MANAGEMENT SCIENCE
Vol. 68, No. 5, May 2022, pp. 3528–3548
ISSN 0025-1909 (print), ISSN 1526-5501 (online)
http://pubsonline.informs.org/journal/mnsc

Spillovers from Online Engagement: How a Newspaper Subscriber's Activation of Digital Paywall Access Affects Her Retention and Subscription Revenue

Adithya Pattabhiramaiah,[a] Eric Overby,[b] Lizhen Xu[b]

[a] Marketing, Scheller College of Business, Georgia Institute of Technology, Atlanta, Georgia 30308; [b] Information Technology Management, Scheller College of Business, Georgia Institute of Technology, Atlanta, Georgia 30308

Contact: adithyap@gatech.edu, https://orcid.org/0000-0002-0501-9937 (AP); eric.overby@scheller.gatech.edu, https://orcid.org/0000-0001-5285-4436 (EO); lizhen.xu@scheller.gatech.edu, https://orcid.org/0000-0002-8603-9891 (LX)

Received: June 21, 2019
Revised: September 16, 2020; December 21, 2020
Accepted: January 22, 2021
Published Online in Articles in Advance: September 8, 2021

https://doi.org/10.1287/mnsc.2021.4092

Abstract. Newspapers are increasingly reliant on subscription revenue as advertising spend shifts to online platforms. Many newspapers have implemented paywalls in an attempt to boost subscription revenue. We study whether and how paywalls can help newspapers boost subscription revenue by retaining existing subscribers. Most major newspapers offer free access to paywalled content to subscribers to the print edition, which may help the newspaper retain subscribers by making their subscriptions more valuable. We leverage variation in whether and when existing subscribers activated access to the paywall of a top 30 North American newspaper. Our identification strategy accounts for self-selection in subscribers' decisions to activate paywall access. We find that a subscriber's activation of digital access decreases the risk of her canceling her subscription by about 31% and increases her subscription revenue by 7%–12%. In other words, digital activation improves subscriber retention and the associated subscription revenue. This suggests a crosschannel spillover in which the online product (the paywalled website) increases customers' valuation for the offline product (the printed newspaper). Our results have implications not only for the newspaper industry but also for firms in other industries that offer subscribers to one product free or subsidized access to a complementary product.

History: Accepted by Kartik Hosanagar, information systems.
Supplemental Material: The data files are available at https://doi.org/10.1287/mnsc.2021.4092.

Keywords: digital monetization • newspaper paywalls • customer retention • matching • synthetic control • causal forest

1. Introduction

As advertising spend has shifted from traditional media to online platforms, newspapers have become increasingly reliant on subscription revenue.[1] A key strategy for newspapers to boost subscription revenue is to implement a digital paywall that requires visitors to the newspaper's website to have a subscription to access content. Indeed, approximately 70% of newspaper sites in the United States and Europe had implemented paywalls by 2019 (Simon and Graves 2019b). There are two basic ways that a paywall can help a newspaper increase subscription revenue. First, the paywall can help the newspaper attract new subscribers. For example, the paywall may prompt visitors who would otherwise consume the newspapers' content for free to subscribe. Second, the paywall can help the newspaper retain existing subscribers. Many newspapers (including the *New York Times*, *Chicago Tribune*, *Washington Post*, *Seattle Times*, and *Atlanta Journal-Constitution*) offer free access to the paywall to subscribers to the print newspaper. This strategy of offering full digital access along with a print subscription may help retain existing print subscribers by increasing the value of their subscriptions, provided that subscribers activate the digital access.

Although both customer acquisition and customer retention are important to increase newspapers' subscription revenue, we focus on retention. Specifically, we study whether an existing newspaper subscriber's activation of digital paywall access affects whether she maintains her subscription (i.e., retention) and the subscription revenue that she generates for the newspaper. Our data contain individual-level subscriber records for each week from February 2013 to March 2017 for a major North American newspaper ranked in the top 30 by circulation. The newspaper implemented a paywall during the early part of the study period that limited access to certain parts of its website. As is the case with the majority of paywall-implementing newspapers in the United States, the newspaper used a bundling strategy in which existing print subscribers were provided free and unlimited access to content on the website, assuming they activated this access. A key strength of our data is that we observe whether and when each subscriber activated digital paywall access. This allows us to measure whether and to what extent a subscriber's activation of digital access affects her retention and the associated subscription revenue.

A challenge for our analysis is that whether a subscriber activates digital access is not random. Subscribers who activate (whom we refer to as activators) and those who do not (nonactivators) may have underlying differences—beyond whether they activated digital access—that influence their subscription behaviors. Our identification (ID) strategy accounts for this potential self-selection problem in multiple ways, including by matching activators and nonactivators with similar consumption preferences (such as whether they receive the paper daily or on weekends only) and subscription patterns (such as subscription term length, whether they pay their subscriptions on time, etc.), controlling for subscriber fixed effects, using instrumental variables methods, and conducting falsification tests. We find that a subscriber's activation of digital access decreases the chance of her canceling her subscription by about 31%. In other words, digital activation improves subscriber retention. Furthermore, the vast majority of activators retained their print + digital subscriptions after activating, rather than switching to digital-only subscriptions. The very few activators who switched to digital only tended to be on relatively short subscription terms. We estimate the subscription revenue impact of the increased retention and show that digital activation is linked to a 7%–12% increase in subscription revenue. Using clickstream data from the newspaper's website, we show that this increase is positively correlated with the degree to which activators took advantage of the content behind the paywall. This suggests that one of the mechanisms driving the effect of activation is that gaining access to otherwise restricted digital news content increases the value of activators' subscriptions, which may explain their higher propensity to maintain their subscriptions. We also find evidence that the relationships between digital activation and subscription behaviors are stronger for subscribers at risk for canceling their subscriptions. Overall, our results point to a crosschannel spillover from the online product (the website) to the offline product (the print newspaper).[2]

Our paper contributes to prior research that has argued that although paywalls may drive ad hoc visitors away from news sites (Chiou and Tucker 2013), they may also contribute new digital subscription revenues and increase print circulation (Pattabhiramaiah et al. 2019). We extend this literature, which has mostly relied on aggregate data, by using highly granular, individual-level subscriber data to show that existing subscribers who activate digital access are more likely to maintain their subscriptions and therefore, to continue generating subscription revenue for the newspaper. We also leverage our individual-level subscriber and clickstream data to explore the mechanisms (including increased subscription value because of access to digital content) driving the effect of activation. The granularity of our data also allows us to examine switching from print + digital to digital-only subscriptions after activation and to document the factors associated with the (minimal) amount of switching that occurs. Our paper also has implications beyond the newspaper industry. Firms often employ bundled product strategies in which subscribers to one product (e.g., a print newspaper) receive free or subsidized access to another complementary product (e.g., a newspaper website) (Gentzkow 2007, 2014). For example, Amazon provides Amazon Prime subscribers with in-store discounts at Whole Foods and free access to premium content on its Prime Video service, and Comcast offers TV subscribers free digital streaming. These bundles can help firms capture positive cross-product/crosschannel spillovers, which are especially important for firms operating in distressed industries such as print news and cable TV.

The rest of the paper is organized as follows. We describe how our work relates to existing literature in Section 2 and provide details of the research setting and our data in Section 3. We present our identification strategy in Section 4 and our results in Sections 5 and 6. We discuss the managerial implications of our findings in Section 7 and present concluding remarks, a discussion of the limitations of the paper, and possible future avenues of research in Section 8.

2. Relationship to the Literature

Our research is broadly related to three streams of research. The first is research on the monetization of digital content, including the effect of paywalls. The second is research on whether news products and channels complement or substitute for one another. The third is research on the value of multichannel customers.

2.1. The Monetization of Digital Content, Including the Effect of Paywalls

Providers of digital news and entertainment content have been aggressively erecting paywalls such as the one we study herein. According to a recent survey, 70% of newspaper sites in the United States and Europe had implemented paywalls by 2019 (Simon and Graves 2019b). The premise is that paywalls can help newspapers increase subscription revenues by attracting new subscribers and retaining existing ones.

Although news publishers are generally united about the importance of paywalls for the industry's survival, they remain hesitant to enforce hard paywalls such as the *Wall Street Journal*'s, wherein the entire news site is cordoned off from nonsubscribers. The industry has been careful to avoid paywall monetization strategies that may be viewed by consumers as overly heavy handed. As of 2019, only about 3% of U.S. newspapers had implemented hard paywalls (Simon and Graves 2019a).

Instead, newspapers typically use paywalls in which access to content is limited on one of two key dimensions: quantity and quality. The "metered" design focuses on quantity by allowing readers free access to all content up to a preset number of monthly articles. Paying subscribers are not subject to the quota and have unlimited access. The "freemium" design focuses on quality. Some content is free to all users, whereas premium content (e.g., op-eds, local news, business news) is only available to subscribers.[3]

Although hybrid approaches exist, the underlying driver of both types of paywalls is the newspaper's interest in monetizing traffic to its website, which is an increasingly popular news resource (George 2008, Seamans and Zhu 2014). This highlights a key trade-off; although paywalls can generate revenue by requiring readers to subscribe, they can also limit the number of readers, which can reduce advertising revenue (Pauwels and Weiss 2008, Chiou and Tucker 2013, Pattabhiramaiah et al. 2019). For example, Oh et al. (2015) document that implementation of the *New York Times* paywall contributed to a drop in readership from online word-of-mouth referrals. In a similar vein, Appel et al. (2020) examine conditions in which content providers should choose a paid subscription-only plan versus providing free access, given the presence or absence of advertising. Some research provides guidance for how firms should use freemium approaches to manage this trade-off between monetizing content and reaching a larger audience. For example, Runge et al. (2016) explore the amount of free content to provide to optimize the resulting mix of product trial, conversions, and word-of-mouth benefits in the context of software applications. Lambrecht and Misra (2017) find that firms should adjust their freemium strategy based on demand; in particular, they should offer more free content when demand is high. Deng et al. (2020) show that providing a free version of a mobile app increases demand for the paid app because customers use the free app to sample before they buy.

We contribute to this stream by studying the effect of subscribers' activation of paywall access on retention and subscription revenue. We show that the paywall can help newspapers retain existing subscribers; after activating access to the paywall, existing subscribers are more likely to maintain their subscriptions. This is important not only because this keeps subscription revenue flowing to the newspaper but also, because retaining subscribers is important for generating advertising revenue, particularly for the print edition (Ingram 2010).

2.2. News Products: Complements or Substitutes

It is possible that subscribers who activate digital paywall access will substitute a digital-only subscription for their print subscription (which includes digital access). As such, our study relates to prior research on whether news products/channels complement or substitute for one another. For example, Xu et al. (2014) and Aral and Dhillon (2020) study use of a newspaper's website and its mobile app and find a complementary relationship between the two. Other research investigates whether news aggregation sites (such as Google and Facebook) substitute for or complement newspapers' websites. Dellarocas et al. (2013) find a substitutive relationship between the amount of article content provided by a news aggregator and the reader's propensity to visit the newspaper's site, whereas Athey et al. (2021) report a complementary relationship between Google News and newspaper websites in Spain. Similarly, Sismeiro and Mahmood (2018) find complementarity in that both direct and referred visits to newspaper websites dropped during a temporary Facebook outage. A related research stream investigates the interplay between content creators who can link to each other's content (Dellarocas et al. 2013) and how this practice affects total newspaper consumption (Roos et al. 2020). We contribute to this literature by examining whether subscribers switch from (i.e., substitute) print + digital subscriptions to digital-only subscriptions after activating digital access.

2.3. Value of Multichannel Customers

In our context, an existing print subscriber's activation of digital access makes her a multichannel consumer of the newspapers' content. Research has shown that multichannel customers are more valuable than are nonmultichannel customers (e.g., Neslin and Shankar 2009, Montaguti et al. 2016). This may be because multichannel customers consume more and thereby naturally use more channels, because multichannel customers are exposed to more marketing messages from the firm, or because multiple channels help the firm provide better service to customers. We should note that nonactivators might also be multichannel consumers of content, although they are limited users of the online channel because of the paywall. So, our analysis essentially compares "fully" multichannel customers to "partly" multichannel customers. We contribute to the literature on multichannel customer value by studying whether the act of digital activation (i.e., the act of becoming a full multichannel customer) affects subscriber value as measured by retention and subscription revenue.

3. Data and Institutional Background

Our data are from a top 30 North American newspaper. Prior to 2013, the newspaper generated subscription revenue from its print newspaper product. Although the newspaper had a website in 2013 (and prior), it did not charge readers to access it. This changed in the first half of 2013, when the newspaper launched a paywall and began selling digital-only subscriptions. Our data contain weekly records for subscribers for the 213-week period from February 24, 2013—which precedes the launch of the paywall—to March 25, 2017.

When launched, the paywall used (and still uses) a "freemium" model. Website visitors could consume an unlimited amount of "unmetered" content (e.g., lifestyle articles and comics). By contrast, "metered" content (e.g., local news and sports news) was restricted. Website visitors could consume only a limited amount of metered content before they were presented with a "stop page." Visitors who were existing subscribers were invited to unlock full access to the metered content by creating and using a login linked to their subscription. This "digital activation" was included with their subscription (i.e., there was no additional fee or change to their subscription price nor was there any direct monetary incentive/reward for activation). The newspaper similarly did not engage in targeted advertising aimed specifically at activators nor employ any sophisticated content filtering techniques that provided different news content to activators and nonactivators on the website.[4] Some of the subscribers in our data activate digital access, and some do not. The key goal of our paper is to study the effect of digital activation for subscribers. Specifically, we ask the following research question: what is the effect of activating digital access on whether a subscriber maintains her subscription (i.e., retention) and the subscription revenue that the subscriber generates for the newspaper?

The data are structured as a subscriber/week panel. Each subscriber is identified via a *subscriber ID*. The data contain the subscriber's *zip code*, along with the average household *income,* average *age,* and *PRIZM (Potential Rating Index for Zip Markets) code* for the zip code. Although we have limited demographic information for subscribers (at the zip code level only), we have rich data on subscriber behaviors. *Date first subscribed* is the date that a subscriber first subscribed to the newspaper. For each subscriber/week, the data contain the subscriber's *account status*, *delivery frequency*, *weekly price*, *EZPay status*, *subscription term*, *subscription expiration date*, and *subscription renewal date*. *Account status* includes the following classifications: current, vacation, grace, and former. "Current" reflects subscribers whose subscription is paid for that week; they receive the printed paper. "Vacation" reflects subscribers whose subscription is current but who are on vacation; these subscribers do not receive the printed paper that week (likely because they are not home to read it), although they are billed for it. "Grace" reflects subscribers whose subscription has lapsed but who still receive (and are billed for) the printed paper that week (i.e., they are in a "grace" period). "Former" reflects subscribers whose subscription has lapsed and who are not in the grace period in that week; they do not receive the printed paper and are not billed. *Delivery Frequency* indicates how frequently the subscriber received the printed newspaper and includes daily (seven days per week), weekend only (three days per week), and Sunday only (one day per week). *Weekly price* is the subscription amount paid by that subscriber/week, which is billed for "current," "vacation," and "grace" weeks but not for "former" weeks. *EZPay status* denotes whether a subscriber allows the newspaper to automatically collect the subscription fee from her credit card or bank account. *Subscription term* indicates whether the subscription covers 13, 26, 52 weeks, etc. *Subscription expiration date* indicates when the subscription term ends, and *subscription renewal date* indicates when the subscriber renewed the subscription for another term.

We determined whether a subscriber activated digital access and if so, the digital activation date, as follows. In addition to the subscriber files, the data contain clickstream data from the newspaper's website from February 13, 2013 to February 19, 2015. These data list each page view during the time period, including the URL (uniform resource locator), a descriptive category of the page (e.g., news, sports, obituary, comics), whether a page was metered/unmetered, date/time accessed, IP (internet protocol) address of the user who accessed the page, the user's browser cookie ID, etc. If a subscriber had activated digital access and was logged in, then her subscriber ID is recorded for each page view. For each subscriber ID in the clickstream data, we recorded the date of the earliest page view tagged with her subscriber ID, which we considered to be the subscriber's digital activation date. Importantly, we do not observe any subscriber IDs in the clickstream data for the earliest weeks in our sample (which corresponded to the prepaywall period for the newspaper). This suggests that the earliest page view that we observe for each subscriber ID represents that subscriber's digital activation date. Indeed, the earliest digital activation dates coincide with the launch of the paywall.[5]

As we discuss in more depth, a key part of our identification strategy is matching activators and nonactivators on their subscription behaviors prior to when the activator activated digital access. In order to have a sufficiently long preactivation period for matching, we dropped from the sample activators who activated before February 23, 2014. This ensured that we observed at least 52 weeks of preactivation subscription activity for each activator. This generated a sample of 19,911 activators and 199,642 nonactivators, each of whose weekly subscription records we observed from February 24, 2013 to March 25, 2017.

3.1. Descriptive Analyses and Model-Free Evidence

In this section, we provide some model-free summaries of our data to explore the possible effects of digital activation. We conduct more formal analysis in the next section.

Table 1 shows descriptive statistics for all subscribers, activators, and nonactivators. The *Account Status* row in Table 1 shows the average number of weeks that each subscriber was in each status. The model-free evidence suggests that activators were more likely to keep their subscriptions current (i.e., not cancel) compared with nonactivators. Activators were more likely to be current (by ~32 weeks), on vacation (by ~9 weeks), or in grace (by ~7 weeks) and less likely to be former (by ~45 weeks) than nonactivators. We also compared *Account Status* for digital activators before and after they activated. Activators were 10.8 percentage points less likely to be in current status after activation (84.6% current before versus 73.7% current after). By the same token, activators were 11.5 percentage points more likely to be in former status after activation (14.3% former before versus 2.8% former after). We considered whether this might reflect a general trend among all subscribers by analyzing the analogous change in account status for nonactivators.

Table 1. Descriptive Statistics

	All subscribers	Nonactivators	Activators
Price[a]	3.85 (2.61)	3.71 (2.60)	5.21 (2.28)
Account Status (# of Weeks)[b]			
Current	135.46 (63.7%)	132.67 (62.3%)	164.25 (77.4%)
Vacation	4.03 (1.9%)	3.474 (1.6%)	9.72 (4.6%)
Grace	10.18 (4.8%)	9.57 (4.5%)	16.24 (7.7%)
Former	63.06 (29.6%)	67.17 (31.6%)	21.87 (10.3%)
Total	212.73	212.88	212.08
Delivery Frequency (# of Weeks)[b]			
Daily	86.48 (57.8%)	80.5 (55.3%)	146.43 (77.0%)
Weekend	5.30 (3.5%)	4.79 (3.3%)	10.41 (5.5%)
Sunday only	57.85 (38.7%)	60.29 (41.4%)	33.31 (17.5%)
Total	149.63	145.58	190.15
n	219,553	199,642	19,911

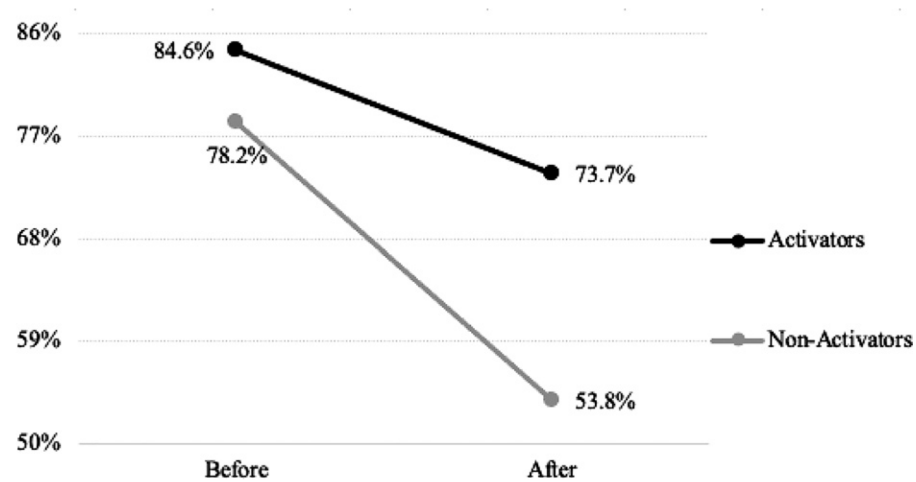
Note. Delivery Frequency statistics are calculated for *Account Status* = current, vacation, or former.
[a]Standard deviations are in parentheses.
[b]Percentages of total are in parentheses.

However, nonactivators do not have an activation date (by definition). In order to compare their subscription trends in the "before" and "after" periods with those of activators, we assigned each nonactivator a *simulated* activation date. We did this by taking a draw from the empirical distribution of activation dates of the activators (e.g., if a given activation date occurred 3% of the time in the distribution, then we assigned that as the simulated digital activation date approximately 3% of the time). This allows us to construct "before" and "after" periods for the nonactivators that are distributionally equivalent to the "before" and "after" periods for the activators. Using the simulated activation dates, we find that nonactivators were even less likely to be in current status after (simulated) activation (78.2% current before versus 53.8% current after) and even more likely to be in former status (14.8% former before versus 40.6% former after). As illustrated in Figure 1, there was an overall trend of increasing cancellation, but it was milder for digital activators. Thus, the model-free evidence suggests that activators were more likely to maintain their subscriptions. To examine the subscription revenue implications of this, we calculated average prices paid by subscribers. As shown in the *Price* row in Table 1, activators paid an average of \$5.21 per week, which is 40.4% more than nonactivators (μ = \$3.71). This is perhaps because activators were more likely to be daily subscribers than were nonactivators (77.0% versus 55.3%). We also compared average weekly prices paid by digital activators before and after they activated. Digital activators paid virtually the same average weekly prices before and after activation (\$5.22 before to \$5.21 after). By contrast, nonactivators paid 19.1% lower average weekly prices after (simulated) activation (\$4.24 before to \$3.43 after). This is depicted in Figure 2.[6] A likely explanation for the steeper price decline for nonactivators is that they were less likely to maintain their subscriptions (and therefore, less likely to keep paying) than were activators.[7] Overall, the model-free evidence suggests that activators are less likely to cancel their subscriptions compared with nonactivators, which results in activators contributing more subscription revenue. We explore this more formally next.

4. Identification Strategy

Whether a subscriber activates digital access is a choice made by the subscriber; it is not randomly assigned. This creates a challenge for identifying the effect of digital activation on retention and subscription revenue. Specifically, activators and nonactivators may have underlying differences—beyond whether they activated digital access—that influence whether they retain their subscriptions and continue to provide subscription

Figure 1. Percentage of "Current" Subscriber Weeks: Activators vs. Nonactivators



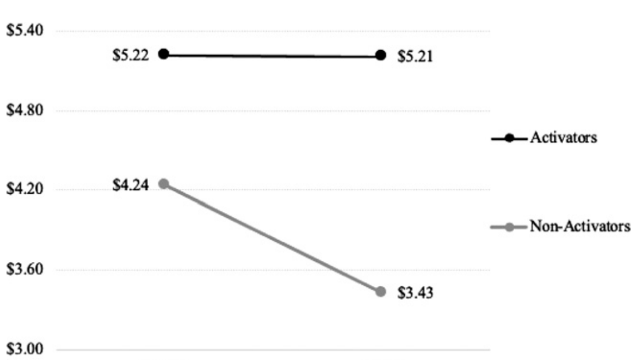
Notes. The "before" and "after" periods for nonactivators are defined based on simulated (i.e., counterfactual) digital activation dates. See the text for details.

revenue to the newspaper. We account for this potential self-selection bias in several ways, as discussed.

4.1. Matching

A key part of our identification strategy is that we match activators to nonactivators with similar subscription behaviors prior to when the activators activated. This limits the risk of self-selection bias by ensuring that activators and nonactivators in our analysis are highly comparable. We focus our analysis on activators who activated digital access no earlier than February 23, 2014; this allows us to observe their subscription behaviors for at least 52 weeks prior to digital activation. We match these activators to nonactivators using different matching methods, including coarsened exact matching (CEM), generalized random forests, and generalized synthetic control.

Figure 2. Average Weekly Price Paid by Subscribers: Activators vs. Nonactivators



Notes. The "before" and "after" periods are defined based on simulated (i.e., counterfactual) digital activation dates. See the text for details.

The 52-week threshold has several advantages. For example, it allows us to observe the subscription behaviors of activators for a long time before they activate digital access, which helps us identify precise matches for them. Because 52 weeks is the longest subscription term in our data, using a 52-week threshold ensures that we observe a full subscription cycle during the preactivation matching period. Note that for the vast majority of the activators in our sample, we observe *more* than 52 weeks in the preactivation period (e.g., for those who activate during the week of March 2, 2014, we observe 53 preactivation weeks; for those who activate during the week of November 30, 2014, we observe 92 preactivation weeks; etc.). Setting the preactivation period threshold to 52 weeks also has the advantage of moving our analysis window away from when the paywall was first introduced. This allows for any general effects of the paywall implementation—which might influence both activators and nonactivators in a way that could bias our results—to dissipate.

4.1.1. Coarsened Exact Matching. Because activators activated at different times, we identified the cohort of subscribers who activated in each week. (There are 53 activator cohorts—one per week—between February 23, 2014 and February 19, 2015, when our clickstream data stop.) We used coarsened exact matching (Iacus et al. 2012) to match the activators in each cohort to nonactivators based on subscription behaviors (*account status*, *delivery frequency*, and *weekly price*) prior to the activation week and other characteristics. We matched on *account status* by calculating the percentage of "current," "vacation," "grace," and "former" weeks (labeled $Pct\ Current_{Pre}$, $Pct\ Vacation_{Pre}$, etc.) as well as the pattern of those weeks (labeled

Pattern $Current_{Pre}$, *Pattern* $Vacation_{Pre}$, etc.) from the beginning of our study period to the activation week. For example, for the activators who activated during the week of April 20, 2014, we observe 60 weeks before the activation week. Assume that an activator in this cohort was in "grace" status for the first 10 weeks and "current" status for the next 50 weeks. As such, *Pct* $Current_{Pre}$ = 0.83 (50 "current" weeks divided by 60 total weeks) and *Pct* $Grace_{Pre}$ = 0.17 for this activator. We quantified the pattern of "current," "grace," etc. weeks by computing the average timing of when these weeks occurred in the preactivation period. For example, the activator in our example has "current" weeks in weeks 11–60 of the preactivation period, which we quantified as $11 \div 60, 12 \div 60, \ldots, 60 \div 60$. We took the average to yield *Pattern* $Current_{Pre}$ = 0.59 (*Pattern* $Grace_{Pre}$ = 0.09 in this example). We coarsened the values of *Pct* $Current_{Pre}$, *Pct* $Vacation_{Pre}$, *Pct* $Grace_{Pre}$, *Pct* $Former_{Pre}$, *Pattern* $Current_{Pre}$, *Pattern* $Vacation_{Pre}$, *Pattern* $Grace_{Pre}$, and *Pattern* $Former_{Pre}$ into bins and matched activators and nonactivators whose values were in the same bins. We used the same process to match on *delivery frequency* (i.e., daily, weekend, Sunday only), thereby allowing us to match on *Pct* $Daily_{Pre}$, *Pct* $Weekend_{Pre}$, *Pct* $Sunday_{Pre}$, *Pattern* $Daily_{Pre}$, *Pattern* $Weekend_{Pre}$, and *Pattern* $Sunday_{Pre}$. To further ensure that activators and nonactivators had similar subscription patterns in the preactivation period, we matched on the average *weekly price* paid during the preactivation period. We also matched activators and nonactivators on several other characteristics, including when they first subscribed to the newspaper (*date first subscribed*), their average *EZPay status* during the preactivation period, and their average *subscription term* during the preactivation period. Last, we used the *subscription renewal date* to determine whether a subscriber renewed her subscription during the activation week, and we matched activators with nonactivators based on whether they renewed their subscriptions during the activation week. This is important because if subscribers activated digital access at the same time as they renewed their subscription, then we might attribute a treatment effect to digital activation when it should be attributed to subscription renewal. Matching on subscription renewal addresses this potential issue. Figure 3 illustrates the basic design.

We used k2k matching (i.e., each activator is matched to one nonactivator). We pooled the matches from the 53 cohorts together, ensuring that each nonactivator is matched to only one activator, even though some nonactivators were suitable matches for activators in more than one cohort. This procedure yielded a matched set of 28,144 subscribers: 14,072 activators matched to 14,072 nonactivators. Because we have weekly data for each subscriber, this yielded a subscriber/week panel containing 5,993,306 observations. Table 2 shows the balance of the matched sample; there is no significant difference between the activators and nonactivators on any of the matching variables.

This procedure allows us to construct proxies for—and match on—difficult to observe variables that might affect subscriber's retention and subscription revenue, such as subscribers' appetite for news (e.g., whether they are a "daily" or "Sunday-only" subscriber), how conscientious subscribers are (e.g., whether they inform the newspaper when they go on vacation, whether they pay on time or let their subscription lapse into "grace" status), subscribers' planning horizons (e.g., whether they commit to 13- or 52-week subscription terms), subscribers' risk aversion (e.g., whether they put their newspaper subscription into vacation status because uncollected newspapers piling up could invite burglars to their home), etc. The procedure also yields fairly strict matches. For example, consider an activator with the following characteristics: (1) first subscribed to the newspaper in

Figure 3. Illustration of the Matching Approach

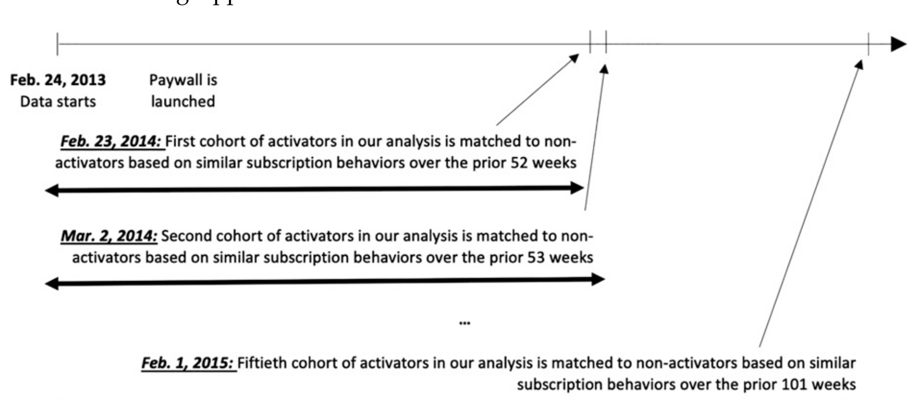


Table 2. Balance of Matching Variables: Coarsened Exact Matching

Variable	Activators Mean (standard deviation)	Nonactivators Mean (standard deviation)	Variable	Activators Mean (standard deviation)	Nonactivators Mean (standard deviation)
Number Current$_{Pre}$	0.884 (0.186)	0.871 (0.187)	*Pattern Current$_{Pre}$*	0.511 (0.090)	0.513 (0.091)
Number Former$_{Pre}$	0.024 (0.14)	0.023 (0.139)	*Pattern Former$_{Pre}$*	0.017 (0.098)	0.017 (0.098)
Number Grace$_{Pre}$	0.062 (0.124)	0.070 (0.127)	*Pattern Grace$_{Pre}$*	0.222 (0.306)	0.225 (0.297)
Number Vacation$_{Pre}$	0.031 (0.066)	0.036 (0.072)	*Pattern Vacation$_{Pre}$*	0.177 (0.29)	0.175 (0.284)
Number Daily$_{Pre}$	0.774 (0.416)	0.774 (0.415)	*Pattern Daily$_{Pre}$*	0.409 (0.219)	0.409 (0.219)
Number Weekend$_{Pre}$	0.034 (0.181)	0.034 (0.181)	*Pattern Weekend$_{Pre}$*	0.018 (0.096)	0.018 (0.096)
Number Sunday$_{Pre}$	0.167 (0.370)	0.167 (0.370)	*Pattern Sunday$_{Pre}$*	0.092 (0.204)	0.092 (0.203)
EZ Pay$_{Pre}$	0.367 (0.472)	0.369 (0.470)	*Dt first subscribed*	14,405.83 (3,704.299)	14,408.260 (3,711.698)
Subs Term$_{Pre}$	18.997 (13.270)	18.987 (13.256)	*Avg Price$_{Pre}$*	5.444 (2.014)	5.431 (2.016)

Note. There were no significant differences in means across groups even at the 1% level.

1999; (2) activated digital access on August 17, 2014; (3) was "current" with a daily subscription (paying $5/week) for 74 of the 78 weeks before digital activation, with 4 weeks in "vacation" status during the summers of 2013 and 2014; and (4) was on a 52-week subscription with EZPay activated. Our procedure matches this activator to a nonactivator with similar subscription behaviors and characteristics, including a similarly long tenure as a subscriber, a similarly moderate level of "vacation" taking over the prior 78 weeks (including similar timing of "vacation" taking), consumption of a similar product (52 week, full price, daily subscription), etc.[8]

4.1.2. Generalized Random Forests and Generalized Synthetic Control. As an alternative to coarsened exact matching, we used the generalized random forest method proposed by Athey et al. (2019) to nonparametrically match activators and nonactivators on observables, including the subscriber's account status, delivery frequency, and zip code.[9] We applied this method to each of the 53 cohorts (based on digital activation week) in our data to estimate the effect of digital activation. An advantage of the generalized random forest method is that it produces an estimate of the treatment effect for each activator. We used these estimates to examine treatment effect heterogeneity, which allows us to explore the mechanisms underlying the relationship between digital activation and retention and subscription revenue. For further robustness to the choice of matching estimator, we also used the generalized synthetic control method (Xu 2017), which extends the synthetic control method (Abadie and Gardeazabal 2003, Abadie et al. 2010) to cases where multiple units are treated at different times. This method allows us to combine multiple nonactivators into synthetic control units whose trends in *weekly price* in the preactivation period are similar to those for the activators. This accounts for unobserved factors that influence whether an activator selects into treatment, to the extent that these factors are captured by trends in *weekly price* in the preactivation period.

4.2. Accounting for Unobservables

Although we match on a rich set of variables over a long time period, it is possible that unobserved differences between activators and nonactivators might still bias our estimation. We address this issue in several ways. First, we include subscriber fixed effects in many of our models, which control for all unobserved, time-invariant characteristics of subscribers that might otherwise bias our results. Second, we conducted subsample analysis to investigate the possibility that our results could be confounded by activators' unobserved preferences for digital content. Third, we use a control function approach with an instrumental variable to account for the possibility that unobserved, time-varying characteristics of subscribers might bias our results. As discussed, we use the level of adoption of online banking in a subscriber's zip code to instrument for the subscriber's choice to activate digital access. Fourth, we use a timing falsification test to examine whether our results could be driven by unobserved, preexisting differences between activators and nonactivators. Last, we conducted sensitivity analysis (e.g., Rosenbaum bounds, unobserved selection on relevant covariates) (Oster 2019) to assess how large an influence any unobserved variables would need to have to overturn our conclusions.

5. Main Analysis and Results

5.1. Effect of Digital Activation on Subscriber Retention

We used several approaches—including descriptive statistics, proportional hazards modeling, and a linear difference-in-differences model—to explore the effect of digital activation on subscriber retention. We used the matched sample produced via the CEM procedure for this analysis so that the activating and nonactivating subscribers in this analysis were highly similar, with the key difference being that the activators activated. As discussed, we combined the 53 cohorts—each containing the activators for that cohort and the nonactivators to whom they are matched—to form the matched sample. Thus, each activator and nonactivator in the matched sample have an "activation week" defined by the cohort to which they belong.

5.1.1. Descriptive Analysis. We calculated the percentage of subscriber/weeks in "former" status both prior to and on/after subscribers' activation weeks. Prior to the activation week, the percentage of "former" subscriber/weeks was 1.21% for activators and 1.24% for nonactivators. These percentages are low because most activators were not in "former" status in the year or more preceding their digital activation; naturally, the nonactivators matched to them were not either. However, on and after the activation week, the percentage of "former" subscriber/weeks was 12.37% for activators and 19.72% for nonactivators. In other words, both groups became more likely to cancel their subscriptions over time (i.e., to go into "former" status), but this likelihood was much larger for nonactivators.

5.1.2. Proportional Hazards Model. We identified the first instance after the activation week in which each subscriber's status becomes "former" (if applicable). We then used a proportional hazards model to examine whether digital activation leads to a lower propensity to cancel (i.e., to go into "former" status). Our results are shown in Table 3 and indicate that this is indeed the case; the coefficient for digital activation is −0.377 (standard error = 0.02). This indicates that activation is associated with a nearly 31.4% (= [1 − exp(−0.377)]%) decrease in the likelihood that a subscriber will cancel.

5.1.3. Linear Probability Difference-in-Differences Model. We created a dummy variable (*Former*) to indicate subscriber/weeks in which *account status* was "former." We ran a difference-in-differences model, shown in Equation (1), to compare the average change in the linear probability of being in "former" status before/after activation for the activators with the analogous change for the nonactivators:

$$Former_{it} = \alpha + \gamma Digital\ Activation_{it} + \kappa_i + (f_t \times c_i) + \epsilon_{it}. \tag{1}$$

$Former_{it}$ denotes whether subscriber i was in "former" status in week t. $DigitalActivation_{it}$ is set to one for the first full week after a subscriber activates digital access and all weeks thereafter; it is set to zero otherwise. κ_i are subscriber fixed effects. f_t are week fixed effects, which we interacted with the 53 cohort indicators (c_i) to allow flexibility in capturing any underlying time trends. ϵ_{it} is the error term, clustered by subscriber to avoid possible contamination of standard errors from autocorrelation. We include fixed effects for subscriber and week to account for cross-sectional and temporal differences in our data. The coefficient of interest is γ. Results are shown in Table 3. We find a positive and significant effect of digital activation of −0.065, which represents a nearly 50% drop in the linear probability of being in "former" status.

We also ran a classic leads-lags difference-in-differences model (Autor 2003). This allowed us to assess whether activators and nonactivators had a similar propensity to be in "former" status during the preactivation period. If they did not, then our results might reflect a continuation of a preactivation difference rather than the effect of digital activation. This specification mirrors the specification shown in Equation (1), except that we replaced $\gamma DigitalActivation_{it}$ with $\sum_{\tau=-10}^{-2} \rho_\tau \times DigitalActivation_{it+\tau} + \sum_{\tau=0}^{10} \rho_\tau \times DigitalActivation_{it+\tau}$. Thus, we have

$$Former_{it} = \alpha + \sum_{\tau=-10}^{-2} \rho_\tau \times DigitalActivation_{it+\tau} + \sum_{\tau=0}^{10} \rho_\tau \times DigitalActivation_{it+\tau} + \kappa_i + (f_t \times c_i) + \epsilon_{it}. \tag{2}$$

Table 3. Effect of Digital Activation on Subscriber Retention and Switching Behavior

	Cancel subscription (transition to "former" subscription status)		Upgrade subscription to more frequent delivery	Downgrade subscription to less frequent delivery	Switch from a print + digital to a digital-only subscription (activators only)
	Hazard model	Linear probability model	Hazard model	Hazard model	Hazard model
Digital activation	−0.377*** (0.02)	−0.065*** (0.003)	−0.049 (0.086)	0.103* (0.056)	
Age (zip code level)					0.25 (0.23)
Income (zip code level)					0.09 (0.20)
Subscription term					−0.54*** (0.17)
Constant		0.129*** (0.001)			

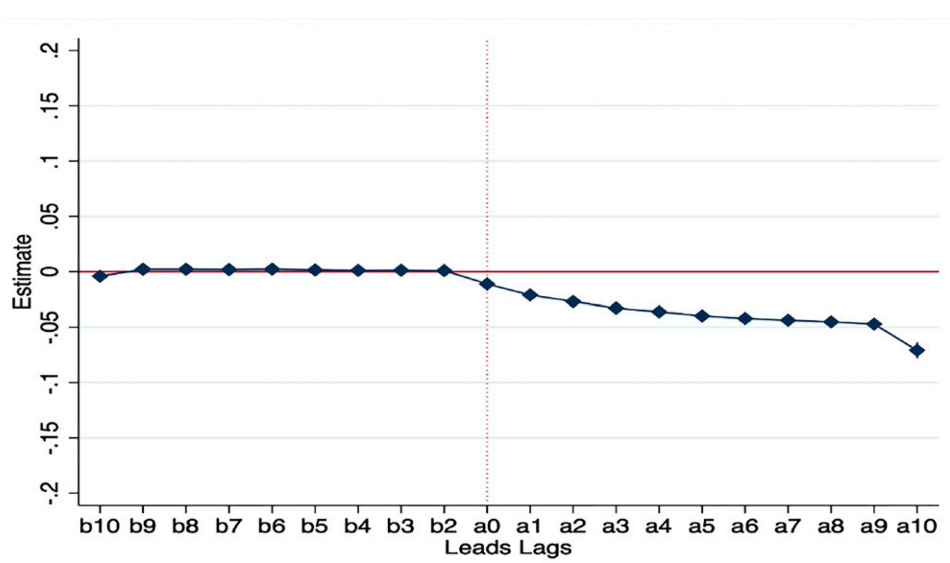
Note. Hazard models are stratified by cohort; results are similar without stratification.
*$p < 0.10$; ***$p < 0.01$.

$DigitalActivation_{it+\tau}$ are dummy variables equal to one for activators 10 or more weeks before activation ($\tau = -10$), 9 weeks before activation ($\tau = -9$), 1 week after activation ($\tau = 1$), 10 or more weeks after activation ($\tau = 10$), etc. We withheld the $\tau = -1$ term to avoid the dummy variable trap; as such, all of the ρ coefficients should be interpreted relative to the week before activation. The lead coefficients (ρ_{-10} to ρ_{-2}) allow us to assess whether the activators were more or less likely to be in "former" status than nonactivators before activation. If our approach is valid, then these terms will be close to zero. The lag coefficients (ρ_1 to ρ_{10}) allow us to assess how the treatment effect evolves over time. Figure 4 plots the lead and lag coefficients. The lead coefficients are quite close to zero, indicating that the matched activators and nonactivators were similarly likely to be in "former" status before the activation week. The lag coefficients show that the effect only becomes apparent after activation (as should be the case) and becomes more negative over time. This makes sense; if nonactivators are more likely to cancel than activators in any given week after activation, then the overall effect size should grow as the weeks pass.[10]

5.1.4. Subscriber Decisions to Switch Subscription Packages. We considered whether digital activation relates not only to the likelihood of subscribers maintaining their subscriptions but also to the likelihood of their switching to more or less frequent—and thereby, more or less expensive—delivery (e.g., from Sunday-only service to daily service or vice versa). We estimated separate proportional hazards models to examine whether activation relates to a subscriber (1) upgrading to more frequent delivery and (2) downgrading to less frequent delivery. We found that the effect of activation was nonsignificant for upgrading and marginally significant for downgrading, although the effect size is small. This is consistent with model-free evidence of minimal switching between delivery frequencies for both activators and nonactivators (see the appendix).

We also examined the possibility that activators were switching from print + digital subscriptions to digital-only subscriptions (i.e., that they were substituting a (more expensive) print subscription for a (cheaper) digital one). This is plausible because digital activation (and a digital-only subscription) allows subscribers to access all content available in the print newspaper. If a subscriber switches to a digital-only subscription after paywall activation, then that is recorded in the data as a change from daily/weekend/Sunday service to digital-only service. Notably, we only observe such switching activity for the activator group. That is because a digital-only subscription requires that the subscriber login to the paywalled website with an account linked to their *Subscriber ID*, which nonactivators do not do (by definition). Only 16 activators (<1%) in our sample made this switch, indicating that very little of this substitution occurs in our data.[11] We explored what factors were associated with an activator switching to digital only via a hazard model. We find that the hazard of switching to digital only is lower for activators who were on longer subscription contracts (see Table 3).

Figure 4. (Color online) Plot of Lead and Lag Coefficients for Estimating the Relationship Between Activation and the Linear Probability of Cancellation



Notes. We withheld the −1 term from the regression to avoid the dummy variable trap. b10–b2 (a0–a10) correspond to $\tau = -10$ through $\tau = -2$, respectively ($\tau = 0$ through $\tau = 10$, respectively) in Equation (2).

Table 4. Effect of Digital Activation on Subscription Revenues

	Coarsened exact matching[a]		Random forest		Generalized synthetic control[a]	
	Estimated	SE	Estimated	SE	Estimated	SE
Digital activation	0.431***	(0.023)	0.529**	(0.150)	0.676**	(0.236)
Increase in subscription revenues, %	7.6		9.5		12.1	

Note. SE, standard error.
[a]Fixed effects for subscribers and weeks are included.
**$p < 0.05$; ***$p < 0.01$.

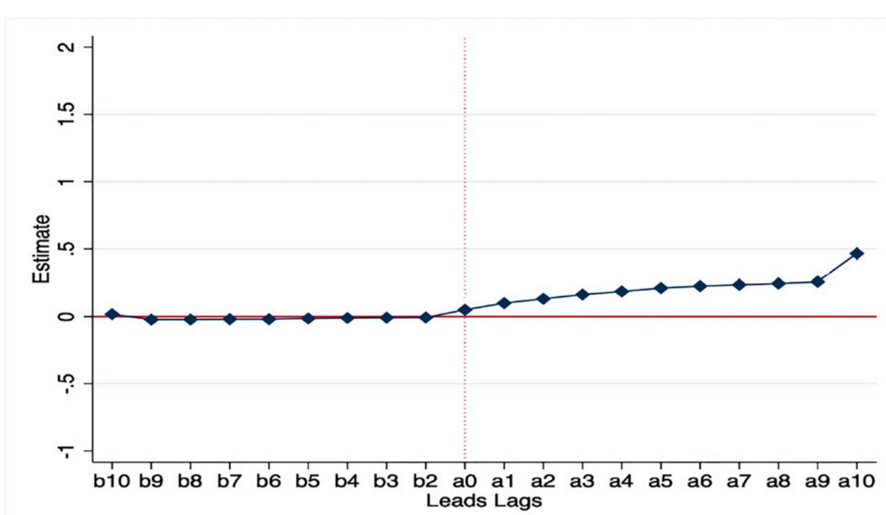
5.2. Effect of Digital Activation on Subscription Revenue

If digital activation improves subscriber retention, then it should also lead to increased subscription revenue. By the same token, if digital activation leads to increased subscription revenue, then that should be because it improves retention, given that we find no evidence that activation is associated with switching to a more expensive product (e.g., from Sunday only to daily service).[12] In this section, we estimate the subscription revenue effect.

5.2.1. Coarsened Exact Matching and Difference-in-Differences Estimation. We estimated the difference-in-differences model shown in Equation (1), except with *WeeklyPrice*$_{it}$ as the dependent variable. We used the matched sample from the CEM procedure. The treatment effect is \$0.43, which corresponds to a 7.69% boost in postactivation subscription revenues for the activator group (see column (1) of Table 4). We also ran a leads/lags model. The results show that there is little difference in subscription revenue between activators and nonactivators prior to activation (see Figure 5). This suggests that the positive revenue effects that we attribute to activation are, in fact, because of activation and not to unobserved differences between activators and nonactivators that precede activation. We also see a positive—and growing—effect of activation on subscription revenue. This is consistent with the retention analysis; over time, as more activators retain their subscriptions relative to nonactivators, activators will contribute more subscription revenue than will nonactivators. Another way to think about the increase in the effect over time is as follows. Note that we are analyzing newspaper subscriptions, such that nonactivators are only likely to cancel after their current subscription ends. Assume that in any given week, a few more nonactivators cancel their subscription compared with activators. Thus, we should see a small treatment effect in the weeks immediately after the activation week, with this effect compounding over time as more nonactivators cancel and add to the number of nonactivators who have already canceled.

5.2.2. Generalized Random Forests. The results of the generalized random forest estimation also yield a positive and significant treatment effect of digital activation. The dependent variable in this analysis is the

Figure 5. (Color online) Plot of Lead and Lag Coefficients Estimating the Relationship Between Activation and Subscription Revenue



Notes. We withheld the −1 term from the regression to avoid the dummy variable trap. b10–b2 (a0–a10) correspond to $\tau = -10$ through $\tau = -2$, respectively($\tau = 0$ through $\tau = 10$, respectively) in Equation (2).

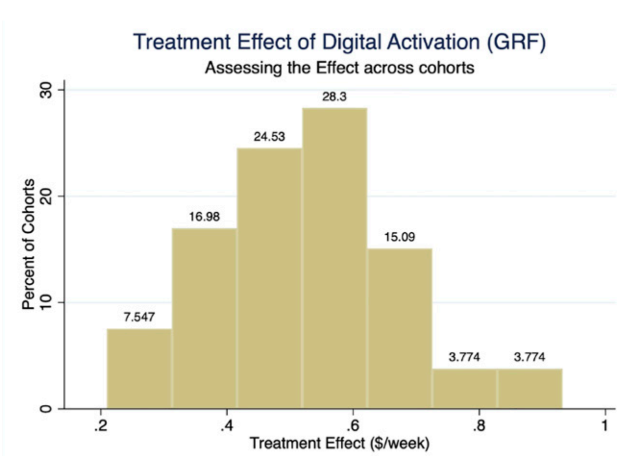
difference in the average weekly subscription revenue before and after the activation week ($AvgPriceChange_i$). We applied this method to each of the 53 cohorts (based on activation week) in our data to estimate the effect of digital activation. We followed the approach suggested by Athey et al. (2019) and grow 4,000 trees for our inference of the treatment effect of digital activation. Following the recommendations of Athey et al. (2019) and to enable sharper insights, we employed a regularized regression that nullified the influence on the treatment effect estimate of covariates that played a limited role in the treatment prediction choice.

We also use "honest" estimation wherein we choose half the sample each for model training and estimation. All model parameters are tuned based on cross-validation in the model training stage, before they are used for inference. We verified that the predictions from the causal forest were well calibrated using balance tests in the "grf" R package. Our results show a positive effect of activating digital access on subscription revenue of an average of $0.53 per subscriber/week (a 9.47% gain). This is shown in the second column in Table 4. A useful feature of the random forests method is that it estimates a treatment effect for each activator—we use these estimates to examine treatment effect heterogeneity in Section 5.4. The distribution of treatment effects is shown in Figure 6; all cohorts show a strong positive treatment effect.

5.3. Effect of Digital Activation on Subscriber Retention and Subscription Revenue: Accounting for Unobservables

Our inclusion of subscriber fixed effects in the difference-in-differences analysis controls for all unobserved characteristics of subscribers that might influence their decision to activate digital access, as long as they do not vary over the time span of our analysis. In this section, we describe our other steps to account for unobserved selection issues.

Figure 6. (Color online) Generalized Random Forest (GRF): Treatment Effect Distribution Across Cohorts



5.3.1. Subsample Analysis. A concern for our analysis is the possibility that activators have unobserved preferences for digital content that cause them to activate digital access. If these unobserved preferences also make subscribers less likely to cancel, then this alternative explanation could account for our results. We conducted subsample analysis to explore this possibility. First, we leveraged the clickstream data to construct a proxy for activators' preference for digital content. Recall that the *Subscriber ID* is recorded in the clickstream data for activators after they activate digital access. (In other words, after a subscriber uses her subscription to activate digital access, her *Subscriber ID* is recorded along with her browsing activity.) We identified all browser cookies associated with each activator's *Subscriber ID* and used those to examine activators' website activity prior to activation. We counted how many pages each activator accessed per week prior to activation ($WebPages_{it}$). We also counted how many times each activator hit the paywall "stop page" prior to activation ($PaywallHits_{it}$), which we identified by inspecting the page URL.[13] We used these measures to proxy for activators' preference for digital content. Second, we estimated the effect of digital activation on retention and subscription revenue for the subsamples of matched activators/nonactivators in which the activators (1) did not hit the paywall stop page prior to activation and (2) did not visit the website prior to activation. Using the proportional hazards model from Section 5.1.2, we find that digital activation is associated with 32% and 27% reduced hazards of cancellation, respectively, for these subsamples. Using the difference-in-differences model from the CEM analysis, we find that digital activation is associated with $0.426 and $0.362 increases in subscription revenue per subscriber/week, respectively, for these subsamples. We interpret this as follows. If the effect we identify is driven purely by preferences for digital content rather than by digital activation, then we should *not* see an effect for activators with zero page views or zero paywall hits prior to activation. Because that is *not* what we see, we conclude that this alternative explanation is unlikely to confound the estimated treatment effects. (Importantly, we cannot measure website activity for nonactivators because their *Subscriber IDs* are not tracked in the clickstream data; if they were tracked, then they would be activators. This is why we leverage differences in website activity within activators, rather than differences in website activity between activators and nonactivators.)

We also ruled out this alternative explanation by leveraging an interesting regularity in the data: specifically, that many activators are in "vacation" status

Table 5. Instrumental Variable Regression

Dependent Variable = Avg Price Change	Estimated	SE	Estimated	SE	Estimated	SE
Treated	0.455***	0.026	0.438***	0.029	0.456***	0.027
Residuals	0.045**	0.018	0.018	0.020	0.037*	0.020
Residuals^2			−0.010	0.011	−0.007	0.010
Cohort FE					Included	
First stage: Logistic regression of activation choice on the instrument						
	Estimated				SE	
Online banking penetration	5.53***				0.599	

Notes. Correlation (*residuals*, *residuals*^{2}) = −0.593. Bootstrapped standard errors (SEs) are reported. FE, fixed effect.
*$p < 0.10$; **$p < 0.05$; ***$p < 0.01$.

when they activate digital access. We believe that this is because these subscribers choose to activate digital access when they know that they will not be at home to receive the printed paper. As such, their activation decision is likely to be driven by the logistical issue of not being at home as opposed to an unobserved preference for digital content that might bias our estimation. We reran the hazard model for the subsample of matched activators/nonactivators in which the activators were in vacation status anytime from one week before to two weeks after the activation week. We find that digital activation reduced the hazard of cancellation by approximately 29% for this subsample. We also reran the difference-in-differences model from the CEM analysis for this subsample, finding an effect of \$0.389. This provides further evidence that the treatment effect that we identify is not confounded by an unobserved preference for digital content.

5.3.2. Instrumental Variables Analysis. We also used an instrumental variable regression combined with the control function approach to account for the possibility that unobserved, time-varying characteristics of subscribers might bias our results. The instrument is the penetration of online banking by zip code ($OnlineBanking_j$), gathered from the Mediamark Research Annual Survey of the American Consumer accessible via the SimplyAnalytics database. The intuition behind the instrument is that the level of online banking penetration in a zip code should be correlated with digital paywall activation in the same zip code because they both reflect the "digital savvy" of residents in that zip code. However, use of online banking should not have a direct effect on what subscribers pay for their newspaper subscriptions.

The control function approach is similar to two-stage least squares, except that the control function approach uses the residuals from the first stage, rather than the fitted values from the first stage, in the outcome equation to correct for possible endogeneity bias (Petrin and Train 2010). In the first stage of the control function approach, we regressed whether a subscriber activated digital access on the $OnlineBanking_j$ instrument in a logistic regression. As shown in Table 5, $OnlineBanking_j$ is a significant predictor of whether a subscriber activated digital access. For the second stage, we regressed $AvgPriceChange_i$ on $DigitalActivation_i$, the residuals from the first stage, and fixed effects for each activation week cohort. Results show an estimated treatment effect of \$0.455.

5.3.3. Timing Falsification Test. We conducted a timing falsification test to investigate further whether the treatment effect of digital activation might simply reflect unobserved differences between activators and nonactivators. The design of this test is illustrated in Figure 7. First, we reran our regressions from the coarsened exact matching procedure after assuming that activation occurred for the activators 26 (and 52) weeks before it actually did (i.e., we set a "fake" activation week) (Hosanagar et al. 2014). We also dropped all weeks following the actual activation week. This allowed us to see if a treatment effect showed up for activators before activation. Second,

Figure 7. Illustration of Falsification Test Using a "Fake" Activation Week

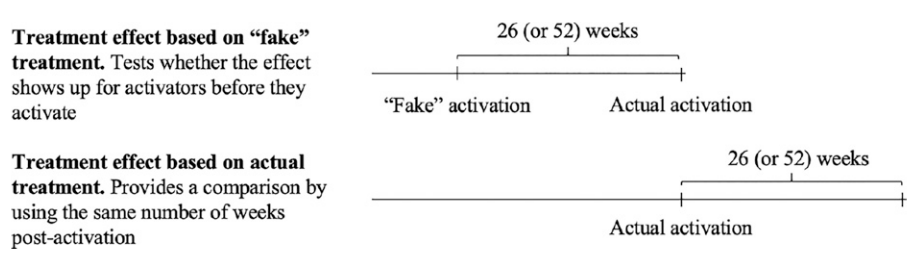


Table 6. Falsification Test Using a "Fake" Activation Week

	26-Week window		52-Week window	
	Actual activation	Fake activation	Actual activation	Fake activation
Digital activation	0.267 (0.015)***	0.001 (0.005)	0.314 (0.017)***	0.014 (0.007)**

$^{**}p < 0.05$; $^{***}p < 0.01$.

we reran our regressions—using the actual activation week—with the postactivation period limited to 26 (and 52) weeks. This ensures that we allowed the same amount of time for an effect to appear after treatment (either fake or actual) in both analyses.

Table 6 shows the results. The treatment effects using the actual activation week are \$0.27 using the 26-week window and \$0.31 using the 52-week window, which are smaller than our focal result. That is because we include more postactivation weeks in the focal analysis, thereby providing more time for the treatment effect to grow as nonactivators continue to cancel over time. The treatment effect using the "fake" activation week is \$0.00 using the 26-week window and \$0.01 using the 52-week window. This suggests that the treatment effect only shows up after activation (as it should), such that any unobserved differences between activators and nonactivators are unlikely to drive the effect that we document.

5.3.4. Examining Sensitivity to Unobserved Selection. We also measured how sensitive our results were to the possibility of unobserved confounders. We calculated the Rosenbaum bounds for our treatment effect of digital activation on subscription revenues (DiPrete and Gangl 2004). We used $AvgPriceChange_i$ as the dependent variable for this analysis and estimated the degree of unobserved selection that would be necessary to overturn the treatment effect. We find that unobservables would need to have between 1.7 and 1.8 times the influence of observables on selection into treatment to overturn the effect (see Table 7). This is above the generally accepted cutoff for proportional selection (Rosenbaum bounds parameter = 1) to justify concerns related to unobserved selection into treatment (Altonji et al. 2005).

To further assess the potential threat of unobserved confounders explaining our results, we follow the approach proposed by Oster (2019). Building on the logic of Altonji et al. (2005), Oster (2019) argues that the robustness of estimates to omitted variable bias can be examined by observing movements in (a) the coefficient of interest and (b) model R^2 from specifications that either include or exclude control variables in a regression. Under the rationale that including "relevant" control variables (those that plausibly contribute to improving model R^2; e.g., temporal fixed effects) would help alleviate omitted variables bias in a regression model (compared with the case when they are excluded), this approach enables researchers to comment on how large the influence of selection on unobservables would need to be, relative to selection on observables, to nullify the treatment effect. Following the recommendations of Oster (2019), we find that the degree of selection on unobservables would need to be 1.797 times that of observables in order to overturn our effect. This is above the generally accepted threshold of 1.0 (which corresponds to equal

Table 7. Rosenbaum Bounds and Oster (2019) Approach to Unobserved Selection

Rosenbaum bounds							Oster (2019) unobserved selection test		
Gamma	sig+	sig−	t-hat+	t-hat−	CI+	CI−		Estimate	SE
1	0.00	0.00	0.624	0.62	0.578	0.672	Treatment effect using the matched sample	0.355***	0.020
1.1	0.00	0.00	0.510	0.74	0.466	0.795			
1.2	0.00	0.00	0.412	0.86	0.371	0.911	Subscriber FE	Included	
1.3	0.00	0.00	0.327	0.97	0.288	1.021	Week FE	Included	
1.4	0.00	0.00	0.253	1.07	0.216	1.127	Model R^2	0.73	
1.5	0.00	0.00	0.187	1.17	0.151	1.227			
1.6	0.00	0.00	0.129	1.26	0.094	1.323			
1.7	0.00	0.00	0.075	1.35	0.041	1.414			
1.8	0.07	0.00	0.026	1.44	−0.008	1.501			
1.9	0.87	0.00	−0.020	1.52	−0.054	1.584			
2	1.00	0.00	−0.063	1.60	−0.098	1.663	Relative degree of selection parameter	1.797	

Notes. Gamma indicates the log odds of differential assignment because of unobserved factors. CI+, upper-bound confidence interval (a = 0.95); CI−, lower-bound confidence interval (a = 0.95); FE, fixed effect; SE, standard error; sig+, upper-bound significance level; sig−, lower-bound significance level; t-hat+, upper-bound Hodges–Lehmann point estimate; t-hat−, lower-bound Hodges–Lehmann point estimate.

$^{***}p < 0.01$.

Table 8. Exploring Heterogeneity in the Treatment Effect of Digital Activation

Dependent Variable = Estimated Treatment Effect for Subscriber	Estimated	SE	Estimated	SE
$Pct\ Current\ weeks_{Pre}$	−0.305***	0.029		
$Pct\ Vacation\ weeks_{Pre}$	−0.589***	0.040		
$Pct\ Grace\ weeks_{Pre}$	−0.216***	0.033		
$Pct\ Former\ weeks_{Pre}$	Baseline			
$Pct\ Daily\ weeks_{Pre}$			−0.157**	0.006
$Pct\ Weekend\ weeks_{Pre}$			−0.074**	0.01
$Pct\ Sunday\ weeks_{Pre}$			Baseline	
$\triangle Metered\ Pages_{(Post-Pre)}$	2.69E-05**	1.11E-05	1.76e-05*	1.09e-05
$\triangle Unmetered\ Pages_{(Post-Pre)}$	2.11E-05	1.33E-05	2.41e-05*	1.31e-05
Intercept	0.384***	0.034	0.250**	0.006
Zip code FE	√		√	

Note. FE, fixed effect; SE, standard error.
*$p < 0.10$; **$p < 0.05$; ***$p < 0.01$.

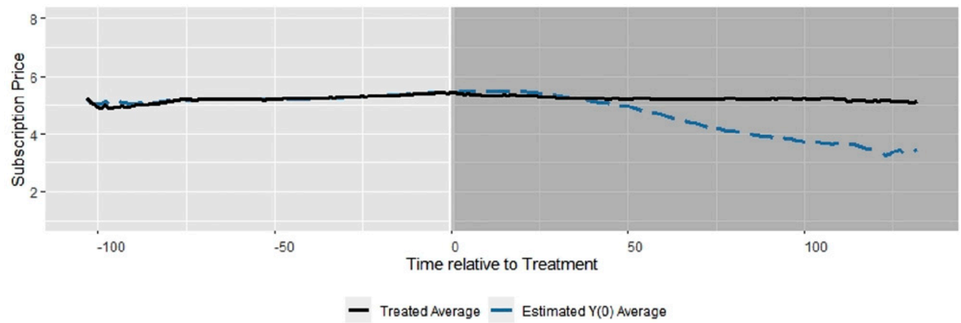
proportional selection on observables and unobservables).[14] This increases our confidence that the treatment effect is unlikely to be driven by selection on unobservables.

5.4. Examining Heterogeneity in the Effect of Digital Activation on Subscriber Revenue

In this section, we explore heterogeneity in the effect of activation on subscription revenue by leveraging the results of the generalized random forest model, which generates an estimated treatment effect for each activator. Understanding this heterogeneity is valuable for newspapers and helps shed light on the mechanisms for why activation leads to increased subscription revenue. We examined differences in activators' loyalty and appetite for news, as measured by their propensity to maintain a current subscription ($Pct\,Current_{Pre}$) for the daily newspaper ($Pct\,Daily_{Pre}$) in the preactivation period. We posit that loyal, high-consumption activators (i.e., those with high values of $Pct\,Current_{Pre}$ and $Pct\,Daily_{Pre}$) will benefit less than other activators. That is because of a ceiling effect; given that digital activation has a positive average treatment effect on subscriber retention, it should generate less benefit for subscribers who already tend to keep their subscriptions current. We also examined differences in activators' consumption of digital content (metered and unmetered) before and after activation. We posit that activators who took substantial advantage of paywall activation—as reflected by the increase in their consumption of metered content after activation—will benefit more than activators who took less advantage.

Using different specifications, we regress the treatment effect estimate for each activator on (a) the percentage of preactivation weeks that she spent in each account status ($Pct\,Current_{Pre}$, $Pct\ Vacation_{Pre}$, and $Pct\,Grace_{Pre}$; we use $Pct\,Former_{Pre}$ as the base case), (b) the percentage of preactivation weeks by delivery frequency ($Pct\,Daily_{Pre}$ and $Pct\,Weekend_{Pre}$; we use $Pct\,Sunday_{Pre}$ as the base case), and (c) the difference in the postactivation (relative to preactivation) average weekly number of metered and unmetered number of articles she consumed ($\Delta MeteredPages_{Post-Pre}$ and $\Delta UnmeteredPages_{Post-Pre}$), which we calculated based on the activators' cookies from the clickstream data. (As discussed, the *Subscriber ID* is recorded for the articles a subscriber consumed postactivation. We used each activator's cookie IDs to identify the articles she consumed preactivation.) We included zip code fixed effects as controls for unobserved local market-specific preferences for newspaper consumption.[15] The results across the different specifications are shown in Table 8. First, relative to the baseline of the percentage of preactivation weeks the activator spent in a former state, we find a negative relationship between $Pct\,Current_{Pre}$ and the size of the treatment effect. This is consistent with a ceiling effect; namely, a subscriber who spent most of her preactivation period maintaining a current subscription was likely already a loyal reader, such that activation would have a relatively small effect on her propensity to cancel. The negative coefficient for $Pct\,Daily_{Pre}$ also suggests a ceiling effect; viz., because daily subscribers already have daily access to the news, the ability to access the news throughout the week via the website may generate less benefit for them (compared with Sunday-only subscribers). One-standard deviation increases in $Pct\,Current_{Pre}$ and $Pct\,Daily_{Pre}$ are each associated with an approximately 50% decrease in the treatment effect. We also find that the activators who consumed more web content postactivation relative to preactivation had larger treatment effects, with the coefficient for $\Delta MeteredPages_{Post-Pre}$ consistently significant across specifications. A standard deviation increase in $\Delta MeteredPages_{Post-Pre}$ is associated with a 3.5%–5% increase in the treatment effect, depending on the specification. This suggests that the effects of activation are larger for activators who took greater advantage of the metered content behind the paywall. Overall, this

Figure 8. (Color online) Generalized Synthetic Control: Revenue over Time for Treated and Synthetic Control Units



analysis suggests that the mechanisms driving the effects of activation are that activation helps the newspaper retain subscribers who might otherwise cancel (with less of an effect on "loyal" subscribers), in part because activation increases the value of a subscription by providing unlimited access to otherwise restricted digital content.

6. Robustness Checks

6.1. Generalized Synthetic Control Analysis for Robustness to Choice of Approach for Matching and Estimation

We examined whether our results for how digital activation affects subscription revenue are robust to using the generalized synthetic control method, which uses a different method for identifying controls for the activators than does CEM or the generalized random forest approach. For this analysis, we used a model specification of the form

$$Weekly\ Price_{it} = \delta\, Digital\ Activation_{it} + \lambda_i' f_t + \varepsilon_{it}, . \quad (3)$$

The coefficient δ captures the average treatment effect of digital activation for the activators. In the generalized synthetic control method, the treated units (i.e., the activators, in our case) are linearly projected onto a multidimensional space spanned by the control units (the nonactivators). The synthetic control units are constructed based on the estimated factors and factor loadings (both of which are estimated from the data) from this projection, which are represented by $f_t = [f_{1t}, \ldots, f_{rt}]'$ and $\lambda_i = [\lambda_{i1}, \ldots, \lambda_{ir}]'$, where r represents the number of factors. The factor loadings are calculated to minimize the difference between the treated units and the control units on the outcome variable (in our case, $WeeklyPrice_{it}$) in the pretreatment period as well as on covariates. Our results (shown in the third column in Table 4) indicate an overall positive effect of digital activation of \$0.68, which represents an approximately 12.08% increase in weekly subscription revenue. Figure 8 shows that the average subscription revenues for activators and nonactivators are similar in the weeks before activation. This indicates that the method worked as designed (i.e., it generated synthetic control units that were comparable with the treated units before treatment).

6.2. Accounting for Local Market Dynamics

It is possible that time-varying and/or location-specific events such as local elections, sporting events (e.g., collegiate athletic victories), etc., could influence both subscribers' decisions to activate digital access and their retention and subscription revenue behaviors. To account for these local market dynamics, we included zip code × week fixed effects in the difference-in-differences model from the CEM analysis. The average treatment effect after including these fixed effects is \$0.412 (standard error = 0.023, $p < 0.001$). This is very close to the estimate without including these fixed effects, which suggests that our results are robust to whether we account for these local market dynamics.

6.3. Sample Inclusion Robustness Check

In our main analysis, we focused on activators for whom we observed at least 52 preactivation weeks. This has several advantages, including helping us construct precise matches, ensuring that we observe at least a full subscription cycle in the pre-period, and shifting our analysis window away from when the paywall was first implemented. However, it also means that we drop subscribers who activated digital access earlier in our analysis period. For robustness, we lowered the 52-week threshold to 10 weeks. This allowed us to include 25,718 more activators (and their matched nonactivators) in the analysis. For this expanded sample, the estimated effect of digital activation is a 48.3% reduction in the risk of cancellation and a 15.81% increase in subscription revenue. These effect size estimates are higher than our focal estimates. This makes sense, given that the longer postactivation window we observe for the newly included subscribers allows more time for the treatment effect to grow, as more of the nonactivators cancel their subscriptions relative to the activators.

7. Managerial Implications

Over the last decade, news publishers, large and small alike, have adopted bundled pricing strategies wherein they offer free, unlimited access to paywalled website content to their print subscribers. This reflects the industry's desire to monetize its digital content in a way that is not perceived by consumers as overly heavy handed. We show empirically that this strategy can help newspapers retain existing subscribers and their associated subscription revenue, provided that the subscribers activate access to the paywall. These benefits to the newspaper seem to be particularly large for subscribers at risk for canceling. We also show that the benefits are larger for activators who take greater advantage of access to the digital content behind the paywall. As such, we recommend that newspapers encourage subscribers not only to activate digital paywall access but also to consume the content behind the paywall. A promising strategy for newspapers to do this is to invest in exclusive and high-quality digital content; see the appendix for evidence supportive of this strategy.

Although newspaper industry experts identify effective digital reader engagement as one of the highest priorities for ensuring the long-term sustainability and survival of newspapers, they also note the challenges posed by the industry's entrenched reliance on legacy print revenues. Notwithstanding the steep advertising losses witnessed by newspapers over the last decade (Seamans and Zhu 2014, Sridhar and Sriram 2015, Pattabhiramaiah et al. 2018), print revenues still make up a lion's share of industry revenues. In fact, industry estimates suggest that, in revenue terms, a print subscriber is worth several times her digital counterpart (Edmonds 2012). Therefore, if digital paywall strategies not only contribute a heretofore untapped source of revenues (i.e., digital subscriptions and digital advertising) but also help retain print subscribers, this has significant implications for the long-term health of the industry. Indeed, our analysis shows that the vast majority of subscribers who activated paywall access retained their print + digital subscriptions rather than switching to digital-only subscriptions, at least in our study time period.

We expect that the retention benefits from digital activation that we document apply to the majority of local newspapers, which comprise about 98% of the 1,287 daily newspapers circulating in the United States (Abernathy 2018). If the paywall also facilitates the acquisition of new subscribers, then its overall effect for the newspaper may be much larger. Finally, encouraging print subscribers' digital activation may enable better cross-selling and upselling opportunities for newspapers. This is because they will be able to observe a wider pattern of crosschannel engagement for digital activators, given that digital activation prompts activators to associate their login/email addresses with their subscriber accounts. In this way, the provision of unlimited digital access to print subscribers could afford further analytic benefits to newspapers.

8. Discussion and Conclusion

Subscription revenue is increasingly important to newspapers as advertising revenue continues to decline. In an attempt to increase subscription revenue, newspapers and other content providers have commissioned paywalls to restrict access to their premium content. By 2019, nearly 70% of newspaper websites had some form of paywall (Simon and Graves 2019b). Paywalls can increase subscription revenue by helping newspapers not only acquire new subscribers but also retain existing ones. It is common for newspapers to offer free access to the paywalled website to their existing print subscribers. A goal of this bundling strategy is to retain subscribers by increasing the value of their subscriptions. We show that this strategy increases subscriber retention and therefore, subscription revenue, provided that subscribers activate their paywall access. This has not been documented previously (to our knowledge) and is particularly important in the newspaper industry, given that print editions have traditionally contributed a lion's share of the industry's revenues but are in a steady decline. Digital activation should also provide the newspaper with the means to link print subscribers' online and offline "path of movement," which should yield additional economic benefits.

We show the benefits of digital paywall activation by analyzing individual-level weekly subscription records from February 2013 to March 2017 for a major North American daily newspaper that ranks within the top 30 by circulation. The newspaper adopted a digital paywall within the initial months of our analysis window and offered its existing subscribers free unlimited access, as long as they activated this access by linking their login/email addresses with their subscriber accounts. We tested the effect of subscribers' activation of digital access on retention and the associated subscription revenue. Because we use observational data, we accounted for potential biases because of self-selection, noting that activators might be different from nonactivators in unobserved ways that affect their subscription behaviors. We leverage a variety of causal-inference techniques, including matching activators with nonactivators based on at least 52 weeks of preactivation subscription activity, to address this potential selection bias. The results suggest that digital activation decreases the risk of subscribers canceling their subscriptions by approximately 31% and contributes a 7%–12% lift in subscription revenue. We find evidence that the mechanisms driving these effects are that activation helps retain subscribers who might otherwise cancel, in part because activation increases the value of their subscriptions by allowing access to otherwise restricted digital content. This suggests a crosschannel spillover from the online product (the website) to the offline product (the print newspaper).

Our research has limitations. First, we are unable to distinguish in our clickstream data website visits from subscribers who did not activate digital access from visits by ad hoc readers. Thus, we cannot precisely compare website behaviors between subscribers who activate digital access and subscribers who do not. Second, given our observational data, we cannot completely eliminate the possibility of our estimates being biased, although our identification strategy involves strict matching and includes subscriber fixed effects, instrumental variables, subsample analyses, falsification tests, and sensitivity analysis. Third, we do not attempt to offer a holistic estimate of the paywall's overall revenue impact, including its possible influence on new subscribers or on advertising revenue. Our results may be viewed as a partial equilibrium effect of the overall influence of digital activation on revenue; expanding beyond this is an opportunity for future research. Another opportunity for future research is to investigate additional mechanisms driving the positive effects of digital activation. For example, if activation helps the newspaper better understand a subscriber's reading habits and interests by virtue of better insight into her multichannel behaviors, then the newspaper's marketing to that subscriber may be more effective, thereby generating retention and subscription revenue benefits. We leave exploration of this (and other) potential mechanisms for future research.

The results from our study are likely to be of interest not only to newspapers but also to firms in the publishing, media, entertainment, and other industries that provide subscribers to their core products with free access to complementary, low-marginal cost products. We show that this bundling strategy can improve subscriber retention, thereby increasing subscription revenue.

Acknowledgments

The authors thank participants at the 2018 Digital Marketing and Machine Learning Conference, the 2019 Marketing Science Conference, and the 2019 Workshop on Information Systems and Economics; seminar participants at the Indian School of Business, Northeastern University, and the University of Georgia; the *Management Science* department editor; associate editor; and three anonymous reviewers for their helpful feedback. The standard disclaimer applies.

Appendix

Additional Model-Free Analysis

The model-free analysis in Section 3.1 shows steeper price declines for nonactivators than for activators. We conclude that this is because activators are more likely to keep their subscription current (and thereby, to keep paying for them) compared with nonactivators. We also used model-free analysis to explore an alternative explanation: that activators were more likely to switch to a more expensive subscription plan (e.g., switching from Sunday only to daily service). (We also explored this in Section 5.1.4 of the text.) As in Section 3.1, we computed the average weekly prices paid by activators and nonactivators before and after activation (actual activation for activators and simulated activation for nonactivators) using only subscriber/weeks coded as "current." This sample restriction allowed us to focus on whether activation affects the prices that subscribers pay when they are subscribed, separate from whether activation affects their subscription cancellation inclination. Figure A.1 shows a similar increase in subscription prices for both activators and nonactivators, which suggests that activators are *not* switching to more expensive subscription plans compared with nonsubscribers.[16]

We extended this analysis by exploring whether activators (and nonactivators) switched to a more or less frequent (thereby, more or less expensive) subscription after activation (e.g., switching from Sunday to daily frequency). We computed the modal delivery frequency for each activator before and after digital activation. We did the same for each nonactivator using simulated activation dates. Table A.1 shows the "transition matrix" of the number of activators and nonactivators with modal frequency "A" before activation and modal frequency "B" after activation (where "A" can equal "B").

As can be seen from the table, there is very little cross-product switching activity. Of the activators, 93.0% had the same modal frequency before and after activation, 5.1% switched to less frequent delivery, and 1.8%

Figure A.1. Average Weekly Price Paid by Subscriber (Current Weeks Only): Activators vs. Nonactivators

Notes. The "before" and "after" periods for nonactivators are defined based on simulated (i.e., counterfactual) digital activation dates. See the text for details.

Table A.1. Counts of Delivery Frequency Transitions for Activators and Nonactivators After Activation (Actual or Simulated)

Modal frequency before digital activation	Modal frequency after digital activation			
	Daily	Weekend	Sunday only	Total
Activators				
Daily	14,047 (94%)	235 (2%)	677 (5%)	14,959
Weekend	94 (9%)	825 (83%)	79 (8%)	998
Sunday only	196 (6%)	56 (2%)	3,307 (92%)	3,289
Total	14,337	1,116	3,793	19,246
Nonactivators				
Daily	77,777 (95%)	734 (1%)	3,208 (4%)	81,719
Weekend	452 (10%)	3,937 (83%)	367 (8%)	4,756
Sunday only	1,440 (2%)	330 (0%)	67,914 (97%)	69,684
Total	79,669	5,001	71,489	156,159

Notes. Analysis is based on subscriber/weeks coded as "current," "vacation," or "grace." Numbers in parentheses are proportions by row (e.g., 94% of activators had a modal delivery frequency of daily before and after activation). Analysis does not include all activators and nonactivators. Excluded activators are those who had more than one modal frequency (e.g., had the same number of daily and weekend weeks) in either the before or after period. Excluded nonactivators are those (a) who did not have any current weeks in the (simulated) after period and (b) who had more than one modal frequency (e.g., had the same number of daily and weekend weeks) in either the before or after period.

switched to more frequent delivery. The corresponding percentages for nonactivators are 95.8%, 2.8%, and 1.4%, respectively. The higher likelihood for activators to switch to less frequent delivery compared with nonactivators might indicate that activation leads to lower prices overall. However, the average price decline for activators who switched to less frequent delivery is less steep than that for the corresponding nonactivators, which would have a countervailing effect. This is shown in Figure A.2, which shows a pattern similar to Figure A.1 but shows before/after average weekly prices (for "current" weeks only) for activators and nonactivators who (a) did not switch their modal frequency, (b) who switched to more frequent delivery, and (c) who switched to less frequent delivery. Overall, switching behavior after digital activation does not appear to be the mechanism driving increases in subscription revenues among the activator group.

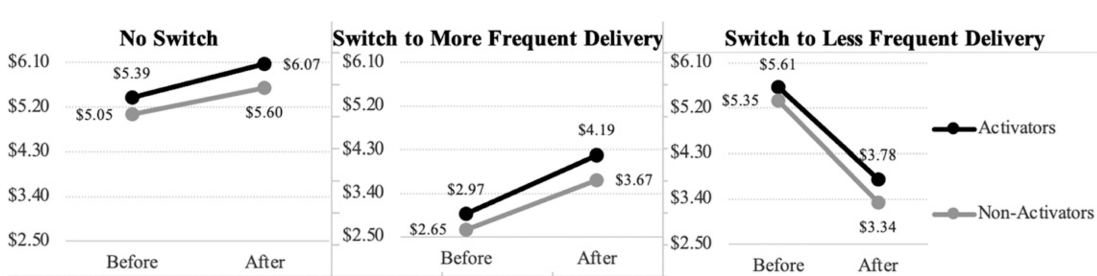
Exploring Changes in Digital News Consumption After Digital Activation

We used the clickstream data to explore how digital news consumption of activators evolved between the preactivation and postactivation periods. We conducted a difference-in-difference-in-differences analysis. The dependent variable in this analysis is the difference between the number of metered and unmetered page views per cookie/week. We considered cookies associated with activators to be treated and all other cookies to be controls. Importantly, we cannot tell whether control cookies belong to subscribers who did not activate paywall access (i.e., nonactivators) or to nonsubscribers. Thus, we lump them together for this analysis, which means that this analysis is structured differently than our other analyses.[17]

We created two dummy variables ("after paywall but before activation" and "after activation") to reflect three time periods: (1) before the paywall went into effect, (2) after the paywall went into effect but before the subscriber activated digital access, and (3) after the subscriber activated digital access. In order to compare the treated and control cookies over these time periods, we needed to assign a simulated paywall activation week to the control cookies (because the website visitors associated with these cookies did not actually activate). As we did for the model-free analysis discussed in Section 3.1, we simulated a digital activation week for each control cookie by taking a draw from the distribution of actual activation dates. We regressed the dependent variable on (1) the "after digital activation" indicator, (2) the "after digital activation" indicator interacted with a treated indicator, and (3) the "after paywall but before activation" indicator interacted with a treated indicator. We also included week fixed effects (to control for time trends) and cookie fixed effects (to control for heterogeneity). The cookie fixed effects and week fixed effects account for the main effects of treatment membership and the after paywall period, respectively. This allows us to assess whether activators increased their consumption of metered content after activating digital access.

We find evidence that supports this reasoning (results are shown in Table A.2). There is a drop in the relative consumption of metered versus unmetered content in the period following the paywall but before the treated user activated digital access. This pattern has face validity in that the paywall implementation may have limited these

Figure A.2. Average Weekly Price Paid by Activators and Nonactivators by Whether They Switched Delivery Frequency (Current Weeks Only)



Notes. The "before" and "after" periods for nonactivators are defined based on simulated (i.e., counterfactual) digital activation dates. See the text for details.

Table A.2. Consumption of Digital News Before and After Paywall Implementation and Digital Activation (Actual or Simulated)

Dependent Variable = *Diff in Pages Viewed*$_{(Metered-Unmetered)}$	Estimated (SE)
(*After Paywall*, *Before Activation*) × *Treated Group*	−0.259* (0.134)
After Activation	−0.027*** (0.006)
After Activation × *Treated Group*	0.597*** (0.031)
Intercept	1.247*** (0.006)
Cookie FE	√
Week FE	√
Observations	77,655,936
Adjusted R^2	0.52

Notes. Standard errors (SEs) clustered at the cookie level are in parentheses. FE, fixed effect.
$^{*}p < 0.10$; $^{***}p < 0.01$.

members' access to metered (premium) content during the time that they had not yet activated digital access. More importantly, on average, activators appear to consume about 2.4 more metered articles per month (i.e., 0.597 articles per week × four weeks) than unmetered articles after digital activation, whereas nonactivators consume fewer metered articles after their (simulated) activation. These results offer a somewhat optimistic outlook for the newspaper's content metering design. If activators increase their engagement with premium (metered) content—which was precisely the ex ante benefit that digital activation afforded these users—it bodes better for the newspaper than the converse. These results also support a case for the newspaper to improve the quality of its premium content, given that this can help them retain subscribers. Overall, these results suggest that after activating digital access, activators engage with premium/metered content on the website, which may have enhanced the value of their subscription. In turn, this may have contributed to their increased propensity to maintain their subscriptions.

Endnotes

[1] See https://www.economist.com/business/2017/10/26/how-leading-american-newspapers-got-people-to-pay-for-news.

[2] It is worth noting that digital activation may yield other benefits beyond those we document. For example, digital activation allows the newspaper to link print subscribers' online and offline behaviors, which may help the newspaper better understand subscribers and implement personalized promotions.

[3] This practice of restricting access to and monetizing premium content has seen increased popularity among entertainment providers as well (e.g., YouTube and Hulu).

[4] We are unable to provide additional details on the paywall implementation, such as the exact implementation date, in order to protect the identity of the newspaper (per the terms of our nondisclosure agreement).

[5] Some subscribers in our data may have activated digital access after February 19, 2015 (when our clickstream data stop). We discuss how this might affect our inference in footnote 10.

[6] One reason that the trend line for activators is above that for nonactivators is that activators are more likely to be daily subscribers, for whom the subscription price is higher.

[7] Please see the appendix for model-free analysis of alternative explanations such as that nonactivators were more likely to switch to a less expensive subscription plan (e.g., switching from daily service to Sunday-only service).

[8] The bins for *subscription term* were 0, (0,6.5), [6.5,13), [13,19.5), [19.5, 26), [26,39), [39, 52), and 52. The bins for *weekly price* were 0, (0,0.5), [0.5,1), [1,2), [2,4), [4,4.5), [4.5, 5), [5, 5.5), [5.5, 6), [6, 6.5), [6.5, 7), [7, 8), [8, 9), [9, 10), [10, 12), [12, 14), [14, 16), and 16. The bins for *date first subscribed* were before 1977, 1977–1985, 1986–1993, 1994–1998, 1999–2003, 2004–2007, 2008–2010, 2011, 2012, 2013, and after 2013. The bins for all other matching variables were 0, (0,0.25), [0.25,0.5), [0.5,0.75), [0.75,1), and 1.

[9] Matching on the zip code is effectively equivalent to matching on our demographic variables (average household income, average age, and PRIZM code), which are reported only at the zip code level.

[10] Our subscription data span 2013–2017, whereas our clickstream data stop in 2015. As a result, it is likely that some of the subscribers who we consider to be nonactivators activated digital access after 2015 (which we do not observe). However, if digital activation increased these subscribers' propensity to retain their subscriptions (which is also likely), then our results—which are based on differences in retention between activators and nonactivators—will be conservative.

[11] This meager rate of substitution is consistent with findings in prior work (e.g., Sridhar and Sriram 2015). Furthermore, Chyi and Ng (2020) report that, on average, the digital-only subscriber base of local newspapers is about 6% of the size of the print subscriber base, even in recent years. This implies that the perceived substitutability between the two product options is not particularly high.

[12] We also do not find any evidence that activators accepted subscription price increases during the study period that exceeded those of nonactivators (see Figure A.1 in the appendix).

[13] Only visits to the metered sections (business, local, and sports news) of the website counted toward the limit of free articles before the user was shown a paywall stop page. From our conversations with managers at the newspaper, the free article limit mostly ranged between 8 and 10 free metered articles per month, as opposed to a strict number.

[14] We use the STATA routine "psacalc" authored by Oster (2019), following a panel difference-in-differences regression (using the "areg" command) on our coarsened exact matched pooled sample. We follow her suggestions to set the maximum model *R*2 (R^2_{max}) to 1.3 times the *R*2 of a model employing the full set of available controls.

[15] The results without zip code fixed effects were essentially identical.

[16] This increase shows that the newspaper increased subscription prices over the analysis period, which is consistent with the trend in the industry overall. Because these price increases applied to both activators and nonactivators, they cannot explain the subscription revenue increase we see from digital activation.

[17] In our subsample analysis in Section 5.3.1, we noted that we cannot measure website activity for nonactivators. That is because we cannot distinguish the activity of nonactivators from that of nonsubscribers. That is why we lump them together in this analysis and use cookies per week as the unit of analysis.

References

Abadie A, Gardeazabal J (2003) The economic costs of conflict: A case study of the Basque Country. *Amer. Econom. Rev.* 93(1): 113–132.

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

Abernathy PM (2018) The loss of newspapers and readers. Accessed September 1, 2020, https://www.usnewsdeserts.com/reports/expanding-news-desert/loss-of-local-news/loss-newspapers-readers/.

Altonji JG, Elder TE, Taber CR (2005) Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *J. Political Econom.* 113(1):151–184. http://dx.doi.org/10.1086/426036

Appel B, Libai B, Muller E, Shachar R (2020) On the monetization of mobile apps. *Internat. J. Res. Marketing* 37(1):93–107.

Aral S, Dhillon PS (2020) Digital paywall design: Implications for content demand and subscriptions. *Management Sci.* 67(4):2381–2402.

Athey SC, Mobius MM, Jenö P (2021) The impact of aggregators on internet news consumption (April 2021). NBER Working Paper No. w28746, National Bureau of Economic Research, Cambridge, MA. Available at SSRN: https://ssrn.com/abstract=3838513.

Athey S, Tibshirani J, Wager S (2019) Generalized random forests. *Ann. Statist.* 47(2):1148–1178.

Autor DH (2003) Outsourcing at will: The contribution of unjust dismissal doctrine to the growth of employment outsourcing. *J. Labor Econom.* 21(1):1–42.

Chiou L, Tucker C (2013) Paywalls and the demand for news. *Inform. Econom. Policy* 25:61–69.

Chyi HI, Ng YMM (2020) Still unwilling to pay: An empirical analysis of 50 U.S. newspapers' digital subscription results. *Digital Journalism* 8(4):526–547.

Dellarocas C, Katona Z, Rand W (2013) Media, aggregators, and the link economy: Strategic hyperlink formation in content networks. *Management Sci.* 59(10):2360–2379.

Deng Y, Lambrecht A, Liu Y (2020) *Spillover Effects and Freemium Strategy in the Mobile App Market* (Social Science Research Network, Rochester, NY). http://dx.doi.org/10.2139/ssrn.3149550.

DiPrete TA, Gangl M (2004) Assessing bias in the estimation of causal effects: Rosenbaum bounds on matching estimators and instrumental variables estimation with imperfect instruments. *Sociol. Methodology* 34(1):271–310.

Edmonds R (2012) Newspapers get $1 in new digital ad revenue for every $25 in print ad revenue lost. Technical report, Poynter Institute. Accessed September 1, 2020, https://www.poynter.org/reporting-editing/2012/newspapers-print-ad-losses-are-larger-than-digital-ad-gains-by-a-ratio-of-25-to-1/.

Gentzkow M (2007) Valuing new goods in a model with complementarity: Online newspapers. *Amer. Econom. Rev.* 97(3):713–744.

Gentzkow M (2014) Trading dollars for dollars: The price of attention online and offline. *Amer. Econom. Rev.* 104(5):481–488.

George LM (2008) The internet and the market for daily newspapers. *B. E. J. Econom. Anal. Policy* 8(1):1–33.

Hosanagar K, Fleder D, Lee D, Buja A (2014) Will the global village fracture into tribes? Recommender systems and their effects on consumer fragmentation. *Management Sci.* 60(4):805–823.

Iacus SM, King G, Porro G (2012) Causal inference without balance checking: Coarsened exact matching. *Political Anal.* 20(01):1–24.

Ingram M (2010) Rupert's paywall is meant to keep people in, not out. *Gigaom* (July 19). Accessed September 1, 2020, https://gigaom.com/2010/07/19/ruperts-paywall-is-meant-to-keep-people-in-not-out/.

Lambrecht A, Misra K (2017) Fee or free: When should firms charge for online content? *Management Sci.* 63(4):1150–1165.

Montaguti E, Neslin SA, Valentini S (2016) Can marketing campaigns induce multichannel buying and more profitable customers? A field experiment. *Marketing Sci.* 35(2):201–217.

Neslin SA, Shankar V (2009) Key issues in multichannel customer management: Current knowledge and future directions. *J. Interactive Marketing* 23(1):70–81.

Oh H, Animesh A, Pinsonneault A (2015) Free vs. for-a-fee: The impact of a paywall on the pattern and effectiveness of word-of-mouth via social media. *Management Inform. Systems Quart.* 40(1):31–56.

Oster E (2019) Unobservable selection and coefficient stability: Theory and evidence. *J. Bus. Econom. Statist.* 37(2):187–204.

Pattabhiramaiah A, Sriram S, Manchanda P (2019) Paywalls: Monetizing online content. *J. Marketing* 83(2):19–36.

Pattabhiramaiah A, Sriram S, Shrihari S (2018) Rising prices under declining preferences: The case of the U.S. print newspaper industry. *Marketing Sci.* 37(1):97–122.

Pauwels K, Weiss A (2008) Moving from free to fee: How online firms market to change their business model successfully. *J. Marketing* 72(3):14–31.

Petrin A, Train K (2010) A control function approach to endogeneity in consumer choice models. *J. Marketing Res.* 47(1):3–13.

Roos JMT, Mela CF, Shachar R (2020) The effect of links and excerpts on internet news consumption. *J. Marketing Res.* 57(3):395–421.

Runge J, Wagner S, Claussen J, Klapper D (2016) Freemium pricing: Evidence from a large-scale field experiment. Working Paper, School of Business and Economics, Institute of Marketing, Humboldt University Berlin. Available at http://dx.doi.org/10.2139/ssrn.2888471.

Seamans R, Zhu F (2014) Responses to entry in multi-sided markets: The impact of Craigslist on local newspapers. *Management Sci.* 60(2):476–493.

Simon F, Graves L (2019a) Across seven countries, the average price for paywalled news is about $15.75/month. Technical report, Niemen Laboratory Report. Accessed September 1, 2020, https://www.niemanlab.org/2019/05/across-seven-countries-the-average-price-for-paywalled-news-is-about-15-75-month/.

Simon F, Graves L (2019b) FACTSHEET—Pay models for online news in the US and Europe: 2019 update. Technical report, Digital News Report. Accessed September 1, 2020, http://www.digitalnewsreport.org/publications/2019/pay-models-2019-update/.

Sismeiro C, Mahmood A (2018) Competitive vs. complementary effects in online social networks and news consumption: A natural experiment. *Management Sci.* 64(11):5014–5037.

Sridhar S, Sriram S (2015) Is online newspaper advertising cannibalizing print advertising? *Quant. Marketing Econom.* 13(4):283–318.

Xu Y (2017) Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Anal.* 25(1):57–76.

Xu J, Forman C, Kim JB, Van Ittersum K (2014) News media channels: Complements or substitutes? Evidence from mobile phone usage. *J. Marketing* 78(4):97–112.