2 = 0.80$) and the second model (weekly) ($p > 0.05$, $R^2 = 0.90$).

In our second placebo test, we vary the artificial BOPS implementation time during the pre-intervention period. For this analysis, we use the weekly aggregation approach since it provides us with more flexibility compared to the monthly aggregation. Note that we have 34 weeks in the original pre-intervention period. To have at least 10 weeks before and 10 weeks after, we replicate our analyses choosing the artificial implementation date between week 10 and week 24. We present the results of this placebo test in Figure 6.

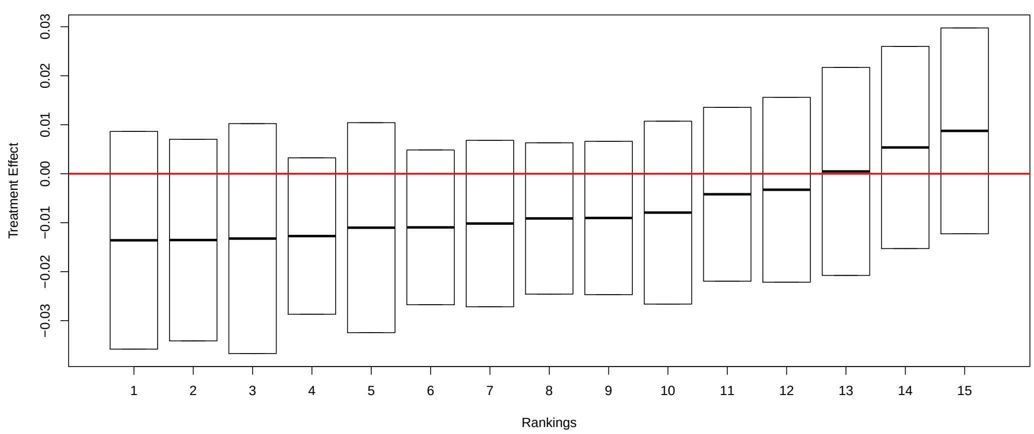

Figure 6: Placebo Test for the BM Channel

Notes: For each trial, we present the estimate for the coefficient and its 95% confidence interval using a box plot. Note that each box plot corresponds to an estimation of equation (3) using STORE-week as the unit of analysis. The confidence interval for our true estimate is [-0.034, -0.012].

Observing the results, we find that there is no trial, which produces a coefficient that is statistically significant at the 5% level. We also observe that there are 12 trials that generate a negative coefficient for the interaction term, but none of these results are statistically significant ($p > 0.1$). Hence, these findings indicate that the negative impact of BOPS on DS-A's BM channel operations does not arise from the structure of the data set.

A.7 Steps for Calculating the Median Customer Distance

To calculate the median/mean customer distance to the closest DS-B store in each DMA, we employ the store location data set for DS-B. We acquired store locations, longitude, and latitude information for all DS-B stores (720 in total) through AggData LLC, which offers data sets regarding current and historical retail store locations. Next, we identify the longitude and latitude information of each U.S. zip code. Note that there exist more than 40K zip codes in the U.S. Subsequently, for each zip code in the U.S., we find the distance to the closest DS-B store using the Haversine great-circle distance between two points on a sphere. Once we calculate the distance between a zip code and the closest DS-B store, we merge this data set with zip code population data set, which we acquired through Unified Data System (UDS.org). In the next step, we identify which DMA each zip code belongs to using a data set we obtained through Truckads.com. Finally, we calculate the weighted median/mean distance for each DMA using the zip code population and distance information within each DMA. After this lengthy process, one can categorize DMAs as either treatment or control group using a specific distance threshold (i.e., 50 miles).

A.8 Using Different Distances for the Online Channel

In this section, we replicate our analysis of the online channel by using different median customer distances to define the treatment and control group DMAs. The additional distances that we investigate include 30, 35, 40, 45 or 60 miles. We report our results in Table 20. For each column in Table 20, we report the findings for the corresponding median customer distance defined in the second row. Since the threshold distance levels change the number of DMAs in the treatment and control groups and we pair the experimental units through propensity score matching, the number of observations (N) is different across models.

Table 20: Robustness Tests for the Online Channel at Different Distances

	ONLINE PURCHASES				
Variables	**(30 Miles)**	**(35 Miles)**	**(40 Miles)**	**(45 Miles)**	**(60 Miles)**
TREAT*AFTER	-0.05^{**}	-0.05^{***}	-0.05^{**}	-0.05^{**}	-0.05^{**}
	(0.016)	(0.016)	(0.017)	(0.017)	(0.018)
TIME-VARYING CONTROLS	YES	YES	YES	YES	YES
YEAR-MONTH FE	YES	YES	YES	YES	YES
DMA FE	YES	YES	YES	YES	YES
N	2,688	2,240	2,336	2,304	2,016
R^2	0.51	0.51	0.49	0.50	0.51

Note: Standard errors (in parentheses) are heteroskedasticity robust and clustered at the DMA level. FE denotes fixed effects.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Observing the results in Table 20, we find that our original findings reported in Table 2 are robust to using different median customer distances.

A.9 Using Different Distances for the BM Channel

In this section, we replicate our analysis of the BM channel by using different distances to define the treatment and control group stores. The additional distances that we investigate include 3, 10, 15 or 20 miles. We report our results in Table 21. For each column in Table 21, we report the findings for the corresponding distance defined in the second row. Since the threshold distance levels change the number of stores in the treatment and control groups and we pair the experimental units through distance approach, the number of observations (N) is different across models.

Table 21: Robustness Tests for the BM Channel at Different Distances

	STORE PURCHASES			
Variables	**(3 Miles)**	**(10 Miles)**	**(15 Miles)**	**(20 Miles)**
TREAT*AFTER	-0.02^{**}	-0.02^{**}	-0.02^{**}	-0.02^{**}
	(0.006)	(0.006)	(0.006)	(0.006)
TIME-VARYING CONTROLS	YES	YES	YES	YES
YEAR-MONTH FE	YES	YES	YES	YES
STORE FE	YES	YES	YES	YES
N	4,320	3,584	3,456	3,360
R^2	0.85	0.86	0.86	0.86

Note: Standard errors (in parentheses) are heteroskedasticity robust and clustered at the store level. FE denotes fixed effects.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Observing the results in Table 21, we find that our original findings reported in Table 4 are robust to using different distances.

A.10 Analysis of the BM Channel with an Alternative Approach

In this section, we replicate our analyses of the BM channel by using an alternative approach in defining the treatment and control groups for the BM channel. Rather than using the distance to the closest DS-B store location, we focus on whether a DS-A store is co-located with a DS-B store. As such, the treatment group consists of DS-A stores that are located in malls where a DS-B store is also present while the control group consists of DS-A stores that are located in different malls.

We report our results in Table 22. The second column (Model-1) reports the findings when we include 129 DS-A stores in the treatment group and 149 DS-A stores in the control group. In the the third column (Model-2), we report our findings after removing 20 stores from the control group such that we have 129 DS-A stores in the treatment group and 129 DS-A stores in the control group. To eliminate the 20 DS-A stores from the control group, we identify each DS-A store's distance to the closest DS-B store using the

Haversine great-circle distance between two points on a sphere. Although these 20 DS-A stores are not located in malls where a DS-B store is also present, their distance to a DS-B store is short such that their mean distance to the closest DS-B store is 1.49 miles. Hence, we remove them from our analyses in Model-2. Note that for the remaining 129 DS-A stores in the control group, the average distance to the closest DS-B store is 75.18 miles.

Table 22: Analysis of the BM Channel with an Alternative Approach

Variables	**STORE PURCHASES Model-1**		**STORE PURCHASES Model-2**	
TREAT*AFTER	-0.02^{**}	(0.006)	-0.02^{**}	(0.006)
TIME-VARYING CONTROLS	YES		YES	
YEAR-MONTH FE	YES		YES	
STORE FE	YES		YES	
REDUCED MODEL	NO		YES	
N	4,448		4,128	
R^2	0.86		0.85	

Note: Standard errors (in parentheses) are heteroskedasticity robust and clustered at the store level. FE denotes fixed effects. REDUCED MODEL refers to employing equal number of units in the treatment and control groups based on distance approach.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Observing the results in Table 22, we find that our original findings reported in Table 4 are robust to using alternative approaches in creating the treatment and control groups.

A.11 Propensity Score Matching for the Online Channel

In Model-2 of Table 2, we report the results of our analyses for the online channel after pairing the experimental units. Here, we explain how we conducted the pairing process using propensity score matching. To do so, we first estimate each DMA's propensity of being in the treatment group through the use of a logistic regression model (LOGIT), where the response variable TREAT is equal to one if the DMA is in the treatment group and zero if the DMA is in the control group. We employ 25 performance measures (see §A.18 for the list of variables) for each designated market area (DMA) and five competitor specific factors. Examining the output of the LOGIT model, we see that residual deviance is 179.75 compared to 264.05 with only the intercept term. Using a well-known goodness of fit approach for limited dependent variables by McKelvey and Zavoina (1975), we also find that the pseudo-R^2 is equivalent to 0.77. Veall and Zimmerman (1996) show that pseudo-R^2 scores generated by McKelvey-Zavoina method were closest to the R^2 value of the underlying linear model. Seeing that the model fits well, we pair the treatment and control group DMAs by using the fitted values reported by the logistic regression model. Note that there are 70 DMAs in the control group. Hence, we choose 70 DMAs from the treatment group of 136 DMAs by pairing them with DMAs in the control group. The propensity score matching helps us conduct the DID analysis by employing DMAs with similar characteristics.

In Table 23, we present the results of the matching process. Note that we break the table into two parts. In the second and third columns, we report the means of variables for the treatment and control group DMAs before we match the DMAs (206 total DMAs). In the fourth and fifth columns, we report the means of variables for the treatment and control group DMAs after we match the DMAs (140 total DMAs).

Table 23: Propensity Score Matching for the Online Channel

	Before Matching		After Matching	
Variables	Treatment	Control	Treatment	Control
HEALTH CARE	16.82	17.44	17.68	17.44
RETAIL TRADE	15.11	15.73	15.48	15.73
ACCOMMODATION	8.37	9.77	7.07	9.77
FINANCE & INSURANCE	6.42	6.76	6.57	6.76
PUBLIC ADMIN	6.48	6.39	6.33	6.39
MANUFACTURING	8.03	9.02	8.22	9.02
TECH SERVICES	8.20	9.61	8.61	9.61
CONSTRUCTION	7.24	6.90	7.51	6.90
WASTE MANAGEMENT	4.04	4.13	4.07	4.13
RECREATION	3.96	3.41	3.82	3.41
GROCERIES	0.18	0.19	0.18	0.19
MEALS & BEVERAGE	0.09	0.07	0.09	0.07
ALCOHOL	0.01	0.01	0.01	0.01
DRUGS & BEAUTY	0.12	0.10	0.12	0.10
MAJOR APPLIANCES	0.01	0.01	0.01	0.01
FURNITURE	0.02	0.02	0.02	0.02
WOMEN CLOTHING	0.03	0.02	0.03	0.02
SPORTING GOODS	0.01	0.01	0.01	0.01
VEHICLES	0.16	0.16	0.16	0.16
UNEMPLOY	5.05	4.79	4.83	4.79
WHITE	0.71	0.76	0.72	0.76
INCOME-1	0.42	0.45	0.42	0.45
INCOME-2	0.30	0.30	0.30	0.30
INCOME-3	0.12	0.11	0.12	0.11
INCOME-4	0.08	0.06	0.08	0.06
STORES	0.64	0.55	0.69	0.55
MEDIAN DISTANCE	3.77	3.72	3.70	3.72
CLOSURES	0.01	0.00	0.00	0.00
OPENINGS	0.01	0.00	0.01	0.00
FOCAL CLOSURES	0.03	0.00	0.03	0.00

Observing Table 23, we see that matching experimental units by employing propensity scores helps us reduce the differences between the treatment and control group DMAs. Before the matching, for example, HEALTH CARE is 16.82 and 17.44 for the two groups, respectively. After the matching, HEALTH CARE is 17.68 and 17.44 for the treatment and control group DMAs, which is around 61% improvement for the difference. We observe similar improvements in differences for other variables in Table 23. After pairing the experimental units, we perform the DID analysis by including only 140 DMAs (70 in each group) in our model instead of the original 206 DMAs. For this model (Model-2), we report our findings in the second column of Table 2 and show that they are consistent with Model-1.

A.12 Random Assignment of Treatment for the Online Channel

Although we conduct propensity score matching analyses and check pre-intervention trends for all of our models, we nevertheless consider that heterogeneity of DMAs could be a concern for our analyses. To rule out these concerns and make sure that our findings are not driven by major differences between the treatment and control groups, we remove DMAs that are more likely to be in the treatment group depending on annual retail sales. First, we identify the mean annual retail sales across the 206 DMAs. Next, we create a subsample by eliminating DMAs generating more than 25.8 million dollars in annual retail sales, which gives us 150 DMAs. Hence, we drop 56 biggest DMAs from our analysis. Note that by employing this subsample, we are able to mimic characteristics of a controlled experiment, in which the treatment is randomly assigned across the experimental units (Gallino and Moreno 2014). We report our results in Table 24.

Table 24: Random Assignment Test for the Online Channel

Variables	**ONLINE PURCHASES Model-1**		**ONLINE PURCHASES Model-2**	
TREAT*AFTER	-0.05^{**}	(0.019)	-0.05^{**}	(0.018)
TIME-VARYING CONTROLS	YES		YES	
YEAR-MONTH FE	YES		YES	
DMA FE	YES		YES	
PROP. SCORE MATCHING	NO		YES	
N	2,400		2,208	
R^2	0.47		0.49	

Note: Standard errors (in parentheses) are heteroskedasticity robust and clustered at the DMA level. FE denotes fixed effects.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The second column (Model-1) reports the findings when we include 81 DMAs in the treatment group and 69 DMAs in the control group. In the the third column (Model-2), we report our findings after pairing DMAs through propensity score matching such that we have 69 DMAs in the treatment group and 69 DMAs in the control group. Observing both models in Table 24, we find that our results are consistent with what we report in Table 2.

A.13 Random Assignment of Treatment for the BM Channel

Just as we did in §A.12, we remove DS-A stores that are more likely to be in the treatment group depending on the population around a ten-mile radius of each store. First, we identify the mean population across the 278 DS-A stores. Next, we create a subsample by eliminating stores with more than 431K residents around a ten-mile radius, which gives us 168 DS-A stores. Hence, we drop 110 DS-A stores from our analysis. We report our results in Table 25.

Table 25: Random Assignment Test for the BM Channel

Variables	STORE PURCHASES Model-1		STORE PURCHASES Model-2	
TREAT*AFTER	-0.02^{**}	(0.006)	-0.03^{***}	(0.008)
TIME-VARYING CONTROLS	YES		YES	
YEAR-MONTH FE	YES		YES	
STORE FE	YES		YES	
REDUCED MODEL	NO		YES	
N	2,688		1,792	
R^2	0.88		0.87	

Note: Standard errors (in parentheses) are heteroskedasticity robust and clustered at the store level. FE denotes fixed effects. REDUCED MODEL refers to employing equal number of units in the treatment and control groups based on distance approach.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The second column (Model-1) reports the findings when we include 56 DS-A stores in the treatment group and 112 DS-A stores in the control group. In the the third column (Model-2), we report our findings after removing 56 stores from the control group such that we have 56 DS-A stores in the treatment group and 56 DS-A stores in the control group. Observing both models in Table 25, we find that our results are consistent with what we report in Table 4.

A.14 Analysis of the Online Channel with an Alternative Approach

In this section, we replicate our analyses of the online channel by using zip codes as the experimental units, which is a more granular analysis compared to the DMA regions. In defining the treatment and control groups, we assign a zip code to the treatment group if the distance of the zip code to the closest DS-B store is less than 5 miles. Otherwise, the zip code is in the control group. At the zip code level, we adopt two different approaches: Zip code-month and zip code-week. We report our results in Table 26.

Table 26: Analysis of the Online Channel using Zip codes

Variables	ONLINE PURCHASES (Zip code-Month)		ONLINE PURCHASES (Zip code-Week)	
TREAT*AFTER	-0.03^{***}	(0.008)	-0.02^{**}	(0.006)
TIME-VARYING CONTROLS	YES		YES	
YEAR-MONTH FE	YES		NO	
YEAR-WEEK FE	NO		YES	
N	70,638		176,983	
R^2	0.15		0.24	

Note: Standard errors (in parentheses) are heteroskedasticity robust and clustered at the DMA level. FE denotes fixed effects.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The second column in Table 26 reports the results of the analysis when we use Zip code-month and pair experimental units through propensity score matching such that we have 2,254 zip codes in the treatment group and 2,254 zip codes in the control group. The third column in Table 26 reports the results of the analysis when we use Zip code-week and pair experimental units through propensity score matching such that we have 1,570 zip codes in the treatment group and 1,570 zip codes in the control group. Observing both models in Table 26, we find that our results are consistent with what we report in Table 2. As an additional robustness check, we also replicate the same analyses using 10 miles to define the treatment and control group zip codes and find that our results are consistent.

Please note that for the propensity score matching, we pair the zip codes by employing a total of five variables: Three at the zip code-level and two at the DMA level. At the zip code level, we use total number of business establishments (ESTABLISH), population (ZPOP), and the number of housing units in the county that the zip code belongs to (COHU). At the DMA level, we use total retail sales (RETAIL SALES) and unemployment rate (DMA UNEMPLOYMENT). The balancing properties show that we achieve at least 61% improvement for each variable that we use in the propensity score matching analysis.

A.15 Analysis of the Online Channel Using Dollar Sales as the Dependent Variable

In this section, we replicate our analysis of the online channel using dollar value of sales as the dependent variable rather than number of purchase transactions. We report our results in Table 27.

Table 27: Online Channel - Dollar Sales as the Dependent Variable

Variables	**ONLINE SALES Model-1**	**ONLINE SALES Model-2**
TREAT*AFTER	-0.06^{***} (0.018)	-0.07^{***} (0.020)
TIME-VARYING CONTROLS	YES	YES
YEAR-MONTH FE	YES	YES
DMA FE	YES	YES
PROP. SCORE MATCHING	NO	YES
N	3,296	2,240
R^2	0.46	0.44

Note: Standard errors (in parentheses) are heteroskedasticity robust and clustered at the DMA level. FE denotes fixed effects.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The second column (Model-1) reports the findings when we include 136 DMAs in the treatment group and 70 DMAs in the control group. In the the third column (Model-2), we report our findings after matching the DMAs using propensity scores such that we have 70 DMAs in the treatment group and 70 DMAs in the control group. Observing both models in Table 27, we find that our results are consistent with what we report in Table 2.

A.16 Analysis of the BM Channel Using Dollar Sales as the Dependent Variable

Just as we did in §A.15, we replicate our analysis of the BM channel using dollar value of sales as the dependent variable rather than number of purchase transactions. We report our results in Table 28.

Table 28: BM Channel - Dollar Sales as the Dependent Variable

Variables	STORE SALES Model-1	STORE SALES Model-2
TREAT*AFTER	-0.01^{**} (0.005)	-0.02^{**} (0.005)
TIME-VARYING CONTROLS	YES	YES
YEAR-MONTH FE	YES	YES
STORE FE	YES	YES
REDUCED MODEL	NO	YES
N	4,448	4,000
R^2	0.80	0.80

Note: Standard errors (in parentheses) are heteroskedasticity robust and clustered at the store level. FE denotes fixed effects. REDUCED MODEL refers to employing equal number of units in the treatment and control groups based on distance approach.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The second column of Table 28 (Model-1) reports the findings when we include 153 DS-A stores in the treatment group and 125 DS-A stores in the control group. In the the third column (Model-2), we report our findings after removing 28 stores from the treatment group such that we have 125 DS-A stores in the treatment group and 125 DS-A stores in the control group. To eliminate the 28 DS-A stores from the treatment group, we identify each DS-A store's distance to the closest DS-B store using the Haversine great-circle distance between two points on a sphere and remove the most distant 28 stores. Observing both models in Table 28, we find that our results are consistent with what we report in Table 4.

A.17 Propensity Score Weighting for the Online Channel

In this section, we replicate our analysis of the online channel using propensity score weighting rather than propensity score matching. Conducting a logistic regression analysis, we first identify propensity scores for each experimental unit (DMA). In this stage, we employ 25 performance measures (see §A.18 for the list of variables) for each designated market area (DMA) and five competitor specific factors. Examining the output the LOGIT model, we see that residual deviance is 179.75 compared to 264.05 with only the intercept term (Pseudo-$R^2 = 0.77$). The goodness of fit measures show that our logistic regression model fits well.

Consistent with previous work (Rosenbaum 1987, Hirano and Imbens 2001, Hirano et al. 2003, Bell et al. 2018), we use propensity scores as sampling weights and use them to reweight the treatment and control group observations so that the two groups will be comparable in terms of the observable characteristics. Note that this analysis is doubly robust since we combine the outcome analysis (i.e., DID) with weighting by the propensity scores (Funk et al. 2011). Following Hirano and Imbens (2001), we define

$$\omega(W,x) = \frac{W}{\hat{e}(x)} + \frac{1-W}{1-\hat{e}(x)},$$

where $W = 1$ represents a treatment group DMA and $\hat{e}(x)$ corresponds to the estimated probability of being in the treatment group. Once we identify the weights for each DMA, we estimate the DID model by including the weights for each experimental unit. We report our results in Table 29.

Table 29: Online Channel - Propensity Score Weighting

Variables	**ONLINE PURCHASES** **Model-1**		**ONLINE SALES** **Model-2**	
TREAT*AFTER	-0.05^{***}	(0.014)	-0.08^{***}	(0.017)
TIME-VARYING CONTROLS	YES		YES	
YEAR-MONTH FE	YES		YES	
DMA FE	YES		YES	
PROP SCORE WEIGHTING	YES		YES	
N	3,296		3,296	
R^2	0.51		0.45	

Note: Standard errors (in parentheses) are heteroskedasticity robust and clustered at the DMA level. FE denotes fixed effects.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The second column (Model-1) reports the findings when we use online purchases (count) as the dependent variable while the third column (Model-2) reports the findings when we use dollar value of online sales as the dependent variable. Observing both models in Table 29, we find that our results are consistent with what we report in Table 2. Please also note that we replicate the whole process using a probit link function and verify that our results are consistent. Furthermore, we also confirm that propensity score weights balance the treatment and control group DMAs properly. Following (Guo and Fraser 2009), we simply compare estimates for covariates from the weighted and unweighted regressions, where the dependent variable is a particular covariate from the logistic (or probit) regression model. For example, running the weighted and unweighted regressions for the WOMEN CLOTHING variable, we observe that the estimate from the unweighted regression is highly significant ($\beta = 0.01$, $p < 0.001$), while the the estimate from the weighted regression is not statistically significant ($\beta = 0.00$, $p > 0.1$).

A.18 List of Variables Used for the Propensity Score Weighting Analysis

In this section, we list and explain the whole set of additional variables that we use in the propensity score weighting analysis. We obtained performance measures for each market area (DMA) from the Standard Rate and Data Service (SRDS) database managed by Kantar Media. After eliminating highly correlated performance measures, we employ 25 DMA-level factors for this analysis. In addition, we also include five competitor specific factors. Table 30 presents the definition of these variables.

Table 30: Full Set of Variables Used in the Propensity Score Weighting

Variable Name	Description
	DMA Level Factors
HEALTH CARE	Number of Health Care Establishments for 1,000 people
RETAIL TRADE	Number of Retail Trade Establishments for 1,000 people
ACCOMMODATION	Number of Accommodation and Food Service Establishments for 1,000 people
FINANCE & INSURANCE	Number of Finance and Insurance Establishments for 1,000 people
PUBLIC ADMIN	Number of Public Administration Establishments for 1,000 people
MANUFACTURING	Number of Manufacturing Establishments for 1,000 people
TECH SERVICES	Number of Technology Service Establishments for 1,000 people
CONSTRUCTION	Number of Construction Establishments for 1,000 people
WASTE MANAGEMENT	Number of Waste Management Establishments for 1,000 people
RECREATION	Number of Entertainment and Recreation Establishments for 1,000 people
GROCERIES	Grocery Sales as a Fraction of Total Retail Sales
MEALS & BEVERAGE	Meal & Unpacked Snack Sales as a Fraction of Total Retail Sales
ALCOHOL	Alcoholic Drink Sales as a Fraction of Total Retail Sales
DRUGS & BEAUTY	Drugs & Beauty Aid Sales as a Fraction of Total Retail Sales
MAJOR APPLIANCES	Major Household Appliance Sales as a Fraction of Total Retail Sales
FURNITURE	Furniture & Sleep Equipment Sales as a Fraction of Total Retail Sales
WOMEN CLOTHING	Women Clothing Sales as a Fraction of Total Retail Sales
SPORTING GOODS	Sporting Good Sales as a Fraction of Total Retail Sales
VEHICLES	Automobile, Truck, and Van Sales as a Fraction of Total Retail Sales
UNEMPLOY	Unemployment Rate for Adult Population
WHITE	Fraction of Residents with White Ethnicity
INCOME-1	Fraction of Households with Income Level less than 50K
INCOME-2	Fraction of Households with Income Level between 50K and 100K
INCOME-3	Fraction of Households with Income Level between 100K and 150K
INCOME-4	Fraction of Households with Income Level greater than 150K
	Competitor-Specific Factors
STORES	Number of Competitor Stores in each DMA
MEDIAN DISTANCE	Median Customer Distance to the Closest Competitor Store in each DMA
CLOSURES	Number of Store Closures by the Competitor in each DMA
OPENINGS	Number of Store Openings by the Competitor in each DMA
FOCAL CLOSURES	Number of Focal Store Closures in each DMA

A.19 High-Value and Low-Value Product Returns for the Online Channel

In this section, we conduct additional analysis for online product returns by categorizing products as low-value and high-value. Recall that the mechanism that we present in the manuscript states that BOPS will have a greater effect on high-value items compared to low-value items due to higher uncertainty associated with high-value purchases. For product returns, this translates into the fact that returns of high-value purchases should decrease more than returns of low-value purchases. This arises because customers that are staying with DS-A are more certain about their purchases. For this analysis, we create a fractional (percentage) variable for low-value transactions and observe its interaction with TREAT $*$ AFTER. To do so, we divide the number of low-value transactions by the total number of transactions using \$100 as the threshold to categorize products as low-value. Note that the correlation among the number of low-value transactions at \$50, \$100, and \$150 cutoff values are between 0.70 and 0.85.

With this analysis, we expect to observe that as the fraction of low-value transactions increases, the decrease in the return rate should diminish. This arises because compared to high-value purchases, not many customers switch to DS-B for low-value purchases. As such, the return rate for low-value purchases should not change much whereas the return rate for high-value purchases should decrease significantly. For the estimation, we will observe the interaction of TREAT$*$AFTER with the fraction of low-value transactions. We report the findings in Table 31.

Table 31: Online Returns and Product Value

Variables	**ONLINE RETURNS Model-1**		**ONLINE RETURNS Model-2**	
TREAT*AFTER	-0.52^{**}	(0.190)	-0.67^{**}	(0.215)
TREAT*AFTER*LOW VALUE	0.46^{*}	(0.220)	0.75^{**}	(0.252)
ONLINE PURCHASES	0.66^{***}	(0.064)	0.74^{***}	(0.050)
TIME-VARYING CONTROLS	YES		YES	
YEAR-MONTH FE	YES		YES	
DMA FE	YES		YES	
PROP. SCORE WEIGHTING	NO		YES	
N	2,848		2,848	
R^2	0.37		0.36	

Note: Standard errors (in parentheses) are heteroskedasticity robust and clustered at the DMA level. FE denotes fixed effects.
$*$ $p < 0.05$, $**$ $p < 0.01$, $***$ $p < 0.001$

The second column (Model-1) in Table 31 reports findings when we run the model without using propensity score weighting while the third column reports the findings when we employ the propensity score weighting. Observing Table 31, we see that the coefficient of the original interaction term TREAT $*$ AFTER is negative and statistically significant in both models. This shows that compared to control group DMAs, ONLINE RETURNS in the treatment group DMAs significantly decrease after BOPS launch by DS-B. When we look at the three-way interaction terms, however, we observe that as the fraction of low-value transactions increases, the decrease in ONLINE RETURNS slows down as the coefficients for the three-way interaction terms are positive and statistically significant in both models.

A.20 Robustness Tests for the Sub-sample Analyses in the Online Channel

In this section, we create fractional (percentage) variables for the loyalty card transactions, online only (exclusive) transactions, and low-value transactions and observe their interactions with TREAT ∗ AFTER. In doing so, we divide the number of transactions for these sub-samples by the total number of transactions. For the low-value items, we use $100 to identify the number of low-value transactions. Please note that the correlation among the number of low-value transactions at $50, $100, and $150 cutoff values are between 0.70 and 0.85 and our results are robust when we use $50 or $150 as cutoff values to categorize products as low-value. We report our findings in Table 32.

Table 32: Interaction Terms for the Online Channel

ONLINE PURCHASES				
Variables	**Model-1**		**Model-2**	
TREAT*AFTER	-0.48^{***}	(0.095)	-0.44^{***}	(0.101)
TREAT*AFTER*LOYALTY	0.21^{**}	(0.074)	0.23^{**}	(0.079)
TREAT*AFTER*ONLINE ONLY	0.53^{**}	(0.168)	0.71^{***}	(0.174)
TREAT*AFTER*LOW VALUE	0.40^{***}	(0.111)	0.35^{**}	(0.119)
TIME-VARYING CONTROLS	YES		YES	
MONTH-YEAR FE	YES		YES	
DMA FE	YES		YES	
PROP. SCORE WEIGHTING	NO		YES	
N	3,296		3,296	
R^2	0.53		0.53	

Note: Standard errors (in parentheses) are heteroskedasticity robust and clustered at the DMA level. FE denotes fixed effects.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The second column (Model-1) in Table 32 reports findings when we run the model without using propensity score weighting while the third column reports the findings when we employ the propensity score weighting. Observing the results, we find that the coefficient of the original interaction term TREAT ∗ AFTER is negative and statistically significant in both models, which is consistent with our original model. Once we check the interaction of LOYALTY, ONLINE ONLY, and LOW VALUE variables with the TREAT ∗ AFTER, we find that coefficients for all three-way interaction terms are positive and statistically significant. Please note that in our main subsample analyses, we show that DS-B's launch of BOPS service offering does not have a statistically significant impact on the loyalty card users and online only items. The results in Table 32 provide further evidence for those findings. For the product value analyses in the manuscript, we show that BOPS service offering has a lower negative influence on the low-value items compared to high-value items due to lower levels of valuation uncertainty for the low-value items. Again, the results in Table 32 provide further evidence for that finding.

A.21 Robustness Tests for the Sub-sample Analyses in the BM Channel

Just as we did in §A.20, we create a fractional (percentage) variable FURNITURE for "furniture and home furnishing" items and observe its interaction with TREAT ∗ AFTER. In doing so, we divide the number of transactions for "furniture and home furnishing" by the total number of transactions in the BM channel. Please note that the analysis for "furniture and home furnishing" items belongs to the BM channel. We report our findings in Table 33.

Table 33: Interaction Term for the Offline Channel

STORE PURCHASES				
Variables	**Model-1**		**Model-2**	
TREAT*AFTER	-0.04^{***}	(0.010)	-0.04^{***}	(0.010)
TREAT*AFTER*FURNITURE	0.36^{**}	(0.118)	0.40^{**}	(0.125)
TIME-VARYING CONTROLS	YES		YES	
MONTH-YEAR FE	YES		YES	
STORE FE	YES		YES	
REDUCED MODEL	NO		YES	
N	4,448		4,000	
R^2	0.86		0.85	

Note: Standard errors (in parentheses) are heteroskedasticity robust and clustered at the store level. FE denotes fixed effects. REDUCED MODEL refers to employing equal number of units in the treatment and control groups based on distance approach.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Observing the results in Table 33, both with and without distance-based matching, we find that the coefficient of the original interaction term TREAT ∗ AFTER is negative and statistically significant in both models, which is consistent with our original model. Observing the interaction of FURNITURE variable with the TREAT ∗ AFTER, we find that the coefficient of the three-way interaction term is positive and statistically significant. Note that in our main sub-sample analysis, we show that DS-B's launch of BOPS service offering does not have any statistically significant impact on the "furniture and home furnishing" items. Hence, the results in Table 33 provide further evidence for that finding.

A.22 Product Value Analysis for the BM Channel

In this section, we conduct product value analysis (high-value vs. low-value) for DS-A's sales at the BM channel. We report the results in Table 34.

Observing Table 34, we find that purchases of high-value items at \$100 and \$150 price thresholds decrease more than those of low-value items at the treatment group stores relative to control group stores. Please note that while the significance is comparable, the magnitude of the decrease for high-value items is greater than that for low-value items. At \$50 price threshold, we find that while the magnitudes of the decrease across both groups are similar, the low-value items has a higher significance than high-value items.

Table 34: High-Value versus Low-Value Analyses for the BM Channel

Price Threshold	**50 Dollar**		**100 Dollar**		**150 Dollar**	
Variables	**LOW**	**HIGH**	**LOW**	**HIGH**	**LOW**	**HIGH**
TREAT*AFTER	-0.02^{***}	-0.02^{*}	-0.02^{***}	-0.04^{***}	-0.02^{***}	-0.06^{***}
	(0.006)	(0.007)	(0.006)	(0.011)	(0.006)	(0.012)
TIME-VARYING CONTROLS	YES	YES	YES	YES	YES	YES
YEAR-MONTH FE	YES	YES	YES	YES	YES	YES
STORE FE	YES	YES	YES	YES	YES	YES
N	4,448	4,448	4,448	4,416	4,448	4,288
R^2	0.87	0.76	0.86	0.77	0.86	0.79

Note: Standard errors (in parentheses) are heteroskedasticity robust and clustered at the store level. FE denotes fixed effects.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In effect, the results we report here for high and low value items at DS-A's BM channel mirrors those that we previously discussed for the online channel. At the BM channel, while BOPS itself offers no additional "uncertainty mitigation" (i.e., valuation uncertainty and product fit uncertainty) that is not already available with store shopping, it still provides product availability information to customers. Store customers value the product availability information enabled through BOPS since it eliminates the risk of a wasted trip to the store. Consequently, DS-A is losing sales to BM customers in addition to online customers.

A.23 Pre-Intervention Trend for the Online Channel

Table 35: Pre-Intervention Trends for the Online Channel

Variables	**ONLINE PURCHASES Model-1**		**ONLINE PURCHASES Model-2**	
TREND	-0.03^{***}	(0.005)	-0.04^{***}	(0.005)
TREAT*TREND	0.00	(0.005)	0.00	(0.005)
TIME-VARYING CONTROLS	YES		YES	
MONTH-YEAR FE	YES		YES	
DMA FE	YES		YES	
PROP. SCORE MATCHING	NO		YES	
N	1,648		1,120	
R^2	0.35		0.35	

Note: Standard errors (in parentheses) are heteroskedasticity robust and clustered at the DMA level. FE denotes fixed effects.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Observing Table 35, we find that the coefficient of the interaction term TREAT*TREND is not statistically significant in any model. Hence, we verify that the treatment and control groups follow parallel trends for the online channel analysis.

A.24 Pre-Intervention Trend for the BM Channel

Table 36: Pre-Intervention Trends for the BM Channel

Variables	STORE PURCHASES Model-1		STORE PURCHASES Model-2	
TREND	0.00*	(0.001)	0.00*	(0.001)
TREAT*TREND	−0.00	(0.002)	−0.00	(0.002)
TIME-VARYING CONTROLS	YES		YES	
MONTH-YEAR FE	YES		YES	
STORE FE	YES		YES	
REDUCED MODEL	NO		YES	
N	2,224		2,000	
R^2	0.86		0.86	

Note: Standard errors (in parentheses) are heteroskedasticity robust and clustered at the store level. FE denotes fixed effects. REDUCED MODEL refers to employing equal number of units in the treatment and control groups based on distance approach.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Observing Table 36, we find that the coefficient of the interaction term TREAT*TREND is not statistically significant in any model. Hence, we verify that the treatment and control groups follow parallel trends for the BM channel analysis.

Table 37: Correlation of Variables for the Online Channel

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. ONLINE SALES	1.00													
2. ONLINE RETURNS	0.98	1.00												
3. ONLINE PURCHASES	0.60	0.51	1.00											
4. NUMBER OF RETURNS	0.90	0.89	0.79	1.00										
5. MEDIAN DISTANCE	-0.11	-0.10	-0.13	-0.12	1.00									
6. TREAT	0.19	0.17	0.24	0.21	-0.43	1.00								
7. TOTAL HOUSEHOLD	0.55	0.46	0.89	0.70	-0.17	0.32	1.00							
8. DMA UNEMPLOYMENT	-0.04	-0.03	-0.07	-0.06	0.00	0.06	-0.04	1.00						
9. DMA RETAIL SALES	0.45	0.40	0.67	0.54	-0.14	0.27	0.86	0.01	1.00					
10. AFTER	0.10	0.04	0.08	0.11	0.00	0.00	0.00	-0.11	-0.16	1.00				
11. FRACTION - LOYALTY CARD	0.20	0.25	-0.10	0.14	0.01	-0.12	-0.12	0.18	-0.09	-0.08	1.00			
12. FRACTION - ONLINE ONLY	-0.04	-0.06	0.15	0.03	-0.01	0.13	0.13	-0.04	0.02	0.14	-0.19	1.00		
13. FRACTION - LOW VALUE	0.14	0.20	-0.08	0.11	0.05	-0.12	-0.06	0.06	0.00	-0.20	0.44	-0.25	1.00	
14. TREATMENT INTENSITY	0.25	0.21	0.38	0.30	-0.40	0.71	0.47	0.00	0.41	0.00	-0.20	0.14	-0.13	1.00

Note: Covariates with high correlations are used in different models (main model, propensity score matching, robustness etc.)

Table 38: Correlation of Variables for the BM Channel

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. STORE SALES	1.00														
2. STORE RETURNS	0.93	1.00													
3. STORE PURCHASES	0.94	0.88	1.00												
4. NUMBER OF RETURNS	0.90	0.96	0.94	1.00											
5. DISTANCE	-0.01	-0.11	-0.01	-0.10	1.00										
6. TREAT	0.19	0.29	0.17	0.27	-0.70	1.00									
7. AFTER	0.04	0.04	-0.05	-0.05	0.00	0.00	1.00								
8. UNEMPLOYMENT	-0.12	-0.09	-0.09	-0.06	-0.10	0.04	-0.25	1.00							
9. POPULATION	0.27	0.35	0.25	0.31	-0.45	0.56	0.00	-0.03	1.00						
10. HOUSEHOLD INCOME	0.16	0.27	0.14	0.24	-0.25	0.35	0.00	-0.11	0.44	1.00					
11. MEDIAN AGE	-0.21	-0.15	-0.21	-0.16	-0.10	0.04	0.00	0.24	-0.13	0.01	1.00				
12. FRACTION - LOYALTY CARD	-0.15	-0.05	-0.05	0.06	-0.14	0.11	-0.09	0.20	-0.06	-0.28	0.12	1.00			
13. FRACTION - LOW VALUE	-0.57	-0.56	-0.32	-0.38	0.09	-0.22	-0.36	0.10	-0.20	-0.16	0.07	0.22	1.00		
14. FRACTION - FURNITURE	0.04	0.08	0.12	0.15	0.14	-0.16	0.07	-0.04	-0.17	-0.02	0.06	0.25	0.08	1.00	
15. TREATMENT INTENSITY	0.20	0.30	0.18	0.28	-0.67	0.90	0.00	-0.01	0.55	0.40	0.04	0.13	-0.21	-0.13	1.00

Note: Covariates with high correlations are used in different models (main model, propensity score matching, robustness etc.)