HARVARD | BUSINESS | SCHOOL

The Value of Descriptive Analytics: Evidence from Online Retailers

Ron Berman
Ayelet Israeli

Working Paper 21-067

The Value of Descriptive Analytics: Evidence from Online Retailers

Ron Berman
The Wharton School, University of Pennsylvania

Ayelet Israeli
Harvard Business School

Working Paper 21-067

Funding for this research was provided in part by Harvard Business School. Research support was provided by the Mack Institute for Innovation Management.

The Value of Descriptive Analytics: Evidence from Online Retailers*

Ron Berman†   Ayelet Israeli‡

December 2021

*We are grateful for helpful comments from Eva Ascarza, Eli Ben-Michael, Kathy Li, Ilya Morozov, Simha Mummalaneni, Davide Proserpio, Dave Reibstein, Christophe Van den Bulte, Jeff Wooldridge, Lynn Wu, and seminar participants at Ben Gurion University, Brock University, Harvard Business School, Hebrew University of Jerusalem, Northeastern University, Northwestern University, UMass Amherst, VU Amsterdam, and participants at the Marketing Analytics Symposium Sydney, the Marketing Science Conference, the Northeastern Marketing Conference, and the Marketing in Israel Conference. We are indebted to the anonymous analytics service provider that provided us with the data. Shawn Zamechek and Hengyu Kuang provided excellent research assistance. Research support from the Mack Institute for Innovation Management is acknowledged.

†The Wharton School, ronber@wharton.upenn.edu.
‡Harvard Business School, aisraeli@hbs.edu.

The Value of Descriptive Analytics: Evidence from Online Retailers

Abstract

Does the adoption of descriptive analytics impact online retailer performance, and if so, how? We use the synthetic difference-in-differences method to analyze the staggered adoption of a retail analytics dashboard by more than 1,500 e-commerce websites, and we find an increase of 4%–10% in average weekly revenues post-adoption. We demonstrate that only retailers that adopt and use the dashboard reap these benefits. The increase in revenue is not explained by price changes or advertising optimization. Instead, it is consistent with the addition of CRM, personalization, and prospecting technologies to retailer websites.

The adoption and usage of descriptive analytics also increases the diversity of products sold, the number of transactions, the numbers of website visitors and unique customers, and the revenue from repeat customers. In contrast, there is no change in basket size.

These findings are consistent with a complementary effect of descriptive analytics that serve as a monitoring device that helps retailers control additional martech tools and amplify their value. Without using the descriptive dashboard, retailers are unable to reap the benefits associated with these technologies.

Keywords: descriptive analytics, big data, difference-in-differences, synthetic control, e-commerce, online retail, martech.

JEL Classification: C55, L25, M31.

1 Introduction

Marketers are often encouraged to invest in analytics-driven decisions, and indeed survey research by Germann et al. (2013) and Germann et al. (2014) reports a positive association between deploying marketing analytics technology and firm performance. Recently, CMOs have been spending 22%–29% of their budgets on marketing technologies (martech) with martech spending set to increase to $122 billion in 2022.[1] This trend led to research that documents the benefits of adopting popular, yet specific, prescriptive technologies such as retargeting and A/B testing (e.g., Lambrecht and Tucker 2013, Koning et al. forthcoming) and contributes to the literature that shows a general positive benefit for firm performance from deploying analytics (Brynjolfsson et al. 2011a, Brynjolfsson and McElheran 2016, Müller et al. 2018).

Analytics technologies are often classified into four categories (Lismont et al. 2017): (i) descriptive (what happened), (ii) diagnostic (why did it happen), (iii) predictive (what will happen next), and (iv) prescriptive (what should be done about it). While academic research often focuses on the potential benefit from predictive or prescriptive technology that uses sophisticated modeling (Pauwels et al. 2009, Hanssens and Pauwels 2016, Wedel and Kannan 2016, Bradlow et al. 2017), most adopting firms typically deploy descriptive analytics and simple key performance indicators (KPIs) dashboards in the sales and marketing departments (Bughin 2017, Delen and Ram 2018, Mintz et al. 2019). Despite the popularity of descriptive analytics dashboards, little is known about how to interpret and turn descriptive metrics into actionable insights, which raises questions about the value of descriptive analytics.

In this paper we set out to measure and document the value of adopting a marketing analytics dashboard. We utilize detailed panel data from over 1,500 online retailers in a variety of industries to test whether there are benefits from adopting the dashboard, and we analyze these benefits. For each retailer we observe multiple performance outcomes and decisions before and after they adopted the dashboard. We focus on trying to provide a causal estimate of the change in revenue that the retailers experience when they adopt the dashboard, followed by an analysis of the decisions that retailers make as a result of the adoption and how these decisions drive changes in customer behavior. Our results can shed light on the additional capabilities that descriptive analytics enables

[1] According to the 2019 Gartner CMO Spend Survey and the 2017–2019 Forester Marketing Technology Services Outlook report, published in April 2018.

for marketers. For practitioners, we provide a benchmark on the benefits they might expect when using a descriptive dashboard, and we illustrate how to potentially best extract these benefits. For researchers, our results shed light on the importance of simple heuristics and provide insights into the mechanisms that drive the performance gains from these heuristics.

How might descriptive analytics benefit retailers? As mentioned, there is no clear evidence that descriptive analytics helps at all, as most prior research did not have appropriate data or could not distinguish between the types of analytics used by companies. If descriptive analytics does provide benefits for retailers, there are a few potential mechanisms that can lead to increased revenue. Our research aims to distinguish among them. First, descriptive dashboards can aid retailers in adjusting traditional marketing levers such as pricing or advertising in response to the KPIs they observe. For example, the common metric of customer acquisition cost (CAC) can be compared to the profit margin of the retailer and can yield an adjustment in advertising to reduce the CAC. We call this mechanism the "direct mechanism."

Second, the descriptive dashboard may be used as a monitoring tool to assess the value of other operations of the retailer but does not drive decisions directly. For example, if a retailer invests in retargeting campaigns or in website personalization for customers, it would be difficult to measure the effects of these investments and optimize parameters without monitoring potentially affected KPIs. In this case descriptive analytics is used as a complement to other martech to assess and amplify the strategies that work best. We refer to this mechanism as a "complementary mechanism."

Finally, descriptive analytics might not operate at all to generate any value for the retailer but may be correlated with other actions the retailer takes. If, for example, when the retailer integrates a descriptive dashboard they also take simultaneous, unrelated actions (such as changing the store design or hiring new managers), then we might observe a correlation between increased performance and analytics adoption. We refer to this mechanism as an "unrelated mechanism."

Prior attempts to investigate the relationship between data analytics and firm performance found 3%–7% higher productivity for firms that adopt data-driven decision making or big data assets. However, these attempts were often limited by access to non-granular or non-primary data[2]

[2]For example, Brynjolfsson et al. (2011a) and Brynjolfsson and McElheran (2016) use survey data, Müller et al. (2018) use general measures of big data assets, and Brynjolfsson et al. (2011a) and Müller et al. (2018) use public firms' financial performance

and did not allow the researchers to gain insight into the different actions that firms have taken to achieve these outcomes, the resulting firm output, or the changes in customer behavior. A further challenge is that the results may be correlational and potentially suffer from endogeneity. For example, one confounder is that more productive firms might have more funds available to invest in analytics. Indeed, even research with detailed primary data—such as Bajari et al. (2019) who find that increased data availability improves the accuracy of firms' demand prediction (with diminishing returns)—suffers from a lack of random assignment of firms into adopting and using analytics, which makes causal interpretation difficult. One exception, although in a very different context, is Anderson et al. (2020) who use a field experiment with small physical mom-and-pop stores in Rwanda and also find benefits to descriptive analytics. In their case the benefit is due to training in data literacy and in interacting with technology products, whereas our paper examines e-commerce firms with existing data collection capabilities.

To overcome the data and identification challenges in prior work, we use detailed panel data from a descriptive marketing analytics dashboard about online retailers that adopted the dashboard. Our data are unique in that they also contain detailed observational data on retailers' behavior and outcomes both before and after the adoption of the analytics service. In particular we observe metrics of the retailers' actions (such as price changes and technology integration), as well as transaction-level data of individual customers. For the main analysis, we adapt the Synthetic difference-in-differences (SynthDiD) method of Arkhangelsky et al. (2021) to scenarios with varying (staggered) treatment times. We also carefully confirm the robustness of our results using multiple empirical methods.

The three mechanisms we describe (direct, complementary, and unrelated) provide us with testable empirical predictions to understand how analytics operate to help retailers. Our data, combined with the SynthDiD method, allow us to provide evidence to investigate these mechanisms. We note that our data do not allow us to perform a full causal mediation analysis, and we therefore cannot interpret all results as fully causal. Further, because our data are collected from firms that chose to adopt the dashboard, selection into treatment may limit the interpretation and generalization of the effects we estimate. Throughout the paper we indicate which assumptions might be needed for a causal interpretation of the evidence, and which results are consistent with the theory but might require further research.

Three major findings emerge from our analysis regarding (i) the overall effect of analytics adoption, (ii) the mechanisms through which the benefits are gained, and (iii) the changes in customer behavior once retailers make decisions based on analytics. First, we find causal evidence for a positive main effect of adopting descriptive analytics technologies on retailer revenue. In our sample, adoption of the analytics service is associated with an increase of 4%–10% in average weekly revenues of the firm. The results are robust to using multiple methods including staggered difference-in-difference (SDiD), SynthDiD, and instrumental variables (IV). We also find that the smallest retailers (in terms of revenues or transactions) benefit most from adoption compared to larger retailers.

Second, we are able to disentangle usage of the dashboard post-adoption from the adoption itself and show that, among adopters of the analytics service, only *users* of the service improve their performance, and these improvements increase with usage of the analytics reports. We do not observe that firms made direct changes to pricing or advertising, but they did invest in adopting new technologies, and particularly in customer relationship management (CRM), personalization, and prospecting technologies. Among those who invested in these technologies, we observe gains in revenues only for firms that actively used the analytics dashboard. Taken together these results suggest that the complementary mechanism of descriptive analytics may be in operation, and that the value of descriptive analytics is that it allows the firm to better control additional martech tools, but it does not necessarily drive direct decisions by retailers. This finding is quite interesting because it shows that marketing actions beyond pricing or advertising changes also have a large potential value if correctly optimized.

The third major finding is that the adoption of analytics results in an increase in the number of transactions, number of new website visitors, number of unique customers, revenue from repeat customers, and the diversity of products purchased, but it did not result in increased basket size per transaction or reduced CAC. Again, all of these changes occur only for those retailers that used the analytics service, and not for those that adopted but didn't use it. As we explain in detail in Section 5, these changes in customer behavior also give more credence to the complementary mechanism through which analytics affects retailer revenue. These results suggest that retailers that adopt descriptive analytics mostly gain from monitoring of complementary investments in additional technologies, which are more likely to drive these changes. Consequently, we observe that while retailers attract more customers to their website and increase repurchase rates among existing customers, the basket size does not change. In this sense, descriptive analytics with complementary technologies affects the extensive margin of customer revenue, but not the intensive margin of profit per customer.

2 Institutional Background

The analysis focuses on online retailers from a variety of industries that operate their own online stores.[3] Many of the retailers in our dataset manufacture and sell their own brands of products. For example, in the Clothing and Fashion categories, many of them produce and sell their individual designs and do not carry clothing from other brands.

Our data come from an analytics service provider that offers a popular analytics dashboard for online retailers. The software collects and analyzes data from the retailer's online store, as well as from other payment and fulfillment channels such as Amazon, Paypal, and Stripe (for sellers who sell through Amazon or use Paypal or Stripe as a checkout mechanism).[4] The analytics service was launched in 2015 and reached a substantial volume of subscribed retailers in 2016. During the majority of the period for which we collected our data, the analytics dashboard was ranked as one of the top analytics dashboards for online stores, giving it substantial visibility among retailers. Once installed, the software collects the retailer's data after installation, as well as historical data. The data are collected daily, at the transaction level, going backwards for up to 24 months before the installation of the dashboard. The service has a basic free version, but most retailers pay for the full range of services. Fees depend on the retailer's annual revenue but are generally lower than 1% of that revenue.[5]

The dashboard uses the transaction-level data to generate metrics and visual reports on aggregate sales, average basket sizes, share of repeat customers' revenues, cost of new customer acquisition, the average customer lifetime value, and many other metrics. Depending on data availability (i.e., whether the retailer also uses Google Analytics or Facebook advertising) more than 20 descriptive metrics are calculated and displayed on a weekly and a monthly level. Compared

[3] Fewer than 7% of these retailers also sell on Amazon.com as third-party sellers.

[4] For confidentiality reasons, we cannot provide identifying information about the analytics service provider.

[5] We only observe a one-time snapshot of the subscription data from late 2018; in that time there are no significant differences between users and non-users of the service in subscription rates.

to basic data reports from Google Analytics or e-commerce hosting providers, the benefits of the analytics service is that it integrates and aggregates data from the retailer's different data sources. However, the analysis is primarily descriptive, providing retailers a view into their performance. Nearly all of the retailers in our data have been using Google Analytics prior to adoption of the analytics dashboard.

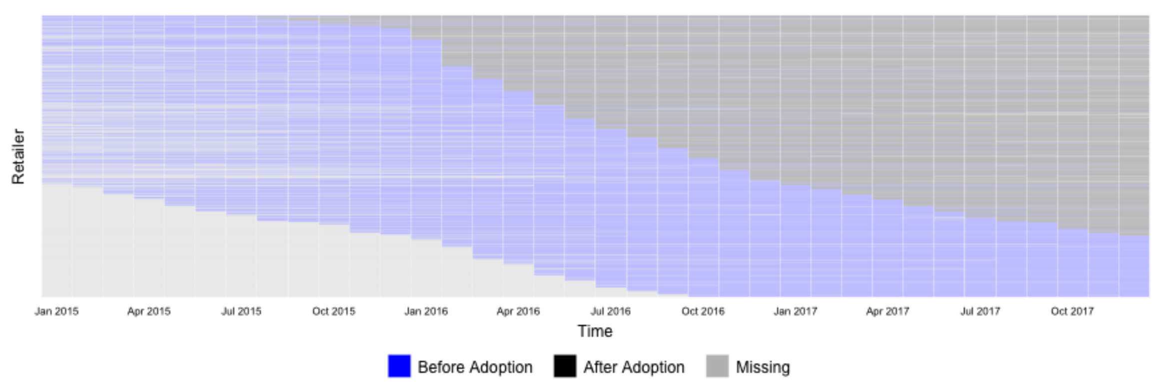
The dashboard presents information and metrics in five main reports: (i) customer acquisition costs, (ii) revenue by hosting platform report, (iii) benchmark reports that compare the retailer's metrics to a set of benchmark retailers selected by industry and revenues, (iv) an executive report that summarizes all reports, and (v) an insights report that uses an algorithm to provide individualized recommendations. Although the insights report has a prescriptive flavor, it was in early development during the time window of the data, and the recommendations were not very useful. For example, a common recommendation was "reduce customer acquisition costs by x%" or "increase the number of visitors to the website by y" without additional details. Additionally, our data also contain detailed information about each retailer's access to each report over time, and the insights report was the least frequently viewed report.

3 Data and Sample Construction

We use data from all retailers that adopted (signed up for) the analytics service between January 2016 and December 2018 and had historic annual revenue of at least USD 100,000. We use retailers that adopted the dashboard in the first 18 months as our focal treated group, and retailers that adopted in the last 18 month as a control-only group. For the treated group, we use retailers that have at least 12 monthly observations, out of which at least one observation occurs in 2015 (before signing up,) one observation in the month of adoption, and one observation after adoption. This yields a total of 1,001 retailers. Some of the retailers use multiple online channels to sell their products (e.g., their own website and Amazon), which yields 1,173 distinct retailer–channel combinations. The data are aggregated to a monthly level. Because retailers adopt the analytics service in a staggered manner, the data for each retailer in this time window can also serve as a control group during the period before their adoption. For the control-only group, in addition to the 12-month requirement, we include only retailers that had at least one observation in 2015 (or in 2016 for the 2018 adopters) to ensure sufficient overlap with the treated group. This yields 508 companies and 679 retailer–channel combinations out of a total of 1,631 companies that adopted the service during these 18 months. In our robustness tests, we leverage a larger portion of the retailers and find consistent results (see Section A.3.1 in the Appendix for details).

Figure 1 presents the treatment variation plot (Imai et al. 2018) that illustrates the dynamics of adoption of the analytics service in the data. Summary statistics for the dataset are reported in Table 1. The industry categorization for the 1,509 retailers in our data are described in Table 2. Table 3 displays the distribution of the different hosting platforms for each of the 1,859 retailer–channel combinations in our data. Most of the retailers in the sample operate their own website, while roughly 7% of the retailers sell on Amazon either as a vendor or a seller. Shopify, BigCommerce, and Magento represent 61% of the retail-channels in our sample. These platforms together held 36% market share in 2016 among the top eight e-commerce hosting platforms for independent sellers.[6] Among the top 100,000 web-stores, they held 37% in 2017.[7]

Figure 1: Treatment variation plot



Each horizontal line corresponds to a retailer–channel combination. Blue regions represent units before adoption of the analytics dashboard. Dark gray regions represent units after adoption of the analytics dashboard. Light gray regions represent units with no data in the time period.

We augment the retailer data with data from four additional sources: (i) data about non-subscribed retailers that visited the analytics service website; (ii) data about usage (login times, reports viewed) of the analytics service by retailers; (iii) data on additional technologies the re-

[6] `https://www.engadget.com/2016-11-03-ecommerce-platform-market-share-looking-at-the-companies-that-d.html`, accessed November 1, 2020.

[7] `https://aheadworks.com/blog/ecommerce-market-2017/`, accessed November 1, 2020.

Table 1: Summary statistics

Variable	Mean	Std. Dev.	Median	N
Panel A: Retailer level				
No. obs. per retailer	49.06	21.31	43.00	1,509
Avg. weekly revenue (USD)	15,187.76	49,324.79	5,220.22	1,509
Avg. no. weekly transactions	188.00	517.85	58.66	1,509
Avg. no. unique customers	169.46	472.64	50.35	1,509
Avg. basket size (USD)	179.61	375.6	87.41	1,509
No. channels per retailer	1.26	0.52	1.00	1,509
Panel B: Observation level				
Avg. weekly revenue (USD)	14,193.84	48,088.17	4,520.89	51,380
Avg. no. weekly transactions	181.74	695.95	49.00	51,380
Avg. no. unique customers	164.46	608.24	42.20	51,380
Avg. basket size (USD)	173.42	407.64	81.65	51,380

tailer installed at their online store, such as advertising tracking and email tracking (collected via Builtwith.com); and (iv) data on historical keyword advertising collected via Spyfu.com, which includes periodic keywords used and ad spend for the US and the UK. Table 4 includes summary statistics of the main variables collected for each of the retailers.

4 Empirical Strategy and Results

This section details the empirical strategy we use to identify treatment effects and then provides estimates of the revenue gains retailers experience when they adopt the analytics dashboard. Section 5 then analyzes the mechanism behind these gains.

4.1 Identification challenges

The observational nature of the data poses several challenges for identification of the treatment effects: (i) every firm in our data is either treated or untreated in each time period, and we do not observe its counterfactual outcome in the unobserved condition; (ii) firms in our data adopt the dashboard in a staggered pattern with different adoption times, requiring careful computation and aggregation of treatment effects (Goodman-Bacon 2018, de Chaisemartin and d'Haultfoeuille

Table 2: Distribution of industries

Industry	Frequency	Percent
Clothing & Fashion	366	24.3%
Health & Beauty	185	12.3%
Food & Drink	110	7.3%
Home & Garden	107	7.1%
Sports & Recreation	93	6.2%
Electronics	89	5.9%
Jewelry & Accessories	87	5.8%
Toys & Games	46	3.0%
Other	426	28.2%
Total	1,509	100%

Table 3: Distribution of hosting platforms

Hosting platform	Frequency	Percent
Shopify	1,010	54.33%
Paypal	484	26.04%
Amazon Seller	129	6.92%
Stripe	114	6.13%
BigCommerce	94	5.06%
Magento	24	1.29%
Amazon Vendor	4	0.22%
Total	1,859	100%

2020, Callaway and Sant'Anna 2021, Wooldridge 2021, Borusyak et al. 2021); (iii) the decision of when to adopt is endogenous and may be correlated with firm expectations of the benefit from the dashboard, resulting in a potential upward bias of our estimates; (iv) we only observe data from firms that eventually adopted the dashboard, and these firms might have been those expecting the highest benefit from the dashboard, leading to a potential upward bias of the estimates due to selection; and (v) a retailer could have implemented other unobserved policies concurrently with adopting the dashboard, and the observed effect cannot be separately attributed to the analytics dashboard.

To address these challenges we utilize two empirical methods: staggered difference-in-differences (SDiD) based on Wooldridge (2021), and cohort-based synthetic difference-in-differences (Synth-DiD) based on Arkhangelsky et al. (2021), which borrows strengths from the SDiD method as well as the synthetic control method (Abadie et al. 2010, Abadie 2021). The technical details of these methods are provided below, but first, we provide an overview of how these methods address the identification challenges.

Challenge (i) of unobserved potential outcomes is addressed by constructing a control group to predict the counterfactual potential outcomes of an adopting firm as if it did not adopt analytics. To do so, we treat the cohorts of retailers that adopted the analytics service in our data window as our focal treatment group. The control group for each adopting cohort is comprised of data from firms that haven't adopted the dashboard yet, or that have adopted the dashboard for the

Table 4: Summary statistics—Additional data

Variable	Mean	Std. Dev.	Median	N
Avg. monthly no. logins	.52	.92	.20	42,925
Monthly indicator for any report views	.08	.27	0	42,925
Monthly no. technologies	55.5	23.42	52	40,754
Google avg. monthly ad costs (USD)	2,531.4	12,275.7	0	18,464
Spyfu monthly advertising spending (USD)	1,838.7	13,193.4	423.9	12,357
Spyfu monthly no. advertising keywords	27.4	44.5	8	7,024

The observations in this table are retailer-month level, since these data are collected at the retailer level. Some data sources do not include all of the retailers in our sample. This leads to a smaller number of observations for the respective variables.

first time outside the data window (Manchanda et al. 2015). Specifically, we designate retailers that adopted the analytics dashboard between January 2016 and June 2017 as the treated cohorts, whereas retailers that adopted between July 2017 and December 2018 serve as control-only cohorts.

This approach requires a conditional parallel-trends assumption on pre-treatment outcomes to hold in the case of SDiD, which we address by controlling for linear cohort-level time trends. The SynthDiD method relaxes the parallel-trends assumption and allows for a more flexible pre-adoption pattern. We test the validity of these methods both visually and statistically.

Challenge (ii) of staggered adoption is addressed by using both the SDiD and the SynthDiD methods to compute treatment effects for each cohort of adopting firms separately, and then aggregating them into an overall average treatment effect on the treated (ATT), resolving issues such as negative weighting of treatment effects identified by, e.g., Goodman-Bacon (2018), de Chaisemartin and d'Haultfoeuille (2020), and Callaway and Sant'Anna (2021).

We address challenge (iii) of endogenous adoption timing in two ways. First, SynthDiD is consistent even if there is unobserved correlation between treatment assignment and firm-level time trends, alleviating much of the concern. Second, we perform an instrumental variables analysis using three instruments for the timing of adoption that likely shift adoption timing but are plausibly unrelated to the retailer's revenue.

Regarding challenge (iv), because all of the retailers in our data adopt the analytics service eventually, they may differ from unobserved non-adopters, causing a selection bias when interpreting our estimates as an average treatment effect (ATE). Our estimates are therefore the effect on retailers that choose to adopt the service, i.e., the average treatment effect on the treated (ATT).[8] Despite this potential limitation, we believe that the ATT is a relevant and appropriate measure to focus on because gaining value from analytics requires engaging with the reports and making data-driven decisions, which are all endogenous decisions. Even if retailers are exogenously assigned to adopt a descriptive analytics solution, we do not expect to see any benefit for those that do not use it.[9] Accordingly, we measure the ATT and expect that similar retailers that are interested in adoption of descriptive analytics will exhibit similar outcomes and experience similar benefits as those in our sample. A second concern of selection is that firms with a higher benefit might choose to adopt with higher probability, and indeed, our estimates are representative for the set of firms in our data. In Section 4.4 we provide more details about treatment effect heterogeneity in our sample. A final concern is that firms with a higher expected benefit might select earlier into treatment, but this is handled by the solutions to challenge (iii) presented above.

Finally, to address challenge (v) of unobserved confounders, we use a unique aspect of our dataset where we observe firms' usage of the dashboard post-adoption. We compare the treatment effect for firms that made use of the dashboard ("users") vs. those that didn't ("non-users") to provide evidence that the benefit we observe can be attributed to the dashboard and not to other confounders.

4.2 Effect of analytics adoption on retailer revenue

4.2.1 Staggered Difference-in-Differences (SDiD)

We start by analyzing the impact of analytics adoption on average weekly revenue using difference-in-differences. When adoption is not staggered a common approach to estimating the ATT is to

[8] In this setting, because the dashboard was available to all retailers and installation requires awareness, the ATT is equivalent to the local average treatment effect (LATE), and not the average treatment effect (ATE) in the population.

[9] A randomized controlled trial would not resolve this issue completely because of compliance and statistical power challenges. For example, Anderson et al. (2020) analyze the impact of basic analytical training on small mom-and-pop stores in Rwanda. Despite using randomized controlled field experiments, Anderson et al. (2020) do not estimate the ATE in the general population because they recruited a specific group of entrepreneurs based on their growth potential. In our setting adopters operate technically sophisticated e-commerce stores and are expected to have these very basic analytical skills.

utilize a two-way-fixed-effects (TWFE) estimator by using OLS:

$$\mathbb{E}[log(Y_{ijt} + 1)] = \alpha_{ij} + \gamma_t + \beta AfterAdopt_{ijt} \tag{1}$$

where Y_{ijt} is average weekly revenue for retailer i in channel j in month t, α_{ij} are retailer–channel fixed effects, γ_t are time fixed effects, and $AfterAdopt_{ijt}$ indicates whether retailer–channel ij adopted the dashboard at or before time t. The coefficient β in this specification estimates the ATT.

Wooldridge (2021) shows that in the case of staggered adoption the TWFE estimator is equivalent to a pooled OLS regression where unit fixed effects are replaced with cohort fixed effects and where the treatment effects are allowed to vary by cohort and time. We additionally allow for flexible linear time trends by cohort, which yields the following specification:

$$\mathbb{E}[log(Y_{ijt} + 1)] = \sum_{r=q}^{T} \lambda_r Adopt_{ijr} + \gamma_t + \sum_{r=q}^{T} \tau_{rt} Adopt_{ijr} \mathbb{I}[t \geq r] + \sum_{r=q}^{T} \phi_r Adopt_{ijr} \times t \tag{2}$$

where $Adopt_{ijr}$ indicates whether retailer–channel ij first adopted the dashboard at time period (cohort) r, $\mathbb{I}(\cdot)$ is the indicator function, λ_r is a fixed effect for adoption cohort r, and τ_{rt} measures the treatment effect at time t for retailer–channels that adopted in time period r. The value q denotes the time period of the first adoption cohort in the data, while T denotes the last time period. The coefficients γ_t are time fixed effects as before, and ϕ_r captures any linear time trends in outcomes of retailer–channels that adopted in cohort r.

The main difference between specifications (1) and (2) is that the unit (retailer–channel) fixed effects α_{ij} are replaced by cohort fixed effects λ_r, and that treatment effects are allowed to be heterogeneous across both cohorts and time. For example, τ_{23} measures the effect at time $t = 3$ for units that adopted in cohort $r = 2$, and it may be different from τ_{33} (effect at time $t = 3$ for cohort 3 units), and from τ_{22} (effect at time $t = 2$ for cohort 2 units). Specification (2) resolves the bias created by specification (1) when staggered adoption with heterogeneous treatment effects is in play. If the true ATT was homogeneous across units and time, or if adoption occurred in a single period, then both specifications would have yielded the same estimate.

We use two-way clustering of standard errors by retailer and month to address serial correlation (Bertrand et al. 2004). Most retailers sell using only one channel—typically their own website. However, we use retailer–channel observations (i.e., a retailer that sells on their own website and also through Amazon has two observations at each time period) because retailers may adopt the dashboard at different times for different online channels. Results are consistent when we aggregate the data to the retailer level.

The coefficient τ_{rt} measures the change in average weekly revenue after the adoption of analytics of cohort r at time t. It is identified by comparing the change in revenue of cohort r adopters from periods before r to period t with the change in revenues in the same time frame for retailers that adopt the dashboard after time t. Unlike Datta et al. (2018), we do not observe a group of non-adopters and hence we rely on future adopters as a control group. The identifying assumption in our SDiD analysis is a conditional parallel-trends assumption—that there were no differential trends in revenues before adoption between retailers that adopted the service and those that did not after conditioning on unit- or time-invariant covariates. We also note that it is unlikely that those retailers that haven't yet adopted the service are indirectly affected by those that adopted the service because adoption is not observable by other firms and there is little competition between the retailers in the sample.

Given the cohort-level estimates $\widehat{\tau}_{rt}$, we can compute a dynamic ATT estimate, $\widehat{\tau}_{\ell} = \frac{\sum_{r=q}^{T-\ell} \widehat{\tau}_{r(r+l)}}{T-\ell-r+1}$, that depends on the length ℓ of exposure to treatment $(0 \leq \ell \leq T-q)$ and measures the average effect ℓ periods after treatment for all treated units. We can also aggregate all the treatment effects into an overall ATT estimate $\widehat{\tau}$:

$$\widehat{\tau} = \frac{\sum_{r=q}^{T} \sum_{t=r}^{T} \widehat{\tau}_{rt}}{(T-q+1)(T-q-2)/2}$$

To ensure that there are sufficient control units in each time period, we truncate our data after June 2017, allowing all firms that adopted after this month to serve as controls only (whereas in each time period, firms in cohorts that have not yet adopted serve as controls before their adoption, similar in spirit to Wang and Goldfarb (2017)). Due to this data truncation, we limit our SDiD analysis to include only observations between July 2015 and June 2017, for retailers that adopted the dashboard in 2016–2018. This results in 36,433 observations of 18 retailer–channel cohorts that adopted the dashboard between January 2016 and June 2017, totalling 1,173 treated units, and 679 control-only units that adopted after June 2017.

The validity of the estimates depends on the validity of the identifying assumption of conditional parallel trends, which we can test because we observe retailer–channels over multiple time periods before adoption. Following Borusyak et al. (2021) we run an OLS analysis that uses only the pre-treatment[10] data and estimate the following specification:

$$\mathbb{E}[log(Y_{ijt}+1)] = \sum_{r=q}^{T} \lambda_r Adopt_{ijr} + \gamma_t + \sum_{\substack{\ell=1-T \\ \ell \neq -7}}^{-2} \delta_\ell Adopt_{ij(t-l)} + \sum_{r=q}^{T} \phi_r Adopt_{ijr} \times t \tag{3}$$

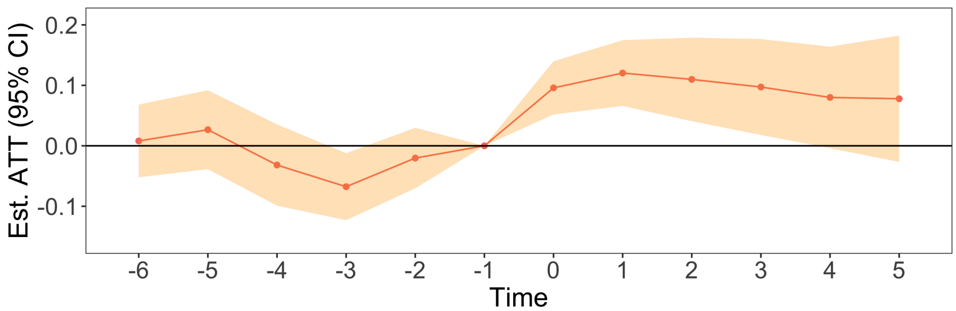
where, as before, λ_r are adoption cohort fixed effects, γ_t are time fixed effects, and ϕ_r are adoption-cohort linear time trends. The indicator $Adopt_{ij(t-l)}$ denotes units that have adopted $|\ell|$ time periods after period t, and the coefficients δ_ℓ are lagged treatment effects that measure the difference in outcomes between treatment and control units ℓ periods before adoption. The effect at lag $\ell = -1$ is normalized to zero, and lag $\ell = -7$ is omitted to serve as the baseline level of outcome, because the data truncation in June 2015 means that the earliest treated cohort of January 2016 has at most six periods of pre-treatment data. We then use a Wald test to test the null hypothesis $\delta_\ell = 0$ for $-6 \leq \ell \leq -2$, that states that the differences in outcomes between treatment and control units in the six periods before adoption are zero.

Figure 2 presents the dynamic event study treatment effect estimates $\widehat{\tau}_\ell$ from specifications (2) and (3), along with their 95% confidence intervals. We observe that most of the pre-treatment estimates contain zero in their confidence intervals (with a violation in period -3), while the treatment effect estimates are positive and inch towards zero in later periods. The Wald test of the null hypothesis of parallel trends is unable to reject the null with a p-value of 0.1079, providing some evidence for the validity of the parallel-trends assumption. Column 1 of Table 5 presents the estimate of the ATT for the first quarter post-adoption, which is valued at .108 [95% C.I.: .056, .161], and Column 2 presents the six-month ATT, which is valued at .098 [95% C.I.: .034, .162]. Because revenues are logged, the six-month ATT estimate suggests an increase in average weekly revenues of roughly 10.3% each month. The unconditional median average weekly revenue before adoption in our data is USD 4,364. Therefore, the median retailer experienced an increase of approximately USD 449.

Equations (2) and (3) allow for flexible cohort-level linear time trends, measured by the coefficients ϕ_r. Such a specification is uncommon in standard analysis that uses a specification such as (1) but is necessary in our setting to obtain parallel trends. When we re-estimate specifications (2)

[10] Borusyak et al. (2021) show that using both pre- and post-treatment data to test for parallel trends conflates multiple identifying assumptions and tries to simultaneously test the parallel-trends assumption and estimate the treatment effects under this assumption. Instead, they recommend using only untreated units (never treated or not yet treated).

Figure 2: SDiD treatment effects



Coefficients and 95% confidence intervals obtained by estimating equations (2) and (3).

Table 5: Effect of analytics adoption on retailer revenue

Methodology	Staggered DiD		SynthDiD		IV (Third stage)
	(3 months)	(6 months)	(3 months)	(6 months)	(one month)
	(1)	(2)	(3)	(4)	(5)
After	.108**	.098**	.056**	.038+	.167**
Adoption	(.026)	(.031)	(.019)	(.020)	(.05)

Significance level: 10% (+); 5% (*); 1% (**).

Point estimates for the treatment effects using: SDiD in Columns (1) and (2) (average effects for three and six months post-adoption); SynthDiD in Columns (3) and (4) (average effects for three and six months post-adoption); and IV in Column (5) (effect for the month of adoption only).

and (3) after omitting the linear time trends coefficients ϕ_r, the analysis yields non-zero estimates of $\widehat{\tau}_\ell$ for most of the pre-treatment periods $\ell = -2, \ldots, -6$. Additionally, the Wald test of the null hypothesis $\delta_\ell = 0$ rejects the null with a p-value of 0.0011. These results raise concerns about the validity of the estimates from SDiD without controlling for time trends at the cohort level.

In the next section we turn to identify the treatment effects using synthetic difference-in-differences, which does not rely on a parallel-trends assumption and also resolves a few additional identification challenges mentioned in Section 4.1.

4.2.2 Aggregated Synthetic Difference-in-Differences (SynthDiD)

The synthetic control method (SCM) uses a weighted average of outcomes from control retailers to predict the outcomes of the adopting retailer "as if" they did not adopt the analytics service. The weights are chosen to optimally match the pre-adoption outcomes of the adopting retailer, and thus they capture any possible trends that might affect identification without requiring a parallel-trends assumption. The difference between the observed outcomes post-adoption and the predicted outcomes are the estimated treatment effects from the method (Abadie et al. 2010, Abadie 2021).

Synthetic control analysis is often applied to a single or small number of treated units with a small number of control units and requires long and balanced pre-adoption panels. In our setting, the choice of pre-adoption panel length creates a tradeoff between bias and variance of the estimates. If we require long pre-adoption panels, the risk of biased estimates is lowered, but the the number of eligible retailers to include in the analysis is smaller, which increases the variance of the estimates. If we shorten the number of pre-adoption time periods, more retailers are eligible to be included in the analysis, but the risk of biased estimates increases. We address this challenge by employing synthetic difference-in-differences (Arkhangelsky et al. 2021), which borrows strengths from both the DiD and the SCM methods. SynthDiD finds optimal weights for control units and pre-treatment periods to minimize the mean squared error of the target ATT to be estimated. Similarly to SCM, SynthDiD uses pre-treatment data of control units to create a synthetic control for the average outcome of treated units and does not rely on a strong parallel-trends assumption for identification. Similarly to DiD, SynthDiD is invariant to unit-level shifts in outcomes and also allows for inference with large panels, even when the pre-treatment period is short.

The SynthDiD method is designed for a balanced panel of units where the treatment timing is identical for all treated units. We adapt the method to estimate the ATT with staggered adoption by estimating a cohort-level ATT and then aggregating the estimates, in a method similar to that used in Section 4.2.1 and also in applications of SCM with staggered adoption (Ben-Michael et al. 2021).

To perform the analysis, for each adoption cohort r we construct a balanced panel where the treatment group is comprised of retailer–channels that adopted the dashboard in period r and have outcome data available $\ell_{\min} = -6$ periods before adoption and $\ell_{\max} + 1 = 6$ periods after, and where the control group is comprised of retailer–channels that have data available for the same time frame but that adopted the dashboard after more than $\ell_{\max}$ periods after cohort r.

If we denote by N_r the set of units in the balanced panel of cohort r, by N_r^{co} the set of units in the control group, and by N_r^{tr} the set of units in the treatment group,[11] then for each cohort r the SynthDiD estimation procedure solves:

$$(\widehat{\tau}_r, \widehat{\alpha}_0, \widehat{\alpha}_{ij}, \widehat{\gamma}_t) = \underset{\tau_r, \alpha_0, \alpha_{ij}, \gamma_t}{\arg\min} \left\{ \sum_{ij \in N_r} \sum_{t=r+\ell_{\min}}^{r+\ell_{\max}} \left(log(Y_{ijt}+1) - \alpha_0 - \alpha_{ij} - \gamma_t - AfterAdopt_{ijt} \cdot \tau_r\right)^2 \widehat{\omega}_{ij} \widehat{\lambda}_t \right\} \tag{4}$$

where τ_r is the average ATT of cohort r in the $\ell_{\max}+1$ periods after adoption, α_0 is an intercept, α_{ij} and γ_t are retailer–channel and time fixed effects as before, and $AfterAdopt_{ijt}$ indicates whether retailer–channel ij adopted the dashboard by time period t.

Equation (4) estimates a two-way-fixed-effect model identical to equation (1) with the addition of unit-specific weights $\widehat{\omega}_{ij}$ and time-specific weights $\widehat{\lambda}_t$. The unit weights $\widehat{\omega}_{ij}$ are selected such that the pre-treatment control outcomes weighted by $\widehat{\omega}_{ij}$ have a similar trend to that of the average outcomes of the treatment units, i.e., for time periods $t < r$,[12]

$$\widehat{\omega}_0 + \sum_{i \in N_r^{co}} \widehat{\omega}_{ij} log(Y_{ijt}+1) \approx \frac{\sum_{i \in N_r^{tr}} log(Y_{ijt}+1)}{|N_r^{tr}|}.$$

The time weights $\widehat{\lambda}_t$ are designed so that the average post-treatment outcome for each of the control units differs by a constant from the weighted average of the pre-treatment outcomes of the same control units, i.e.,

$$\widehat{\lambda}_0 + \sum_{t=r+\ell_{\min}}^{r-1} \widehat{\lambda}_t log(Y_{ijt}+1) \approx \frac{\sum_{t=r}^{r+\ell_{\max}} log(Y_{ijt}+1)}{\ell_{\max}+1}.$$

The unit weights $\widehat{\omega}_{ij}$ serve the same role as in the standard SCM to align the pre-exposure trends in the outcomes of treated and control units. The time weights $\widehat{\lambda}_t$ balance pre-exposure time periods with post-exposure ones—if a specific pre-exposure period is more predictive of post-exposure outcomes, it will receive a higher weight. Arkhangelsky et al. (2021) show that under quite general conditions, including correlation between treatment assignment and unit-level time trends, as well as heterogeneous treatment effects, $\widehat{\tau}_r$ is a consistent and an asymptotically normal estimator of

[11] That is, $N_r = N_r^{co} \cup N_r^{tr}$.

[12] We omit the full details of the estimation which can be found in Section 2 of Arkhangelsky et al. (2021). Briefly, the method uses regularized ridge regression to regress the average of treatment unit outcomes on control unit outcomes, while constraining the resulting weights to sum to 1.

τ_r as long as the combination of the number of control units and pre-adoption periods is large enough compared to the combination of the number of treated units and post-adoption periods.[13] We believe that these assumptions hold in our setting since lowering the number of pre-adoption periods indeed increases the number of control units faster than the number of treated units.

Similarly to the analysis in Section 4.2.1, we truncate the data between July 2015 and December 2017, and obtain 18 cohort-level ATT estimates $\widehat{\tau}_r$ for cohorts that adopted between January 2016 and June 2017 using the `synthdid` package in the R programming language.[14] We then compute the weighted average ATT as $\widehat{\tau} = \frac{\sum_r N_r^{tr} \cdot \widehat{\tau}_r}{\sum_r N_r^{tr}}$. Standard errors for each $\widehat{\tau}_r$ are estimated using the jackknife (Algorithm 3 of Arkhangelsky et al. (2021)), or the placebo method (Algorithm 4 of Arkhangelsky et al. (2021)) if a cohort has only one treated unit. Standard errors for the overall ATT are computed as a weighted average of the cohort-level standard errors. Across the 18 cohorts there are 982 treated retailers that have sufficient observations for the analysis. Table A-1 in the Appendix details the breakdown of the number of treated and control units used in each cohort. On average there are 55 treated units and 485 control units in each cohort.[15]

Columns 3 and 4 of Table 5 present the estimated ATT from SynthDiD in the three and six months after adoption of the dashboard, with the first quarter's ATT valued at .056 [95% C.I.: 0.019, 0.090], and the six-month ATT valued at .038 [95% C.I.: -.001, 0.078]. We also present an event study plot in Figure 3 that shows the pre-treatment fit of SynthDiD between the treatment and control units, as well as the evolution of the treatment effect over time.[16] We observe that SynthDiD produces better pre-treatment fit between the trends of the treated and control units compared to SDiD. The treatment effect estimates after adoption are smaller in magnitude than those achieved with SDiD. The effects are positive in the first few months after adoption and then attenuate in later periods.[17] The six-month treatment effect of .038 is equivalent to an average increase of 3.9% in revenues, or USD 170 for the median retailer in the data.

Section A.3.1 of the Appendix presents multiple analyses that demonstrate robustness of our

[13] The exact conditions appear in Assumption 2 of Arkhangelsky et al. (2021).

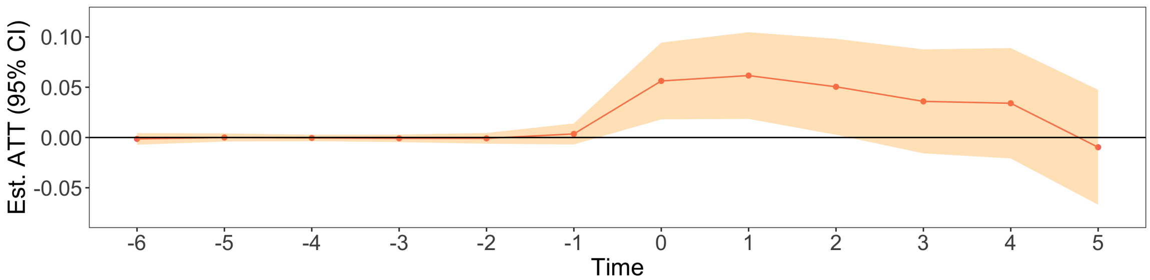
[14] The data window is longer in this analysis to allow a balanced panel of the same length for all cohorts. Using the same truncation rule as in Section 4.2.1 doesn't affect the results qualitatively.

[15] The control group is larger than the treated group because it is comprised of all later adopters from multiple future cohorts and not just units from one cohort.

[16] Appendix A.1 describes the methodology used to construct this plot.

[17] Column 1 of Table A-2 of the Appendix reports the period-by-period estimates that correspond to Figure 3.

Figure 3: SynthDiD treatment effects



SynthDiD ATTs of adopting analytics on revenues over time. Time 0 indicates month of adoption. Other times are relative to adoption.

SynthDiD estimates to different numbers of pre-adoption lags ($\ell_{\min}$) and post-adoption leads ($\ell_{\max}$) used to create the balanced panel, aggregation of the data to the retailer level instead of retailer–channel level, and using an additional sample where 2017 adopters serve as treated and 2018 adopters serve as controls.

Because SynthDiD ensures an appropriate number of control units, accounts for potential pre-adoption trends, and allows for some level of endogenous treatment selection and timing, we use SynthDiD as the main analysis method in the remainder of the paper.

4.3 Addressing potential selection bias

In this section we assess the sensitivity of our findings to selection. We ask how the estimate that weekly revenues increase by an average of 3.9% in the six months after adoption of the dashboard might be affected by endogenous timing of the adoption or by selection into treatment. First, we use an instrumental variables analysis to alleviate concerns about endogenous adoption timing. Second, we present evidence that adopting the dashboard without using it does not produce an observed increase in performance, which should alleviate concerns about unobserved confounding factors. The results in this section allow us to rule out the unrelated mechanism described in Section 1.

4.3.1 Instrumental variables analysis of endogenous adoption timing

As described in challenge (iii) in Section 4.1, firms may choose to adopt the dashboard when they anticipate the greatest benefit. We address this issue using instrumental variables (IV) analysis.

Our IV strategy hinges on identifying times of increased awareness and attention to the analytics service among online retailers and on assuming that such increased awareness drives some of the adoption exogenously. Specifically, we use three time-varying exogenous factors that plausibly impact the timing of analytics adoption but that are uncorrelated with the revenue of the retailer. The first IV utilizes the adoption rates among other retailers within each retailer's industry, and the other two IVs use data from the website of the *analytics service provider* (not the retailers) to measure interest in signing up to the dashboard from other websites in the retailer's geographic region and hosting channel. We view these three IVs as exogenous shifters of adoption timing that are not correlated with the performance of an adopting retailer. Web Appendix WA-1 provides details of the data used to construct the three instruments and discusses their relevance for shifting adoption and the validity of the exclusion restriction assumption.

Because retailer–channels adopt the dashboard only once, the instruments cannot continue to be relevant adoption shifters post-adoption. We therefore limit the scope of our analysis to include only observations up to the first month of adoption (including that month). Because we use an IV as the identification strategy, we consequently treat adoption timing as random and no longer rely on pre-adoption matching of trends between treatment and control. This allows us to use all adopting cohorts in the IV analysis.

Our IV strategy has two limitations. First, the instruments would ideally have exogenous variation at the retailer–channel level in each time period. However, our instruments only have variation at an industry-region-platform level, implying that two units from the same region, in the same industry, and using the same platform will have identical instruments. Luckily, these three instruments together create enough sub-groups so that the variation is almost at the retailer–channel level (generating 6,686 unique values for 12,115 observations in this analysis). Second, because the estimates limit the data to incorporating only one period after adoption, the IV estimates measure the ATT only during the first month of adoption. For these two reasons, we consider the IV estimates only as supporting evidence for the positive effect of adoption on performance outcomes that we found using SDiD and SynthDiD.

Because the endogenous adoption variable is binary, we utilize the three-step estimation procedure proposed in Deng et al. (2019) based on Wooldridge (2010), Wooldridge (2005), and Wooldridge (2019). The first step is a probit model predicting the decision to adopt the dashboard. The resulting predicted probability is then used as an instrument in a two-step IV procedure. The estimation procedure and results are detailed in Web Appendix WA-2. The 2SLS instrument derived from the first stage passes the Stock-Yogo, Cragg-Donald, Anderson-Rubin, and Stock-Wright weak instrument tests.

Column 5 of Table 5 reports the IV analysis results that confirm our previous findings and show a revenue increase of 18.2% [95% CI: 7.1%, 30.5%]. Because the IV analysis only estimates effects for the month of adoption, the estimates are substantially larger than those obtained in the SDiD analysis presented in Figure 2 and in the SynthDiD analysis presented in Section 4.2.2.[18] Note that the IV analysis uses the same data used in the SynthDiD analysis. When limiting the data to the data used in the SDiD analysis, the IV estimate is .099 with a p-value of 0.057. That suggests a revenue increase of 10.4% [95% CI: -.003%, 22.3%].

4.3.2 Disentangling adoption and usage

A second issue we examine is potential simultaneity bias in the estimate (challenge (v) in section 4.1). Firms that adopted the dashboard could have made other concurrent changes, such as changing their management team or their website design, and these actions could have generated the change in performance outcomes we observe. If this concern is valid, we would expect firms that adopted the dashboard, but *did not* use it, to exhibit increased performance as well.

Our data are unique in that they include the dashboard login times and the reports accessed by each retailer, which allow us to separately estimate the effects of the dashboard's adoption by level of usage. We examine three dimensions of usage: any usage, intensity of dashboard logins, and report viewership. Because the access and login information are not available at the retailer–channel level, but rather at the retailer level, we aggregate observations to the retailer level and use the first time of adoption as the retailer's adoption time. We use the SynthDiD method to analyze the differential treatment effects based on each of the different aspects of usage.

First, we identify "users" as those retailers that ever logged on to the analytics service. Roughly 10% of retailers that adopted the service never used it, and we call them "non-users." We then compare post-adoption treatment effects across the groups of users and non-users to examine whether

[18] To further evaluate these results, Table W-3 in the Web Appendix provides a direct comparison and reports the results of an OLS regression using the same observations and specifications as the IV regressions, without instrumenting for the Adoption indicator. The standard OLS effect size is less than half of the IV's third-stage estimate.

retailers that use the service exhibit different outcomes than those that do not.[19] Figure 4 reports the treatment effects of all time periods from six months before adoption to six months after adoption. Before adoption, for both users and non-users most point estimates are statistically indistinguishable from zero (those that are different than zero are very close economically to zero). After adoption, only the point estimates of users are larger than zero. Additionally, the ATT for the six months post-adoption are .046 [95% CI: 0.002, 0.090], and -.11 [95% CI: -0.247, 0.019] for users and non-users, respectively.[20] When we compute the confidence interval for the difference between the ATTs in the six months post-adoption, we find a difference of 0.16 [95% CI: 0.090, 0.230], which is significantly different from zero with a p-value less than 0.001. In fact, we find that for every time period in the six months after adoption, the point estimates for the group of users and non-users are statistically different from each other with a p-value of less than 0.001, except for time period 1 where the p-value for the differences equals 0.017.

Reassuringly, we find that our positive effects are driven by the retailers that use the analytics dashboard. Although we are not able to completely rule out simultaneity of actions by the retailer, a correlation between usage and performance is a necessary condition for analytics to have a causal effect. Because our prior estimates were an overall average for all adopters, we find that the effects are larger for those that adopted the service and became users of the service.

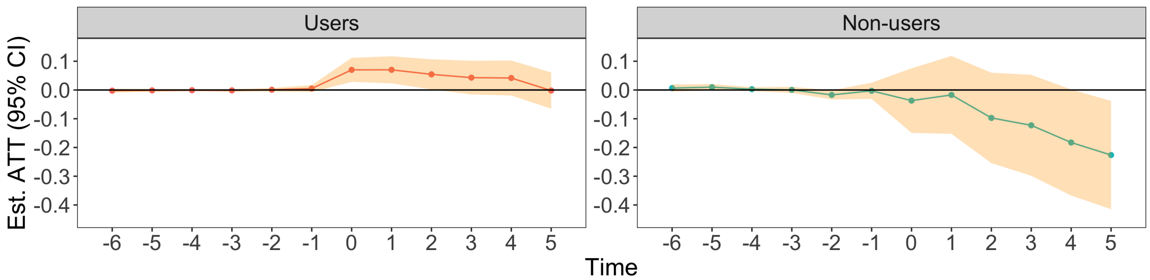
Next, we examine the relationship between the intensity of dashboard usage and performance outcomes. For each retailer, we compute the average monthly number of dashboard events and split the sample of retailers to those above and below the median average post-adoption usage.[21] We use SynthDiD to compute the ATTs for each intensity subgroup. Column B of Table A-3 in the Appendix presets the results, demonstrating that intensity of analytics service usage matters.

[19] Appendix A.2 details the method of computing subgroup effects. While the number of non-users is relatively small, SynthDiD is designed to compute treatment effects even if only one unit is treated, as long as there are enough control units. Therefore, the method works well in this setting.

[20] For pre-adoption matching, SynthDiD targets the ATT of the entire sample and thus estimates for individual time periods might be inaccurate while their average is an unbiased estimate. The method also does not optimize the pre-adoption fit for each sub-group, as can be seen in Figure 4, and each subgroup is compared to a mix of units from both groups as the control. We should therefore focus on the difference between the subgroups and not the absolute values of the period-level effects.

[21] We use median split rather than quartiles or deciles to ensure pre-adoption matching for the treated subgroup. Results are similar but noisier if we split the data into more quantiles.

Figure 4: Treatment effects of users versus non-users



ATTs (and 95% confidence intervals) of adopting analytics on revenues over time. Time 0 indicates the month of adoption. Other times are relative to adoption. Column A of Table A-3 in the Appendix presets the corresponding coefficients.

Overall, retailers with a larger average number of monthly events reap additional increases in revenues compared to those with less activity. (For example, the ATT for revenues six months after adoption is -.018 [95% CI: -0.071, 0.035] for the below-median usage group vs. .099 [95% CI: 0.031, 0.167] for the above-median group.)

Finally, we examine the usage of reports. The dashboard provides five main reports, as detailed in Section 2. We split the data to those that use reports and those that do not. We conduct two different tests using the reports data. First, we generate an indicator variable for whether the retailer examined any of the reports during a particular month. Nearly 80% of retailers viewed a report at least once, but there is substantial variation in report usage over time. Second, we create monthly indicator variables for whether a retailer examined a particular type of the five reports described above, and then for each retailer we compute the average number of monthly reports they examined. In the first test we compare SynthDiD estimates between retailers that viewed at least one report and retailers that did not view any reports. In the second test, we again compare the revenue of firms with below- and above-median average numbers of reports viewed.

We find that those retailers that accessed reports exhibit higher increases in revenues compared to those that did not, and retailers that accessed reports more frequently had larger gains compared to those that accessed reports less frequently. Columns C and D of Table A-3 in the Appendix present the results for these SynthDiD analyses. (For example, the ATT for revenues six months after adoption is -.009 [95% CI: -0.064, 0.028] for the below-median report usage group vs. .077 [95% CI: 0.012, 0.142] for the above-median group.)

To summarize, we find that the intensity of dashboard usage is associated with increased revenue. If retailers adopted other methods that increased their performance simultaneously to the adoption of the dashboard, we would not have found that the increase in revenues is associated with dashboard usage.

4.4 Heterogeneity of the impact of the analytics service

Given the evidence for the effect of adopting the analytics service, we turn to ask whether the benefits are distributed uniformly across firms. In particular, we look at the heterogeneity of the effect based on firm size. We use SynthDiD to estimate heterogeneous effects by comparing the ATTs between subgroups of large and small firms.

Because we do not have an external measure of firm size (such as number of employees), we use revenues and number of transactions as proxies for size. To avoid using the same observations for both determining firm size as well as matching pre-trends, we make sure there is enough time difference between observations used to determine size and those used to match pre-trends. To do this, we focus on firms that adopted in the first half of 2017, and we use two proxies for retailer size: i) median split of the average monthly revenues in 2015; and ii) median split of the average number of transactions in 2015. Table A-4 in the Appendix presents the results. For both revenue and transaction medians, retailer–channels with below-median size exhibit a statistically significant increase in revenues after adoption, while the above-median retailers exhibit marginal and small decreases or no statistically significant changes in performance after adoption. Specifically, the ATT for revenues six months after adoption is .224 [95% CI: 0.064, 0.384] for the below-median revenue-based size group vs. -.033 [95% CI: -0.108, 0.042] for the above-median group. Similarly, the ATT for revenues six months after adoption is .220 [95% CI: 0.069, 0.370] for the below-median transaction-based size group vs. -.019 [95% CI: -0.110, 0.073] for the above-median group. Additionally, we compute the confidence intervals for the differences between the ATTs in the six months post-adoption between the below- and above-median firms and find statistically significant differences. Therefore, we conclude that smaller retailers reap more of the benefits of dashboard adoption. This effect might be expected with descriptive analytics—larger and well-established retailers might have already optimized their performance and experience lower returns from this type of analytics.

5 Mechanism

Because we observe evidence of a positive effect of adopting descriptive analytics, we expect the adoption of analytics to drive better retailer decisions which would translate to changes in customer behavior. These changes in customer behavior would be associated with generating higher revenues. In this section we first examine which decisions retailers make after they adopt descriptive analytics, and we further examine whether the observed changes in customer behavior are consistent with the predicted changes from the actions taken. The analysis allows us to provide evidence for the potential mechanisms through which descriptive analytics operate to drive the gains in revenue described in Section 1.

The three mechanisms we described previously (direct, complementary, and unrelated) provide us with testable empirical predictions to understand how analytics operate to help retailers. For the direct mechanism, we would expect to see changes in retailer decisions that are unrelated to the integration of additional technologies, and we would expect to see these changes only for retailers that make use of the analytics dashboard (users) versus those that adopt analytics but do not login to look at the reports (non-users).

For the complementary mechanism, we would expect to see that the highest gain from adopting analytics appears for retailers that adopted additional technologies, but this gain will only happen if they also made use of the descriptive analytics dashboard. That is, if descriptive analytics is not complementary to the other technologies, we would not expect to see a difference in the benefit generated among users and non-users of the analytics dashboard. Other evidence that would be consistent with this mechanism are changes in customer behavior that are predicted by the addition of specific technologies but that are unrelated to other actions of the retailer.

Finally, for an unrelated mechanism we would observe that adopting analytics presents a gain in performance regardless of usage of the dashboard, and regardless of what other actions the retailer takes as a result of adopting analytics. If an unrelated process is the mechanism behind the effect, the effect will also disappear when we instrument for the timing of adoption.

We note that, ideally, we would like to perform a causal mediation analysis (Imai et al. 2010a;b, Pearl 2014) to disentangle between the different mechanisms. However, our data do not provide enough power and exogenous variation to properly test these claims in a causal framework. Our results are therefore indications for consistency between our theory and observed behavior, but more research with better data will be needed to fully explore it.[22]

Descriptive analytics does not provide retailers with specific guidance on which actions to take. Even the availability of simple benchmarks, which may tell a retailer whether they are underperforming on some metric, does not prescribe how to improve that metric. One possible exception is customer acquisition cost (CAC)—retailers often know the value of a customer (in terms of the average margin), and if the CAC is higher than the desired margin of the retailer, it is obvious they should change their advertising spending. However, it is less clear how this change should be made (e.g., decreased budget or change in allocation). We therefore analyze the main observable actions that retailers can take to affect customer behavior: (i) changing prices; (ii) reallocating and optimizing advertising; (iii) in-store personalization; and (iv) driving traffic to the store.

If prices in the store change, or if personalization is improved, we expect the purchased baskets to change, either in size or in product composition. If advertising allocation or optimization changes, and if technologies that drive customers (new or repeating) to the websites change, we expect to see a change in the composition of customers that make purchases after the adoption of analytics.

5.1 Decisions driven by the adoption of analytics

To analyze the changes in observable retailer actions after adoption, we use the SynthDid approach. The application of the method is similar to that of Section 4.2.2, but in the current analysis the dependent variables are the level of the actions taken (e.g., amount of discounts, price variation, or advertising allocation) during the observation window.

In a few cases, such as those of advertising and pricing, our data are limited to retailers that had this data available and made it accessible to the analytics service. This limits the analysis only to retailers that had connected their Google Analytics advertising data or Shopify store data at the time of the data collection, resulting in 87% of the retailers using Google Analytics for the advertising analysis and 54.3% of the retailers using Shopify for the pricing analysis. We have confirmed that the previous main results of the effect of analytics hold for each subsample.

For pricing, we analyze the transaction data of the retailers and compute five measures as the dependent variable: i) the average price for all the products sold by a retailer in a particular month;

[22] An RCT where analytics capabilities are allocated randomly to retailers will not suffice for this analysis. The mediators themselves need to be randomized as well.

ii) the average monthly product price weighted by total quantity of sales for each product; iii) the average variance in prices for each product, to reflect changes in prices; iv) the average variance in prices for each product weighted by total quantity of sales for each product; and (v) the average discount rate. None of these variables yield statistically significant differences after adoption of the analytics service (for neither users nor non-users of the dashboard).

For advertising, we use the retailers' own data collected through Google Analytics and estimated spending data collected through SpyFu.com. We look at (i) the total monthly advertising spend of a retailer reported by each of these sources; (ii) the monthly number of search terms each retailer targeted; and (iii) the number of new display ad copies each month. The Google Analytics advertising spending data do not exhibit any statistical difference in spending after adopting analytics. This is the case for all retailers regardless of analytics usage.[23] The SpyFu data also did not reveal any meaningful changes in advertising behavior after adoption.[24]

Turning to analyze the changes to the retailer's online store, we collected longitudinal data from Builtwith.com on the adoption of different web technologies by the retailers.[25] Examples of web technologies include A/B testing, advertising retargeting, product recommendation, and personalization. The data covers 97% of the retailers in our dataset. We first examine whether the overall number of different technologies installed on a website increases with the adoption of analytics, as well as how the number varies by category, for technologies related to retail management: Analytics & Tracking (565 different technologies), Advertising (532), E-commerce (263), E-mail hosting (121), and Payment (71). For the Builtwith.com data, we only observe the adoption of technology and not the usage. However, many of these technologies do not require active usage, but only installation.

Table A-5 in the Appendix reports the SynthDiD coefficients for overall technologies as well as the three leading technologies for both users and non-users.[26] Although the pre-trend matching for non-users is noisier, we still detect increases in adoption of the Analytics & Tracking and E-commerce technologies also among that group. The results suggest that, with the adoption of descriptive analytics, retailers increase the overall number of technologies installed on their

[23] We caution that the Google Analytics advertising data are sparser so they include lower coverage of time periods.

[24] We were able to identify the monthly advertising budget for 874 of the retailers (58%) and the number of terms and display ads for 486 of the retailers.

[25] This analysis excludes the focal analytics service, which is not detected by Builtwith.com.

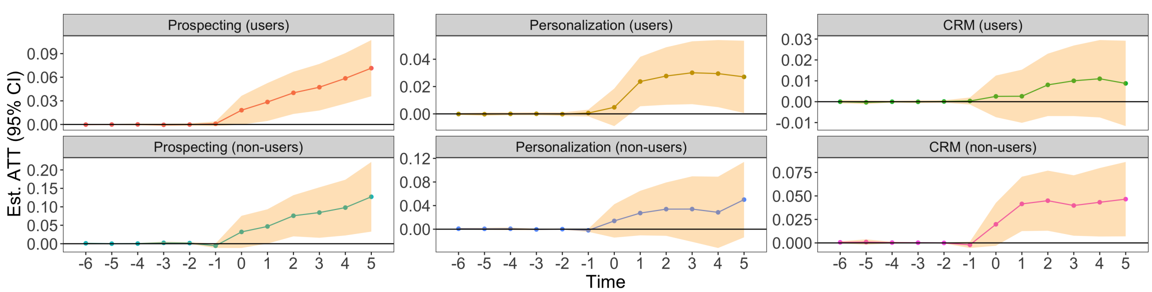
[26] The remaining two categories do not show significant effects.

online store and, in particular, adopt more Analytics & Tracking, Advertising, and E-commerce technologies.

To better understand retailers' decisions, we further break down the Analytics & Tracking, Advertising, and E-commerce categories into sub-categories of more specific technologies. We obtain six meaningful sub-categories: CRM, website design and optimization, lead generation and prospecting, personalization, other advertising technologies, and other E-commerce technologies. Prospecting technologies are focused on attracting new users to websites, while personalization technologies are focused on optimizing the on-site experience of existing visitors and retargeting previous visitors.

Using the same approach as before, we analyze the number of technologies in each sub-category that retailers use as the dependent variable. We find an increase in prospecting technologies, in personalizing technologies, and in CRM technologies, for both users and non-users. The other sub-categories do not exhibit significant changes after analytics adoption. Figure 5 displays the results for personalization, prospecting, and CRM technologies. Note that while CRM for users and personalization for non-users have a positive pattern post-adoption, they are not statistically significant.

Figure 5: Adoption of sub-technologies



ATTs (and 95% confidence intervals) of adopting analytics on peronalization, prospecting, and CRM technologies over time. Time 0 indicates month of adoption. Other times are relative to adoption.

We conclude that the adoption of descriptive analytics was associated with the adoption of additional CRM, personalization, and prospecting technologies. There is no evidence of association with price changes, advertising allocation changes, changes in advertising-spending or adoption of other technologies. Notably, the increase in the specific technology adoption exists for both adopters who used the dashboard (users) and adopters who did not use it (non-users). One explanation is that retailers were encouraged to adopt both a descriptive dashboard and CRM, personalization, and prospecting technologies. However, the fact that we only observe an increase in revenue for adopters who are users of the dashboard implies that adding these technologies by themselves would not be a main contributor to a gain in revenue.

5.2 Changes to customer behavior

The observed increased usage of CRM, personalization, and prospecting technologies is predicted to contribute to changes in customer behavior. Specifically, enhanced prospecting should increase the number of new visitors to the store and new customers (those visitors who convert) at the store. Improved personalization is predicted to increase repeat visits through retargeting, and the diversity of products purchased through recommender systems (Brynjolfsson et al. 2011b), as well as conversion rates from visits into purchases. Improved CRM technologies should increase repeat purchase rates. If pricing were to change, we would have expected to see a change in basket size (the amount purchased) and potentially the number of products purchased. Advertising optimization would have affected the number of new visitors to the store and would also have shown an improved CAC.

We use the following dependent variables to test for changes in customer behavior which are predicted by the observed changes in retailer decisions. We focus on monthly measures of: (i) average number of weekly transactions (*Transactions*); (ii) average number of weekly unique customers (*Uniques*); (iii) number of new visitors as counted by Google Analytics (*New*); (iv) average basket size in USD (*Basket*); (v) average weekly amount of revenue from repeat customers (*Repeat*); (vi) the CAC; (vii) the conversion rates from website visits to transactions[27] (*Conversion*);[28] (viii) the unique number of products sold (*Products*); and (ix) the product concentration measured using the Herfindahl Index (HHI) (*ProductHHI*).[29]

Figures 6 and 7 report the results of a SynthDiD analysis where the above variables were used as

[27] We thank an anonymous reviewer for suggesting this analysis.

[28] For *New* and *Conversion*, we could only compute three lag and four lead periods due to data constraints. While these measures are not perfect, the results are consistent with other metrics.

[29] Product-level data are limited to Shopify transactions. *ProductHHI* measures what fraction of sales each SKU generates, squares that figure, and sums those up to create a measure of concentration between 0 and 1 for each retailer-month.

dependent variables. We only report those variables that exhibit statistically significant changes in outcomes after adoption. Figure 6 presents the results only for those retailers that used the analytics service. For users, we observe an increase in *Transactions*, *Uniques*, *New*, *Repeat*, and *Conversion*, and a decrease in *ProductHHI*. The three-month ATTs are significantly different from zero for all variables, and the six-month ATTs are significantly different from zero for *Transactions*, *Repeat*, and *ProductHHI*. These changes are consistent with our predictions that CRM, personalization, and prospecting technologies contribute to a changing mix of customers in the store, but they do not necessarily change how much customers spend. Additionally, the changes in *ProductHHI* suggest that the assortment of products purchased has changed after adoption. This change in product assortment can result from two sources we observe in the results: prospecting technologies allow the retailer to better target new customers (without changing advertising budgets), which in turn increases the heterogeneity of customer tastes for products, and CRM and personalization efforts increase the likelihood of existing customers to try new products. This interpretation is consistent with previous research findings on the impact of technology adoption on product diversity in online retailing (Brynjolfsson et al. 2011b, Oestreicher-Singer and Sundararajan 2012).

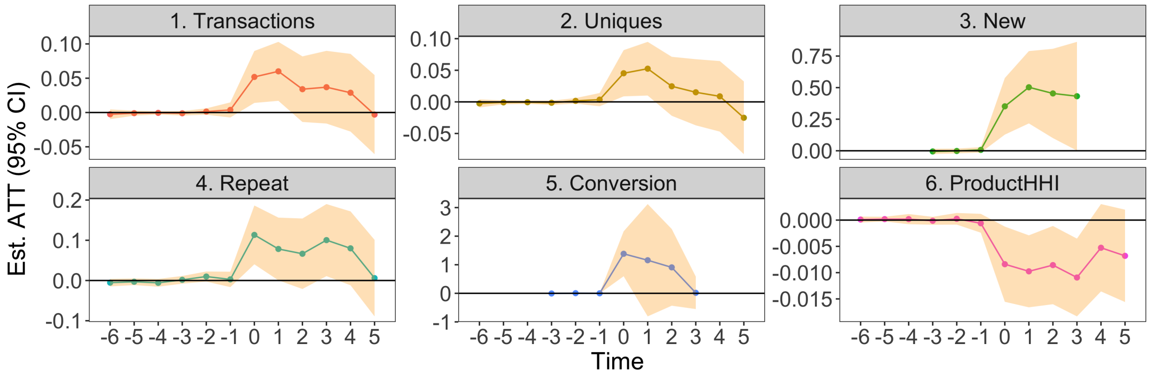
Finally, we do not find evidence for changes in *Basket* or the CAC, which is consistent with the absence of changes in prices and advertising after the adoption of analytics reported in section 5.1. The fact that we observe an increase in revenue without a change in basket size suggests that the increases are in the extensive margin of customer revenue, but not the intensive margin of profit per customer.

Figure 7 reports the results of the same analysis but for adopting retailers that did not use the dashboard.[30] As can be seen in the figures, while the pre-adoption period matching worked well for all outcome variables, retailers that did not use the service exhibit null effects for all of the variables. None of the three-month ATTs are different from zero, and among the six-month ATTs the only variable significantly different from zero is *Repeat* (estimated at -0.176 with p-value 0.071). This is particularly striking because this effect was positive for users of the dashboard, and because non-users invested in CRM technologies, which would presumably increase repeated revenues.

The results are consistent with our prediction on how different firm actions contribute to the observed gain in revenues. As before, we would caution that these results should be interpreted

[30]For *New*, we use low frequency of usage because there were not enough retailers that did not use the service to perform an analysis for this variables.

Figure 6: Customer behavior outcomes for users



ATTs (and 95% confidence intervals) of adopting analytics on customer behavior-related outcomes over time. The analyses are limited to retailers that used the dashboard (or had high frequency of usage for the *New* variable). Time 0 indicates month of adoption. Other times are relative to adoption.

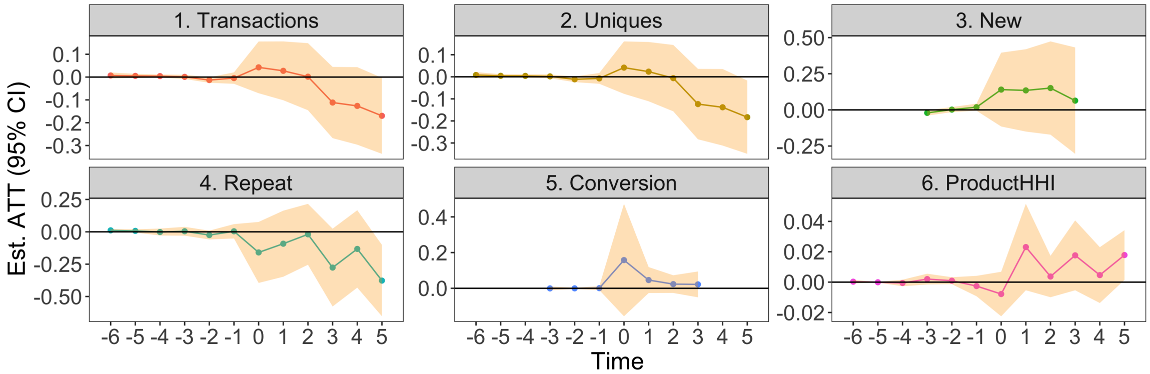
carefully because it is difficult to disentangle the effect of two simultaneous actions (e.g., pricing changes and recommender systems) that may contribute to the same outcome (e.g., changes in basket size).

5.3 Discussion

One surprising result from the mechanism analysis is that the gain in revenue after adopting descriptive analytics is not achieved through changes in pricing or advertising, which are often the easiest changes for retailers to make. Given the nature of the descriptive dashboard, however, it would indeed be hard to come up with a specific price or advertising allocation using simple KPIs. Potentially, once retailers realize that the descriptive dashboard doesn't provide specific recommendations, they might also install other technologies that automate price and advertising changes. However, we don't observe any changes that are consistent with such actions (not in pricing and advertising decisions, nor in customer behavior).

Overall, the evidence suggests that while all retailers adopt the focal analytics service and are also likely to adopt additional CRM, personalization, and prospecting technologies, only those retailers that *use* the service also exhibit an increase in the number of new visitors, a reduction in the concentration of sold products, and a corresponding increase in their performance outcomes, such as revenues and transactions. Therefore, we conclude that the descriptive dashboard and

Figure 7: Customer behavior outcomes for non-users



ATTs (and 95% confidence intervals) of adopting analytics on customer behavior-related outcomes over time. The analyses are limited to retailers that did not use the dashboard (or had low frequency of usage for the *New* variable). Time 0 indicates month of adoption. Other times are relative to adoption.

the CRM, personalization, and prospecting technologies are complementary but that they require usage to realize the benefits of the analytics service. Overall, it appears that a benefit of using a descriptive dashboard is that it allows retailers to evaluate how different technologies impact their outcomes and fine-tune them. Put together with the results on usage and the IV results, we conclude that the mechanism behind improvement in firm outcomes due to descriptive analytics adoption is the complementary mechanism.

In the context of our conclusions, there are two alternative interpretations that emerge and can potentially explain the results. First, non-users of the dashboard may also be non-users of the technologies, which may explain why they do not accrue the benefits of these technologies. Because we do not observe usage of technologies we are unable to rule that out. In that case, the benefit for users may be simply due to usage and not due to the dashboard adoption. However, our IV strategy that shifts timing of the dashboard adoption but not the technologies adoption, suggests that the improvement in firm outcomes is due, at least partially, to the adoption of analytics. In addition, many of the CRM, personalization, and prospecting technologies do not require human intervention after the initial setup, which also alleviates this concern (e.g., the Facebook Pixel technology). Second, it is possible that the decision-maker in the firm who adopts the dashboard also controls other technology adoption but does not control pricing or advertising decisions. Although we do not observe the number of employees in each firm, most of the firms in the data are small and are likely to have 1 or 2 employees. Moreover, as demonstrated in Section 4.4, most of the benefit is accrued by the smaller firms in our sample. Therefore, we believe it is typically the same person who makes all of the aforementioned decisions, which may mitigate this concern.[31]

6 Conclusion

Although the interest in marketing analytics technologies and their impact has been tremendous in the past few years, causal evidence for the efficacy of these technologies is surprisingly rare. This is partially due to lack of data, but also due to the inability to explain what drives the benefits that are observed. In this paper we describe the effect of the adoption of a descriptive analytics service by a wide variety of online retailers, and we make an effort to provide causal estimates of this effect. Our unique dataset allowed us to not only provide estimates of the value created by descriptive analytics but also to explore the mechanism behind that value creation.

The results of our analysis show that the adoption of the analytics service in our sample of firms increases weekly revenue by an average of 4%–10% in the six months after adoption. Although this range is wide and results from using multiple estimation approaches, we demonstrated that the positive effect is substantial and robust using multiple methods and alternative analyses. In addition, we provide evidence consistent with the interpretation that the dashboard benefits are accrued indirectly, likely by using the dashboard as a monitoring tool to assess the impact of other technologies.

The magnitude of our estimates is economically meaningful, at 4%–10%, suggesting an increase of USD 170–449 per week for a median retailer. Prior literature that investigated the relationship between data analytics and firm performance found a positive relationship of 3%–7% greater productivity for firms that adopt data-driven decision making or big data assets (e.g. Brynjolfsson et al. (2011a), Brynjolfsson and McElheran (2016)), and (Müller et al. 2018)). However, their adoption and outcome measures were coarser. Recent papers that investigate the adoption of analytics technologies using more detailed data found larger effect sizes, in line with our findings. Koning et al. (forthcoming) find an increase of 10% in page views after adoption of AB testing technologies by startup. Runge and Nair (2021) find an average increase of 42% in conversion rates among low-spending firms and 21.5% in high-spending firms after adopting randomized control trial tools

[31] We thank an anonymous referee for pointing out this alternative interpretation.

on Facebook. Finally, Anderson et al. (2020) find an increase of 45% in sales and 36% in profits after adoption of descriptive analytics by micro entrepreneurs in Rwanda.

There are a few potential concerns regarding the interpretation of our results. The first concern regards the generalizability of our results because we analyze the adoption of a single analytics service. Regarding this concern we note that, first, because the analytics service provides a descriptive dashboard and does not incorporate algorithmic recommendations or predictions, we believe it is representative and not very different from other descriptive solutions. Second, this specific analytics dashboard was featured as a top selection by Shopify, one of the most widely used e-commerce hosting platforms, which indicates that it is not a small analytics provider. Third, while the focus is on adoption of one focal analytics service, our data and analysis include 1,509 e-commerce firms from varying industries and countries. These are not small firms, but ones with annual revenues of $100,000 or more, and nearly all of them have been using Google Analytics prior to adoption of the service.[32] Therefore, we believe that our findings are generalizable to adoption of other descriptive dashboards by a variety of firms.

The second concern is about the endogenous timing of adoption which may cause simultaneity and threaten the causal interpretation of our findings. Potentially, firms could have selected an ideal time to adopt in which they believe that the dashboard would be most effective, or firms made many changes (e.g., hired more skilled employees) and one of these changes was adding an analytics solution, with analytics having no direct impact on firm actions or performance (the unrelated mechanism in Section 1). A unique result that we provide in our analysis is to show that the *usage* of analytics, and not its adoption per se, is what drives the improved firm performance. Further, the SynthDiD method we employ is robust to correlation between unobserved time-varying factors and a retailer's decision to adopt, and the IV strategy that shifts the timing of adoption also shows robust results, suggesting that the improvement in firm outcomes is due, at least partially, to the adoption of analytics.

A third concern is that our estimates might be biased due to selection of the retailers that chose to adopt the dashboard. Indeed, Section 4.4 shows that there is heterogeneity in the effect, with smaller retailers experiencing a larger benefit from the dashboard compared to larger one. This

[32]For comparison, in our data retailers had an average basket size of $180 with a median of $87, and annual revenues with an average of $735,000 and a median of $250,000. Shopify's retailers in 2019 had an average basket size of $67–$101 depending on region and an average annual revenue of $74,000 (Shopify 2019)

result is not surprising given that larger retailers have probably already optimized much of their operation and have less to benefit from simpler analytics. The result also alludes to the potential threat of selection bias to our analysis—the results would generalize to firms that are similar to those that chose to adopt the dashboard in our data, and might not generalize to very large firms (which are rare in our data), or those not interested in using the dashboard, as the results for non-users show. To further alleviate concerns about the causal relationship between retailer decisions and outputs, one could use causal mediation analysis (Imai et al. 2010a;b, Pearl 2014). However, the endogeneity of the mediators in our sample (firm actions) has prevented us from performing this analysis, which is left for future work. Additionally, there might be other mediating firm actions that we do not observe.

Building on the main effect that we identify, we focused on disentangling different potential avenues through which analytics may benefit retailers. The research on big data analytics (e.g., Brynjolfsson et al. 2011a, LaValle et al. 2011, Wamba et al. 2015, Akter et al. 2016, Brynjolfsson and McElheran 2016, Seddon et al. 2017) does not provide details beyond strategic and organizational considerations on how firms derive their observed benefits. Partially this is due to lack of detailed firm data, but it is also due to using aggregated data from many industries, where firms within the sample are difficult to compare. Focusing on online retailers provided a better ability to inspect these companies and their actions. Specifically, some of the major decisions for retailers are their advertising, pricing, and assortment choices. Because there isn't a clear theoretical argument for how firms should best capitalize on their investments in analytics, our findings may provide some guidance.

We do not find any changes in firms' actions with regard to pricing strategies or advertising spending, but we do find changes in the resulting assortment of purchases.[33] These changes may be due to the firm changing the inventory of products it sells, but they may also be due to the firm affecting the type of products consumers are exposed to or the type of consumers the firm attracts. Although the data cannot rule out (or support) the former explanation, the analysis of web technologies on the site provides further evidence that supports the latter—the adoption of analytics increases retailer integration of CRM, personalization, and prospecting technologies. These technological changes, coupled with changes in assortment and new visitors, are consistent

[33] Jin and Sun (2019) also do not find price changes when analytics are adopted. Their setting is competitors on one specific online shopping platform.

with the finding that the customer's basket size does not change, but the number of consumers as well as repeat revenues both increase.

One conclusion from our results is that retailers should not expect to generate actionable insights from descriptive dashboards easily. That is, descriptive analytics is not an "install and forget" solution but, rather, one that requires continuous monitoring, and from which the benefits may accrue over time with experience, but also with additional investment. The analysis shows that both users and non-users of the dashboard adopt additional technologies, but only the users of the dashboard experience benefits that these technologies are likely to provide. That is, only users saw increases in new visitors, in the number of unique products sold, in revenue from repeat customers, and overall in the number of transactions and revenue.

Why are descriptive analytics solutions so popular then? Although they rarely provide recommendations and leave the user to generate their own insights from the data, they provide retailers with a simple way to monitor and assess the performance of different decisions, thus enabling marketers to extend the range of actions they can take and to integrate new technologies. In turn, this research suggests future avenues for researchers to create better predictive and prescriptive solutions to solve these challenges for marketers.

References

A. Abadie. Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*, 59(2):391–425, 2021.

A. Abadie, A. Diamond, and J. Hainmueller. Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*, 105(490):493–505, 2010.

S. Akter, S. F. Wamba, A. Gunasekaran, R. Dubey, and S. J. Childe. How to improve firm performance using big data analytics capability and business strategy alignment? *International Journal of Production Economics*, 182:113–131, 2016.

S. J. Anderson, P. Chintagunta, R. Kaul, and N. J. Vilcassim. Impact of marketing analytics 1.0 on entrepreneur and firm performance: Field experiment evidence from Rwanda. Working paper, 2020.

D. Arkhangelsky, S. Athey, D. A. Hirshberg, G. W. Imbens, and S. Wager. Synthetic difference-in-differences. *American Economic Review*, 111(12):4088–4118, 2021.

P. Bajari, V. Chernozhukov, A. Hortaçsu, and J. Suzuki. The impact of big data on firm performance: An empirical investigation. *AEA Papers and Proceedings*, 109:33–37, 2019.

E. Ben-Michael, A. Feller, and J. Rothstein. Synthetic controls and weighted event studies with staggered adoption. *arXiv preprint arXiv:1912.03290*, 2021.

M. Bertrand, E. Duflo, and S. Mullainathan. How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics*, 119(1):249–275, 2004.

B. S. Blum and A. Goldfarb. Does the internet defy the law of gravity? *Journal of International Economics*, 70(2):384–405, 2006.

K. Borusyak, X. Jaravel, and J. Spiess. Revisiting event study designs: Robust and efficient estimation. *arXiv preprint arXiv:2108.12419*, 2021.

E. T. Bradlow, M. Gangwar, P. Kopalle, and S. Voleti. The role of big data and predictive analytics in retailing. *Journal of Retailing*, 93(1):79–95, 2017.

E. Brynjolfsson and K. McElheran. Data in action: Data-driven decision making in U.S. manufacturing. Center for Economic Studies Working Paper 16-06, 2016.

E. Brynjolfsson, L. Hitt, and H. H. Kim. Strength in numbers: How does data-driven decision making affect firm performance? Working paper, 2011a.

E. Brynjolfsson, Y. Hu, and D. Simester. Goodbye Pareto principle, hello long tail: The effect of search costs on the concentration of product sales. *Management Science*, 57(8):1373–1386, 2011b.

J. Bughin. Ten big lessons learned from big data analytics. *Applied Marketing Analytics*, 2(4): 286–295, 2017.

B. Callaway and P. H. C. Sant'Anna. Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230, 2021.

H. Datta, G. Knox, and B. J. Bronnenberg. Changing their tune: How consumers' adoption of online streaming affects music consumption and discovery. *Marketing Science*, 37(1):5–21, 2018.

C. de Chaisemartin and X. d'Haultfoeuille. Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–2996, 2020.

D. Delen and S. Ram. Research challenges and opportunities in business analytics. *Journal of Business Analytics*, 1(1):2–12, 2018.

Y. Deng, A. Lambrecht, and Y. Liu. Spillover effects and freemium strategy in mobile app market. Working paper, 2019.

F. Germann, G. L. Lilien, and A. Rangaswamy. Performance implications of deploying marketing analytics. *International Journal of Research in Marketing*, 30(2):114–128, 2013.

F. Germann, G. L. Lilien, L. Fiedler, and M. Kraus. Do retailers benefit from deploying customer analytics? *Journal of Retailing*, 90(4):587–593, 2014.

A. Goodman-Bacon. Difference-in-differences with variation in treatment timing. Technical report, National Bureau of Economic Research, 2018.

C. Han. Efficiency comparison of random effects two stage least squares estimators. *Economic Letters*, 148:59–62, 2016.

D. M. Hanssens and K. H. Pauwels. Demonstrating the value of marketing. *Journal of Marketing*, 80(6):173–190, 2016.

K. Imai, L. Keele, and D. Tingley. A general approach to causal mediation analysis. *Psychological Methods*, 15(4):309, 2010a.

K. Imai, L. Keele, and T. Yamamoto. Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science*, 25(1):51–71, 2010b.

K. Imai, I. S. Kim, and E. Wang. Matching methods for causal inference with time-series cross-section data. Working paper, 2018.

Y. Jin and Z. Sun. Information acquisition and the return to data. Working paper, 2019.

R. Koning, S. Hasan, and A. Chatterji. Experimentation and startup performance: Evidence from a/b testing. *Management Science*, forthcoming.

A. Lambrecht and C. Tucker. When does retargeting work? Information specificity in online advertising. *Journal of Marketing Research*, 50(5):561–576, 2013.

S. LaValle, E. Lesser, R. Shockley, M. S. Hopkins, and N. Kruschwitz. Big data, analytics and the path from insights to value. *MIT Sloan Management Review*, 52(2):21, 2011.

J. Lismont, J. Vanthienen, B. Baesens, and W. Lemahieu. Defining analytics maturity indicators: A survey approach. *International Journal of Information Management*, 37(3):114–124, 2017.

P. Manchanda, G. Packard, and A. Pattabhiramaiah. Social dollars: The economic impact of customer participation in a firm-sponsored online customer community. *Marketing Science*, 34 (3):367–387, 2015.

O. Mintz, Y. Bart, P. Lenk, and D. Reibstein. Drowning in metrics: How managers select and trade-off metrics for making marketing budgetary decisions. *SSRN Electronic Journal 3502600*, 2019.

O. Müller, M. Fay, and J. vom Brocke. The effect of big data and analytics on firm performance: An econometric analysis considering industry characteristics. *Journal of Management Information Systems*, 25(2):488–509, 2018.

G. Oestreicher-Singer and A. Sundararajan. Recommendation networks and the long tail of electronic commerce. *MIS Quarterly*, 36(1):65–83, 2012.

K. Pauwels, T. Ambler, B. H. Clark, P. LaPointe, D. Reibstein, B. Skiera, B. Wierenga, and T. Wiesel. Why, what, how, and what research is needed? *Journal of Service Research*, 12(2): 175–189, 2009.

J. Pearl. Interpretation and identification of causal mediation. *Psychological Methods*, 19(4):459, 2014.

J. Runge and H. Nair. Exploration in action: The role of randomized control trials in online demand generation. *SSRN Electronic Journal 3794028*, 2021.

P. B. Seddon, D. Constantinidis, T. Tamm, and H. Dod. How does business analytics contribute to business value? *Information Systems Journal*, 27(3):237–269, 2017.

Shopify. Shopify's state of commerce report, June 2019. https://news.shopify.com/shopify-unveils-first-state-of-commerce-report, accessed November 1, 2020.

S. F. Wamba, S. Akter, A. Edwards, G. Chopin, and D. Gnanzou. How "big data" can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics*, 165:234–246, 2015.

K. Wang and A. Goldfarb. Can offline stores drive online sales? *Journal of Marketing Research*, 54(5):706–719, 2017.

M. Wedel and P. Kannan. Marketing analytics for data-rich environments. *Journal of Marketing*, 80(6):97–121, 2016.

J. Wooldridge. Fixed-effects and related estimators for correlated random-coefficient and treatment-effect panel data models. *Review of Economics and Statistics*, 87(2):385–390, 2005.

J. Wooldridge. *Econometric analysis of cross section and panel data*, pages 937–940. The MIT Press, 2010.

J. Wooldridge. Correlated random effects models with unbalanced panels. *Journal of Econometrics*, 211(1):137–150, 2019.

J. Wooldridge. Two-way fixed effects, the two-way mundlak regression, and difference-in-differences estimators. *Available at SSRN 3906345*, 2021.

A Appendix

A.1 Constructing the SynthDiD Event Study Plot

SynthDiD is designed to minimize the MSE of a target estimated ATT and not to separately measure effects in specific time periods. The event study plot we construct is therefore used to illustrate the effect, although aggregating the standard errors that appear in this plot into an ATT requires taking serial correlation into account which is done automatically by SynthDiD but cannot be done using the standard errors from the plot (or from the associated period-by-period standard errors we report in Appendix A.3).

To compute the treatment effects pre-adoption, we compute for each adoption cohort r and each time period t between $r + \ell_{\min}$ and $r - 1$:

$$\widehat{\tau}_{r(t-r)} = \left(\frac{\sum_{i \in N_r^{tr}} log(Y_{ijt} + 1)}{|N_r^{tr}|} - \left(\widehat{\omega}_0 + \sum_{i \in N_r^{co}} \widehat{\omega}_{ij} log(Y_{ijt} + 1) \right) \right) \cdot \widehat{\lambda}_t.$$

Because the weights $\widehat{\lambda}_t$ for time periods before adoption sum up to 1, summing up $\widehat{\tau}_{r(t-r)}$ yields values which are approximately zero, which shows a good fit between treatment outcomes and the synthetic control pre-adoption. The standard errors for these values are computed using the jackknife method (or in case there is only one treated unit, the placebo method). The values for each cohort are then averaged, and the standard errors are aggregated appropriately.

Computing the cohort-level effects post-adoption is done in a similar manner. For each adoption cohort r and each time period t between r and $r + \ell_{\max}$ we compute:

$$\widehat{\tau}_{r(t-r)} = \left(\frac{\sum_{i \in N_r^{tr}} log(Y_{ijt} + 1)}{|N_r^{tr}|} - \left(\widehat{\omega}_0 + \sum_{i \in N_r^{co}} \widehat{\omega}_{ij} log(Y_{ijt} + 1) \right) \right).$$

In this case we do not weight the estimated effect by $\widehat{\lambda}_t$ because post-adoption the weight is just $\frac{1}{\ell_{\max+1}}$. Averaging the effects within cohorts and the resulting averages across cohorts then produces the ATT reported by SynthDiD. The standard error for the aggregate ATT is computed using the jackknife method (or the placebo method as before) directly for the aggregated value, which takes into account potential serial correlation between effects across time.

A.2 Computing subgroup effects using SynthDiD

There are three possible approaches to compute subgroup effects using SynthDiD. The first is to split the data into separate subsets for each subgroup and compute the SynthDiD effects for each subset. This approach therefore compares each treated unit to control units in the same subgroup. The disadvantage of this method is that each subset will use a different control group resulting in a different synthetic control and thus the effects cannot be compared across subgroups.

The second approach is to use a control group comprised of all non-adopters (regardless of subgroup membership) in each balanced panel cohort analysis, and then switch the treatment units to the subgroup to be analyzed and perform a SynthDiD analysis. The advantage of this method is that the control weights $\widehat{\omega}_i$ computed for each treated subgroup will match the pre-trends the best, but the disadvantage is that every control unit will receive a different weight for each treatment subgroup being analyzed, which again makes comparing the results harder, as the synthetic control used to estimate effects is different.

The third approach, which is the one we implement and report,[34] uses all the data to estimate the SynthDiD control weights $\widehat{\omega}_i$, but it then computes the treatment effects $\widehat{\tau}_{r(t-r)}$ which are defined in Section A.1 using *only* the treated subgroup units as N_r^{tr}. This has the advantage that all units have the same synthetic control, but it has the disadvantage that the pre-trend match might not be as good, because it was designed to match the average outcome of all treated units and not just a subgroup.

A.3 Additional details and analyses using SynthDiD

Table A-1 shows the number of treated and control units in each cohort.

[34] The first approach yielded similar but noisier results.

Table A-1: Number of units used in SynthDiD analysis

Cohort	Treatment	Control
Jan 2016	57	431
Feb 2016	131	446
Mar 2016	58	418
Apr 2016	72	406
May 2016	71	430
Jun 2016	67	417
Jul 2016	55	413
Aug 2016	49	435
Sep 2016	53	497

Cohort	Treatment	Control
Oct 2016	49	494
Nov 2016	67	538
Dec 2016	56	524
Jan 2017	30	551
Feb 2017	18	557
Mar 2017	39	566
Apr 2017	33	576
May 2017	45	539
Jun 2017	32	494
Average	54.6	485

A.3.1 Main Effect: Robustness

We perform a series of robustness tests to our main effect results. These are available in Table A-2. In this table, column "Baseline" presents the results corresponding to Figure 3 in the paper.

First, due to the importance of pre-period synthetic control matching, we show that our effects are robust to the choice of the number of lags and leads; we display overall robust results for additional 12-month windows with seven, eight, and nine pre-adoption lags (see columns "Lag-1," "Lag-2," and "Lag-3" in the appropriate tables).

Second, to verify that the effects we measure are not due to the aggregation at the retailer-platform level, we aggregate the data to the retailer level, redefine "after adoption" to occur once the first hosting platform data are added to the dashboard, and estimate the models at the retailer level. Corresponding results appear in each table as the "Company-level" column.

Finally, to demonstrate the generalizability of our results we collected additional data on 1,091 retailers that adopted the service in 2018. For this sample we only observe a subset of the variables and thus use it only for robustness tests of our main effect. We use these firms to repeat the analysis without the 2016 adopters, using all of 2017 retailers as the treatment group and the 2018 retailers as the control-only group. To ensure sufficient overlap, we restrict the sample to include retailers that had at least one observation in 2016. This yielded 765 firms out of the 958 firms that adopted the dashboard in 2017 as treated (833 retailer–channel combinations), and for 2018 we could use 757 firms out of 1,091 (886 retailer–channel combinations). Column "2017/2018 sample" reports the results. The results are similar to the baseline, albeit with slightly larger effect sizes post-adoption. Additionally, the effect stays significantly positive for one additional time period.

As the table shows, the SynthDiD method matches the treatment and control units such that the difference between them is indistinguishable from zero for all of the different specifications.

Table A-2: SynthDiD Robustness

Lag	Baseline	Lag-1	Lag-2	Lag-3	Company-level	2017/2018 sample
-9				-0.0003 (0.002)		
-8			-0.001 (0.002)	-0.0005 (0.002)		
-7		-0.001 (0.002)	-0.0004 (0.002)	0.00002 (0.001)		
-6	-0.001 (0.003)	-0.001 (0.002)	-0.001 (0.001)	-0.0003 (0.001)	-0.001 (0.003)	-0.001 (0.003)
-5	0.00000 (0.002)	-0.00001 (0.002)	-0.0001 (0.001)	0.00003 (0.001)	0.00000 (0.002)	-0.0002 (0.001)
-4	-0.0004 (0.002)	-0.0001 (0.001)	0.0001 (0.001)	0.00004 (0.001)	-0.0003 (0.002)	-0.001 (0.001)
-3	-0.001 (0.002)	-0.001 (0.002)	-0.001 (0.002)	-0.001 (0.002)	-0.001 (0.002)	0.0001 (0.002)
-2	-0.001 (0.003)	-0.001 (0.003)	0.0002 (0.002)	-0.0001 (0.003)	-0.001 (0.003)	-0.003 (0.003)
-1	0.004 (0.005)	0.004 (0.005)	0.002 (0.006)	0.002 (0.005)	0.004 (0.006)	0.005 (0.006)
0	0.056** (0.019)	0.053** (0.019)	0.048* (0.019)	0.036+ (0.019)	0.057** (0.020)	0.087** (0.022)
1	0.062** (0.022)	0.059** (0.022)	0.041+ (0.022)	0.017 (0.022)	0.060** (0.023)	0.075** (0.027)
2	0.050* (0.024)	0.037 (0.024)	0.021 (0.025)	-0.009 (0.025)	0.036 (0.026)	0.082** (0.030)
3	0.036 (0.026)	0.037 (0.026)	0.013 (0.027)		0.023 (0.029)	0.062* (0.031)
4	0.034 (0.028)	0.030 (0.029)			0.014 (0.030)	0.049 (0.031)
5	-0.010 (0.029)				-0.029 (0.031)	0.042 (0.032)
3 months	0.056** (0.019)	0.050** (0.019)	0.037* (0.019)	0.014 (0.018)	0.051* (0.020)	0.081** (0.023)
6 months	0.038+ (0.020)				0.027 (0.022)	0.066** (0.024)

Significance level: 10% (+); 5% (*); 1% (**).

A.3.2 Additional tables

The tables in this section present period-by-period estimates and the three- and six-month ATTs. Table A-3 presents estimates of the effect of adoption on revenues, conditional on usage intensity, and Table A-4 presents estimates of the effect conditional on firm size.

Table A-5 presents the analysis of sub-technologies for the mechanism section.

Table A-3: Usage results

	A: Usage dummy		B: Usage Intensity		C: Report Usage		D: Report Intensity	
Lag	Users	Non-users	Below	Above	Users	Non-users	Below	Above
-6	-0.002	0.007	0.001	-0.004	-0.002	0.0005	0.001	-0.004
	(0.003)	(0.006)	(0.004)	(0.006)	(0.004)	(0.006)	(0.004)	(0.005)
-5	-0.001	0.010+	0.001	-0.002	-0.001	0.006	0.003	-0.004
	(0.002)	(0.006)	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)	(0.004)
-4	-0.001	0.003	-0.003	0.003	0.0004	-0.003	-0.002	0.002
	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.005)	(0.002)	(0.002)
-3	-0.001	0.0004	-0.006*	0.006*	-0.002	0.001	-0.006*	0.005+
	(0.002)	(0.006)	(0.003)	(0.003)	(0.002)	(0.004)	(0.003)	(0.003)
-2	0.001	-0.017*	-0.003	0.002	-0.001	-0.004	-0.004	0.002
	(0.003)	(0.008)	(0.004)	(0.004)	(0.003)	(0.007)	(0.004)	(0.004)
-1	0.005	-0.003	0.010	-0.005	0.005	-0.001	0.008	-0.002
	(0.006)	(0.014)	(0.007)	(0.009)	(0.006)	(0.011)	(0.007)	(0.009)
0	0.070**	-0.037	0.039	0.086**	0.069**	0.008	0.049+	0.068*
	(0.021)	(0.057)	(0.025)	(0.033)	(0.023)	(0.043)	(0.025)	(0.032)
1	0.070**	-0.017	0.028	0.111**	0.070**	0.018	0.042	0.084*
	(0.024)	(0.069)	(0.029)	(0.035)	(0.025)	(0.055)	(0.031)	(0.033)
2	0.054*	-0.097	0.0002	0.094*	0.046+	-0.006	0.003	0.082*
	(0.027)	(0.080)	(0.033)	(0.040)	(0.028)	(0.062)	(0.033)	(0.040)
3	0.043	-0.123	-0.029	0.106*	0.037	-0.034	-0.032	0.099*
	(0.030)	(0.090)	(0.036)	(0.046)	(0.031)	(0.068)	(0.037)	(0.044)
4	0.042	-0.183+	-0.043	0.108*	0.033	-0.059	-0.029	0.075
	(0.031)	(0.094)	(0.037)	(0.050)	(0.033)	(0.071)	(0.037)	(0.048)
5	-0.002	-0.226*	-0.103**	0.090+	-0.0001	-0.147*	-0.089*	0.055
	(0.032)	(0.096)	(0.038)	(0.051)	(0.033)	(0.075)	(0.039)	(0.049)
3 months	0.065**	-0.050	0.022	0.097**	0.062**	0.007	0.031	0.078**
	(0.021)	(0.060)	(0.025)	(0.031)	(0.022)	(0.047)	(0.026)	(0.030)
6 months	0.046*	-0.114+	-0.018	0.099**	0.042+	-0.037	-0.009	0.077*
	(0.023)	(0.068)	(0.027)	(0.035)	(0.024)	(0.052)	(0.028)	(0.033)

Significance level: 10% (+); 5% (*); 1% (**).

Table A-4: Heterogeneity by size

		Revenue medians		Transaction medians	
Lag	Sample	Below	Above	Below	Above
-6	-0.0001	-0.015	0.014+	-0.014	0.012
	(0.006)	(0.009)	(0.009)	(0.010)	(0.008)
-5	0.0005	-0.002	0.003	-0.008	0.008
	(0.004)	(0.007)	(0.004)	(0.005)	(0.005)
-4	-0.001	-0.005	0.003	-0.008+	0.005**
	(0.002)	(0.004)	(0.002)	(0.004)	(0.001)
-3	0.001	-0.001	0.003	0.004	-0.001
	(0.003)	(0.005)	(0.003)	(0.004)	(0.003)
-2	-0.003	0.005	-0.010	0.003	-0.008
	(0.005)	(0.007)	(0.007)	(0.006)	(0.007)
-1	0.003	0.018	-0.013	0.023+	-0.016
	(0.009)	(0.014)	(0.012)	(0.013)	(0.012)
0	0.056	0.142*	-0.014	0.227**	-0.076*
	(0.034)	(0.061)	(0.038)	(0.061)	(0.036)
1	0.101*	0.201*	0.013	0.223**	0.002
	(0.041)	(0.079)	(0.038)	(0.079)	(0.038)
2	0.119*	0.254*	0.010	0.265**	0.012
	(0.055)	(0.099)	(0.058)	(0.096)	(0.061)
3	0.113*	0.278**	-0.034	0.252**	-0.001
	(0.055)	(0.098)	(0.055)	(0.089)	(0.069)
4	0.086	0.240+	-0.045	0.193+	0.004
	(0.062)	(0.125)	(0.056)	(0.116)	(0.070)
5	0.038	0.234*	-0.126*	0.159	-0.052
	(0.062)	(0.114)	(0.064)	(0.106)	(0.078)
3 months	0.092*	0.199**	0.002	0.238**	-0.021
	(0.038)	(0.073)	(0.033)	(0.071)	(0.034)
6 months	0.086*	0.224**	-0.033	0.220**	-0.019
	(0.043)	(0.082)	(0.038)	(0.077)	(0.047)

Significance level: 10% (+); 5% (*); 1% (**).

Column 1 reports the period-by-period estimates for the sample of firms that adopted the service between January 2017 and June 2017. The next columns use the same sample of firms and separate them to subgroups based on performance in 2015.

Table A-5: Technology Adoption

	Overall		Analytics&Tracking		Advertising		E-commerce	
Lag	Users	Non-users	Users	Non-users	Users	Non-users	Users	Non-users
-6	0.00003	-0.001	-0.0001	0.0003	-0.00005	0.001	-0.001	0.001+
	(0.0004)	(0.001)	(0.001)	(0.0004)	(0.0002)	(0.001)	(0.001)	(0.001)
-5	-0.0001	0.001*	0.00004	0.0001	-0.00002	-0.0001	0.0001	-0.002
	(0.0002)	(0.0003)	(0.0004)	(0.0004)	(0.0003)	(0.0001)	(0.001)	(0.002)
-4	0.00000	0.00000	-0.0002	0.0005	-0.00001	-0.0002	0.0001	-0.0001
	(0.00001)	(0.00000)	(0.001)	(0.001)	(0.0003)	(0.0003)	(0.0002)	(0.001)
-3	-0.0001	0.0004	-0.0001	0.00000	0.0002	-0.0002	-0.0001	-0.0002
	(0.0002)	(0.0005)	(0.0002)	(0.0003)	(0.001)	(0.001)	(0.0003)	(0.001)
-2	-0.0001	-0.0001	-0.0002	-0.001	0.001	-0.002	-0.0001	0.0002
	(0.0003)	(0.0005)	(0.0004)	(0.001)	(0.001)	(0.003)	(0.0003)	(0.001)
-1	0.0002	-0.0001	0.001	0.0001	-0.001	0.002	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)	(0.004)
0	0.004	0.017+	0.013*	0.026	0.019+	0.009	0.007	0.006
	(0.004)	(0.009)	(0.006)	(0.017)	(0.010)	(0.024)	(0.005)	(0.012)
1	0.010+	0.021*	0.022*	0.053*	0.048**	0.021	0.011	0.005
	(0.006)	(0.010)	(0.009)	(0.021)	(0.013)	(0.023)	(0.007)	(0.013)
2	0.012+	0.022+	0.029**	0.075**	0.047**	0.026	0.019*	0.026
	(0.006)	(0.012)	(0.010)	(0.023)	(0.015)	(0.027)	(0.008)	(0.020)
3	0.008	0.010	0.025*	0.067*	0.052**	0.019	0.021*	0.038
	(0.007)	(0.014)	(0.011)	(0.026)	(0.017)	(0.033)	(0.009)	(0.025)
4	0.006	0.001	0.025*	0.067*	0.048**	0.005	0.022*	0.055*
	(0.008)	(0.015)	(0.011)	(0.029)	(0.017)	(0.034)	(0.009)	(0.026)
5	0.002	-0.0004	0.030*	0.058+	0.060**	-0.007	0.019+	0.041
	(0.009)	(0.016)	(0.013)	(0.032)	(0.019)	(0.037)	(0.010)	(0.027)
3 months	0.009+	0.020*	0.021**	0.051**	0.038**	0.018	0.012*	0.012
	(0.005)	(0.009)	(0.008)	(0.019)	(0.011)	(0.022)	(0.006)	(0.013)
6 months	0.007	0.012	0.024**	0.058**	0.045*	0.012	0.016*	0.028+
	(0.006)	(0.011)	(0.009)	(0.022)	(0.013	(0.026)	(0.007)	(0.017)

Significance level: 10% (+); 5% (*); 1% (**).

Web Appendix—IV Analysis

WA-1 Construction of IVs and identifying assumptions

We construct three instruments: an industry instrument, a platform instrument, and a geographic region instrument. The industry instrument calculates for each retailer in each period t what fraction of retailers within their industry (excluding themselves) already adopted the dashboard. This would capture industry time trends that possibly affect the retailer's adoption decision but are likely uncorrelated with their individual performance.

The other two instruments are constructed using traffic to the analytics provider's website and not to the individual retailer websites. New visitors to the analytics provider's website are likely online retailers that are potential clients of the dashboard and not end consumers who buy items at the retailers' websites. Figure W-1 in Web Appendix WA-2 illustrates the traffic we leverage. The number of unique *new visitors* to the analytics provider's website measures how many retailers are interested in potentially signing up for the analytics service in that specific month. We expect such increased attention to be correlated with earlier adoption by adopting retailers. The variation in the number of new visitors to the analytics provider's website stems from online mentions, articles, and advertising about the provider's capabilities, and should be uncorrelated with the performance of a specific retailer. Hence, we also expect the exclusion restriction to hold for this instrument. For the platform instrument, we use the the number of unique new visitors that arrived at the service provider's website through three hosting platforms: Shopify, BigCommerce, and Other. The analytics dashboard was periodically listed as a top popular service for retailers on some of the hosting platforms, and this IV captures variation in attention from these appearances. Such prominence would have increased attention to the dashboard and was likely to influence a retailer's adoption timing without affecting performance except via adoption.[1] Figure W-2 in Appendix WA-2 illustrates the variation in this instrument.

The geographic instrument is similar to the platform instrument but measures new visitors from the geographic region of the retailer's headquarters.[2] In this case, this variable captures the effect

[1] We omitted traffic from channels with targeted advertising because they could violate the exclusion restriction by attracting retailers that were already interested in adopting analytics.

[2] We were unable to identify the location of 102 of the retailers. Thus we define this variable for 1,059 of the retailers, which represent 1,267 of the retailer–channel combinations.

of local media and "buzz" about the analytics service, which in turn is more likely to be consumed by retailers of closer geographic proximity (Blum and Goldfarb 2006). Table W-1 in Appendix WA-2 displays the definition of regions and distribution of retailers in our sample, and Figure W-3 in Appendix WA-2 illustrates the variation in this instrument.

A potential threat to the validity of the exclusion restriction for the geographic IV is country-specific performance waves that cause companies in regions that do well financially to both visit the analytics provider's website and to sign up. This concern is mitigated thanks to the short time span of the data, which did not exhibit any such waves to the best of our knowledge. Additionally, our results are robust to adding region fixed effects.[3] We also use the platform IV which is less likely to suffer from hosting platform performance waves. Industry-level performance waves are mitigated using industry fixed effects.

WA-2 IV estimation procedure

In the first stage, we estimate a random effects probit model for the decision to adopt the analytics service. We estimate the following model:

$$Pr(Adopt_{ijt} = 1) = \Phi\left(\alpha + \beta X_{ijt-1} + \gamma Z_{ijt} + \delta \bar{X}_{ij} + \theta \bar{Z}_{ij} + \gamma_t + \mu_{ij}\right) \tag{1}$$

where $Adopt_{ijt}$ indicates whether retailer i adopted the service in channel j in month t. Observations after adoption time t are not used to estimate Equation (1) because the adoption decision is made once. X_{ijt-1} are retailer–channel–time control variables that include eight industry dummies, six channel dummies, the number of other channels of retailer i (excluding channel j), and the lag DV of the relevant DV from the next stage for each regression. Z_{ijt-1} are the instrumental variables, which indicate the adoption rate of the analytics service among retailers within the same industry (industryAdoption), and the number of new visitors to the service provider's website from the hosting platform (referralVisits) and from the geographic region (regionVisits). Following Deng et al. (2019), and as suggested by Wooldridge (2005) and Wooldridge (2019), we include $\bar{X}_{ij}$ and $\bar{Z}_{ij}$, the mean value of variables per retailer–channel for time-varying variables. γ_t are month-year fixed effects. Due to the large number of retailers, and to address the incidental parameter problem

[3] We do not include the 65 region fixed effects in our main regression specification due to efficiency concerns and due to the incidental parameter problem in probit analysis. However, the main effect results using region fixed effects are virtually identical to the reported results.

with fixed effects in probit models, we include μ_{ij} as retailer–channel random effects instead of fixed effects.

Following the first stage, the predicted probability of adoption, $\widehat{Adopt}_{ijt}$, is used as an instrument for the endogenous adoption indicator $AfterTreatment_{ijt}$ in a two-stage least squares regression. This regression comprises the second and third stages of our estimation procedure, and is specified as:

$$log(Y_{ijt}+1) = \alpha + \beta \mathbf{AfterTreatment_{ijt}} + \delta X_{ijt-1} + \theta \bar{X}_i j + \gamma_t + \mu_{ij} + \epsilon_{ijt} \tag{2}$$

where the variables are defined as in the previous equations for each retailer i with channel j in month t. When estimating this equation we use all observations up to the first observation after adoption. Because we use an IV to alleviate endogeneity concerns, we use all adoptions that occur in 2016 or 2017 to estimate equations (1) and (2). Note that this approach allows us to incorporate month-year fixed effects because there are adoptions and controls in each month and year (up to January 2018). These specifications increase the number of adoptions in our data (because all retailers adopt) compared to the limitation we had in the sample in Section 4.2.1. For each retailer–channel, the controls are every retailer–channel that have not yet adopted, and the effects are identified off individual retailer–channel variation.

Figure W-1: Traffic to the service provider's website

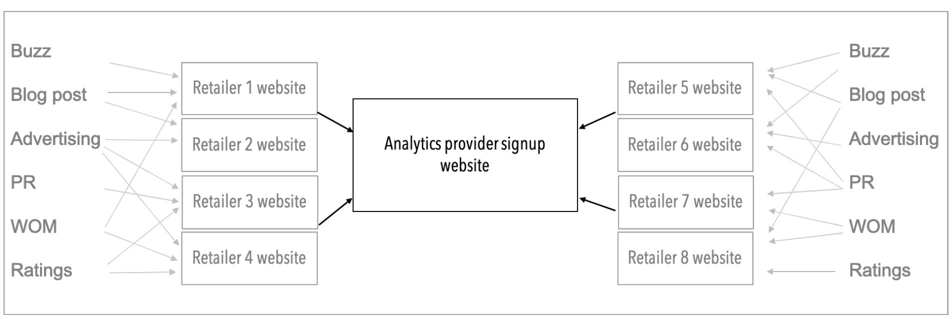


Illustration of the traffic leveraged as IVs. We use the traffic to the focal service provider sign-up page (illustrated with the black thick arrows) as a proxy for various "buzz" channels about the analytics service provider (gray thin arrows).

Table W-1 presents the distribution of retailers' geographic regions. Figures W-2 and W-3 present the variation in the traffic-based instruments over time.

Table W-1: Distribution of retailers' regions

Region	Frequency	Percent	**Region**	Frequency	Percent
California	216	20.4%	New Jersey	14	1.3%
United Kingdom	99	9.3%	Washington	14	1.3%
Australia	85	8.0%	Oregon	13	1.2%
Canada	71	6.7%	Colorado	12	1.1%
New York	55	5.2%	Georgia	12	1.1%
Texas	28	2.6%	Illinois	12	1.1%
Virginia	28	2.6%	Israel	11	1.0%
Florida	25	2.4%	Utah	11	1.0%
New Zealand	21	2.0%	Michigan	10	0.9%
Germany	17	1.6%	North Carolina	10	0.9%
Pakistan	15	1.4%	Pennsylvania	10	0.9%
India	14	1.3%	Other	256	24.2%
			Total	1,059	100.0%

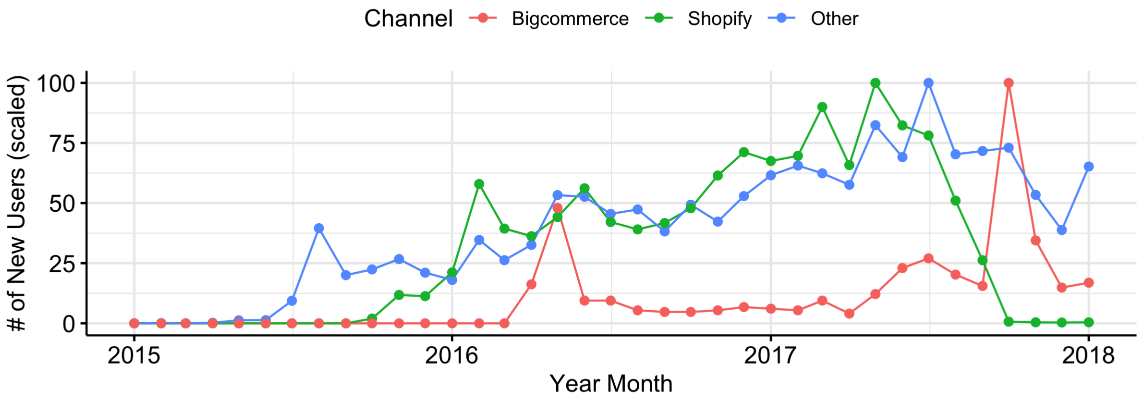
Distribution of retailers' geographic regions. In the US, we report states; otherwise countries are reported. In this table, locations with fewer than 10 retailers were consolidated into "other" for confidentiality reasons and in the interest of space. In the estimation we use the actual locations and not the "other" category.

WA-3 IV estimation results

Panel A of Table W-2 presents the first-stage results and shows that our IVs are positively correlated with adoption as expected. Prior to estimating the two-step IV, we perform a Durbin-Wu-Hausman endogeneity test using the constructed IV, $\widehat{Adopt}_{ijt}$. The test produces p-values consistent with evidence of endogeneity (p-value < 0.0001). We also confirm that our instruments are not weak using Stock-Yogo, Anderson-Rubin, and Stock-Wright tests.

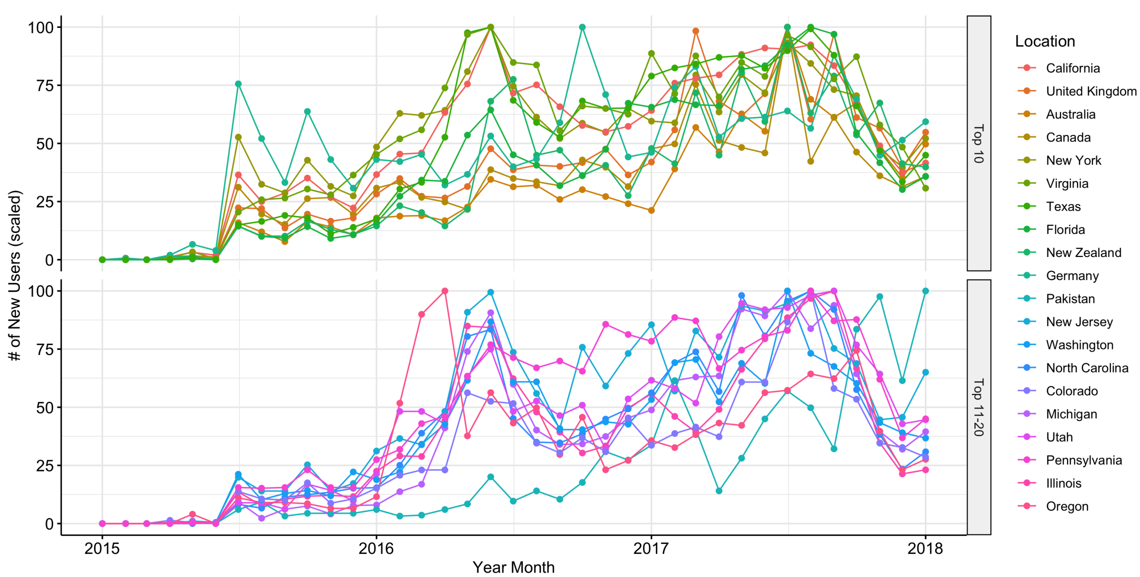
Panel B of Table W-2 reports the second-stage results of our IV estimation using the Baltagi error component 2SLS estimators (EC2SLS) (Han 2016). The F-statistic on the instrument in the second stage is significantly different from zero. Additionally, the IV variable $\widehat{Adopt}_{ijt}$ is positively associated with adoption. Column 1 of Table W-3 reports the third-stage IV results, and Column 2 reports comparable OLS estimates using the same sample as the IV analysis.

Figure W-2: Hosting platform-based instrument over time



The variation in traffic we leverage for our hosting platform-based IV. For each channel, we plot the number of new users that arrived from that channel, scaled such that the maximum number of users is 100 for each channel.

Figure W-3: Location-based instrument over time



The variation in traffic we leverage for our location-based IV. For the top 20 locations in terms of number of companies in the data, we plot the number of new users that arrived from that location, scaled such that the maximum number of users is 100 for each location.

Table W-2: IV first and second steps

Panel A: First-Stage Probit (DV: Adoption dummy)	
Region visits	.00028+
	(.00015)
Referral visits	.000081**
	(.00001)
Industry adoption	3.4*
	(1.4)
Log (Lag revenue)	.11**
	(.041)
# other channels	-.23*
	(.098)
Panel B: Second Stage (DV: Adoption dummy)	
$\widehat{Adopt}_{ijt}$	.143**
	(.019)
F-Stat	199.60

Table W-3: IV and OLS estimates

Specification	IV (third stage)	OLS
After adoption	.167**	.076**
	(.05)	(.019)
Log (Lag revenue)	.59**	.54**
	(.016)	(.017)
# other channels	.0053	.004
	(.010)	(.013)
Observations	12,115	12,115

Significance level: 10% (+); 5% (*); 1% (**).
All models include (i) retailer–channel random effects, (ii) industry, channel, and month fixed effects, and (iii) controls for retailer–channel means.