Supporting Online Materials For

# Advertising Effectively Influences Older Users:

## How Field Experiments Can Improve Measurement and Targeting

Randall A. Lewis and David H. Reiley

As noted in the body of the paper, we delivered a total of three campaigns for this retailer to the treatment group over a period of several months. The first campaign was the largest and, hence, provides the cleanest opportunity to estimate a statistically significant effect of the ads, so we concentrate on that campaign in this paper. The other two campaigns provide an opportunity for robustness checks; however, we expected them to have weaker effects due to delivering a much smaller number of impressions. Also, there was no re-randomization between campaigns, so to the extent that advertising may have persistent effects, the effects of Campaigns 2 and 3 will not be separately identified from the unknown long-run effects of Campaign 1. Table 1 summarizes the delivery of ad impressions in the three different campaigns.

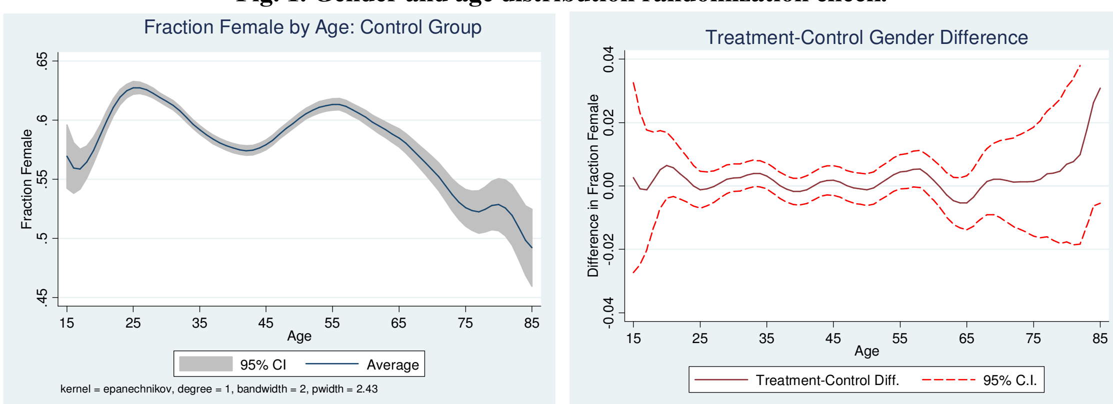
**Table 1. Summary statistics for the three campaigns.**

| | Campaign 1 | Campaign 2 | Campaign 3 | All 3 Campaigns |
|---|---|---|---|---|
| Time Period Covered | Early Fall '07 | Late Fall '07 | Winter '08 | |
| Length of Campaign | 14 days | 10 days | 10 days | |
| Number of Ads Displayed | 32,272,816 | 9,664,332 | 17,010,502 | 58,947,650 |
| Number of Users Shown Ads | 814,052 | 721,378 | 801,174 | 924,484 |
| % Treatment Group Viewing Ads | 63.7% | 56.5% | 62.7% | 72.3% |
| Mean Ad Views per Viewer | 39.6 | 13.4 | 21.2 | 63.8 |

The sales data include separate measures of offline and online purchases for each individual each week. Sales amounts include all purchases that the retailer could link to each individual customer in the database, using such information as the name on a shopper's credit-card during checkout at the store. To the extent that these customers sometimes make purchases that cannot be tracked by the retailer (for example, using cash and not identifying themselves), our estimate may underestimate the total effect of advertising on sales. However, the retailer believes that it accurately tracks at least 90% of purchases for these customers.

The text reports summary statistics that support valid randomization, especially along the dimensions of gender and age. Here, we investigate this issue in additional detail. In Fig. 1, we plot the fraction of the treatment and control groups that were female by age and computed the difference. Neither plot suggests any anomalous treatment-control differences. The left panel shows the fraction of the sample at each age group that was female for the control group. For customers around ages 25 and 55, there are disproportionately more women than men relative to other age groups. The right panel shows the difference between the treatment and control groups regarding the fraction of the sample that is female for each age group. The lack of statistical difference from zero indicates a valid randomization with respect to gender.

**Fig. 1. Gender and age distribution randomization check.**




Campaigns 2 and 3, which were shown to the treatment and control groups following the campaign, corroborate both the effect of the ads on sales and with respect to age. In terms of the average effect of the ads across all customers (Table 2 and Table 3), all three campaigns exhibited effects of the ads on sales of approximately 3% on both online and offline channels, with roughly 80-90% of the effect of the ads coming through the offline channel, in line with the offline sales volume of 84% of total sales for the control group. Thus, large retailers which do most of their business offline can reap benefits both online and offline from advertising online.

**Table 2. Ad effects in levels and sales percentages computed by linear regression.**

| | Effect of Online Ads on Sales* | | | | Ad Effect as % of Sales** | | |
|---|---|---|---|---|---|---|---|
| | Total | Offline | Online | % Offline | Total | Offline | Online |
| Campaign #1 | 0.061<br>(0.037) | 0.048<br>(0.035) | 0.014<br>(0.013) | 78% | 3.3% | 3.1% | 4.7% |
| Campaign #2 | 0.061<br>(0.044) | 0.052<br>(0.042) | 0.009<br>(0.013) | 85% | 3.0% | 3.0% | 2.8% |
| Campaign #3 | 0.029<br>(0.028) | 0.023<br>(0.026) | 0.006<br>(0.008) | 80% | 3.1% | 2.9% | 3.9% |
| All 3 Campaigns | 0.152<br>(0.069) | 0.123<br>(0.064) | 0.029<br>(0.022) | 81% | 3.2% | 3.0% | 3.8% |

* Each estimate was computed using a regression with sales as the dependent variable and a treatment grou indicator and online and offline sales from the three weeks preceding the first campaign as independent

** Sales levels correspond to the average purchase amount for the control group.

**Table 3. Ad effects in levels and sales percentages computed by simple treatment-control differences.**

| | Effect of Online Ads on Sales* | | | | Ad Effect as % of Sales** | | |
|---|---|---|---|---|---|---|---|
| | Total | Offline | Online | % Offline | Total | Offline | Online |
| Campaign #1 | 0.053<br>(0.038) | 0.046<br>(0.035) | 0.007<br>(0.013) | 87% | 2.9% | 3.0% | 2.3% |
| Campaign #2 | 0.054<br>(0.044) | 0.050<br>(0.042) | 0.004<br>(0.013) | 93% | 2.7% | 2.9% | 1.2% |
| Campaign #3 | 0.028<br>(0.028) | 0.023<br>(0.026) | 0.005<br>(0.008) | 82% | 2.9% | 2.9% | 3.3% |
| All 3 Campaigns | 0.134<br>(0.070) | 0.119<br>(0.064) | 0.016<br>(0.023) | 88% | 2.8% | 2.9% | 2.1% |

* Each estimate is the difference between the treatment and control group average sales for each category of sales for each campaign.

** Sales levels correspond to the average purchase amount for the control group.

Following the discovery that the advertising seemed to affect women more and the offline channel more, we examined the heterogeneous differences in purchasing behavior for the three weeks prior to the campaign (Fig. 2) and for each of the three campaigns separately (Fig 3, and Fig 4, and Fig 5).

We now examine the robustness of the result that the elderly purchase more in response to the ads for campaigns 2 and 3. At first glance the pre-experiment sales results (Fig. 2) appear to partially explain the main results of the paper as the elderly in the treatment group have slightly higher sales than the control group. However, we would like to highlight the fact that less than 5% of the customers purchase in any given week and there are few repeat purchasers. In addition, while the treatment group appears to purchase more than the control for the elderly around age 80, the location of this differential appears to occur at slightly different ages than it does during campaign 1. Thus, we conclude that the statistical variation prior to the experiment does not explain our results which rely on purchases by different people during subsequent weeks.

An examination of the variation in the time dimension allows us to further test the robustness of the results. We compute an experimental difference-in-differences (DID) estimate by subtracting the pre-experiment average weekly sales from the average weekly sales during the campaign for each individual and then comparing these averages across the treatment and control groups. In our experimental DID, we find very similar results to our estimates presented in the main text (Fig 6). We have relegated these results to this appendix to simplify the exposition by avoiding descriptions of the pre-experimental sales.

Upon completing the in-depth decomposition of the results for all three campaigns, we discovered several other marginally significant regions among the estimates for Campaigns 2 and 3. However, we hesitate to rush to any conclusions, due to the risk of committing multiple type I errors. We specifically demand a greater level of significance from our primary results to avoid any spurious conclusions arising from multiple-hypothesis testing problems.

**Fig. 2. Nonparametric plots of sales versus age for the three weeks preceding the experiment.** The average purchasing behavior for each age group validates the randomization. The graphs are oriented in a grid with the three columns representing age group purchases for females, males, and both males and females, respectively, and the three rows representing online, offline, and combined sales. The dark lines are local differences in averages computed by locally linear regression using an Epanechnikov kernel with bandwidth of four years. The dashed lines above and below the difference in averages correspond to asymptotic 95% pointwise confidence intervals.

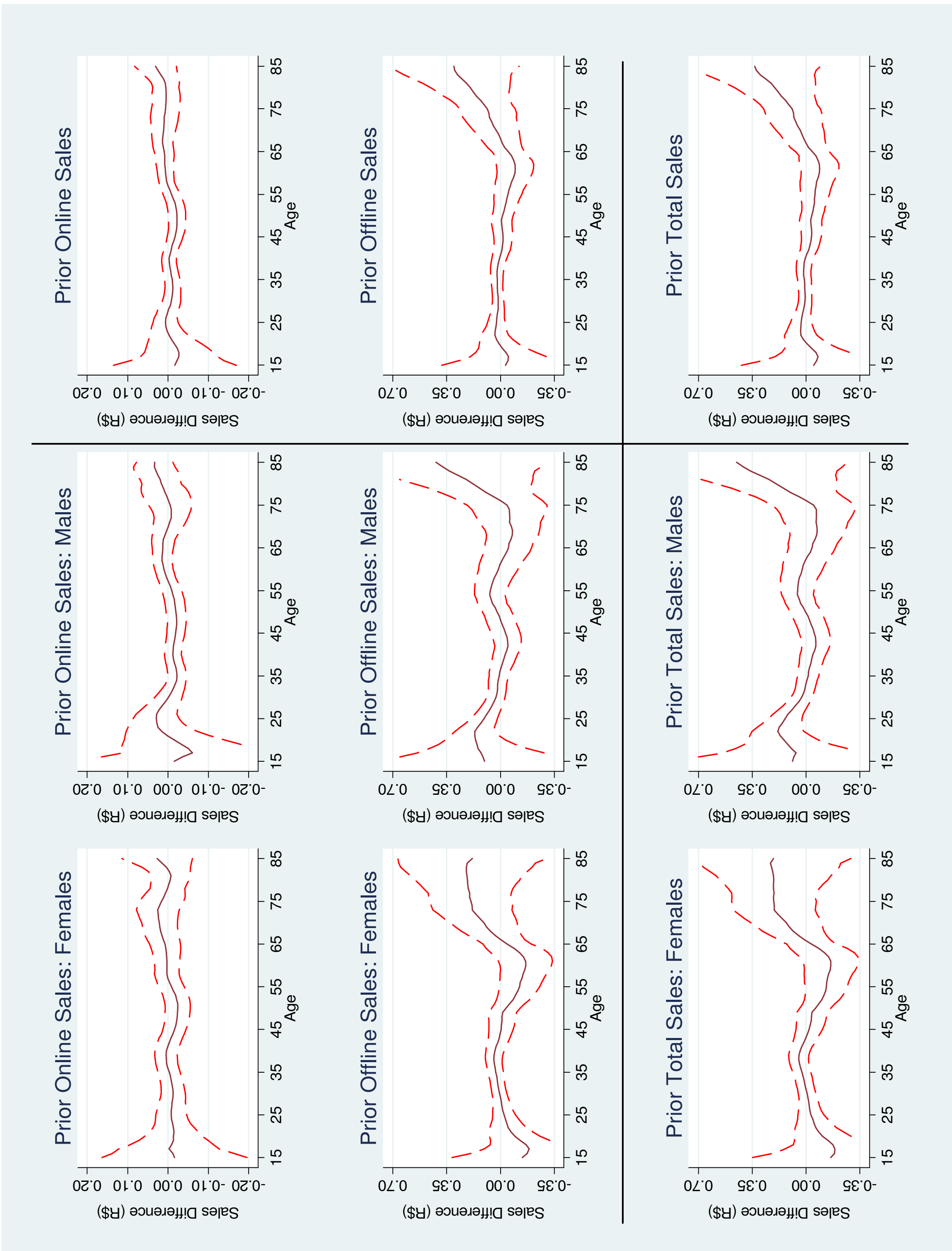


**Fig. 3. Nonparametric plots of Sales versus age for the two weeks during campaign 1.** The average purchasing behavior for each age group shows an effect most pronounced among older women. The graphs are oriented in a grid with the three columns representing age group purchases for females, males, and both males and females, respectively, and the three rows representing online, offline, and combined sales. The dark lines are local differences in averages computed by locally linear regression using an Epanechnikov kernel with bandwidth of four years. The dashed lines above and below the difference in averages correspond to asymptotic 95% pointwise confidence intervals.

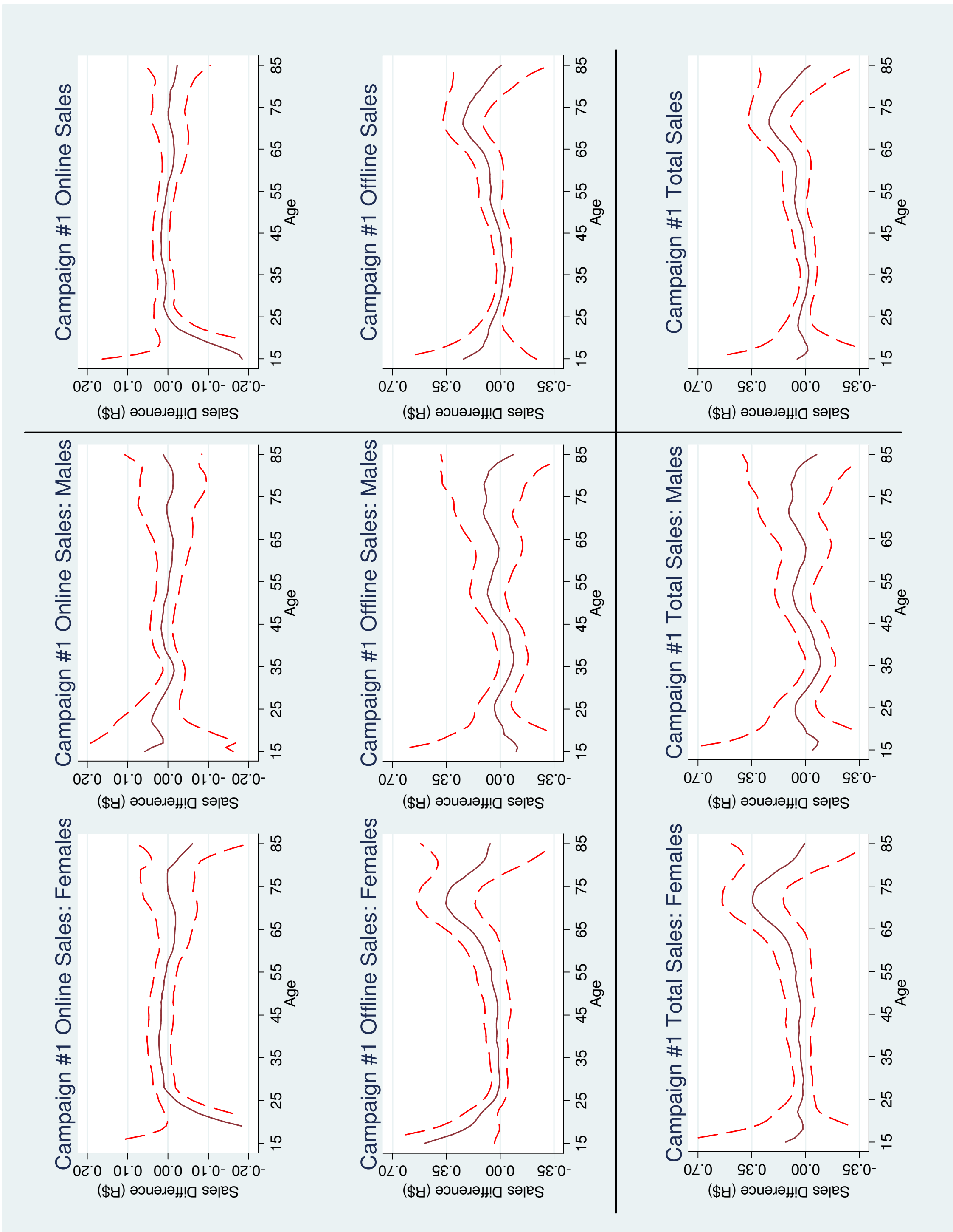


**Fig. 4. Nonparametric plots of sales versus age for the ten days during campaign 2.** The average purchasing behavior for each age group shows a weak effect that is most pronounced among older women. The graphs are oriented in a grid with the three columns representing age group purchases for females, males, and both males and females, respectively, and the three rows representing online, offline, and combined sales. The dark lines are local differences in averages computed by locally linear regression using an Epanechnikov kernel with bandwidth of four years. The dashed lines above and below the difference in averages correspond to asymptotic 95% pointwise confidence intervals.

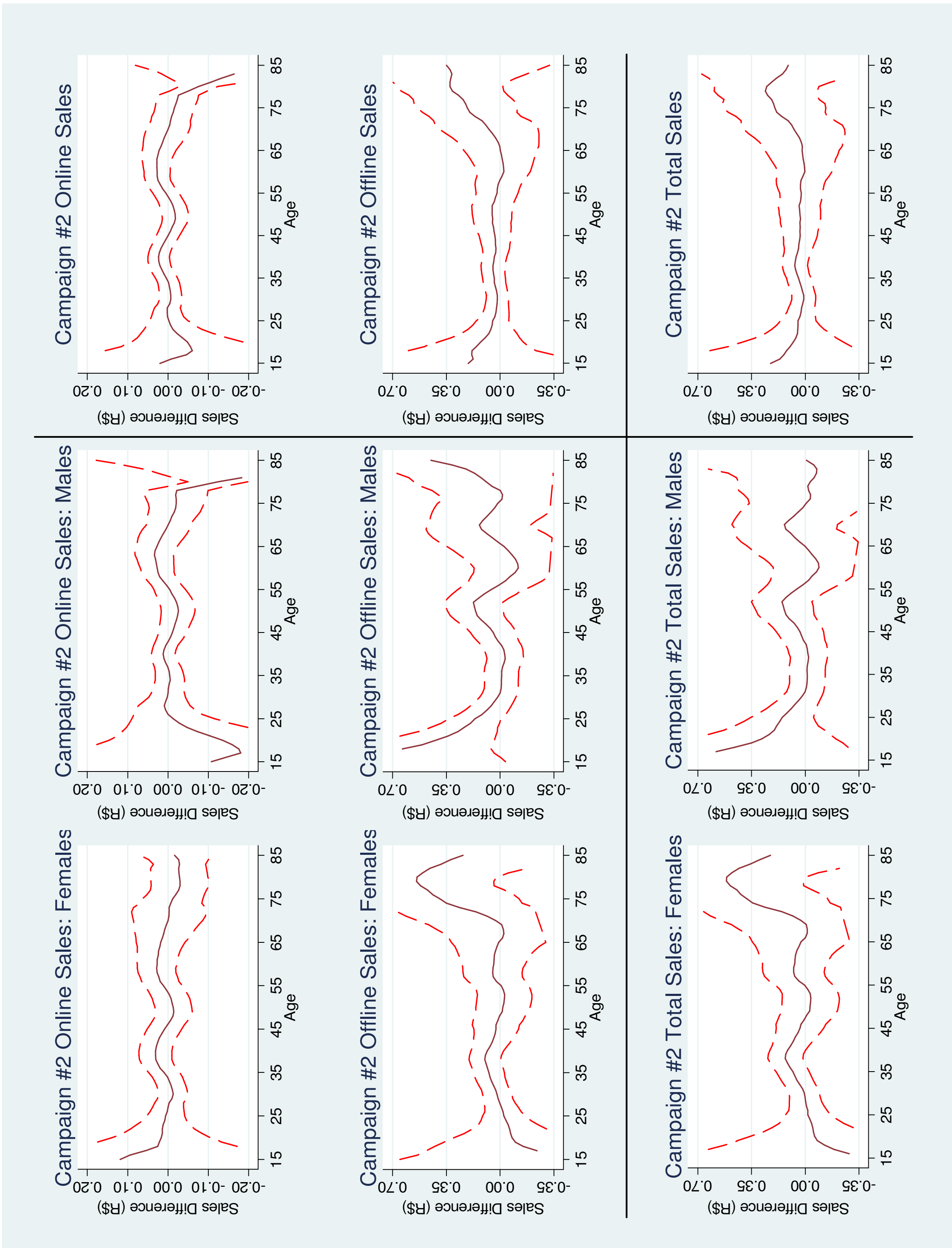


**Fig. 5. Nonparametric plots of sales versus age for the ten days during campaign 3.** The average purchasing behavior for each age group shows a weak effect that is most pronounced among older women. The graphs are oriented in a grid with the three columns representing age group purchases for females, males, and both males and females, respectively, and the three rows representing online, offline, and combined sales. The dark lines are local differences in averages computed by locally linear regression using an Epanechnikov kernel with bandwidth of four years. The dashed lines above and below the difference in averages correspond to asymptotic 95% pointwise confidence intervals.

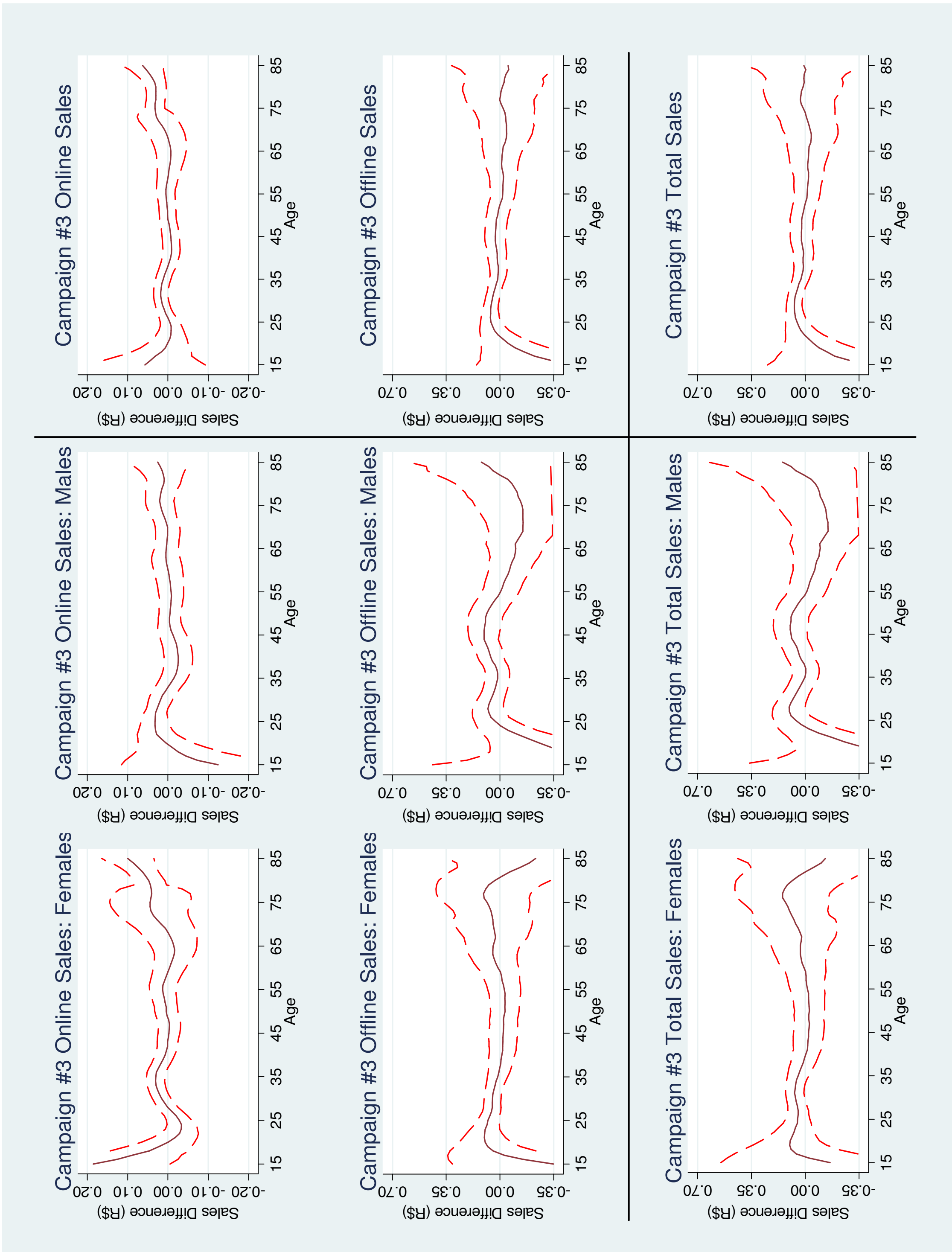


**Fig. 6. Nonparametric plots of difference between pre-campaign and campaign 1 weekly sales versus age.** The average purchasing behavior for each age group shows a strong effect that is most pronounced among older women. The graphs are oriented in a grid with the three columns representing age group purchases for females, males, and both males and females, respectively, and the three rows representing online, offline, and combined sales. The dark lines are local differences in averages computed by locally linear regression using an Epanechnikov kernel with bandwidth of four years. The dashed lines above and below the difference in averages correspond to asymptotic 95% pointwise confidence intervals.