HARVARD | BUSINESS | SCHOOL

The Comprehensive Effects of a Digital Paywall Sales Strategy

Doug J. Chung
Ho Kim
Reo Song

Working Paper 19-118

The Comprehensive Effects of a Digital Paywall Sales Strategy

Doug J. Chung
Harvard Business School

Ho Kim
University of Missouri

Reo Song
California State University

Working Paper 19-118

The Comprehensive Effects of a Digital Paywall Sales Strategy

Doug J. Chung, Harvard University

Ho Kim, University of Missouri, St. Louis

Reo Song, California State University, Long Beach*

ABSTRACT

This paper explores the multiple and comprehensive effects of a digital paywall sales strategy, an increasingly common means of go-to-market for media firms. Specifically, we examine the effects of a digital paywall on a media firm's two sources of income—subscription and advertising—across its two channels—traditional and digital. We compile a unique data set from multiple sources that contain detailed data on 79 major U.S. print media firms; and, for causal inference, we utilize a synthetic control method to distinguish the true effect from naturally occurring time trends. In addition, we take into account demand spillover—substitution vs. complementarity—across channels, as well as factors that moderate such spillover effects. We find that, although heterogeneous across media firms, a paywall sales strategy can lead to positive demand substitution from digital to traditional channels, especially for firms with large circulation and uniqueness of content. Furthermore, uniqueness of content reduces the decline in digital demand, moderating the loss in digital advertising revenue while increasing digital subscription revenue. Overall, the effectiveness of a digital paywall varies by both the source and the channel of income across media firms with different characteristics.

Keywords: sales strategy, digital paywall, media industry, newspaper, demand substitution, spillover effect, synthetic control.

* All authors contributed equally, and their names are listed in alphabetical order.

1. Introduction

The Internet offers abundant free resources. Users watch free videos on YouTube, share pictures for free on Instagram, read online news for free at news aggregation sites, and enjoy free online games. It is not surprising that the Internet has long been considered free. Thus, when media firms (e.g., newspapers) built websites to share their articles, most allowed free access for Internet users, anticipating that free online content would generate a source of new revenue from online digital advertising.

Decades after the rollout of free websites, media firms have started to reconsider their initial digital sales strategies. For example, *The New York Times* successfully introduced a paid online content model, or a "digital paywall," in March 2011. Since then, many media firms, particularly those traditionally focused on print content (newspaper firms), have followed suit, adopting a pay-for-online-content model. As of 2015, 78% of U.S. print media firms[1] with a daily circulation of over 50,000 were using a digital subscription model (Williams, 2016). This research aims to examine the comprehensive effect—on subscriptions[2] and advertising for both traditional[3] and digital channels, respectively—of a digital paywall sales strategy[4] for media outlets.

The main motivation behind U.S. print media firms' transition to a digital paywall sales strategy can be explained by the consistent and rapid decline in print circulation. The estimated total U.S. daily print circulation in 1990 was 62 million and 63 million for weekdays and Sundays,

[1] We refer to media outlets traditionally focused on print-version newspapers as print media firms.
[2] We use the term subscription or circulation to refer to paid media content—i.e., the purchase of a digital subscription or a traditional print-version newspaper. For print versions, strictly speaking, circulation is typically larger than the number of subscriptions, as it includes newsstand and kiosk sales in addition to home-delivery (subscription) sales. However, the difference is minimal, as most sales come from subscriptions. Hence, hereafter, we use both terms interchangeably.
[3] We interchangeably use the terms traditional and print to refer to a non-digital channel.
[4] We define a go-to-market strategy involving the use of a pay-for-view pricing scheme for a firm's online content as the "digital paywall sales strategy" because the pricing scheme of digital content (paywall) directly and indirectly affects the sales outcome (from subscriptions and advertising) of both the digital and traditional channels.

respectively. By 2017, daily circulation had decreased to 31 million and 34 million for weekdays and Sundays, respectively (Pew Research Center, 2017). Coinciding with print circulation, advertising has also declined. Publicly traded U.S. print media firms experienced an average decrease of 9.0% per annum in print advertising revenue from 2010 to 2017. Although digital advertising revenue makes up an increasingly large portion of total advertising revenue, the amount is still relatively small compared to that of print advertising. Digital advertising revenue accounted for only 29% of the total advertising revenue of publicly traded U.S. print media firms in 2017. Furthermore, the growth in digital advertising revenue has not offset the decline in traditional advertising revenue, as the U.S. newspaper industry's total advertising revenue fell from $26 billion in 2010 to $18 billion in 2017 (Pew Research Center, 2017).

Hence, it is not surprising that many traditional print media firms have pursued a digital paywall sales strategy to seek a new source of revenue: digital subscription revenue. However, such a shift in strategy can alter existing revenue streams across different channels. In this research, we investigate the comprehensive effects of a paywall sales strategy in the U.S. newspaper industry. Specifically, we examine the effects of a digital paywall on two sources of revenue—subscription and advertising—across the two channels of distribution—print and digital. In doing so, we explore demand spillovers—substitution vs. complementarity—across channels with the adoption of a paywall, as well as factors that moderate such spillover effects. Combined, we seek to understand the joint elements that affect the success of a digital paywall sales strategy. To the best of our knowledge, this research is among the first to comprehensively examine all sources of revenue across both the digital and traditional channels with the adoption of a digital paywall.

A digital paywall sales strategy, by definition, creates a new source of income—digital subscription revenue. In addition to digital subscription revenue, the adoption of a paywall may

induce demand spillover onto print (Anand, 2016; Pattabhiramaiah, Sriram, & Manchanda 2017). If the print and digital versions of a newspaper are substitutes, we can expect the spillover to be positive (Gentzkow, 2007). In contrast, if they are complements, we can expect the spillover to be negative (Xu et al., 2014; Koukova, Kannan, & Ratchford, 2008). The spillover effect likely differs across heterogeneous media firms, and, thus, the direction and magnitude of the spillover are empirical questions.

In addition to revenue sourced from subscriptions, advertising revenue from both print and digital channels is an important source of income for media firms. The adoption of a paywall would likely decrease digital advertising revenue because of the reduction in website traffic, although the magnitude can be either minimal (Pauwels & Weiss, 2008) or significant (Chiou & Tucker, 2013). Print advertising revenue would likely depend on the magnitude and direction of the demand spillover. **Table 1** shows the expected directional effects of revenue source by channel with the adoption of a digital paywall.

To examine the comprehensive effects of a digital paywall sales strategy for traditional print media firms, we compile a unique data set from multiple sources. We collect the print circulation of the top 79 U.S. print media firms and their print and digital subscription prices over a ten-year period (from October 2007 to September 2017). We also compile the corresponding firms' monthly web traffic in terms of pageviews and paywall timing. In addition, we collect firm characteristics, including their size, political views, content uniqueness, and consumer (reader) demographics. Furthermore, we acquire both print and digital advertising revenues from Nielsen to compute advertising rates for each channel.

For causal inference, we utilize the synthetic control method (Abadie & Gardeazabal, 2003; Abadie, Diamond, & Hainmueller, 2010; Tirunillai and Tellis 2017), a generalized version of

difference-in-differences (DID), to control for heterogeneous time trends by different media firms. For each paywalled media firm (43 firms), the synthetic control constructs an artificial, non-paywalled firm (from a pool of 36 firms that did not change their paywall status during the analysis period) that is deemed similar to the focal paywalled firm. We compare the print circulation and website pageviews of a focal paywalled firm to those of its synthetic non-paywalled counterpart to estimate the demand for content in traditional and digital channels, respectively. Furthermore, we take into account a firm's pricing decision that can simultaneously occur with the adoption of a paywall to distinguish the effect of paywall from that of price. In addition, we compare the types of consumers—and, thus, their price sensitivity—pre- and post-paywall to infer the counterfactual print subscription price had a firm not adopted a digital paywall. As mentioned, a print media firm's revenue source typically consists of four components (see **Table 1**)—that is, subscription and advertising revenue by channel, print and digital. Using the model estimates and the print and digital advertising rates, we evaluate the effect of a paywall by revenue source and channel.

The results show that the effectiveness of a paywall varies according to a media firm's characteristics and content. A firm's uniqueness of content positively affects spillover to print demand and alleviates the decrease in digital demand. In terms of firm revenue, we find that large media firms (e.g., *The New York Times*, *Chicago Tribune*, and *Los Angeles Times*) significantly benefit, whereas small-sized firms (e.g., the *Sun-Sentinel*, *Chicago Sun-Times*, and *The Denver Post*) do not benefit by pursuing a digital paywall sales strategy. In addition, the increase/decrease in revenue source (subscription and advertising) by channel (pring and digital) differs substantially by firm. For example, the success of the digital paywall for *The New York Times* comes not only from an increase in digital subscriptions, but also from an increase in print subscriptions and advertising because of the positive spillover to print demand; this more than offsets the negative

effect of a paywall on digital advertising due to a decrease in web traffic. In contrast, the success of the digital paywall for the *Chicago Tribune* comes largely from an increase in print subscriptions and advertising (because of the positive spillover) and less so from an increase in digital subscriptions.

The noteworthy success of these media firms contrasts with the results of other print media firms for which the decrease in digital advertising revenue largely offsets the positive effect of a digital paywall. For many media firms in our sample, the net effect of a paywall in the digital channel is negative, as the newly generated digital subscription revenue is offset by a significant decrease in digital advertising revenue. Hence, were it not for the positive spillover effect onto print subscription and advertising, the net effect of a paywall would be negative. This fact shows the importance of understanding the comprehensive effects of a digital paywall sales strategy, beyond the direct effect on digital subscription and advertising revenues.

The rest of the paper is organized as follows. Section 2 explains the data used for our empirical analyses, and Section 3 presents the model and estimation methodology. Sections 4 and 5 discuss the results and counterfactual analyses, respectively. Section 6 concludes.

2. Data

From multiple sources, we compile a unique and comprehensive dataset for our empirical analyses. The data include U.S. print media firms' print circulation, digital pageviews, print and digital subscription prices, firm characteristics, and consumer (reader) demographics. Print circulation and subscription prices (for both print and digital) are collected from publishers' statements reported to the Alliance for Audited Media (AAM). We obtain the pageviews from Alexa Internet, a media intelligence subsidiary of Amazon. The temporal frequency (time interval) on pageviews

is monthly, whereas that of print circulation and prices is semiannual, because the AAM requires print media firms to report information only in March and September of each year.

The analysis period spans ten years, from October 2007 to September 2017.[5] During this period, 52 of the top 100 U.S. print media firms (in terms of print circulation) introduced and maintained a paywall.[6] Among this set, nine firms had missing data. Thus, for reliable analysis, we turn our attention to the 43 paywalled print media firms without data issues. To apply the synthetic control method for causal inference, we need a set of firms whose paywall status did not change during the analysis period. There are 45 such firms in the list of top 100 U.S. print media firms. Among this set, nine firms are discarded due to missing data. Therefore, our data set consists of 79 U.S. print media firms: 43 paywalled firms whose paywall effects we analyze; and 36[7] firms from which a synthetic control firm is constructed for each of the 43 paywalled firms.

Figure 1a shows the average (daily) print circulation and the average (annual) print subscription price of the 79 print media firms over the analysis period. The average print circulation decreased from 275,594 in March 2008 to 106,525 in September 2017, while the average print subscription price increased from $205 to $573 during the same period. **Figure 1b** shows the average (monthly) pageviews per million internet users and the average (annual) digital subscription price from January 2010 to September 2017. During that period, the average pageviews decreased from 49,198 in January 2010 to 14,920 in September 2017, while the average digital subscription price increased from $1.93 to $78.[8]

[5] The data on pageviews span only seven years, from 2010 to 2017, as data are available starting only from 2010.
[6] Three print media firms introduced but then canceled a paywall during this period.
[7] The donor pool consists of firms that have not changed paywall stratus during our analysis period: 2007 – 2017. Hence, the donor pool includes not only non-paywalled firms but also those that have adopted a paywall before 2007. There are three such firms: *The Wall Street Journal*, *Albuquerque Journal*, and *Arkansas Democrat-Gazette*.
[8] Note that the average digital subscription price is computed over all 79 firms. If a firm does not have a paywall in a certain period, its digital subscription price is set to be zero. As such, the average digital subscription price of paywalled firms is higher than the average price shown in **Figure 1b**.

Table 2 shows the descriptive statistics of firm characteristics and reader demographics across the 79 print media firms. We utilize a number of variables to capture content characteristics. The political slant index (Gentzkow & Shapiro, 2010) captures the firms' degree of political conservativeness. For example, *The Wall Street Journal*'s slant index is 0.4829, and the paper is perceived to be more conservative than *The New York Times*, which has a slant index of 0.4271. The uniqueness index (Kim, Song, & Kim, 2018) captures the proportion of exclusive content, with a higher index indicating higher exclusive content. In addition, we capture firms' reputation and brand equity using their size in terms of print circulation measured in 2010, the most recent year before any of the firms in our data introduced a paywall. We obtain reader demographics data from Nielsen-Scarborough surveys that include distributions of age, income, and educational attainment of each media firm's reader base.

3. Model

3.1 Base Framework

Obtaining an accurate effect of a paywall for a media firm (outlet) is a causal inference problem in which there are two states (paywall adoption and no adoption) along with a corresponding sales outcome[9] (circulation and pageviews for print and digital channels, respectively) for each state. Because we do not observe the outcome for the other state that a specific firm is not in (i.e., we do not observe the outcome that a paywalled firm would have obtained if it had not adopted a paywall), we cannot directly identify the causal effect of a paywall. In such a setting, the *potential outcome* approach has been widely used as a framework for analyzing causal effects (Athey & Imbens, 2017).

[9] Although pageviews are not directly related to sales (subscriptions), they affect a firm's advertising revenue, and, thus, we use the term sales outcome (or simply outcome) to refer to either print circulation or website pageviews in our empirical context.

For clarity, we first define the notations and then specify the model. Let $y_{it}^{(0)}$ be the outcome of media firm i in period t if the firm does not have a paywall in that period. That is, $y_{it}^{(0)}$ is the outcome of firm i in period t that corresponds to the (control) state: without a paywall. Similarly, let $y_{it}^{(1)}$ be the outcome of the same firm in the same period if it has a paywall. That is, $y_{it}^{(1)}$ is the outcome of the same firm in the same period that corresponds to the (treatment) state: with a paywall. The outcome of interest in our empirical setting is either print circulation or digital pageviews. Let the subscription price (either for the print or the digital channel, depending on the outcome of interest) of firm i in period t without and with a paywall be $p_{it}^{(0)}$ and $p_{it}^{(1)}$, respectively; and, without loss of generality, let the price without a paywall be proportional to the price with a paywall such that

$$p_{it}^{(0)} = \rho_{it} p_{it}^{(1)} = (\rho + \xi_{it}) p_{it}^{(1)},$$

where $E(\xi_{it}) = 0$.

Based on the above notation, we model the outcome of media firm i at time t, without and with a paywall, as

$$y_{it}^{(0)} = \alpha_i + \lambda_t + \beta^{(0)} p_{it}^{(0)} + \boldsymbol{\phi}_i' \mathbf{F}_t + \varepsilon_{it}, \tag{1.1}$$

$$y_{it}^{(1)} = \alpha_i + \lambda_t + \beta^{(1)} p_{it}^{(1)} + \delta_{it} + \boldsymbol{\phi}_i' \mathbf{F}_t + \varepsilon_{it}, \tag{1.2}$$

respectively, where α_i is the base outcome; λ_t is the aggregate time effect common across all firms; $\beta^{(0)}$ and $\beta^{(1)}$ are the price elasticity without and with a paywall, respectively; and δ_{it} is the effect of the paywall on the outcome of interest. The vector $\mathbf{F}_t$ represents any unobserved, potentially time-varying, external factors that affect outcome; and ε_{it} is the idiosyncratic error term. Regarding the aggregate time effect (λ_t) and exogenous factors ($\boldsymbol{\phi}_i'$), we adopt the commonly used

assumption in treatment methods[10] that these effects are the same across the two parallel worlds. Firm i may change its price with the introduction of a paywall; thus, the difference between the two outcomes ($y_{it}^{(1)} - y_{it}^{(0)}$) encompasses the effects of both paywall and price change that occur simultaneously with the adoption of a paywall. By differencing Equations (1.1) and (1.2), we obtain

$$\begin{aligned} y_{it}^{(1)} - y_{it}^{(0)} &= \beta^{(1)} p_{it}^{(1)} - \beta^{(0)} p_{it}^{(0)} + \delta_{it} \\ &= \left(\beta^{(1)} - \rho_{it}\beta^{(0)}\right) p_{it}^{(1)} + \delta_{it}. \end{aligned}$$

Note that the differencing removes the base outcome, the aggregate time effect, and any unobserved external factors.

Now suppose that firm i at time t has installed a paywall, and let y_{it} and p_{it} be the *observed* outcome and price, respectively. We observe $y_{it}^{(1)}$ and $p_{it}^{(1)}$ during the firm's paywall periods (i.e., $y_{it} = y_{it}^{(1)}$ and $p_{it} = p_{it}^{(1)}$) such that

$$\begin{aligned} y_{it} = y_{it}^{(1)} &= \alpha_i + \lambda_t + \beta^{(1)} p_{it}^{(1)} + \delta_{it} + \boldsymbol{\phi}_i' \mathbf{F}_t + \varepsilon_{it} \\ &= y_{it}^{(0)} + \left(\beta^{(1)} - \rho_{it}\beta^{(0)}\right) p_{it} + \delta_{it} \text{ for } t \geq \tau_i, \end{aligned} \tag{2}$$

where τ_i is the period in which firm i introduced its paywall. Note that $y_{it}^{(0)}$ is the *unobserved* counterfactual outcome for $t \geq \tau_i$. Based on Equation (2), we can identify the effect of the paywall δ_{it}, conditional on the counterfactual outcome. To compute the counterfactual outcome, we turn to the synthetic control approach, a generalized version of DID (Abadie, Diamond, & Hainmueller, 2010).

3.2 Synthetic Control Method

[10] Examples include DID, synthetic control, and propensity score matching.

Suppose that there are J media firms that have *not* introduced a paywall during the analysis period. These firms comprise the "donor pool" from which a synthetic clone is constructed for a paywalled firm. Specifically, for paywalled firm i, consider a $(J \times 1)$ vector of weights $\mathbf{w}_i = (w_{i1}, w_{i2}, \dots, w_{iJ})'$ such that $w_{ij} \geq 0$ for all j and $w_{i1} + w_{i2} + \cdots + w_{iJ} = 1$. The synthetic control method constructs a synthetic matching firm for a given paywalled firm by finding the vector of optimal weights $(w_{i1}^*, w_{i2}^*, \dots, w_{iJ}^*)$ that satisfies the following identity conditions:

(I-1) The time-constant characteristics of the focal firm (z_{ik} for characteristic $k = 1, 2, \dots, K$) are the same as the weighted characteristics of the control firms (i.e., $z_{ik} = \sum_{j=1}^{J} w_{ij}^* z_{jk}$ for all considered characteristics).

(I-2) The outcome of the focal firm during its pre-paywall period $(y_{i1}, y_{i2}, \dots, y_{i,\tau_i-1})$ is the same as the weighted outcome of the control firms during the same pre-paywall period $(y_{it} = \sum_{j=1}^{J} w_{ij}^* y_{jt}, for\ t < \tau_i)$.

Once the optimal weights are known, the counterfactual outcome of firm i in a post-paywall period t (i.e., the outcome that the firm would have obtained in a post-paywall period t if it had not adopted a paywall in τ_i) is simply the weighted average of the non-paywalled firms' outcome at t:

$$\hat{y}_{it}^{(0)} = \sum_{j=1}^{J} w_{ij}^* y_{jt}. \quad (3)$$

In typical empirical settings, there is often no set of weights $\mathbf{w}_i$ that exactly satisfies conditions (I-1) and (I-2). In such cases, the weights are selected by minimizing the distance between the synthetic control and the focal firm so that the identity conditions best hold. Let $\mathbf{x}_i$ be a vector of firm i's characteristics and its average pre-paywall sales outcome: $\mathbf{x}_i = (z_{i1}, z_{i2}, \dots, z_{iK}, \bar{y}_{i,\tau_i-1})'$, where $\bar{y}_{i,\tau_i-1} = \sum_{t=1}^{\tau_i} y_{it}$. For a non-paywalled firm j, let $\mathbf{x}_{0j}$ be the combined vector of j's characteristics and its average sales outcome until $\tau_i - 1$ (the pre-paywall

period of firm i): $\mathbf{x}_{0j} = \left(z_{j1}, z_{j2}, \ldots, z_{jK}, \bar{y}_{j,\tau_i-1}\right)'$, where $\bar{y}_{i,\tau_i-1} = \sum_{t=1}^{\tau_i} y_{it}$. Let $\mathbf{X}_0$ be a $(K+1) \times J$ sized matrix whose j'th column is $\mathbf{x}_{0j}$. For the best-fitting synthetic control, we choose $\mathbf{w}_i^*$, which minimizes the distance $\| \mathbf{x}_i - \mathbf{X}_0\mathbf{w}_i \|$, subject to a weight matrix $\mathbf{V}_i$ such that

$$\mathbf{w}_i^* = \underset{\mathbf{w}_i}{\operatorname{argmin}}\left(\sqrt{(\mathbf{x}_i - \mathbf{X}_0\mathbf{w}_i)'\mathbf{V}_i(\mathbf{x}_i - \mathbf{X}_0\mathbf{w}_i)}\right), \tag{4}$$

where the weight matrix $\mathbf{V}_i$ determines the relative importance of firm characteristics in constructing a synthetic control for firm i. Following Abadie and Gardeazabal (2003) and Abadie, Diamond, and Hainmueller (2010), we choose $\mathbf{V}_i$, which minimizes the mean squared prediction error of firm i's pre-paywall sales outcome.

The synthetic control method provides several advantages over the popular DID approach in our empirical setting. For DID to be valid, a researcher needs to identify a control group against which to evaluate the treatment group. Typically, the control group is arbitrarily hand-picked, and, thus, the DID approach works only when determining the control group is relatively straightforward and obvious (e.g., Chevalier & Mayzlin, 2006). This clearly is not the case in our empirical setting, where we have various firms whose control group is not well known *a priori*. Furthermore, the synthetic control method provides a data-driven approach in constructing a matching unit. More importantly, the synthetic control method relaxes a critical assumption of DID that the unobserved common factors are constant over time (i.e., $\mathbf{F}_t = \mathbf{F}$ in Equations (1.1) and (1.2)). Since our data span ten years, it is highly likely that the unobserved common factors vary over time.

3.3 Heterogeneous Effects of Paywall

To capture the heterogeneous effects across different media firms, we model the effect of paywall as a function of firm characteristics ($\mathbf{m}_i$) such that

(5) $$\delta_{it} = \gamma_0 + \boldsymbol{\gamma}'\mathbf{m}_i + c_i + \eta_{it},$$

where γ_0 is the common base effect; $\mathbf{m}_i$ is the vector of observed firm characteristics; and $\boldsymbol{\gamma}$ is the vector of corresponding parameters. The factor c_i represents any time-constant unobserved components that influence the effect of paywall, such as omitted firm characteristics. The time-varying unobserved term η_{it} captures any other idiosyncrasies and are assumed to be independently and identically distributed across firms and time with properties $E(\eta_{it}|\mathbf{m}_i, c_i) = 0$ and $Var(\eta_{it}|\mathbf{m}_i, c_i) = \sigma_\eta^2$. For our empirical application, we construct the vector $\mathbf{m}_i$ with the following variables: conservative slant; content uniqueness; and firm size in terms of the logarithm of the print circulation as of 2010[11] as a measure of reputation and brand equity as a media outlet.

The composite effect of the unobserved factor c_i includes a firm's omitted content characteristics, as well as the local market's willingness to accept a paywall. To the extent that these unobserved variables may be correlated with the observed variables ($\mathbf{m}_i$), the estimation results would be biased if the omitted variables were not properly controlled for. To mitigate this bias, we include proxy variables that are closely related to the omitted variables (Wooldridge, 2010). In this regard, firms' content characteristics and reader demographics can be a reasonable proxy for the omitted variables. First, media firms tailor language (Banville, 2016) and political slant (Gentzkow & Shapiro, 2010) to meet the needs of the local readers. Thus, various unobserved content characteristics may also vary across local markets, and variables that represent local market characteristics can explain the variation. Second, the acceptance of a paywall in a local market is likely to depend on local demographics such as income and education; thus, variables that represent local markets can also explain the variation in local readers' willingness to accept a paywall. Based on this rationale, we model c_i to be a function of reader demographics such that:

[11] We use the 2010 value because the firms in our sample started adopting paywalls in 2011.

(6) $$c_i = \boldsymbol{\theta}'\mathbf{d}_i + \omega_i \quad ,$$

where the vector $\mathbf{d}_i$ includes demographic characteristics—such as age, education, and income—of media firm i's focal market.

By inserting Equations (3), (5), and (6), Equation (2) simplifies to:

(7) $$y_{it} - \hat{y}_{it}^{(0)} = \gamma_0 + \beta p_{it} + \boldsymbol{\gamma}'\mathbf{m}_i + \boldsymbol{\theta}'\mathbf{d}_i + \varsigma_{it},$$

$$\beta = \beta^{(1)} - \rho\beta^{(0)},$$

$$\varsigma_{it} = \omega_i + \eta_{it} \, .$$

The composite error term ς_{it} is serially correlated within a media firm because of the existence of ω_i: $E(\varsigma_{it}\varsigma_{it-1}|\mathbf{m}_i, \mathbf{d}_i) = Var(\omega_i)$; however, it is contemporaneously uncorrelated across firms because the common time effect (λ_t), and unobserved external factors ($\boldsymbol{\phi}_i'\mathbf{F}_t$) are eliminated through the differencing of Equations (1.1) and (1.2).

Note that both the decision to adopt a paywall and prices are likely endogenous. That is, the unobserved component in Demand (ε_{it} in Equations (1.1) and (1.2)) is correlated with the firm's pricing and paywall decision. For example, suppose that there is a sudden increase in demand because the firm hired a renowned editor. Then, the firm would likely offer higher prices and consider the timing to be right for a digital paywall. Such endogeneity issues are taken care of by differencing Equations (1.1) and (1.2) to remove the unobserved components. However, if the effectiveness of paywall differs by unobserved firm characteristics (the ς_{it}s), and firms set prices accordingly, prices would still be endogenous in Equation (7). To address this endogeneity issue, we adopt an instrumental variable approach. We use the wage in media occupations as an instrumental variable for both print and digital prices in their respective demand equations. We obtain the wages (specifically for editors in media occupations: occupation code 27-3041) from the Bureau of Labor Statistics (BLS). Because the BLS surveys wages of various occupation

groups by different geographic locations, the instrumental variable for our empirical analysis varies across both firms and time. Thus, we account for both cross-sectional and cross-time variations in firm prices such that the first-stage predicted values vary across firms. The standard identification argument for the use of input costs applies—labor costs (specifically for editors) are related to prices but not directly related to demand. As the population's income likely affects demand, we include income among our demographic variables ($\mathbf{d}_i$) in Equation (7). Thus, our instrumental approach would rely on variations in the wage difference between editors and the general population. We estimate Equation (7) separately for the two outcome variables: subscriptions and pageviews for print and digital channels, respectively.[12]

4. Results

We first discuss the results of the synthetic control and then examine the results of print and digital demand.

4.1 Synthetic Control

For illustration of the synthetic control, we show the results of *The New York Times* (NYT), which adopted a paywall in March 2011. **Table 3** shows the estimated weights—from Equation (4)—assigned to the 36 control firms (the firms in the donor pool) to construct the optimal synthetic control unit for the NYT. Although there are 36 potential contributors, only a fraction of them contribute to the synthetic control of the NYT. In terms of print demand (circulation), *USA Today* contributes the most, at 48%, followed by *The Record* and *The Wall Street Journal* at 24% each. In terms of digital demand (pageviews), most of the influence comes from *The Wall Street Journal*. As mentioned, the weighted average of the control firms' sales outcome (either circulation for print

[12] We do not jointly estimate the two equations because our sample periods and time intervals are different for each channel.

demand or pageviews for digital demand) would be the counterfactual outcome if the NYT had not adopted a paywall.

Figure 2a compares the actual and counterfactual print circulations of the NYT over time. Consistent with the overall industry trend (see **Figure 1a**), the NYT's actual print circulation significantly decreased throughout the periods. However, the synthetic control results suggest that the NYT's print circulation would have decreased even more if it had not introduced a digital paywall (the dashed line). Hence, the synthetic control method indicates that the NYT's paywall actually decelerated the decreasing trend in print circulation. Similarly, **Figure 2b** compares the NYT's actual and counterfactual monthly pageviews, normalized per million Internet users. Starting in January 2010, the actual monthly pageviews significantly decreased, which is consistent with the overall industry trend (see **Figure 1b**).[13] However, in contrast to print circulation, the synthetic control results suggest that the NYT's pageviews would have decreased less if it had not introduced a paywall (the dashed line).

We apply the synthetic control method to all paywalled media firms in our sample to calibrate the effect of paywall on both print and digital demand. **Figure 3a** shows the distribution of the effect of paywall on print circulation—that is, the difference between the actual and counterfactual print circulation for each firm, in each period, after the inception of a paywall. Although the average effect of paywall on print circulation is positive (4% increase), there is considerable heterogeneity in its effect. Similarly, **Figure 3b** shows the distribution of the effect of paywall on pageviews—that is, the difference between the actual and counterfactual pageviews

[13] The industry trend in pageviews includes those of both paywalled and non-paywalled newspapers. Hence, the increase in the number of paywalled newspapers contributes to the declining trend.

for each media firm after the inception of a paywall.[14] The average effect of paywall on pageviews is significantly negative (25% decrease). That is, on average, paywalled firms lost about 25% of their traffic (pageviews) due to the inception of a paywall. Again, a large degree of heterogeneity exists in the effect of paywall. We discuss the cause of this heterogeneity in the following sections.

4.2 The Effects of a Paywall on Print Demand

We apply the two-stage least squares (2SLS) method to estimate the model and use the White robust standard error to account for the heteroskedasticity in the error structure. Columns (I) and (II) of **Table 4** show the results of the first-stage equation with regard to print and digital subscription prices, respectively, in the 2SLS estimation. As expected, input (editor) wages are positively related to both the print and digital subscription prices, implying that firms increase price with an increase in costs.

Columns (I) – (III) of **Table 5** show the results of several model specifications of Equation (7) on print demand in terms of circulation. Column (I) shows the result without addressing price endogeneity. Column (II) shows the result using instrumental variables to correct for price endogeneity—our main model. Column (III) shows the result of the fixed-effects model using instrumental variables. Comparing the three columns, we find the following. First, there are meaningful differences in parameter estimates between columns (I) and (II), suggesting the importance of addressing price endogeneity. Specifically, when endogeneity is not addressed, the coefficients on price becomes noticeably smaller in magnitude. Second, the price coefficient of the fixed-effects model with endogeneity correction (column (III)) is not statistically different from

[14] Figure 3(b) shows that, for some firm-month combination, a paywall has a positive effect on digital demand . This is because monthly pageviews are highly volatile across time where, in certain months for some firms, pageviews are unusually large. However, the effect sizes are negligible (within the error range), and when averaged across months, the effect of paywall is significantly negative.

that of the suggested model (column (II)). The fixed-effects model controls for arbitrary correlation between price and unobserved firm-specific effects. Hence, the not different price coefficients between the two model specifications suggest that our main model obtains unbiased estimates and, at the same time, provides inferences on the heterogeneous effects of paywall across firms, utilizing the cross-sectional variation in the data.

The results indicate that a digital paywall sales strategy can be effective in increasing print demand (substitution spillover effect), especially for media firms with high reputation and equity (firms with a large print circulation) and unique content. The size of a firm provides a readily accessible and, thus, diagnostic measure of reputation. Hence, as in other industries, the reputation of a firm is an important factor that determines consumer demand and, in our context, spillover (from digital to print) demand. Similarly, firms with unique content induce positive spillovers to print demand with the adoption of a paywall.

Note that the estimated positive coefficient ($\widehat{\beta^{(1)} - \rho}\beta^{(0)} = 0.139$) on print subscription price does not mean that the effect of price on demand is positive. Rather, the estimated coefficient shows that media firms set their print subscription prices in relation to the market-level price elasticity without ($\beta^{(0)}$) and with a paywall ($\beta^{(1)}$). Thus, the estimated price coefficient provides valuable insights into how firms would have set their print subscription prices had they not adopted a paywall: $\rho = \frac{\beta^{(1)} - 0.139}{\beta^{(0)}}$. Utilizing the data pre- and post-paywall via firms, we estimate the price elasticities before and after a paywall to be -0.978 and -0.614, respectively.[15] The two elasticity

[15] To obtain the pre-paywall price elasticity, we estimate the following regression for the pre-paywall periods: $y_{it} = \alpha_i^{(0)} + \beta^{(0)} p_{it} + \varepsilon_{it}$, where y_{it} is the log of print circulation and p_{it} is the log of print subscription price of firm i in period t; thus, $\beta^{(0)}$ is the pre-paywall price elasticity of print demand. Similarly, to obtain the post-paywall price elasticity, we estimate the following regression for the post-paywall periods: $y_{it} = \alpha_i^{(1)} + \beta^{(1)} p_{it} + \varepsilon_{it}$, where y_{it} and p_{it} are defined as above; thus, $\beta^{(1)}$ is the post-paywall price elasticity of print demand. In both analyses, we use the wage of editors as an instrument for price and compute robust standard errors.

estimates imply that a paywalled media firm would have charged a substantially lower price for its print subscription ($\hat{\rho} = \frac{-0.614 - .139}{-0.987} \approx 0.76$) had it not introduced a paywall.

4.3 The Effects of a Paywall on Digital Demand

Columns (IV) – (VI) of **Table 5** show the estimation results of several model specifications of Equation (7) for digital demand in terms of pageviews. Column (IV) shows the estimation result without addressing the endogeneity of price. Column (V) shows the estimation result with endogeneity correction—our suggested model. Column (VI) shows the result of the fixed-effects model using instrumental variables. Similar to that of print demand, the price coefficient becomes larger in magnitude when corrected for endogeneity and is not different from that of the fixed-effects model. The results indicate that, in terms of alleviating the decline in digital demand, a digital paywall sales strategy is effective for firms with unique content.

The statistically insignificant price coefficient ($\widehat{\beta^{(1)} - \rho\beta^{(0)}} = 0$) provides further insight into how firms set digital subscription prices. The insignificance of the price coefficient ($\hat{\rho} = \frac{\hat{\beta}^{(1)}}{\hat{\beta}^{(0)}}$) suggests that a paper's digital subscription price *without* a paywall would be proportional to the ratio of the two price elasticities: $\hat{\beta}^{(1)}$ and $\hat{\beta}^{(0)}$. Because the aggregate price elasticity at the market level is likely smaller in magnitude (less price-sensitive) with a paywall than that without it ($\beta^{(0)} < \beta^{(1)} < 0$), $\hat{\rho}$ would be less than one. In addition, because of transition (search) costs (the monetary cost of browsing on a firm's website), it is realistic to assume that although the digital subscription price is zero before a paywall, the actual price (or cost) inferred by the reader is not zero, but some small positive figure related to search costs. Hence, our digital demand estimates provide inferences on how media firms set digital subscription prices proportional to the ratio of price elasticities and consumer search costs.

5. The Comprehensive Effects of a Digital Paywall on Firm Revenue

As described in **Table 1**, pursuing a digital paywall sales strategy affects a media firm's different source of revenue—subscription and advertising—across two channels—print and digital. We first discuss subscription and advertising revenues across the two channels. Then, we discuss the magnitude of the comprehensive effect of a digital paywall for various heterogeneous media firms.

5.1 Subscription Revenue

The introduction of a digital paywall affects subscription revenue in two ways: (1) it creates a new source of revenue—digital subscription revenue; and (2) it affects print subscription either positively or negatively (depending on the direction of the demand-spillover effect). To compute the marginal effect on print subscription revenue, we utilize the counterfactual print demand (through synthetic control) and the counterfactual print subscription price (by assuming $\rho_{it} = 0.76$, as discussed in Section 5.2), assuming that the firm did not adopt a paywall. We then compare the counterfactual print subscription revenue with the actual subscription revenue. Digital subscription revenue is computed by simply multiplying the number of digital subscriptions by the digital subscription price. Because digital subscription revenue did not exist before a paywall, all digital subscription revenue can be considered the effect of the paywall.

5.2 Advertising Revenue

For a media firm that considers a digital paywall, a key interest would be its effect on advertising revenue. A digital paywall changes print and digital subscriptions and, thus, would affect print and digital advertising revenues, as subscription demand is closely related to advertising demand (Armstrong 2006).

To obtain the marginal effect of paywall on advertising revenue—both print and digital—we utilize the counterfactual demand (obtained through synthetic control), assuming that the media

firm did not introduce a paywall. We first compute the marginal change in demand by comparing the counterfactual demand with the actual demand in both channels. Then, we obtain the print and digital advertising rates for each firm-period combination utilizing advertising revenue such that:

$$PrintAdRate_{it} = \frac{PrintAdRevenue_{it}}{PrintCirculation_{it}}$$

$$DigitalAdRate_{it} = \frac{DigitalAdReve\quad_{it}}{Pageviews_{it}},$$

where $Pageviews_{it}$ is the pageviews for media firm i's website at time t. To compute the effect of a paywall on print advertising revenue, we multiply the print advertising rate by the difference between the actual and counterfactual print circulation. Similarly, we multiply the digital advertising rate by the difference between the actual and counterfactual pageviews to compute the effect of a paywall on digital advertising revenue.

5.3 Total Firm Revenue

Table 6 shows the change in revenue by source and channel for select firms whose print and digital advertising data are available. All figures represent a marginal change in a firm's annual revenue as a result of adopting a digital paywall. A substantial amount of heterogeneity exists in terms of the comprehensive effect of a paywall, ranging from a 24% increase to a 12% decrease in total firm revenue. Large-sized firms such as *The New York Times*, the *Chicago Tribune*, and the *Los Angeles Times* significantly benefited, whereas small-sized firms, such as the *Sun- Sentinel* and the *Orlando Sentinel* did not benefit from pursuing a digital paywall sales strategy.

The results reveal some interesting insights regarding the source of the revenue streams with the adoption of a paywall (see **Figure 4**). For example, the success of *The New York Times* came from a large increase in print and digital subscription revenues. That is, by introducing a paywall, *The New York Times* decelerated the declining trend in print subscriptions and

substantially increased its digital subscription base, more than offsetting the noticeable decrease in digital advertising revenue (due to the decrease in web traffic). In contrast, the success of *The Des Moines Register* came mainly from an increase in print subscriptions and print advertising revenue and less so from an increase in digital subscription revenues, the direct source of income generated from a digital paywall.

The noticeable success of these firms contrasts with the null or unfavorable outcomes of firms such as the *Sun-Sentinel*, the *Chicago Sun-Times*, the *Orlando Sentinel*, and *The Denver Post*. Again, the cause of the non-positive outcome differs by firm. The *Sun-Sentinel* lost a significant amount in digital advertising revenue, offsetting the increase in digital and print subscription revenue, whereas *The Denver Post* failed to sustain print advertising revenue (negative spillover to print demand). Surprisingly, for most print media firms, the net effect of a paywall in the digital channel is negative, as the newly generated digital subscription revenue is offset by a significant decrease in digital advertising revenue. That is, the decrease in web traffic after a paywall is significant to the level at which the decrease in the digital advertising revenue is larger than the increase in digital subscription revenue. If it were not for the positive spillover effect onto print demand (and, thus, onto print advertising revenue), the net effect of a paywall would have been mostly negative. This fact shows the importance of understanding and accounting for the comprehensive effect of a digital paywall, beyond the direct effect on digital subscriptions and digital advertising revenues.

6. Conclusion

More than two decades have passed since media firms (newspapers and TV networks) began to publish online digital content via their websites. In the early days, online media content was offered free of charge as a means to increase the social awareness and prestige of the firm. Later, while

still maintaining free online content, media firms transitioned to a digital advertising business model, anticipating that the traffic generated by free digital content would produce sizable income via digital advertising (banner and display ads by third parties). However, for the print media industry, the increase in digital adverting was not enough to offset the decrease in print demand, negatively affecting both print subscriptions and advertising revenues. Thus, many print media firms decided to pursue a new pricing and go-to-market strategy—a digital paywall sales strategy—utilizing a digital paywall to transform their business model from a free-for-online-content to a pay-for-online-content model.

This research examines the comprehensive effects of a digital paywall sales strategy—specifically, how the adoption of a digital paywall affects a media firm's two main sources of income—subscription and advertising—across its two channels—traditional and digital. To conduct our analysis, we compile a unique and comprehensive dataset from multiple sources that contain detailed data on a majority of the top print media firms in the U.S.; and, for inference, we utilize a synthetic control method to disentangle the causal effect of a paywall from naturally occurring trends. Furthermore, to distinguish the effect of paywall from that of price, our model encompasses a firm's pricing decision that can occur simultaneously with the adoption of a paywall. In addition, we explore demand spillovers—substitution vs. complementarity—across channels and factors that moderate such spillover effects.

For a typical print media firm, the decision to introduce a paywall likely affects both its digital and print businesses, and, thus, it is important to understand how the free-to-paid transition of a firm's online content affects the firm's other trade through print subscriptions and advertising. The relation between different products (or services) has important implications for business decision making. For example, if two products complement one another, firms can lower the price

of one good to increase the sales of the other good. In contrast, if two products are substitutes, firms can raise the price of one good to increase the sales of the other good. In 2011, Amazon.com introduced the Kindle Fire at $199, below its cost (Milian, 2011), anticipating that users of Kindle Fire would become members of the Amazon Prime service or purchase more books from the company. In the print media industry, studies find that the relation between the different channels (print and digital) can either be complementary or substitutive (Deleersnyder et al., 2002; Gentzkow, 2007; Xu et al., 2014). This study examines demand relation via different channels in totality across heterogeneous firms to assess the comprehensive effect of a digital paywall. To the best of our knowledge, this study is among the first to do so.

Overall, we find that the adoption of a digital paywall can lead to positive demand substitution from digital to traditional channels, particularly for large-sized firms and those with unique content. In addition, unique content mitigates the decrease in digital demand, alleviating the loss of digital advertising income while adding a new source of income—digital subscription revenue. The effectiveness of a digital paywall, however, varies by the source of income via channel across heterogeneous media firms.

This study has several limitations. First, while print media firms typically offer diverse price tiers, we do not take into account the complex pricing structure. For feasibility, in our empirical analysis, we instead use the basic home delivery price, as reported in publisher statements. Second, each media firm allows different numbers of free articles per month. While both print and digital demand may be influenced by free articles, we cannot include this information due to the lack of data. Third, to promote the adoption of a digital paywall, media firms run limited-time offers and promotions, but we cannot incorporate the effects of such efforts, again due to limited data. Finally, the analyses are conducted at the aggregate and not at the

individual level and, thus, do not incorporate consumer heterogeneity. If detailed individual-level data are available, one can examine factors that determine consumers' decision making when faced with a digital paywall, which can provide insights into segmentation, promotion and pricing to better utilize a digital paywall. Although not addressed in this research due to data limitations, the abovementioned issues would be an exciting venue for future research.

In summary, this research provides a practical application to examine the comprehensive effects—subscriptions and advertising across digital and traditional channels—of a digital paywall sales strategy, a popular means of go-to-market for many media firms. We believe that our results will guide media firms in pursuit of a digital paywall to better position their online content to alleviate the decrease in digital advertising while increasing print subscriptions and advertising to achieve greater sales in both the traditional and digital channels. Although our empirical application is conducted for the U.S. print media industry, the method of analysis can be extended to other industries involving a free-to-pay transition of a firm's online content, which can bring additional insights and, thus, can also be an exciting avenue for future research.

REFERENCES

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

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

Anand, Bharat (2016), *The Content Trap: A Strategist's Guide to Digital Change*, Random House.

Armstrong, Mark (2006), "Competition in Two-Sided Markets," *The RAND Journal of Economics*, 37 (3), 668-691.

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

Banville, Lee (2016), *Covering American Politics in the 21st Century: An Encyclopedia of News Media Titans, Trends, and Controversies*, ABC-CLIO: Santa Barbara, CA.

Chevalier, Judith and Dian Mayzlin (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research* 43 (August), 345-354.

Chiou, Lesley and Catherine Tucker (2013), "Paywalls and the Demand for News," Information Economics and Policy, 25, 61-69.

Deleersnyder, Barbara, Inge Geyskens, Katrijn Gielens, and Marnik G. Dekimpe (2002), "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, 337-348.

Gentzkow, Matthew (2007), "Valuing New Goods in a Model with Complementarity: Online Newspapers," The American Economic Review, 97 (3), 713-744.

Gentzkow, Matthew and Jesse M. Shapiro (2010), "What Drives Media Slant? Evidence from U.S. Daily Newspapers," Econometrica, 78 (1), 35-71.

Kim, Ho, Reo Song, and Youngsoo Kim (2018), "Free-to-Paid Transition of Online Content Providers: An Analysis of Paywall Rollouts by Newspapers," Working paper.

Koukova, Nevena T., P. K. Kannan, and Brian T. Ratchford (2008), "Product from Bundling: Implications for Marketing Digital Products," *Journal of Retailing,* 84 (2), 181-194.

Milian, Mark (2011), Amazon to Lose Money on Each Tablet Sale , Analysts Say, CNN. http://edition.cnn.com/2011/09/30/tech/gaming-gadgets/kindle-fire-price/index.html.

Pattabhiramaiah, Adithya, S. Sriram, and Puneet Manchanda (2017), "Paywalls: Monetizing Online Content," forthcoming at the *Journal of Marketing*.

Pew Research Center (2017), "Newspaper Fact Sheet," http://www.journalism.org/fact-sheet/newspapers/.

Pauwels, Koen and Allen Weiss (2008), "Moving from Free to Fee: How Online Firms Market to Change Their Business Model Successfully," *Journal of Marketing*, 72 (May), 14-31.

Tirunillai, Seshadri and Gerard J. Tellis (2017), "Does Offline TV Advertising Affect Online Chatter? Quasi-Experimental Analysis Using Synthetic Control," *Marketing Science*, 36 (6), 862-878.

Williams, Alex T. (2016), "Paying for Digital News: The rapid adoption and current landscape of digital subscriptions at U.S. newspapers," American Press Institute, https://www.americanpressinstitute.org/publications/reports/digital-subscriptions/single-page/.

Xu, Jiao, Chris Forman, Jun B. Kim, and Koert Van Ittersum (2014), "News Media Channels: Complements or Substitutes? Evidence from Mobile Phone Usage," *Journal of Marketing*, 78(July), 97-112.

Table 1. The Effect of a Digital Paywall on Revenue Source by Channel

Revenue source / channel	Print	Digital
Subscription (or circulation)	Increase (decrease) if print and digital channels are substitutes (complements)	Increase from zero (without a paywall)
Advertising	Increase (decrease) if print and digital channels are substitutes (complements)	Decrease because of reduction in web traffic

Table 2. Media Firm Characteristics and Reader Demographics

		Mean	Median	St. Dev.	Minimum	Maximum
Firm Characteristics						
Conservative Slant		0.447	0.448	0.032	0.350	0.514
Content Uniqueness		2.244	2.213	0.911	0.100	4.812
Print Circulation as of 2010		225,889	144,294	313,789	75,615	2,061,142
Reader Demographics						
Age	18 – 24	0.115	0.106	0.034	0.063	0.190
	25 – 34	0.165	0.167	0.023	0.120	0.213
	35 – 44	0.159	0.162	0.019	0.103	0.197
	45 – 54	0.177	0.178	0.016	0.141	0.214
	55 and Over	0.384	0.384	0.052	0.283	0.468
Income	Below $50,000	0.434	0.435	0.067	0.282	0.563
	50,000 – 74,999	0.176	0.179	0.026	0.108	0.228
	75,000 – 99,999	0.149	0.146	0.023	0.100	0.217
	100,000 – 149,999	0.137	0.136	0.028	0.080	0.198
	150,000 and Over	0.103	0.087	0.050	0.038	0.253
Education	High School or Less	0.353	0.356	0.061	0.211	0.543
	Some College	0.320	0.324	0.046	0.165	0.410
	College and More	0.327	0.317	0.068	0.212	0.571

Table 3. An Illustration of Synthetic Control: The Weights for The New York Times

Control Firms (Donor pool)	Contribution (Weight)	
	(I) Print Circulation	(II) Web Traffic
1. Albuquerque Journal		
2. Arizona Daily Star		
3. Arkansas Democrat-Gazette		
4. Austin American-Statesman		
5. Boston Herald	0.005	
6. Daily News		
7. Dayton Daily News	0.020	
8. Fort Worth Star-Telegram		
9. Grand Rapids Press		
10. Intelligencer Journal/Lancaster New Era		
11. Las Vegas Review-Journal		
12. New York Post	0.031	
13. Pittsburgh Post-Gazette		
14. Pittsburgh Tribune-Review		
15. Press-Register		
16. San Antonio Express-News		
17. St. Louis Post-Dispatch		
18. St. Petersburg Times (Tampa Bay Times)		
19. The Advocate		
20. The Birmingham News		
21. The Commercial Appeal		
22. The Florida Times-Union		
23. The Journal News		0.032
24. The Oklahoman		
25. The Oregonian		
26. The Post-Standard		
27. The Press-Enterprise		
28. The Record	0.243	
29. The Roanoke Times		
30. The Star-Ledger		
31. The Tennessean		
32. The Times of Northwest Indiana		
33. The Times-Picayune		
34. The Virginian-Pilot		
35. The Wall Street Journal	0.242	0.968
36. USA Today	0.484	
Total	1.000	1.000

Table 4. The First-Stage Regression (in 2SLS) Results

	Dependent Variable	
	(I) Log of Print-Subscription Price	(II) Log of Digital-Subscription Price
Constant	-0.587	-7.242***
	(4.378)	(2.505)
Log of the Input (Editor) Wage	0.812***	0.247***
	(0.118)	(0.025)
Conservative Slant	-0.721	-0.037
	(1.889)	(2.096)
Content Uniqueness	-0.070	0.022
	(0.067)	(0.068)
Log of Print Circulation	-0.174	0.274
	(0.185)	(0.167)
Age: 25 – 34	-2.552	1.808
	(3.538)	(2.681)
Age: 35 – 44	3.523	0.170***
	(3.491)	(4.070)
Age: 45 – 54	1.924	2.120
	(4.541)	(5.218)
Age: 55 and over	-0.493	3.646*
	(1.856)	(1.898)
Education: High School or less	-1.573	4.819***
	(4.973)	(1.507)
Education: Some College	0.815	0.201
	(3.254)	(2.788)
Income: 50,000 – 74,000	1.573	3.386
	(4.973)	(3.818)
Income: 75,000 – 99,000	-1.484	3.230
	(3.367)	(2.730)
Income: 100,000 – 149,000	-1.654	3.620
	(3.349)	(3.926)
Income: 150,000 and more	4.548	5.800*
	(3.409)	(3.148)
Number of Observations	411	2376

***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$. Standard errors are in parenthesis. The results of the owner dummy variables are omitted for brevity.

Table 5. The Effect of a Digital Paywall by Firm Characteristics and Reader Demographics

	Print Demand (Circulation)			Digital Demand (Web Traffic)		
	(I)	(II)	(III)	(IV)	(V)	(VI)
	Without Endogeneity Correction	Endogeneity Correction	Fixed effects with Endogeneity Correction	Without Endogeneity Correction	Endogeneity Correction	Fixed effects with Endogeneity Correction
Constant	-0.722	-1.184		0.021	-2.965	
	(0.877)	(1.027)		(2.492)	(3.927)	
Log of Print-Subscription Price	0.073***	0.139**	0.143***			
	(0.018)	(0.068)	(0.068)			
Log of Digital Edition Price				-0.113	-0.728	-0.743
				(0.089)	(0.527)	(0.535)
Conservative Slant	0.186	0.223		1.273	1.267	
	(0.498)	(0.496)		(1.523)	(2.583)	
Content Uniqueness	0.038***	0.043***		0.095***	0.109**	
	(0.012)	(0.014)		(0.037)	(0.050)	
Log of Print Circulation	0.051	0.061*		0.009	0.182	
	(0.033)	(0.034)		(0.108)	(0.175)	
Age: 25 – 34	1.031	1.153*		0.116	1.344	
	(0.685)	(0.681)		(1.981)	(3.508)	
Age: 35 – 44	-2.393***	-2.641***		2.163	2.309	
	(0.693)	(0.760)		(1.748)	(2.984)	
Age: 45 – 54	-0.300	-0.456		-1.954	-0.584	
	(0.727)	(0.768)		(2.884)	(4.093)	
Age: 55 and over	-0.252	-0.275		0.055	2.457	
	(0.336)	(0.345)		(1.236)	(2.853)	
Education: High School or less	0.227	0.326		-0.789	2.174	
	(0.397)	(0.422)		(1.353)	(2.767)	
Education: Some College	-1.091*	-1.223**		-3.423***	-3.087	
	(0.589)	(0.581)		(1.213)	(2.866)	
Income: 50,000 – 74,000	-0.178	-0.188		1.610	3.421	
	(0.870)	(0.883)		(2.409)	(2.639)	
Income: 75,000 – 99,000	1.310**	1.377**		0.524	2.592	
	(0.605)	(0.629)		(1.747)	(3.221)	
Income: 100,000 – 149,000	0.744	0.862		3.407**	5.618*	
	(0.541)	(0.544)		(1.526)	(2.908)	
Income: 150,000 and more	-1.042	-1.393**		-1.237	2.468	
	(0.636)	(0.703)		(1.603)	(4.207)	
Instrument for Price	No	Yes	Yes	No	Yes	Yes
Number of Observations	411	411	411	2376	2376	2376
R-squared: Within	0.1157	0.1157	0.0103	0.0108	0.0108	0.0000
R-squared: Between	0.4225	0.4095	0.0002	0.3883	0.1713	0.0210
R-squared: Overall	0.3082	0.2832	0.0290	0.1761	0.0549	0.0031

***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$. Standard errors are in parentheses.

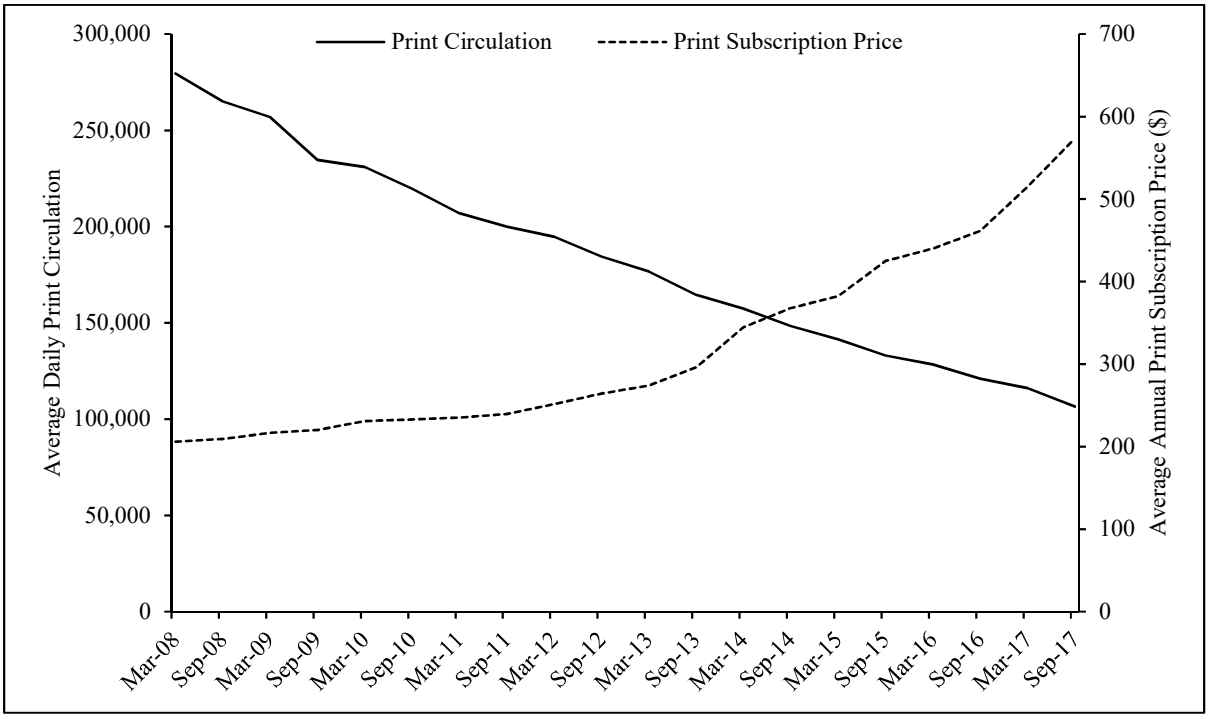
Table 6. The Comprehensive Effect of a Digital Paywall Sales Strategy: Firm Revenue by Source and Channel with the Adoption of a Digital Paywall ($1,000)

Media Firm	Total		Print Subscription	Digital Subscription	Print Advertising	Digital Advertising
	Amount Change*	% Change†	Amount Change*	Amount Change*	Amount Change*	Amount Change*
The New York Times	235,788	20.00%	166,988	93,380	35,720	-60,300
Chicago Tribune	94,492	19.00%	56,542	1,704	35,133	1,112
Los Angeles Times	93,966	12.50%	60,788	12,176	35,983	-14,981
Star Tribune	39,295	24.30%	23,507	7,389	10,863	-2,464
The Sacramento Bee	24,726	15.30%	22,569	338	4,431	-2,612
The Des Moines Register	16,845	18.00%	12,675	619	4,512	-960
The News & Observer	14,631	16.90%	12,745	332	2,157	-603
The Buffalo News	13,696	12.70%	8,694	1,147	4,363	-508
The Charlotte Observer	11,802	10.00%	11,577	202	-268	291
The Indianapolis Star	11,161	8.30%	12,864	1,263	881	-3,847
Democrat and Chronicle	7,044	5.20%	9,342	427	-959	-1,765
The Baltimore Sun	6,209	3.50%	12,609	2,042	-2,277	-6,164
The Courier-Journal	5,337	5.10%	9,138	601	-1,511	-2,892
The Post and Courier	3,921	8.20%	5,154	276	603	-2,113
The Denver Post	-3,765	-1.00%	11,010	271	-13,871	-1,175
Orlando Sentinel	-7,090	-9.50%	3,350	149	-4,461	-6,128
Chicago Sun-Times	-8,498	-4.60%	9,302	292	-5,398	-12,694
Sun Sentinel	-24,899	-11.90%	8,685	718	-12,600	-21,702

* The figures represent the difference between firms' actual (with a paywall) and counterfactual (without a paywall) annual revenue by source (subscription and advertising) and channel (print and digital).
† The figures represent the percentage change in firms' revenue (relative to the total *actual* annual revenue) as a result of a paywall.

Figure 1. Aggregate Trends

(a) Print Circulation and Print Subscription Price



(b) Web Traffic and Digital Subscription Price

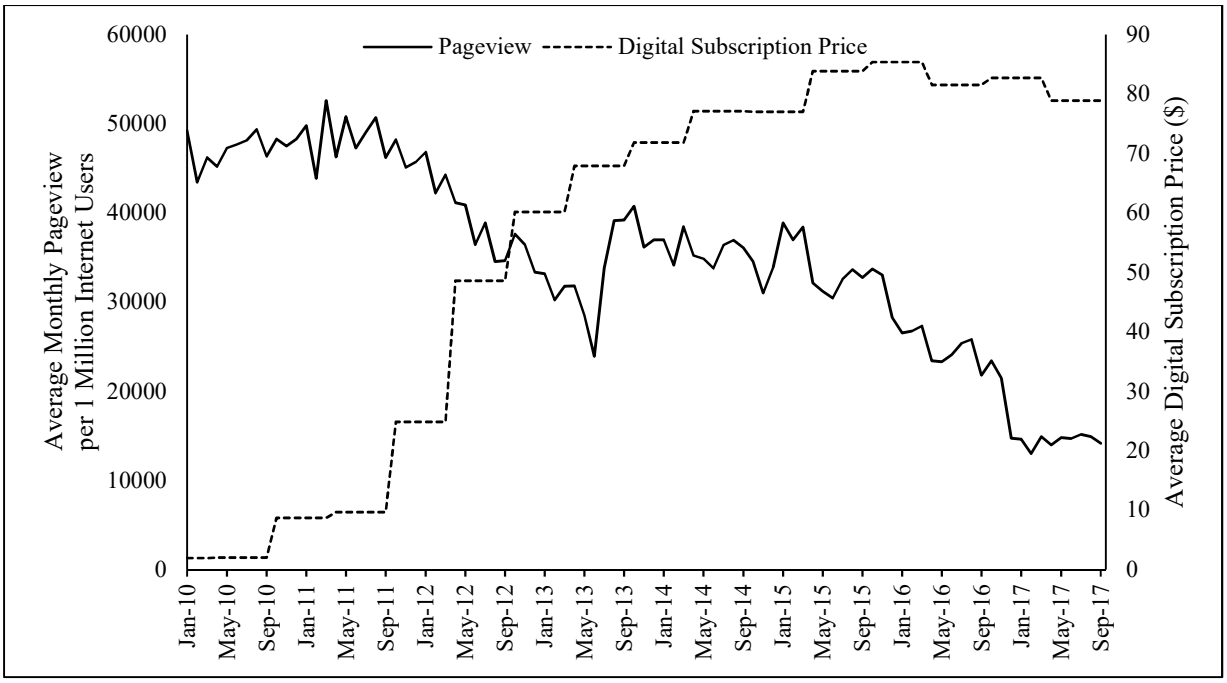
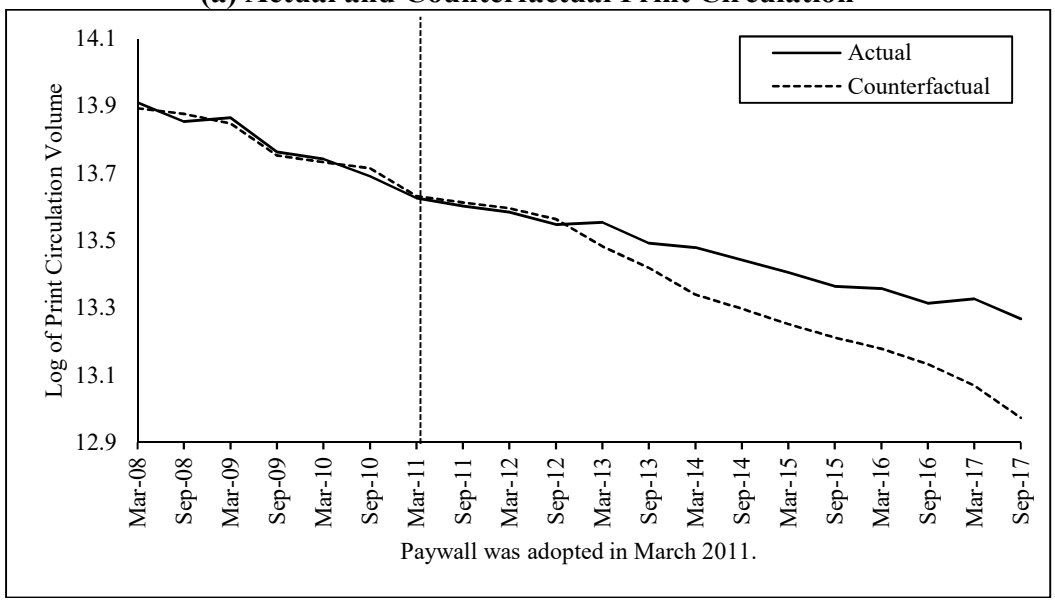


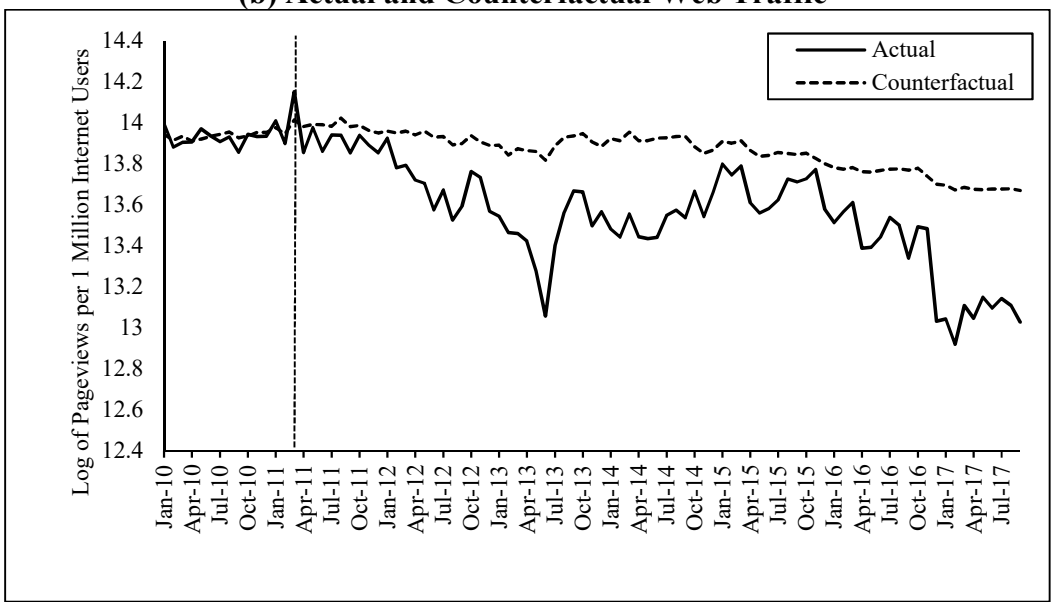
Figure 2. Synthetic Control for the New York Times

(a) Actual and Counterfactual Print Circulation



The New York Times adopted its paywall in March 2011. The vertical dotted line marks the paywall adoption period.

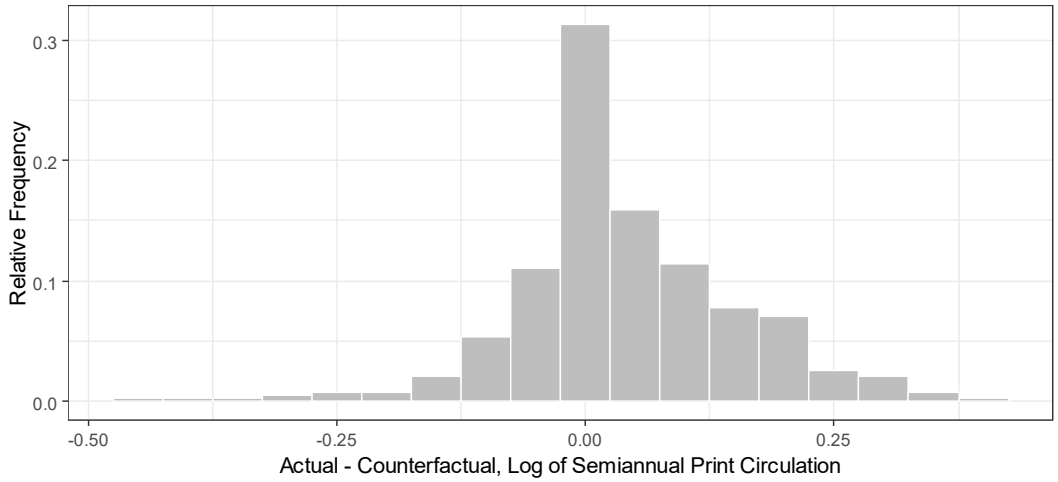
(b) Actual and Counterfactual Web Traffic



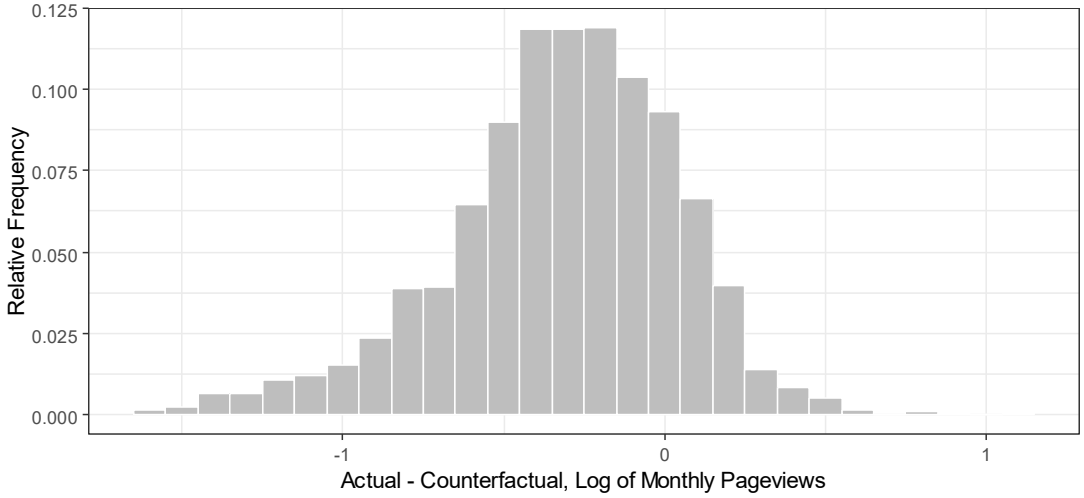
The New York Times adopted its paywall in March 2011. The vertical dotted line marks the paywall adoption period.

Figure 3. The Distribution of the Effect of Paywall on Demand

(a) Print Demand (Semiannual Circulation) (Number of Observations: 411)



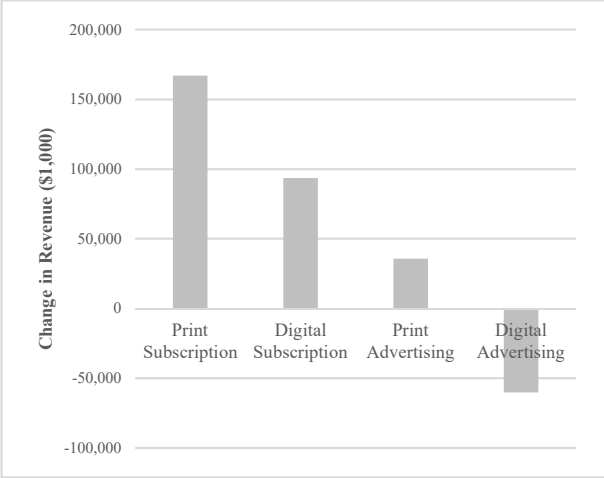
(b) Digital Demand (Monthly Pageviews) (Number of Observations: 2376)



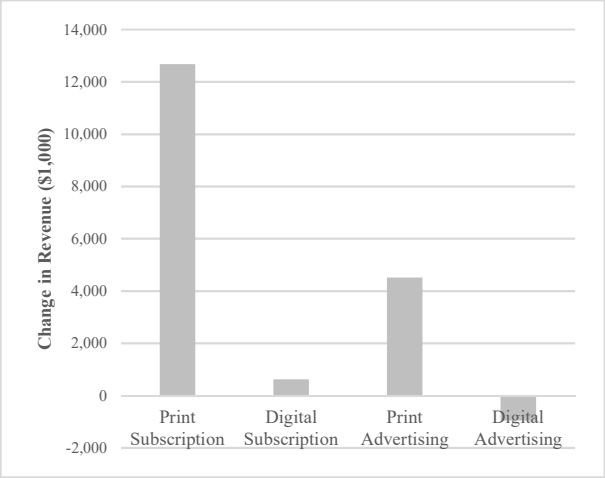
Note: Some newspaper-month combinations have positive effects in 3(b). This is because monthly pageviews are highly volatile—i.e., pageviews are unusually large in some months for some newspapers. However, the effect sizes are negligible (suggesting that positive effects are most likely within the error range) and when averaged across months, the effect of paywall is significantly negative.

Figure 4. The Change in Revenue by Source and Channel (Select Firms)

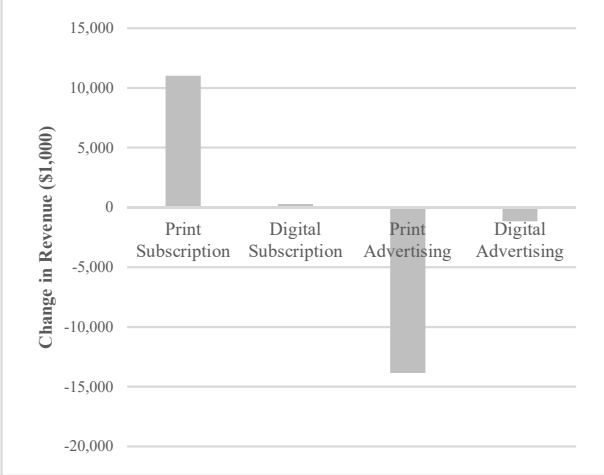
a) The New York Times



b) The Des Moines Register



c) The Denver Post



d) Sun Sentinel