Linking Clicks to Bricks: Spillover Benefits of Online Advertising

Mi Zhou, Vibhanshu Abhishek, Edward H. Kennedy, Kannan Srinivasan, and Ritwik Sinha
Linking Clicks to Bricks: Spillover Benefits of Online Advertising

Mi Zhou
Carnegie Mellon University, Heinz College
mzhou1@andrew.cmu.edu

Vibhanshu Abhishek
University of California – Irvine, Paul Merage School of Business
vibs@uci.edu

Edward H. Kennedy
Carnegie Mellon University, Department of Statistics & Data Science
edward@stat.cmu.edu

Kannan Srinivasan
Carnegie Mellon University, Tepper School of Business
kannans@andrew.cmu.edu

Ritwik Sinha
Adobe Systems
risinha@adobe.com

October 2018
Abstract

Businesses have widely used email ads to directly send promotional information to consumers. While email ads serve as a convenient channel that allows firms to target consumers online, are they effective in increasing offline revenues for firms that predominantly sell in brick-and-mortar stores? Is the effect of email ads, if any, heterogeneous across different consumer segments? If so, on which consumers is the effect highest? In this research, we address these questions using a unique high-dimensional observational dataset from one of the largest retailers in the US, which links each consumer’s online behaviors to the item-level purchase records in physical stores. We use a doubly robust estimator (DRE) that incorporates nonparametric machine learning methods and allows us to perform causal estimation on observational data. Using the DRE we find that on average receiving email ads can increase a consumer’s spending in physical stores by approximately $1.49, although the effect is heterogeneous across number of ads. Additionally, we find that the increased offline sales result from increased purchase probability and a wider variety of products being purchased. Further, we use a data-driven approach to demonstrate that the effect of email ads is heterogeneous across different consumer segments. Interestingly, the effect is highest among consumers who have fewer interactions with the focal retailer recently (i.e., lower email opening frequency). Overall, our results suggest a reminder effect of email ads. Receiving email ads from the retailer can generate awareness and remind the consumer of the retailer’s offerings of various products and services, which gradually increase the consumer’s purchase probability in the retailer’s physical stores. These findings have direct implications for marketers to improve their digital marketing strategy design and for policy makers who are interested in evaluating the economic impact of prevalent email advertising.

Keywords: Online advertising, offline sales, email ads, doubly robust estimator, machine learning, causal inference
1. Introduction

With the widespread use of the internet, businesses have been using email ads to directly send promotional information to consumers. According to a survey, over 80% of firms worldwide reported investing in email ads to reach consumers (Wattal et al. 2012). Recent statistics show that there are 3.7 billion email users in the world as of 2017, and the number of email users in the U.S. is estimated to reach nearly 255 million by the end of 2020 (Statista 2017). It is reported 91% of consumers check email at least once a day on their smartphones (Entrepreneur 2015). Email is likely to remain one of the main marketing channels in the country. Roughly half of the marketers in the U.S. intend to increase their spending on email marketing, and about 48% are expected to maintain their budgets. In line with this projection, email advertising spending in the U.S. is expected to increase from 270 million U.S. dollars in 2015 to 350 million U.S. dollars in 2019 (Statista 2017). Moreover, according to a recent survey among marketing professionals in the U.S., 63% of respondents said their priority in marketing investment was email advertising, which was ranked 2nd out of 18 different advertising channels (Target Marketing 2016). Given the popularity and importance of email advertising, most marketers are confident about its profitability. During a 2016 survey, 36% of responding marketers stated that email marketing generated significant return on investment (ROI), and 44% of them stated that email marketing generated some ROI (Salesforce Research 2016). However, as prior research points out, the profitability of email advertising could vary substantially across firms and consumers (Zhang et al. 2017). Not surprisingly, how to better target email ads to consumers online in order to improve sales has been an important business problem, which is critically dependent on a comprehensive understanding of heterogeneous causal effects of email ads on consumer behaviors.

However, given the abundance of research in the marketing literature about digital advertising, most prior studies related to email advertising have focused on understanding how the nature of content in an email influences consumers’ responses (i.e., Ansari and Mela 2003, Wattal et al. 2012), whereas the economic impact of email ads on actual revenues has not been much researched (Sahni et al. 2017) and there is little evidence on how to optimize the delivery of email ads based on its heterogeneous effects on different consumers. While most of research on digital advertising has focused on search and display ads, given the performance of emails ads, they need to be examined in greater detail. In addition, the existing papers largely focus on the effect of email advertising on online sales. Given the fact that most large brick-and-mortar retailers use email ads to reach their consumers but a substantial portion of their sale revenues are still created offline (i.e., most of their products are
sold in physical stores), this points to a gap as the effect of email ads on offline sales is not well understood, and given the aforementioned statistics, this impact can be substantial. What makes this problem even more challenging, and has potentially served as an impediment to research on this topic, is the difficulty in matching each consumer’s offline purchase records to that consumer’s online activities for estimating these cross-channel effects.

To realize the importance of this fact, consider that US retailers’ online advertising spending accounts for over one-fifth of all US online advertising spending, more than in any other industry over the past several years (eMarketer 2015). However, according to recent statistics in retail sales, only 6.7% of consumers’ total purchases were generated online, whereas more than 90% of their total purchases were still generated in physical stores, offline (US Census Bureau 2015). Hence, one pertinent question for retailers is whether their investment in online advertising has an impact on their offline revenues. If the effectiveness of online advertising goes beyond e-commerce, i.e. to offline sales, managers ignoring such cross-channel effects may devise suboptimal marketing strategies. If not, then retailers might have been wasting ad dollars online.

It is well known that researchers face challenges in measuring the cross-channel effect of online advertising on offline sales mainly due to a lack of high-quality single-source data at consumer level (Johnson et al. 2017, Lobschat et al. 2017, Abhishek et al. 2015). Though there is a stream of research on omni-channel retailing that investigates the relationship between consumers’ online and offline behaviors, the causal effects of online advertising on offline sales are still understudied in the extant literature (Johnson et al. 2017, Lobschat et al. 2017). For example, although we can easily track consumers’ online activities (e.g., receiving an email) using tools such as cookies, it is extremely difficult to match each consumer’s online activities to that consumer’s purchase records in physical stores, which is needed for the estimation of cross-channel effects at the consumer individual level. To the best of our knowledge, there are only three notable works in recent literature that have provided some empirical evidence for the effect of display advertising on offline sales (Lewis and Reiley 2014, Johnson et al. 2017, Lobschat et al. 2017). However, there are several reasons why these studies cannot shed light on the omni-channel effect of emails ads.

Firstly, display ads are often designed to build brand awareness, whereas the content of email ads usually consists of promotional information. Secondly, display ads and email ads differ in their mechanisms of how they affect consumer behavior. For instance, display ads are found to usually have an early impact on the consumer as they encourage the consumer to visit the firm’s website, suggesting
an indirect effect on the consumer’s purchase (Abhishek et al. 2017, Lobschat et al. 2017), whereas email ads can directly increase a consumer’s online ticket purchase (Sahni et al. 2017). Last but not least, the two types of ads are often targeted at different types of consumers. Display ads often target consumers more broadly based on their proclivity to visit certain websites, so advertisers can cast a wider net using these types of ads. However, email ads can often reach a much more focused group of consumers, who have willingly allowed the firm to contact them due to their preference for the firm’s products.

Given these differences between display ads and email ads, it is imperative to examine the effect of email ads on offline sales independently due to the importance of emails as a marketing channel. To achieve this objective, we focus on the following empirical questions in this paper. First, can email ads create a statistically and economically significant improvement on both online and offline revenues for the retailer? To the best of our knowledge, this is the first paper that examines the cross-channel question for emails. Second, is the causal effect of email ads on offline sales heterogeneous across different consumer segments? If so, on which consumers is the effect of email ads most favorable? To answer these empirical questions, we have partnered with an industry-leading data analytics company to obtain a proprietary large-scale individual-level dataset (about 3 TB) from one of the largest domestic retailers in the US. The dataset includes detailed records for all the marketing, visitation and e-commerce data (online), and item-level transaction records in physical stores (offline) for each consumer. We believe that the depth and breadth of this data is quite unique in the marketing literature.

Measuring email ads’ effectiveness is complicated by the potential endogeneity problems of advertising (i.e., targeting). Well-designed experiments can ideally overcome endogeneity issues (Gordon et. al, 2019), but in practice, experiments can also be non-representative, costly, and time consuming (Eckles and Bakshy, 2017). Hence, it is important for researchers to take advantage of high-quality observational datasets available to many firms, and to leverage advanced statistical methods to address potential endogeneity issues in observational datasets. In this paper we try to reconcile two opposing viewpoints by Eckles and Bakshy (2017) who favor observational methods under certain conditions and Gordon et. al (2019) that favors randomized experiments, and show that in our case observational data might be appropriate for drawing causal inference. In recent years, the machine learning literature has seen a burgeoning stream of research that has developed new methods to discover causal inference using high-dimensional datasets (see Athey and Imbens (2017) for an
overview on the literature of machine learning in econometrics). However, to the best of our knowledge, most of these state-of-the-art approaches are underused in applied research fields. Thus, another goal of this paper is to adapt some recent methodological developments to address marketing problems. In particular, we introduce a doubly robust estimator (DRE) that incorporates nonparametric machine learning methods to estimate average treatment effects (Kennedy 2016, Chernozhukov et al. 2017) and heterogeneous treatment effects (van der Laan 2013) of email ads on offline sales. DRE protects against bias due to model misspecification, avoids the reliance on unrealistic parametric distributions, and reduces “curse of dimensionality” caused by big data. We hope that these techniques can enable researchers to obtain more credible and robust causal estimates using very granular data on consumers.

In our analysis, we find that email ads have a statistically and economically significant positive effect on consumers’ spending in the retailer’s digital and physical stores. However, the effects are heterogeneous across channels and number of ads. While receiving one or two email ads does not significantly affect a consumer’s offline purchase amount, receiving three email ads increases a consumer’s offline spending by approximately $11.82. This finding suggests a reminder effect of email ads: Receiving more emails from the retailer can attract the consumer’s attention and gradually register the retailer’s various types of products in the consumer’s memory, which can nudge the consumer, making him or her more likely to consider the retailer’s products on the next purchase occasion. This also supports some industry reports that some of the roles marketing emails play in the consumer purchase process include the provision of additional incentives to purchases and as reminders for purchases that need to be made (Statista 2017). Moreover, we find that the increased offline sales due to email ads are from a positive impact on consumer’s purchase probability couple with a wider variety of products being purchased by the consumer. The emails ads have a relatively smaller effect on online sales and the effect (weakly) increases as the numbers of ads increase.

We also find evidence for heterogeneous effects of email ads on offline sales using a loss-based machine learning approach. The effect of email ads is highest among consumers who have had fewer interactions with the retailer in recent weeks, which is consistent with the reminder effect of email ads. For consumers who may have forgotten about the retailer and have not had much interaction with the retailer recently, receiving email ads can increase the chances of the consumer getting reminded of and considering the retailer’s offerings for purchase. These findings have important implication for managers. Traditionally, retailers sending promotional email ads would want
to target consumers who have already shown an interest in their products (i.e., retargeting consumers who have purchased or visited recently). Contrary to this traditional belief, our results suggest that retailers should consider targeting consumers who may have forgotten about the retailer recently to remind them about the retailer’s offerings of various products and services. This retargeting fosters an accumulation of consumer awareness, with positive effects on the consumer’s purchase probability, which can improve offline sales.

Overall, this paper provides direct managerial guidance for firms investing in email ads, especially for retailers that predominantly sell through physical stores, offline. Moreover, this paper makes a methodological contribution by demonstrating some state-of-the-art machine learning methods for estimating average treatment effects using high-dimensional observational data and extending the methodology to estimate heterogeneous treatment effects. We believe these advanced statistical approaches can enable researchers to produce more credible causal estimates using observational data and steer marketing decisions in the right direction.

2. Literature Review

Our research is related to two broad streams of literature: (i) email marketing, and (ii) cross-channel effects of online advertising. We discuss these two streams below.

Email Marketing

This research is related to the email marketing literature, which is limited and does not directly address our research questions. Wattal et al. (2012) studied the economic benefits of personalization in email ads. They found that when firms use product-based personalization, consumers respond positively, but consumers respond negatively when firms are explicit in their use of personally identifiable information, highlighting consumers’ concerns over the use of information in personalization. Kumar et al. (2014) studied the impact of marketing activities on the propensity of a consumer getting in and out of email marketing lists. Sahni et al. (2017) ran field experiments on an online ticket resale platform and found that emails can significantly increase a consumer’s ticket purchase on the platform. Our paper is closely related to Zhang et al. (2017), who studied customers’ email open and purchase behaviors by using a unified hidden Markov and copula framework. Based on parametric predictive modeling using data of 200 users, the authors found that email-active customers are not necessarily associated with a higher number of purchase occasions. Using actual offline sales records of over 100,000 consumers, our paper adds to this stream of literature by
providing empirical evidence of causal effects of email ads on actual offline and online sales based on nonparametric causal modeling incorporating machine learning methods.

Cross-Channel Effects of Online Advertising

More broadly, this research is closely related to the literature that investigates the cross-channel effects of online advertising. From a managerial perspective, knowing the full return from online advertising expenditures is necessary for developing an efficient marketing plan. Solely measuring the effects online could result in misjudgment of the impact of advertising. The importance of cross-channel effects of online advertising has long been recognized in both academia and industry, which motivates a stream of research on omni-channel retailing that investigates the correlation between consumers’ online and offline behaviors. However, there is not enough research in this domain that has causally quantified the cross-channel effects of online advertising at individual level, due to a lack of high-quality individual-level data (i.e., difficulty in linking the data of consumers’ online activities and the records of their offline behaviors at individual level). The following discussion notes the most relevant works in chronological order.

Naik and Peters (2009) used aggregated advertising spending data and found that online advertising has positive influence on increasing consumers’ offline visits to dealer showrooms as well as online visits to dealer’s website. Wiesel et al. (2011) used aggregated-level data of advertising costs and sales revenues and found that using Google AdWords has positive profit impact on offline sales. Chan et al. (2011) found that a large portion of the customer lifetime value from customers acquired through paid search is due to transactions in offline channels, providing further support for the importance of cross-channel management.

Danaher and Dagger (2013) measured individual-level exposure to multiple media by surveying consumers and asking them to recall which media channels they watched or read and then linking back to firm-supplied customer contact history and sales transactions. Their results show that email and sponsored search have an influence on purchase outcomes. This paper is a first step in addressing a difficult but important contemporary marketing problem, but the data only included a restricted subset of consumers (i.e., 3007 female loyalty program members between 25 and 54 years of age living in Australia’s three largest cities who accepted to take an online survey to complete questionnaires for a reward of 100 loyalty points), and self-reported media exposure recall may suffer from consumers’ memory recall bias, which has certain limits on the accuracy of comparing the relative effectiveness of each advertising channel.
Dinner et al. (2014) used aggregated data from an upscale clothing retailer and found that search advertising is effective in increasing offline sales. However, the authors admit that their analysis is based on descriptive data; thus, their analysis finds associations rather than establishes direct causality, since using aggregated advertising expenditures as independent variables instead of consumers’ actual advertising exposures leads to potential omitted-variable bias and simultaneity bias.

In addition to several observational studies using aggregated-level data that may lead to systematically biased estimates due to aggregation bias (Abhishek et al. 2015), there are three recent studies that have conducted randomized field experiments to investigate the effects of online display advertising on offline sales. Lewis and Reiley (2014) conducted a randomized experiment with Yahoo! users to investigate the effects of brand image advertising. They demonstrate that online display advertising can profitably increase both online and in-store purchases. Unfortunately, without making a difficult to test difference-in-differences assumption, they did not obtain statistically significant experimental estimates. Johnson et al. (2017) used another larger randomized field experiment with Yahoo! users to find statistically significant evidence that online display advertising can increase retailers’ offline sales. In addition, Lobschat et al. (2017) have recently analyzed an observational dataset to show that banner advertising can increase consumers’ online website visit incidence, and indirectly increase offline sales through website visits. Since all these studies examine the impact of display ads, their effect sizes are expected to be small (or insignificant). More recently, Kalyanam et al (2018) use a meta experiment and show weak positive effects of sponsored ads on offline sales.

To summarize, our paper adds to the above literature in several ways. First, using a unique large-scale individual-level dataset and modern machine learning approaches, this research quantifies the economic impact of email ads on actual offline sales, providing the first empirical evidence on the cross-channel effect of email ads. Second, existing studies typically do not allow for heterogeneity in online advertising effectiveness (Lewis and Reiley 2014, Johnson et al. 2017), or simply redefine consumers into binary categories (Lobschat et al. 2017). In this research, we use data-adaptive methods to demonstrate heterogeneous effects of email ads on offline sales across different consumer segments. In doing so, we shed light on applicable digital marketing targeting strategies.
3. Data

We have obtained a proprietary large-scale individual-level dataset of one of the largest retailers in the home and office supplies businesses in the US.\(^1\) This retailer has both a web store (online) and a chain of physical stores (offline). The retailer sells many different products and services such as office supplies, business machines, furniture, and business services. All the products in our dataset are categorized into nine major categories, as shown in Table 1.

We have data for eight months between December 2012 and August 2013, for 23 million customers (about 3 TB), containing all the marketing, visitation and e-commerce data (online), and transaction data in physical stores (offline). This large-scale dataset is unique as it is collated by combining separate data sources using a proprietary entity disambiguation technique developed by our partner, a leading data analytics company in the US. This firm leverages mapping points such as email, loyalty cards, postal addresses, cookies, and other online and offline activities to unify the sources to a common grouping key. It provides a holistic view of consumers’ interactions with the retailer across different channels including online marketing, online properties, and brick-and-mortar locations. We are thus able to better understand how different channels jointly influence consumer behavior both online and offline. To our knowledge, the amount of consumer data that we process is much larger than that of any previous papers in related research. Seventy-eight percent of the consumers in our dataset have only purchased in physical stores, even though they consistently received online ads during the observation period, which shows that although retailers invest a lot in online advertising, most consumers still prefer to purchase offline. In this study, we focus on existing consumers who had opted into receiving emails from the focal retailer.

In our exploratory analysis of the whole dataset, we find that among the 23 million consumers, 11.3 million were not exposed to advertising activities during the observational period, although we still observe their purchase records in offline stores. The remaining 11.7 million consumers received at least one email ad during the observation period. The average weekly spending of these 11.7 million consumers is consistently higher than that of the remaining 11.3 million consumers in our dataset. Since there might be some systematic differences between these two groups of consumers, in this study we select for analysis only those 11.7 million consumers who received at least one email ad

\(^1\) To preserve anonymity, we cannot disclose the name of the retailer.
during the observation period. Thus, restricting the sample to consumers who had agreed to accept email ads from this focal retailer will enable us to provide a cleaner estimate as well as a lower bound of the effectiveness of email ads on consumers’ purchase.

Table 1. Product Category Summary

<table>
<thead>
<tr>
<th>Product Category</th>
<th>Number of Products</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computers &amp; Accessories</td>
<td>1,857,123</td>
</tr>
<tr>
<td>Food &amp; Housewares</td>
<td>422,716</td>
</tr>
<tr>
<td>Furniture</td>
<td>404,377</td>
</tr>
<tr>
<td>Office Supplies</td>
<td>246,817</td>
</tr>
<tr>
<td>Business Machines</td>
<td>149,578</td>
</tr>
<tr>
<td>Technology Supporting Services</td>
<td>82,378</td>
</tr>
<tr>
<td>Business Services</td>
<td>71,657</td>
</tr>
<tr>
<td>Store Supplies</td>
<td>26,060</td>
</tr>
<tr>
<td>Other</td>
<td>21,891</td>
</tr>
</tbody>
</table>

Figure 1 illustrates the number of consumers exposed to this retailer’s email ads and display ads, respectively, in every week during the observation period. For this retailer email marketing is a major channel through which the retailer sends promotional information to its consumers. These email ads are sent in bulk to consumers. Every week during our observation period about 6 to 8 million consumers received email ads from this retailer, whereas only 0.2 to 0.5 million consumers were exposed to display ads every week. Therefore, in this research we focus on the effects of email ads on consumers’ offline purchase behaviors.

Figure 1 illustrates the number of consumers exposed to this retailer’s email ads and display ads, respectively, in every week during the observation period. For this retailer email marketing is a major channel through which the retailer sends promotional information to its consumers. These email ads are sent in bulk to consumers. Every week during our observation period about 6 to 8 million consumers received email ads from this retailer, whereas only 0.2 to 0.5 million consumers were exposed to display ads every week. Therefore, in this research we focus on the effects of email ads on consumers’ offline purchase behaviors.

Figure 1. Number of Consumers Receiving Ads Every Week (color online)

Figure 2-a and 2-b illustrate the patterns of email ads schedule at daily level and weekly level, respectively. Figure 2-a illustrates the total number of email ads sent to consumers every day. This
figure shows that the weekly schedule of email ads is relatively consistent; this retailer usually sent email ads nearly every Monday, Tuesday, Wednesday, and Friday. Direct marketers often refer to this regular schedule of planned campaigns as advertising cadence (Zantedeschi et al. 2016). This limits the possibility of simultaneity between advertising exposures and periods with large unexplained shocks in demand, which could result in biased estimated effects of email advertising. While this retailer regularly sent email ads every week, they were sent to different consumers each time. Figure 2-b illustrates that the number of email ads each consumer received every week differs over time, which shows considerable customer level variation in the emails sent.

After conducting an exploratory analysis of the whole dataset, we randomly draw a sample of about 100,000 consumers and create a weekly panel for these consumers in 10 consecutive weeks (from March 11, 2013 to May 19, 2013). The summary statistics for the weekly panel is shown in Table 2. Note that in our dataset, purchase amount and quantity could be negative values, which represent net returns of merchandise. We do not exclude these observations from our analysis, as advertising could influence purchases even for consumers who return more than they purchase during a given time period. This approach is consistent with the prior literature (Lewis and Reiley 2014). Moreover, in this sample, we observe that only 3,192 consumers (3%) were exposed to display ads during the target week. The distribution of display ad exposures is relatively skewed. Therefore, in our robustness checks, we remove these 3% of consumers to provide cleaner estimates, which show qualitatively consistent results.

![Figure 2-a (daily level). Total Number of Email Ads Sent Out Every Day (color online)](image)

---

2 The choice of restricting the analysis to 10 weeks of data is due to the computational complexity of estimating our model. Furthermore, we feel 10 weeks of data for analysis is adequate because the retailer uses past 6 weeks of data to make digital advertising decisions. The results are qualitatively similar for longer lengths of the panel.
Figure 2-b (weekly level). Number of Email Ads a Consumer Receives Every Week (color online)

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std.</th>
<th>Min</th>
<th>Median</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>#Email_sent</td>
<td>2.59</td>
<td>1.93</td>
<td>0.00</td>
<td>3.00</td>
<td>20.00</td>
</tr>
<tr>
<td>#Email_opened_times</td>
<td>0.82</td>
<td>2.37</td>
<td>0.00</td>
<td>0.00</td>
<td>82.00</td>
</tr>
<tr>
<td>#DisplayAds_impressions</td>
<td>2.57</td>
<td>24.99</td>
<td>0.00</td>
<td>0.00</td>
<td>2188.00</td>
</tr>
<tr>
<td>Coupon used (quantity)</td>
<td>0.11</td>
<td>0.89</td>
<td>-1.00</td>
<td>0.00</td>
<td>56.00</td>
</tr>
<tr>
<td>Coupon used (amount)</td>
<td>-0.51</td>
<td>5.57</td>
<td>-467.70</td>
<td>0.00</td>
<td>74.58</td>
</tr>
</tbody>
</table>

Weekly purchase amount (US $) and quantity:

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std.</th>
<th>Min</th>
<th>Median</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amount – Offline</td>
<td>4.14</td>
<td>34.78</td>
<td>-500.00</td>
<td>0.00</td>
<td>4732.00</td>
</tr>
<tr>
<td>Amount – Online</td>
<td>1.61</td>
<td>30.21</td>
<td>-5.53</td>
<td>0.00</td>
<td>3953.00</td>
</tr>
<tr>
<td>Quantity – Offline</td>
<td>0.49</td>
<td>3.44</td>
<td>-8.00</td>
<td>0.00</td>
<td>529.00</td>
</tr>
<tr>
<td>Quantity – Online</td>
<td>0.08</td>
<td>1.33</td>
<td>0.00</td>
<td>0.00</td>
<td>201.00</td>
</tr>
</tbody>
</table>

Weekly purchase amount in each product category (US $):

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std.</th>
<th>Min</th>
<th>Median</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business Machines</td>
<td>2.43</td>
<td>23.21</td>
<td>-450.00</td>
<td>0.00</td>
<td>2600.00</td>
</tr>
<tr>
<td>Business Services</td>
<td>0.50</td>
<td>18.75</td>
<td>-34.99</td>
<td>0.00</td>
<td>4732.00</td>
</tr>
<tr>
<td>Food &amp; Housewares</td>
<td>1.46</td>
<td>15.26</td>
<td>-77.98</td>
<td>0.00</td>
<td>1748.00</td>
</tr>
<tr>
<td>Computers &amp; Accessories</td>
<td>0.96</td>
<td>21.69</td>
<td>-500.00</td>
<td>0.00</td>
<td>3664.00</td>
</tr>
<tr>
<td>Office Supplies</td>
<td>0.63</td>
<td>6.80</td>
<td>-89.49</td>
<td>0.00</td>
<td>787.70</td>
</tr>
<tr>
<td>Furniture</td>
<td>0.25</td>
<td>10.49</td>
<td>-180.00</td>
<td>0.00</td>
<td>2360.00</td>
</tr>
<tr>
<td>Store Supplies</td>
<td>0.03</td>
<td>0.77</td>
<td>-95.99</td>
<td>0.00</td>
<td>30.07</td>
</tr>
<tr>
<td>Technology Supporting Services</td>
<td>0.15</td>
<td>5.55</td>
<td>-93.01</td>
<td>0.00</td>
<td>499.00</td>
</tr>
<tr>
<td>Other</td>
<td>0.00</td>
<td>0.42</td>
<td>-94.50</td>
<td>0.00</td>
<td>49.98</td>
</tr>
</tbody>
</table>

Weekly purchase quantity in each product category:

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std.</th>
<th>Min</th>
<th>Median</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business Machines</td>
<td>0.06</td>
<td>0.44</td>
<td>-4.00</td>
<td>0.00</td>
<td>21.00</td>
</tr>
<tr>
<td>Business Services</td>
<td>0.07</td>
<td>2.44</td>
<td>-3.00</td>
<td>0.00</td>
<td>525.00</td>
</tr>
<tr>
<td>Food &amp; Housewares</td>
<td>0.14</td>
<td>2.58</td>
<td>-4.00</td>
<td>0.00</td>
<td>775.00</td>
</tr>
<tr>
<td>Computers &amp; Accessories</td>
<td>0.02</td>
<td>0.24</td>
<td>-3.00</td>
<td>0.00</td>
<td>21.00</td>
</tr>
<tr>
<td>Office Supplies</td>
<td>0.09</td>
<td>0.87</td>
<td>-9.00</td>
<td>0.00</td>
<td>94.00</td>
</tr>
<tr>
<td>Furniture</td>
<td>0.00</td>
<td>0.12</td>
<td>-3.00</td>
<td>0.00</td>
<td>14.00</td>
</tr>
</tbody>
</table>
4. Empirical Model

4.1. Model Setup

We consider a setup following the potential outcome framework (Rubin 1974) where there are $N$ consumers, indexed by $i = 1, \ldots, N$. Let $Y_i \in \mathbb{R}$ denote the outcome variable – consumer's offline spending in a randomly chosen target week (May 13–19, 2013), which is a common week without any special holiday events.\(^3\) Let $A_i \in \{0,1\}$ denote a binary treatment (e.g., receiving one more email ad), whose effect is in question. In particular, the treatment group will consist of consumers receiving $k + 1$ email ads the target week, and the control group will consist of consumers receiving $k$ email ads the same week ($k = 0,1,2,3$). These two groups allow us to estimate the changing marginal effect of an email ad (i.e., the effect of sending one additional email ad) on the consumer's offline purchase. Importantly, the number of email ads a consumer received every week during the observation period differs over time. For example, a consumer who received three emails the second week might have received one email the week before. Such variation might be due to the changing targeting rules of the retailer over time or some explicit randomness in the email campaign.\(^4\) Moreover, all the existing consumers had opted in for receiving emails from the retailer and did not opt out of email communication during the observation period; however, the consumers could not decide how often they received emails, which provides some exogenous variation in the number of promotional emails they received every week. We summarize the distribution of email ads that were sent in bulk by the retailer in the target week in Table 3. While some consumers receive more than 4 email ads, their fraction is relatively small.

\(^3\) We perform the analysis with other weeks as the target week and obtain qualitatively similar results.

\(^4\) E.g. the firm might implement rules such as send a campaign to 3 million random customers.
Table 3. Number of Email Ads Received by Consumers in the Target Week

<table>
<thead>
<tr>
<th>Number of Email Ads Received in the Target Week</th>
<th>Number of Consumers</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>27,324</td>
</tr>
<tr>
<td>1</td>
<td>18,353</td>
</tr>
<tr>
<td>2</td>
<td>3,913</td>
</tr>
<tr>
<td>3</td>
<td>11,720</td>
</tr>
<tr>
<td>4</td>
<td>43,718</td>
</tr>
</tbody>
</table>

We postulate the existence of a pair of potential outcomes for each consumer, \((Y_{i}^0, Y_{i}^1)\), which are the outcomes that would have been observed for a particular consumer \(i\) under control \((A_i = 0)\) and treatment \((A_i = 1)\), respectively. Note that we consider four distinct treatments here: (i) \(A^1\): sending the email ad to customers who have received no emails, (ii) \(A^2\): sending the second email ad to customers who have received the first ad, (iii) \(A^3\): sending the third email ad to customers who have received two ads, and finally, (iv) \(A^3\): sending the third email ad to customers who have received two ads.\(^5\) The realized outcome for consumer \(i\) is the potential outcome corresponding to the treatment received:

\[
Y_{i}^{obs} = \begin{cases} 
Y_{i}^0 & \text{if } A_i = 0, \\
Y_{i}^1 & \text{if } A_i = 1.
\end{cases}
\]

We also observe a substantial amount of data about consumers, which is denoted by \(L_i\), a \(D\)-dimensional vector of covariates \((D = 189)\). The covariates \(L_i\) are generated from consumers’ historical purchase records in each product category (e.g., weekly spending amount and purchase quantity in product categories such as Business Machines, Food, Office Supplies, Furniture, etc.), their coupon usage behavior (e.g., quantity and amount of coupons used), and their advertising exposures and responses (e.g., number of email ads received, number of times they opened email ads, number

\(^5\) The rational for choosing the incremental email as a treatment is because the consumers who receive no emails are relatively similar to consumers who receive 1 email, but are considerably different from consumers who receive 3 emails. If we matched consumers who receive no emails to consumers who received 3 emails, we run into issues of data sparsity. Due to this choice, our results should always be interpreted as the effect of yet another email ad and not the net effect of receiving a certain number of email ads.
of display ads impressions) in each of the previous nine weeks before the target week. In total, there are 189 covariates generated from historical data of each consumer’s behavior.

Summarizing, our data consist of $Z_i = (Y_i^{obs}, A_i, L_i)$, for $i = 1, \ldots, N$. We assume that the observed data consist of an independent and identically distributed sample of $Z = (Y, A, L)$ according to some unknown probability distribution $P_0$ so that there is no interference between consumers, which is often called the consistency condition, formalized as $A = a \Rightarrow Y = Y^a$. Moreover, we suppose that each consumer has a non-zero probability to receive treatment level $A = a$ such that $p(A = a | L = l) > 0$ whenever $p(L = l) > 0$, which is often called the positivity condition.

The average treatment effect (ATE) is then given by

$$\psi = \mathbb{E}[Y^1 - Y^0] = \mathbb{E}(\mu(1) - \mu(0)),$$

where $\mu(a) = \mathbb{E}(Y|A = a)$. Several additional assumptions should be met for us to correctly estimate the ATE correctly. The first condition that should be satisfied is the Stable Unit Treatment Value Assumption (SUTVA), i.e. one customer should get exactly one treatment and it should impact the outcome of any other customer. The second important condition that should be satisfied is that the treatment is randomly assigned to customers. While the first assumption is satisfied in our condition, advertisers do not send ads to consumers randomly and the treatment might be endogeneous.

This endogeneity can arise due to three important factors as summarized by Gordon et al (2019) – (i) user activity, (ii) firm targeting, and (iii) competition bias. Activity bias arise due to the fact that some consumers might receive more ads as they appear online more often, but are also more likely to convert online. Similarly, advertisers might show a higher number of ads to consumers who are more likely to convert and ATE might be biased upwards as a consequence. Finally, most online ads are shown only after an advertiser wins an ad auction, and a selection might be introduced based on the competitors’ strategy. In our context of email ads, activity bias and competition bias are not relevant as email ads are sent directly to consumers. However, the email ads could potentially be targeted based on individuals’ characteristics and past buying behavior. As a consequence, we need to adjust for potential confoundedness to estimate the ATE.

4.2. Experimental v/s Observation Data

Estimating the causal impact of advertising has been a matter of significant interest in the marketing literature. Typically, researchers have used one of two approaches: (i) randomized
experiments and (ii) matching estimators such as PSM on observational data. In recent times, there has been considerable debate on the appropriateness of the two approaches in the literature. There are several papers that discuss this issue but two important papers, by Eckles and Bakshy (2017) and Gordon et. al (2019), offer opposing viewpoints in this literature.

On one hand, Eckles and Bakshy (2017) suggest that the estimation on high-dimensional observational data returns estimations that are extremely similar to estimates from a randomized experiment. Particularly, their evaluation results have shown that high-dimensional models adjusting for detailed records of individuals’ past behaviors can almost entirely remove the bias. Their paper demonstrates that high-dimensional datasets and statistical learning techniques can be used to significantly improve causal inference using observational data, which avoids the potential high costs and infeasibility of randomized field experiments in various settings. On the other hand, Gordon et. al (2019) argue that many unobservable factors make exposure to advertising endogenous. As a result, observational methods such as propensity score matching might not be able to overcome the data limitation and suffer from some of the biases mentioned in Section 4.1. They examine 15 large scale experiments and conclude that estimates from commonly used observational techniques are significantly different from the experimental estimates, and hence, biased. In their context, this bias arises because the treatment is mostly driven by unobserved variables and the observed features are unable to account for a significant variation in the treatment across individuals. Both these papers use data from Facebook, but there is an important distinction. The number of features (3700) considered in Eckles and Bakshy (2017) considerably outstrips the features (54) considered in Gordon et. al (2019). However, both these papers concur that the bias in the observational estimates is essentially a function of the effect of unobserved heterogeneity on the treatment, the more correlated the omitted variables are with the treatment and the observed outcomes, higher the bias.

Although Gordon et. al (2019) argue that the bias in the observational studies arise due to the correlation of the treatment with the unobserved heterogeneity, this bias might also arise due to another. More specifically, it can arise because of the curse of dimensionality associated with commonly used matching estimators as documented in the computer science literature (Li et. al, 2016). Li et. al (2016) provide a comprehensive survey of the issue associated with matching estimators such as PSM, and show the extent of the bias increases as the dimensionality of the data increases. We also verify this finding in our data. Following the design in Frolich et al. (2015), we test the curse of dimensionality in the following experiment.
We create a synthetic dataset by matching each treated sample (i.e., the customer who received one email ad to its counterfactual in the control group (i.e., customers who received no email ad) using PSM on the 198-dimensional vector \(L\). The treated samples are then discarded and do not play any further role in the experiment. The matched control samples form the pseudo-treated “population”, and the remaining control samples form the control “population”. Then, we repeatedly draw samples with replacement out of the “populations”, consisting of 50% pseudo-treated and 50% controlled units. We choose consumer’s offline spending in that week as the response variable. Note that the true causal effect in this experiment is zero, as the pseudo-treated units did not (by default) receive any treatment. Using a PSM estimator, we estimate that the treatment effect is $3.01 [1.44, 4.72], which is significantly larger than zero (ground truth). This indicates that parametric matching techniques can be biased due to the curse of dimensionality, and not only due to confoundedness (which is the only reason Gordon et. al (2019) point out). Hence, we concur with Gordon et. al (2019) that traditionally used observational methods might not be appropriate, but they might be inappropriate not only due to omitted variables, but also due to the high-dimensionality of the data, particularly in the context of marketing big data.

Nevertheless, we argue that the observational data used in this study is appropriate for estimating the causal effect of email ads. In order to do so, it is important that, conditioned upon a set of observed pretreatment covariates \(L\), the treatment is independent of potential outcomes. This no unmeasured confounding condition is formalized as \(A \perp Y^a | L\). In our setting, we use a large number of pretreatment covariates generated from detailed records of consumers’ past behavior to adjust for potential confoundedness as part of our estimation strategy for drawing causal inferences using observational data. Based on our discussion with our data provider and managers at the focal retail, this proprietary dataset includes all the information this retailer had about these consumers during the observation period. Furthermore, marketing tools provided by our partner firm relied solely on data from the prior 6 weeks to make advertising decisions. Therefore, our empirical models are able to account for all the observable characteristics and behaviors on consumers that the retailer considered when sending out email ads in bulk. These characteristics of our data generating process greatly reduces the threat of targeting bias in our estimation and enhances the validity of our causal statements.\(^6\) Furthermore, to test the threat of omitted variables, we employ an approach suggested

\(^6\) Another reason that the proposed approach should work better in the context of email ads is that unlike display or search advertising where there is a lot of heterogeneity in unobserved factors that drive user behavior. On the other hand, email
by Gordon et. al (2019), which tests how much of the variation in the treatment can be explained by omitted variables. We regress the treatment $A$ on the observed covariates $L$ and observe that most of the variation in the treatment is explain by the observed variables ($R^2 = 0.85$). This indicates that the effect of the omitted variables on the treatment cannot be too large, and the no confounding assumption is likely satisfied. Also, based on Eckles and Bakshy (2017), we can conclude that our data can be used for causal analysis as (i) it contains significant number of covariates that captures the variation in the treatment effect, and (ii) past purchase behavior is indicative of future purchase behavior as pointed out in the prior literature (Dekimpe and Hanssens, 1995).

Therefore, we use a high-dimensional dataset generated from detailed consumer behaviors in the past to adjust for potential confoundedness (using ensemble machine learning methods), and estimate the causal impact of advertising. Note that this is similar in spirit to propensity score matching (PSM) that is quite commonly used in the marketing literature, however the proposed estimator does not suffer from the curse of dimensionality as explained below.

### 4.3. Identification Strategy

We select a doubly robust estimator (DRE) that incorporates machine learning methods to estimate average treatment effects and heterogeneous or conditional treatment effects of email ads on consumers’ offline purchase using a unique high-dimensional consumer-level dataset (Kennedy 2016, Chernozhukov et al. 2017, van der Laan 2013). In recent years, researchers have made new developments in econometrics for addressing causality issues using machine learning methods (Athey and Imbens 2017). Machine learning methods provide important new tools to improve estimation of causal effects in high-dimensional settings. As high-dimensional datasets are becoming commonplace in many research domains, such as digital marketing, it is important for researchers to flexibly control for a large number of covariates. However, even though statisticians and econometricians have made significant progress on modern methodologies, to the best of our knowledge, most of these state-of-the-art approaches are highly underused in the marketing literature. DRE is considered as one of the most effective causal estimation methods in observational studies (Athey et al. 2017, Athey and Imbens 2017). It has many appealing properties for making causal inference in our setting.
First, DRE helps protect against bias due to model misspecification. It relies on two nuisance functions: one for the conditional mean of outcomes (modeling the association between the potential outcomes and the covariates) and one for the propensity score (modeling the association between the treatment indicator and the covariates). As long as the estimation for either of them is consistent, the resulting estimator for the average treatment effect is consistent, thus making it “doubly robust” (Robins and Rotnitzky 1995). The non-parametric nature of our estimator addresses a strong concern about model uncertainty typically associated with online advertising pointed out by Lewis and Rao (2015).

Second, we can use nonparametric machine learning methods to estimate the nuisance functions (i.e., using ensemble models to predict propensity scores and the outcome variables), which allows the data-generating process to be completely unrestricted. We can thus avoid relying on unrealistic parametric distributions required in many traditional approaches such as maximum likelihood estimators and Bayesian models. Moreover, by using machine learning methods, we can also reduce the sensitivity to the “curse of dimensionality” by flexibly controlling for a large number of covariates as part of our identification strategy for making causal inferences (Kennedy et al. 2017, Eckles and Bakshy 2017).

Third, DRE achieves efficiency in the sense of reaching the semi-parametric efficiency bound of Hahn (1998). It will converge at the fastest possible rate and be approximately unbiased and normal, allowing straightforward construction of valid confidence intervals for the parameters of interest, even when we use flexible nonparametric machine learning estimates for nuisance functions. Overall, all these appealing properties of DRE significantly enhance the credibility of our causal analysis made from the observational dataset.

5. Average Treatment Effect

DRE is based on the concept of the influence function to estimate average treatment effects of email ads on offline and online sales. The concept of the influence function is a foundational object of statistical theory that allows us to characterize a wide range of estimators and their efficiency. There is a deep connection between asymptotically linear estimators for a given model and the influence functions under that model. In some sense, if we know one then we know the other. Thus, if we can find all the influence functions for a given model, we can characterize all asymptotically linear estimators for that model. For a comprehensive review on semiparametric theory and empirical
processes in causal inference, we refer to Kennedy (2016). In the following part, we briefly discuss the influence function for average treatment effect (ATE), and then we discuss how to construct the estimator for ATE using its efficient influence function.

5.1. Influence Function

Our target parameter is the average treatment effect (ATE):

$$\psi = \mathbb{E}[Y^1 - Y^0] = \mathbb{E}\{\mu(L, 1) - \mu(L, 0)\},$$

where $Y$ denotes the outcome variable (i.e., consumer's offline spending in a randomly chosen target week), and $A \in \{0, 1\}$ denotes a binary treatment (i.e., 1 denotes receiving the treatment and 0 otherwise). We let

$$\mu(l, a) = \mathbb{E}(Y | L = l, A = a)$$

denote the outcome regression function (i.e., using predictive models to predict the outcome variable such as consumer’s offline spending given a set of covariates and the treatment assignment). Also let

$$\pi(l) = P(A = 1 | L = l)$$

denote the propensity score (i.e., using predictive models to predict the probability of a consumer being assigned to the treatment group given a set of covariates). Therefore, the confounding factors $L$ affect the treatment variable $A$ via the propensity score $\pi(l)$ and the outcome variable via the function $\mu(l, a)$. Both of these nuisance functions are unknown and potentially complicated. We consider estimating them via the use of nonparametric machine learning methods (i.e., using ensemble modeling to construct the nuisance functions), which allows the data-generating process to be completely unrestricted, and can flexibly control for a large number of covariates.

It has been shown that under a nonparametric model of $Z = (Y, A, L) \sim P_0$ where the distribution $P_0$ is unrestricted, the efficient influence function for our target parameter $\psi$ is given by

$$\varphi(Z; \psi, \eta) = m_1(Z; \eta) - m_0(Z; \eta) - \psi,$$

where

$$m_a(Z; \eta) = m_a(Z; \pi, \mu) = \frac{I(A = a)\{Y - \mu(L, a)\}}{a\pi(L) + (1 - a)(1 - \pi(L))} + \mu(L, a),$$

with $\eta = (\pi, \mu)$ being the nuisance functions for this problem. For a full proof of the efficient influence function for $\psi$, we refer to Kennedy (2016). Another closely related work, by Chernozhukov
et al. (2017), has also demonstrated this using Neyman orthogonal scores, and proposed a closely related estimator named double machine learning estimator, which is essentially a DRE.

Let \( \mathbb{P}_n = n^{-1} \sum \delta_{Z_i} \) denote the empirical distribution of the data, where \( \delta_z \) is the Dirac measure that simply indicates whether \( Z = z \). It has been shown that \( \hat{\psi} = \hat{\psi}(\mathbb{P}_n) \) is asymptotically linear with influence function \( \varphi \) if the estimator can be approximated by an empirical average in the sense that

\[
\hat{\psi} - \psi = \mathbb{P}_n\{\varphi(Z)\} + o_p(1/\sqrt{n}),
\]

where \( \varphi \) has mean zero and finite variance (i.e., \( \mathbb{E}\{\varphi(Z)\} = 0 \) and \( \mathbb{E}\{\varphi(Z)^2\} < \infty \)). Here \( o_p(1/\sqrt{n}) \) employs the usual stochastic order notation so that \( X_n = o_p(1/r_n) \) means \( r_nX_n \xrightarrow{p} 0 \) where \( \xrightarrow{p} \) denotes convergence in probability. Importantly, by the classical central limit theorem, our estimator \( \hat{\psi} \) with influence function \( \varphi \) is asymptotically normal with

\[
\sqrt{n}(\hat{\psi} - \psi) \xrightarrow{d} N(0, \mathbb{E}\{\varphi(Z)^2\}),
\]

where \( \xrightarrow{d} \) denotes convergence in distribution. Hence, if we know the influence function for an estimator, we know its asymptotic distribution, and we can easily construct confidence intervals and hypothesis tests.

The above discussion shows that given an estimator \( \hat{\psi} \), we can learn about its asymptotic behavior by considering its influence function \( \varphi(Z) \). On the other hand, if we find the influence function for a given model, we can use it to construct an estimator for that model. Therefore, now we can use the known influence functions to construct an estimator for our target parameter \( \psi \). Given the known efficient influence function \( \varphi(Z; \psi, \eta) \) that depends on the target parameter \( \psi \) as well as a nuisance parameter \( \eta = (\pi, \mu) \), we can construct an estimator by solving the estimating equation \( \mathbb{P}_n\{\varphi(Z; \psi, \eta)\} = 0 \) in \( \psi \), where \( \hat{\eta} \) is estimated via the use of nonparametric machine learning methods. After some simple algebra, the doubly robust estimator is given by

\[
\hat{\psi} = \mathbb{P}_n\{m_1(Z; \hat{\eta}) - m_0(Z; \hat{\eta})\}
\]

with

\[
m_a(Z; \eta) = m_a(Z; \pi, \mu) = \frac{I(A = a)(Y - \mu(L, a))}{a\pi(L) + (1 - a)(1 - \pi(L))} + \mu(L, a).
\]
In observational studies, the covariates $L$ are usually high-dimensional, and little is known about the propensity score and outcome regression functions $\pi$ and $\mu$, in which case it makes sense to use flexible, nonparametric, data-adaptive methods to estimate them. In our case, we use ensemble modeling to construct the nuisance function estimator $\hat{\eta} = (\hat{\pi}, \hat{\mu})$. This estimator optimally combines three machine-learning methods by estimating the nuisance function as weighted averages of estimations from Lasso, regression tree, and random forest. The weights are restricted to sum to 1 and are chosen so that the weighted average of these methods gives the lowest average mean squared out-of-sample prediction error estimated using cross-validation. It can be easily shown that as long as either $\hat{\pi}$ or $\hat{\mu}$ is consistent, the resulting estimator $\hat{\psi}$ for the average treatment effect is consistent, thus making it “doubly robust”, which protects against bias due to model misspecification (Kennedy 2016).

5.2. Estimation Process

Based on the constructed estimator $\hat{\psi}$, we make use of sample-splitting in our estimation process, which helps avoid overfitting that can easily result from the application of complex flexible machine learning methods (Chernozhukov et al. 2017). The specific estimation steps are the following.

Step 1: Let $K$ be a fixed integer. Form a $K$-fold random partition of $\{1, \ldots, N\}$ by dividing it into equal parts $I_k^K_{k=1}$, each of size $N/K$, assuming that $N$ is a multiple of $K$. For each set $I_k$, let $I_k^c$ denote all observation indices that are not in $I_k$.

Step 2: Construct $K$ estimators

$$\hat{\psi}(I_k, I_k^c), \quad k = 1, \ldots, K,$$

that employ the machine learning estimators for the nuisance functions

$$\hat{\eta}(I_k^c) = (\hat{\pi}(l; I_k^c), \hat{\mu}(l, a; I_k^c)),$$

where each estimator $\hat{\psi}(I_k, I_k^c)$ is defined as the root $\psi$ of

$$\frac{1}{n} \sum_{i \in I_k^c} \hat{\phi}(Z; \psi, \hat{\eta}(I_k^c)) = 0,$$

which is the estimating equation $\mathbb{P}_n \{\phi(Z; \psi, \hat{\eta}) = 0 \text{ for } \mathbb{E}\{\phi(Z)\} = 0.$

Step 3: Average the $K$ estimators to obtain the final estimator:
\[ \hat{\psi} = \frac{1}{K} \sum_{k=1}^{K} \hat{\psi}(l_k, I_k). \]

An approximate standard error for this estimator is \( \hat{\sigma}/\sqrt{N} \), given by the asymptotically normal distribution of \( \hat{\psi} \), with

\[ \hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^{N} \hat{\varphi}_i^2, \]

\[ \hat{\varphi}_i = \hat{\varphi}(A_i, \hat{\psi}, \hat{I}_{k(i)}), \] and \( k(i) = \{k \in \{1, ..., K\}: i \in I_k\} \). An approximate \((1 - \alpha) \times 100\) percent confidence interval is given by

\[ \hat{\psi} \pm \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \frac{\hat{\sigma}}{\sqrt{N}}. \]

### 5.3. Results

In Table 4, we report our main results for the average treatment effects (ATE) of email ads on consumers’ purchase behaviors. We report results based on ensemble machine learning modeling for estimating the nuisance functions used in forming the influence functions’ estimating equations. The raw set of covariates is used as features in the machine learning models. We present results based on 5-fold cross fitting. To reduce the disproportionate impact of extreme propensity scores in the model, we trim the propensity scores with the cutoff points of 0.05 and 0.95.\(^7\)

<table>
<thead>
<tr>
<th>Table 4. Estimation Results for the Average Treatment Effects of Email Ads on Offline Sales</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Control Group</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>(k = 0)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>(k = 1)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>(k = 2)</td>
</tr>
</tbody>
</table>

\(^7\) We also test different choices of cutoff points in robustness checks.
Turning to the results in Table 4, we first report the OLS estimates for ATE of email ads on offline sales in column (1). The OLS estimate suggests that there is no effect of the first email, but a significant positive effect of the second email and a significant negative effect of the third email. These estimates are clearly not valid, as there are neglected confounding variables. We show these estimates to serve as a reference and point out that the magnitude and the direction of the bias are quite significant. When we turn to the estimates that flexibly account for confounding factors using DRE reported in column (2), we see that they are substantially different from the baseline that does not account for confounding. On average, receiving one email ad has a small negative effect on the consumer’s offline spending but this effect is weakly significant. Similarly, receiving two email ads does not have a significant effect. However, receiving three email ads significantly increases the consumer’s offline spending by approximately $11.82. We would like to point out that these effects are significantly higher in magnitude as compared to the effect of display ads estimated by Lewis and Reiley (2014). This disparity directly points to the effectiveness of email ads over display ads, as both Lewis and Reiley (2014) and our study look at the effect of advertising on existing customers. The sudden increase in the sales from the third ad suggest a potential reminder effect of email ads that works after a certain threshold: Receiving more emails from the retailer gradually attracts the consumer’s attention and registers the retailer’s various products in the consumer’s memory, which can make the consumer more likely to consider the retailer’s products on the next purchase occasion. We believe the threshold effect exists because the advertising stock needs to build up to a certain high level before the consumer is motivated enough to walk into the brick-and-mortar store. Finally, we observe that the effect of email ads plateaus out, and the fourth ad has no effect on offline sales. The average effect of an email ad sent during this period is increasing weekly sales by $1.49.

From a managerial perspective, it is important to note the discrepancy between the OLS and DRE. For example, the firm is more likely to send the second email to consumers who are more likely to convert, and not accounting for this targeting will result in biased estimates. This would result in a suboptimal market plan, hurting firm’s ROI from email advertising.
A unique contribution of this research is that we use consumers’ actual purchase amount both online and offline to examine the causal effect of email ads. To our knowledge, in prior literature on email marketing, most studies use certain proxies instead of actual sales, such as number of purchase (Zhang et al. 2017), email opening and click-through (Wattal et al. 2012), and the propensity of a consumer getting in and out of email marketing subscription, except that Sahni et al. (2017) uses actual online ticket sales. We believe it is imperative to provide evidence on the effect of email ads on actual sales both online and offline, as consumers often go through a purchase funnel during their shopping process. Even though we know email ads may increase consumers’ engagement in initial stages, an increase in consumer engagement with email ads does not necessarily guarantee a proportional increase in the retailer’s revenue due to the well-known purchase funnel effect. Thus this research addresses this important gap in the marketing literature. Furthermore, unlike existing literature, we examine the effect of a cadence of email ads and identify the threshold and satiation effects as the consumers receive a sequence of ads.

To better understand the mechanism driving the results in column (2), we conduct four additional analyses, presented in Table 5 where we use the same DRE with different dependent variables.

### Table 5. Estimation Results for the Average Treatment Effects of Email Ads on Consumer Behaviors

<table>
<thead>
<tr>
<th>Control Group</th>
<th>Treated Group</th>
<th>Doubly Robust Estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td>(k email ads)</td>
<td>(k+1 email ads)</td>
<td>(1)</td>
</tr>
<tr>
<td>Coupon Redeemed</td>
<td>Purchase Incidence</td>
<td>Product Variety</td>
</tr>
<tr>
<td>k = 0</td>
<td>k + 1 = 1</td>
<td>0.14*</td>
</tr>
<tr>
<td>k = 1</td>
<td>k + 1 = 2</td>
<td>-0.70** (0.25)</td>
</tr>
<tr>
<td>k = 2</td>
<td>k + 1 = 3</td>
<td>-4.78*** (0.63)</td>
</tr>
<tr>
<td>k = 3</td>
<td>k + 1 = 4</td>
<td>0.28* (0.13)</td>
</tr>
</tbody>
</table>
Column (1) reports the effect of email ads on the amount of coupon values redeemed by the consumer. Note that the value of coupons redeemed is represented in negative numbers in transaction records, which will affect the interpretation of the results accordingly. Since the retailer uses email ads to directly send promotional information to consumers (i.e., coupon codes), if these promotions successfully attract consumers’ attention and increase purchase, we should expect to see that consumers receiving more email ads are more likely to use coupons in their purchase. In our analysis, we find that receiving email ads gradually increases the value of coupons redeemed by the consumer before tapering off after the third email. On average, receiving two email ads slightly increases a consumer’s coupon redemption by $0.70, and receiving three email ads, by $4.78 approximately. These findings are consistent with our expectation of a reminder effect. The effect for the fourth email is (weakly) positive, which we believe is capturing the satiation effects and potential aversion effect when receiving too many ads in the same week.

Column (2) reports the effect of email ads on consumers’ purchase incidence, where we create a binary dependent variable, with 1 indicating a consumer has made at least one offline purchase in the target week, and 0 otherwise. We find that email ads gradually increase the consumer’s offline purchase incidence, as expected. However, there is a sharp increase after three emails and the third email increases the consumer’s offline purchase incidence by 0.235, pointing to the threshold effect mentioned earlier. Interestingly, by combining the results in columns (1) and (2), we find that receiving two email ads slightly increases the consumer’s purchase probability, by 0.034, suggesting a small reminder effect with positive impact on offline sales, and slightly increases a consumer’s coupon redemption amount, by $0.70, which has a negative impact on offline revenues. Hence, these two effects together may explain the results in column (2) in Table 4: why receiving two email ads slightly increases the consumer’s purchase probability, but does not significantly affect offline sales. On the other hand, the third email leads to a significantly high purchase incidence and a much higher coupon redemption value, but the increased incidence is more than able to make up for the high redemption coupon value resulting in an overall increase in spending by $11.82. Finally, the purchase incidence decreases with the fourth email ad indicating a potential aversion effect that has been documented in the literature (Braun and Moe, 2013).

Column (3) reports the effect of email ads on product variety in the consumer’s purchase, where we use a dependent variable indicating the number of unique product categories the consumer has purchased in the target week, with the maximum value, 9, indicating the consumer has purchased
at least one product in each of the nine different categories (see Table 1 for a summary of product categories). We find that receiving three email ads significantly increases the consumer’s purchased product variety as the average number of categories purchased increases by 0.242. This finding suggests that the incremental offline sales result from not only an increased purchase probability but also a wider variety of products being purchased by the consumer. However, this effect comes into play only when the consumer receives a substantial number of ads.

Finally, we also examine the effects of email ads on online sales for comparison in column (4). The effect of email ads on online sales is in general smaller as compared to offline ads, even though it follows an increasing pattern similar to offline sales. The first email ad has a (weak) positive effect on online sales, unlike on the offline channel (where the effect is negative). This observation might be due to the fact that after reading their email, it is easier for the consumers to purchase a product online, but quite unlikely for them to walk into a store to make a purchase. The effect of the second and third emails is insignificant, but the fourth email has a strong positive effect on sales, suggesting an accumulation effect (Braun and Moe, 2013) that surpasses a higher threshold that is required for (online) spending.

6. Heterogeneous Treatment Effect

In many cases, when researchers are evaluating a policy or treatment, a policy or treatment might have different costs and benefits if applied in different settings. Gaining insight into the nature of such heterogeneous treatment effects can be useful, and it is important to know the applications where the treatment effects are most favorable. In the digital marketing context, understanding the heterogeneity in different consumer segments has been playing a central role in many marketing activities because the effect of online advertising often varies across different consumer segments. Not accounting for such heterogeneity can lead to potential biased results on the effectiveness of online advertising. Unfortunately, existing studies on the causal effects of online advertising typically report average effects and do not allow for heterogeneity (Danaher and Dagger 2013, Lewis and Reiley 2014), or simply predetermine consumers into binary categories (Lobschat et al. 2017), possibly due to data limitations.

In this section, we leverage the power of big data to examine two questions: (1) Is the causal effect of email ads on offline sales heterogeneous across different consumer segments? (2) If so, on which consumer segment is the effect highest? Answering these questions can help retailers to more
accurately evaluate the effectiveness of their marketing campaigns and optimize their marketing strategies in the future.

We select the approach proposed by van der Laan (2013) to estimate heterogeneous effects based on DRE and regression trees. The standard approach for partitioning consumers into groups would be to include interaction terms between different covariates in a regression model. This method works well when there are a small number of covariates, but quickly breaks down when the number of consumer characteristics is very large. Moreover, adding interaction terms also relies on assumptions of linear additivity and unrealistic distributions on the error term, both of which would reduce the credibility of causal inferences made from large-scale observational data. Hence, in this research, we select the loss-based machine learning method proposed by van der Laan (2013). This method is a data-driven approach to partition our high-dimensional data into subpopulations that differ in the magnitude of the treatment effects of email ads. As this approach is purely non-parametric, it does not assume any unrealistic parametric distributions for consumer heterogeneity.

Our approach is closely related to Athey and Imbens (2016), who propose an approach named causal trees that adapts the machine learning method of regression trees. The difference between these two approaches is that ours accounts for potential confoundedness in observational data using DRE, whereas the causal tree approach is for randomized experimental studies and needs to be modified before it can be applied to observational studies. Next, we discuss the loss function proposed in van der Laan (2013), and then the loss-based machine learning method for estimating heterogeneous treatment effects.

6.1. Loss Function

Let $V$ be a function of $L$. Our target parameter is the heterogeneous treatment effect, or conditional treatment effect:

$$\psi(V) = \mathbb{E}[Y^1 - Y^0 | V].$$

Define the loss-function

$$L_{\pi, \mu}(\psi)(Z) = (D_1(\pi, \mu)(Z) - \psi(V))^2.$$ 

This loss function is indexed by nuisance parameters $(\mu, \pi)$ required to evaluate

$$D_1(\pi, \mu) = \frac{2a - 1}{a \pi(L) + (1 - a)[1 - \pi(L)]} \left( Y - \mu(L, a) \right) + \mu(L, 1) - \mu(L, 0).$$
In fact, \( D_1(\pi, \mu) - \mathbb{E}[Y^1 - Y^0] \) is the efficient influence function for the parameter \( \psi = \mathbb{E}[Y^1 - Y^0] \), and it has the property that \( \mathbb{E}[D_1(\hat{\mu}, \hat{\pi}) \mid V] = \psi(V) \) if either \( \hat{\pi} \) or \( \hat{\mu} \) is consistent (and under the positivity condition). Due to this property, it follows that if either \( \hat{\pi} \) or \( \hat{\mu} \) is consistent, then

\[
L_{\hat{\pi}, \hat{\mu}}(\hat{\psi}) = D_1(\hat{\pi}, \hat{\mu})^2 + \hat{\psi}^2(V) - 2D_1(\hat{\pi}, \hat{\mu})\hat{\psi}(V)
\]

\[
= D_1(\pi, \mu)^2 + \psi^2(V) - 2\psi(V)\hat{\psi}(V)
\]

\[
= (\hat{\psi} - \psi)^2(V) + D_1(\hat{\mu}, \hat{\pi})^2 - \psi^2(V).
\]

Hence, this proves that, if either \( \hat{\pi} \) or \( \hat{\mu} \) is consistent, then the true risk of this loss function \( L_{\hat{\pi}, \hat{\mu}}(\hat{\psi}) \) equals \( (\hat{\psi} - \psi)^2(V) \) up till a constant (not depending on the candidate \( \hat{\psi} \)) so that \( L_{\hat{\pi}, \hat{\mu}}(\hat{\psi}) \) is minimized over \( \hat{\psi} \) by the true \( \psi \). Moreover, it can be shown that the empirical mean of the loss efficiently estimates the true underlying risk, making \( L_{\hat{\pi}, \hat{\mu}}(\hat{\psi}) \) a double robust and efficient loss function for this true underlying squared error risk. That is, \( L_{\hat{\pi}, \hat{\mu}}(\hat{\psi}) \) (and its cross-validated counterpart as used in cross-validation) is a double robust locally efficient estimator of the true underlying risk (up till the irrelevant constant), under regularity conditions.

### 6.2. Estimation Process

The loss function \( L_{\hat{\pi}, \hat{\mu}}(\hat{\psi}) \) shows that we can apply any least-squares regression algorithm to regress \( D_1(\hat{\pi}, \hat{\mu})(Z) \) on \( V \) (van der Laan 2013). In this research, we apply the regression trees approach (Breiman et al. 1984), a highly data adaptive machine learning algorithm, to estimate heterogeneous treatment effects. Among common machine learning methods, regression trees are a natural choice for partitioning into subgroups (Athey and Imbens 2017). Consider a regression with two covariates. The value of each covariate can be split so that it is above or below a certain level. The regression tree approach would consider which covariate should be split, and at which level, so that the sum of squared residuals is minimized. With many covariates, these steps of choosing which covariate to split, and where to split it, are carried out sequentially, thus resulting in a tree format. The tree eventually results in a partition of the data into groups, defined according to values of the covariates, where each group is referred to as a leaf. In the simplest version of a regression tree, we would stop this splitting process once the reduction in the sum of squared residuals is below a certain level. To better interpret the splitting rules, we have created a number of RFM variables, including recency indicators (whether the consumer has made a purchase in recent \( n \) weeks), frequency variables
(the total number of weeks that the consumer has made a purchase), monetary values categories (whether the consumer is a high-, low-, or no-spending consumer in recent n weeks), and several categorical variables indicating the frequency of the consumer’s interactions with previous advertising campaigns (i.e., receiving email ads in previous weeks, email ads opening behaviors, display advertising exposures, coupon usage frequency). These variables have been commonly used in the prior marketing literature.

When machine-learning methods are applied to estimating heterogeneous treatment effects, they in effect search over many covariates and subsets of the covariate space for the best fit. As a result, such methods may lead to spurious findings of treatment effect differences. In clinical medical trials, pre-analysis plans must be registered in advance to avoid the problem that researchers will be tempted to search among groups of the studied population to find one that seems to be affected by the treatment, and may instead end up with spurious findings. In social sciences, the problem of searching across groups becomes more severe when there are many covariates. To avoid these potential spurious results, we rely on sample-splitting, which is similar to the honest estimation used in Athey and Imbens (2016): One sample is used to determine the optimal partition of the covariates space (the tree structure), and another to estimate treatment effects for each subpopulation so that the confidence intervals are valid no matter how many covariates are used in estimation.

Since we also make use of sample splitting in the estimation process for nuisance parameters, as described before, we divide our sample into three equal parts as our estimation process includes three procedures. Specifically, in the first step, the first sample part is used to estimate nuisance functions based on ensemble machine learning modeling. The estimated nuisance functions, including the propensity score $\hat{\pi}$ and the outcome regression function $\hat{\mu}$, are used to construct the estimator for $D_1(\pi, \mu)$ in the second and the third sample parts. In the second step, the second sample part is used to construct the regression tree that determines the optimal partition of the covariates space. In the last step, based on the constructed tree structure from the second step, we use the third sample part to estimate heterogeneous treatment effects and conduct hypothesis tests about the magnitude of differences in treatment effects across subsets of the population. The sample-splitting process ensures that the confidence intervals are valid for the estimated heterogeneous treatment effects.
6.3. Results

For managers, it is of utmost interest to know how the effect of online advertising on their offline sales varies across different consumer segments. In this section, we discuss the heterogeneity in the causal effects of email ads on offline sales using the loss-based machine learning approach. We construct a regression tree for each control-treated group, respectively, which will partition consumers into identifiable segments according to different splitting rules. The results are displayed in Figures 5–8. Overall, we find significant heterogeneous effect of emails ads on offline sales.

In general, we find that email ads are significantly more effective among consumers with fewer interactions with the retailer in recent weeks. For example, Figure 6 shows that among consumers in leaf 5, those with lower email opening frequency in the recent three weeks, the effect is 8.09, although the ATE of receiving two email ads on offline sales is insignificant. Moreover, Figure 7 shows that for consumers who received fewer email ads in the recent two weeks (leaf 1), receiving three email ads increases their spending in physical stores by 19.43, significantly larger than the estimated ATE (11.82). Overall, this finding is broadly consistent with Lobschat et al. (2017), who find that banner advertising is most suitable to generate awareness among non-recent online consumers for firms that predominantly sell offline. It is also generally consistent with the findings in Sahni et al. (2017) that show the effect of the offers in emails is significantly higher among individuals who did not transact on an online ticket resale platform.

Figure 5. Treated (Control) Group: 1 (0) Email Ads
Figure 6. Treated (Control) Group: 2 (1) Email Ads
Our findings have direct implications for retail managers to improve their targeting strategies by sending email ads to the right consumers online in order to create an improvement in revenues offline. Based on our estimation results for heterogeneous treatment effect in Figures 5–8, we can easily identify on which consumers the email ads are most effective for increasing offline sales and create a personalized e-mail marketing schedule. If we reallocate these promotional emails to target only consumers who would have been positively affected, then we can create an improvement in offline revenues. For example, Figure 5 shows that for consumers in leaf 3, receiving one email ad decreases a consumer’s offline spending by $2.73, with a 95% confident interval of [0.20, 5.26]. Figure 6 shows that among consumers in leaf 5, receiving two email ads increases a consumer’s offline purchase by $8.09 [5.56, 10.62]. Similarly, Figure 7 shows positive treatment effects of receiving three email ads among consumers in leaf 1 and leaf 4. For these two groups of consumers who only received two email ads, if we send them three email ads instead, then the corresponding revenue increase will be $19.43 [26.90, 21.96] and $77.99 [64.87, 91.11]. Hence, the email marketing strategy of the firm can be personalized based on historical consumer features to maximize consumer sales across channels.

7. Robustness Check

To check whether our model and findings are robust, we have conducted four sets of robustness checks.
Firstly, for the estimation results of the ATE of email ads on offline sales and online sales reported in Table 4 and 5, we use the cutoff points of 0.05 and 0.95 on propensity scores in order to reduce the disproportionate impact of extreme values in the data at the cost of reducing sample size. To test the sensitivity of the main results to the selection of cutoff points, we also use different choices of cutoff points to corroborate the main results. As shown in columns (1)–(4) in Table 6 and Table 7, the results show qualitatively consistent ATE estimates for both offline and online sales, suggesting that our results remain robust to the selection of cutoff points. The difference in results is explained by the 80-20 rule, i.e. a disproportionate number of sales is generated by a few customers and the inclusion or exclusion of these customers can have a significant impact on the estimates of ATE.

Secondly, we conduct a robustness check based on the subset of consumers who were not exposed to display ads during the target week. For the focal retailer, we have learned that email ads are used much more frequently and are thus the focus of this study. As shown in Figure 1, every week about 6 to 8 million consumers received email ads, whereas only 0.2 to 0.5 million consumers were exposed to display ads. However, one may still worry that without controlling for consumers’ same-period exposure to display advertising, the effectiveness of email ads may be overestimated. Therefore, we choose a subset of consumers without same-period display advertising to alleviate this concern. In our sample, there are 105,028 consumers in total, as shown in Table 3. Among them, 3,192 consumers (3%) were exposed to display ads during the target week, whereas 101,836 consumers (97%) were not. Thus, we select this subset of 97% consumers and conduct the same estimation for the average treatment effects of email ads on consumers’ offline and online purchase. The estimation results are reported in column (5) in Table 6 and Table 7. They show qualitatively consistent estimates, suggesting that our main results remain robust.

Table 6. Robustness Check - Offline Sales

<table>
<thead>
<tr>
<th>Control Group</th>
<th>Treated Group</th>
<th>Doubly Robust Estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td>(k email ads)</td>
<td>(k+1 email ads)</td>
<td>(1) [0.03, 0.97] (2) [0.04, 0.96] (3) [0.05, 0.95] (4) [0.06, 0.94] (5) No Display Ads Exposure</td>
</tr>
<tr>
<td>k = 0</td>
<td>k + 1 = 1</td>
<td>-1.29* (0.53) -1.26* (0.53) -1.23* (0.52) -0.94* (0.41) -0.87* (0.38)</td>
</tr>
<tr>
<td>k = 1</td>
<td>k + 1 = 2</td>
<td>-0.37 -0.82 -0.19 0.67 -1.39</td>
</tr>
</tbody>
</table>
Table 7. Robustness Check - Online Sales

<table>
<thead>
<tr>
<th>Control Group</th>
<th>Treated Group</th>
<th>Doubly Robust Estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td>(k email ads)</td>
<td>(k+1 email ads)</td>
<td>(1) (2) (3) (4) (5)</td>
</tr>
<tr>
<td>k = 0</td>
<td>k + 1 = 1</td>
<td>[0.03, 0.97] [0.04, 0.96] [0.05, 0.95] [0.06, 0.94]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.39* 0.45* 0.39* 0.28 0.24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.18) (0.16) (0.16) (0.17) (0.14)</td>
</tr>
<tr>
<td>k = 1</td>
<td>k + 1 = 2</td>
<td>(1.30) (1.39) (1.47) (1.55) (1.43)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.86 0.61 0.31 0.12 -0.39</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.30) (1.39) (1.47) (1.55) (1.43)</td>
</tr>
<tr>
<td>k = 2</td>
<td>k + 1 = 3</td>
<td>(5.80) (6.21) (3.28) (2.05) (2.16)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.45 -1.05 1.09 2.62 3.57</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(5.80) (6.21) (3.28) (2.05) (2.16)</td>
</tr>
<tr>
<td>k = 3</td>
<td>k + 1 = 4</td>
<td>6.03*** 5.81*** 5.61*** 5.28*** 4.50***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.18) (0.18) (0.17) (0.18) (0.14)</td>
</tr>
</tbody>
</table>

Note: Standard errors are in parentheses.
Significance level: *** 0.001, ** 0.01 * 0.05

Thirdly, one of the limitations of this study is that we do not observe the content of the emails during the observation period, which may affect consumers’ response. To alleviate this concern, we conducted a survey on Amazon Mechanical Turk (AMT) using emails’ content data collected in 2017. From our discussion with the data scientist in charge of this dataset, we know that these emails were bulk emails sent to consumers without any personalization of the content. Moreover, since the focal retailer sells more than 1.7 million products and services, the email content is designed in a form of a catalog instead of emphasizing specific products. Figure 9 and 10 illustrate two examples. We...
collected email content data during the same week in 2017 as that of the observation period in our dataset.

We observed five different emails sent out to consumers. In our AMT survey, we thus created five different groups. AMT participants were randomly assigned to one of the five groups (each with a different email), and asked to evaluate both the title and the content of the email. There were 328 effective submitted surveys. The number of effective submissions in each group ranges from 63 to 71. The results showed no significant difference across the five email ads, in terms of both email title and email content. Specifically, ANOVA analysis shows no significant difference in participants’ responses across the five groups, such as their purchase intention, perceived promotion attractiveness, and
willingness to open the email. In addition, there is no significant gender or age difference across the five groups.

Fourthly, to further corroborate that our results and findings are not driven by a specific product’s sale, we estimate the ATE of email ads on each product category’s sales. The results are reported in Table 8. The results show qualitatively consistent findings across most product categories such as food and computer accessories, suggesting that our main findings remain robust.

### Table 8. Estimation Results for the Average Treatment Effects of Email Ads on Consumer Behaviors

<table>
<thead>
<tr>
<th>Control</th>
<th>Treated</th>
<th>Doubly Robust Estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>k email ads</td>
<td>k+1 email ads</td>
</tr>
<tr>
<td></td>
<td>Business Machines</td>
<td>Business Services</td>
</tr>
<tr>
<td>k = 0</td>
<td>-0.75* (0.35)</td>
<td>-0.03 (0.08)</td>
</tr>
<tr>
<td>k = 1</td>
<td>1.13 (0.83)</td>
<td>0.20 (0.43)</td>
</tr>
<tr>
<td>k = 2</td>
<td>0.39 (5.74)</td>
<td>-0.59 (2.69)</td>
</tr>
<tr>
<td>k = 3</td>
<td>0.87 (0.54)</td>
<td>1.56*** (0.17)</td>
</tr>
</tbody>
</table>

Note: Standard errors are in parentheses.
Significance level: *** 0.001, ** 0.01, * 0.05

### 8. Conclusion

This research improves our understanding of the causal effect of email ads on consumer purchase behaviors. Using a unique high-dimensional individual-level dataset, we address two important research questions: (1) What is the causal effect of email ads on consumer’s spending in offline stores? (2) Is the effect of email ads heterogeneous across different consumer segments? If so, on which consumers is the effect most favorable? Overall, this research makes two major contributions.

First, from a managerial perspective, our results shed light on estimating causal effects of email ads on offline sales, which have not been quantified in the extant literature. We show that, in our
setting, receiving email ads has a statistically and economically significant effect on a consumer’s spending in physical stores. On average, receiving email ads from the retailer can increase a consumer’s weekly spending in the retailer’s offline stores by approximately $1.49. What’s more, we have shown that the positive effect of email ads is heterogeneous across different consumer segments. We find that the effect is significantly higher among consumers who have not interacted much with the retailer recently, such as those with lower email ads opening frequency in recent weeks. Contrary to the traditional belief that retailers may want to target consumers who have shown a clear interest in the retailer (i.e., retargeting individuals who have made a recent purchase or visit), our results suggest that retailers should target consumers who might be unaware of or might have forgotten about the retailer recently. Sending email ads to these individuals can remind them of the retailer’s offerings of a wide variety of products and thus increase their purchase probability in the retailer’s physical stores, which can improve the retailer’s offline revenues. This research has direct managerial implications for firms that use email ads to communicate with consumers, especially for retailers that predominantly sell through offline channels. This research also has direct implications for policy makers who are interested in evaluating the economic impact of email advertising, which is a prevalent tool adopted by many businesses.

Second, from a methodological perspective, this research contributes to the marketing literature by introducing DRE that naturally incorporates flexible machine learning methods to estimate average treatment effects and heterogeneous effects using observational datasets. These state-of-the-art statistical methods have been proved to be able to enhance the credibility and robustness of causal inference made from observational datasets. As fine-grained individual-level data become increasingly common due to big data, we hope that researchers can take advantage of these modern approaches to produce more credible and robust causal estimates when addressing research questions using observational datasets readily available in various contexts.

We acknowledge some limitations of our study which give rise to interesting future research directions. First, although our data represent a large sample of the retailer’s existing consumers, these consumers are those who had purchased from the retailer prior to the observation period and do not include newly joined consumers. Future research can thus examine how email ads affect new consumers’ decision-making, as they may have a different mindset during their shopping process compared to the existing consumers, who are already familiar with the focal retailer. Second, we do not observe consumers’ exposure to offline advertising (e.g., TV advertising) during the observation
period. Future research should ideally account for potential interaction effects between online advertising and offline advertising on consumer behaviors at the individual level. Thirdly, we do not observe the nuisance effect of email ads in this study as none of the consumers unsubscribe to the marketing emails (as observed in our data). However, as the intensity of email ads increase, consumers might be likely to unsubscribe, pointing to an interesting direction for future study. Finally, while our findings shed some light on the causal effect of email ads on offline sales in the retail industry, future research may uncover the heterogeneous effect across different industries to yield further insights.

**Acknowledgements**

We sincerely thank Marketing Science Institute and Adobe for supporting this research.

**References**


