Retail Advertising Works!

Measuring the Effects of Advertising on Sales via a Controlled Experiment on Yahoo!

Randall Lewis
MIT and
Yahoo! Research

David Reiley
Yahoo! Research and
University of Arizona

FTC & NU Microeconomics Conference
November 2008

Outline of Presentation

I. Why Is It Hard to Measure Effects of Ads?
II. Designing an Experiment
III. Our Data on Sales and Advertising
IV. Basic Treatment Effects on Campaign #1
V. Persistence over Time
VI. Detailed Results on Campaign #1
VII. Conclusion

Advertising's effects on sales have always been very difficult to measure.

"Half the money I spend on advertising is wasted; the trouble is I don't know which half."

-John Wanamaker
(Department store merchant, 1838-1922)

Advertisers do not have good measures of the effects of brand image advertising.

- *Harvard Business Review* article by the founder and president of ComScore (Abraham, 2008) illustrates the state of the art for practitioners:
  - Compares those who saw an online ad with those who didn't.
  - Potential problem: the two samples do not come from the same population.
  - Example: Who sees an ad for eTrade on Google?
    - Those who search for "online brokerage" and similar keywords.
    - Does the ad actually cause the difference in sales?
  - Correlation is not the same as causality.

Measuring the effects of advertising on sales has been difficult for economists as well as practitioners.

- The classic technique: econometric regressions of aggregate sales versus advertising.
  - Practitioners call this Marketing Mix Modeling.
  - A textbook example of the “endogeneity” problem in econometrics (see Berndt, 1991).
  - But what causes advertising to vary over time?
  - Many studies flawed in this way.

We have just seen two ways for observational data to provide inaccurate results.

- Aggregate time-series data
  - Advertising doesn't vary systematically over time.
- Individual cross-sectional data
  - The types of people who see ads aren't the same population as those who don't see ads.
  - Even in the absence of any ads, they might well have different shopping behavior.

- When existing data don't give a valid answer to our question of interest, we should consider generating our own data.

An experiment is the best way to establish a causal relationship.

- Systematically vary the amount of advertising: show ads to some consumers but not others.
- Measure the difference in sales between the two groups of consumers.
- Like a clinical trial for a new pharmaceutical.
- Almost never done in advertising, either in online or traditional media.
  - Exceptions: direct mail, search advertising.

Our understanding of advertising today resembles our understanding of physics in the 1500s.

- Do heavy bodies fall at faster rates than light ones?
- Galileo's key insight: use the experimental method.
- Huge advance over mere introspection or observation.

Marketers often measure effects of advertising using experiments…

- … but not with actual transaction data.
- Typical measurements come from questionnaires:
  - “Do you remember seeing this commercial?”
  - “What brand comes to mind first when you think about batteries?”
  - “How positively do you feel about this brand?”
- Useful for comparing two different “creatives.”
- But do these measurements translate into actual effects of advertising on sales?

A few previous experiments measured the effects of advertising on sales.

- Experiments with IRI BehaviorScan (split-cable TV)
  - Hundreds of individual tests reported in several papers:
    - Abraham and Lodish (1995)
    - Lodish et al. (1995a,b)
    - Hu, Lodish, and Krieger (2007)
  - Sample size: 3,000 households.
  - Hard to find statistically significant effects.
- Experiments with direct-mail catalog frequency
  - Anderson and Simester (2008)
  - Sample size: 20,000 households.
  - Increased mailings produce higher short-run sales, but the effects are partially offset by reductions in long-run sales.

- Our experiment will study 1.6 million individuals.

Some studies derive valid insights from non-experimental panel data.

- Observing a panel of individuals over time can help solve the problem of individual heterogeneity.
  - Monitor how individuals' purchase behavior changes over time, as advertising changes.
  - Observe not just the level of sales across individuals, but also the changes in sales over time across individuals.
- Examples: Ackerberg (2001, 2003)
  - Individual diaries of TV ads viewed.
  - Sample of 2,000 households.
  - Evidence that new-product advertising has informative effects: more impact on those who never before purchased.

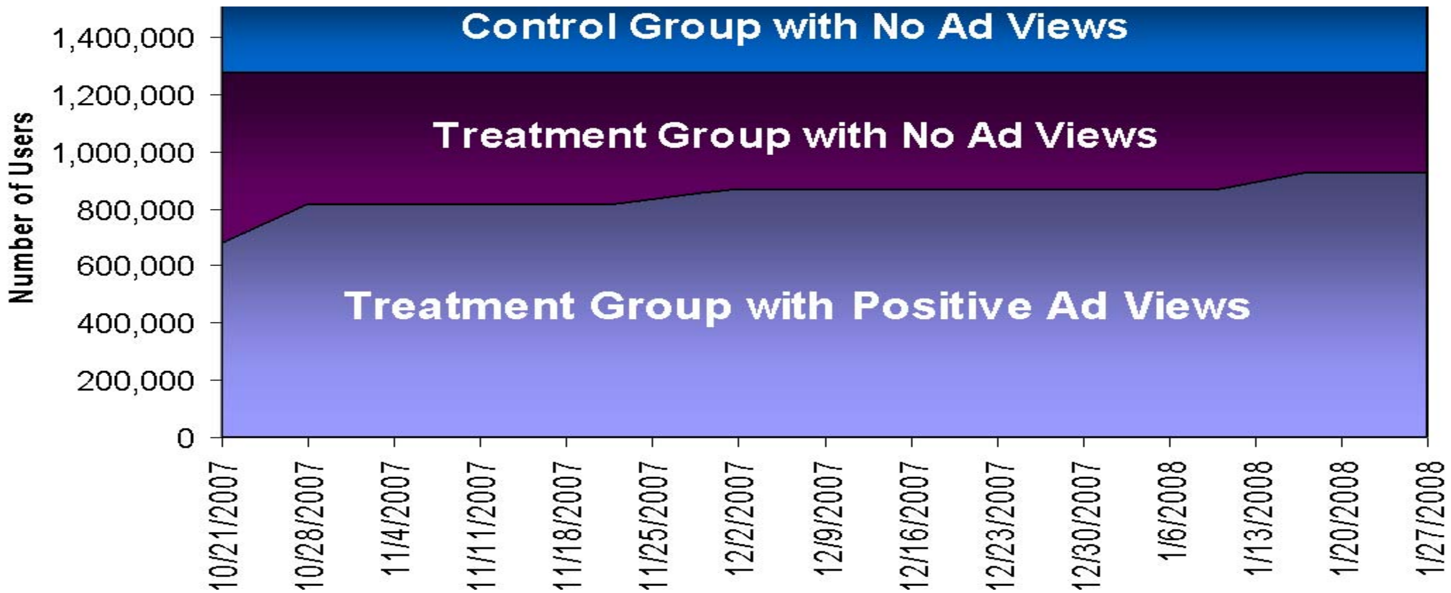
Our study will combine a large-scale experiment with individual panel data.

- We match Yahoo! ID database with nationwide retailer's customer databases
  - 1,577,256 customers matched
- 80% of matched customers assigned to the treatment group
  - Allowed to view 3 ad campaigns on Yahoo! from the retailer
- Remaining 20% assigned to the control group
  - Do not see ads from the retailer
- Ad campaigns are "Run of Yahoo! network" ads
- Following the online ad campaigns, we received both online and in-store sales data: for each week, for each person
  - Third party de-identifies observations to protect customer identities
  - Retailer multiplied all sales amounts by a scalar factor

Ads were shown across the Yahoo! network, similar to this Netflix ad.

By the end of the three campaigns, over 900,000 people had seen ads.

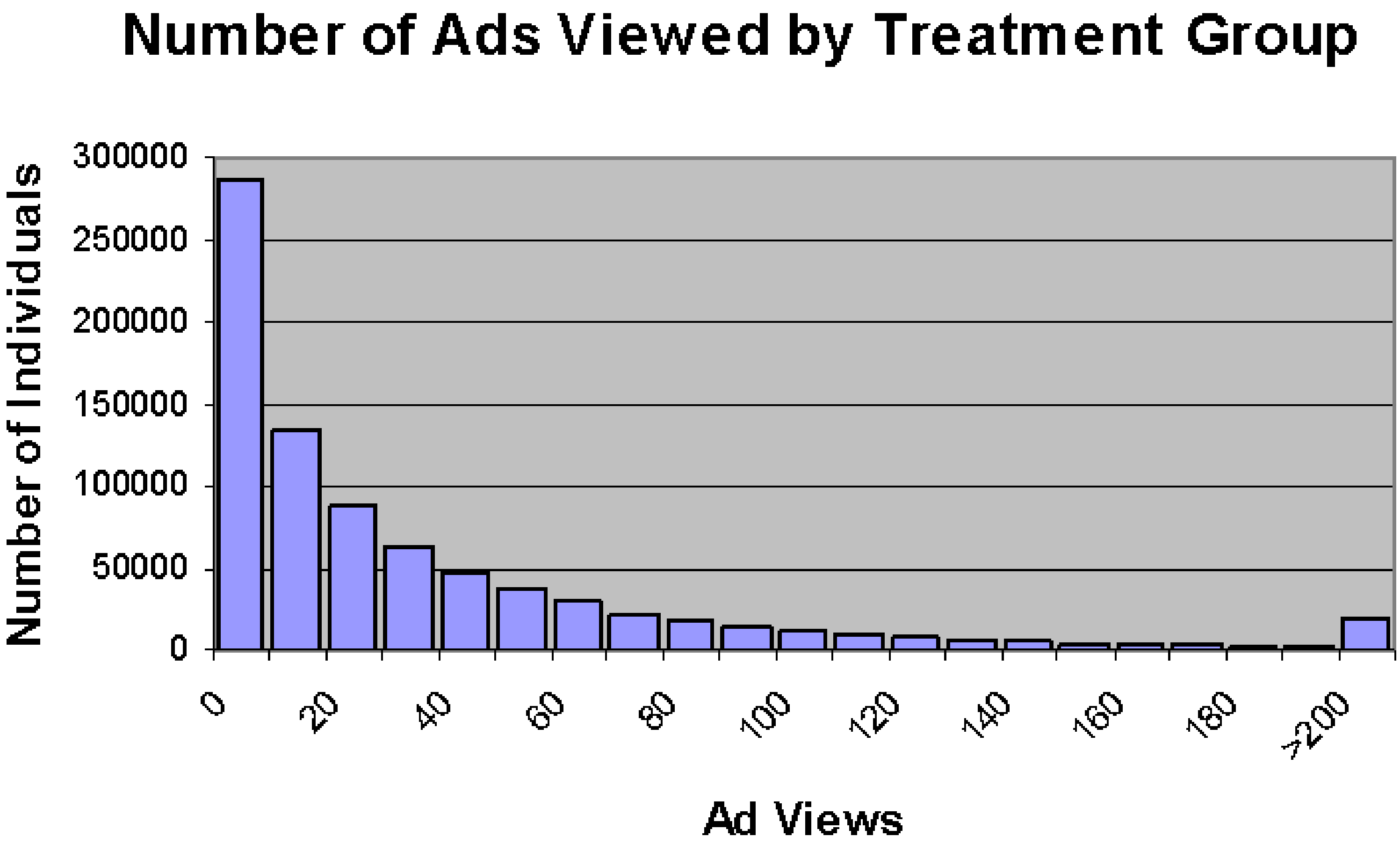
	Campaign 1	Campaign 2	Campaign 3	All 3 Campaigns
Time Period Covered	Early Fall '07	Late Fall '07	Winter '08	
Length of Campaign	14 days	10 days	10 days	
Number of Ads Displayed	32,272,816	9,664,332	17,010,502	58,947,650
Number of Users Shown Ads	814,052	721,378	801,174	924,484
% Treatment Group Viewing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



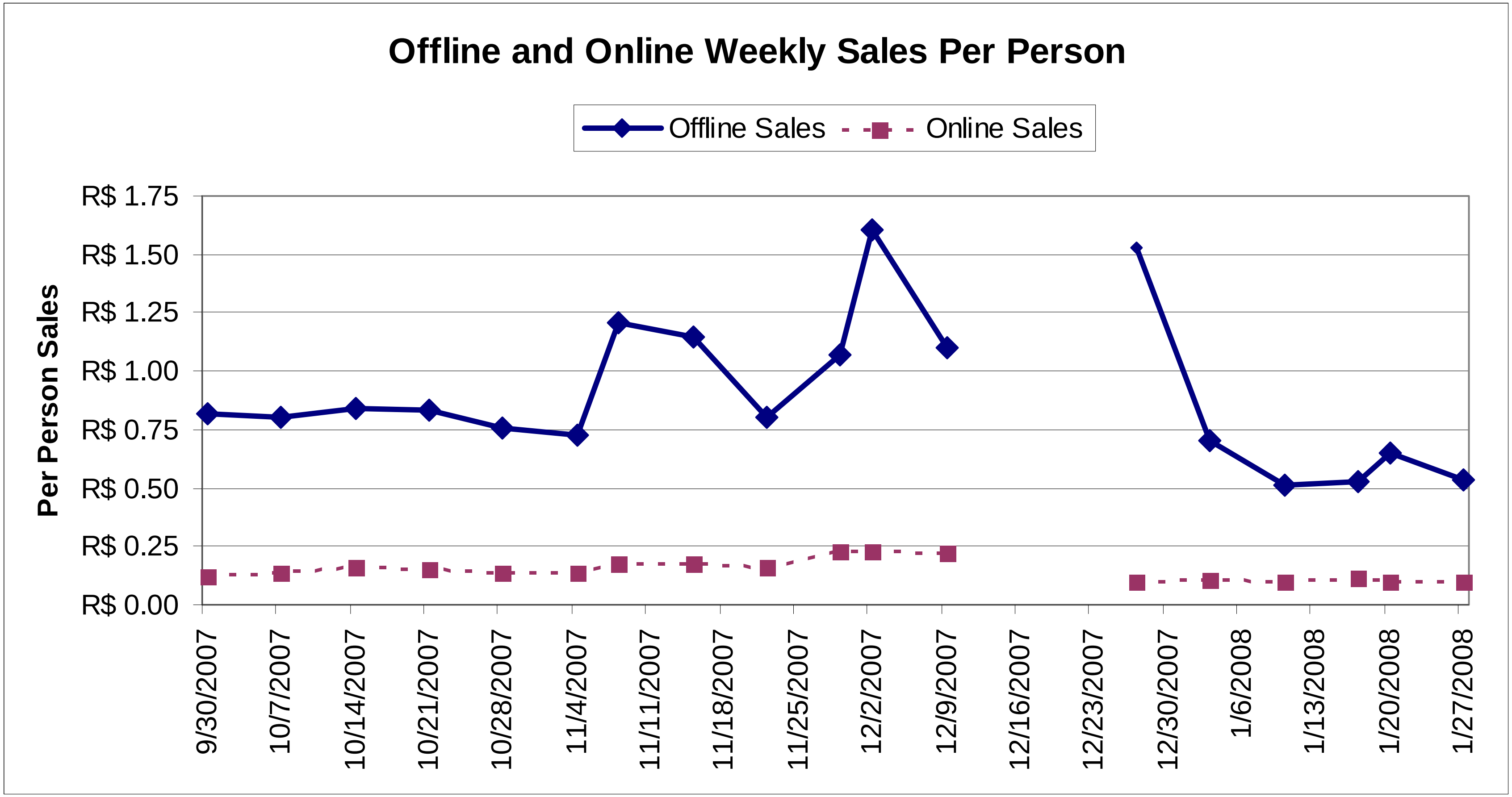
Descriptive statistics for Campaign #1 indicate valid treatment-control randomization.

	Control	Treatment
% Female	59.5%	59.7%
% Retailer Ad Views > 0	0.0%	63.7%
% Yahoo Page Views > 0	76.4%	76.4%
Mean Y! Page Views per Person	358	363
Mean Ad Views per Person	0	25
Mean Ad Clicks per Person	0	0.056
% Ad Impressions Clicked (CTR)	-	0.28%
% People Clicking at Least Once	-	4.59%

We see a skewed distribution of ad views across individuals.



In-store sales are more than five times as large as online sales, and have high variance across weeks.



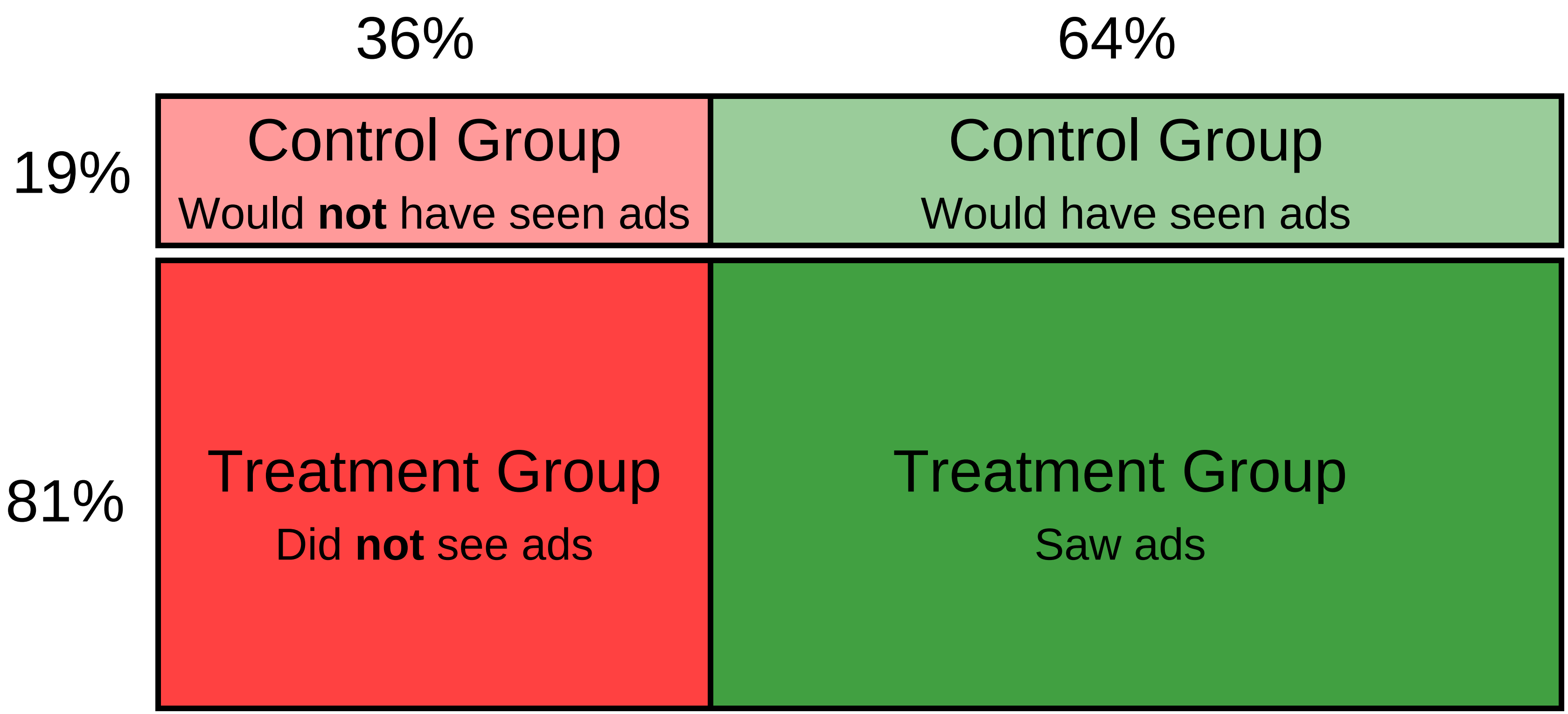
Sales vary widely across weeks and include many individual outliers.

		Mean Sales	Std. Dev.	Min	Max	Transactions
Campaign #1						
09/24	3 Weeks Before	**R$ 0.939**	14.1	-932.04	4156.01	42,809
10/01	2 Weeks Before	**R$ 0.937**	14.1	-1380.97	3732.03	41,635
10/08	1 Week Before	**R$ 0.999**	14.3	-1332.04	3379.61	43,769
10/15	Week 1 During	**R$ 0.987**	13.5	-2330.10	2163.11	43,956
10/22	Week 2 During	**R$ 0.898**	13.3	-1520.39	2796.12	40,971
10/29	Week 1 Following	**R$ 0.861**	13.3	-1097.96	3516.51	40,152
Campaign #2						
11/02	3 Weeks Before	**R$ 1.386**	16.4	-1574.95	3217.30	52,776
11/09	2 Weeks Before	**R$ 1.327**	16.6	-654.70	5433.00	57,192
11/16	1 Week Before	**R$ 0.956**	13.4	-2349.61	2506.57	45,359
11/23	Week 1 During	**R$ 1.299**	16.7	-1077.83	3671.75	53,428
11/30	Week 2 During (3 Days)	**R$ 0.784**	14.0	-849.51	3669.13	29,927
12/03	Week 1 Following	**R$ 1.317**	16.1	-2670.87	5273.86	57,522
Campaign #3						
12/21	3 Weeks Before	**R$ 1.635**	17.9	-2051.39	2521.88	62,454
12/28	2 Weeks Before	**R$ 0.812**	13.0	-1238.83	1870.99	49,144
01/04	1 Week Before	**R$ 0.616**	11.7	-1120.77	3400.54	38,265
01/11	Week 1 During	**R$ 0.644**	11.7	-1118.58	3939.81	36,321
01/18	Week 2 During (3 Days)	**R$ 0.322**	7.5	-588.84	1437.17	18,238
01/21	Week 1 Following	**R$ 0.636**	11.5	-2336.83	3300.97	33,724

N=1,577,256 observations per week

Not all of the treatment-group members browsed Yahoo! enough to see the retailer's ads.

- Only 64% of the treatment group browsed enough to see at least one ad in Campaign #1. Our estimated effects will be "diluted" by 36%.
- We expect similar browsing patters in the control group, but cannot observe which control-group members would not have seen ads.



Descriptive statistics show a positive increase in sales due to ads.

	During Campaign (2 weeks) **Mean Sales/Person**
Control:	**R$ 1.84** (0.03)
Treatment:	**1.89** (0.02)

- But the effect is not statistically significant.
- One reason is the 36% dilution of the treatment group.

Suppose we had no experiment, and just compared spending by those who did or did not see ads.

	During Campaign (2 weeks) Mean Sales/Person
Control:	R$ 1.84 (0.03)
Treatment:	1.89 (0.02)
Exposed to Retailer’s Ads: [64% of Treatment Group]	**1.81** (0.02)
Not Exposed to Retailer’s Ads: [36% of Treatment Group]	**2.04** (0.03)

- We would conclude that ads decrease sales by R$0.23!
- But this would be a mistake, because here we’re not comparing apples to apples.

Pre-campaign data shows us that the non-experimental sales differences have nothing to do with ad exposures.

	Before Campaign (2 weeks) Mean Sales/Person		During Campaign (2 weeks) Mean Sales/Person
Control:			R$ 1.84 (0.03)
Treatment:			1.89 (0.02)
Exposed to Retailer's Ads: [64% of Treatment Group]	**1.81** (0.02)	→	1.81 (0.02)
Not Exposed to Retailer's Ads: [36% of Treatment Group]	**2.15** (0.03)	→	2.04 (0.03)

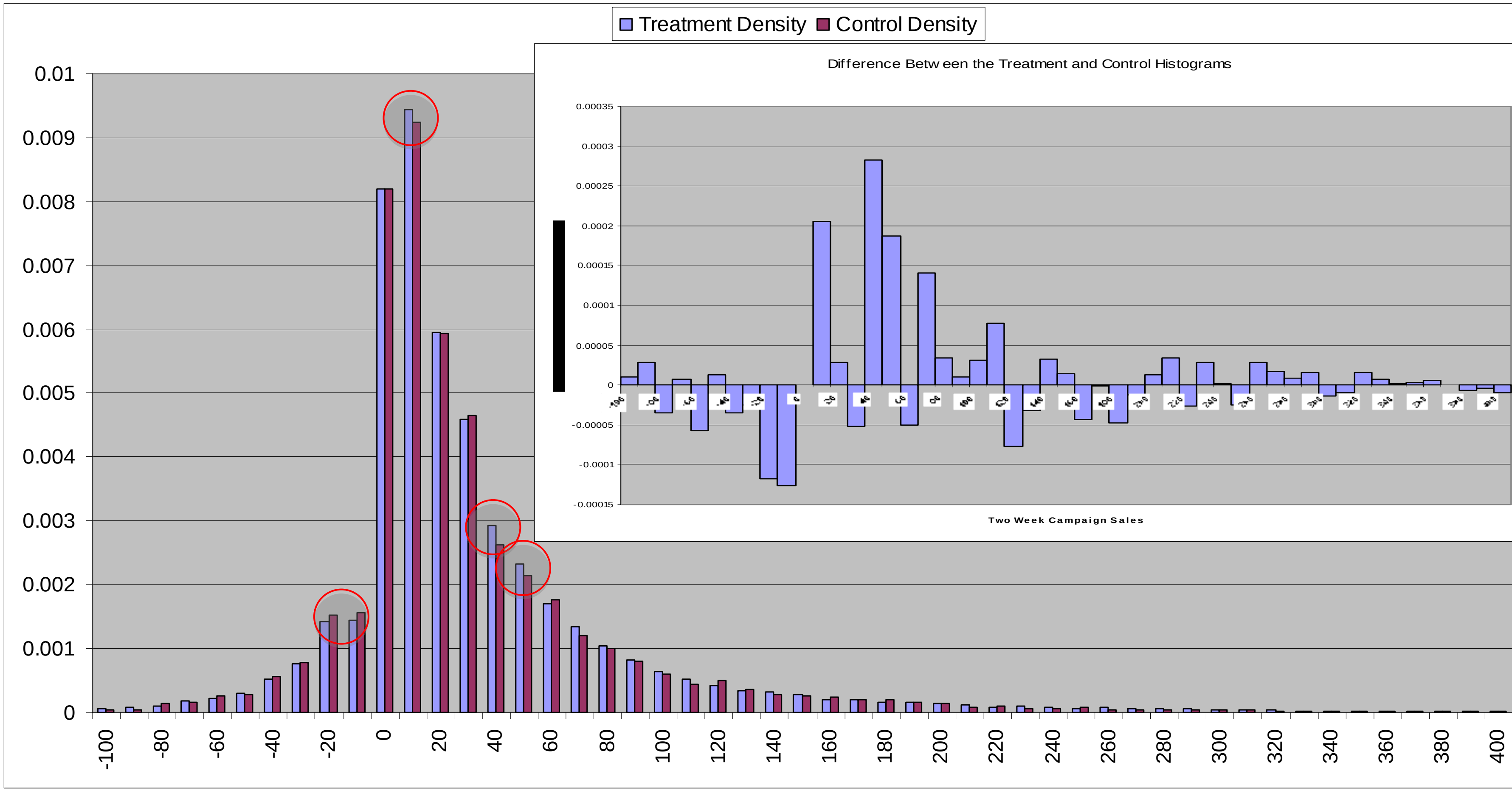
- People who browse enough to see ads also have a lower baseline propensity to purchase from the retailer.
- Potential mistake solved with experiment, panel data.

Ad exposures appear to have prevented a normal decline in sales during this time period.

	Before Campaign (2 weeks) Mean Sales/Person		During Campaign (2 weeks) Mean Sales/Person	Difference (During – Before) Mean Sales/Person
Control:	R$ 1.95 (0.04)	→	R$ 1.84 (0.03)	**-R$ 0.10** (0.05)
Treatment:	1.93 (0.02)	→	1.89 (0.02)	
Exposed to Retailer's Ads: [64% of Treatment Group]	1.81 (0.02)	→	1.81 (0.02)	**R$ 0.00** (0.03)
Not Exposed to Retailer's Ads: [36% of Treatment Group]	2.15 (0.03)	→	2.04 (0.03)	**-R$ 0.10** (0.04)

- Control-group sales fall.
- Unexposed treatment-group sales fall.
- Treated-group sales stay constant.

Instead of just means, let’s look at the treatment effect on the *distribution* of purchase amounts.



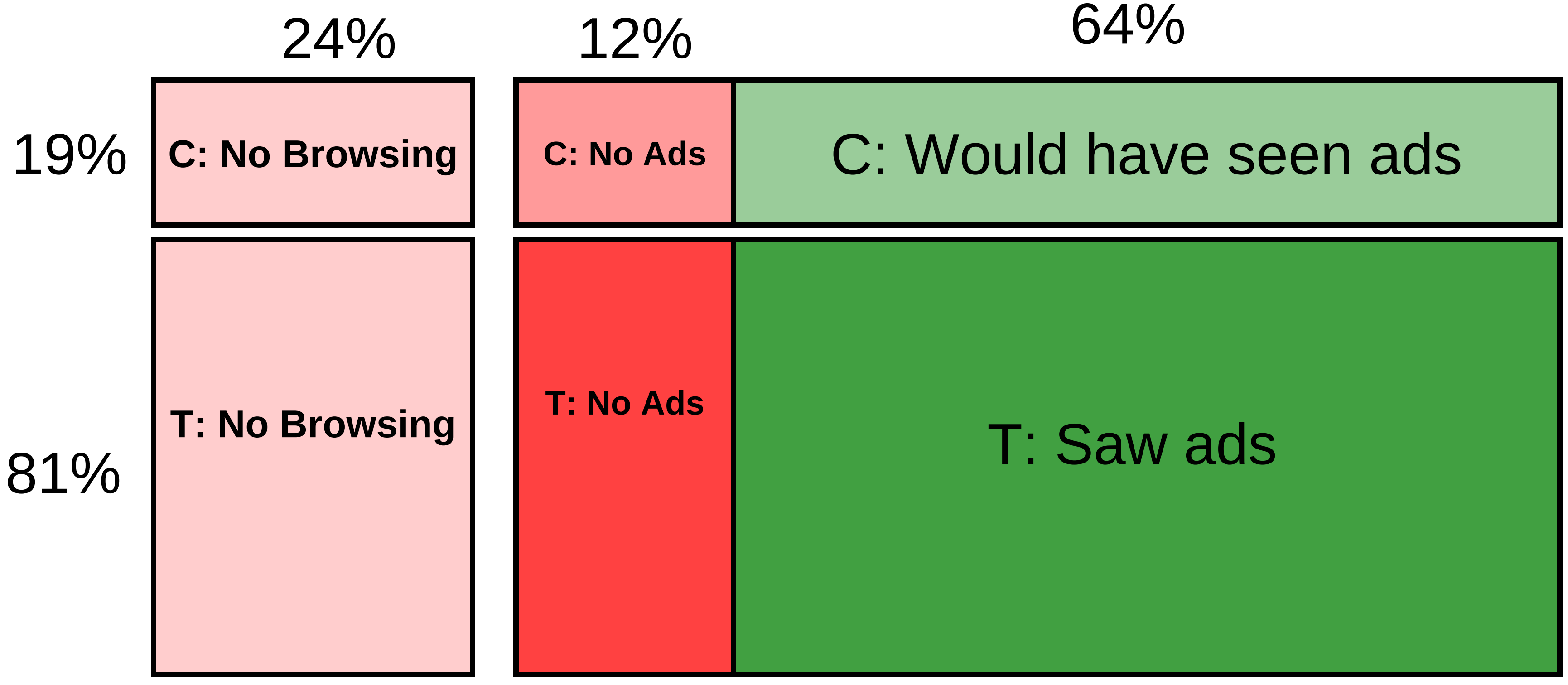
*Purchase amounts of zero not displayed.

We use three different methods to estimate the effect of advertising on those who see ads.

- Compare sales between treatment and control.
  - We can't observe who are the 36% of people who would not have seen ads.
  - We correct for 36% dilution in measurement.
- Repeat the above, but exclude those 24% of individuals with zero Yahoo! page views during the campaign.
  - We *can* observe who are the 24% of people who did not browse the Yahoo! network at all.
  - Again correct for dilution in measurement (17%).
- Difference in difference: compare before/after purchase amounts between treated and untreated individuals.

We have three different groups of individuals to consider.

- We can't completely separate green from red, so we have noise in our estimates.
- We can eliminate the 24% who don't browse Yahoo! at all.
  - But the data are imperfectly matched.



The first two techniques look only at sales during the two-week campaign.

- Recall that for the treatment group:
  - 24% did not browse Yahoo! at all.
  - 12% browsed Yahoo!, but not enough to see these ads.
  - 64% saw these ads.
- Simple difference: Compare treatment minus control.
- Rescaling: Divide by 0.64 or 0.83 to compute the effect of the treatment on the treated.
  - Rescale both the estimate and the standard error.

	Treatment-Control	Excluding Page Views=0
Simple Difference	R$ 0.053	R$ 0.078
	(0.038)	(0.045)
Rescaled: Effect on Treated	**0.083**	**0.093**
	(0.059)	(0.054)

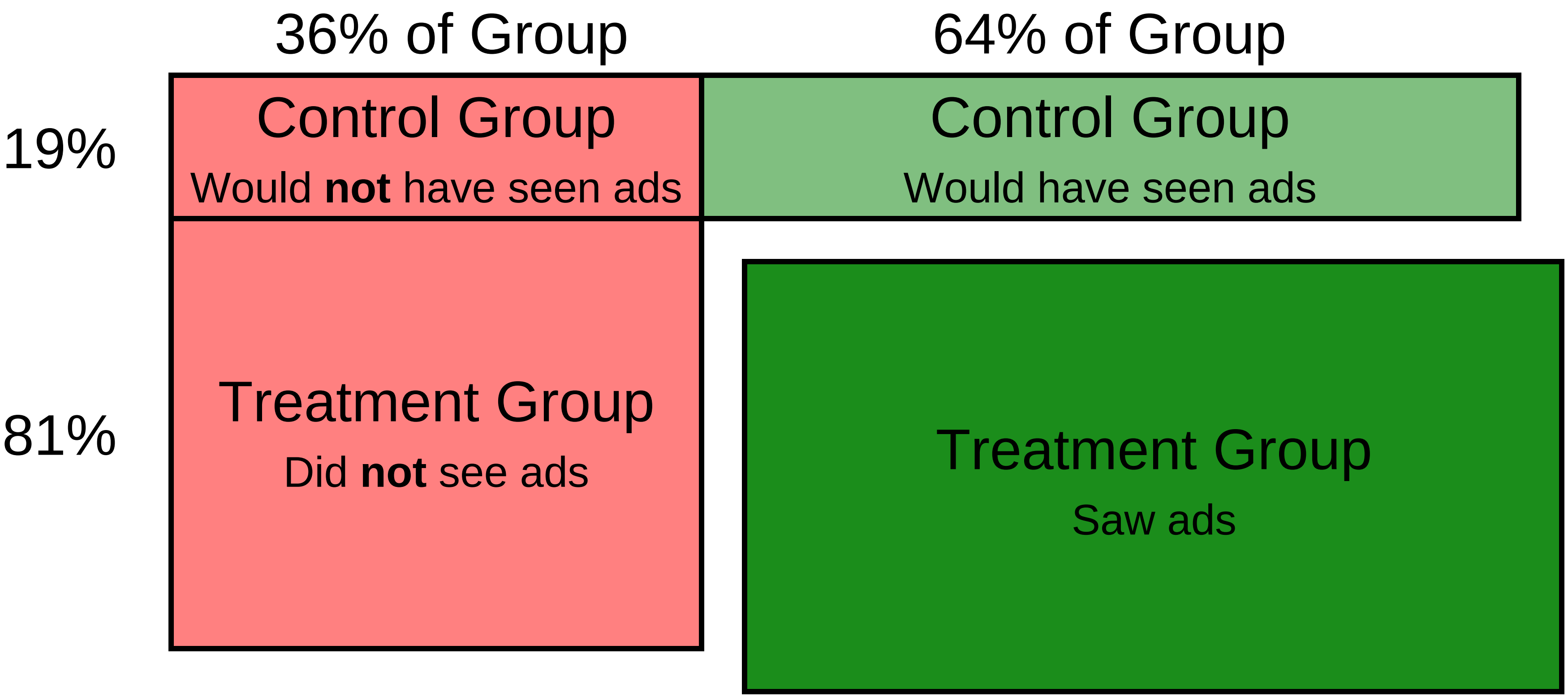
Our third technique uses the data's panel structure to look at pre-post differences in sales.

- We wish to control for unobserved heterogeneity in shopping, which is correlated with Yahoo! browsing behavior.
  - Assume these differences are constant over time.
- We do so by looking at pre-post differences in sales for individuals.
- Now we pool together the control group with the no-ads part of the treatment group, and compare to those treated with ads.

DID controls the group and individual heterogeneity across time.

$$Sales_{i,t} = \gamma_t SawAds_{i,t} + \beta_t + \alpha_i + \varepsilon_{i,t}$$

$$\Delta Sales_i = \gamma_t SawAds_{i,post} + \Delta\beta + \Delta\varepsilon_i$$



Our difference-in-difference estimate yields a statistically and economically significant treatment effect.

- Estimated effect per customer of viewing ads:
  - **Mean = R$ .102, SE = R$ .043**
- Estimated sales impact for the retailer:
  - **R$83,000 ± 70,000**
    - 95% confidence interval.
    - Based on 814,052 treated individuals.
    - Compare with cost of about **R$20,000**.

Our difference-in-difference model passes a specification test.

- To use DID, we assume that the heterogeneity of the two groups doesn't change over time in a way that could be correlated with changes in advertising.
- This allows us to pool together the control group with the untreated (no ads) portion of the treatment group.
- To test this assumption, we test the hypothesis that the control group and the untreated portion of the treatment group have the same before-after difference in sales.
  - The difference between these two means is R$0.001 (p=0.988).
  - Thus, we cannot reject the hypothesis that our DID model is correctly specified.

A reminder shows us why we feel comfortable pooling the two groups.

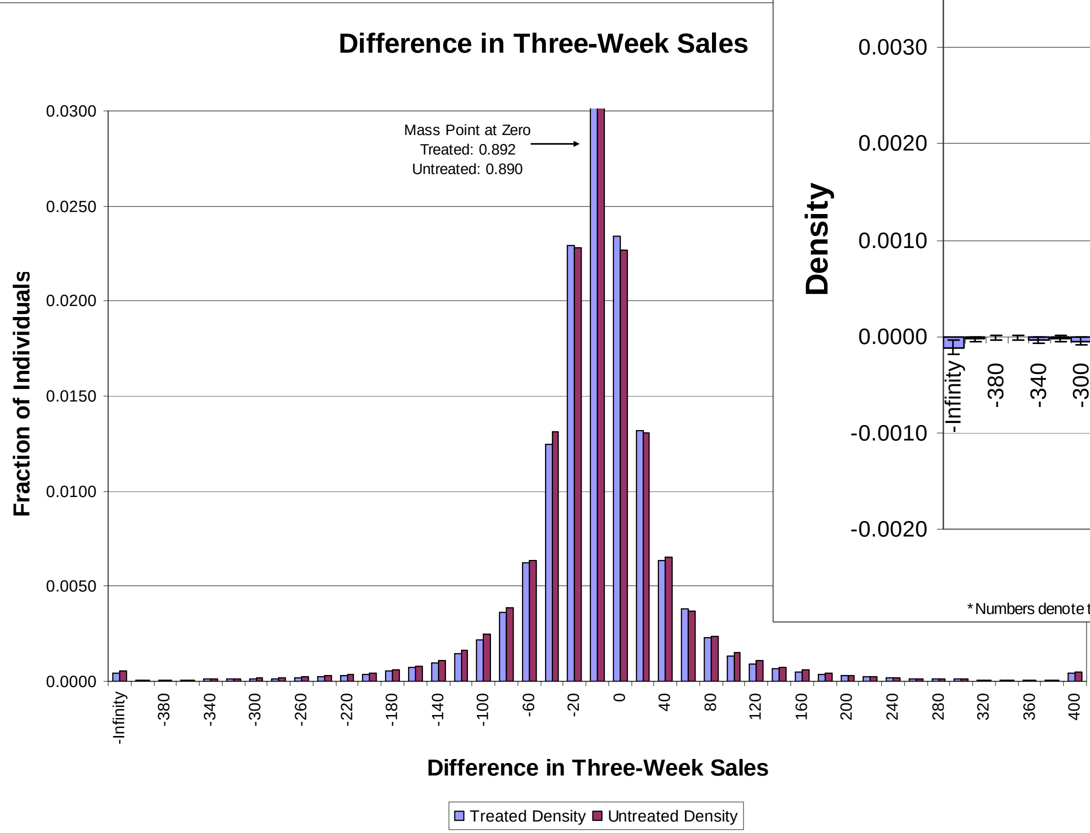
	Before Campaign (2 weeks) Mean Sales/Person		During Campaign (2 weeks) Mean Sales/Person	Difference (During – Before) Mean Sales/Person
Control:	R$ 1.95 (0.04)	→	R$ 1.84 (0.03)	**-R$ 0.10** (0.05)
Treatment:	1.93 (0.02)	→	1.89 (0.02)	
Exposed to Retailer's Ads: [64% of Treatment Group]	1.81 (0.02)	→	1.81 (0.02)	**R$ 0.00** (0.03)
Not Exposed to Retailer's Ads: [36% of Treatment Group]	2.15 (0.03)	→	2.04 (0.03)	**-R$ 0.10** (0.04)

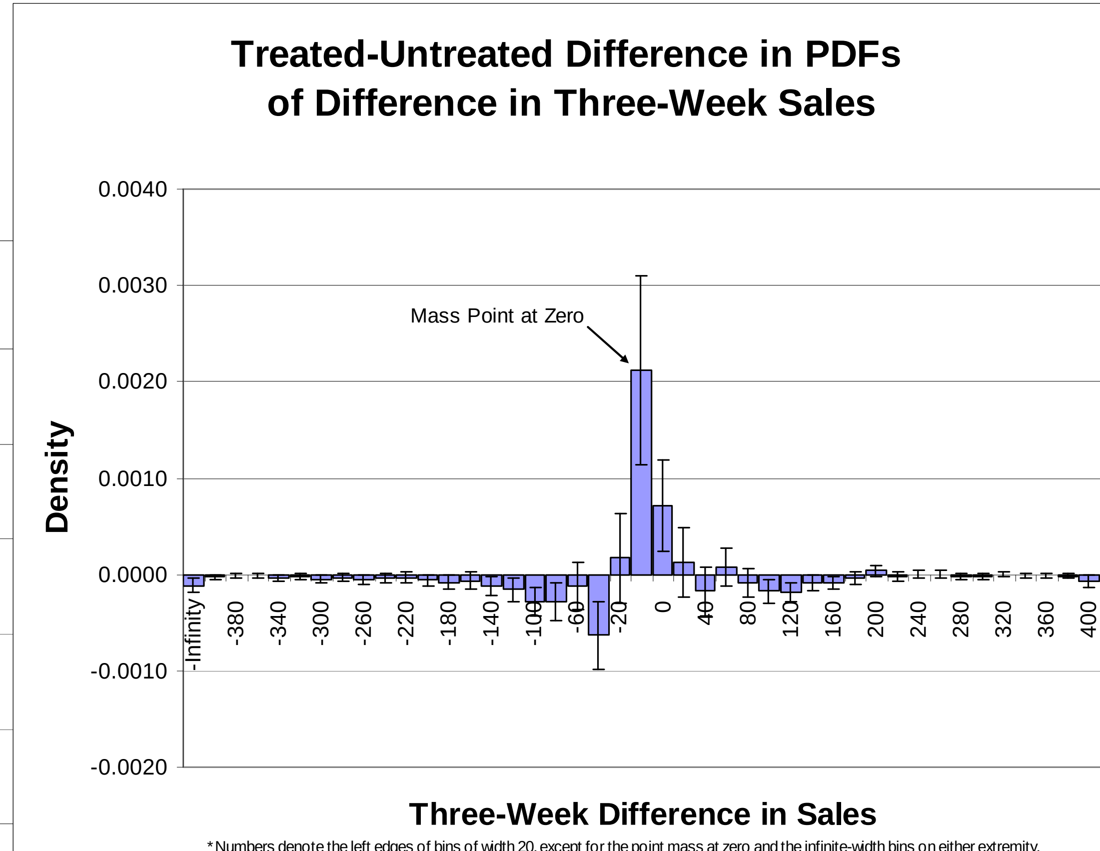
What happens after the two-week campaign is over?

- Positive effects during the campaign could be followed by:
  - Negative effects (intertemporal substitution)
  - Equal sales (short-lived effect of advertising)
  - Higher sales (persistence beyond the campaign)
- We can distinguish between these hypotheses by looking at the week following the two weeks of the campaign.

Pre-post differences in sales show positive effects for treated versus untreated individuals.

Three-week Sales Differences



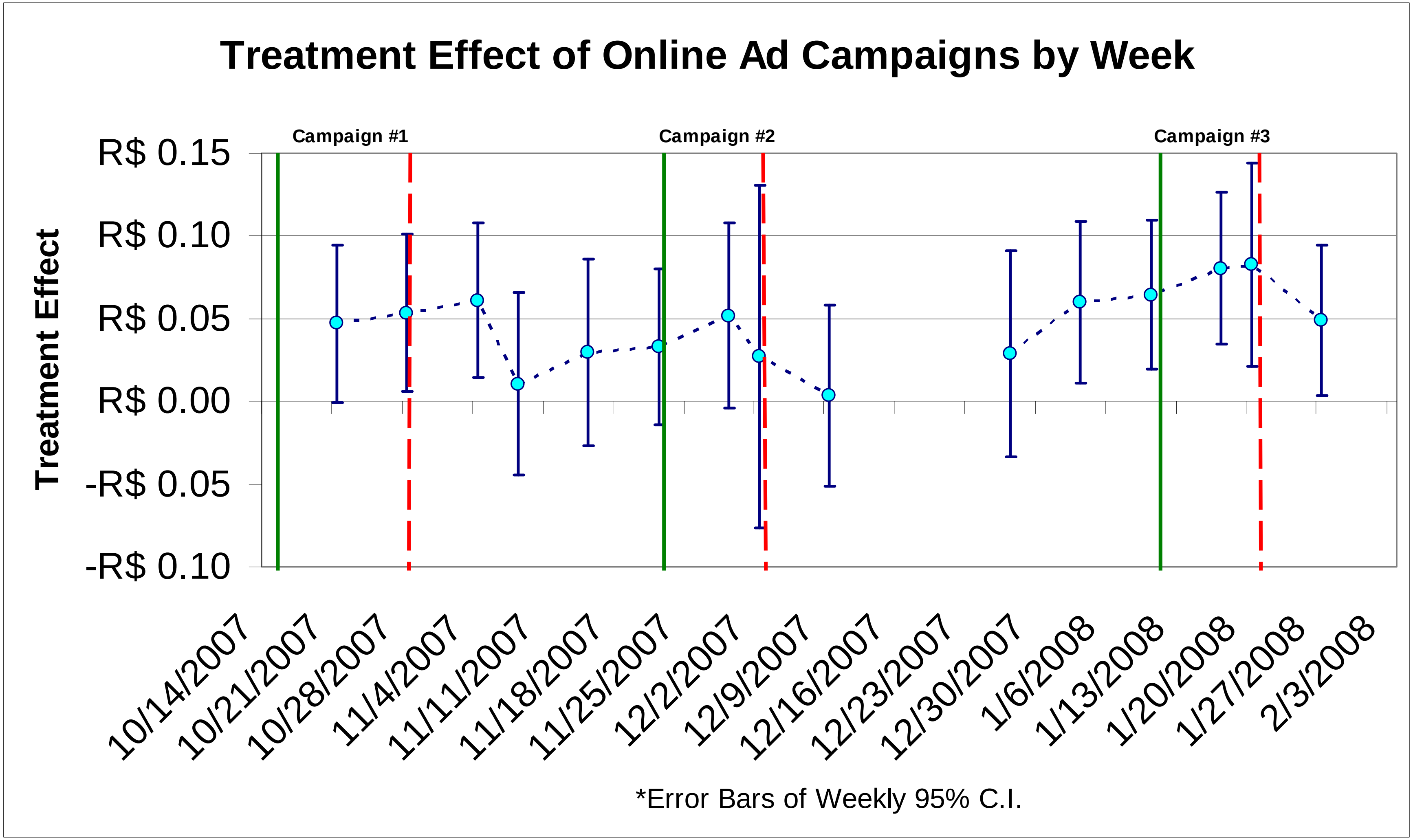


Treated-Untreated Differences in Three-week Sales

The 3-week estimates suggest persistent effects.

- Third week DID estimate confirms persistence of sales beyond the campaign.
  - Three-week DID treatment effect: R$0.166 (0.052).
  - Compare to two-week DID estimate: R$0.102 (0.043).
  - Single-week treatment effect: R$0.061 (0.024).

Strong persistence: we find that DID estimates are consistently positive, even several weeks after the ads.



We find that weekly estimates are consistently positive for 15 weeks.

	Treatment Effect*	Robust S.E.
Campaign #1		
Week 1 During	**R$ 0.047**	0.024
Week 2 During	**R$ 0.053**	0.024
Week 1 Following	**R$ 0.061**	0.024
Campaign #2		
3 Weeks Before	**R$ 0.011**	0.028
2 Weeks Before	**R$ 0.030**	0.029
1 Week Before	**R$ 0.033**	0.024
Week 1 During	**R$ 0.052**	0.029
Week 2 During (3 Days)	**R$ 0.012**	0.023
Week 1 Following	**R$ 0.004**	0.028
Campaign #3**		
3 Weeks Before	**R$ 0.029**	0.032
2 Weeks Before	**R$ 0.060**	0.025
1 Week Before	**R$ 0.064**	0.023
Week 1 During	**R$ 0.080**	0.023
Week 2 During (3 Days)	**R$ 0.035**	0.013
Week 1 Following	**R$ 0.049**	0.023

N=1,577,256 obs. per week**

* For purposes of computing the treatment effect on the treated, we define "treated" individuals as having ever seen an ad in one of these campaigns up to that point in time.

** Estimates for Campaign #3 involves mismatched observations due to an imperfect merge to compute the difference in differences. Lewis (2008) derives the methods used to compute these estimates.

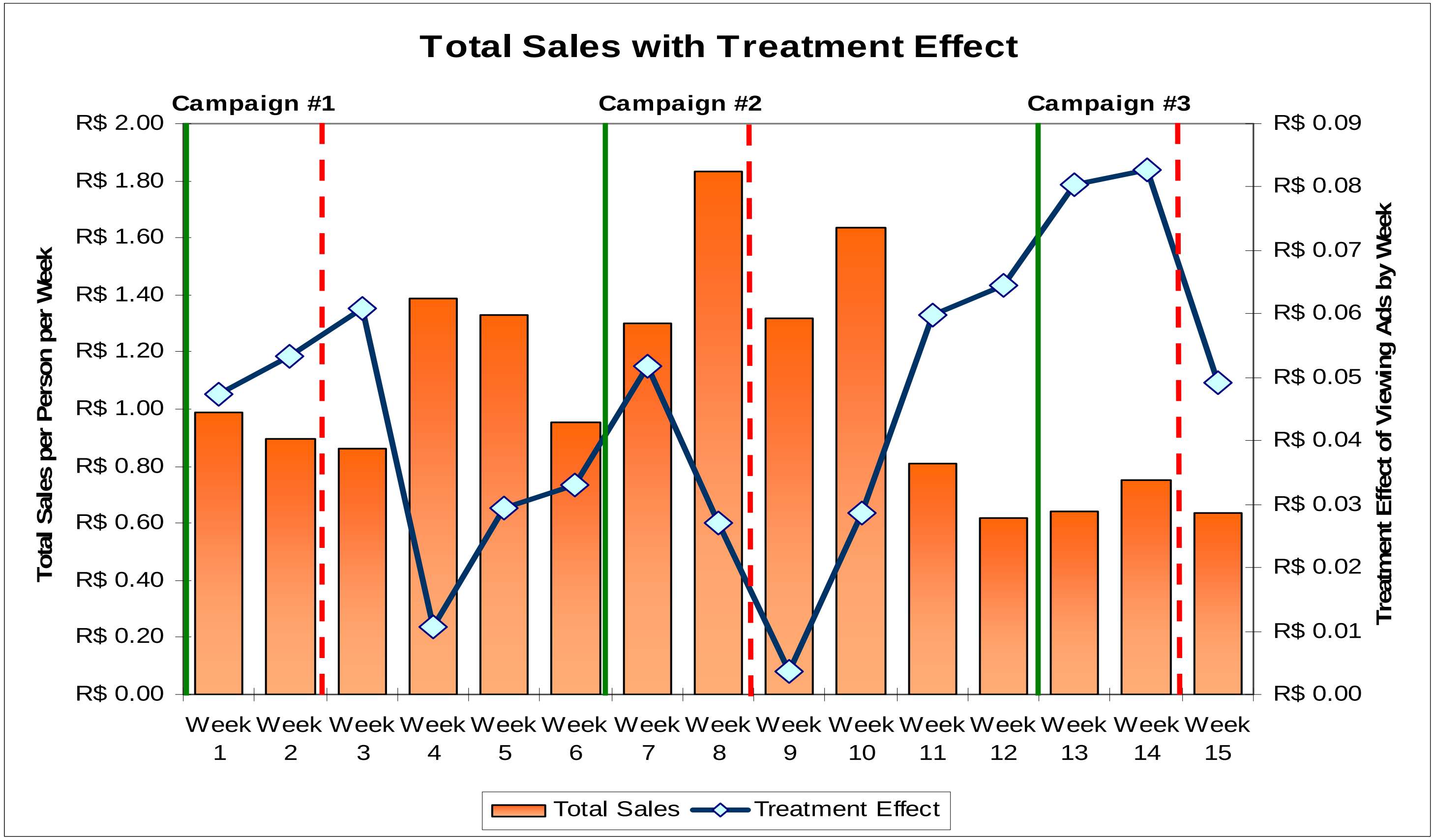
Cumulative effects indicate a large return relative to the cost of ads.

	Treatment Effect	Robust S.E.	t-stat	P(|t|>T)
Average Weekly Effect				
Simple Average (OLS)	**R$ 0.045**	0.0140	3.25	0.001
Efficient Average (GLS)	**R$ 0.048**	0.0136	3.53	0.000
Cumulative Effects over All 3 Campaigns				
Cumulative Sales	**R$ 0.532**	0.196	2.72	0.007
Simple Aggregate Effect (OLS)	**R$ 0.611**	0.188	3.25	0.001
Efficient Aggregate Effect (GLS)	**R$ 0.645**	0.183	3.53	0.000
Length of Measured Cumulative Effects		13 wks. 3 days		

- Best estimate: R$0.65 times 864K individuals.
- Total revenue impact: R$560K±310K.
- Total cost of ads: R$51K.
- Large return to online retail-image advertising!

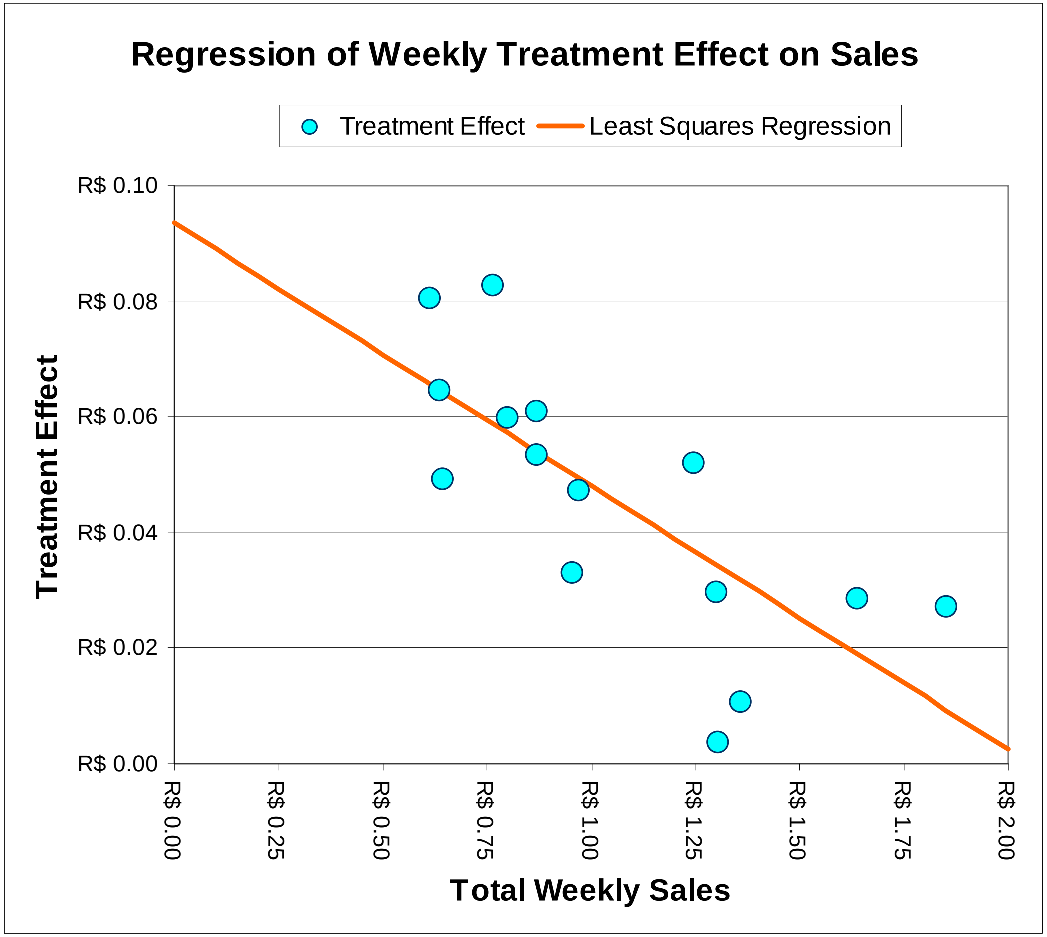
The treatment effect appears to be larger when total sales are smaller.

- Effect of viewing ads varies over 15 weeks.
  - During weeks with higher sales, effect of viewing ads is smaller.
  - During weeks with lower sales, effect of viewing ads is larger.



The advertising treatment effect appears to be countercyclical.

- Online advertising may help smooth out fluctuations in sales by getting people to buy more when sales are down.



Next we estimate separate effects for the effect on offline and online sales.

	Total Sales	Offline Sales	Online Sales
Ads Viewed	**R$ 0.166**	**R$ 0.155**	**R$ 0.011**
[63.7% of Treatment]	(0.05)	(0.05)	(0.02)

- As before, these are DID estimates.
- We see that 93% of the total effect on sales comes through offline sales.

Do we capture the effects of ads by measuring only clicks? No.

	Total Sales	Offline Sales	Online Sales
Ads Viewed	**R$ 0.166**	**R$ 0.155**	**R$ 0.011**
[63.7% of Treatment]	(0.05)	(0.05)	(0.02)
Ads Viewed Not Clicked	**R$ 0.139**		
[92.8% of Viewers]	(0.05)		
Ads Clicked	**R$ 0.508**		
[7.2% of Viewers]	(0.16)		

- Clickers buy more, as one would expect.
- But viewers have an increase in sales that represents 78% of the total treatment effect.

The effect on non-clickers occurs in stores, not in the online store.

	Total Sales	Offline Sales	Online Sales
Ads Viewed	**R$ 0.166**	**R$ 0.155**	**R$ 0.011**
[63.7% of Treatment]	(0.05)	(0.05)	(0.02)
Ads Viewed Not Clicked	**R$ 0.139**	**R$ 0.150**	-R$ 0.010
[92.8% of Viewers]	(0.05)	(0.05)	(0.02)
Ads Clicked	**R$ 0.508**		
[7.2% of Viewers]	(0.16)		

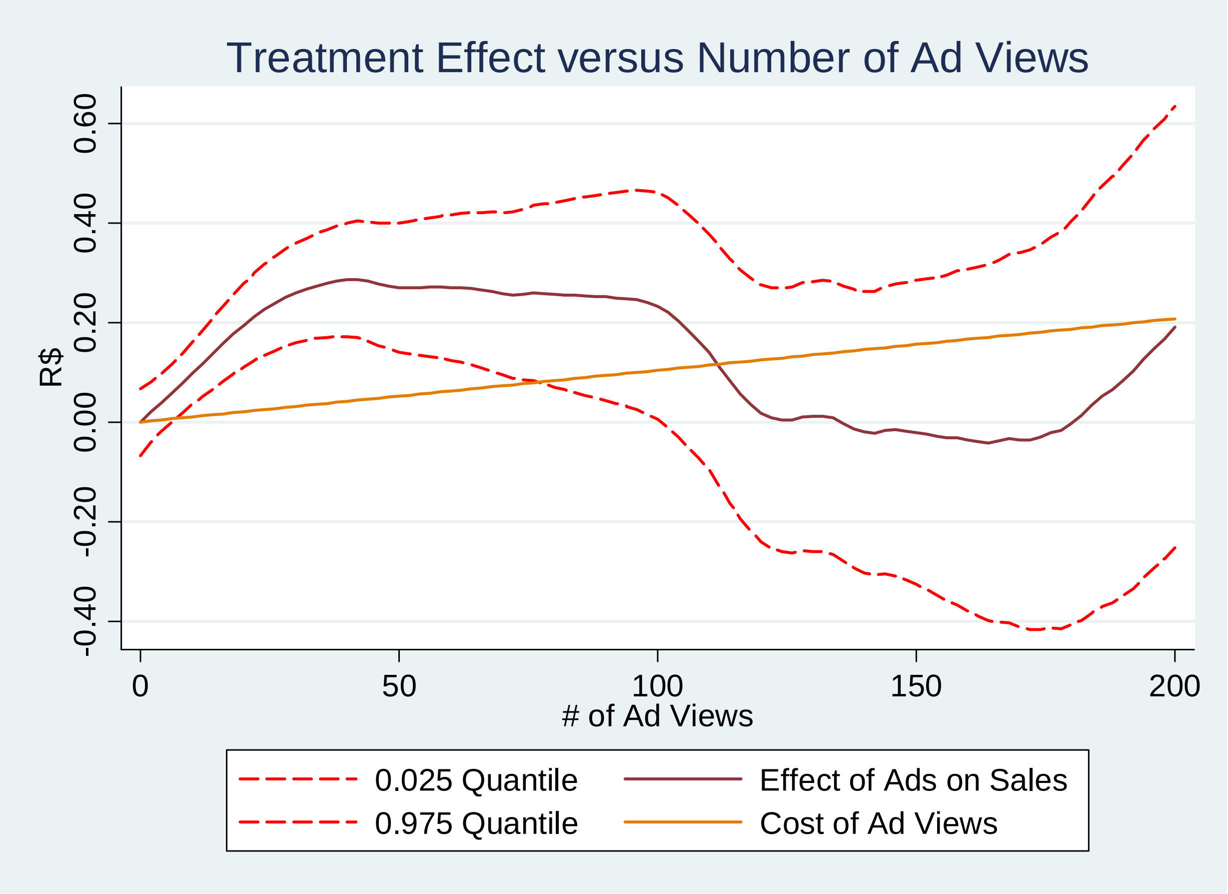
The effect on clickers occurs both offline and online.

	Total Sales	Offline Sales	Online Sales
Ads Viewed [63.7% of Treatment]	**R$ 0.166** (0.05)	**R$ 0.155** (0.05)	**R$ 0.011** (0.02)
Ads Viewed Not Clicked [92.8% of Viewers]	**R$ 0.139** (0.05)	**R$ 0.150** (0.05)	-R$ 0.010 (0.02)
Ads Clicked [7.2% of Viewers]	**R$ 0.508** (0.16)	R$ 0.215 (0.16)	**R$ 0.292** (0.04)

- Those who click on the ads buy significantly more online.
- The estimate on offline sales is too imprecise to be statistically significant.

The effect of online display ads depends on browsing behavior.

The largest effect of the advertising was on customers who browsed enough to see between 1 and 100 ads.



The effect consists of both an increase in basket size and higher purchase probability.

- ¼ effect comes from a larger number of customers making purchases.
- ¾ effect comes from larger average purchases.

	3-Week DID Treatment Effect	Treated Group Level*
Pr(Transaction)	**0.10%**	**6.48%**
	-0.05%	
Mean Basket Size	**R$ 1.75**	**R$ 40.72**
	-0.74	
Revenue Per Person	**R$ 0.166**	**R$ 2.639**
	-0.052	

* Levels computed using all individuals who saw at least one ad during the 2-week campaign and all sales figures from 3 weeks following the start of the campaign.

Conclusion:
Retail Advertising Works!

1. Online display advertising increases both online and offline sales. Approximately 5% increase in revenue.
2. Effects are persistent across many weeks.
3. Estimated effects of advertising are *inversely* correlated with weekly sales volume (countercyclical).
4. Total revenue effect more than 10X the cost of ads.
5. Views without clicks still produce large results for offline sales. Clicks predict online sales.
6. Optimal frequency may be much higher than in traditional media: perhaps on the order of 100 impressions.
7. Positive effects both on basket size (75% of the effect) and probability of transaction (25% of the effect).

Future advertising experiments will provide more insights.

- Replicate these results with other retailers.
- Investigate the effectiveness of targeting.
  - Demographics
  - Geographic
  - Online behavior
  - Past sales
- How does frequency of exposure matter?
  - Experiment with frequency capping.
- Competitive effects of advertising.

Why don’t firms experiment more?

- The flaws with analysis of observational data are subtle.
- Managers don’t often think like scientists.
- When you experiment, you’re admitting you don’t already know the right answer.
- When you experiment, one of the things you try will turn out to be “the wrong thing to do.”
- It’s risky to try something different than the norm in your field.