Digital Paywall Design: Implications for Subscription Rates & Cross-Channel Demand

Sinan Aral & Paramveer Dhillon*
Sloan School of Management
Massachusetts Institute of Technology
{sinan|dhillon}@mit.edu
(Draft Date: 06/03/2017)

Abstract

Most online content publishers have moved to subscription-based business models regulated by digital paywalls. But, the implications of digital paywall design for total demand, cross channel demand, revenues and profits, are not well understood. We therefore utilized micro-level user activity data from The New York Times (NYT) together with multiple quasi-experiments to conduct the largest ever study of the causal impact of paywall design changes in the (1) quantity (the number of free articles) and (2) breadth (the number of available sections) of free content content available through the paywall on cross-channel demand and subscription rates. The results confirm an economically significant impact of the newspaper's paywall design on cross-channel demand: doubling the quantity of free content on the NYT mobile application increases cross-channel demand on the browser by 16.8%. But, doubling the breadth of free content available on the NYT mobile application reduces readership on the browser by 37.1%. Paywall design also has an economically significant impact on subscription rates. Reductions in the quantity of free content available and increases in the breadth of free content available directly caused about 78-80% of the total subscriptions observed during our seven month study. These findings can help structure the scientific discussion about digital paywall design and help managers optimize digital paywalls to maximize readership, revenue and profit.

*PD is a postdoctoral associate and SA is the David Austin Professor of IT and Marketing at the MIT Sloan School of Management. The authors would like to thank Catherine Tucker, Dean Eckles, Christophe Van den Bulte & Miguel Godinho de Matos for helpful comments. All errors remain our own.

1 Digital Paywall Design

The Internet has unmistakably disrupted content production and distribution and transformed the way news and other content are consumed. It is therefore now vital for content producers, like newspapers, to develop viable digital strategies to manage consumption and monetization across digital and traditional print channels as well as between digital channels. Until a decade ago, the main sources of revenue for publishers were advertisements (both print and digital) and print circulation. However, increased competition and reduced advertising margins have forced news-media outlets to adopt subscription-based business models to increase online circulation revenue (Casadesus-Masanell and Zhu, 2010). Popular news-media outlets like the Wall Street Journal (WSJ), the New York Times (NYT), and the Los Angeles Times, for example, have all moved to subscription-based models in order to survive (Tartakoff, 2009).

Subscription-based models are primarily regulated by digital paywalls, which mediate access to content and typically provide some free content to non-subscribers each month. As an experience good, readers must experience digital content to value it (Shapiro and Varian, 2013). Newspapers that implement digital paywalls therefore hope readers will sample their free content and, over time, value the content enough to subscribe.

When designing a digital paywall, content providers face a tradeoff: On one hand, they don't want to give away too much free content such that they don't appropriate enough value. On the other hand, they don't want to give away too little, such that potential subscribers are unable to sample the quality of the content and contribute to advertising revenue. This raises an important question: What and how much content should publishers give away to non-subscribers for free? Or, more broadly, how do the various design decisions at the heart of a digital paywall drive cross-channel readership, subscriptions, revenues and profits?

Digital paywall design can be thought of as an n-dimensional design space, with each dimension representing a parameter the designer can specify. For example, the designer can specify any number of the following parameters: **1) Quantity**: The number of free articles that non-subscribers can access in each time-period. Publishers of digital content can implement an all-or-nothing paywall (e.g. WSJ, The Financial Times (FT) and The Economist) which allows access to content for subscribers only, or a more lenient paywall (e.g. the NYT or Boston Globe) which gives access to some free articles to non-subscribers each time-period (e.g. 10/month in the case of NYT and 5/month in the case of Boston Globe). **2) Breadth**: The breadth of content that non-subscribers can access. Non-subscribers can either have access to all content across all sections (high breadth) or a limited subset of content, like popular news or politics, (low breadth) while only subscribers have access to more niche content like in-person player interviews as in the case of ESPN. **3) Temporal Differentiation**: The temporal inclusion or exclusion of content, such as full access to non-subscribers only on weekends, or only to monthly/quarterly special issues. **4) Archives**: The ability to access the digital archives of the newspaper. For instance, non-subscribers could only be given access to today's newspaper. **5) Curation**: Offering exclusive curation facilities like summaries of articles or forum access for conversations with the newsroom journalists. **6) Referrals**: Whether the paywall should allow free referrals to the newspaper's website from search engines, social media and news aggregators – sometimes referred to as a "porous" paywall.

Unfortunately, the design of digital paywalls remains a dark-art, as evidenced by newspapers' continuous tinkering with paywall designs (Wang, 2015). It has become pivotal for newspapers and other content providers to not only understand the anatomy of digital paywall design but also the mechanisms by which the various design parameters (e.g. quantity, breadth and referrals) impact variables of managerial interest (e.g subscriptions, reveneues and cross-channel demand). In addition, publishers and advertisers are interested in understanding interactions between mobile and desktop channels as this implicates the design of content offerings for each channel (e.g. different paywall offerings in each channel or offering more differentiated content through some channels than through others).

This paper attempts to open the black box of digital paywall design. We study the impact of arguably the two most critical parameters of a digital paywall—the quantity and breadth (or diversity) of free content available to non-subscribers—by exploiting quasi-experimental variation in these parameters at the NYT to measure their impact on cross-channel demand and subscription rates. A two-dimensional representation of this paywall design space is presented in Figure 1. This work makes several contributions to our understanding of digital and mobile disruptions in online content industries.

First, we econometrically identify the dynamics of content consumption across the mobile app

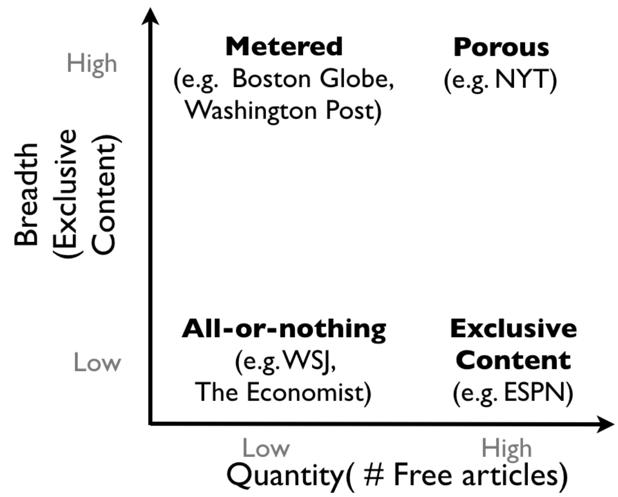

Figure 1: The paywall design space: Various strategies

and browser channels (two of the most prominent digital channels), which sheds light on whether these channels are complements or substitutes. It could be that readers have a limited time budget for consuming news and that the channels are not differentiated such that one channel substitutes for the other. Alternatively, if the channels are content differentiated and if reading one increases readers' interest in the other, the channels could be complements, increasing total consumption. Whether these channels are complements or substitutes in practice is an important empirical question for both academics and managers.

Second, we estimate the impact of paywall design on the NYT's subscriber base. Until now there has only been anecdotal evidence that digital paywalls affect subscriptions (Jackson, 2013; Kumar et al., 2013). But, there exists no rigorous econometric quantification of such effects. Moreover, it is unclear whether paywalls increase or decrease subscription rates. The introduction of a paywall may repel potential subscribers because they get no chance to sample the content first. Alternatively, it may entice more subscriptions if the quality of the newspaper's journalism is known to be high or if the content is differentiated enough to justify the subscription fee. We adjudicate this debate with fine grained behavioral data at the user level.

Third, our work sheds light on the role different devices play in our content consumption behaviors. It is unclear how users consume content differently when using different digital devices (Boik et al., 2016). Tablets provide a more 'laid-back' and immersive experience than smartphones (Kumar et al., 2013) but their size reduces their mobility. Our work quantifies the impact of the use of different mobile devices on the dynamics of news consumption.

Our results suggest a nuanced relationship between paywall design, cross-channel demand and subscription rates. An increase in readership on the mobile app leads to a corresponding increase in readership on the browser, demonstrating a complementarity between these two channels. However, an increase in the breadth of content consumed on the mobile app leads to a decrease in readership on the browser, suggesting more breadth creates substitution across these channels. The availability of more diverse content allows readers to choose content aligned with their preferences which leads to higher overall engagement. We also find important heterogeneous effects across different user segments. For example, the complementarity between the channels is significantly less for heavy mobile app users than light mobile app users, and also less for tablet users compared to non-tablet users. The effect of an increase in the diversity of free content available on the mobile app cannabilizing readership on the browser is even stronger for tablet users than for non-tablet users. Our results also suggest a statistically and economically significant impact of paywall design on subscription rates. Decreases in the quantity of free articles available and increases in the breadth of free articles available caused 78-80% of the subscriptions we observed during our seven month study.

In summary, our work econometrically identifies whether the mobile app and the browser are complements or substitutes; measures the impact of changes in the quantity and breadth of content offered through the digital paywall on cross-channel demand; and causally estimates the impact of paywall design on subscription rates.

2 Theory & Literature

Newspapers want to increase demand for their content to increase their subscription and advertising revenues. There is heterogeneity in readers' willingness to pay for content—the high willingness to pay readers generate revenue through subscriptions[1] and the low willingness to pay readers generate (albeit less) revenue through advertising. The goal is to maximize revenue by regulating subscription rates, web traffic and advertising and subscription prices through the use of a digital paywall.

[1]Actually, high willingness to pay readers i.e. subscribers, also generate advertisement revenue as they also see some advertisements. However, it dwarfs the subscription revenue they generate.

Consider the following thought exercise: Let $i = 1, \ldots, I$ be the total number of articles published by the newspaper on a given day and let a_i denote the q dimensional vector of attributes of the i^{th} article e.g. topic, political lean, author, tone, writing style etc. On the consumption side, assume that $j = 1, \ldots, J$ readers visit the newspaper website and consume content on a given day and that the j^{th} reader has a q dimensional vector preference vector u_j which represents their preference for the corresponding attribute of an article.

The demand for content is a function of the expected match between an article's attributes and the reader's preferences i.e. $m_{ij} = a_i \cdot u_j$—the better the match (m_{ij}), the higher the demand. Readers have uncertainty regarding the quality of the match of the newspaper's content with their preferences, so each article consumed by the reader updates their expectation of the quality of the match. This makes it pivotal for newspapers to decide how much free content to give away via the digital paywall. If the volume of free content is too low, readers will have high uncertainty regarding the potential fit of the newspaper to their tastes and their willingness to pay for content might not be higher than the subscription cost. On the other hand, if the volume of free content is too high, publishers risk not holding back enough content that would rationalize subscribing[2]. This is Arrow's Information Paradox applied to digital journalism (Arrow, 1962).

Publishers have two paywall design parameters that regulate users' content sampling—the quantity of free articles (n) and the diversity of free articles (d) i.e. which sections free articles are from. By changing n and d, publishers change the distribution of articles from which readers can sample content and hence the preceived match between the content and the readers' preferences. Though publishers could pursue a variety of digital content strategies by adjusting these two parameters, it is an empirical question what the impact of these strategies are on content demand and revenue.

Our work therefore contributes to the literature on the impact of match quality on demand. (Roos and Shachar, 2014) study this in context of movies and (Roos et al., 2015) model matches in internet news consumption. We also contribute to the literature on digital paywalls (Chiou and Tucker, 2013; Lambrecht and Misra, 2015). This literature has shown that having a digital paywall reduces the demand for content and hence visits to a website. Further, (Lambrecht and Misra,

[2]Therefore, the extreme case of giving all content for free to maximally reduce uncertainty regarding the quality of match is also not feasible as then the newspaper will have to rely on just the advertisement revenue, which is highly unreliable

2015) show how the firms that could offer both free and paid content should adjust their content in periods of high demand. To our knowledge, there has been no theoretical or empirical work on how the key digital strategy decisions at the heart of digital paywall design impact readership or revenues.

Due to the proliferation of mobile devices, these days, readers consume content via multiple digital channels e.g. desktop, smartphone, tablet and, even on the smartphone or tablet, they can use the mobile web browser or the mobile app. Publishers want to increase demand across all digital channels, as demand drives revenues from subscriptions and advertisements. It is unclear, however, how a particular digital strategy impacts same channel or cross channel demand, or readers' expectations about the match between their tastes and the content being offered and therefore their willingness to pay. So, newspapers are increasingly interested in understanding how paywall design shapes cross-channel demand and subscription rates.

Here our work contributes to the empirical literature on consumption dynamics across different channels (Ansari et al., 2008; Avery et al., 2012; Geyskens et al., 2002; Athey et al., 2014; Deleersnyder et al., 2002). This work has mostly focused on the connection between digital and brick-and-mortar channels (Brynjolfsson et al., 2009; Ellison and Ellison, 2006; Goolsbee, 2001; Prince, 2007) and between digital news-media outlets and traditional print media (Athey et al., 2014; Deleersnyder et al., 2002). There has also been some work that quantifies interdependencies among mobile-only channels (Bang et al., 2013; Ghose et al., 2013) and between the tablet and other channels (Xu et al., 2015). Perhaps, the work that comes closest to ours in this literature is (Xu et al., 2014) which studies the impact of the introduction of the Fox News mobile app on the mobile browser channel.

The main finding from this literature is that we observe substitution when the channels have duplicate capabilities (Deleersnyder et al., 2002), when one of the channels offers new capabilities (Alba et al., 1997) or when people have limited time to spend on media consumption (Deleersnyder et al., 2002). Complementarity is observed when a channel provides additional utility to consumers above and beyond existing channels. This can come about in several ways, for instance one channel can serve as a form of advertisement for the other channel (Avery et al., 2012) or a channel can offer different comparative advantages, for example when the digital channel offers more variety and lower prices whereas brick-and-mortar stores offer convenience or avoidance of shipping costs (Forman et al., 2009). Whether the various digital channels are complements or substitutes depends on users' perceptions. If a channel replaces the capabilities of another channel, then it might cannibalize its demand. But if a channel only contains a subset or summaries of the content accessible via the other channel, then the channels might be complements as users might find the lead to a story on one channel and later visit the other channel for the complete story (Xu et al., 2014).

Along similar lines, it is also unclear what the impact of a paywall design change would be on the size of publishers' subscription base. The subscription base would stay the same if the readers do not subscribe, which could happen for several reasons, e.g. the newspaper might not have given away enough content for free that there is high uncertainty regarding the quality of the match or it might have given too much content for free that it obviated the need to subscribe. The subscription base could increase if the quality of the free content has given a strong enough signal to increase the readers' willingness to pay to a value higher than the cost of the subscription bundle.

During our study, the New York Times (NYT) made several changes to their digital paywall strategy with respect to the quantity and diversity of free content offered. We exploit this quasi-experimental variation in both the quantity and diversity of content available via the digital paywall to study its impact on content demand and subscription rates. Our work is the first to study the causal impact of fine grained design choices at the heart of the digital paywall i.e. how the amount of free articles and diversity of free content allowed impacts newspapers' readership and revenue across various digital channels. This has direct actionable implications for publishers' digital content strategies.

2.1 A simple model of cross-channel demand

Consider a model of content consumption in which, in each time-period t, a given user i has a news budget I_{it} which they spend reading news stories on the NYT and other news outlets. Each news story has a "type" v ($\in 1, \ldots, V$) e.g. which section does it belong to—Top News, Sports, Weddings etc. Further, each news-story can be consumed via different modes m ($\in$ **B**rowser, **A**pp (Mobile)).

Users can subscribe to the NYT by paying a subscription fee P—and get unfettered access to all of the content. The subscription status of a user, is denoted by $S_{it} = (0, 1)$.

There is a different cost associated with each news-story based on its type, the mode by which it is consumed and whether the user is a subscriber or not $C_{it}^{v,m,S}$. The main factors that contribute to this cost are the time costs, access costs (not having access at all or having only limited access, for instance via a digital a paywall) and the cost of associated workarounds (e.g. using a different browser or deleting cookies to bypass the digital paywall etc). Users can also consume news stories at a different news outlet at a cost of $\bar{C}_{it}$. The cost of consumption faced by subscribers is only the time cost, and is lower than the cost faced by non-subscribers: $C_{it}^{v,m,S=1} \leq C_{it}^{v,m,S=0}, \bar{C}_{it} \;\forall\; v, m$.

The vector of overall news consumption by user i in time-period t is denoted by $R_{it}^{v,m}$ and the consumption from other news outlets is denoted by $\bar{R}_{it}$. We assume a Cobb-Douglas utility function for the user's news consumption with constant returns to scale (Athey et al., 2017)[3].

$$
\begin{aligned}
U_{it} &= \left[\prod_{v \in V, m \in \{B,A\}} \left(R_{it}^{v,m} \right)^{\alpha_{v,m}} \right]^{\tau_i \tau_t} \bar{R}_{it}^{1-\tau_i \tau_t} \\
\text{such that, if } S_{it} = 1, \qquad I_{it} &\geq \left[\sum_{v \in V} \sum_{m \in \{B,A\}} R_{it}^{v,m} C_{it}^{v,m,S=1} \right] + \bar{R}_{it} \bar{C}_{it} + P \\
\text{if } S_{it} = 0, \qquad I_{it} &\geq \left[\sum_{v \in V} \sum_{m \in \{B,A\}} R_{it}^{v,m} C_{it}^{v,m,S=0} \right] + \bar{R}_{it} \bar{C}_{it}
\end{aligned}
$$

The above formulation assumes that a user's share of time spent reading news is decomposed into day/seasonal effects (τ_t) and any individual-level idiosyncrasies (τ_i) in news consumption. The utility function also implies that the users' preferences are stable over time. So, a utility maximizing user will consume a constant share of news stories of a particular type, using a particular mode, as long as the cost of reading different types of news do not change.

Using the first-order-conditions, it is straightforward to derive the demand of news stories of type v in mode m by both subscribers and non-subscribers.

[3]The constant returns to scale assumption places restrictions on the news consumption process, but can be relaxed in a more general model, which is not the main focus of this paper.

For non-subscribers, we get:

$$R_{it}^{v,m*} = \frac{\tau_i \tau_t \alpha_{v,m} I_{it}}{C_{it}^{v,m,S=0}} \tag{1}$$

and for subscribers, we get:

$$R_{it}^{v,m*} = \frac{\tau_i \tau_t \alpha_{v,m} (I_{it} - P)}{C_{it}^{v,m,S=1}} \tag{2}$$

In our setting, the NYT changed their paywall offerings in terms of the number and breadth of free articles offered. In other words, their policy changes directly manipulated the cost of consuming stories of different types. Several consequential propostions follow directly from this model:

Proposition 1. *When the cost of reading one type of article decreases, the individuals affected will increase their consumption of that type of article.*

Proof: This follows directly from the demand bundle derived above as lowering the costs $C_{it}^{v,m,S=0}, C_{it}^{v,m,S=1}$, increases the consumption $R_{it}^{v,m}$ for that type of article.

Proposition 2. *Individuals with a high news-budget i.e. high I_{it} who want to read more, will become subscribers.*

Proof: The proof follows from the fact that a user will subscribe when they get more utility from reading news stories as a subscriber than by consuming news stories as a non-subscriber. In other words:

$$U_{it}(R_{it}^{v,m*} | S_{it} = 1) \geq U_{it}(R_{it}^{v,m*} | S_{it} = 0) \tag{3}$$

Substituting the demand from Equations 1 and 2 in 3 and simplifying, we get

$$\prod_{v \in V, m \in \{B,A\}} \left(\frac{I_{it} - P}{C_{it}^{v,m,S=1}} \right)^{\alpha_{v,m}} \geq \prod_{v \in V, m \in \{B,A\}} \left(\frac{I_{it}}{C_{it}^{v,m,S=0}} \right)^{\alpha_{v,m}} \tag{4}$$

Combining this with the fact that $C_{it}^{v,m,S=1} \leq C_{it}^{v,m,S=0}$, we can see that users with high I_{it} will become subscribers. This simple model describes the broad conditions under which consumers will subscribe and increase or decrease their consumption across different channels and devices as the cost of consumption changes. With this simple model of consumption and subscription behavior to guide our expectations, we now turn to our empirical estimation.

3 Empirical Setting

We use seven months of fine-grained, individual-level data from New York Times (NYT) to study the impact of changes in the quantity and breadth of content accessible via the digital paywall on cross-channel demand and subscription rates. Specifically, we studied the dynamics of content consumption in two predominant digital channels: native mobile applications (mobile apps) and the browser (both desktop and mobile) and how this dynamic drives subscriptions.

About 70% of the NYT's advertising revenue comes from print advertising, but the NYT saw a 3.4% decline in print advertisement revenue in 2015 (Ingram, 2015). Further, it only saw a 1% increase in print circulation in the same period (Ingram, 2015). The $400 million a year it makes in digital revenue (about half from digital advertising and the other half from subscriptions) is not enough to support operating costs of $1.5 billion and a newsroom of 1000 reporters. The revenues from various streams for Q2-2016 are: $45 million from digital advertising, $86 million from print advertising, $219 million from print circulation and $56 million from digital-only subscriptions[4]. Dating back to 1996, the NYT has experimented with charging for content several times, though these attempts were mostly unsuccessful (Kumar et al., 2013). The threat posed by digital and mobile disruptions to the NYT coupled with its scale and the quality of journalism make it a highly representative testbed to study the design of a digital paywall and its affect on demand and subscription rates across different channels such as the mobile app and browser.

The mobile app has a user-friendly interface and can harness the device and platform specific characteristics of the operating system, which the browser cannot. However, this comes at a cost: Since the browser environment is generic it allows easy portability and integration of content across digital platforms, whereas the marginal impact of adding content to the mobile app is higher as it requires non-trivial code development (Mikkonen and Taivalsaari, 2011; Lionbridge-Technologies-Inc, 2012; Summerfield, 2011). For these reasons, mobile apps typically have less diverse content than is accessible from the desktop or mobile browser. We observe this on the mobile app of the NYT as well[5]. But, the causal identification of the relationship between consumption on the mobile app and browser channels is difficult when using observational data

[4] `http://www.nytimes.com/2016/07/29/business/media/new-york-times-co-q2-earnings.html`

[5] (Xu et al., 2014) also observe 17/22 providers offering less diverse content on their mobile app compared to the browser.

as there are endogeneity concerns due to correlated unobservables which affect consumption on both the mobile app and browser. So, we need to identify exogenous demand shocks that can perturb demand in one channel without directly affecting demand on the other channel. It is also unclear how the various elements of paywall design impact subscription rates. Quantifying the impact of paywalls on subscription rates is extremely important for managers to help them efficiently design the paywall, but such quantification is also riddled with endogeneity concerns.

In this paper we use several changes to the NYT's digital paywall settings, which changed the quantity of content available (from unlimited to 3 articles/day i.e. *high n* to *low n*) and the breadth (or diversity) of the content available (i.e. comparing access to only the Top News and Video sections to access to all the sections i.e. *low d* to *high d*) only on the mobile app channel as quasi-experimental variation in the demand (quantity and diversity) for content on the mobile app. Since these policy changes were temporally separated, they allow us to disentangle the impact of changes in quantity (unlimited to 3/day) from changes in diversity (Top News and Videos to all the sections). These quasi-experiments allow us to causally identify the connection between the mobile app and browser channel, quantify the impact of paywall design changes on cross-channel content consumption and estimate the impact of the paywall on subscription rates.

4 Data

Our data consists of user-level activity from the New York Times (NYT) for the seven month period from April to October 2013. The data track the browsing behavior of 525 million unique visitors who accumulated over 2.5 billion page views during this period. For the analysis in this paper, we construct a panel of 808,785 users who are either registered (which gives them the ability to comment on articles, save articles for future reading and get personalized content recommendations) or subscribed and who accessed the newspaper's website at least once from a mobile app or browser. We exclude anonymous users, who are only identified by browser cookies, because they cannot be tracked reliably across devices.[6]

[6]The effect of this limitation is small as registered users are those who have shown interest in the NYT's brand of journalism and have a higher propensity to become subscribers as compared to "ones-and-done" anonymous users.

4.1 Variable Construction

Readership and Diversity Variables: We construct the readership variables, readership on the app (R^A_{it}) and readership on the browser (R^B_{it}) as the number of articles read by user i on day t on the app and browser respectively. The choice of the unit of time as a day is primarily due to the perishable nature of news content. The diversity variable D^A_{it} was constructed by counting the number of unique sections from which user i consumed articles on day t. For instance, if user #10 read seven articles via the mobile app and three articles via the desktop and mobile browsers on day 76, and the seven articles read on the mobile app were from the sections Top News, Sports, Weddings, and the Op-Ed section and the three articles read on the browser were all from Politics, then, the variables will be coded as $R^A_{10,76} = 7$, $R^B_{10,76} = 3$, $D^A_{10,76} = 4$. After the users have exhausted their limit of free articles, they see pop-ups urging them to become a subscriber in order to continue reading. The variable $Popups_{it}$ encodes the number of pop-ups seen by the user i on day t. Note that the users only started seeing pop-ups once the paywall change was in effect as described later.

Subscription Variable: The subscription variable S_{it} indicates the subscription status of user i on day t. Readers can subscribe to one of the four available bundles (1) all digital, (2) all digital and home delivery, (3) web and smartphone, (4) web and tablet. For simplicity and the ease of interpretability we pooled all the subscription bundles and coded the subscription status as a binary variable as equal to 0 if the user was registered and 1 if the user was a subscriber. We also estimated the heterogeneity of treatment effects across the sub-populations of subscribers with bundles (3) and (4), as the treatment (change in paywall) could be applicable to these subscribers, (e.g. if the web and smartphone bundle subscribers also access content via a tablet).

4.2 Summary Statistics

We created a daily panel of registered and subscribed users taking into account their readership activity on the browser and mobile app as well as their subscription status at the user-day (it) level. In our panel of 808,785 registered users, 54% of the users did not declare their gender. Of the remaining 46%, 61% were males and 39% were females. Similarly, 41% of the users declared their location. 91% of them were from U.S., 2% from Canada and the remaining 7% from the rest of the world (195 different countries).

All of the users accessed the newspaper websites at least once from the browser or mobile app. In our sample, 67% of users had only one mobile device (iPhone, iPad, Android or iPodTouch), 31% had two devices and the remaining 2% is split between users with three or four devices. The devices used were split 10%, 44%, 45% and 1% between Android, iPhone, iPad and iPodTouch respectively.

In total, users in our panel read a total of 130,587,022 articles via the mobile app and 142,439,549 articles via the desktop/mobile browsers from 158 unique sections. Tables 1 and 2 show the summary statistics and correlations between the various readership and diversity variables. As can be seen from the summary statistics, the variance of the readership and diversity variables is greater than their mean, suggesting a long-tail.

Variable	Mean (μ)	SD (σ)	Minimum	Maximum
Readership on app (R_{it}^{A})	.71	2.32	0	36
Readership on browser (R_{it}^{B})	.81	2.42	0	108
Diversity of content on app (D_{it}^{A})	.37	1.06	0	20
Subscription Pop-ups on app ($Popups_{it}$)	.006	.13	0	14
# Registered Users (R)	179,695	31818	143,222	245,692
# Subscribed Users (S)	629,090	31818	563,093	665,563

Table 1: Table showing summary statistics of the relevant readership and diversity variables. Results are computed for a panel of (users) n=808,785, (days) t=212 resulting in a total of 171,462,420 user-day observations.

Variable	1	2	3	4
1. Readership on app (R_{it}^{A})	1.00	-	-	-
2. Readership on browser (R_{it}^{B})	.052	1.00	-	-
3. Diversity of content on app (D_{it}^{A})	.403	.051	1.00	-
4. Subscription Pop-ups on app ($Popups_{it}$)	.072	-.004	.06	1.00

Table 2: Table showing correlations (Pearson) between the various variables of interest.

4.3 Quasi-experiment: A policy change in the digital paywall

The NYT launched their digital paywall in 2011. Since that time they have instituted a porous paywall through which unsubscribed users can read a fixed number of articles every month (currently 10) and in addition can access more articles each month if they are referred to those articles through social media websites or search engines.

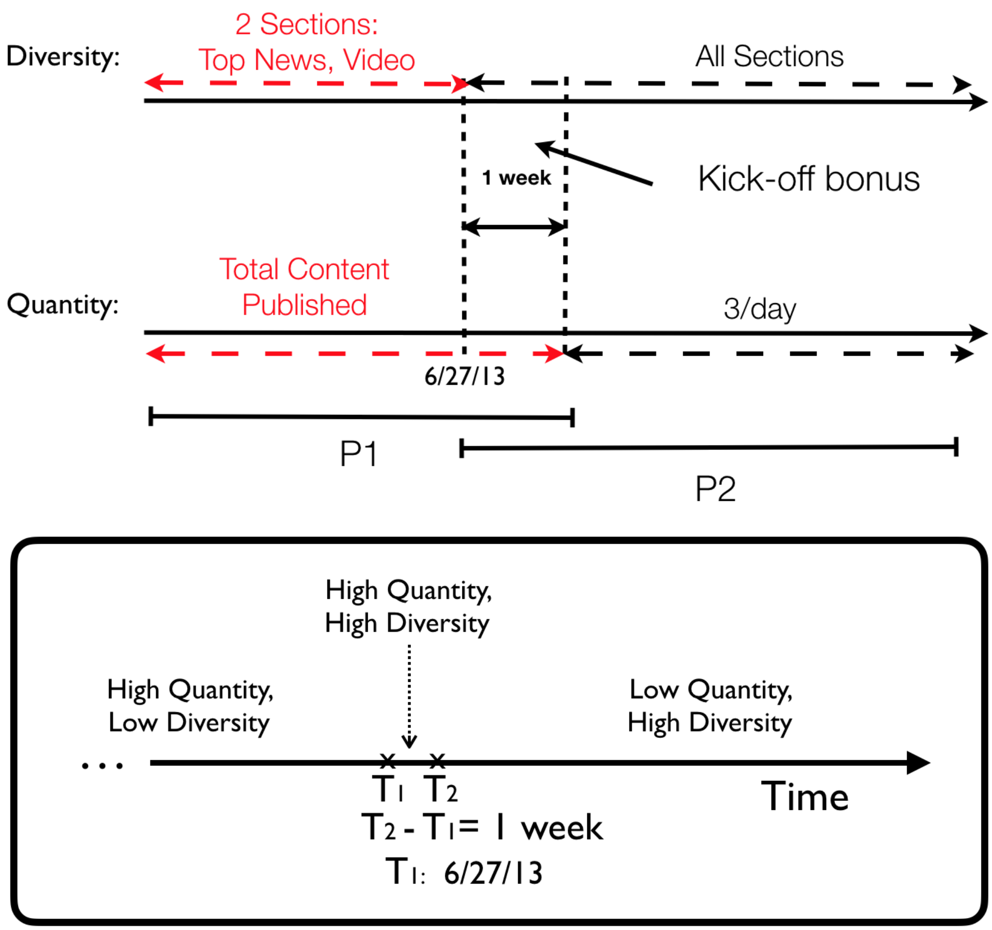

Figure 2: Details of the Paywall setting change by our partner newspaper. Note: “High Quantity”= Total Content Published, “High Diversity”= All Sections, “Low Quantity”= 3 articles per day, “Low Diversity”= Top News and Video sections.

NYT digital content is distributed through three channels: (1) the main website (`www.nytimes.com`) accessible from desktop computers (2) the mobile website (`www.mobile.nytimes.com`) accessible via browsers on mobile devices (smartphones and tablets) and (3) the mobile app which can be installed on smartphones and tablets of all varieties. The NYT paywall allowed ten free articles per month via channels (1) and (2), whereas users could read an unlimited number of articles through channel (3), but only from the *Top News* and *Video* sections of the mobile application.

However, on June 27, 2013, the NYT started metering their mobile apps (channel (3)) such that unsubscribed users could only read 3 articles per day. At the same time, those articles could now be accessed from any section and not just from the *Top News* and *Video* sections. To kick off the update, users had a 1 week trial period where they could still access unlimited articles as before but from any section and after the trial period, they were only able to access 3 articles per day. If, after hitting their quota, a user tries to access more articles, they see a pop-up in the mobile app urging them to become a subscriber. This change in the paywall's settings was exogenous to the readership on the browser channel and serves as a demand shock for the content consumed on the mobile app. Not every user in our sample downloaded the updated version of the mobile app at the same time. This heterogeneity in the timing of updates, which is not correlated with any observable differences between users, provides exogenous shocks to consumption that vary across both users i and time t. Figure 2 displays the details of the quasi-experiment, which we describe in more detail below. We use this quasi-experimental variation in the quantity and diversity of the content available via the digital paywall to identify the impact of these two key elements of paywall design on cross-channel demand and subscription rates.

5 Model Specifications

We are interested in digital paywall design and, in particular, the impact of the quantity and breadth/diversity of content offered through the paywall on cross-channel demand and subscription rates.

5.1 Cross-Channel Demand

In order to establish and estimate the effect of changes in the amount of content consumed in the mobile app channel on consumption in the browser channel, we estimate the following specification with the user-day panel that we constructed:

$$R_{it}^{B} = \alpha + \beta_A R_{it}^{A} + \beta_D D_{it}^{A} + \beta_S S_{it} + \beta X_{it} + \gamma_i + \delta_t + \epsilon_{it} \tag{5}$$

where γ_i and δ_t represent user and time fixed effects respectively for user i and day t.

These specifications estimate the effect of (1) the total amount of readership and (2) the breadth/diversity of readership, in the mobile app channel on readership in the browser channel. Readership is heavily impacted by whether the user is registered or subscribed. Subscribers have unlimited access to content and have signaled appreciation for the quality of journalism of the newspaper by paying a subscription fee, so they consume more content than registered users. We control for this by including the time-varying subscription status of user i, S_{it} in our specification. The time-varying user-specific covariate X_{it} is the number of the remaining free articles on the browser in this time period. Person fixed effects difference out individual specific factors and time fixed effects control for temporal variation, trends due to seasonality and that popularity of certain kinds of content with varying coverage during news-cycles, which could significantly impact readership.

5.2 Subscription Rates

Next, we causally estimate the impact of a change in digital paywall settings on subscription rates. There are two main mechanisms by which the changes in quantity and diversity of content available via digital paywall can affect the subscription rates. First, after the users have exhausted their limit of free articles, they see pop-ups urging them to become a subscriber in order to continue reading. The annoyance cost associated with seeing the pop-up several times and the inability to enjoy the quality journalism of the newspaper could cause the users to subscribe. Second, the ability to consume more diverse content could change the user's prior expectations about the match of the newspaper's content with their preferences and entice them to subscribe.

In the specification below (Equation 6), the variable $Popups_{it}$ represents the number of times user i tries to access more than the allowed limit of free articles and sees pop-ups in period t. Note that the users could only see pop-ups after the paywall change that constricted the quantity of content was in effect i.e. time-period after T_2 in Figure 2.

We estimate the following specification with the user-day panel that we constructed:

$$S_{it} = \alpha + \beta_A Popups_{it} + \beta_D D^A_{it} + \beta X_{it} + \gamma_i + \delta_t + \epsilon_{it} \tag{6}$$

where γ_i and δ_t represent user and time fixed effects respectively for user i and day t.

This specification estimates the effect of (1) readership and (2) the diversity of readership on subscription status (S_{it} 0=registered/1=subscribed) of user i at time t. The time-varying user-specific covariate X_{it} is whether they received a gift subscription during this time period. Person fixed effects difference out individual specific factors and time fixed effects control for temporal variation which could be driving subscription rates.

5.3 Heterogeneous Effects: Tablet users

The wide proliferation of mobile devices has allowed content to be consumed via multiple channels like smartphones and tablets. These channels differ vastly in their form factors, user experiences and prevalence of use. Several authors have already documented the more immersive experience provided by tablets and its impact on online retailing and news consumption (Xu et al., 2015; Kumar et al., 2013).

Here we investigate if our estimates of cross-channel demand and the propensity to subscribe are different for the sub-population of readers who use a tablet to consume content and therefore estimate the following specification, with other estimation details remaining the same:

$$
\begin{aligned}
R^B_{it} &= \alpha + \beta_A R^A_{it} + \beta_{ATab} R^A_{it} \times \mathbb{1}(Tablet)_i + \beta_D D^A_{it} + \beta_{DTab} D^A_{it} \times \mathbb{1}(Tablet)_i + \beta_S S_{it} + \beta X_{it} \\
&+ \gamma_i + \delta_t + \epsilon_{it} \\
S_{it} &= \alpha + \beta_A Popups_{it} + \beta_{ATab} Popups_{it} \times \mathbb{1}(Tablet)_i + \beta_D D^A_{it} + \beta_{DTab} D^A_{it} \times \mathbb{1}(Tablet)_i + \beta X_{it} \\
&+ \gamma_i + \delta_t + \epsilon_{it}
\end{aligned}
$$

5.4 Heterogeneous Effects: Heavy Content Consumers

"Heavy" content consumers are interesting because these users will likely be the most directly impacted by the change in paywall, especially by the decrease in the number of articles available. It's unclear what their new behavior will be, as they can substitute use of the mobile app with use of the browser or they can complement it by also reducing their browser consumption. In addition, they can also convert from a registered user to subscribed user and continue consuming similar amounts of content as before the paywall change.

So, we investigate if there is a systematic difference in the readership and subscription patterns of the sub-population of users who are heavy users of the mobile app. We operationalize this as users who read more than the average number of articles read by our sample population.

We estimate the following specifications:

$$
\begin{aligned}
R^B_{it} &= \alpha + \beta_A R^A_{it} + \beta_{AH} R^A_{it} \times \mathbb{1}(Heavy - App - User)_i + \beta_D D^A_{it} + \beta_{DH} D^A_{it} \times \mathbb{1}(Heavy - App - User)_i \\
&+ \beta_S S_{it} + \beta X_{it} + \gamma_i + \delta_t + \epsilon_{it} \\
S_{it} &= \alpha + \beta_A Popups_{it} + \beta_{AH} Popups_{it} \times \mathbb{1}(Heavy - App - User)_i + \beta_D D^A_{it} \\
&+ \beta_{DH} D^A_{it} \times \mathbb{1}(Heavy - App - User)_i + \beta X_{it} + \gamma_i + \delta_t + \epsilon_{it}
\end{aligned}
$$

5.5 Identification

There are two main sources of endogeneity in our setting. First, several of the variables of interest (R^A_{it}, D^A_{it} & $Popups_{it}$) are jointly determined creating simultaneity. Second, correlated unobservables, which affect consumption on both the mobile app and the browser, pose a threat to identification and make the subscription status variable S_{it} in Equation 5 endogenous. Any unobservable time shock, for instance owing to the popularity of the news, which increases or decreases the readership of a subset of readers on both the browser and the mobile app, is a potential source of endogeneity. We overcome these threats to causal identification by using the quasi-experimental variation described in the previous section to derive instruments for the readership, diversity and subscription status variables.

Instruments for Readership and Diversity: We use paywall changes in the quantity and diversity of content available to non-subscribers to derive our instruments. To tease apart the impact of the change in quantity (from unlimited access to 3 articles per day) and the change in diversity (from access to Top News and Video to access to all the sections), we decompose the change into two phases—one in which only the quantity of content changed (P1 (Figure 2 (a))) and the other in which only the breadth/diversity of available content changed (P2 (Figure 2 (a))).

The paywall change in the mobile app was rolled out as part of an update available to download on day 88 of our observation period. However, not every user in our sample downloaded the updated version of the mobile app on the same day that it became available as shown in Figure 3(a). Variation in when users used their phones, used the app and agreed to update prompts creates exogenous variation in when users updated their apps. This differential updating pro-

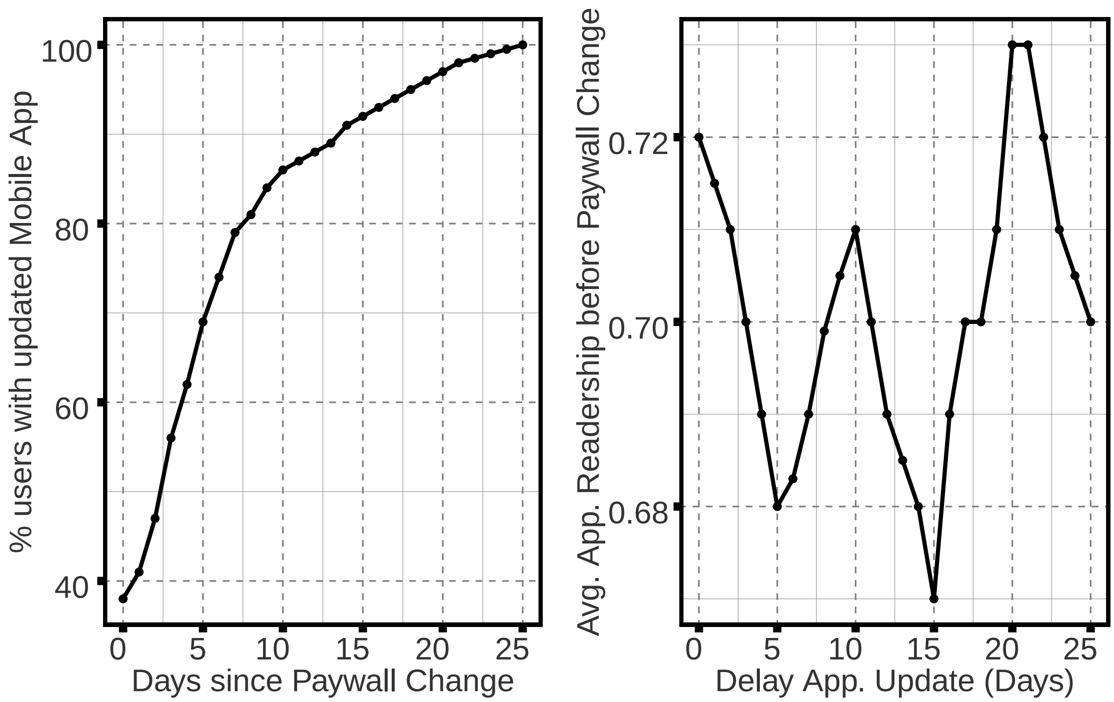

Figure 3: (a). Differential updating of the mobile app by users. (b). Levels of mobile app activity prior to the paywall change and the delay in updating the mobile app.

vides useful variation across users and time. Due to the possibility of ignoring this variation across users for the registered users who receive the treatment (paywall change), we chose an instrumental variable based identification strategy as opposed to a regression discontinuity design.

Although the differential updating of the mobile app provides variation across users, it could be that users endogenously choose when to update the app, for example, early updaters might be frequent users of the app, in which case there could be non random selection into exposure to the instrument. However, as shown in Figure 3(b), there is no systematic pattern in users' delays in updating the app and their readership activity prior to the paywall change, indicating that there is no systematic correlation between when users choose to update the app and their consumption behavior.

The instruments for the endogenous variables R^A_{it}, $Popups_{it}$ are operationalized as a binary indicator variable $\mathbb{1}(\text{Quantity})_{it}$ and the instruments for the endogenous variable D^A_{it} is operationalized as a binary indicator variable $\mathbb{1}(\text{Diversity})_{it}$. After the paywall change, the readers see pop-ups urging them to subscribe as a result of constriction of the quantity of the content available via the paywall from all the content to 3 articles a day. Hence $\mathbb{1}(\text{Quantity})_{it}$ also serves as an instrument for the variable $Popups_{it}$ in our specification for subscription status (Equation 6).

Note that in our specifications for estimating heterogeneous effects in Section 5.3 & 5.4, the interaction terms of the endogenous variables e.g. $R^A_{it} \times \mathbb{1}(Tablet)_i$ are also endogenous, so as is standard practice, their instruments are derived by interacting the original instrument with the relevant heterogeneity, for instance $\mathbb{1}(Quantity)_{it} \times \mathbb{1}(Tablet)_i$ is used as an instrument for $R^A_{it} \times \mathbb{1}(Tablet)_i$.

Instrument for Subscription Status: NYT allows people to gift subscriptions to their friends/relatives via their website (`nytimes.com`) as well as via third-party vendors—LivingSocial (`livingsocial.com`), Groupon (`groupon.com`) & Fab (`fab.com`). Gift subscriptions are one-time subscriptions and are typically for a duration of 3 months, 6 months, or an year.

Gifts provide an exogenous shock to the subscription status and hence are correlated with the variable S_{it} since it flips to 1 on the day that the gift subscription is activated.

The exclusion restriction holds since the timing of gifts is close to random and there is an element of surprise in gifts, so the propensity to gift is not correlated with the error term[7]. Further, as shown in Figure 4, there is a time-lag between the dates that the gift subscription was purchased and when it was activated, however there is no systematic pattern between the propensity of readers to activate their gift subscription and their past readership (as can be seen in Figure 4(b)). Hence, we use gifts as an instrument for the subscription status. The instrumental variable $Gift_{it}$ is coded as a binary variable, which is 1 on the day that the gift subscription was purchased.

In our data we have a total of 11423 gift subscriptions, out of which 3664 were purchased via LivingSocial, 6 via Groupon, 22 via Fab and the remaining 7731 were purchased organically from the NYT website. In the results section we show that gifts are a strong instrument. Table 3 summarizes the list of all the instruments.

[7]There is some chance of violation of the exclusion restriction if avid NYT readers ask their friends/relatives to gift them a subscription, however we expect such cases to be isolated and hence won't bias our parameter estimates significantly.

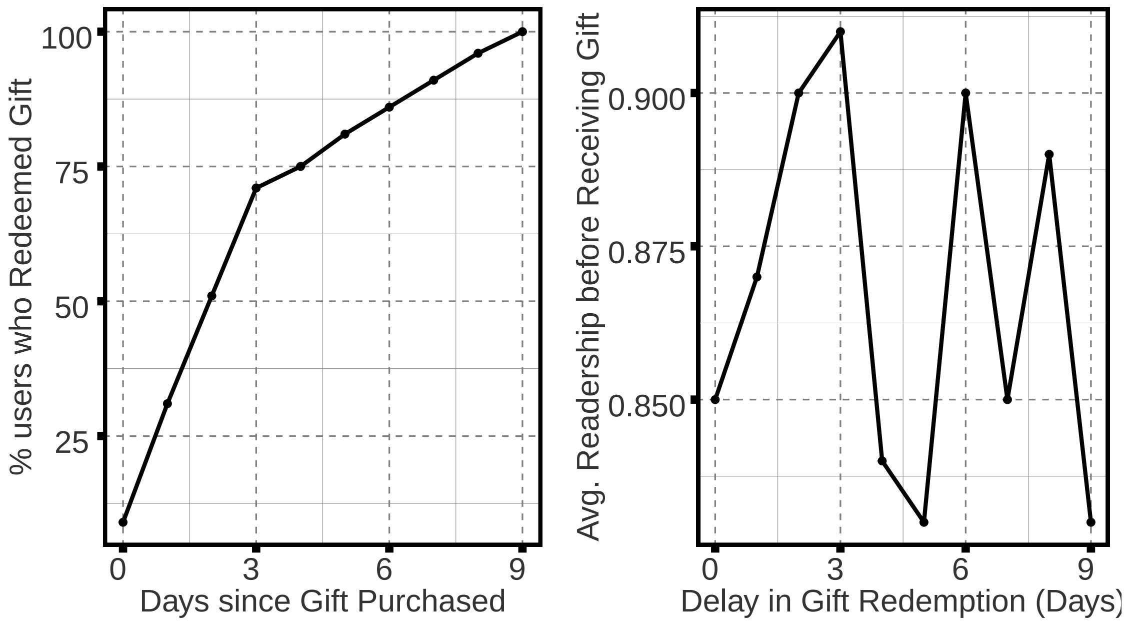

Figure 4: (a). Differential redemption of the gift subscription. (b). Readership prior to receiving the gift subscription.

Variable	Instrument
For specification in Equation 5	
R^A_{it}	$\mathbb{1}(\text{Quantity})_{it}$
D^A_{it}	$\mathbb{1}(\text{Diversity})_{it}$
S_{it}	$Gift_{it}$
For specification in Equation 6	
$Popups_{it}$	$\mathbb{1}(\text{Quantity})_{it}$
D^A_{it}	$\mathbb{1}(\text{Diversity})_{it}$

Table 3: List of instruments used for each endogenous variable. *Note:* 1). $\mathbb{1}(\cdot)$ indicates binary instrument.

5.6 Estimation

Standard linear econometric theory hinges on several normality assumptions (Angrist and Pischke, 2008). However, our readership and diversity variables are heavy tailed counts (Zhang et al., 2014)—most readers do not read anything on most days and read lots of articles on certain days. So, we adopt a standard procedure used in econometrics to address this problem—we log-transform the readership and diversity counts as $\log(R^{\{\cdot\}}+1)$. This has implications for the interpretation of our regression coefficients. Owing to the log-transformation of both the dependent and independent variables, the regression coefficients are interpreted as a 1% increase in the independent variable leading to a β% increase in the dependent variable.

In some of our specifications, the dependent variable is binary e.g. subscription status (S_{it}). In such cases, the preferred specification is either a logit or probit. However, estimating such models with millions of individual and time fixed effects is extremely challenging for any software and takes extremely long to converge to the correct solution. Several researchers have proposed using standard linear probability models (LPM) for the estimation of such specifications (Goldfarb and Tucker, 2011; Angrist and Pischke, 2008), especially if the distribution of the binary indicator variable is not extremely skewed e.g. 99% 0s and 1% 1s (which could lead to the predicted probabilities lying outside the [0, 1] interval). In our case, all of the predicted probabilities lie between 0 and 1, and therefore LPM with robust standard errors will yield unbiased and consistent estimates (Horrace and Oaxaca, 2006). Hence, we estimate our specifications with the binary dependent variable using LPM. Note that since in this case only the independent variables are log-transformed, the regression coefficients are interpreted as a 1% increase in the independent variable leading to a $\beta \times log(\frac{1.01}{1})$ increase in the dependent variable. We estimate all of our instrumental variable specifications using two stage least squares (2SLS) (Angrist and Pischke, 2008).

6 Results

6.1 Model Free Evidence

Figure 5 shows the impact of the paywall change on all the key economic variables—R^A_{it}, D^A_{it} and R^B_{it} for the entire panel of users. There is a slight decrease in the readership on the mobile app R^A_{it} and a slight increase in the diversity of content consumed. The reason for only this mild change is that the majority of the users in our panel are subscribers and hence their consumption behavior will not be impacted by the change in paywall. The paywall change affects only the registered users and some subscribers.[8] So, in order to get a closer look at the magnitude of the impact of the paywall, we need to observe the behavior of the registered users.

Figure 6 shows the same plot but only for the registered users. As expected, there is a steep

[8]The paywall might affect subscribers who only have the web+smartphone or web+tablet subscription bundles, in case they consume content via tablet or smartphone respectively. The paywall may also affect the subscribers who churn and are hence only reduced to the status of registered users.

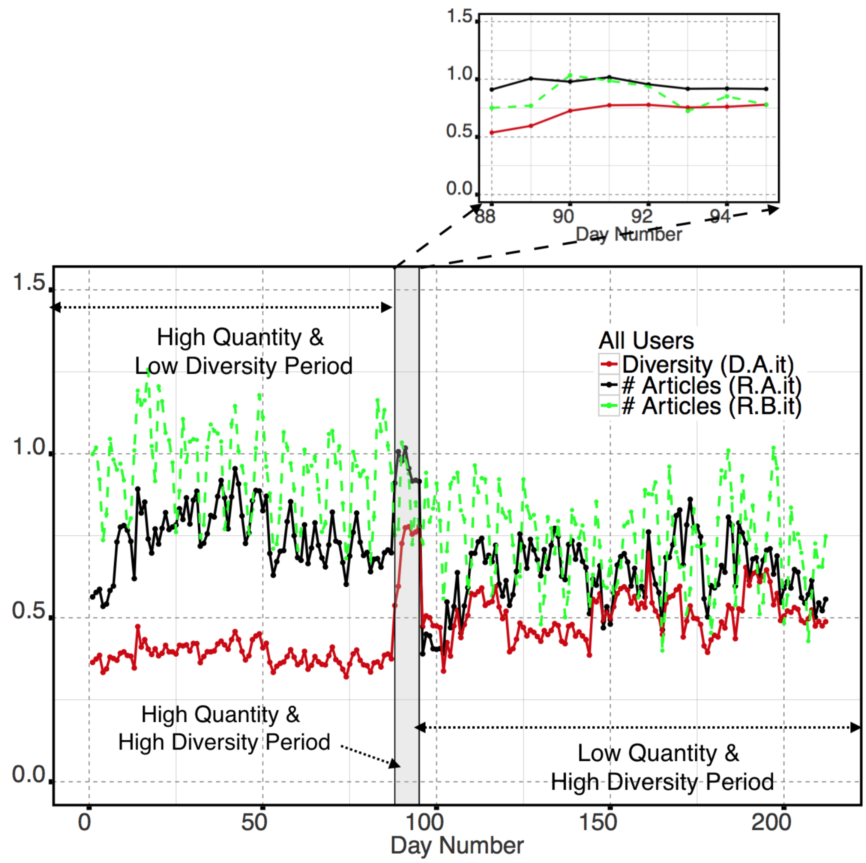

Figure 5: (Best viewed in color) Average number of articles read and the sections from which they were consumed for the entire panel i.e. registered users, subscribers and switchers (who changed their subscription status during the observation period). Note: 1). "High Quantity"= Total Content Published, "High Diversity"= All Sections, "Low Quantity"= 3 articles per day, "Low Diversity"= Top News and Video sections. 2). For ease of display, this plot only shows the users who updated the app on day 1.

decrease in the readership on the mobile app R_{it}^{A}. The average diversity of content consumed D_{it}^{A} only increased slightly after the change (in absolute terms), but it is a much larger change relatively, since now far fewer articles are being read in the first place. Also, we can see a dip in the content consumption on the browser R_{it}^{B} which suggests an impact of the paywall change on cross-channel content consumption. It is hard to discern if this change is driven by the change in quantity or diversity or both and we need econometric estimation to quantify those effects.

Figure 7 shows the average number of articles read by users that are not from the Top News and Video sections, and gives us a better picture of the increased variety of the content being consumed. Before the paywall change readers only had access to the Top News and Video sections, hence the number of articles read from other sections was zero. However, given the

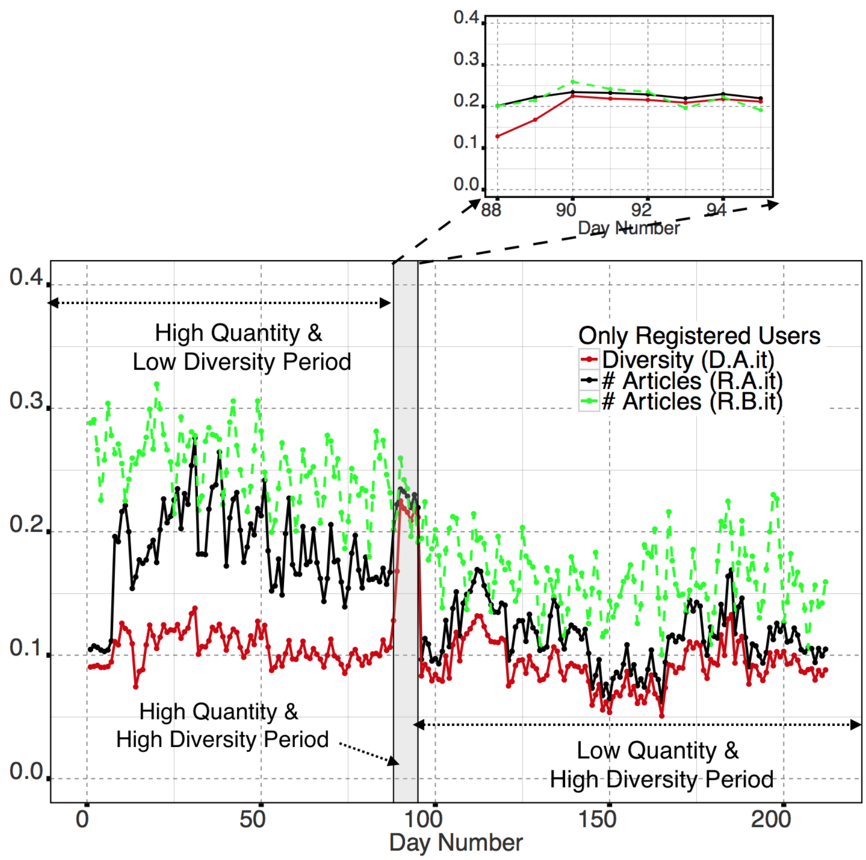

Figure 6: (Best viewed in color) Average number of articles read and the sections from which they were consumed for the registered users i.e. users who stayed registered throughout the observation period. Note: 1). "High Quantity"= Total Content Published, "High Diversity"= All Sections, "Low Quantity"= 3 articles per day, "Low Diversity"= Top News and Video sections. 2). For ease of display, this plot only shows the users who updated the app on day 1.

opportunity to consume diverse content, the users consumed more diverse content.

Table 4 shows the number of registered and subscribed users *one day before* the paywall change came into effect and on the *last day* of the observation period. Clearly, we see an economically significant reduction in the number of registered users and a corresponding increase in the number of subscribed users. We can not, as yet, attach a causal interpretation to this model free evidence or quantify how much of this increased subscriber base is caused by the paywall change.

6.2 Cross-Channel Demand

Baseline Regression Specification: We estimate the specification in Equation 1 via OLS. The results presented in Table 5 indicate that after controlling for user and day level fixed effects,

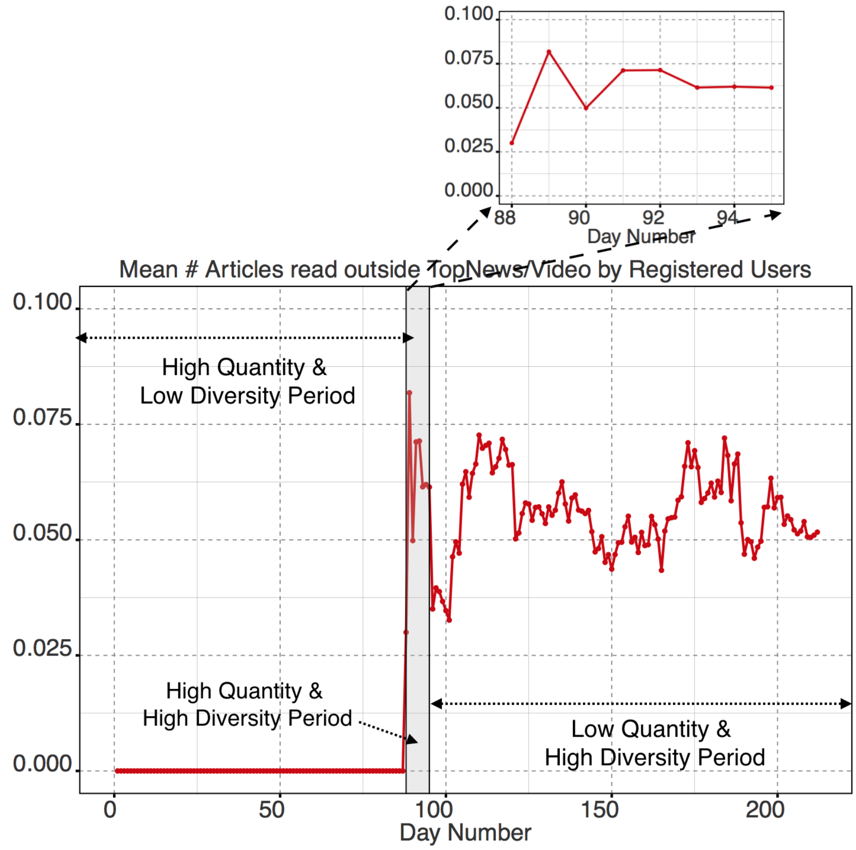

Figure 7: Plot showing increased variety of content consumed after the paywall change. Note: 1). "High Quantity"= Total Content Published, "High Diversity"= All Sections, "Low Quantity"= 3 articles per day, "Low Diversity"= Top News and Video sections. 2). For ease of display, this plot only shows the users who updated the app on day 1.

there is a positive relationship between increases in readership in the mobile app channel and readership in the browser channel. The results also indicate that an increase in the diversity of content available in the mobile app channel is associated with an increase in the readership in the browser channel. Also, not surprisingly, we see that being a subscriber is associated with higher readership (a positive coefficient on the S_{it} variable). The baseline results suggest that there is indeed a relationship between the readership on the mobile app and the browser channel and that there are potentially interesting dynamics of the impact of diversity on readership.

Causal Estimation using Instrumental Variables: Next we turn to estimating the specification in Equation 1 by instrumenting the endogenous variables R_{it}^{A} and D_{it}^{A} with the paywall changes in quantity and diversity respectively. The variable S_{it} is instrumented by gift subscriptions given to the registered users.

Variable	Before Paywall Change	After Paywall Change
# Registered Users	180,074	144,584
# Subscribed Users	628,711	664,201

Table 4: Table showing the change in subscriptions before (1 day prior) and after (last day of observation period) the paywall change was implemented. Note: We can not show the entire time-series of subscriptions owing to privacy concerns.

	Model 1	Model 2	Model 3
R_{it}^{A}	.012***	.010***	.009***
	(.000)	(.000)	(.000)
D_{it}^{A}	-	.007***	.005***
		(.000)	(.000)
S_{it}	-	-	.070***
			(.000)
User Fixed Effects	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes
R^2	.52	.55	.56
F-statistic	8.3×10^5	7.3×10^5	6.9×10^5

Table 5: OLS Regressions with dependent variable R_{it}^{B} **p < 0.05,***p < 0.01. Results are computed for a panel of (users) n=808,785, (days) t=212 resulting in a total of 171,462,420 user-day observations. *Note:* 1). Standard errors (shown in parenthesis) are clustered at the level of users. 2.) All the readership and diversity variables are log-transformed.

Instrumental variable (IV) estimation hinges on two key assumptions (Angrist and Pischke, 2008). First, the instrument(s) should be correlated with the endogenous variable and second, the instrument(s) should have no direct effect on the dependent variable (also known as the *exclusion restriction*). It is straightforward to see that these assumptions are satisfied for our specifications as the paywall changes are supply-side shocks that affected the readership and diversity only in the mobile app channel.

The results for IV estimation are shown in Table 6 and the magnitudes of effects are both economically and statistically significant. The instruments are strong, as can be seen by the Cragg-Donald Wald Statistic (Cragg and Donald, 1993; Stock and Yogo, 2005). We notice that a 1% increase in the readership on the mobile app R_{it}^{A} causes approximately 0.168% increase in the readership on the browser R_{it}^{B}. This suggests a complementary relationship between the browser and the mobile app channel as readers might start reading an article on one channel and continue on the other.

Interestingly, the results also indicate that the availability of diverse content on the mobile app decreases readership on the browser (a 1% increase in diversity on the app causing approximately 0.371% decrease in readership on the browser). An increase in the diversity of content available on the mobile app causes a decrease in readership on the browser.

Dependent Variable →	R_{it}^{B}
R_{it}^{A}	.168***
	(.016)
D_{it}^{A}	-.371***
	(.046)
S_{it}	.071***
	(.014)
User Fixed Effects	Yes
Time Fixed Effects	Yes
F-statistic	1.2×10^{7}
Weak identification Test:	
(Cragg-Donald Wald Statistic/p-value)	$.39 \times 10^{3}/.000$

Table 6: Instrumental Variable (IV) Regressions estimated using 2SLS **$p < 0.05$,***$p < 0.01$. Results are computed for a panel of (users) n=808,785, (days) t=212 resulting in a total of 171,462,420 user-day observations. *Note:* 1). Standard errors (SEs) are clustered at the level of users. SEs are computed using Bartlett kernel of bandwidth 3 and are robust to heteroskedasticity and autocorrelation 2.) All the readership and diversity variables are log-transformed.

The Mobile Browser vs The Mobile App: With increasing interest in the design of mobile content consumption platforms (e.g. apps and mobile web browsers) (Evans, 2016), it has become important for publishers to know whether they are in synergy or if one is cannibalizing the other. Until now, we have defined and operationalized the readership on the browser R_{it}^{B} by the total number of articles read by a user on a given day using the desktop and mobile browsers. So, next we define our browser readership using just the readership on the mobile browser R_{it}^{MobWeb}. Everything else, including all the estimation details, stay the same as the specification in Equation 1.

The results are shown in Table 7 and we find substitution between the mobile browser and the mobile app. However, when there is an increase in the diversity of content available on the mobile app, it increases readership on the mobile browser.

A potential explanation for this substitution effect is that people have limited time budget while consuming media content (Deleersnyder et al., 2002), which is even more salient in the case of mobile devices, owing to the "dynamism" associated with their usage. However, when more diverse content is available, we see increased usage of the browser. Potentially, more diverse content has the effect of engaging users, which makes them explore that topic via the browser channel as well.

Dependent Variable →	R_{it}^{MobWeb}
R_{it}^{A}	-.180***
	(.006)
D_{it}^{A}	.465***
	(.019)
S_{it}	-.024***
	(.005)
User Fixed Effects	Yes
Time Fixed Effects	Yes
F-statistic	7.8×10^4
Weak identification Test:	
(Cragg-Donald Wald Statistic/p-value)	$.39 \times 10^3/.000$

Table 7: Instrumental Variable (IV) Regressions estimated using 2SLS **$p < 0.05$,***$p < 0.01$. Results are computed for a panel of (users) n=808,785, (days) t=212 resulting in a total of 171,462,420 user-day observations. *Note:* 1). Standard errors (SEs) are clustered at the level of users. SEs are computed using Bartlett kernel of bandwidth 3 and are robust to heteroskedasticity and autocorrelation 2.) All the readership and diversity variables are log-transformed.

6.3 Subscription Rates

Baseline Regression Specification: The results presented in Table 8 (estimation of the specification in Equation 2 via OLS) indicate that a 1% increase in the readership on the mobile app (and hence a 1% increase in the number of pop-up messages urging users to subscribe, assuming a fixed total readership) is correlated with a decrease in the likelihood of the average user to subscribe by ($-.001 \times log(1.01) \approx -4 \times 10^{-6}$). However, an increase in diversity of content consumption on the mobile app is correlated with an increase in the probability of becoming a subscriber.

Causal Estimation using Instrumental Variables: Since the readership and diversity variables R_{it}^{A}, D_{it}^{A} are endogenous, we instrument them as earlier by the paywall changes in quantity and diversity respectively. The IV estimation of the impact of readership and diversity on the subscription rates allow us to attach causal interpretations to our estimates of the impact of those variables on subscription rates. Also, since the readership and diversity on the mobile app are

	Model 1	Model 2	Model 3
$Popups_{it}$	-.001***	-.001***	-.001***
	(.000)	(.000)	(.000)
D_{it}^{A}	-	.018***	.016***
	-	(.000)	(.000)
$Gift_{it}$	-	-	-.318***
	-	-	(.003)
User Fixed Effects	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes
R^2	.02	.03	.03
F-statistic	1.0×10^5	1.1×10^5	4.7×10^5

Table 8: OLS Regressions with dependent variable S_{it} **p < 0.05,***p < 0.01. Results are computed for a panel of (users) n=808,785, (days) t=212 resulting in a total of 171,462,420 user-day observations. *Note:* 1). Standard errors (shown in parenthesis) are clustered at the level of users. 2.) All the readership and diversity variables are log-transformed.

instrumented by the paywall changes in quantity and diversity, it allows us to estimate the causal impact of paywall changes on subscription rates, assuming that the impact of paywall changes on subscription rates is only mediated by readership and the diversity of the content consumption.

The results of our IV estimation are shown in Table 9 indicate that both the changes in the paywall—the reduction in number of free articles to only 3/day and an increase in the diversity of content offered on the mobile app—had an economically and statistically significant positive causal impact on subscription rates. We notice that a 1% increase in the number of pop-ups on the mobile app $Popups_{it}$ and a 1% increase in the diversity of content consumption on the mobile app D_{it}^{A} both increase the probability of subscribing by approximately .016.

Back-of-the-envelope calculations show that these paywall changes causally impacted between 28019 and 28337 i.e. approximately 78-80% of the total 35490 subscriptions (see Table 4. 180,074-144,584=35490) that happened during the observation period[9].

Converting Partial Subscribers to Full Subscribers: In the specification described in Equation 2, we defined the subscription status S_{it} as a binary indicator variable indicating whether the reader is either registered or subscribed. However, the NYT has several subscription bundles

[9]We see in the data that the average number of articles read by the registered users after the paywall dropped from 0.18 to 0.14, which is a 22.2% drop. Moreover, we notice that for every 3 less articles read, on average the readers saw 1 pop-up (since not every registered user tried to read more than 3 articles), so there was a (22.2/3)% increase in pop-ups. Similarly, the average diversity of the consumed content increased from 0.12 to 0.125 i.e. approximately 4% increase. Combining this with the fact that there were 180,074 registered users one day before the introduction of the paywall and estimates from Table 9, we can estimate the paywall changes to causally impact between 28019 and 28337 subscriptions in our observation period.

Dependent Variable →	S_{it}
$Popups_{it}$	1.64***
	(.012)
D^{A}_{it}	1.43***
	(.011)
$Gift_{it}$	.318***
	(.015)
User Fixed Effects	Yes
Time Fixed Effects	Yes
F-statistic	8.8×10^3
Weak identification Test:	
(Cragg-Donald Wald Statistic/p-value)	$8.7 \times 10^3/.000$

Table 9: Instrumental Variable (IV) Regression estimated using 2SLS **p $<$ 0.05,***p $<$ 0.01. Results are computed for a panel of (users) n=808,785, (days) t=212 resulting in a total of 171,462,420 user-day observations. *Note:* 1). Standard errors (shown in parenthesis) are clustered at the level of users. 2.) All the readership and diversity variables are log-transformed.

(1) all digital (2) all digital and home delivery, (3) web and smartphone, (4) web and tablet. Subscribers with bundles (1) and (2) have unfettered access to all the content across all the digital channels (desktop, smartphone, tablet). However, subscribers with bundles (3) and (4) might not have full access across all the digital channels. For instance, subscribers with bundles (3) and (4) will be affected by the paywall changes should they access the content using a tablet and smartphone respectively.

In other words, in the previous specification the dependent variable measured the conversion of a registered user to a subscribed user with any of the four subscription bundles. Here we define a new specification, where we measure the conversion of registered users or partial subscribers (Bundles (3), (4)) to a full-subscriber (Bundles (1), (2)). The dependent variable (F_{it}) codes 0 as registered users or partial subscribers and 1 as full subscribers. All other details stay the same. This new specification is economically relevant to the NYT because the full subscription bundles (1), (2) are priced higher than bundles (3) and (4), and ultimately they want every subscriber to be a full subscriber, holding the costs of customer acquisition constant. We expect the results of this new specification to be similar to the earlier specification, but the magnitudes to be smaller as the paywall changes only affect some of the partial subscribers in some channels.

The results for IV estimation are shown in Table 10 and the magnitudes of the effects are again economically and statistically significant. As expected, the direction of effects is the same as earlier but the magnitude of the effects have been approximately halved for all coefficients.

Dependent Variable →	F_{it}
$Popups_{it}$	.787***
	(.006)
D^A_{it}	.775***
	(.006)
$Gift_{it}$	-.014***
	(.007)
User Fixed Effects	Yes
Time Fixed Effects	Yes
F-statistic	9.5×10^3
Weak identification Test:	
(Cragg-Donald Wald Statistic/p-value)	8.7×10^3

Table 10: Instrumental Variable (IV) Regression estimated using 2SLS **$p < 0.05$,***$p < 0.01$. Results are computed for a panel of (users) n=808,785, (days) t=212 resulting in a total of 171,462,420 user-day observations. *Note:* 1). Standard errors (SEs) are clustered at the level of users. SEs are computed using Bartlett kernel of bandwidth 3 and are robust to heteroskedasticity and autocorrelation 2.) All the readership and diversity variables are log-transformed.

6.4 Cross-Channel Demand and Subscription Rates of Tablet users.

Our causal estimates of average treatment effects show that there is complementarity between the readership on the mobile app and the browser and that an increase in diversity of the content available on the mobile app steals readership from the browser. Here we estimate the treatment effects on the sub-population of readers who consume the content via tablets.

Results shown in Table 11 indicate that a 1% increase in the readership by the tablet users on the mobile app decreases the readership on the browser by .061% relative to non-tablet users and that a 1% increase in the diversity on the mobile app decreases the content consumption on the browser by .062% relative to non-tablet users.

The results highlight an economically significant finding of increased substitution between the readership on the mobile app and browser by the tablet users. In addition, we also see an accentuation of the impact of diversity on the mobile app on the readership on the browser—with the availability of more diverse content, the tablet users decrease their readership on the browser significantly compared to the non-users. This finding is similar to a finding by (Xu et al., 2015) of substitution between the tablet and desktop channels, however, in our case the browser consumption can also be on a mobile device via a mobile browser (in our data, 3.8% of total browser visits were mobile) and is not solely confined to desktop.

The results with the subscription status S_{it} (Table 12) as the dependent variable show that the tablet users are more likely to be impacted by the change in quantity and diversity of the content available on the mobile app and become subscribed users. A 1% increase in the readership on mobile app increases the probability of tablet users subscribing by .019 and a 1% increase in the diversity on mobile app increases the probability of tablet users subscribing by .039.

The mechanism behind our findings could be that tablets have bigger screens and a more immersive user interface than smartphones, so the users might prefer not to switch to a different channel. Decrease in quantity of content available may simply have the annoyance effect which might make the users subscribe. The increased propensity of tablet users to subscribe after more diverse content is available might be attributed to the ability of the users to explore new type of content and coming to appreciate the journalism of the newspaper and hence subscribing.

Dependent Variable $\rightarrow$	R_{it}^B
R_{it}^A	.186***
	(.021)
$R_{it}^A \times \mathbb{1}(Tablet)_i$	-.061***
	(.010)
D_{it}^A	-.253***
	(.050)
$D_{it}^A \times \mathbb{1}(Tablet)_i$	-.062***
	(.014)
S_{it}	.057***
	(.013)
User Fixed Effects	Yes
Time Fixed Effects	Yes
F-statistic	1×10^7
Weak identification Test:	
(Cragg-Donald Wald Statistic/p-value)	$.2 \times 10^3/.000$

Table 11: Instrumental Variable (IV) Regression estimated using 2SLS **p < 0.05,***p < 0.01. Results are computed for a panel of (users) n=808,785, (days) t=212 resulting in a total of 171,462,420 user-day observations. *Note:* 1). Standard errors (SEs) are clustered at the level of users. SEs are computed using Bartlett kernel of bandwidth 3 and are robust to heteroskedasticity and autocorrelation 2.) All the readership and diversity variables are log-transformed.

Dependent Variable $\rightarrow$	S_{it}
$Popups_{it}$	1.62***
	(.014)
$Popups_{it} \times \mathbb{1}(Tablet)_i$	.242***
	(.027)
D^A_{it}	2.47***
	(.024)
$D^A_{it} \times \mathbb{1}(Tablet)_i$	1.50***
	(.028)
User Fixed Effects	Yes
Time Fixed Effects	Yes
F-statistic	6.4×10^3
Weak identification Test:	
(Cragg-Donald Wald Statistic/p-value)	$3.2 \times 10^3/.000$

Table 12: Instrumental Variable (IV) Regression estimated using 2SLS **p < 0.05,***p < 0.01. Results are computed for a panel of (users) n=808,785, (days) t=212 resulting in a total of 171,462,420 user-day observations. *Note:* 1). Standard errors (SEs) are clustered at the level of users. SEs are computed using Bartlett kernel of bandwidth 3 and are robust to heteroskedasticity and autocorrelation 2.) All the readership and diversity variables are log-transformed.

6.5 Do "Heavy" Content Consumers behave differently?

The results in Table 13 indicate that the degree of complementarity exhibited by the heavy users of mobile app is relatively lower than the light users by more than six times (.989% vs .150%). A potential explanation for this behavior is that the heavy users of the mobile app predominantly consume their content by mobile devices—smartphone and tablet, and hence make less use of browsers overall.

Also, as can be seen from our estimation with the subscription status S_{it} as the dependent variable (Table 14), the heavy app users are less likely to subscribe compared to the light users, as they are perhaps heavy category consumers and hence substitute NYT with other news outlets. A 1% increase in the number of pop-ups on mobile app decreases the probability of heavy users subscribing by .002 compared to the light users.

6.6 Robustness Checks

We check the robustness of our estimates in several ways. First, we consider alternative ways of measuring readership other than the number of articles (stories) read. Second, we conduct falsification tests by examining endogenous variables in the future, when they should have no

Dependent Variable →	R^B_{it}
R^A_{it}	.989***
	(.048)
$R^A_{it} \times \mathbb{1}(Heavy - App - User)_i$	-.839***
	(.035)
D^A_{it}	1.49***
	(.401)
$D^A_{it} \times \mathbb{1}(Heavy - App - User)_i$	1.00***
	(.354)
S_{it}	.257***
	(.078)
User Fixed Effects	Yes
Time Fixed Effects	Yes
F-statistic	1.7×10^5
Weak identification Test:	
(Cragg-Donald Wald Statistic/p-value)	14.19/.000

Table 13: Instrumental Variable (IV) Regression estimated using 2SLS **$p < 0.05$,***$p < 0.01$. Results are computed for a panel of (users) n=808,785, (days) t=212 resulting in a total of 171,462,420 user-day observations. *Note:* 1). Standard errors (SEs) are clustered at the level of users. SEs are computed using Bartlett kernel of bandwidth 3 and are robust to heteroskedasticity and autocorrelation 2.) All the readership and diversity variables are log-transformed.

causal effect on demand or subscription rates. Third, we consider alternate functional forms of our specification, in particular poisson and negative binomial count models for the instrumental variable regression.

Alternate Definition of Readership Variables: In our analysis we have operationalized the readership variables R^A_{it} and R^B_{it} by the number of articles read by the users. Another related measure of readership or engagement can be the number of visits or clicks (V^A_{it}, V^B_{it}) made by the reader on the newspaper's website. More precisely, these count the number of clicks the reader made on the front-page of the newspaper or browsing section-fronts and is strictly greater than or equal to the corresponding readership variable $R^{A/B}_{it}$. It might help to think of $V^{A/B}_{it}$ as a noisy version of the corresponding readership variable.

We re-estimate the results in Table 6 with every other aspect of estimation being the same as earlier, except replacing the readership variables by the visits/clicks variables. As can be seen from the estimates in Table 15, the estimates of all the variables are comparable in magnitude and sign and hence in economic significance.

Falsification Test: In the previous section we gave evidence that suggests exogeneity of our

Dependent Variable $\rightarrow$	S_{it}
$Popups_{it}$	1.23***
	(.269)
$Popups_{it} \times \mathbb{1}(Heavy - App - User)_i$	-.208***
	(.070)
D^A_{it}	1.13***
	(.194)
$D^A_{it} \times \mathbb{1}(Heavy - App - User)_i$	-.701***
	(.207)
User Fixed Effects	Yes
Time Fixed Effects	Yes
F-statistic	430.8
Weak identification Test:	
(Cragg-Donald Wald Statistic/p-value)	117.85/.000

Table 14: Instrumental Variable (IV) Regression estimated using 2SLS **p < 0.05,***p < 0.01. Results are computed for a panel of (users) n=808,785, (days) t=212 resulting in a total of 171,462,420 user-day observations. *Note:* 1). Standard errors (SEs) are clustered at the level of users. SEs are computed using Bartlett kernel of bandwidth 3 and are robust to heteroskedasticity and autocorrelation 2.) All the readership and diversity variables are log-transformed.

instruments, in particular that there is no systematic relationship between the propensity of the users to update the mobile app and their readership behavior. In order to alleviate any remaining concerns about the efficacy of our procedure, we perform a falsification test. We estimate the specification described in Equation 7, where $T_1 = 88$ i.e. the day on which the paywall change was first rolled out (cf. Figure 3(b)). Essentially, we have shifted the endogenous variables to a period in the future after the paywall change, and if our procedure is correct, there should be no relationship between the explanatory variables and the dependent variable.

$$R^B_{it} = \alpha + \beta_A R^A_{it+T_1} + \beta_D D^A_{it+T_1} + \beta_S S_{it+T_1} + \beta X_{it} + \gamma_i + \delta_t + \epsilon_{it} \tag{7}$$

We perform the estimation as earlier using 2SLS, except that the endogenous variables and the instruments are shifted in time. Results in Table 16 show that there is no significant relationship between readership on the browser and the endogenous explanatory variables—$R^A_{it+T_1}, D^A_{it+T_1}, S_{it+T_1}$.

Poisson and Negative Binomial Model Specifications: For the instrumental variable (IV) estimations we used the standard 2SLS estimation procedure with the log-transformed readership and diversity variables. Another way of modeling count variables is using poisson and negative binomial regressions. Poisson regressions model count variables but the Poisson distribution is

Dependent Variable →	V_{it}^{B}
V_{it}^{A}	.259***
	(.027)
D_{it}^{A}	-.418***
	(.114)
S_{it}	.967***
	(.028)
User Fixed Effects	Yes
Time Fixed Effects	Yes
F-statistic	8.3×10^5
Weak identification Test:	
(Cragg-Donald Wald Statistic/p-value)	1.7×10^3/.000

Table 15: Instrumental Variable (IV) Regression estimated using 2SLS **$p < 0.05$,***$p < 0.01$. Results are computed for a panel of (users) n=808,785, (days) t=212 resulting in a total of 171,462,420 user-day observations. *Note:* 1). Standard errors (SEs) are clustered at the level of users. SEs are computed using Bartlett kernel of bandwidth 3 and are robust to heteroskedasticity and autocorrelation 2.) All the readership and diversity variables are log-transformed.

equidispersed—its mean is equal to its variance—which is a limiting assumption in our case, as the readership variables are overdispersed—have higher variance compared to their means—and hence have a long-tail. So, we also check the robustness of our estimates by estimating a negative binomial regression.

The estimation was performed using the control function approach (Wooldridge, 2015) and the standard errors were obtained via bootstrapping (1000 samples)[10]. Table 17 shows the estimation results. As can be seen, the impacts of the various variables have the same signs and statistical significance as in Table 6. Note that a direct comparison of the magnitudes of the coefficients in Poisson and Negative binomial regression with OLS is not straightforward owing to the non-linearity of these estimation procedures.

7 Discussion

We conducted the largest ever study of the impact of paywall design changes in the (1) quantity of free articles and (2) breadth of sections from which those articles can be read on cross-channel demand and subscription rates using micro-level user activity data. Changes to the quantity and

[10]We performed the two-stage procedure, where in the first stage we regressed the instruments on the endogenous explanatory variables, and in the second stage we estimated a non-linear (Poisson or negative binomial) regression which also included the residuals as well as the endogenous variables from the first stage.

Dependent Variable $\rightarrow$	R^B_{it}
$R^A_{it+T_1}$	.421*
	(.11)
$D^A_{it+T_1}$	.207
	(.22)
S_{it+T_1}	.691*
	(.17)
User Fixed Effects	Yes
Time Fixed Effects	Yes
R^2	.27
F-statistic	4.8×10^6
Weak identification Test:	
(Cragg-Donald Wald Statistic/p-value)	2.74/.09

Table 16: Instrumental Variable (IV) Regression for the Falsification Test estimated using 2SLS *p $<$ 0.10,**p $<$ 0.05,***p $<$ 0.01. Results are computed for a panel of (users) n=808,785, (days) t=212 resulting in a total of 171,462,420 user-day observations. *Note:* 1). Standard errors (SEs) are clustered at the level of users. SEs are computed using Bartlett kernel of bandwidth 3 and are robust to heteroskedasticity and autocorrelation 2.) All the readership and diversity variables are log-transformed.

breadth of free content available via the paywall on the mobile app provided exogenous shocks to the readership variables for identification.

Results suggest the existence of a highly nuanced connection between the mobile app and the browser. An increase in readership on the mobile app leads to a corresponding increase in readership on the browser, suggesting complementarity. However, an increase in diversity on the mobile app leads to a decrease in the readership on the browser, suggesting diversity has a substitutive effect across channels. This effect could be attributed to more diverse content leading to a better match with consumer preferences, and hence higher engagement, which increases the cost of switching channels. The complementarity between the channels is significantly less for heavy mobile app users and tablet users actually substitute between the channels. The impact of an increase in the diversity on the mobile app stealing readership from the browser is even more pronounced for tablet users.

This is also the first paper to causally establish and estimate the impact of two important parameters of digital paywall design—quantity and diversity—on the probability of converting a registered user into a subscriber. Assuming that the effect of paywall changes on subscription rates is solely mediated by readership, our results suggest that a decrease in the number of free

Dependent Variable →	Poisson: R^B_{it}	Negative Binomial: R^B_{it}
R^A_{it}	.234***	.317***
	(.011)	(.031)
D^A_{it}	-.691***	-.572***
	(.003)	(.056)
S_{it}	.872***	.346***
	(.000)	(.002)
User Fixed Effects	Yes	Yes
Time Fixed Effects	Yes	Yes
Log Likelihood	-0.17×10^7	-7.21×10^8
Wald χ^2	2.31×10^8	6.45×10^7

Table 17: Poisson and Negative Binomial Instrumental Variable (IV) Regressions estimated using the control function approach **$p < 0.05$,***$p < 0.01$. Results are computed for a panel of (Users) n=808,785, (Time Period) t=212 resulting in a total of 171,462,420 User-Time observations. *Note:* 1). Standard errors (shown in parenthesis) are clustered at the level of users.

articles (and hence an increase in the number of pop-ups urging to subscribe) and an increase in the diversity (number of sections from which those articles can be consumed), causally increases the probability of a registered user becoming a subscriber. Back-of-the-envelope calculations show that the paywall changes in quantity and diversity causally impacted 78-80% of the subscriptions that happened during the observation period. These results are both statistically and economically significantly. Tablet users are more likely to be driven by the paywall to convert to subscribers. However, the heavy users are less likely to converted to subscribers by the paywall.

A behavioral story consistent with our findings is that due to the ever-increasing penetration of mobile devices, especially among the youth, they are becoming the preferred mode of content consumption compared to desktop computers. This paradigm shift is due to the "dynamism" offered by the mobile devices—the ability to consume content on-the-go—for example, while waiting for the server to serve a latte at Starbucks or while standing in a crowded subway train. However, not all mobile devices are created equal and serve the same purpose or people. Smartphones are more portable owing to their small size and their small screen sizes can create cognitive difficulties in consuming content (Ghose et al., 2012), hence people use them to skim through news articles but actually complete the reading in a more "static" environment of a home desktop computer, leading to a complementarity effect. On the other hand, tablets provide a more relaxed and immersive environment for readers and are less portable, hence users prefer them over desktops for content consumption, creating a substitution effect.

Heavy users of the mobile app are less likely, whereas the users who own a tablet are more likely to convert to subscribers compared to the average user. One might have expected the heavy mobile app users to have higher propensity to subscribe since they are likely to see more pop-ups urging them to subscribe. This counter-intuitive finding could be explained by the fact that heavy users of NYT are in general heavy category users i.e. they consume lots of news from a variety of sources. So, with the introduction of the paywall change that led to constriction in the quantity of content accessible, they switched to other news outlets. Tablet users could have increased propensity to subscribe owing to the immersive experience of the tablets which make these users value the brand experience of the newspaper more than other users, or it could also simply be explained by the fact that tablet owners are usually high-earners and might have more disposable income[11].

7.1 Implications for Paywall Design

The particular paywall change discussed in the paper was successful for the NYT based on the data. Overall, it did decrease the total number of article impressions on both the mobile app and the browser by a combined amount of 0.26 articles per individual per day which amounts to a total decrease of $\approx$ 4.45 million impressions during our observation period. A crude back-of-the-envelope calculation assuming two advertisements per page and the average rate of \$10.50 as the CPM[12] suggests a loss of around $\$940,000$ in digital ad-revenue during our observation period. However, as we saw in the results section, the paywall change causally led to an increase of about 28000 subscribers. Combining this with the average cost of a subscription bundle of around \$80, it amounts to a revenue of about \$2.24 million. So, the paywall introduction had a healthy net-positive impact on the bottom-line of NYT as the loss in ad-revenue was more than offset by the increase in the revenue from subscriptions.

Based on our results, we suggest that the newspapers in general should not focus on the short-term objective of maximizing the ad-revenue as they face severe competition from Google and Facebook to attract the ad money. Rather, they should strive to convert the online visitors to paid subscribers by offering differentiated content modulated via digital paywalls. As we saw in

[11] `http://www.marketingcharts.com/online/us-tablet-ownership-update-january-2014-39508/`
[12] http://www.nytimes.com/marketing/selfservice/help.html

this paper, the digital paywalls which match their free content offerings to readers' preferences by letting them choose the content they want to consume are effective in increasing the newspapers' subscription bases.

7.2 Limitations and Future Research

Although our work improves our understanding of several underpinnings of a modern day newspaper's digital strategy, it is not without limitations. First, we only explored two of the design parameters of the digital paywall, namely quantity and diversity. In order to fully navigate the strategic landscape, it is important to understand the trade-offs associated with other paywall design choices e.g. archival, curation access. Second, in our analysis we ignored the anonymous users and considered only the registered and subscribed users, as it is hard to reliably track those users across devices. It would be interesting to perform a study which uses more sophisticated digital fingerprinting to examine these users. Third, our results are for a relatively small window of time (about 7 months), almost equally split before and after the paywall change. As part of future work, it will be interesting to quantify the long-term impacts of paywall design changes on readership and subscriptions. It will also be interesting to see if the patterns of complementarity between the mobile app and browser channels persist over time. We hope our work will inspire future research to overcome these limitations in pushing the limits of our understanding of the relationship between digital paywall design and multi-channel content consumption.

References

Joseph Alba, John Lynch, Barton Weitz, Chris Janiszewski, Richard Lutz, Alan Sawyer, and Stacy Wood. Interactive home shopping: consumer, retailer, and manufacturer incentives to participate in electronic marketplaces. *The Journal of Marketing*, pages 38–53, 1997.

Joshua D Angrist and Jörn-Steffen Pischke. *Mostly harmless econometrics: An empiricist's companion*. Princeton university press, 2008.

Asim Ansari, Carl F Mela, and Scott A Neslin. Customer channel migration. *Journal of Marketing Research*, 45(1):60–76, 2008.

Kenneth Arrow. Economic welfare and the allocation of resources for invention. In *The rate and direction of inventive activity: Economic and social factors*, pages 609–626. Princeton University Press, 1962.

Susan Athey, Emilio Calvano, and Joshua S Gans. The impact of the internet on advertising markets for news media. *Rotman School of Management Working Paper*, (2180851), 2014.

Susan Athey, Markus M Mobius, and Jenő Pál. The impact of aggregators on internet news consumption. 2017.

Jill Avery, Thomas J Steenburgh, John Deighton, and Mary Caravella. Adding bricks to clicks: Predicting the patterns of cross-channel elasticities over time. *Journal of Marketing*, 76(3):96–111, 2012.

Youngsok Bang, Dong-Joo Lee, Kunsoo Han, Minha Hwang, and Jae-Hyeon Ahn. Channel capabilities, product characteristics, and the impacts of mobile channel introduction. *Journal of Management Information Systems*, 30(2):101–126, 2013.

Andre Boik, Shane Greenstein, and Jeffrey Prince. The empirical economics of online attention. 2016.

Erik Brynjolfsson, Yu Hu, and Mohammad S Rahman. Battle of the retail channels: How product selection and geography drive cross-channel competition. *Management Science*, 55(11):1755–1765, 2009.

Ramon Casadesus-Masanell and Feng Zhu. Strategies to fight ad-sponsored rivals. *Management Science*, 56(9):1484–1499, 2010.

Lesley Chiou and Catherine Tucker. Paywalls and the demand for news. *Information Economics and Policy*, 25(2):61–69, 2013.

John G Cragg and Stephen G Donald. Testing identifiability and specification in instrumental variable models. *Econometric Theory*, 9(02):222–240, 1993.

Barbara Deleersnyder, Inge Geyskens, Katrijn Gielens, and Marnik G Dekimpe. How cannibalistic is the internet channel? a study of the newspaper industry in the united kingdom and the netherlands. *International Journal of Research in Marketing*, 19(4):337–348, 2002.

Glenn Ellison and Sara Fisher Ellison. Internet retail demand: Taxes, geography, and online-offline competition. Technical report, National Bureau of Economic Research, 2006.

Ben Evans. Mobile ate the world. 2016. URL `http://www.slideshare.net/a16z/mobile-is-eating-the-world-2016`.

Chris Forman, Anindya Ghose, and Avi Goldfarb. Competition between local and electronic markets: How the benefit of buying online depends on where you live. *Management Science*, 55 (1):47–57, 2009.

Inge Geyskens, Katrijn Gielens, and Marnik G Dekimpe. The market valuation of internet channel additions. *Journal of Marketing*, 66(2):102–119, 2002.

A Ghose, S Han, and S Park. Analyzing the interdependence between web and mobile advertising: a randomized field experiment. Technical report, Working paper, 2013.

Anindya Ghose, Avi Goldfarb, and Sang Pil Han. How is the mobile internet different? search costs and local activities. *Information Systems Research*, 24(3):613–631, 2012.

Avi Goldfarb and Catherine E Tucker. Privacy regulation and online advertising. *Management Science*, 57(1):57–71, 2011.

Austan Goolsbee. Competition in the computer industry: Online versus retail. *The Journal of Industrial Economics*, 49(4):487–499, 2001.

William C Horrace and Ronald L Oaxaca. Results on the bias and inconsistency of ordinary least squares for the linear probability model. *Economics Letters*, 90(3):321–327, 2006.

Mathew Ingram. Here's what the new york times' 1 million subscriber number means. 2015. URL `http://fortune.com/2015/08/06/new-york-times-paywall/`.

Jasper Jackson. Digital media strategies 2013: The rise and rise of the paywall. 2013. URL `https://www.themediabriefing.com/article/digital-media-strategies-rise-of-paywalls`.

Vineet Kumar, Bharat Anand, Sunil Gupta, and Felix Oberholzer-Gee. The new york times paywall. *Harvard Business School Case*, 512-077, January 2013.

Anja Lambrecht and Kanishka Misra. Fee or free: When should firms charge for online content? *Available at SSRN 2307961*, 2015.

Lionbridge-Technologies-Inc. Mobile web apps vs. mobile native apps: How to make the right choice. 2012. URL http://www.lionbridge.com/files/2012/11/Lionbridge-WP_MobileApps2.pdf.

Tommi Mikkonen and Antero Taivalsaari. Apps vs. open web: The battle of the decade. In *Proceedings of the 2nd Workshop on Software Engineering for Mobile Application Development*, pages 22–26, 2011.

Jeffrey T Prince. The beginning of online/retail competition and its origins: An application to personal computers. *International Journal of Industrial Organization*, 25(1):139–156, 2007.

N. S. Ramnath. The real reason why myntra moved to an app-only strategy. 2015. URL http://www.foundingfuel.com/slideshow/the-real-reason-why-myntra-moved-to-an-apponly-strategy.

Jason MT Roos and Ron Shachar. When kerry met sally: Politics and perceptions in the demand for movies. *Management Science*, 60(7):1617–1631, 2014.

Jason MT Roos, Carl F Mela, and Ron Shachar. The effect of links and excerpts on internet news consumption. *Available at SSRN 2678938*, 2015.

Carl Shapiro and Hal R Varian. *Information rules: a strategic guide to the network economy*. Harvard Business Press, 2013.

James H Stock and Motohiro Yogo. Testing for weak instruments in linear iv regression. *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg*, 2005.

Jason Summerfield. Mobile website vs. mobile app: Which is best for your organization? 2011. URL http://www.hswsolutions.com/services/mobile-web-development/mobile-website-vs-apps/.

Joseph Tartakoff. Taking the plunge: How newspaper sites that charge are faring. 2009. URL http://www.cbsnews.com/news/paidcontent-taking-the-plunge-how-newspaper-sites-that-charge-are-faring/.

Shan Wang. Tribune publishing will be moving all of its newspapers to a metered paywall model next year. 2015. URL http://www.niemanlab.org/2015/11/tribune-publishing-will-be-moving-all-of-its-newspapers-to-a-metered-paywall-model-next-year/.

Jeffrey M Wooldridge. Control function methods in applied econometrics. *Journal of Human Resources*, 50(2):420–445, 2015.

Jiao Xu, Chris Forman, Jun B Kim, and Koert Van Ittersum. News media channels: complements or substitutes? evidence from mobile phone usage. *Journal of Marketing*, 78(4):97–112, 2014.

Kaiquan Xu, Jason Chan, Anindya Ghose, and Sangpil Han. Battle of the channels: The impact of tablets on digital commerce. *Management Science, Forthcoming*, 2015.

Yao Zhang, Eric T Bradlow, and Dylan S Small. Predicting customer value using clumpiness: From rfm to rfmc. *Marketing Science*, 34(2):195–208, 2014.