Paywalls: Monetizing Online Content[*]

Adithya Pattabhiramaiah[†]

S. Sriram[‡]

Puneet Manchanda[§]

Oct, 2018

ABSTRACT

In recent years, many providers of news and entertainment have been exploring the possibility of monetizing online content. In the context of newspapers, the paywall instituted by the New York Times starting in March 2011 is a well-publicized case in point. While the premise behind paywalls is that the subscription revenue can potentially be a new source of income, the externalities that might arise as a consequence of this pricing change are unclear. We study two potential externalities of newspaper paywalls and compare them against the new direct subscription revenue generated. The first externality that we consider is the effect of a paywall on the *engagement* of its online reader base. The second externality is the *spillover effect* on the print version of the newspaper. If readers view print and online versions of a newspaper as substitutes, increasing the price of the latter is likely to increase the demand for the former. Moreover, many newspaper paywalls offer bundles wherein print subscribers are provided free access to the online newspaper. Therefore, the value that a reader derives from the print subscription could be higher subsequent to the erection of the paywall. As a result, paywalls are likely to have a positive spillover effect on print subscription, and consequently, circulation. We document the sizes of the two externalities for the New York Times paywall and compare them with the direct subscription revenue generated. We comment on implications for newspapers and online content providers who are seeking mechanisms to monetize digital content.

The movement of content (news, music, TV etc.) to digital media has had a large and (mostly) negative impact on the economics of multiple industries over the last 15 years. This impact is largely a function of the fact that while consumption of digital content is growing by leaps and bounds, the willingness to pay for this content is very low. As a result, firms in these industries have been exploring different options to monetize this content. For example, in the newspaper industry, firms have been scrambling to find ways to monetize their online platforms, mainly to offset the loss in circulation and advertising revenue from their legacy offline (print) properties (George, 2008; Sweney, 2015; Seamans and Zhu, 2014). Since access to the online version of newspapers has traditionally been provided free to readers, revenue from the online channel has come solely from advertising. Although online ad revenues have been growing steadily, this growth has not been sufficient to compensate for the offline ad revenue losses (industry estimates suggest for every advertising dollar gained online, newspapers lose 16 advertising dollars offline (Thompson, 2013). Thus, in recent years, newspapers have tried to tap into a new source of online revenue via imposing an access and/or a consumption fee, commonly implemented as a "paywall." As of 2014, nearly 75% of all newspapers had implemented paywalls (Marsh, 2014), including some high-profile national newspapers such as the New York Times, the LA Times and the Washington Post. However, the expectation that paywalls provide an additional source of revenue needs to be evaluated in the light of the fact that newspapers are complex businesses, with multiple interconnected parts. Specifically, newspapers are platforms bringing readers and advertisers together, with the two sides having a direct impact on each other. This effect is even more nuanced as online user type (loyal versus casual) may react differently to any changes (Tornoe, 2016, Mutter, 2015, Masnick, 2016). In addition, for (legacy) newspapers, the online and offline properties do not exist in isolation.

The heterogeneity in terms of user type and the interconnectedness of these various parts leads to multiple externalities that arise when a paywall is implemented, making it hard to pinpoint its overall impact. In this paper, we do this by explicitly considering two externalities. The first externality is the *engagement effect*, or the impact of the paywall on reader engagement (measured via activity and consumption patterns on its website) for both loyal and casual readers.[1] Note that

---

[1] http://www.experian.com.au/blogs/marketing-forward/2012/06/29/paywalls-drive-readers-away/. Note that our conceptualization here of the "engagement effect" corresponds better to the industry standard for measuring digital engagement, i.e., consumer engagement with the content (product-level engagement) vs. broader measures of firm-level engagement (e.g., number of user referrals or app downloads).

both sets of readers are important to the newspapers. Broadly speaking, loyal users bring in more subscription revenue while causal readers (typically a majority of traffic) bring in more advertising revenue, while also increasing the "footprint" of the newspaper. Any change in reader engagement is likely to have an impact on online advertising revenues. Second, and perhaps more important, we consider the effect of the online paywall on print (offline) readership and subscriptions. We call this the *spillover effect.* The spillover effect can arise because of two reasons: (a) substitution effect: readers view print and online versions of NYT as substitutes and hence an incease in the price of the online version can increase the demand for print, or (b) bundling effect: subscribing to the print version provided free access to the online content, thus the erection of the paywall increased the value of print subscription. While there is some discussion in the popular press regarding the first externality and its corresponding role in influencing newspaper advertising revenues, there is little documentation about its direction and magnitude. In addition, the second externality, or the spillover effect, has been virtually ignored. A useful baseline to examine the impact of these externalities is the incremental online subscription revenue from paywalls, which we call the *direct effect.* Thus, our approach provides a holistic assessment of the implementation of the paywall. This is in contrast to popular press reports (highlighted above) that while mentioning the direct effect, do not provide estimates of its direction and magnitude. More importantly, they typically ignore the spillover effect completely.

We investigate the two externalities for the paywall commissioned by the New York Times (henceforth, NYT) in March 2011. Ideally, there are two possible research designs that could help to identify the causal impact of the paywall. In the first such design, if there are a large number of newspapers serving independent markets, we could assign a random sample of these newspapers to the treatment condition by introducing a paywall and compare the quantities of interest for the treated newspapers with those experienced by the control newspapers. For the second ideal design, we could treat a random sample of readers of the NYT to the paywall and compare their subsequent response to those of the untreated readers. However, the implementation of either design is unrealistic. We therefore attempt to infer the effect of the paywall by leveraging non-experimental behavioral/market data by considering the change in online readership metrics and print circulation for the NYT subsequent to the commission of the paywall. In order to parse out temporal trends in online news consumption that are common to all newspapers, we compare these changes for the

New York Times against those experienced by national newspapers of similar popularity, i.e., USA Today, the Washington Post, the Wall Street Journal, Chicago Tribune and the New York Daily News (henceforth, USAT, WP, WSJ, CT and NYDN respectively). The availability of granular data on news consumption at the individual user level allows us to test the validity of the key identifying assumption behind such a design - that there be no impact of the treatment on the control units. We do this in two ways. First, we verify in the raw data that reader substitution between NYT and the control newspapers was minimal. In addition, we compare the results of an analysis based on data from a sub-sample of users that are *exclusive* users of NYT or the control newspapers, with those for our full sample.

A further potential concern with such an analysis is that the control newspapers may not be strictly comparable to NYT because they experienced different temporal trends prior to the erection of the paywall. Under such a scenario, it would be difficult to parse out the effect of the paywall from the naturally occuring differences in readership over time. In order to address this issue, we employ the synthetic control method (Abadie et al., 2010; Abadie and Gardeazabal, 2003). Intuitively, this approach creates a "synthetic control unit," computed as a weighted combination of all the control units. The weights are chosen such that the synthetic control closely matches the treated unit in terms of pre-period trends and other covariates. Thus, the synthetic control method naturally satisfies the parallel trend assumption required for inferring the causal effect of the paywall. As Athey and Imbens (2017) note, the synthetic control methodology is one of the most prominent advances in causal identification and policy evaluation over the last 15 years. It has also been receiving increased attention in the marketing literature (Tirunillai and Tellis, 2017; Guo et al., 2017; Li, 2017).

In order to infer the engagement effect of the paywall, we use web analytics data from comScore, which tracks the visitation behavior of panelists to NYT and the control newspapers. The availability of granular information on newspaper consumption at the individual level (as noted above) allows us to rule out the possibility that the effect of the NYT paywall on its online visitation (if observed) was driven by cross-newspaper substitution of readers. For the spillover effect, we obtain the weekday and weekend print circulation data for NYT, WP and USAT from the annual audit reports published by the Alliance of Audited Media. All our analyses are conducted at the designated market area or DMA level. We control for cross-sectional differences in tastes

for newspaper readership by including controls in the form of DMA-specific fixed effects for each newspaper. In addition, we include controls for temporal evolution in preference for consuming online and print news. Therefore, our identification relies on how the differences in the variables of interest between NYT and the control newspaper changed subsequent to the paywall, after controlling for cross-sectional differences between these newspapers as well as broader temporal trends in news consumption. Furthermore, the variant of the synthetic control method that we employ allows us to leverage the rich variation in readership available at the DMA level for each of our control newspapers, to improve the identification of the paywall's effect on the treated newspaper.

Our results reveal that the number of unique visitors decreased by 16.8% as a result of the paywall. This drop was also accompanied by a significant reduction in engagement metrics such as visits, pages consumed, and duration per visitor. In addition, while heavy users of the NYT website (defined based on their pre-paywall usage) reduced their visits and pages viewed significantly subsequent to the paywall, the corresponding effect on the behavior of light users was not pronounced. Thus, the adverse effect of the paywall was driven mainly by the behavior of heavier users. Further, we leverage the combination of web visitation and referring domain information available in our dataset to roughly classify users into subscribers/non-subscribers, and try to examine the paywall's impact on the activities of these groups. We find that while the adverse effect was indeed more pronounced among non-subcribers, this effect is dampened among heavy users who are likely to have subscribed. Together, these results suggest that the quantity of advertising impressions that could be served at the NYT website decreased as a result of the paywall.

With regard to the spillover effect, we find that the introduction of the paywall arrested the decline of print subscriptions for the NYT. As a result, compared to the counterfactual scenario of no paywall, the newspaper witnessed between a 1-4% lift in readership in both weekday and weekend subscriptions when we use national newspapers such as the USAT and the WP as the control group. Further, we attempt to parse out the substitution and bundling explanations for the positive spillover effect by exploring the differences in the spillover effect for subscriptions versus single copy sales. Since single copy sales did not grant free access to online content (while subscriptions did), any spillover effect should be a result of the substitution effect. On the other hand, the spillover effect for subscriptions should be due to a combination of substitution and bundling effects. Our results suggest that the positive spillover effect was more likely driven by the

bundling effect rather than the substitution effect.[2]

While related work (e.g., Kim and Song, 2017) has investigated the impact of paywalls on unique visitors, our research differs on multiple dimensions. First, in addition to the number of unique visitors, we investigate the impact of the paywall on industry-standard metrics of user engagement such as as pages consumed, visits per visitor, pages per visitor, and duration per visitor. Second, we exploit the disaggregate nature of our data to pinpoint the heterogeneity in the effect of the paywall on usage segments. Third, we isolate and quantify the spillover effect of the paywall on the legacy product (print subscriptions) and discuss the plausible mechanism driving this effect. The last aspect, i.e., the spillover effect is a hitherto unexplored consequence of paywalls.

There are two broad implications of our findings. First, the results suggest that monetization of online content, especially in the form of metered paywalls might suppress usage among loyal consumers. This can have implications for future growth potential of the firm. The loss of heavy users might hamper user generated content creation, which might be detrimental to platforms such as newspapers who may rely on such content in the future. Second, for media firms, a surprising, and usually overlooked, insight from our research is that the monetization of online content can have positive spillover effects for offline consumption. In situations where the offline channel is significantly more lucrative than its online counterpart (which is the case for newspapers and television), charging a fee for online content might arrest the erosion of offline revenues. In summary, our paper proposes a framework that will help managers in evaluating the various implications of monetizing digital content.

The rest of the paper is organized as follows. First, we investigate the effect of the paywall on online readership. Next, we consider the spillover effect of the paywall on print readership and discuss the possible mechanisms behind the observed effect. We conclude with some comments regarding the implications of these findings for the broad issue of monetizing online content.

---

[2]We thank an anonymous reviewer for pointing out this possibility and encouraging us to further explore the mechanism behind the spillover effect.

Engagement Effect: Paywall and Online Readership

The first externality of the paywall that we consider is the engagement effect.[3] This effect is likely to arise because erection of the paywall can adversely affect the number of visitors to a newspaper's website (Chiou and Tucker, 2013). For example, prominent national (San Francisco Chronicle, Dallas Morning News) and international newspapers (Sun, Toronto Star) have withdrawn their paywalls reportedly due to big losses in traffic.[4] Similarly, several local U.S. newspapers (e.g., Memphis Commercial Appeal, Columbia Tribune) also witnessed considerable decrease in traffic with the erection of the paywall (Blankenhorn, 2013). Moreover, as some paywalls, e.g., the New York Times', limit the number of articles that can be viewed for free, they can also reduce reader engagement by lowering both the average number of pages viewed by a visitor, as well as the duration of their visits.

Data

We use the web analytics data collected by comScore from Jan 2010 through May 2013 for our investigation. Given that NYT launched its paywall on March 28, 2011, our data span a reasonably wide window before and after the intervention. The web analytics data track the online activities of comScore panelists and include information on the websites visited by each panelist, date and time of the visit, number of pages viewed and the time spent on each website. In addition, we have information on the ZIP code where each panelist resides. To provide a benchmark for inferring the effect of the paywall on traffic to the news website, we extracted information regarding activities on six websites: NYTimes.com, WashingtonPost.com, USAToday.com, WSJ.com, ChicagoTribune.com, and NYDailyNews.com. This resulted in a sample of 75,174 representative individuals identified by comScore's sampling strategy.[5]

Our primary interest is in studying how news consumption at NYT changed after its paywall was erected. In order to parse out any changes in consumption that might have occurred as a result

---

[3]Note that the engagement effect captures the effect of the paywall on the digital consumer's (including both subscribers and non-subscribers) engagement with the online newspaper while the direct effect is based on the behavior of only subscribers.

[4]http://www.reuters.com/article/us-newscorp-sun-paywall-idUSKCN0SO25R20151030

[5]For details on the sampling strategy, see http://www.comscore.com/Media/Files/Misc/comScore-Unified-Digital-Measurement-Methodology-PDF.

of the general trend in news consumption, we use WP, USAT, WSJ, CT and NYDN as part of the control group. We chose these five newspapers as reasonable "controls" because similar to NYT, WP, USAT, WSJ, CT and NYDN are national newspapers. Moreover, the readership bases of the six newspapers are comparable (see Table 1) and represent the top set of U.S. news websites in terms of online traffic (2015 PEW Journalism Report).[6] In addition, WP, USAT, WSJ, CT and NYDN did not see any changes in the pricing of online content during the period of our analysis. While the WSJ always had a paywall, the other newspapers in our control group also commissioned paywalls in the time period after our analysis window. Table 2 lists the paywall launch dates for each newspaper in our analysis set.[7] Thus, we intend to use the readership trends in WP, USAT, WSJ, CT and NYDN after March 2011 to project the trend that NYT would have experienced had it not instituted the paywall. This, in turn, would help us understand the causal effect of the paywall instituted by NYT.

We begin by aggregating the web visitation data across panelists within a geographic market (DMA) to a monthly level. Our analysis includes the panelists residing in the top 25 DMAs (see Table 3) - these comprise over 70% of NYT's readership base.[8]

Based on industry practice and prior research (e.g., Lambrecht and Misra, 2017), we use the following four metrics of online news consumption to measure and capture engagement: number of unique visitors, number of visits per visitor, pages viewed per visitor, and the average time spent per visitor on the website.

Empirical Analysis

Recall that we propose to study the effect of the NYT paywall by using a basket of five newspapers - USAT, WP, WSJ, CT and NYDN - as controls. In order to address the concern that these three newspapers may have different temporal trends than NYT, we employ the synthetic control method. The synthetic control method permits the pooling of a combination of untreated units

---

[6]http://www.journalism.org/files/2015/04/FINAL-STATE-OF-THE-NEWS-MEDIA1.pdf

[7]WP announced a paywall in June 2013 though the paywall was not effectively enforced until Dec 2013 (Volokh, 2014).

[8]Abadie et al. (2010) highlight that the synthetic control method works well with balanced panels where there is no missing data. When we considered the full set of 202 DMAs, they included gaps in our dependent measures, which led to the synthetic control method not working. Therefore we restrict our analysis to the top 25-DMAs. When we estimated the effect of the paywall using a panel regression (differences-in-differences) model using the full set of 202 DMAs available in our dataset, the estimates were very close to the estimates for only the top-25 DMAs (see Table A.1 in Appendix A) - this was true for all of our dependent variables of interest.

to create a composite control that the treated unit can be compared against. The central idea behind the synthetic control method is that the outcomes of the control units can be weighted so as to construct the counterfactual treatment-free outcome for the treated unit. The weights are chosen such that the treated unit and synthetic control have similar outcomes and covariates over the pre-treatment period. Therefore, intuitively, the synthetic control method projects the treated units into a multi-dimensional space spanned by the control units in a way that they are matched on pre-treatment outcomes. Thus, the treated and control units are rendered "more comparable" by adjusting the loadings on each of the dimensions (which in this literature are also called factors).

More technically, a synthetic control for a single treated unit is formed by finding the vector of weights W^* that minimizes $[(X_1 - X_0W)\prime V(X_1 - X_0W)]$ subject to the weights in W being positive and summing to 1, where X_1 and X_0 contain the pre-treatment outcomes and covariates for the treated unit and control units respectively, and V captures the relative importance of these variables as predictors of the outcome of interest. Intuitively, the coefficient of interest (the parameter governing the treatment effect) is estimated by choosing W that forces the synthetic control to be as close to the treated unit as possible.

A primary benefit of a synthetic control estimator is that it reduces the reliance of the results on the parallel trends assumption that difference-in-difference/panel estimators are predicated on (Abadie et al., 2010; Xu, 2017; Tirunillai and Tellis, 2017). Thus, our identification of the effect of the paywall on NYT's online visitation does not rely on the control newspapers neccessarily following a similar trend. Our model specification can incorporate strict non-parametric controls in the form of newspaper market fixed effects to account for idiosyncratic differences in tastes for each newspaper in each market. In addition, they also include fixed effects for each month in the data so as to capture the influence of any common (across newspaper) time trends.

We study the effect of the paywall on multiple treatment units, with each DMA constituting a different treatment unit. In our analysis, we employ the generalized synthetic control estimator (Xu, 2017), which is a variant of the synthetic control estimator. A key advantage of employing a generalized synthetic control method is the ability to handle data with multiple treated units (DMAs in our case). The generalized synthetic control method leverages information on differences in control newspaper readership at different markets to construct a synthetic control unit for NYT's readership in each market, which effectively enhances the reliability of inference by increasing the

8

size of the control group from (number of control brands) to [(number of control brands) x (number of markets)]. (Xu, 2017). For ease of exposition, we omit subscript j which indexes DMAs, though, unless indicated otherwise, we estimate the model using DMA level data on newspaper readership. Thus, each newspaper n has a further $j = 1...J$ DMA data points.

We specify the model for online newspaper readership as:

$$ln(Q_{knt}) = \gamma_{kn}D_{knt} + x'_{nt}\alpha + \lambda'_n f_t + \varepsilon_{knt} \tag{1}$$

where k indexes the online metric (visits, pages, duration etc.) for newspaper n in month t. D_{knt} is an indicator variable that turns on for all months following the introduction of the paywall for only the treated newspaper. The term γ_{kn} is the coefficient of interest and captures the heterogeneous treatment effect of the paywall along metric k on newspaper n at time t. The term x'_{nt} is a vector of observed covariates and α is the corresponding vector of unknown parameters. The term $f_t = [f_{1t}, f_{2t}, ..., f_{rt}]'$ is a vector of r unobserved common factors, while $\lambda'_n = [\lambda_{n1}, ..., \lambda_{nr}]'$ denotes the corresponding factor loadings.[9] While the treated and control units are influenced by the same set of factors, and the number of factors is fixed throughout the analysis period (t = month 1 through month 48), each newspaper x DMA combination can have a different set of loadings on the r factors. Note that cross-sectional controls in the form of newspaper x DMA fixed effects and time (month) fixed effects can be considered as two special cases of the unobserved factors by setting $f_t = 1$ and $\lambda_n = 1$ respectively. In all our model specifications, we impose additive two-way fixed effects, a very strict non-parametric way of accounting for the possibly evolving nature of unobservables specific to treated and control units (Xu, 2017 p. 60). We discuss the steps involved in the generalized synthetic control estimation in the next sub-section. Since the dependent variable is specified in logarithms, we can compute the percentage change in the readership metric for NYT as a result of the paywall as $(exp(\gamma_{kn}) - 1)$.[10]

An additional advantage of the generalized synthetic control method is its ability to report readily interpretable uncertainty estimates around the treatment effect. Traditional inference in the synthetic control method is performed via placebo tests - which involves a procedure of "syn-

---

[9]The terms 'factors' and 'factor loadings' in the generalized synthetic control method are borrowed from the literature on interactive fixed effects models in economics (Bai, 2009). The time-varying coefficients are also referred to as (latent) factors while the unit-specific intercepts are labeled as factor loadings.

[10]We add a small constant term to get around instances of zeroes in our dependent variable as we take logs.

thetically" assigning treatment to control units, chosen one at time at random from the donor pool (i.e., the set of untreated newspapers), to compute a distribution of treatment effects. This enables us to assess whether the estimated treatment effect is larger than the collection of simulated treatment effects in placebo tests where no effect should exist. The generalized synthetic control method "automates" this procedure of running placebo tests and provides readily interpretable uncertainty estimates in the form of standard errors and confidence intervals around the estimated treatment effect - while preserving the efficiency of the estimation algorithm (Xu, 2017). In addition, the generalized synthetic control method has built-in safeguards to ensure that the results are robust in the presence of serial correlation. The estimator obtains uncertainty estimates around the treatment effect, using a parametric bootstrap procedure via resampling of the residuals, conditional on observed covariates and unobserved factors and factor loadings. This method allows the preservation of the serial correlation within the units, thus avoiding underestimating the standard errors due to serial correlation. The detailed algorithm describing the implementation of the parametric bootstrap procedure is available in Xu (2017), p. 65.

Next, we discuss a few possible threats to the validity of our estimation of the causal effect of the paywall on online news consumption. The first issue is whether treatment was anticipated i.e., whether the launch of the paywall was anticipated by consumers in a way that motivated either: (a) elevated levels of news consumption on the site right before the paywall went up, or (b) avoidance of the website in this period on account of the NYT's decision to commission a paywall. To test both of these possibilities, we performed simple checks by comparing trends in NYT visitation patterns over a narrow window immediately preceding paywall launch. A paired (across-DMAs) two-tailed t-test comparing NYT page consumption levels per visitor (visits per visitor), during March 2011 - the month before NYT's paywall commission, with its preceding month of Feb 2011, had a t-value of -0.99 (-0.84) with df=165, while the analogous t-values for the tests comparing these measures for Feb 2011 with Jan 2011 numbers were 0.72 (0.62). Thus, we find limited evidence of anticipatory effects in visitation/news consumption behavior at NYT in advance of the paywall.

A second issue to consider is whether the design of the NYT paywall was chosen strategically by the firm. Specifically, we are unable to separate out the effect of the paywall's introduction from the effect of changes to the newspaper's quality that might have been prompted by the paywall. As Doctor (2013) notes, there is little evidence that the quality of NYT's online offering changed

concomitantly with the paywall. This gives us confidence that the rigorous temporal controls included in the model are appropriate to account for changes in our dependent measures witnessed over time unrelated to paywall commission. Furthermore, it may be important to parse out the effects of any purely coincidental strategic actions by the NYT at the time of the paywall, i.e., they would have occurred even if the paywall were not introduced. If this were to be an issue, the observed effects of the paywall should be unlikely to hold for other similarly sized national newspaper(s) that also launched paywalls. To explore this, we use a similar research design to evaluate the effect of the paywall that the LA Times (LAT) launched in March 2012. To the extent that the estimated effects for the LAT paywall (commissioned in a completely different time window) are similar to those for NYT, we can gain some confidence that the results are both unlikely to be contaminated by coincidental strategic changes unrelated to the paywall, and can be generalized to similar newspapers. Nevertheless, similar to the majority of empirical research focused on inferring causal effects from non-experimental data (e.g., Goldfarb and Tucker, 2014; Chevalier and Mayzlin, 2006; Goldfarb and Tucker, 2011), we are unable to rule out completely the possibility that there may be factors unobservable to us that may have played a role in NYT's paywall launch decision.

Results

We present the results of the model (Equation 1) intended to explore the paywall effect on online readership in Table 4. These results suggest that the NYT paywall had a negative and statistically significant effect on engagement metrics, i.e., number of visits, pages visited, and the duration per visitor. In addition, the number of unique visitors decreased by 16.8% as a result of the paywall.

In terms of the differential effects on light vs. heavy consumers of online news, there are two alternative views. On the one hand, Mutter (2015) suggests that paywalls are likely to deter casual visitors and/or readers with low willingness to pay for online content.[11] On the other hand, a metered paywall such as the one erected by NYT is likely to impose a constraint only for heavy users. Therefore, the paywall is likely to have an adverse effect on the more engaged readers of NYT. Notwithstanding the ambiguity regarding whether the paywall is likely to have a greater

---

[11]Casual visitors may especially perceive the popup reminders intimating them of the available number of free articles (before encountering the paywall) as detrimental to their experience.

effect on visits among heavy or light users, the debate highlights the importance of considering the effects on these groups separately.

We investigate whether the paywall had a differential impact on light vs. heavy users by dividing panelists into two groups based on their pre-paywall usage. Specifically, we classify a panelist as a heavy user if their pre-paywall average number of pages accessed at NYT was higher than the median value of 4.1 pages. We first examine the impact on unique visitors - the results in Table 5 reveal that the paywall adversely impacted the number of unique visitors by 11.3% among light users and 57.2% among heavy users. We then turn to the engagement metrics - these suggest that the impact of the paywall was more negative for the heavy users.

We further verify the sensitivity of the results to alternative characterizations of heavy versus light usage. Rather than classifying panelists into heavy vs. light users based on a median split, we perform this classification based on their actual usage. Specifically, we classify a panelist as a heavy user if their average number of visits to NYT is greater than a certain number of pages. The paywall imposed a limit of 20 articles per month that could be accessed without payment. As our data contain information on the number of pages accessed by each panelist, and not the number of articles, we try different page thresholds under the assumption that a typical NYT article has about 1200 words, ranging from 1 - 2 pages (Pew Research Center, 2014; NYT Op-Ed Submission Directive, Hall, 2013).

In our empirical analysis, we use three thresholds: 20, 30 and 40 pages. Note that the definition of heavy usage becomes more stringent as we move from a threshold of 20 pages to 40 pages. Therefore, comparing the results across alternative thresholds will help us assess how the effect of the paywall changes with the degree of heavy usage. We present the results from this analysis in Table 6. Overall, we find consistent results (across these thresholds) that the paywall instituted by NYT had an adverse effect on engagement among heavy users.

There are two potential explanations for the stronger adverse effect of the paywall among heavy users.[12] First, the paywall can deter the ability of engaged users to share their content with others as the recipients may find it harder to read this content under the paywall (Maher, 2015). While we cannot formally test if this mechanism is indeed driving our results, previous research e.g., Oh

---

[12]A third explanation has to do with the "hassle cost" of having to repeatedly login to verify subscription status. However, modern web browsers allow for the saving of login credentials (typically via cookies) and so this is unlikely to play a big role.

et al. (2015) has documented a reduction in online word of mouth activity pertaining to popular newspaper articles in the periods following the paywall. Second, as noted earlier, heavy users are more likely to be constrained by the limit on the number of free articles imposed by the paywall, leading to the asymmetric reduction in engagement. However, note that the number of articles constraint should not apply to heavy users who subscribe to the NYT (either online or offline).

In order to understand whether the constraint imposed by the paywall on non-subscribers drove the decrease in engagement among heavy users, we need to study how subscribers and non-subscribers responded to the erection of the paywall. However, the comScore data do not contain information on whether a panelist was a subscriber to the NYT. Typically, only the publisher is privy to the proprietary information on subscription status. Therefore, we adopt an alternative approach by inferring subscribers based on post-paywall usage of users. As noted above, the NYT paywall allows a user to view 20 articles a month without subscription. Thus, we can identify a panelist as a subscriber based on whether they accessed more than 20 articles in a month. Using the same logic as above, we translate this article limit to page limits - 20 and 40. However, this count needs to be adjusted for traffic that came in via social media sites such as Facebook and/or search engines such as Google - the paywall's "leaky" design did not charge this against the 20 article limit.[13]

Fortunately, the comScore data contain information on the source (referring) website from which a user accessed NYT. This enables us to identify the number of accessed pages that would be counted towards the limit for each user. In other words, we drop all page views via referrals (from search engines, news aggregators and social media sites) and only consider the number of *directly* accessed pages (20 or 40). Once a user crosses this page limit in a given month, we classify her as a subscriber for all subsequent months.[14] It is important to note that classifying subscribers based on this strategy is likely to be noisy. However, by considering a wide range of thresholds to define subscribers, we are able to assess the robustness of our results to this noisiness in classification. We aggregated the data to the newspaper level for all months for this analysis, as we encountered

---

[13]A 2013 study found that nearly 66% of users reported social media as their primary source of news, with 47% of users surveyed identifying Facebook as their main source of news (Lichterman, 2016), highlighting the importance of accounting for the referring medium while analyzing page visits.

[14]The premise behind treating a user as a subscriber for all months subsequent to reaching the 20 article limit is that subscribers do not necessarily need to cross the 20 article limit every month. However, we acknowledge that this definition assumes that users do not terminate their subscription subsequently.

estimation challenges with the synthetic control method with DMA level data. Abadie et al. (2010) note that, to prevent its applicability where inappopriate, the synthetic control method employs a safeguard in that it fails to provide a result when the counterfactual units (i.e., the weighted combination of untreated units) fall outside an acceptable region (the convex hull) as governed by the treated units - the DMAs in our case. We thus aggregate up to the newspaper level for this analysis to get around this problem.

The above analysis allows us to examine whether the adverse effect of the paywall is restricted to non-subscribers. Specifically, we focus on the effect of the paywall for four groups of customers: (heavy vs. light users defined based on their pre-paywall usage) x (subscribers vs. non-subscribers defined based on their post-paywall activity, counting only *direct* visits to the NYT website). The results from this analysis are in Table 7. As can be seen from the table, the adverse effect of the paywall is not pronounced among subscribing users. Overall, we consistently find that among heavy users, non-subscribers reduce their activity on NYT more than subscribers - in line with the intuition that (a) the act of subscribing attenuates the drop in engagement for heavy users, or (b) users that anticipate using NYT more, tend to subscribe to the paywall.

Robustness Checks

Assessing the Inclusion and Exclusion of the WSJ from the donor pool

An underlying assumption behind the synthetic control method is that contributors to the donor pool (i.e., the untreated newspapers) should not have experienced treatment *during the analysis period.* However, as we discussed earlier, the WSJ had launched a paywall in 1996, well before the period of our analysis. Since the WSJ's paywall was in place well before the erection of NYT's paywall in 2011, it is unlikely that the WSJ's paywall operations may have interfered with how the NYT paywall influenced the engagement of its users. Nevertheless, to be conservative, we reestimated all our models using the generalized synthetic control method by omitting WSJ from the donor pool. We present the results where we examined the effect of the NYT paywall on light and heavy users, for a cutoff of 40 pages, in Table A.2 in Appendix A. These results suggest that the key findings remain unaltered when we excluded WSJ from the donor pool.

Assessing Substitution between the New York Times and Control Newspapers

A potential concern with our analysis is that the limits imposed by the NYT paywall might have

induced some of its readers to substitute to the control newspapers. If such substitution exists, we will be double counting the effect of the paywall in our analyses wherein we treat the control newspapers as being unaffected by the treatment.

We verify if substitution is bound to be problematic in our context in two ways. First, we examine whether there is model-free evidence of substitution by considering how users change their online reading habits of the control newspapers when they modify their online usage at the NYT subsequent to the paywall. If NYT and the control newspapers are substitutes, we should observe that a decrease (increase) in usage of the NYT should be associated with an increase (decrease) in usage of the control newspapers. In Table 8, we present a two-way frequency tabulation of the number of individuals in our sample that demonstrated an increase, decrease or no change (within 5%) in their consumption levels (as measured by the number of pages consumed) from the pre paywall period to the post paywall period across treated and control newspapers. The numbers indicate that the majority of users did not change their usage of the control newspapers even when they changed their consumption of news content at NYT subsequent to the paywall. This gives us confidence that substitution is unlikely to have affected our results.

Second, we employed the generalized synthetic control method on a restricted sample of exclusive users in our dataset who accessed _either_ the NYT or one of the control group newspapers in either period (i.e., users who used the NYT _and_ any one of the control group newspapers in either pre or post paywall periods, are excluded from the analysis). To construct our dataset for this analysis, we aggregated the individual level newspaper consumption data to the DMA-month level for each newspaper. Approximately 27% of users in our sample accessed both treated and control newspapers in either the pre/post paywall periods. We present summary statistics in Table 9, comparing the full sample with the exclusive sample - the two datasets are alike on our key measures of interest, in the pre-period. If the results are robust for this set of exclusive consumers of each newspaper for whom we can rule out substitution, we can infer that substitution is unlikely to have biased the results from the analysis based on the broader sample of users.

We present these results based on this sample of exclusive users in Appendix Table A.6 which suggest that the paywall adversely impacted engagement among heavy users of NYT. Therefore, we contend that our key results are not driven by users substituting between newspapers as a result of the paywall instituted by NYT.

Assessing the role of subscriber-acquisition focused promotional advertising by the newspapers

A potential concern with the analysis is that the estimated treatment effect of the paywall includes the effect of the structural change in price as well as any associated promotions that NYT might have initiated comcomitantly to recruit subscribers. To the extent that these additional promotional efforts (if they exist) are a result of NYT introducing the paywall, the estimated treatment effect may be interpreted as a consequence of the implementation of the paywall. Nevertheless, we examine the role of the newspaper's advertising by including acquisition focused promotion (with subscriber-acquisition focused ad expenditure as a proxy) as a covariate in our analysis. To this end, we collect time-series data on advertising expenditures focusing on subscription-drives by these newspapers at the national level, from Kantar Media's Ad$pender database. We test for the robustness of our results on users classified into light/heavy based on activity levels, to the inclusion of ad spending as a covariate. We find that the results are substantively unaffected - see Appendix Table A.9. In addition, we find that the effect of promotion (subscriber acquisition) focused advertising spending by the newspapers is not significant in all cases, after incorporating rigorous controls for time trends etc. This is intuitive because we already include rich non-parametric controls in the form of two-way fixed effects effects while estimating our generalized synthetic control models.[15]

Generalizability

Our results suggest that the paywall instituted by NYT adversely affected engagement among its heavy users. However, since our analysis is based on data from one newspaper, it is not clear if they can be generalized to other contexts. In order to explore whether similar results are likely to hold for other national newspapers, we consider the paywall instituted by LA Times in March 2012. To this end, we adopt a research design similar to the one discussed above by using data on online visitation to LATimes.com (henceforth LAT) among comScore panelists. We stratify users into light and heavy users, based on a median (4.4 pages) split, based on their pre-period activity levels on LAT's website, similar to our approach for the NYT. We use USAT, WP, WSJ, CT and NYDN as part of the donor pool. We present the results from this analysis in Appendix Table A.10. These results suggest that the key results that the paywall adversely impacted engagement among heavy users of NYT are replicated for the LA Times (see Appendix Table A.10). This provides

---

[15]These results also reinforce our confidence in the ability of the strict non-parametric controls in the form of month dummies in all other model specifications to serve as reasonable controls for any coincidental subscriber-acquisition related promotions by newspapers.

us some confidence that the key findings documented in this paper may not be unique to NYT. Furthermore, they also enhance our confidence that the observed effect of NYT's paywall on its visitation patterns is unlikely to have accrued on account of factors that merely coincided with the paywall rollout but were unrelated to the paywall launch decision (such as the newspaper's decision to change its font size on the website, or invest in its newsroom with an objective of improving the general quality of its news offerings).

The Spillover Effect: The Paywall and Print Readership

The spillover effect of the online paywall on print readership can arise via two possible mechanisms. First, if readers view print and online versions of a newspaper as substitutes, increasing the price of the latter is likely to increase demand for the former. Second, many newspaper paywalls, including the one instituted by the New York Times, offer print subscribers free access to the online newspaper. Such a bundled pricing strategy suggests that the value a reader derives from print subscription is likely to have increased subsequent to the erection of the paywall. As a result, paywalls can have a positive spillover effect on print subscription, and consequently circulation. In addition to the positive benefit from generating revenue from readers, the paywall may allow newspapers to boost their print ad revenues by projecting a higher circulation to is advertisers. This is especially important given that an average print reader brings in 16 to 228 times more in advertising revenue than an online reader (Thompson, 2013, Blodget, 2011). Thus the effect of a positive spillover on readership will be larger than the additional revenue generated from the online side.[16]

Data

We obtained data on print circulation from the Alliance for Audited Media's (henceforth AAM) annual Audit Reports for the years 2005 through 2013. As in the case of online visitation data, we collected this information for the NYT and three other newspapers with similar circulation - USAT,

---

[16]Given the proliferation of digital devices on which news content can be consumed, the paywall could induce switching behavior *within* online channels e.g., the website versus a mobile app, see Dhillon and Aral (2016).

WP and the WSJ.[17] AAM reports the circulation data at the annual level. Therefore, we have six years of data prior to the erection of the paywall (i.e. 2005 through 2010) and three years after the paywall (i.e., 2011 through 2013). We collected these data at the DMA level for 202 DMAs in the U.S. The circulation data are further broken down by weekdays vs. weekends. Next, we also collected these circulation data for the most popular local newspaper in the 25 largest DMAs. The idea is to verify the robustness of the results by treating local newspapers (as opposed to the national newspapers listed above) as the control group.

Model-free Evidence

We present the average circulation numbers before (2005 through 2010) and after (2011 through 2013) the paywall for NYT and the three other national newspapers in Table 10. Based on pre-period circulation as an evaluation metric, USAT was the most popular newspaper with 2.45 million subscribers, followed by the WSJ (1.92 million) and the NYT (1.69 million). Table 3 provides a list of the top 25 DMAs for NYT by circulation. These DMAs account for approximately 75% of print NYT's circulation, in the average year in our data. Across the 202 DMAs in our sample, the average DMA had about 11,091 (7,734) USAT subscribers compared with 5,146 (4,080) NYT subscribers in the pre (post) period.

Turning to the temporal pattern, paid circulation of U.S. print newspapers decreased consistently during our analysis period.[18] Our data from the four national newspapers: the NYT, WP, USAT, and the WSJ exhibit a similar pattern. In Table 10, we present the average annual (i.e., year-on-year) growth rates for these newspapers. These data suggest that on average, the three newspapers saw their circulation figures decline by 3.0% to 7.9% during the period of our analysis.

To examine the extent of possible substitution between print versions of NYT and control newspapers, we exploit the fact that we have circulation and subscription data for the print newspapers across many DMAs. Using these data, we perform analyses in the same spirit as in the case of online readership to understand if there is any substitution between print versions of these news-

---

[17]Due to the sparsity of print circulation data at the DMA level, we are unable to include a broader basket of newspapers for our analysis of the spillover effect. Note that it is more likely for national newspapers outside of the top 5 (such as even CT and NYDN) to have online visitors from a broader set of DMAs, than print subscribers living in regions farther out from their core circulation markets (Chicago and New York respectively, for the two example newspapers).

[18]http://www.marketingcharts.com/traditional/global-newspaper-circulation-and-advertising-trends-in-2013-43338/attachment/wan-ifra-newspaper-circs-ad-trends-in2013-june2014/

papers. The premise is that different DMAs varied in the extent to which the print readership of NYT changed over time. If there is indeed substitution between the various newspapers, markets that saw a steep decline in NYT print readership should also have experienced a steep increase in the readership of the control newspapers. Overall, change in NYT readership is not associated with a concomitant change in the readership of other newspapers (Table A.7 in Appendix A). This suggests that there is limited concern about substitution between newspapers as a result of the paywall, contaminating our characterization of the spillover effect.

Next, we consider the average annual growth rates for these newspapers prior to the NYT paywall and post paywall. We present these results in Table 10. These results highlight two aspects of the print circulation data. First, prior to the erection of the paywall, NYT saw its circulation decrease at rates similar to those experienced by other national newspapers. Second, we find that between 2011 and 2013, WP, USAT, and WSJ saw steeper declines in their circulation than during the 2005-2010 period. This pattern is consistent with the steep decline in print circulation experienced by the U.S. newspaper industry in the last decade (Edmonds, 2016). Contrary to this pattern, the results in Table 10 imply that NYT experienced lower declines during this period. Together, these data patterns are suggestive of a positive spillover effect of the paywall on print circulation. In Figure 1, we present a histogram of the average percentage year-on-year change in weekday print circulation for the NYT over our analysis duration to illustrate the cross-sectional variation in the data - the plot indicates that the majority of markets experienced a small percentage decline in circulation over time (the average percentage change across markets in NYT's weekday print circulation ranges between 2-3% as shown in Table 10).[19] Thus, the effect is unlikely to be driven by the presence of outliers. We now formalize the analysis by including controls for potentially differential rates of evolution of print circulation across the newspapers. Specifically, we include DMA specific linear and quadratic time trends to account to differential temporal evolution in readership across the DMAs in our data. Thus, we estimate the effect of the paywall intervention on print readership by exploiting the residual variation in shares after accounting for those motivated by changes in seasonal changes to print readership at the market level.

---

[19]While this plot indicates an overall year-on-year drop of 2%-4% in print circulation, a few specific DMAs (e.g., Bend, OR) seem to witness large percentage gains because of their very low circulation base.

Empirical Analysis

We employ a combination of approaches to estimate the effect of the NYT paywall on its print circulation. First, we use the generalized synthetic control method to estimate the effect of the NYT paywall, which is very similar in spirit to the models discussed so far for online visitation, with the exception that we use annual data in this case. However, results from the generalized synthetic control can be less reliable when the pre-treatment observation window is rather short (T<10 periods). In such cases, Xu (2017) suggests that results from the generalized synthetic control method should be validated against alternative estimation methods that are less dependent on the need for long observation windows. Therefore, as a robustness check, we specify a panel model using the same dataset, to examine the effect of the paywall on print readership R of newspaper n in market j in year t:

$$R_{njt} = \lambda_n + \lambda_j + \mu I_\tau + \delta I_{n=NYT} \times I_\tau + \vartheta p_{njt} + \wp_j\, t + \Upsilon_j\, t^2 + \varepsilon_{njt} \tag{2}$$

where I_τ: is a time-indicator signifying pre/post paywall launch and takes on the value of 1 post-paywall and 0 otherwise. We use print readership of USAT and the WSJ to establish a baseline/control for the effect of NYT's paywall (we do not use WP as DMA level circulation data for this newspaper was not available in the AAM database). We use two dependent variables (as part of the vector R) consisting of the weekday and weekend newspaper print circulation share (i.e., the % readership in each market, which is constructed by dividing the market level circulation by the number of households) for the analysis. The terms λ_j and λ_n capture reader preference for newspaper consumption in market j and for newspaper n, respectively (controlling for differences in taste for readership of different newspapers and in different markets), while μ captures any time specific effects of the post-period (common to both newspapers). δ is our coefficient of interest and measures the causal average effect of the paywall on NYT print newspaper readership. We also control for market specific time trends in newspaper readership using a parametric function [the $(\wp_j\, t + \Upsilon_j\, t^2)$ term in equation (2)].

The basic premise behind the DiD specification is that the temporal trends in print circulation for the control newspapers post-paywall will inform us about the corresponding trends for the NYT

had the paywall not been instituted. We therefore verify if the NYT and the control newspapers experienced similar temporal trends prior to the erection of the paywall. To do this, we regress pre-period Sunday readership for all newspapers on a newspaper specific linear year-time trend, after including DMA and newspaper fixed effects and clustering standard errors across DMAs to account for any serial correlation in readership. We do not see significant differences in the annual circulation trends for NYT vs. USAT ($F(1,200)=0.04$, $p=0.83$) and NYT vs. WSJ ($F(1,200)=1.05$, $p=0.31$).

Results

We present the results from the method of generalized synthetic control in Table 11 for the full sample of 202 DMAs, as well as only the top-25 DMAs, for comparison. Overall the results indicate that the effect of the paywall is positive and statistically significant, with the effect on circulation share ranging from 0.38 - 0.52 share points.[20]

We report the results from the panel regression with individual newspapers as controls (2) in Table 12. The results reveal a significantly positive coefficient on the (NYT x paywall) interaction term for both weekday and weekend circulation shares. This implies that the paywall had a positive effect on the offline readership of NYT, either in terms of slower rate of decline compared to the control newspapers or even growth. This result is consistent with the model-free evidence presented in the previous section. The estimates suggest that the NYT paywall had a positive effect on its print circulation to the extent of 0.18 - 0.68 share points, representing between a 1.05 - 3.98% lift in print subscriptions compared to the counterfactual scenario without the paywall. Thus, we see the impact of firm actions online on the behavior of its customers offline (similar results for a non-media market have been also been documented in van Nierop et al., 2011).

In order to assess the robustness of our estimates to the choice of control group, we consider an alternative analysis with the most popular *local* newspaper in each market as the baseline/control. Since we could not obtain credible circulation numbers for local newspapers in each of the 202 DMAs, we restrict our analysis to the top 25 DMAs. We present the results from this analysis

---

[20]The results presented here do not account for the role played by subscription prices in determining circulation share. Across all of our specifications, the size of our estimates of the spillover effect computed with subscription price as a covariate were identical (within 0.01-0.02 share points) with their counterparts with price omitted. We chose to omit price as a covariate because it was statistically insignificant in all cases and did not add materially to model fit.

in Table 13. Consistent with the results from the analysis with national newspapers used as controls, we find that the NYT paywall had a positive effect on the offline circulation of NYT. However, note that the magnitude of this effect is larger at 2.27 circulation share points when we use the local newspapers as controls (for reference, the effect of the paywall for the top 25 DMAs, considering USAT as the control newspaper, was between 0.62-0.55 share points for weekday and weekend circulation, when we estimated a panel regression model - see Table A.1 Appendix A). This larger effect can perhaps be rationalized by the steeper drop in print circulation witnessed by local newspapers in relation to national newspapers.[21]

In sum, we find that the print readership of NYT benefited from the paywall, potentially in the form of lower attrition relative to other similar newspapers. In other words, the NYT witnessed a positive and significant spillover effect of the online paywall on its print edition. There are two possible mechanisms governing this finding. The first mechanism is substitution, i.e., readers might have viewed the print and online versions of the NYT as substitutes. As a result, increasing the price of the online version by erecting a paywall might have had a positive effect on the demand for the print version of the newspaper. The second mechanism is bundling arising from the fact that NYT offered bundled versions of the newspaper wherein print subscribers received free access to the digital content. In fact, this bundle was priced very close to the digital-only subscription, thereby rendering it more attractive than digital-only subscription.[22] This bundling might also have increased the demand for print subscription after the paywall was erected, thereby also positively influencing the newspaper's ability to attract print advertising.

Interestingly, the bundling mechanism only works for print subcriptions sales but not for single copy sales. As a result, any spillover effect on single copy sales should be solely attributable to substitution between print and online versions of NYT, while the spillover effect on subscription sales should be a composite of both the substitution and bundling mechanisms. Therefore, by comparing the spillover effects for single copy sales versus subscriptions, we can comment on whether the substitution or bundling mechanism drove the spillover effect. To this end, we collected quarterly national level data on single copy sales and circulation of the print version for the newspapers in our sample from AAM's semi-annual publisher's statements. We use these data in a panel regression

---

[21]http://www.nytimes.com/2010/04/27/business/media/27audit.html

[22]Please see Table A.8 in Appendix A for details on the various subscriptions offered by the NYT around the time of hte paywall.

to investigate the prevalence of these two mechanisms.

The results from this analysis (Table 14) show an overall positive effect of the paywall amounting to of 29.3% of the newspaper's total circulation (i.e., a combination of subscription and single copy sales). However, when we consider subscription and single copy sales separately, we find an overall positive and significant effect of the paywall on subscription sales but an insignificant effect on single copy sales (p=0.459). This suggests that the primary driver of the positive spillover effect of the paywall on print readership was the bundling mechanism, although we are unable to conclusively demonstrate that substitution did not at all play a role.[23]

Generalizability

As in the case of our analysis of online visitation, these results are based on data from one newspaper. Therefore, we explore whether these results are generalizable to other newspapers of similar size. To this end, we compiled zipcode level print circulation data for the LAT, in the period surrounding its paywall commission (May 2012) and for a control newspaper, the WP as it did not operate a paywall during our analysis window. We were unable to perform this analysis using USAT as an alternative control group because USAT data are not available at the zipcode level. We chose to collect zipcode level data both because it was the lowest level of aggregation reported in the AAM Audit Reports for LAT, and also because the relatively smaller national coverage of the LAT restricted the DMA level data to less than 10 DMAs. We present the results of this panel regression analysis in Appendix Table A.11. We find a significant positive spillover effect (between 4.5-4.8 share points, higher in magnitude than the corresponding number for NYT) for the paywall erected by LAT. These results are similar in spirit to our finding for the NYT paywall. These results helps us place more confidence in our documentation of a positive spillover effect of the NYT paywall on its print circulation.

Overall, these results suggest a positive effect of the paywall on print newspaper circulation for NYT - a positive significant *spillover effect* of the paywall. Thus, our results are consistent with the view that the paywall may be serving a very important objective for this industry viz. stemming the decline in print readership. As discussed earlier, 65-80% of revenues for newspapers such as the NYT and LAT are obtained from the print edition of the newspaper. In addition, preserving a

---

[23]This conclusion is based on the assumption that although subscription and single copy sales address different segments of the newspaper's readership base, they responded similarly to the erection of the paywall in terms of their print readership.

print reader is believed to be at least 16 times as valuable, in revenue terms, than an online reader. Thus, the spillover effect of newspaper paywalls may play a large role in preserving a legacy source of revenue via slowing the decline in print readership.

Managerial Implications

Our paper proposes a framework that will help managers in evaluating the various implications of monetizing digital content. Specifically, we document that, in addition to considering the obvious direct effect of paywalls on subscription revenue, managers need to consider (a) how such a monetization approach would alter user engagement, and (b) the spillover effect of the paywall on the offline channel. Of these, the insight that managers need to consider the spillover effect of digital monetization on legacy media channels is a surprising and often overlooked implication. Overall, our empirical analyses highlight three key findings of relevance to managers:

1. The paywall instituted by NYT drove away some readers, as evidenced by a decline in the number of unique visitors to its website after the paywall.

2. In the period following the paywall, previously heavy readers of the NYT visit the website less often and also spend a shorter time on the website. Although these adverse effects are attenuated among readers who are likely to have subscribed to the newspaper, these findings imply that paywalls might pose a challenge to the greater objective of increasing engagement among online readers.

3. There is a positive significant effect of the paywall on the newspaper's print circulation, indicating that the spillover effect serves as a sizeable benefit.

In sum, there are two positive consequences of the paywall: (a) the incremental online subscription revenues generated by the paywall, and (b) arresting the decline in print circulation and the corresponding benefits from circulation and advertising revenue. On the other hand, decrease in engagement might have had an adverse effect on the newspaper's digital advertising revenues. In what follows, we discuss the overall implications of these findings for the overall financial performance of the NYT.

First, let us consider the direct effect of the paywall in the form of increased subscription revenues. In each quarter following the launch of its paywall in March 2011, NYT witnessed a

steady increase in the number of paid subscribers (2). NYT is reported to be successful in amassing a sizable base of over 500,000 digital subscribers in just eighteen months after the paywall was set up (Haughney, 2013). In addition to this direct effect, the paywall may also influence the newspaper's online advertising revenues indirectly in multiple ways. As discussed earlier, the paywall resulted in lower engagement in online content. Lower engagement and traffic leads to a lower quantity of ad impressions that can be served on the newspaper's website. Thus, relative to the period before the paywall, this will lead to lower advertising revenue. However, as a result of the paywall, the newspaper is likely to have richer information on subscribing visitors, increasing its ability to serve targeted ads. Moreover, subscribing visitors, by virtue of their revealed willingness to pay for digital content, are likely to be more attractive to advertisers. In the absence of the paywall, advertisers would not have been able to directly identify such high willingness to pay users. Therefore, the paywall can potentially help a newspaper charge higher ad rates per impression (typically measured in terms of cost per mille or CPM) as a result of the improved quality of the served ad impressions. There are early indications from the results of survey-based journalism research that advertisers are willing to pay higher advertising rates for their ads in paid online newspapers.[24] Therefore, the net effect of paywalls on online advertising is likely to depend on the relative magnitudes of the changes in the quantity and quality of ad impressions subsequent to the paywall.

If online advertising revenues did indeed decline as a result of the paywall, this resulting lower cash flow can hamper the newspaper's ability to invest in quality. Consequently, the decline in quality can lead to a further decline in readership, thereby driving the quality-driven circulation spiral (Gabszewicz et al., 2007).[25] While our data do not allow us to comment on the circulation spiral, media reports have lauded the the paywall as a net positive contributor to the NYT's revenues (Doctor, 2013). Therefore, any adverse changes in the quality of online content because of the paywall are unlikely.

Next, we consider the spillover effect of the paywall on the print newspaper. This spillover effect can have positive revenue implications both from the reader and advertiser sides. Nevertheless, recall that we discussed two plausible mechanisms behind the spillover effect: substitution effect wherein readers abandon the online version and switch to print as a result of the paywall, and the

---

[24]http://sabramedia.com/blog/newspapers-battle-between-paywall-and-advertising

[25]We thank an anonymous reviewer for pointing out the role of declining circulation/advertising in driving the circulation spiral.

bundling effect wherein readers who would have otherwise subscribed to the online paywall instead find print subscription with free online access more attractive. Note that even if it was not labeled as such, a similar "bundled option" of consuming print and online news was available to the reader prior to the paywall, although the online newspaper was free at that time. Thus, the benefits to the newspaper arising from such a bundled pricing plan should be attributed to the paywall, as the bundle would not have existed otherwise. Given our finding that bundling was probably the main driver of the spillover effect of the paywall, the positive benefit from the reader side is likely to have been somewhat limited. If the spillover effect of the paywall were to be driven by the substitution explanation, all the increase in print readership would be deemed as incremental, likely implying a larger revenue gain. Therefore, we can view our analysis as a conservative assessment of the magnitude of the spillover effect.

Under the scenario where the spillover effect is mostly driven by bundling, most of the positive benefit would be derived from increase in print advertising. Given that an average print reader generates $126 in print advertising, this increase can be sizable (see Appendix B for a rough calculation of the revenue gain from the spillover effect). However, there are two potential caveats to this positive outlook. First, if advertisers are actively switching between print and online versions of the newspaper, it is possible that some of the calculated increase in print advertising might be a result of advertiser substitution away from online advertising at the NYT. However, as Sridhar and Sriram (2015), Salmon (2009) and Hartung (2010) note, such cross-channel substitution is likely to be small.

Second, if bundling is the main driver of the spillover effect, it is might be argued that readers who subscribe to the print plus online bundle may, in reality, end up throwing away the print newspaper and only consume digital news. If this is true, advertisers might not view the corresponding subscriptions numbers as credible, thereby calling into question any corresponding gains in print advertising revenues. However, comparison of the options suggests that the price of even the cheapest print option was greater than that of a digital-only access plan. Therefore, it is unlikely that a NYT reader interested only in its online newspaper, subscribed to the print newspaper just to gain access to the digital version (i.e., with no intention of consuming the print version). Rather, bundling is likely to have helped in retaining some marginal print subscribers who were contemplating moving away from the print version. Furthermore, advertisers still continue to rely

on the readership numbers for print newspapers that are compiled and audited by the Alliance for Audited Media.[26] Therefore, in the short-term, newspapers are strictly better off by projecting a larger print readership base to advertisers even if there are questions regarding the extent to which their readers actually consume the print newspaper. Nevertheless, advertisers may view these circulation numbers differently in the long-term.

In conclusion, the above discussion highlights that the managerial implications span outcomes related to subscriptions, product design (e.g., bundling), pricing, and advertising revenues. More importantly, an often ignored consequence is that these implications encompass both the digital and offline channels.

Discussion

This paper advances a framework that can help managers in evaluating the various implications of monetizing digital content. The notion of monetizing online content is a problem that extends beyond the context of newspapers. Recently, television content providers and educational institutions have been grappling with the issue of designing appropriate monetizing strategies for their online content.

The first key insight from our work is that one needs to consider the spillover effect of online monetization on offline content consumption. This is especially critical if, as in our setting, there is a positive spillover of charging for online content on offline revenues. In our case, the analysis suggests that this positive spillover was due primarily to the bundling of online and offline content. We conjecture that a bundling strategy that provides free access to online content with the subscription to offline content might be reasonable when (a) the marginal cost of online content delivery is relatively low, (b) the offline channel is significantly more lucrative in terms of ad revenues, and (c) it is important to prevent channel partners for offline content (e.g., cable companies for television content) from feeling threatened by the online content.

Second, a digital monetization strategy might indirectly facilitate the creation of a broader, and more comprehensive "view" of audience engagement with both online and offline product offerings. This will happen when offline subscribers start linking their email ID with their online subscription

---

[26]http://www.ads-on-line.com/newbasiccourse/Products/; https://yourbusiness.azcentral.com/circulation-vs-advertising-revenue-14360.html

accounts, in order to authenticate their status. The ability to create such a comprehensive database may offer various long-term benefits for firms implementing various digital monetization strategies.

Lastly, the design of metered plans brings to surface the debate regarding whether heavy users should be "penalized" by the platform, as is typically done in freemium pricing plans. This might motivate heavy users to migrate from the platform or curtail their usage causing, thereby having deleterious consequences for overall engagement as well as sharing and propogation of content. Providing some value-added services exclusively to subscribers might be a viable strategy to circumvent this problem. We hope that our findings and this discussion engender future investigation in this area.

Conclusion

Newspaper paywalls are becoming an increasingly prevalent phenomenon, with nearly 75% of newspapers in the U.S. having either implemented or actively considering setting up a paywall. The popular belief is that paywalls may provide a much welcome new source of revenue: online subscriptions. However, as suggested by various surveys of newspaper readers, newspapers stand the risk of driving away readers who are not willing to pay for online news. As online ad revenues are heavily linked to newspaper readership, newspapers also stand to putting these revenues at risk if the paywall leads to heavy reader attrition. Thus, the overall impact of setting up newspaper paywalls is far from obvious. In this study, we employ data on online and print readership of NYT to assess the overall impact of the paywall it instituted in March 2011. We find that NYT's paywall appears to have driven away some readers, as evidenced by a decline in the number of unique visitors to its website after the paywall. In addition, our results suggest that following the paywall, previously active readers of the NYT visit the website less often and also spend a shorter time on the website, implying that paywalls may pose a challenge with the greater objective of generating increased engagement of online readers. We find a positive significant effect of the paywall on the newspaper's print circulation, indicating that the spillover effect serves as a sizeable benefit, in addition to the incremental online subscription revenues generated by the paywall. Overall, this research is the first of its kind to offer empirical evidence for positive overall economic returns accrued to information media firms from the decision to charge readers for access to online content.

However, our work does have a few limitations. First, without access to individual level data that includes payment (subscription) status online and offline as well as advertising revenues, we cannot delve deeper into the reasons for the increased revenue. Second, as discussed earlier, the aggregate nature of our data only permits us to offer logical conjectures regarding the mechanism governing the spillover effect, versus providing a precise quantification of the role played by all the plausible alternative mechanisms. Third, our analysis is unable to offer normative/prescriptive guidelines for setting up paywalls or managing their timing as our estimates are conditional on the firm's decision to charge readers for online news content. Fourth, while we are to show that our results are consistent across two large national newspapers, they may not extend directly to other media properties. We hope that future work can overcome these limitations.

References

Abadie, Alberto, Alexis Diamond and Jens Hainmueller (2010), 'Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program', *Journal of the American Statistical Association* **105**(490), 493–505.

Abadie, Alberto and Javier Gardeazabal (2003), 'The Economic Costs of Conflict: A Case Study of the Basque Country', *American Economic Review* **93**(1), 113–132.

Athey, Susan and Guido W. Imbens (2017), 'The State of Applied Econometrics: Causality and Policy Evaluation', *Journal of Economic Perspectives* **31**(2), 3–32.

Bai, Jushan (2009), 'Panel Data Models With Interactive Fixed Effects', *Econometrica* **77**(4), 1229–1279.

Blankenhorn, Dana (2013), 'The Newspaper Paywall Delusion'.
URL: *http://www.thestreet.com/story/11886411/1/the-newspaper-paywall-delusion.html*

Blodget, Henry (2011), 'A Print Reader Is Worth 228-Times As Much As An Online Reader - And Other Fun Facts About The New York Times'.
URL: *http://www.businessinsider.com/new-york-times-print-versus-online-2011-5*

Chevalier, Judith and Dina Mayzlin (2006), 'The Effect of Word of Mouth on Sales: Online Book Reviews', *Journal of Marketing Research* **XLIII**(August), 345–354.

Chiou, Leslie and Catherine Tucker (2013), 'Paywalls and the Demand for News', *Information Economics and Policy* **25**, 61–69.

Dhillon, Paramveer and Sinan Aral (2016), 'Digital Paywall Design: Effect on Subscription Rates & Cross-Channel Demand', *Working paper* .

Doctor, Ken (2013), 'The newsonomics of the New York Times' Paywalls 2.0'.
URL: *http://www.niemanlab.org/2013/11/the-newsonomics-of-the-new-york-times-paywalls-2-0/*

Edmonds, Rick (2016), 'Newspaper declines accelerate, latest Pew Research finds, other sectors healthier'.
URL: *http://www.poynter.org/2016/newspaper-declines-accelerate-latest-pew-research-finds-other-sectors-healthier/416657/*

Fan, Ying (2013), 'Ownership Consolidation and product Characteristics: A Study of the U.S Daily Newspaper Market', *American Economic Review* **103**(5), 1598–1628.

Gabszewicz, Jean J., P G Garella and Nathalie Sonnac (2007), 'Newspapers' Market Shares and the theory of the Circulation Spiral', *Information Economics & Policy* **19**, 405–413.

George, Lisa (2008), 'The Internet and the Market for Daily Newspapers', *The B.E. Journal of Economic Analysis & Policy* **8**(1).

Goldfarb, Avi and Catherine E. Tucker (2011), 'Advertising Bans and the Substitutability of Online and Offline Advertising', *Journal of Marketing Research* **48**(2), 207–227.

Goldfarb, Avi and Catherine Tucker (2014), 'Conducting Research with Quasi-Experiments: A Guide for Marketers', *SSRN Electronic Journal* .
URL: *http://www.ssrn.com/abstract=2420920*

Guo, Tong, S Sriram and Puneet Manchanda (2017), ''Let the Sun Shine in': The Impact of Industry Payment Disclosure on Physician Prescription Behavior', *working paper, SSRN ID=2953399* .

Hall, Tracy (2013), 'NYT Op-Ed Submission Directive', *NYT* .
URL: *http://www.nytimes.com/2013/10/14/opinion/op-ed-and-you.html?mcubz=0*

Hartung, Adam (2010), 'Don't fear Cannibalization: Embrace Future Solutions - Netflix, Apple Ipad, Newspapers'.
URL: *http://adamhartung.com/dont-fear-cannibalization-embrace-future-solutions-netflix-apple-ipad-newspapers/*

Haughney, Christine (2013), 'Digital Subscriptions Bolster Newspapers' Slipping Circulation Figures', *The New York Times* .
URL: *http://www.nytimes.com/2013/05/01/business/media/digital-subscribers-buoy-newspaper-circulation.html*

Ingram, Matthew (2015), 'New York Times ad revenue slides, but digital subs and native ads are growing', *Fortune* .
URL: *http://fortune.com/2015/10/30/new-york-times/*

Johnston, Michael (2014), 'What Are Average CPM Rates in 2015?'.
URL: *http://monetizepros.com/display-advertising/average-cpm-rates/*

Kim, Ho and Reo Song (2017), 'Free-to-Paid Transition of Online Content Providers: An Empirical Analysis of U.S. Newspapers' Paywall Rollout', *Working paper* .

Lambrecht, Anja and Kanishka Misra (2017), 'Fee or Free: When should Firms Charge for Online Content?', *Management Science* **63**(4), 1150–1165.

Li, Kathleen (2017), 'Statistical Inference for Average Treatment Effects Estimated by Synthetic Control Methods', *working paper* .

Lichterman, Joseph (2016), 'Nearly half of U.S. adults get news on Facebook, Pew says'.
URL: *http://www.niemanlab.org/2016/05/pew-report-44-percent-of-u-s-adults-get-news-on-facebook/*

Maher, John (2015), 'How and Why To Share Content That's Behind A Paywall', *McDougall Interactive* .
URL: *http://www.mcdougallinteractive.com/blog/content-development/share-content-behind-a-paywall/*

Marsh, Peter (2014), 'The state of paid content: For free, for a fee, or somewhere in between'.

URL: *http://www.inma.org/blogs/ahead-of-the-curve/post.cfm/the-state-of-paid-content-for-free-for-a-fee-or-somewhere-in-between*

Masnick, Mike (2016), 'Lots Of Newspapers Discovering That Paywalls Don't Work'.
URL: *https://www.techdirt.com/articles/20160815/09353135247/lots-newspapers-discovering-that-paywalls-dont-work.shtml*

Mutter, Alan (2015), 'Newsosaur: Tear Down That Paywall'.
URL: *http://www.editorandpublisher.com/columns/newsosaur-tear-down-that-paywall/*

Oh, Hyelim, Animesh Animesh and Alain Pinsonneault (2015), 'Free Versus For-a-Fee: The Impact of a Paywall on the Pattern and Effectiveness of Word-of-Mouth via Social Media', *MIS Quarterly* **40**(1), 31–56.

Pattabhiramaiah, Adithya, S Sriram and Shrihari Sridhar (2018), 'Rising Prices under Declining Preferences: The case of the U.S. Print Newspaper Industry', *Marketing Science* **Forthcoming**.

Pew Research Center, PEJ (2014), 'Annual Report of American Journalism'.
URL: *http://assets.pewresearch.org/wp-content/uploads/sites/13/2017/05/24141554/State-of-the-News-Media-Report-2004-FINAL.pdf*

Rochet, Jean-Charles and Jean Tirole (2003), 'Platform Competition in Two-Sided Markets', *Journal of the European Economic Association* **1**(4), 990–1029.

Salmon, Felix (2009), 'Newspaper self-cannibalization datapoint of the day'.
URL: *http://blogs.reuters.com/felix-salmon/2009/07/30/newspaper-self-cannibalization-datapoint-of-the-day/*

Seamans, Robert and Feng Zhu (2014), 'Responses to Entry in Multi-Sided Markets: The Impact of Craigslist on Local Newspapers', *Management Science* **60**(2), 476–493.

Sridhar, Shrihari and S. Sriram (2015), 'Is Online Newspaper Advertising Cannibalizing Print Advertising?', *Quantitative Marketing and Economics* **13**(4), 283–318.

Sweney, Mark (2015), 'Time spent reading newspapers worldwide falls over 25% in four years', *The Guardian* .
URL: *http://www.theguardian.com/media/2015/jun/01/global-newspaper-readership-zenithoptimedia-media-consumption*

Tadena, Nathalie (2015), 'New York Times Swings to Profit', *Wall Street Journal* .
URL: *https://www.wsj.com/articles/new-york-times-swings-to-profit-1446124971*

Thompson, Derek (2013), 'This Is the Scariest Statistic About the Newspaper Business Today'.
URL: *http://www.theatlantic.com/business/archive/2013/03/this-is-the-scariest-statistic-about-the-newspaper-business-today/274125*

Tirunillai, Seshadri and Gerard Tellis (2017), 'Does Offline TV Advertising Affect Online Chatter? Quasi-Experimental Analysis Using Synthetic Control', *Marketing Science* **Forthcoming**.

Tornoe, Rob (2016), 'Digital Publishing: For Some Publishers, Paywalls are Gaining Traction', *Editor & Publisher* .

URL: *http://www.editorandpublisher.com/columns/digital-publishing-for-some-publishers-paywalls-are-gaining-traction/*

van Nierop, J. E. M., P. S. H. Leeflang, M. L. Teerling and K. R. E. Huizingh (2011), 'The impact of the introduction and use of an informational website on offline customer buying behavior', *International Journal of Research in Marketing* **28**(2), 155–165.

Volokh, Eugene (2014), 'Paywall reprieve: Volokh content will be free and easy for everyone until Dec. 31, free for many (and free for all via RSS/Twitter/Facebook) after that'.
URL: *https://www.washingtonpost.com/news/volokh-conspiracy/wp/2014/07/25/paywall-reprieve-volokh-content-will-be-free-and-easy-for-everyone-until-dec-31-free-for-many-and-free-for-all-via-rsstwitterfacebook-after-that/*

Xu, Yixing (2017), 'Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models', *Political Analysis* **25**(1), 57–76.

Tables and Figures

Newspaper	Rank in 2010	Rank in 2011	Rank in 2013	Circulation in 2010	Circulation in 2013
Wall Street Journal	1	1	1	1,752,693	2,378,827
USA Today	2	2	3	1,671,539	1,674,306
New York Times	3	3	2	1,086,293	1,865,318
LA Times	4	6	4	1,078,186	653,868
Washington Post	5	5	8	763,305	474,767
Chicago Tribune	7	7	10	657,690	414,930
NY Daily News	6	6	6	701,831	516,165

Comparison of NYT, WP and USAT's newspaper readership (print + online)

Table 1: Top Newspapers in the U.S. by circulation

Source: Alliance for Audited Media's annual Newspaper Audit Reports;
http://www.thepaperboy.com/usa-top-100-newspapers.cfm

Newspaper	Paywall launch date
NYT	Mar 2011
WSJ	1996
Washington Post	June 2013 (enforced Dec 2013)
Chicago Tribune	Feb 2016
USA Today	Oct 2017
NYDailyNews	Feb 2018

Table 2: Paywall Launch Dates for the Newspapers in our Sample

Rank	DMA	Rank	DMA	Rank	DMA	Rank	DMA	Rank	DMA
1	New York	6	Washington D.C.	11	Minneapolis - St. Paul	16	Tampa - St. Pete - Sarasota	21	Cleveland
2	Los Angeles	7	San Fran - Oakland - San Jose	12	Phoenix	17	Orlando - Daytona Beach - Melbourne	22	Pittsburgh
3	Chicago	8	Dallas - Ft. Worth	13	Detroit	18	Indianapolis	23	Miami - Ft. Lauderdale
4	Boston	9	Atlanta	14	Houston	19	Denver	24	Sacramento - Stockton - Modesto
5	Philadelphia	10	Seattle - Tacoma	15	Portland	20	Hartford - New Haven	25	Charlotte

Table 3: Top 25 DMAs for NYT

	ln(Uniq. Visitors)		ln(Pages)		ln(Visits per visitor)		ln(Pages per visitor)		ln(Duration per visitor)	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
NYT x Paywall	-0.184**	0.029	-0.428**	0.073	0.010	0.125	-0.104	0.127	-0.112	0.148
#Obs-Treated					1025					
#Obs-Control					5125					

** p<0.01, *p<0.05, + p<0.1;

Standard errors are obtained from a placebo test, and are bootstrapped with 1000 replications.

Two-way fixed effects for DMAxnewspaper and month are included. The treatment effect is evaluated at the mean counterfactual.

Table 4: Effect of the paywall on NYT Online Visitation, aggregate data, generalized synthetic control

	ln(Unique Visitors)		ln(Pages)		ln(Visits per visitor)		ln(Pages per visitor)		ln(Duration per visitor)	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Light x NYT x paywall	-0.120^{+}	0.065	0.023	0.125	0.022	0.044	0.086	0.084	0.289*	0.113
Heavy x NYT x paywall	-0.858**	0.150	-3.560**	0.660	-0.632**	0.180	-0.884**	0.316	-0.390	0.362
#Obs-Treated					1025					
#Obs-Control					5125					

** $p<0.01$, *$p<0.05$, + $p<0.1$

Standard errors are obtained from a placebo test, and are bootstrapped with 1000 replications.

Two-way fixed effects for DMAxnewspaper and month are included. The treatment effect is evaluated at the mean counterfactual.

Table 5: NYT paywall on Online Visitation - breakup by Activity Level, median split, aggregate data, generalized synthetic control

20 pages

	ln(Unique Visitors)		ln(Pages)		ln(Visits per visitor)		ln(Pages per visitor)		ln(Duration per visitor)	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Light x NYT x paywall	-0.117*	0.058	0.095	0.131	-0.040	0.050	0.142	0.097	0.302**	0.115
Heavy x NYT x paywall	-0.751**	0.115	-3.110**	0.373	-0.788**	0.165	-1.508**	0.310	-1.806**	0.393
#Obs-Treated	1025									
#Obs-Control	5125									

30 pages

	ln(Unique Visitors)		ln(Pages)		ln(Visits per visitor)		ln(Pages per visitor)		ln(Duration per visitor)	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Light x NYT x paywall	-0.149**	0.060	-0.691**	0.297	-0.019	0.045	0.109	0.088	0.283*	0.111
Heavy x NYT x paywall	-0.571**	0.086	-2.677**	0.260	-1.082**	0.168	-2.010**	0.270	-2.106**	0.284

40 pages

	ln(Unique Visitors)		ln(Pages)		ln(Visits per visitor)		ln(Pages per visitor)		ln(Duration per visitor)	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Light x NYT x paywall	-0.154**	0.061	-0.063	0.132	-0.039	0.045	0.060	0.082	0.237*	0.113
Heavy x NYT x paywall	-0.148	0.138	-1.994**	0.202	-0.988**	0.133	-1.667**	0.171	-1.783**	0.198
#Obs-Treated	1025									
#Obs-Control	5125									

** p<0.01, *p<0.05, + p<0.1

Standard errors are obtained from a placebo test, and are bootstrapped with 1000 replications.

Two-way fixed effects for DMAxnewspaper and month are included. The treatment effect is evaluated at the mean counterfactual.

Table 6: NYT paywall on Online Visitation of users with varying activity levels, aggregate data, generalized synthetic control

Table 7: NYT paywall on Online Visitation - Tracking behaviors of Subscribers and Non-Subscribers, aggregate data, generalized synthetic control

| | 20 pages | | | | | | | | | | 40 pages | | | | | | | | | |
| | ln(Unique Visitors) | | ln(Pages) | | ln(Visits per visitor) | | ln(Pages per visitor) | | ln(Duration per visitor) | | ln(Unique Visitors) | | ln(Pages) | | ln(Visits per visitor) | | ln(Pages per visitor) | | ln(Duration per visitor) | |
	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Light x Non-Subs x NYT x paywall	-0.170**	0.061	-0.144	0.126	-0.039	0.043	0.003	0.089	0.082	0.118	-0.164**	0.058	-0.236+	0.133	-0.052	0.041	-0.067	0.083	0.025	0.115
Light x Subs x NYT x paywall	-0.548**	0.219	-3.142	2.662	0.223	0.533	-0.941	1.157	-4.815	2.846	-1.124	0.727	-0.373	5.351	1.825	1.048	2.682	1.792	3.888	2.723
Heavy x Non-Subs x NYT x paywall	-0.520**	0.079	-2.158**	0.204	-0.629**	0.130	-1.640**	0.250	-1.803**	0.269	-0.136	0.096	-1.334**	0.181	-0.638**	0.101	-1.238**	0.129	-1.301**	0.145
Heavy x Subs x NYT x paywall	-0.032	0.021	-0.152+	0.082	-0.108+	0.061	-0.321*	0.147	-0.230	0.243	0.004	0.018	-0.035	0.059	-0.013	0.032	-0.050	0.054	-0.008	0.081
#Obs-Treated											1025									
#Obs-Control											5125									

** $p<0.01$, * $p<0.05$, + $p<0.1$

Standard errors are obtained from a placebo test, and are bootstrapped with 1000 replications.

Two-way fixed effects for DMAxnewspaper and month are included. The treatment effect is evaluated at the mean counterfactual.

# Users		NYT △Pages $_{(Pre-Post)}$				NYT △Pages $_{(Pre-Post)}$		
	USAT	Increased	No Change	Decreased	WP	Increased	No Change	Decreased
Control group $\triangle_{(Pre-Post)}$	Increased	62	142	1	Increased	77	206	2
	No Change	630	67588	3745	No Change	609	68361	3874
	Decreased	4	1823	1179	Decreased	10	986	1049
	WSJ	Increased	No Change	Decreased	CT	Increased	No Change	Decreased
Control group $\triangle_{(Pre-Post)}$	Increased	48	85	2	Increased	9	41	1
	No Change	643	69045	4282	No Change	685	68670	4537
	Decreased	5	423	641	Decreased	2	842	387
	NYDN	Increased	No Change	Decreased				
Control group $\triangle_{(Pre-Post)}$	Increased	55	109	4				
	No Change	635	68358	4240				
	Decreased	6	1086	681				

Table 8: Two-way frequency table of change in newspaper visitation from pre to post
Exploring Substitution across Treated and Control Newspapers

(Pre period, per quarter)	Full Sample		Non-overlapping users	
	NYT	Control	NYT	Control
Visits	16.06	8.30	16.16	8.39
Pages	37.82	27.53	38.11	27.89
Duration	102.83	35.97	103.24	36.20

Table 9: Summary statistics for Full Sample and Exclusive Sample (users who accessed either NYT or one of the control newspapers, but not both)

Print Circulation for each Newspaper						
	NYT Weekend	NYT Weekday	USAT Weekend	USAT Weekday	WSJ Weekend	WSJ Weekday
Pre	1,686,020	1,034,263	2,454,332	2,207,041	1,919,427	2,039,218
Post	1,407,170	819,372	1,742,403	1,554,420	1,474,160	1,502,907
Percent change (pre to post)	-16.54%	-20.78%	-29.01%	-29.57%	-23.20%	-26.30%
Avg year on year percent change (2005-2013)	-3.28%	-1.89%	-6.54%	-6.19%	-4.39%	-5.20%
Avg year on year percent change (2005-2010)	-3.18%	-4.19%	-4.41%	-3.47%	-2.59%	-1.59%
Avg year on year percent change (2011-2013)	-3.44%	1.93%	-10.11%	-10.72%	-7.37%	-11.23%

Table 10: Print circulation trends for each newspaper

	All DMAs				Top 25 DMAs			
DV=	Weekday circulation share (%)		Weekend circulation share (%)		Weekday circulation share (%)		Weekend circulation share (%)	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
NYT x Paywall	0.35**	0.02	0.34**	0.03	0.40**	0.11	0.32*	0.14
N-treated			202				25	
N-control			404				50	

** p<0.01, *p<0.05, + p<0.1

All Models include two-way (newspaper x market and year) fixed effects; Newspaper sample: NYT (treated), USAT and WSJ (donor pool).

Table 11: Effect of paywall on print readership, generalized synthetic control

	All DMAs, USA Today as control group				All DMAs, WSJ as control group			
DV=	Weekday circulation share (%)		Weekend circulation share (%)		Weekday circulation share (%)		Weekend circulation share (%)	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
NYT x Paywall	0.46**	0.02	0.50**	0.03	0.26**	0.02	0.20**	0.02
DMA dummies, DMA specific linear and quadratic time trends	√		√		√		√	
R^2	0.63		0.49		0.70		0.67	

** p<0.01, *p<0.05, + p<0.1; standard errors are clustered by DMA

Table 12: Robustness check: Effect of paywall on print readership, panel regression

	Top 25 DMAs, most popular local newspaper in each market as control group			
DV=	Weekday circulation share (%)		Weekend circulation share (%)	
	Est.	SE	Est.	SE
NYT x Paywall	2.28**	0.70	2.27*	1.06
Time trend	-0.60	0.20	-0.89	0.32
DMA dummies	√		√	
R^2	0.86		0.87	

** p<0.01, *p<0.05, + p<0.1

Table 13: Effect of the paywall on print readership - using Local Newspapers as the control group, panel regression

	Total Print Circulation considering Subscription and Single Copy Sales	Subscription Sales	Single-Copy Sales
Effect of the paywall	29.3 %	23.31 %	n.s.
Controls (Fixed effects for each newspaper, year fixed effects specific to each newspaper, seasonality controls in the form of quarter of the year fixed effects included)	√	√	√

Table 14: Exploring the mechanism behind the spillover effect of the NYT paywall on its print circulation
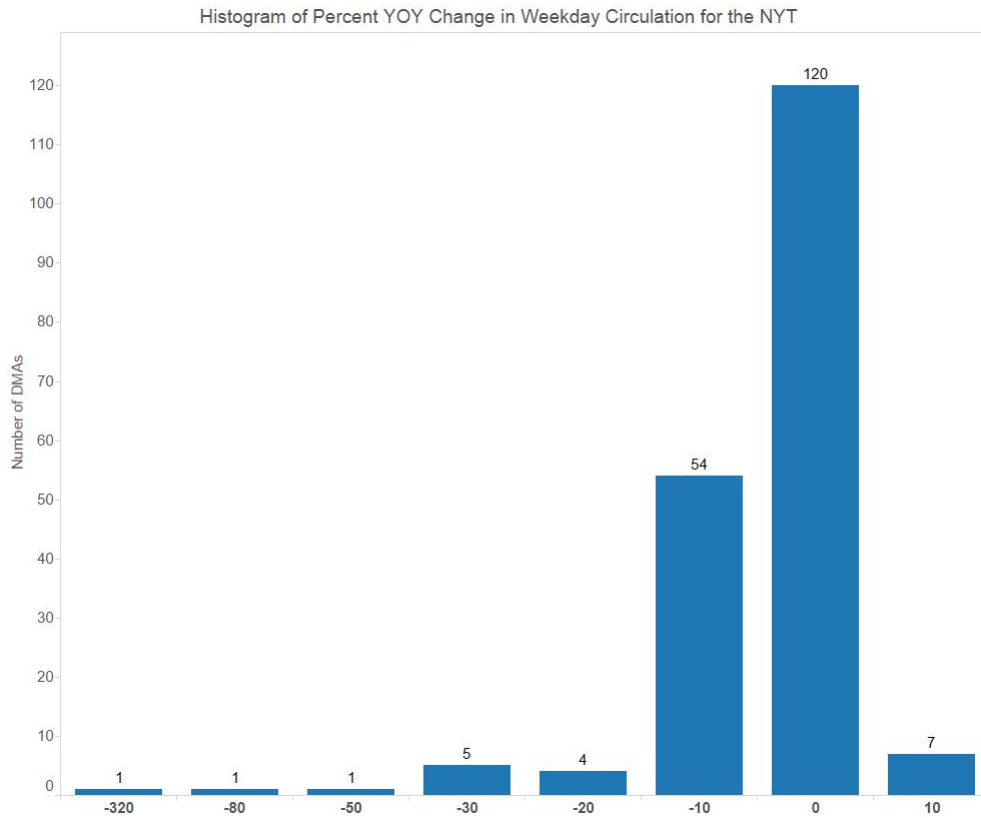
Figure 1: Change in NYT Weekday Print Circulation
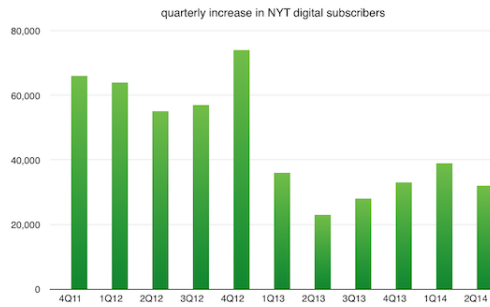
Source: http://www.cjr.org/the_audit/the_new_york_timess_paywall_ha.php

Figure 2: Growth in paid subscribers to NYT's website

Web Appendix

Appendix A - Robustness Checks

1) Estimating the spillover effect on the print newspaper using a panel regression

DV=	Top 25 DMAs, USA Today as control group				Top 25 DMAs, WSJ as control group			
	Weekday circulation share (%)		Weekend circulation share (%)		Weekday circulation share (%)		Weekend circulation share (%)	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
NYT x Paywall	0.62**	0.10	0.55**	0.13	0.42**	0.06	0.21*	0.09
DMA dummies, DMA specific linear and quadratic time trends	√		√		√		√	
R^2	0.58		0.49		0.84		0.74	

** p<0.01, *p<0.05, + p<0.1; standard errors are clustered by DMA

Table A.1: Robustness Check: estimating the spillover effect on the print newspaper, top 25 DMAs, panel regression

2) Examining the impact of excluding WSJ from the donor pool (set of control group newspapers)

| | 40 pages - including WSJ | | | | | | | | | | 40 pages - excluding WSJ | | | | | | | | | |
| | ln(Unique Visitors) | | ln(Pages) | | ln(Visits per visitor) | | ln(Pages per visitor) | | ln(Duration per visitor) | | ln(Unique Visitors) | | ln(Pages) | | ln(Visits per visitor) | | ln(Pages per visitor) | | ln(Duration per visitor) | |
	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Light x NYT x paywall	-0.154**	0.061	-0.063	0.132	-0.039	0.045	0.060	0.082	0.237*	0.113	-0.183**	0.059	-0.126	0.133	0.070	0.122	0.033	0.084	0.189[+]	0.098
Heavy x NYT x paywall	-0.148	0.138	-1.994**	0.202	-0.988**	0.133	-1.667**	0.171	-1.783**	0.198	-0.391**	0.079	-2.041**	0.206	-1.024**	0.115	-1.712**	0.182	-1.900**	0.213
#Obs-treated	1025																			
#Obs-control	4100																			

** $p<0.01$, * $p<0.05$, + $p<0.1$

Standard errors are obtained from a placebo test, and are bootstrapped with 1000 replications.
Two-way fixed effects for DMAxnewspaper and month are included. The treatment effect is evaluated at the mean counterfactual.

Table A.2: Assessing the impact of excluding WSJ from the donor pool

3) Assessing the impact of the NYT paywall in the short term (one quarter, two quarters after paywall rollout)

SHORT TERM EFFECT OF THE PAYWALL

20 pages, first month

	ln(Unique Visitors)		ln(Pages)		ln(Visits per visitor)		ln(Pages per visitor)		ln(Duration per visitor)	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Light x NYT x paywall	0.125	0.083	0.232	0.144	0.056	0.052	0.216*	0.104	0.318*	0.150
Heavy x NYT x paywall	-0.287**	0.065	-1.301**	0.171	-0.428**	0.122	-0.806**	0.160	-0.869**	0.208
#Obs-Treated	375									
#Obs-Control	1875									

20 pages, first two months

	ln(Unique Visitors)		ln(Pages)		ln(Visits per visitor)		ln(Pages per visitor)		ln(Duration per visitor)	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Light x NYT x paywall	0.069	0.073	0.354*	0.164	0.051	0.054	0.203+	0.108	0.299*	0.142
Heavy x NYT x paywall	-0.318**	0.064	-1.434**	0.153	-0.544**	0.123	-0.960**	0.162	-0.952**	0.210
#Obs-Treated	400									
#Obs-Control	2000									

20 pages, first three months

	ln(Unique Visitors)		ln(Pages)		ln(Visits per visitor)		ln(Pages per visitor)		ln(Duration per visitor)	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Light x NYT x paywall	0.084	0.062	0.160	0.153	0.069+	0.043	0.213*	0.099	0.339**	0.139
Heavy x NYT x paywall	-0.319**	0.062	-1.436**	0.167	-0.546**	0.127	-0.961**	0.161	-0.955**	0.205
#Obs-Treated	425									
#Obs-Control	2125									

20 pages, first six months

	ln(Unique Visitors)		ln(Pages)		ln(Visits per visitor)		ln(Pages per visitor)		ln(Duration per visitor)	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Light x NYT x paywall	0.048	0.055	0.159	0.157	0.093*	0.042	0.260**	0.090	0.404**	0.125
Heavy x NYT x paywall	-0.383**	0.066	-1.645**	0.204	-0.659**	0.170	-1.075**	0.208	-1.074**	0.228
#Obs-Treated	500									
#Obs-Control	2500									

** $p<0.01$, *$p<0.05$, + $p<0.1$

Standard errors are obtained from a placebo test, and are bootstrapped with 1000 replications. The treatment effect is evaluated at the mean counterfactual.

Two-way fixed effects for DMAxnewspaper and month are included.

Table A.3: Short term effect of the NYT paywall on Online Visitation of users with varying activity levels, aggregate data, generalized synthetic control

4) Examining cross-newspaper subtitution - Two-way frequency table of visitation behavior between the treated and control newspapers (visits and duration on the website)

# Users	NYT $\triangle$Visits $_{(Pre-Post)}$			NYT $\triangle$Visits $_{(Pre-Post)}$		
	USAT	Increased	No Change	Decreased	WP	Increased	No Change	Decreased
Control group $\triangle_{(Pre-Post)}$	Increased	30	57	0	Increased	23	40	0
	No Change	219	72208	1571	No Change	225	72424	1559
	Decreased	0	525	564	Decreased	1	326	576
	WSJ	Increased	No Change	Decreased	CT	Increased	No Change	Decreased
Control group $\triangle_{(Pre-Post)}$	Increased	22	20	0	Increased	3	18	0
	No Change	227	72598	1767	No Change	246	72373	1919
	Decreased	0	172	368	Decreased	0	399	216
	NYDN	Increased	No Change	Decreased				
Control group $\triangle_{(Pre-Post)}$	Increased	22	23	0				
	No Change	227	72264	1773				
	Decreased	0	503	362				

Table A.4: Two-way frequency table of change in newspaper visitation from pre to post
Exploring Substitution across Treated and Control Newspapers

# Users	NYT $\triangle$Duration $_{(Pre-Post)}$			NYT $\triangle$Duration $_{(Pre-Post)}$		
	USAT	Increased	No Change	Decreased	WP	Increased	No Change	Decreased
Control group $\triangle_{(Pre-Post)}$	Increased	48	127	4	Increased	54	137	4
	No Change	873	65974	5445	No Change	854	66189	5280
	Decreased	13	1626	1064	Decreased	26	1401	1229
	WSJ	Increased	No Change	Decreased	CT	Increased	No Change	Decreased
Control group $\triangle_{(Pre-Post)}$	Increased	46	65	5	Increased	10	79	3
	No Change	878	67035	5738	No Change	920	66736	6031
	Decreased	10	627	770	Decreased	4	912	479
	NYDN	Increased	No Change	Decreased				
Control group $\triangle_{(Pre-Post)}$	Increased	81	236	17				
	No Change	839	65840	5663				
	Decreased	14	1651	833				

Table A.5: Two-way frequency table of change in newspaper visitation from pre to post
Exploring Substitution across Treated and Control Newspapers

5) Examining possible substitution between the treated and control newspapers by considering behaviors of only those users who accessed EITHER the treated or one of the control newspapers in the pre or post periods

	20 pages						30 pages						40 pages					
	ln(Visits per visitor)		ln(Pages per visitor)		ln(Duration per visitor)		ln(Visits per visitor)		ln(Pages per visitor)		ln(Duration per visitor)		ln(Visits per visitor)		ln(Pages per visitor)		ln(Duration per visitor)	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
NYT x paywall x Light	-0.010	0.043	0.197**	0.025	0.414**	0.027	0.064*	0.018	0.137**	0.025	0.336*	0.027	0.041*	0.018	0.093**	0.024	0.284**	0.027
NYT x paywall x Heavy	-0.742**	0.002	-1.207**	0.003	-1.300**	0.005	-0.629**	0.002	-1.000**	0.002	-1.079**	0.005	-0.435**	0.001	-0.735**	0.002	-0.805**	0.002
#Obs-treated							1025											
#Obs-control							5125											

** p<0.01, *p<0.05, + p<0.1

Standard errors are obtained from a placebo test, and are bootstrapped with 1000 replications. Two-way fixed effects for DMAxnewspaper and month are included. The treatment effect is evaluated at the mean counterfactual.

Table A.6: NYT paywall on Online Visitation - Restricted sample of exclusive users created by eliminating users who accessed both NYT and any one of our control newspapers, from the sample, in either the pre or the post period.

6) Examining possible substitution behaviors between treated and control print newspapers

# Users		NYT △Wkday circ$_{(Pre-Post)}$				NYT △Wkday circ$_{(Pre-Post)}$		
	USAT	Increased	No Change	Decreased	WSJ	Increased	No Change	Decreased
Control group	Increased	147	29	11	Increased	131	31	10
	No Change	2	8	0	No Change	14	7	0
△$_{(Pre-Post)}$	Decreased	1	1	0	Decreased	5	0	1

Table A.7: Two-way frequency table of change in newspapers' weekday circulation from pre to post Exploring Substitution across Treated and Control Newspapers

7) Information on different subscription options offered by the NYT after the paywall:

	Subscription Option 1	Subscription Option 2	Subscription Option 3
$DIGITAL$	\$3.75 per week for web + mobile app access (\$195/year)	\$5 per week for web + iPad app access (\$260/year)	\$8.75 per week for all-access plan (\$455/year)
$PRINT^\sharp$	\$5 for Sunday only home delivery	\$6.5 per week for weekend home delivery	\$11.5 per week for daily home delivery

$\sharp$ - all NYT print subscriptions were provided free digital access -

https://www.theatlantic.com/business/archive/2011/03/new-york-times-erects-pay-wall/348883/

Table A.8: Subsciption options offered by the NYT around the time of the paywall

7) Examining the impact of promotional ad spending by the NYT:

| | | Excluding ad spending | | Including ad Spending | | | |
| | | | | Effect of the paywall | | Effect of Prom. Focused Ad Spending | |
		Est.	SE	Est.	SE	Est.	SE
	Unique Visitors	-0.117*	0.058	-0.103^{+}	0.056	0.026	0.034
	Pages	0.095	0.131	-0.324	0.251	0.126	0.090
LIGHT	Visits per visitor	-0.040	0.050	-0.099	0.078	-0.045	0.039
	Pages per visitor	0.142	0.097	0.179^{+}	0.093	0.068	0.060
	Duration per visitor	0.302**	0.115	0.352**	0.119	0.011	0.072
	Unique Visitors	-0.751**	0.115	-0.760**	0.117	-0.016	0.037
	Pages	-3.110**	0.373	-3.067**	0.374	0.082	0.093
HEAVY	Visits per visitor	-0.788**	0.165	-0.828**	0.191	0.071	0.053
	Pages per visitor	-1.508**	0.310	-1.284**	0.228	0.085	0.058
	Duration per visitor	-1.806**	0.393	-1.510**	0.323	0.099	0.070

** $p<0.01$, *$p<0.05$, + $p<0.1$; Heavy classification based on 20 pages of NYT pre-paywall usage.

Standard errors are obtained from a placebo test, and are bootstrapped with 1000 replications.

Two-way fixed effects for DMAxnewspaper and month are included. The treatment effect is evaluated at the mean counterfactual.

Table A.9: Examining the role of subscriber-acquisition related promotions by newspapers

8) Assessing generalizability of the effect of the paywall on NYT, by investigating the impact of the paywall launched by the LAT in Mar 2012:

| | ln(Unique Visitors) | | ln(Pages) | | ln(Visits per Visitor) | | ln(Pages per Visitor) | | ln(Duration per Visitor) | |
	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
LAT x paywall x Light	0.066	0.084	0.126	0.145	0.012	0.051	0.004	0.108	0.150	0.125
LAT x paywall x Heavy	-0.272**	0.112	-0.644**	0.168	-0.111**	0.079	-0.204**	0.109	-0.232**	0.127

N-treated	1025
N-control	5125

** $p<0.01$, *$p<0.05$, + $p<0.1$

Standard errors are obtained from a placebo test, and are bootstrapped with 1000 replications.

Two-way fixed effects for DMAxnewspaper and month are included. The treatment effect is evaluated at the mean counterfactual.

Table A.10: Assessing Generalizability: examining the effect of the LA Times Paywall, median split, generalized synthetic control.

9) Exploring the spillover effect of the LAT paywall on its print circulation

WP as Control group	Weekend		Weekday	
	Est.	SE	Est.	SE
Paywall	-6.32**	0.60	-4.72**	0.94
LAT x Paywall	4.58**	0.87	4.79**	1.36
Zipcode specific linear and quadratic trends	$\checkmark$		$\checkmark$	
Newspaper dummies	$\checkmark$		$\checkmark$	
R^2	0.89		0.84	

** p<0.01, *p<0.05, + p<0.1

Table A.11: Assessing Generalizability: examining the effect of the LA Times Paywall on print readership, panel regression.

Appendix B - Revenue Impact of the NYT paywall

As discussed earlier, the paywall resulted in lower engagement in online content, especially among heavy users. Lower engagement and traffic leads to a lower quantity of ad impressions that can be served on the newspaper's website. Thus, relative to the period before the paywall, this will lead to lower advertising revenue. However, as a result of the paywall, the newspaper is likely to have richer information on subscribing visitors, increasing its ability to serve *targeted* ads. Moreover, subscribing visitors, by virtue of their revealed willingness to pay for digital content, are likely to be more attractive to advertisers. In the absence of the paywall, advertisers would not have been able to directly identify such high willingness to pay users. Therefore, the paywall can potentially help a newspaper charge higher ad rates per impression (typically measured in terms of cost per mille or CPM) as a result of the improved quality of the served ad impressions. Therefore, the net effect of paywalls on online advertising (which we term the *indirect* effect) is likely to depend on the relative magnitudes of the changes in the quantity and quality of ad impressions subsequent to the paywall.

In order to study this we use the online advertising data described above. Our data consist of advertising expenditure and advertising impressions sourced from comScore's AdMetrix package. We first provide a plot of the temporal evolution of ad revenues for the NYT as well as for the broad category (total ad revenues invested in online newspapers in the U.S., as reported by the Newspaper Association of America) in Figure 3. We see that NYT's total digital ad revenue increased in the period following the paywall, just as the category spending did. More specifically, we find that

while the NYT's digital advertising grew at an average annual rate of 169.4% between 2009 and 2013, the corresponding increase for online newspaper advertising was about 9.08% during this period.

The average online ad rate (CPM or the cost per 1000 impressions) for the NYT increased by around 32% in the period following the paywall. As such, this is in contrast to the decrease in CPMs experienced by online display advertising during this period (Johnston, 2014). Prima facie, this might indicate that advertisers were willing to pay a premium for ads placed at NYTimes.com, possibly as a result of superior quality of ad impression served on the website. However, the fact that CPMs increased post-paywall does not necessarily imply that advertisers were willing to pay higher rates per impression - it is also conceivable that the NYT increased its CPMs in anticipation of the changes in the quality of impressions, although advertisers did not perceive such quality improvements.

In order to understand the impact of the paywall on NYT's online ad revenues, we run two sets of analyses: First, we track online advertising in a panel regression of logged online ad expenditure at the NYT considering corresponding ad spend on all US online newspapers as a comparison group (data we compiled from naa.org). This allows us to track the evolution of NYT's online advertising relative to that experienced by the overall category of online newspaper advertising. In the second specification, we regress a proxy for the number of digital ad impressions served at the NYT (in logs) on a paywall indicator, after controlling for time trends and seasonality in advertising.

Formally, we used the following specification:

$$A_{nt} = \alpha I_\tau + \theta_1 t + \theta_2 t^2 + \nu_t + \epsilon_{nt}, \tag{3}$$

where the dependent variable A consists of the logarithm of the ratio of total online ad spending (in \$) on the NYT in each month t to the corresponding category level ad spending (advertising spending on all online newspapers in the U.S.) in the first specification, and the logarithm of ad impressions for the NYT in the second specification.

In essence, the first specification allows us to track how NYT ad revenues evolved in relation to that of U.S. online newspapers. We control for temporal changes in advertising by using a parametric function of linear and quadratic month trends. We also include month of the year fixed

effects to capture seasonal variations in advertising behavior. In addition, we explore the impact of including separate month of the year fixed effects for the pre/post periods, as a robustness test of whether the paywall motivated seasonal shifts in ad spending levels on the NYT. The coefficient α captures the effect of the paywall on NYT's ad revenues, or the *indirect effect.*

We present the results in Table A.12. We find that the paywall had a negative effect of around 48.90% (calculated as [exp(-0.67)-1] from the estimate in column (1) in Table A.12) of post-period ad expenditure, which remains consistent when we consider a shorter time window before and after the paywall (see Table A.13). Results from the second model specification indicate that the paywall had a significant impact on the number of advertising impressions on the NYT website.[27] Considered against the backdrop of the finding that the paywall had a negative impact on the number of impressions served following the paywall (on account of the loss of the heavy user segment), one can rationalize the drop in advertising as arising from the reduced *quantity* of impressions - in line with the results from the second specification. This suggests that the effect of the paywall on online advertising due to changes in quantity of impressions served (the *quantity effect*) dominated the corresponding change to advertising due to the *quality effect.*

Overall, a decrease in circulation/readership can lead to lower advertising, especially in markets where advertising is known to be closely linked to the size of the reader base - such as newspapers (Rochet and Tirole, 2003; Fan, 2013; Pattabhiramaiah et al., 2018). This resulting lower cash flow can hamper the newspaper's ability to invest in quality. The resulting decline in quality can lead to a further decline in readership, thereby driving the quality-driven circulation spiral (Gabszewicz et al., 2007). In our context, if the paywall decreased online readership and led to lower online lower advertising, this might result in lower quality of content and decrease readership further.[28] While our data do not allow us to comment on the exact mechanism behind the circulation spiral, as we discuss below, the paywall had a net positive impact on NYT's revenues. Therefore, we do not foresee any adverse changes in the quality of online content because of the paywall.

In sum, our results of lower online engagement of NYT's readers after the paywall are consistent with the observed decrease in online advertising. An essential caveat to these results may be in order. Our choice of NYT was motivated mainly based on media reports of its success with executing

---

[27]We are thankful to an anonymous reviewer for guiding us to pursue this line of enquiry.

[28]We thank an anonymous reviewer for pointing out the role of declining circulation/advertising in driving the circulation spiral.

a paywall. While these media reports did not discuss specifics of the advertising gains for NYT from the paywall, our results appear to be inconsistent with the broad claims in media reports lauding the all-around success of NYT's paywall (Doctor, 2013). In fact, our findings are more consistent with industry reports that have lamented the losses in ad revenues accrued to newspaper firms after they erected digital paywalls (Ingram, 2015; Tadena, 2015).

So what is the total impact of the NYT's paywall on its overall revenues? In order to answer this question, we employ industry data as well as our model estimates to perform back of the envelope calculations to infer the revenue impact of the paywall. We first consider online subscriptions, which is a new source of revenue to newspapers on account of the paywall. At the end of our analysis period in 2013, 500,000 readers had signed up for NYT's paid membership. While the NYT offered various pricing tiers for different subscription plans ($3.75 per week for access to NYTimes.com, $5 per week for online+iPad access, $8.75 per week for unlimited access on all devices),[29] we do not have information on how many consumers signed up for each of these plans. Thus, we use the price of the cheapest plan ($3.75 per week) to arrive at the most conservative estimate for NYT's online subscription revenues. Using this metric, we compute that the NYT gained approximately $97.5 million in online subscription revenues in 2013. We next discuss the impact on online advertising. Given that the NYT lost approximately 48.90% in online revenues (compared to the category level baseline of online newspaper advertising), we can attribute a $7.34 million revenue loss (48.90% of the $15.02 million in online ad revenues) to the paywall.

Next, we consider the spillover effect of the paywall on the print newspaper. Recall that we discussed two plausible mechanisms behind the spillover effect: substitution effect wherein readers abandon the online version and switch to print as a result of the paywall, and the bundling effect wherein readers who would have otherwise subscribed to the online paywall instead find print subscription with free online access more attractive. Given our finding that bundling was probably the main driver of the spillover effect, we base our calculations by considering only bundling.[30]

In order to quantify the magnitude of the spillover effect, we consider two plausible scenarios. First is the most likely scenario based on the modal options chosen by NYT readers. In this case,

---

[29]http://www.nytimes.com/subscriptions/Multiproduct/lp5558.html – retrieved May 2013.

[30]If the spillover effect of the paywall were to be driven by the substitution explanation, all the increase in print readership would be deemed as incremental. This would likely imply a larger revenue gain. Therefore, we can view our analysis as a conservative assessment of the magnitude of the spillover effect.

we assume that the typical marginal reader subscribes to the modal weekend print option at a cost of $6.5 per week. Further, we assume that, in the absence of bundling online access with print, she would have chosen the modal/most affordable digital subscription option at $3.75 per week. Thus, the marginal revenue benefit of the paywall would be $2.75 (i.e., $6.5-$3.75) per subscriber per week. This, leads to a net incremental subscription revenue of $31.3 million. In addition, if we consider the incremental advertising revenue of $126 that each print reader of NYT generates (compiled based on the NYT's 2013 annual report), this would yield an additional $27.6 million in benefits to NYT. The total spillover effect accruing from incremental print subscriptions and advertising would thus be $58.9 million.

Second, we consider a more conservative scenario wherein the reader chooses between: i) the cheapest print option (Sunday only, costing $5 per week), and ii) the mid-priced digital option (which offered digital access on a browser+iPad for a fee of $5 per week). Given that the two prices are identical, there would be no marginal benefit from subscription revenues. Therefore, spillover effect of the paywall in this case would likely arise only from print advertising, i.e., $27.6 million. To put these results in context, for every $1 generated in online subscription revenue as a result of the paywall, the NYT lost $0.08 in online advertising revenue as a result of the indirect effect. At the same time, it gained between $0.28-$0.60 as a result of the positive effect of the paywall on the print newspaper.

There are two potential caveats to the quantification of the spillover effect. First, if advertisers are actively switching between print and online versions of the newspaper, it is possible that some of the calculated increase in print advertising might be a result of advertiser substitution away from online advertising at NYT. However, as Sridhar and Sriram (2015), Salmon (2009) and Hartung (2010) note, such cross-channel substitution is likely to be small. Second, the calculation of the incremental advertising revenues assumes that any increase in readership is immediately monetizable in the form of higher advertising revenues. In reality, it might take some time before this increase in ad revenues as a result of the change in readership can be realized.

The net benefit from the paywall under the conservative (modal) scenario is the sum of the three revenue components: direct effect via online subscription, indirect effect on online advertising and spillover effect on print readership, and consequently, advertising. Based on our calculations, this amounts to a gain of $117.7-$149.1 million, which represents between a 6.4%-8.1% of NYT's total

revenues in 2013. Thus, this research is one of the first to offer empirical evidence for a positive economic return from newspaper paywalls, by documenting that the NYT paywall was responsible for at least a 6% gain in its total revenues within a period of two years since its inception.
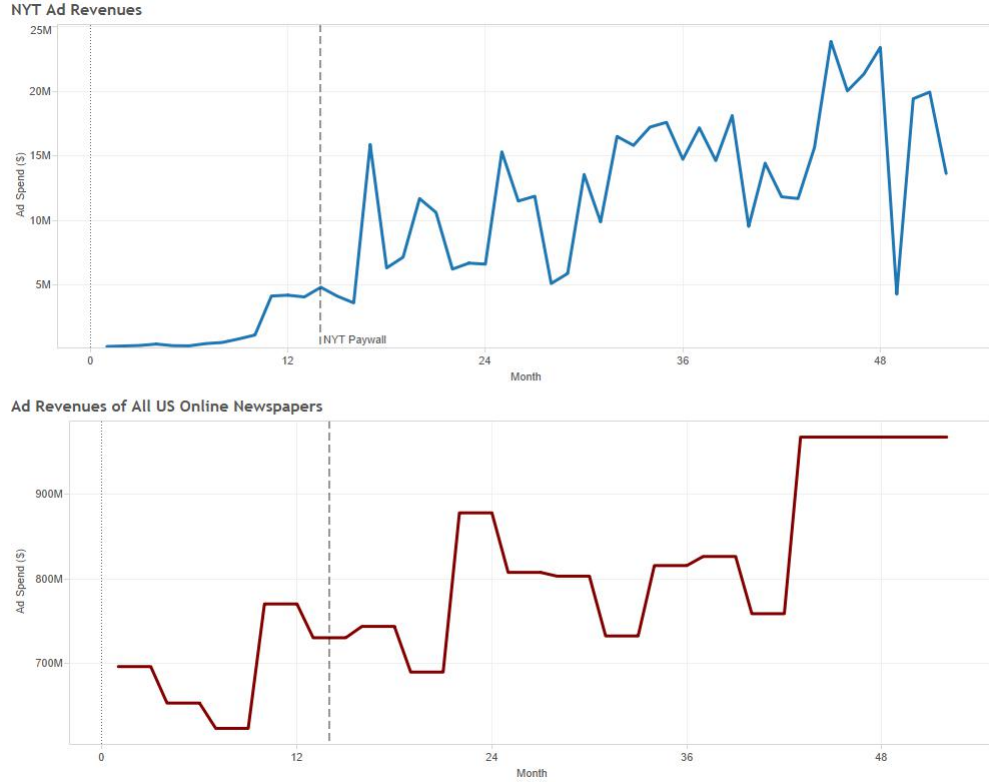

Figure 3: Temporal Evolution of NYT's Online Ad Revenues when compared with the Online Newspaper Industry's.

DV =	log of Ad Revenues						log of Ad Impressions					
	(1)		(2)		(3)		(4)		(5)		(6)	
Parameter	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Paywall	-0.67**	0.25	-0.59*	0.29	-1.45*	0.73	-0.96**	0.26	-0.90**	0.30	-1.52**	0.61
Month Trend	0.25**	0.02	0.25**	0.02	0.26**	0.02	0.26**	0.02	0.26**	0.02	0.26**	0.02
Month Trend - Quadratic	-3.1e-3**	3.0e-4	-3.1e-3**	3.3e-4	-3.2e-3**	3.4e-4	-3e-3**	3e-4	-3e-3**	3e-4	-3e-3**	4e-4
Month of the year fix ef.			✓						✓			
Separate month of the year fix. ef for pre/post					✓						✓	
Adj.R^2	0.89		0.90		0.92		0.88		0.89		0.91	

** p<0.01, *p<0.05, + p<0.1 ; Robust SE's are reported.

Table A.12: Online Advertising Regression

DV=	Log of Ad Revenues			
	(1)		(2)	
	1 Qtr pre/post		2 Qtrs pre/post	
	Est.	SE	Est.	SE
Post	-1.079**	-0.244	-1.150**	-0.318
Linear and quad. trends	√		√	
N	6		12	
Adj. R^2	0.902		0.672	

+ p<0.1 * p<0.05 ** p<0.01

Table A.13: Short term effect of the paywall on Online Advertising