Pricing Online Content: Fee or Free?*

Anja Lambrecht[+]

Kanishka Misra[±]

- Preliminary and Incomplete -

- Please do not Circulate -

March 30, 2014

Keywords: Paywall, Pricing, Online Media, Internet, Electronic Commerce, Free Content, Paid Content

* We are grateful to Rajesh Chandy, S. Sriram, Catherine Tucker, seminar participants at Technical University Munich, Groningen University, MIT and participants at the Theory and Practice in Marketing 2013 Conference, the Choice Symposium 2013 and at the 2013 Marketing Science conference for their comments and suggestions.
[+] Anja Lambrecht, London Business School, London, UK; alambrecht@london.edu.
[±] Kanishka Misra, Ross School of Business, Michigan; kanishka@umich.edu.

Pricing Online Content: Fee or Free?

Abstract

Many online content providers aim to compensate for a loss in advertising revenues by charging consumers for access to online content. However, such a choice is not straightforward since subscription fees typically deter customers, further reducing advertising revenues. As of yet, academic research offers little guidance on whether firms indeed benefit from charging for content and, in particular how firms should optimally implement such a "fee" model.

In this research, we empirically examine and quantify a content provider's trade-off between advertising and subscription revenues. We build a unique data set from the sports' website ESPN.com. ESPN.com offers the majority of content for free but charges a membership fee for a subset of articles. We collect data on the number of free and paid articles per day and sport, as well as demand for each type of article per day and sport over a 13 months period.

We estimate how the number of free and paid articles affects viewership of the site and empirically quantify a firm's trade-off between advertising and subscription revenues. Our approach controls for a wide range of demand shifters and possible endogeneity of the number of articles the firm offers on any day. We find that, on average, the firm should not adjust the amount of paid content. However, we find strong differences across sports' seasons: the marginal paid article increases revenue in the off season but decreases revenue in regular season. We determine the increase in revenue if the firm followed our suggested strategy. More generally, our results suggest that online content providers should carefully identify temporal variation in demand and flexibly adjust the amount of paid content they offer rather than setting a static paywall.

1 Introduction

The future of the media industry is widely believed to depend on the ability of media companies to monetize content online. However, for well over a decade, the prevalent view has been that “information wants to be free” and that consumers are unwilling to pay for content online (Edgecliffe-Johnson 2009). This is supported by research showing that consumers respond negatively to even small monetary fees (Shampanier et al. 2007; Ascarza et al. 2012), making it difficult to charge even small amounts for digital content.

Yet, plummeting advertising revenues across the media industry (see Figure 1) force companies to identify new and additional sources of revenue: In December 2008, The Tribune, owner of the Chicago Tribune and LA Times, filed for bankruptcy protection. In 2009, the New York Times’ credit crisis prompted a piece questioning its continued existence (Hirschorn 2009). Then, in August 2013 The Washington Post was sold to Jeff Bezos as “for much of the past decade, The Post has been unable to escape the financial turmoil that has engulfed newspapers” (Farhi 2013). Many regional newspapers, such as the Miami Herald and the San Francisco Chronicle face similar financial trouble.[1]

Just how new and additional sources of revenues should be opened remains unclear. Charging for online content adds subscription revenue (Pauwels and Weiss 2007) but can significantly deter consumers, leading to lower advertising revenues. For example, Chiou and Tucker 2012 find that visits to websites of local newspapers fell by 73% after the introduction of a paywall.

[1] http://www.realclearpolitics.com/lists/top_10_newspapers_in_trouble/miami_herald.html?state=play

Acknowledging the trade-off between subscription and advertising revenues, firms have in recent years experimented with a wide range of revenue models that include giving away all content for free (e.g., washingtonpost.com), charging for all content (e.g., thetimes.co.uk) and offering some content free of charge but charging for a subset of content (e.g. ESPN.com, faz.net, nyt.com)[2]. Some firms experiment with different strategies: The NYT initially offered all content for free but switched to a paid model with 20 and later 10 free articles per month. The Wall Street Journal required a subscription, later changed to a largely free version but reverted to a partly paid model.

Importantly, while the industry norm is to follow a static rule on how much content is free or paid (e.g., all paid, 10 free per month), firms can more flexibly adjust the amount of paid content they offer. For example, at the Wall Street Journal only subscribers can 'unlock' a selection of articles and this selection varies by day. Theoretically, such policies are promising if consumer demand for paid content varies over time. But as of yet there is little evidence on how specifically the trade-off of advertising versus subscription revenues plays out, and whether firms benefit from varying the amount of paid content they offer.

In this research, we empirically examine and quantify a content provider's trade-off between advertising and subscription revenues. We evaluate whether a firm should follow a static policy or dynamically adjust the amount of paid content it offers, and how such variation can be implemented.

We build a unique data set from the sports website ESPN.com. ESPN.com offers the majority of content for free but charges a membership fee for a subset of articles. The

[2] ESPN.com charges for a subset of articles, faz.net charges for historic articles, nyt.com charges for

number of paid articles varies by day and by sport. Via a web crawler, we collect data on the number of free and paid articles per day and sport over a thirteen months period. We complement this data with the number of unique visitors, page views, that is the number of instances a user visited any of the firm's web pages, and time spent for each type of article per day and sport.

We document how consumer demand varies across periods with different levels of demand, here sports' seasons. We then estimate how unique visitors to the firm's paid section, as a proxy for subscribers, and page views on the site as a proxy for advertising revenues respond to paid content. We control for a wide range of demand shifters and possible endogeneity of the number of articles the firm offers on any day.

We find that, on average, paid articles increase the number of visitors to the paid section, and thus subscribers, while reducing overall advertising revenues from page views on the site. Using these estimates we quantify the impact of adding an additional paid article on the firm's revenues. Our estimates suggest that on average the marginal increase in subscription revenue is statistically indistinguishable from the marginal decrease in advertising revenue, suggesting that the firm should not adjust the amount of paid content. However, we find strong differences over time when accounting for factors that exogenously vary consumer demand, here sports' seasons. Specifically, the marginal paid article increases revenues in off season but decreases revenues in regular season. For the post season we find mixed results across the different sports.

Our findings suggest that adding paid content may sometimes – but not always – be good for firms. Indeed, based on our results, it seems likely that many online content

any article that exceeds a monthly allowance of 10 free articles (on NYT, see also Kumar et al. 2013).

providers that currently use a static policy may benefit from re-adjusting their pricing strategy to dynamically respond to consumer demand.

More broadly, our results suggest that to increase revenue from online content, firms need to examine in detail what drives underlying consumer behavior and, ideally, identify in real-time periods of unusually high (or low) demand to then flexibly respond by adjusting the amount of paid content. This requires firms to track consumer usage behavior. In some instances, firms may have access to individual-level data, such as for consumers who have signed up for the paid section. For nonsubscribers this can often prove difficult because consumers may not allow cookies, cookies expire or simply because consumers use multiple devices. Our empirical approach illustrates how firms can use aggregate data to align their paid content offering with consumer demand.

2 Relationship to previous literature

Our research adds to three different but related streams of research. First, research using analytical modeling shows that offering both a paid and a free component can allow firms to implement quality differentiation, versioning or second-degree price discrimination (Shapiro and Varian 1998; Bhargava and Choudhary 2001). Additionally, while paid content reduces the number of page impressions and so advertising revenues, it can also lead to a positional disadvantage in advertising markets since advertisers are willing to pay a premium to firms with a high expected share of loyal consumers (Athey et al. 2011). Research further demonstrates that the trade-off between free and paid content varies with competition (Godes et al. 2009) and advertising effectiveness (Halbheer et al. 2013). Likewise, consumer heterogeneity affects the effectiveness of a content provider's revenue model. Specifically, with heterogeneous consumers, firms should combine pay-per-view and advertising revenues but offer options to consumers (Prasad et al. 2003).

A second, emerging, stream of research aims to provide empirical insights relevant to an online content provider's choice of revenue model (for an overview on online revenue models, see Lambrecht et al. 2013). While an online content provider targeted towards marketing professionals finds that moving from free to fee can be profitable (Pauwels and Weiss 2007), visits to online news sites can fall by as much as 73%, following the introduction of a paywall (Chiou and Tucker 2012)[3]. But lacking detailed data on advertising revenues or users' website activities, research to date has been unable to examine whether an increase in subscription revenue would off-set such losses in advertising revenues from reduced page views. Additionally, as of yet empirical studies have been limited to a static setup where either all or none of the content was free and have been unable to explore the effect of demand variation on paid content strategies.

More broadly, our work is motivated by a third stream of literature that focuses on the state of the newspaper industry. This research is concerned with an increase in subscription prices alongside declining circulation of print newspapers (Seamans and Zhu 2013; Pattabhiramaiah et al. 2013) and the effect on their ability to price discriminate (Angelucci et al. 2013). In studying how media firms can monetize their content online, we aim to contribute to the discussion on how media firms can build sustainable revenue models, given that consumers' attention increasingly shifts online.

3 In the case reported by Chiou and Tucker 2012, online readership declined dramatically following the introduction of a monthly fee of $9.95. The strong response to such a relatively small fee is perhaps not surprising given findings that consumers typically perceive the benefits associated with free products, compared to those of paid products, as higher than would be expected based on the price change alone (Shampanier et al. 2007; Ascarza et al. 2012).

In sum, while research so far has offered some broad guidelines on a content provider's choice of '*fee or free*', it does not provide conclusive evidence on whether and when firm benefits from charging for online content. Our research seeks to fill this gap. We quantify an online content provider's trade-off between subscription and advertising revenues, and examine whether the firm would benefit from dynamically adjusting the amount of paid content. We examine how such dynamic strategies can build on insights on heterogeneity in consumer willingness-to-pay over time and across individuals.

3 Data

3.1 Empirical setting

Our empirical study is set in the context of the sports' website ESPN.com. ESPN.com is the website of the US sports TV network ESPN and owned by Disney. ESPN.com provides a wide range of coverage on sports and sport events, including news and background reports. Following we refer to ESPN.com simply as ESPN.

The ESPN website has a main homepage plus homepages for each sport. The homepages display only title and links to articles but no abstracts or full articles. Importantly, ESPN offers two types of articles. Regular articles, available free of charge to all consumers (hereafter free articles) and "Insider" articles (hereafter paid articles), available only to consumers who pay a membership fee. On each sport's homepage, paid articles are easily recognizable through a small orange "in"-icon. The number of paid articles varies across days and sports.

In our empirical analysis, we focus on six different sports that typically offer both paid and free articles: College Basketball (CBA), College Football (CFB), Baseball

(MLB), Basketball (NBA), Football (NFL) and Hockey (NHL). We abstract form sports such as NASCAR and tennis that did not offer paid articles during our observation period.

3.2 Website Content and User Activity

A typical challenge in analyzing the effectiveness of “free” versus “fee” strategies, is the difficulty to obtain data that discloses detailed usage information alongside pricing strategies (Pauwels and Weiss 2007) while also controlling for industry-wide demand. We circumvent this challenge by combining multiple data sets. Our data capture, for a period of thirteen months, per day and sport the number of free and paid articles featured on the firm’s sport-specific homepage, the number of unique visitors to the paid and free sections on the firm’s website and the number of page views in both sections. They also include, on a day and sport level, unique visitors and page views to competitive sites. This means, that while we do not have user-level data, our data are disaggregate on the day-and-sport level. We next describe in more detail the different data sets we use.

3.2.1 Website Content of ESPN

First, we use a web scraper to collect on a daily basis the number of free and paid articles on each of the six sports’ homepages at ESPN from December 2010 to December 2011. As free articles, we collect all links with the url-format espn.go.com/sportname. As paid articles, we collect all links with the url-format insider.espn.go.com/sportname.[4] We then identify links that remain on a sport’s homepage for a very long time period (more than 100 days). These links typically do not represent content-based news articles but provide general information that often does not change over time (e.g., links to pages on the NBA draft for previous years or games timetables). We count as articles all links that appear on the sport's home page for less than 100 days. As the first part of Table 1, Panel A indicates, a sport's homepage displays 34 articles on average per day of which 25 are free and 9 paid.

We next explore the recency of articles. On average across all days and sports, 39% of free and 25% of paid articles displayed every day are new content whereas 61% of free and 75% of paid articles have already been displayed the previous day. On any day, the average age of free articles displayed is 11 days and the average age of paid articles is 7 days. This suggests that while the firm updates content over time updating happens gradually.

We compare free and paid articles in more detail. For a sample period of seven days (November 9 – 15, 2011), we collect data on the length (measured as the number of words) of all free and paid articles featured in the two most prominent sections of the sports' homepages (Sections "Headlines" and "Top Stories") as well as in the "Insider" section that lists a selection of paid articles. While paid articles are on average longer, the standard deviation in article length is high and more so for free articles (Table 2). This is a result of a high number of very short free articles: 10% of free but no paid articles have less than 200 words. We compare all 274 paid articles to the top 274 free articles, by number of words, and find that in this subset free articles are on average longer. This suggests that both the paid and the free section feature many detailed articles.

---

[4] This metric abstracts away from content on other websites that the sport's homepage links to, such as Twitter, and blogs that come with a different url-format. These links are always free.

Lastly, we broadly look at the type of articles that are featured in both sections. We find that the free section includes both news and editorial content (e.g., comments on a team's performance) whereas the paid section focuses on editorial content and more in-depth news reports (e.g., interview with a coach). This makes sense since readers could easily substitute news articles by an article from a competing site whereas this is more difficult for editorial content or in-depth reporting.

3.2.2 User Activity on ESPN

In our second data set, we obtain, for the same time period, daily data from Comscore on consumer activity by sport. This includes the number of unique visitors, the number of pages viewed and total time spent for both free and paid articles. We do not have access to consumer-level data. Consistent with our definition of free and paid articles we use the url-formats espn.go.com/sportname and insider.espn.go.com/sportname to identify website activities.

Comscore collets its data based on an online panel of consumers whose web activities they follow. They then weigh the individual-level observations to obtain a data set that is representative of the US population. This approach means that our data sometimes record zero visitors (mostly to the insider section or to one of the competing websites, see below) even though the true number for the US population is nonzero. Since these numbers are hard to interpret and since in our empirical estimation we take logs of the key variables, we exclude these 218 out of a total of 2,250 day-sport observations.

Panel A in Table 1 reflects that significantly more individuals visit the free section than the paid section of the site. It also illustrates that each unique visitor to the free section visits on average 5.3 pages and each visitor to the paid section visits 2.1 pages, in line with the fact that the site offers significantly more free than paid articles. The time visitors spend per page is similar across paid and free articles.

3.2.3 User Activity on Competing Sites

As a third data set, to control on a daily and per-sport level for industry-wide demand for sport news, we obtain from Comscore data per day and sport on website activities for the main competing sports website, sports.yahoo.com.[5] Yahoo offers its content for free. The data include the number of unique visitors, the number of page views and total time spent per day and sport. Table 1, Panel B documents that page views per visitor and time spent per visitor at Yahoo is comparable to those for free ESPN articles (5.5 vs 5.2 for page views per visitor; 6.2 vs 5.2 for time spent per visitor).

To further measure demand for ESPN news on a particular sport, we collect from Google Trends data on the number of searches for 'ESPN + sport' for every day in our data. We scale the data to numbers between 0 and 100.

3.2.4 Seasonalities

We next collect data on the seasons by sport and examine whether the demand for sport news varies by a sport's season. Each sport has three seasons. The off season is the period when no games are scheduled. Note that in the off season there are still sports news such as free agency signing and drafts, and scores for any pre-season games results which are not considered in the teams' final performance. The regular season is the period when scheduled games are played. Participation in these games is based on the planned schedule and so is independent of performance. During post season playoffs and

[5] Google ad planner estimates the reach of ESPN.com is 12% and that of sports.yahoo.com as 18%.

a sport's final games are played (e.g., playoff in the professional sports MLB, NBA, NFL, NHL; the bowl season for college football and March madness in college basketball). Table 3 summarizes the key variables by season.

The number of free and paid articles displayed varies more strongly within than across seasons. As would be expected, we observe a large variation in demand for articles across seasons. All our measures indicate that demand for news is lowest in off season. Average demand is similar in regular and post season.

Lastly, we collect data on sport events as possible demand controls. This includes the number of games played in each sport on each individual day, the date of the final game within each sport, for professional sports, the dates of the draft and, for college sports, college signing day. We also collect the date of the NBA lockout in the 2011 season.

3.3 Subscription and Advertising Revenues

We next describe the subscription plans that ESPN offers. We then describe in detail how we estimate an implicit average price per visit to the paid section using the weighted average subscription price and information on visitors to the paid section. In Section 5 we discuss the robustness of the underlying assumption that monthly visit frequency remains constant over time.

Customers can sign up for one of three membership plans to access paid articles. A two year membership costs $2.50 per month, a yearly membership plan charges $3.33 per month, and a monthly membership $6.95 per month. We obtain data from Comscore on the number of customers that sign up for each of the membership plans for December 2010 to December 2011. This suggests that 47% of customers choose the yearly plan, 35% choose the 2-year and 13% choose the monthly plan.[6] This gives us an average subscription revenue of $40.44 per year. Note that while our data give us reliable information about the average attractiveness of the plans, the number of individuals signing up for any plan in any month is low so we are unable to report representative data on total monthly new subscribers at ESPN.

We know that ESPN had 640,000 subscribers in 2011 (ESPN 2012) and, according to Comscore, a total of 55 million unique daily sport-visits to the paid section. This means that each subscriber returned 86.62 times a year to the paid section for each sport they visited. Note that this estimate is across sports and days, for example, visiting the NFL and NBA sport pages on the same day will count as two separate visits. Explicitly this means we have on average 86.62 day-sport visits for each subscriber in a year or 7.22 day-sport visits per month. The effective price per visit therefore amounts to $0.47 per day and sport.

ESPN features advertising on all webpages, including its homepage, the homepage for each sport and the page for each article. On each page it typically displays one ad, independently of whether an article is free or paid.[7] From Comscore we obtain estimates on ESPN's monthly advertising revenues as well as page views from December 2010 to December 2011.[8] We use this data to compute the monthly price per 1000

---

6 Additionally, 4% signed up for a holiday offer in December 2010 and 1% for a trial in October 2011.

7 We counted the number of display ads per article for MLB and NBA on a single day (June 28, 2011). On average, these articles display one ad. It is likely that revenue from sponsored links is negligible, so we do not include sponsored links. Comscore also does not provide estimates for revenues from sponsored links which further suggests that such revenues are negligible.

8 Comscore estimates are based on projected ad spend costs. This means the advertising revenue they report approximates the net advertising cost not the gross cost that is quoted on ratecards and often substantially higher. The data is predominantly inputted by agencies and so it reflects the actual payments to ESPN rather than gross pay-outs by advertisers that may include costs for agency services.

impressions. On average, ESPN's revenue per 1000 impressions is $11.51. Prices vary over time with a minimum of $8.34 (in April) and a maximum of $15.45 (in December 2011). We were able to verify the average, minimum and maximum advertising prices with ESPN.

4 Empirical Analysis

4.1 Model-Free Evidence

The key strength of our empirical setting is that it allows us to combine three types of data: First, detailed data on consumers' usage of *both* free and paid content. Second, data that capture variation in pricing – here, through the amount of paid content offered per day and sport. Third, variables that measure, and so allow us to empirically control for, industry-wide demand. Importantly, since our data is aggregate by day and sport, we observe variations in behavior both within and across time. Unlike previous research, we are therefore in the unique position that we can estimate the effect of paid content on both subscription revenues through the analysis of visits to the paid section of the site, and on advertising revenues through the analysis of page views on the site while controlling for variation in industry-wide demand. We next provide model-free evidence for the effect of paid content on consumer behavior that translates into an online content provider's revenue streams.

Theory predicts that the number of unique visitors to a paid section should increase as the number of paid articles in that section increases. This is because consumers derive utility from articles. Since consumers pay a lump-sum to visit the paid section, a consumer's net utility of a visit to the paid section on any day is the sum of the utilities from all paid articles minus the time pro-rated subscription fee. As the firm adds a paid article, the consumer's expected utility from subscribing to the paid section marginally increases. As a result, the consumer becomes more likely to subscribe. It is this marginal impact of a paid article on unique visitors to a sport on a day that we will later focus on in our estimation.

Two behavioral mechanisms may explain why a consumer's utility from subscribing may increase. First, the expected utility from visiting the paid section on any day increases in the number of paid articles offered because more content is available to view.[9] A second and complementary view is that consumers have heterogeneous preferences for articles. As the number of available articles increases, the likelihood that a consumer finds an article that fits their preferences increases. As a result, the expected utility from subscription increases. In our empirical analysis, we focus on how the number of paid articles impacts the number of unique paying visitors and so jointly capture the effect of both mechanisms.

We examine in our data the relationship between the number of paid articles on a day and sport and the number of unique visitors to the paid section for that sport that day. Figure 3, Panel A plots for any number of paid articles with at least two observations in a sport the median of normalized unique visitors to the paid section. We obtain this measure by computing the difference between the number of unique visitors on any day in a sport and the corresponding number of unique visitors on the Yahoo site that day and in that sport. We then subtract the average number of unique visitors to the paid section

[9] There are several ways by which consumers can learn about the availability of paid articles for a sport. The headline of a paid article may be featured on the firm's homepage or on the sport's homepage. Also, links to an article reported by search engines such as Google indicates whether an article is paid.

of this sport to further normalize this number by sport. As predicted, we find that unique visitors increase in the number of paid articles.

We next analyze how page views change with the number of paid articles. As each page view allows the firm to display ads, advertising revenues increase with page views. We expect that paid articles should reduce page views. First, there may be 'cannibalization' in terms of which types of articles are displayed on the sport's homepage. Second, non-subscribing customers may be upset about the sheer number (or share) of paid articles and as a result of reactance view less pages. Put differently, non-subscribers may have a disutility from paid articles that increases in the number of paid articles the firm offers. Third, paid articles affect the number of unique visitors to the site. While subscribers are more likely to visit, the utility from visiting for non-subscribers may decrease when only less appealing content is offered for free. As a result, they may visit the site less often. Since in our sample despite the high number of subscribers in absolute terms, 96% of visitors are non-subscribing and 98% of page views are in the free section, we expect the negative impact on non-subscribing consumers to dwarf a positive effect from subscribers on page views. [10]

[10] We acknowledge that there may be other long-term effects of paid articles on consumer behavior. For example, consumers may learn about the firm's policy over time and their expectations about the future number of paid articles may affect their behavior. In this research, we abstract from such possibilities.

We examine in our data whether, as theory predicts, page views fall with the number of paid articles. Figure 3, Panel B plots the corresponding relationship for paid articles and the median of total page views on the focal site for every level of paid articles, where page views are normalized by page views on the Yahoo site for that sport and by the average number of page views in that sport on our focal site.

Together, the results in Figure 3 illustrate the basic trade-off that online content providers face. As firms increase the number of paid articles on the site unique visitors to the paid section, and by consequence subscription revenues seem to increase but page views appear to decline, potentially resulting in lower advertising revenues.

We examine whether the relationship between paid articles and unique visitors to that section, respectively total page views, holds across sports' seasons. Figure 4 plots the relationship between paid articles and the median of normalized unique visitors, respectively normalized page views, for each article level by whether a sport is off season, in regular season or in post season. Again, unique visitors and page views are normalized using the corresponding Yahoo measures and the averages within a sport.

We find that the relationship between the number of paid articles and unique visitors to the paid section, respectively page views, varies across seasons. The increase of unique visitors in paid articles is weakest in the off season and regular season and most pronounced in the post season. While the slope-coefficients for off season and regular season are not significantly different, the slope-coefficient in post season is significantly different ($p<0.01$) from both other slope-coefficients.

Page views mirror this effect. The slope for the off season is not significantly different from zero. This becomes negative in the regular season, a trend which continues in the post season. Due to the large variation we observe in the post season, the slope coefficient is greater than in the regular season with a p-value of only 0.07.

We propose that the pattern we observe across seasons is a result of the variation in consumer demand and valuation for sport news. In off season, demand is generally lowest. It seems likely that only individuals with a very high valuation for sport news visit the site. The lack of any negative effect on page views could be related to individuals' greater likelihood to sign up for the paid section. Additionally, it could be a result of a lack of outside options that provide sport news such as TV.

The data in Table 3 illustrate that in regular season a large number of consumers join the site. However, relative to the off season this does not translate into an increase in visitors to the paid section, possibly because these are consumers with an, on average, lower valuation of sport news or because during regular season many additional ways are available to consume sport news, such as TV. In line with both of these interpretations, we find that during regular season page views respond more negatively to the number of paid articles the firm posts on its site. It appears that in post season, paid articles become more effective in attracting consumers, relative to either off season or regular season. This may be the case if for some consumers' the valuation of sport news strongly increases at this time when the most important games are played, e.g. whose teams are in the finals. However, the negative relationship of paid articles to page views in post season indicates that a large group of consumers still responds negatively to paid content.

We try to identify empirical regularities that shed light on how the firm allocates paid content to their website. We estimate three separate regressions with the number of paid articles as dependent variable. First, we include on the right hand side only sports' seasons as demand shifters. We find an R2 of 0.002. Second, we only include as demand shifters the variables we additionally collected: whether a day was a non-working day, whether a game was played that day, the scaled measure of Google searches, and indicators for draft, NBA lockout and whether a final game was played that day. For this model we find a R2 of 0.131. For our third estimation that uses as explanatory variables exclusively a set of sports dummies we find a R2 of 0.448. These results provide suggestive evidence that the firm varies the number of paid articles by season.

In the next section, we present a more comprehensive empirical analysis. We control for covariates to better capture the variation in the data, establish a causal relationship between the number of paid articles and unique visitors to the paid section, respectively total page views, and quantify the effect of paid articles on subscription and advertising revenues. We evaluate whether a firm would benefit from varying paid content across season and compare this to a policy that varies paid content across sports.

4.2 Average effect of paid articles

SUR: We use our data across sports and days to jointly estimate the effect of paid articles on unique visitors to the firm's site, on unique visitors to the paid section of the site and on the total number of page views on the firm's site. Our estimation is on the day-sport level.

We analyze the effect of paid articles on the three decisions a consumer makes: (1) whether to visit the firm's site, (2) whether to visit the paid section and (3) how many pages to view. We estimate a seemingly unrelated regression (SUR) to allow for a correlated error structure across all three equations.

$$
\text{(1)}\quad
\begin{aligned}
&\text{share(Firm)}_{it} = \beta_1^F PaidArt_{it} + \beta_2^F FreeArt_{it} + \beta_{3-7}^F Controls_{it} + \beta_8^F \ln\left(UniqVisYa_{it}\right) + \\
&\quad \beta_9^F Google_{it} + \beta_{10}^F RegSeason_{it} + \beta_{11}^F PostSeason_{it} + \beta_{12}^F Sport_i + \varepsilon_{it}^F \\
&\text{share}(PaidS)_{it} = \beta_1^P PaidArt_{it} + \beta_2^P FreeArt_{it} + \beta_{3-7}^P Controls_{it} + \beta_8^P \ln\left(UniqVisYa_{it}\right) + \\
&\quad \beta_9^P Google_{it} + \beta_{10}^P RegSeason_{it} + \beta_{11}^P PostSeason_{it} + \beta_{12}^P Sport_i + \varepsilon_{it}^P \\
&\text{PageViews}_{it} = \beta_1^V PaidArt_{it} + \beta_2^V FreeArt_{it} + \beta_{3-7}^V Controls_{it} + \beta_8^V \ln\left(PageViewsYa_{it}\right) + \\
&\quad \beta_9^{PV} Google_{it} + \beta_{10}^V RegSeason_{it} + \beta_{11}^V PostSeason_{it} + \beta_{12}^V Sport_i + \varepsilon_{it}^V
\end{aligned}
$$

The first part of Equation (1) captures the effect of paid articles on the overall number of visitors to the site (covariates with superscript *F*). The dependent variable share(Firm)_{it} is the log-transformation of the ratio of unique visitors to the ESPN site for a sport *i* on day *t* to the number of all US citizens that search for sport news in 2011, a total of 162,027,797. (We obtain this number by multiplying two figures, the US population in 2011 of 311,591,917, see http://www.infoplease.com/ipa/A0004986.html, and the share of the population that look for sport news online of 52%, see http://edition.cnn.com/2010/TECH/03/01/social.network.news/.) It thus represents the share of the total relevant population that ESPN attracted that day for that sport.

The second part of Equation (1) refers to visits to the paid section of the site. Here, the dependent variable, $\text{share}(PaidS)_{it}$, represents the log-transformed share of all ESPN visitors to a sport *i* on a day *t* that visited the paid section of that sport that day (covariates with superscript *P*). It thus represents unique visitors to the paid section, conditional on the number of individuals that visit the free section of the site. The third part of Equation (1) has PageViews_{it} as dependent variable. It captures the effect on the number of page views within a sport *i* on day *t* within the firm's website (covariates with superscript *V)*, including those in the free and the paid section.

Across all three equations, $PaidArt_{it}$ represents the number of paid and $FreeArt_{it}$ the number of free articles for sport *i* on day *t* (here, we omit superscripts for ease of discussion). $Controls_{it}$ represents a vector of control variables, including (1) a dummy that captures whether any games are played in sport *i* on day *t*, (2) whether during off season, there is a draft of players in a sport on a particular day (for college sports, it captures the national sign-up day), (3) a control for the NBA lockout during the 2011/12 season, (4) whether on a day *t* there was the final game in a sport *i* and (5) a variable that captures whether day *t* is a non-working day, that is a weekend day or a public holiday.

We further control for demand shocks that may not be captured by our controls so far but similarly affect all firms in the market using as controls the number of visitors to Yahoo as the major competing site for sport *i* on day *t*, $UniqVisYa_{it}$, and respectively the number of page views on Yahoo for sport *i* on day *t*, $PageViewsYa_{it}$. We additionally use data captured from Google Trends to control for demand shocks that may be unique to the focal firm, ESPN. $Google_{it}$ measures the number of Google searches for 'ESPN + sport' scaled between 0 and 100. To further account for a possible shift in demand by season, we include dummies for whether a sport is in regular season or post season. Lastly, we include fixed effects by sport.

Column (1) in Table 4 displays the results. It shows that the number of paid articles does, on average, not attract consumers to the firm's website. However, in line with our model-free analysis, paid articles are indeed effective in converting users who visit the free section into visitors to the paid section and so ultimately into subscribers. We use these two estimates to compute the overall effect of paid articles on visitors to the paid section. Based on the logit share formulation, we determine the percentage change in unique visitors to the paid section from adding a paid article as $PercChangeVis = \beta_1^F\left(1-\text{share(Firm)}_{median}\right)+\beta_1^P\left(1-\text{share(PaidS)}_{median}\right)$. This gives us that, after controlling for demand, an additional paid article increases the number of unique paying visitors by 5%. The estimates suggest that, at the median of 19,269 unique visitors to the paid section, an additional paid article increases viewership in the paid section by 991 customers.

Conversely, paid articles decrease page views. Specifically, the results suggest that an additional paid article increases page views by 0.8%. This means that at the median of 2,190,945 page views, an additional paid article decreases page views by 18,108. This result illustrates the trade-off apparent from our earlier descriptive analysis that paid content may attract subscription revenues but decreases page views, and thus ultimately advertising revenues.

Our results likewise suggest that free articles fulfill three roles: First, they attract consumers to the firm's website. Second, they increase the number of page views on the firm's site. Third, the results indicate that free articles may convert consumers into subscribers, potentially because sampling high-quality free content increases the probability that a consumer will sign up for the paid section.

Accounting for endogeneity through 3SLS: A key strength of our previous specification is that it controls for a wide range of demand shifters. For example, increased demand on days on which games are played would be captured by the control for gameday. Or, if people tend to watch more TV and consume less online sport news on weekends, then this should be captured by the nonworking-day control. Alternatively, if there is a piece of unexpected sport news that generally affects demand, such an effect would be captured by the activity on the Yahoo site whereas shocks that would affect the attractiveness of the ESPN site only should be captured by our variable $Google_{it}$ derived from Google Trends.

However, even after controlling for a wide range of observable demand shocks there could possibly be demand shocks that ESPN observes but not the researcher. ESPN could then use this information when deciding on the number of free or paid articles on that day. An example could be a breaking news story that is unique to ESPN (e.g. ESPN signing a new Monday Night Football deal with the NFL on September 8, 2011). Note that anecdotal evidence suggests that rather than knowing the revenue-optimizing paywall, firms experiment with respect to their paid content strategy.[11] Nonetheless, we turn to an instrumental variable estimation to control for such possible endogeneity. We use as an instrument the number of free and paid articles that ESPN displayed the previous day that is the day before such news were known.

Our instrumenting strategy builds on the insight that on any given day, the firm does not update the full set of articles it displays for any sport but instead retains a subset of articles that were displayed the previous day. On average across days and sports, 75% of all paid and 61% of all free articles were retained from the previous day. Displaying an article for more than a single day makes sense as long as potential readers do not visit every day. Indeed our data indicate that customers visit the firm's website on average every 4.2 days, meaning that an article initially displayed the previous day will still be of

11 The design of the New York Times paywall seems to be based more on trial-and-error than robust optimization (http://www.poynter.org/latest-news/mediawire/167147/changes-to-new-york-times-

interest to many customers visiting today. Continuing to display an existing article is attractive for the firm since it incurs zero marginal cost of production on the second day.

In the case of the ESPN-NFL deal on September 8, we would use the number of free and paid articles displayed on September 7 to instrument for the number of free and paid articles on September 8, 2011. Indeed, while ESPN featured an article on the deal on September 8 there was no such report on the day before the deal was announced.

Using as instrument for the number of paid and free articles on any day the number of paid and free articles displayed the previous day means we assume that, after controlling for the extensive set of our demand shifters, yesterday's free and paid articles affect unique visitors and page views today only through the number of articles today and not in any other way. This assumption would be problematic if the firm would not immediately publish a new article but delay publishing until the next day, possibly because of anticipated demand that day. For example it would be a problem, if, hypothetically, ESPN would expect greater traffic to its site on September 9 and hold back reporting the news for a day. However, the market for online news is highly competitive and, by its nature, competes on real-time information. As a result delaying news does not seem a likely strategy.[12] Note that our instrumenting strategy also assumes that the firm is myopic, meaning that the decision whether or not to publish an article

---

paywall/). The examples provided in the introduction further suggest that firms as of yet are not necessarily aware of the optimal paid content strategy.

12 Articles forwarded through email or social media are likely consumed by recipients on the same day. However, we cannot rule out that consumption may occur the next day. Then, our identifying assumption that yesterday's articles affect unique visitors and page views today only through the number of articles today would not be valid. We checked whether this affects the results. We rerun the by-season specification that we discuss in Section 4.3 and include as additional control unique visitors to the paid section the previous day, respectively page views the previous day. The coefficients are broadly similar in sign, significance and size to those that we present in Table 6.

may be affected by the expected demand on the same day but not by expected demand on the following day.

We re-estimate our previous specification accounting for the possible endogeneity of the number of paid and free articles, using as instruments the lagged number of paid and free articles in that sport on that day.

Column (2) in Table 4 displays the results of the 3SLS estimation with endogenous explanatory variables. Similarly to before, we find that paid articles do not affect the number of visitors to the site but do significantly shift the number of visitors to the paid section. This effect is slightly smaller in size than when not accounting for the possible endogeneity of articles but our main effect still holds. Specifically, the estimate now suggests that, calculated at the median value, an additional paid article increases viewership in the paid section by 766 unique visitors. Similarly to before, paid articles significantly reduce page views on the site overall, here by 24,394. In a further unreported specification we find that our results are robust to including as additional instrument the age of the previous day's free and paid articles.

We confirm in three separate instrumental variable regressions that the number of free and paid articles is endogenous. Here, the dependent variable are either the share of individuals visiting the firm's site, the share of visitors going on to the paid section or page views. In line with our 3SLS estimation, we use as instruments the number of paid and free articles on the previous day as instruments. In all three estimations, an F-test of the significance of excluded instruments strongly rejects zero ($p<0.001$). The Kleibergen-Paap rk statistic suggests that we can reject the hypothesis that the first stage is underidentified ($p<0.001$). Our instruments are therefore good predictors of the number of free and paid articles.

Effect on the firm's revenue: We evaluate whether offering an additional article is profitable for the firm. We use several data points in addition to our parameter estimates. First, we use the fact that alongside each article the firm displays on average one ad, meaning that each page view leads to one ad impression. Second, as laid out in Section 3.3 we compute the weighted average of subscription fees across plans as $3.37 per month. Third, as illustrated in Section 3.3, we rely on the fact that $0.47 revenue can be attributed, on average, to a visit to the paid section.

Note that this analysis translates unique visitors into subscribers holding constant unique monthly visits per subscriber and so the implicit price per visit. We acknowledge that the increase in unique visitors we observe as a result of an increase in paid articles might be due to a change in visit frequency by the same set of subscribers rather than to an increase in subscribers. Section 5 will examine this possibility.

We use our estimates to compute the financial impact for the firm of adding one marginal paid article. To do so, we bootstrap from the distributions of the three key coefficients which are the effect of paid articles on Share ESPN, on Share Paid Section and on Page Views. For each day-sport observation, we take 10,000 draws from these distributions.

From the two share regressions, we estimate the derivative of the share of visitors to the paid section with respect to the number of paid articles. This means, we follow the same log transformation as initially described in Section 4.2 but now do this on the level of each day-sport observation. Multiplying this with market size, we obtain the estimated marginal increase in paid visitors for that day-sport for a particular draw. We then multiply the increase in visitors with the revenue per visit that is \$0.47. This gives us the additional subscription revenue the firm could earn from offering one more insider article. Note that in our analysis we assume that quality of articles remains constant for a marginal paid article.

To understand the impact on advertising revenues, we multiply the derivative of the page views with respect to number of paid articles and the price per 1000 page views.[13] We initially focus on the average price of \$11.51 but later check the robustness of our results to the price of \$8.34 and of \$15.45 per 1000 impressions, the minimum and the maximum monthly ad prices during our observation period.

For each day and sport, we compute the derivative of the total revenue (subscription plus advertising) to ESPN with respect to the number of paid article for every draw and every day-sport. In Figure 4 we plot the distribution of the marginal revenue impact of adding one paid article at the average price per 1000 impressions of \$11.51. In Table 5, we present the median across draws per day-sport observation and the percentage of days with statistically significant negative or positive impact marginal revenue.

We focus in our discussion on the results of the specification that accounts for the endogeneity in the number of paid and free articles. Both Figure 4 and Table 5 illustrate that there is a large variation across the data. Specifically, they show that on average only

[13] We assume that the firm charges the same price for a page view in the free and the paid section. We cannot conclusively rule out that ESPN charges a higher price for advertising in the paid section. But since ESPN already offers access to a highly targeted audience it is unlikely they would charge a significant premium for access to subscribers. Note also that an advertiser's willingness to pay is often lower for smaller audiences (Athey et al. 2011), suggesting further that the firm would not be able to charge a premium to advertisers in the paid section.

a small percentage of day-sport observations is significantly different from zero. These data also show that, on average, adding a paid article would reduce revenues by $129. We check the robustness of the results to advertising prices per 1000 impressions of $8.34 and $15.45. We find that broadly our results hold. They are likewise similar when we estimate the monetary value of adding a paid article based on the specification that does not account for endogeneity of paid articles.

In sum, these results demonstrate that, on average, the firm should not change the number of paid articles displayed.

4.3 Exogenous variation of demand

Our descriptive analysis indicates that consumer response to sport news varies across a sport's season, with potentially more pronounced responses in regular and post season than in off season. We next empirically tease apart the effect of paid articles across seasons.

Similarly to the previous section, we estimate the effect of paid articles on the number of visitors to the firm, to the paid section and on total page views in a 3SLS framework but allow the effect of free and paid articles to vary by season. In Table 6, Column (1) reports the results in a specification where the number of articles is treated as exogenous. In Column (2) we again report estimation results where we treat the number of paid and free articles as endogeneous. We use as instruments the number of paid and free articles on the previous day by season. The two set of estimation results look broadly similar in direction and significance of their coefficients.

We again confirm that the number of free and paid articles is endogenous in this regression. Similarly to before, we test this separately for each of the main dependent variables, Share ESPN, Share Paid Section and Page Views using an instrumental variable regression with the number of paid and free articles on the previous day by season as instruments. For all three equations, an F-test of the significance of excluded instruments strongly rejects zero ($p<0.001$). The Kleibergen-Paap rk statistic suggests that we can reject the hypothesis that the first stage is underidentified ($p<0.001$). Our instruments are therefore good predictors of the number of free and paid articles.

We focus our discussion on the 3SLS results and so account for the endogeneity of paid and free articles. Column (2) of Table 6 demonstrates in detail the role of paid and free articles for the firm. Based on the coefficients that measure how paid articles affect the share of consumers that visit the firm's site and the share of visitors that go on to the paid section, we compute the aggregate effect of paid articles on unique visitors to the site by season. Similarly to before, this is based on the logit share formulation. We find that paid articles increase the share of visitors that come to the paid section by 5% in off season, by 3% in regular season and by 7% in post season. At the median number of unique visitors per season, this translates into an increase in 882 visitors in the off season, 573 visitors in the regular season and 1,307 in the post season. This variation across seasons broadly reflects our earlier analysis in Figure 3, Panel B which demonstrated that the increase of unique visitors to the paid section is strongest during post season.

Conversely, the effect of paid articles on page views turns more negative as we move from off season to regular and post season. Specifically, one additional paid article decreases page views by 3.1% (6.0%) in the regular (post) season. At a median of 5,084,165 (3,858,686) page views this is a reduction by 157,582 (232,721) page views in the regular (post) season. At an average price per 1,000 impressions of $11.51, this translates into a loss of advertising revenues of $1,814 in the regular and $2,679 in the post season. Importantly, these results are again consistent with Figure 3, Panel B which demonstrated an increasingly negative effect of paid articles on page views as we move from regular season to post season. Note that we cannot conclusively rule out that article quality might change across season. But this seems unlikely given that, as Table 3 demonstrates, the firm does not seem to be strategic in managing article quantity across seasons.

We likewise examine the role of free articles. The first part of Table 6 that focuses on Share ESPN clearly illustrates that free articles are effective in attracting visitors to the site. This is most effective in the off season, possibly because visitation during regular and post season is less driven by the number of articles and more by recent events. Interestingly, only during regular season are paid articles effective in converting consumers into paying visitors.

Effect on the firm's revenue: We compute the effect on the firm's revenue separately for off season, regular season and post season. We follow the same approach as outlined in Section 4.2. Table 7 displays the median across all draws of the financial impact in US dollars of an additional paid article for each day-sport observations in our data as well as the percentage of day-sport observations that are positive or negative at a 5% significance level. The results for the SUR and the 3SLS estimation are largely consistent. They demonstrate that during off season the firm would clearly benefit from increasing the number of paid articles with a revenue increase per paid article between $475 and $630 per day and sport. Across all observations and advertising prices, the vast majority of observations show a positive impact of paid articles. The first panels of

Figure 6 and Figure 7 plot the histogram of all medians of the day-sport observations, based on the SUR and 3SLS estimations respectively for the post season. Again, this confirms that the large majority of observations are positive.

In regular season, however, the effect of an additional paid article is largely negative, on average between $628 and $1952. For the results based on the 3SLS estimation and price levels of $11.51 and $15.45, we find that for at least 90% of day-sport observations, the firm should decrease the number of paid articles. The histograms in the second panels of Figure 6 and Figure 7 again confirm this. This means that during regular season the firm mostly benefits from decreasing the number of paid articles.

In post season, this pattern is less clear. For the average price for advertising, we find a negative impact of adding a marginal article. However, reducing the number of articles benefits the firm on only 63% of day-sport observations. Our earlier coefficients suggest that this is a result of an increase in the effectiveness of paid articles in converting visitors to the firm's website into subscribers during post season.

Given that 46% of observations in our data lie in the off season and 44% in the regular season, our results mean that on roughly half of the days the firm should increase and on roughly half of the days it should decrease the number of paid articles. As such our results indicate that accounting for variation across time is important in optimizing a firm's revenue streams from offering content online.

We check whether our results are robust to an alternative specification where we measure the effect of the percentage of paid articles, instead of the number of paid and free articles, on our outcome variables. We use as instruments the percentage of paid articles on the previous day. We find that our results hold: In the off season the firm should increase the number of paid articles on 96% of days whereas in the regular season the firm should decrease the number of paid articles by one on 99% of days.

5 Robustness checks

Translating unique visitors to the paid section into subscribers: Throughout our empirical analysis we translate unique visitors into subscribers holding constant the number of unique monthly visits. It is, however, possible that the increase in unique visitors we observe as a result of an increase in paid articles is not due to an increase in subscribers but instead to a change in the visit frequency by the same set of subscribers. Using additional data provided by Comscore, we conduct two analyses that support our assumption that an increase in daily unique visitors indeed translates into a greater number of subscribers.

First, we focus on unique monthly visitors. The data presented in Table 3 suggest that the number of unique visitors on any day is greatest during post season. If this effect is indeed due to an increase in subscribers, then the result should hold when using more aggregate monthly data. We use data on unique monthly visitors to the paid section by sport. Here, an individual is counted once per month and sport, independently of how often they visit the site during that month. In a linear regression, we regress season dummies on monthly unique visitors by sport controlling for sports. We find a positive and significant effect of post season (p=0.002) on unique monthly visitors and insignificant effects for off season and regular season. This confirms that variation in the number of unique daily visitors that we observe across seasons translates into a variation of monthly visitors and so most likely can be traced back to an increase in subscribers.

Second, we use data on a consumer's average number of visits per sport and month. We estimate a linear regression with the average visit frequency as dependent and unique monthly visitors per sport as independent variable, controlling for sports. Our results suggest that visit frequency does not vary with unique monthly visitors (coefficient -0.000019, p=0.889). This result further supports that the variation in unique daily visitors we observe indeed comes from an increase in subscribers.

Variation of revenue effects by sport: For the post season, our analysis in Table 7 and Figure 7 indicate a great variation in the revenue effects. We ask whether this variation in revenue effects might be due to variation across sports. We therefore redo the analysis of revenue underlying Figure 7 by season and sport. We omit displaying the results for the off season as they consistently reflect the pattern displayed in Figure 7 with 100% of observations being positive.

Figure 8 displays the results for the regular season by sport. It illustrates that the results hold by sport. Across all sports, the large majority of observations is positive, meaning that independently of sports the firm should decrease the number of paid articles in regular season. Figure 9 illustrates variation across sports in the post season. Specifically, the results indicate that the firm would benefit from decreasing the number of paid articles in post season for all sports, except for NFL and College Football.

6 Conclusion

The last decade has seen many media companies, such as newspapers, struggle and it is generally believed that their future hinges on their ability to implement a sustainable revenue model online. However, solving the basic trade-off that lies in gaining subscription revenues by offering paid content at the cost of even lower advertising revenues is not obvious.

In this research, we empirically examine and quantify a content provider's trade-off between advertising and subscription revenues. We evaluate whether a firm should follow a static policy or flexibly adjust the amount of paid content it offers, and how such variation should be implemented.

We build a unique data set from the sports website of ESPN where we combine data on the number of free and paid articles offered per sport over time with metrics of consumer demand such as unique visitors and page views, for both free and paid articles. We also control for industry-wide demand by tracking usage at the major competitive websites. We estimate how the number of free and paid articles affects viewership of the site. We empirically quantify the impact of the number of paid articles on the increase in the number of subscribers, and the decrease in total page views and evaluate whether the company would benefit from adding paid content.

Our results suggest that on average the marginal increase in subscription revenue is statistically indistinguishable from the marginal decrease in advertising revenue. Thus, on average, the firm should not adjust the amount of paid content. However, we find strong differences when we allow our results to vary by an indicator of exogenous demand variation. Specifically, we find that the marginal paid article increases revenues in the off season but decreases revenue in regular season. For the post season, we further find that the effectiveness of paid content varies across sports.

In sum, our findings suggest that adding paid content may sometimes – but not always – be good for firms. We conclude from this that many online content providers that have recently experimented with fee models but typically use a static policy may benefit from re-evaluating their pricing strategy to flexibly respond to consumer demand.

Our results also suggest that firms should carefully identify the key dimensions along which demand varies. We find that variation in demand over time can be an important factor – possibly more so than variation in demand over different types of subject matters. While in our setting, variation in demand over time arises from sports' seasonalities, demand shocks can tilt the balance between advertising and subscription revenues in many other instances. Traffic on a general news site might greatly increase on election days because customers who typically do not consume political news – and would not sign up for paid content – like to visit. In fact, both the New York Times and the Wall Street Journal lifted their paywalls during the 2012 election. Similarly, following the 2013 bombings in Boston the Boston Globe temporarily lifted its paywall. Our results suggest that the sites may have concluded that additional advertising revenues from the sudden influx of low-valuation customers (who would not pay a subscription) may outweigh loss of subscription revenues.

One implication of our findings is that online content providers can greatly benefit from investing into data analytics to identify in real-time periods of unusually high (or low) demand to then respond by adjusting the amount of paid content. Such analytics may partly happen on consumer-level data, for example, when the firm can track behavior of subscribers across their site. However, often firms may not be able to track individual consumers: consistently identifying nonsubscribers through cookies may be difficult if consumers do not allow cookies, cookies expire or consumers use multiple devices. Our approach illustrates how firms can use aggregate data to align their paid content offering with consumer demand.

Of course there are limitations to our work. Our study focuses on the immediate, short-term effects from offering paid content. There may be additional, long-term effects that we are not able to account for. Further, we do not analyze the effect of differences in article quality. Future research on optimal pricing strategies for online content providers could examine how willingness-to-pay varies with differences in article type or quality. Lastly, our study is set in an industry where many firms (still) offer all content for free. It is possible that in settings were all or most competitors charge for access to their content a subscription model may more generally appear to be optimal. This would then raise a new set of questions, such as how consumers trade off between fee-paying online sites and other media, e.g. cable or satellite television. We leave such questions for future research.

Bibliography

Angelucci, Charles, Julia Cage, and Romain De Nijs (2013), "Price Discrimination in a Two-Sided Market: Theory and Evidence from the Newspaper Industry," Working paper.

Ascarza, Eva, Anja Lambrecht, and N.J. Vilcassim (2012), "When Talk is "Free": An Analysis of Subscriber Behavior under Two- and Three-Part Tariffs," Journal of Marketing Research, 49 (6), 882-99.

Athey, Susan, Emilio Calvano, and Joshua S. Gans (2011), "The Impact of the Internet on Advertising Markets for News Media," Working paper.

Bhargava, Hemant K. and Vidyanand Choudhary (2001), "Information Goods and Vertical Differentiation," Journal of Management Information Systems, 18 (2), 89-106.

Chiou, Lesley and Catherine Tucker (2012), "Paywalls and the Demand for News," Information Economics and Policy, forthcoming.

ESPN (2012), "ESPN Digital Media Breaks Ground in 2011," Press release (http://espnmediazone.com/us/press-releases/2012/01/espn-digital-media-breaks-ground-in-2011/).

Godes, David, Elie Ofek, and Miklos Sarvary (2009), "Content vs. Advertising: The Impact of Competition on Media Firm Strategy," Marketing Science, 28 (1), 20-35.

Halbheer, Daniel, Florian Stahl, Oded Koenigsberg, and Donald R. Lehmann (2013), "Digital Content Strategies," Working paper.

Hirschorn, Michael (2009), "End Times," The Atlantic, January/February 2009.

Kumar, Vineet, Bharat Anand, Sunil Gupta, and Felix Overholzer-Gee (2013), "The New York Times Paywall," Harvard Business School Case.

Lambrecht, Anja, Avi Goldfarb, Alessandro Bonatti, Anindya Ghose, Dan Goldstein, Randall Lewis, Anita Rao, Navdeep Sahni, and Song Yao (2013), "How Do Firms Make Money Online?," Working paper.

Pattabhiramaiah, Adithya, S. Sriram, and Shrihari Sridhar (2013), "Rising Prices under Declining Preferences: The case of the U.S. Print Newspaper Industry," Working paper.

Pauwels, Koen and Allen Weiss (2007), "Moving From Free to Fee: How Marketing Can Stimulate Gains and Stem Losses for an Online Content Provider," Journal of Marketing, 72 (3).

Prasad, Ashutosh, Vijay Mahajan, and Bart Bronnenberg (2003), "Advertising versus pay-per-view in electronic media," International Journal of Research in Marketing, 20, 13-30.

Seamans, Robert and Feng Zhu (2013), "Responses to Entry in Multi-Sided Markets: The Impact of Craigslist on Local Newspapers," Working paper.

Shampanier, Kristina, Nina Mazar, and Dan Ariely (2007), "Zero as a Special Price: The True Value of Free Products," Marketing Science, 26 (6), 745-57.

Shapiro, Carl and Hal H. Varian (1998), "Versioning: The Smart Way to Sell Information," Harvard Business Review, 76(Nov/Dec), 106-14.

Table 1: Articles and Activity on ESPN and Yahoo

			Percentiles		
Panel A: ESPN	Mean	Std. Dev.	5th	Median	95th
Articles					
All	33.8	17.8	15	30	70
Free	24.8	17.3	10	19	62
Paid	9.0	3.8	4	8	17
Unique Visitors					
All pages	648,713	504,733	107,745	490,503	1,564,764
Paid pages	28,309	42,819	1,504	19,269	78,292
Page views					
All pages	3,665,127	3,873,991	391,988	2,190,945	10,663,158
Free pages	3,596,655	3,852,824	361,481	2,126,869	10,602,879
Paid pages	68,472	159,223	1,871	34,319	225,185
Page views per unique visitor					
All pages	5.2	2.4	2.6	4.5	9.7
Paid pages	2.1	1.7	1.1	1.6	4.4
Time spent (min)					
All pages	3,583,409	3,991,603	289,428	2,032,662	10,862,223
Free pages	3,528,311	3,973,873	272,430	1,985,387	10,772,211
Paid pages	55,098	137,155	772	26,303	183,661
Time spent per page (min)					
All pages	0.9	0.3	0.6	0.9	1.4
Free pages	0.9	0.3	0.6	0.9	1.4
Paid pages	0.8	0.5	0.2	0.8	1.8
Time spent per unique visitor					
Free pages	5.2	3.4	2.3	4.3	10.8
Paid pages	1.7	1.4	0.3	1.4	3.7
Panel B: Yahoo					
Unique Visitors	879,752	989,628	54,791	562,754	2,826,432
Page views	4,451,955	5,244,159	285,452	2,231,225	14,160,282
Page views per unique visitor	5.5	3.3	2.2	4.6	12.2
Time spent (min)	5,073,084	6,404,012	209,335	2,429,279	16,166,617
Time spent per unique visitor	6.2	5.3	2.1	4.9	14.1

N=2032

Table 2: Length of Free and Paid Articles

	Mean (word count)	Std.dev. (word count)	N
Articles overall			
Free	965	837	824
Paid	1332	654	274
Top 274 per type (by length)			
Free	1832	921	274
Paid	1332	654	274
Category: Top Stories			
Free	1392	980	402
Paid	1241	538	139
Category: Headlines			
Free	615	404	481
Paid	1561	1047	46
Category: Insider			
Paid	1404	587	148

Note: Word counts for 11/9 - 11/15; sometimes articles are listed in more than one category.

Table 3: Data by Sports' Seasons

	Off season			Regular season			Post season		
	Median	Mean	Std. dev.	Median	Mean	Std. dev.	Median	Mean	Std. dev.
Number of articles									
All	24	25.6	9.5	34	40.9	19.6	35	40.1	22.2
Free	16	16.6	8.0	25	31.8	19.5	25	31.6	21.2
Paid	8	8.9	4.3	9	9.1	3.4	9	8.5	3.1
Unique visitors									
ESPN all pages	299,375	362,540	282,980	884,217	899,314	542,883	851,424	856,498	434,384
ESPN paid pages	18,975	30,060	56,203	18,817	24,588	22,822	23,792	36,363	37,386
Yahoo	259,349	534,452	752,980	910,998	1,163,744	1,102,731	963,030	1,208,856	912,929
Page views									
ESPN all pages	1,134,415	1,522,005	1,663,506	5,084,165	5,765,398	4,511,573	3,858,686	4,268,230	2,687,423
ESPN free pages	1,063,301	1,437,158	1,550,101	4,986,436	5,719,069	4,497,247	3,702,846	4,178,381	2,667,945
ESPN paid pages	36,446	84,847	220,913	30,697	46,329	52,391	45,585	89,849	116,629
Yahoo	1,048,170	1,933,525	2,667,499	5,531,379	6,835,527	6,278,702	4,911,359	5,520,920	3,754,829
N	932			891			209		

Table 4: Effect of Paid Articles on Visitors and Page Views

	(1) OLS			(2) IV		
	Coef.	Std. Err.		Coef.	Std. Err.	
Share ESPN						
Free Articles	0.003	0.001	***	0.006	0.001	***
Paid Articles	0.003	0.003		0.002	0.003	
Regular Season	0.453	0.025	***	0.425	0.025	***
Post Season	0.396	0.031	***	0.373	0.032	***
Nonworkingday	-0.240	0.017	***	-0.248	0.017	***
Gameday	0.072	0.025	***	0.046	0.026	*
Googlescaled	0.023	0.001	***	0.024	0.001	***
ln(Yahoo Unique Vis	0.165	0.009	***	0.161	0.009	***
Draft	-0.075	0.100		-0.090	0.106	
Lockout	-0.007	0.039		0.034	0.040	
Final Game	-0.165	0.148		-0.199	0.149	
Constant	-8.643	0.115	***	-8.662	0.119	***
Share Paid Section						
Free Articles	0.003	0.002	**	0.008	0.002	***
Paid Articles	0.050	0.008	***	0.040	0.009	***
Regular Season	-0.817	0.074	***	-0.834	0.076	***
Post Season	-0.510	0.094	***	-0.513	0.095	***
Nonworkingday	-0.079	0.050		-0.085	0.050	*
Gameday	-0.115	0.075		-0.158	0.078	**
Googlescaled	-0.006	0.003	**	-0.006	0.003	**
ln(Yahoo Unique Vis	-0.046	0.029		-0.054	0.030	*
Draft	1.251	0.299	***	1.319	0.315	***
Lockout	-0.637	0.117	***	-0.602	0.119	***
Final Game	0.079	0.442		0.005	0.443	
Constant	-2.853	0.390	***	-2.752	0.399	***
ln(Page Views)						
Free Articles	0.002	0.001	**	0.004	0.001	***
Paid Articles	-0.008	0.003	**	-0.011	0.004	***
Regular Season	0.588	0.033	***	0.566	0.033	***
Post Season	0.349	0.041	***	0.333	0.041	***
Nonworkingday	-0.268	0.022	***	-0.270	0.022	***
Gameday	0.172	0.033	***	0.163	0.034	***
Googlescaled	0.033	0.001	***	0.033	0.001	***
ln(Yahoo Unique Vis	0.186	0.013	***	0.187	0.014	***
Draft	0.052	0.132		0.091	0.138	
Lockout	0.123	0.051	**	0.149	0.052	***
Final Game	-0.408	0.194	**	-0.438	0.194	**
Constant	11.224	0.194	***	11.184	0.202	***
N		2032			2007	
R-2 Share ESPN		0.853			0.851	
R-2 Share Paid Section		0.300			0.297	
R-2 ln(Page Views)		0.832			0.832	
Significance of first-stage regressions					significant at 0.001	
First stage R-2					0.65 - 0.84	

Fixed effects by sport included but not displayed for readability. *** p<0.01, **p<0.05, *p<0.1

Table 5: Percent of Observations with Positive/Negative Significant Effect

Price per 1000 impressions	Statistic	SUR, no endogeneity	3SLS, with endogeneity
8.34	Median $ per day	$787	$162
	% days sig (5%) negative	4%	13%
	% days sig (5%) positive	51%	32%
11.51	Median $ per day	$488	($129)
	% days sig (5%) negative	7%	19%
	% days sig (5%) positive	42%	22%
15.45	Median $ per day	$141	($455)
	% days sig (5%) negative	10%	25%
	% days sig (5%) positive	34%	15%

Table 6: Effect of Paid Articles on Visitors and Page Views by Season

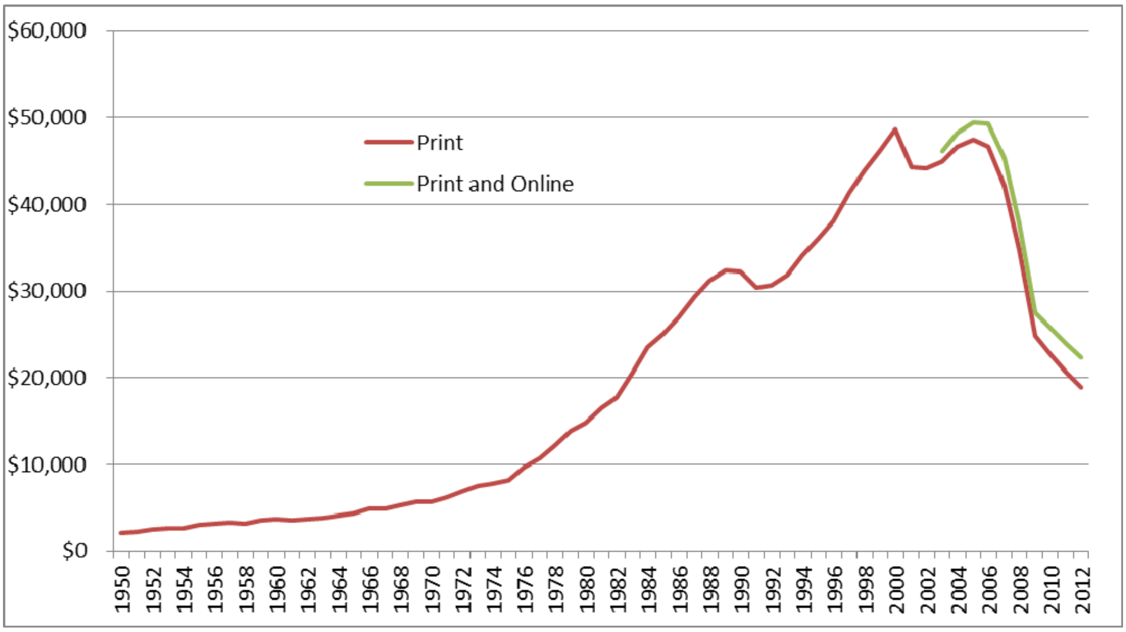
	(1) OLS			(2) IV		
	Coef.	Std. Err.		Coef.	Std. Err.	
Share ESPN						
Free Articles -Off Season	0.010	0.001	***	0.010	0.001	***
Free Articles -Regular Season	0.002	0.001	***	0.006	0.001	***
Free Articles -Post Season	0.004	0.001	***	0.005	0.001	***
Paid Articles -Off Season	0.018	0.003	***	0.018	0.004	***
Paid Articles -Regular Season	-0.016	0.004	***	-0.019	0.004	***
Paid Articles -Post Season	-0.025	0.008	***	-0.034	0.010	***
Regular Season	0.917	0.052	***	0.849	0.062	***
Post Season	0.849	0.078	***	0.940	0.087	***
Nonworkingday	-0.242	0.016	***	-0.254	0.017	***
Gameday	0.046	0.025	*	0.015	0.026	
Googlescaled	0.025	0.001	***	0.026	0.001	***
ln(Yahoo Unique Visitors)	0.158	0.008	***	0.154	0.009	***
Draft	-0.067	0.098		-0.102	0.104	
Lockout	-0.029	0.039		0.033	0.042	
Final Game	-0.217	0.145		-0.247	0.147	*
Constant	-8.787	0.114	***	-8.799	0.118	***
Share Paid Section						
Free Articles -Off Season	0.000	0.004		0.002	0.005	
Free Articles -Regular Season	0.004	0.002	**	0.011	0.003	***
Free Articles -Post Season	0.003	0.003		0.005	0.004	
Paid Articles -Off Season	0.037	0.010	***	0.031	0.011	***
Paid Articles -Regular Season	0.060	0.011	***	0.050	0.013	***
Paid Articles -Post Season	0.121	0.024	***	0.107	0.029	***
Regular Season	-1.106	0.160	***	-1.205	0.188	***
Post Season	-1.268	0.238	***	-1.164	0.264	***
Nonworkingday	-0.071	0.050		-0.087	0.051	*
Gameday	-0.091	0.076		-0.152	0.081	*
Googlescaled	-0.008	0.003	**	-0.007	0.003	**
ln(Yahoo Unique Visitors)	-0.038	0.029		-0.048	0.030	
Draft	1.266	0.299	***	1.306	0.316	***
Lockout	-0.623	0.119	***	-0.537	0.128	***
Final Game	0.207	0.444		0.157	0.448	
Constant	-2.778	0.391	***	-2.701	0.400	***
ln(Page Views)						
Free Articles -Off Season	0.009	0.002	***	0.009	0.002	***
Free Articles -Regular Season	0.000	0.001		0.003	0.002	*
Free Articles -Post Season	0.000	0.001		0.002	0.002	
Paid Articles -Off Season	0.009	0.004	**	0.006	0.005	
Paid Articles -Regular Season	-0.029	0.005	***	-0.031	0.006	***
Paid Articles -Post Season	-0.040	0.011	***	-0.060	0.013	***
Regular Season	1.093	0.069	***	1.022	0.080	***
Post Season	0.923	0.102	***	1.051	0.113	***
Nonworkingday	-0.271	0.021	***	-0.276	0.022	***
Gameday	0.147	0.033	***	0.135	0.035	***
Googlescaled	0.034	0.001	***	0.035	0.001	***
ln(Yahoo Page Views)	0.178	0.013	***	0.179	0.014	***
Draft	0.055	0.130		0.077	0.136	
Lockout	0.107	0.052	**	0.141	0.055	**
Final Game	-0.451	0.192	**	-0.507	0.193	***
Constant	11.054	0.194	***	11.036	0.202	***
N		2032			2007	
R-2 Share ESPN		0.861			0.858	
R-2 Share Paid Section		0.305			0.299	
R-2 ln(Page Views)		0.839			0.838	
Significance of first-stage regressions					significant at 0.001	
First stage R-2					0.77 - 0.98	

Fixed effects by sport included but not displayed for readability. *** p<0.01, **p<0.05, *p<0.1

Table 7: Percent of Observations with Positive/Negative Significant Effect by Season

Price per 1000 impressions	Statistic	SUR no endogeneity Season Off	SUR no endogeneity Season Regular	SUR no endogeneity Season Post	3SLS with endogeneity Season Off	3SLS with endogeneity Season Regular	3SLS with endogeneity Season Post
8.34	Median $ per day	$550	($628)	($137)	$475	($857)	($754)
	% days sig (5%) negative	0%	73%	27%	0%	82%	52%
	% days sig (5%) positive	100%	1%	13%	97%	0%	0%
11.51	Median $ per day	$590	($1,050)	($444)	$499	($1,338)	($1,421)
	% days sig (5%) negative	0%	84%	38%	0%	90%	63%
	% days sig (5%) positive	100%	0%	5%	95%	0%	0%
15.45	Median $ per day	$630	($1,591)	($775)	$533	($1,952)	($2,137)
	% days sig (5%) negative	0%	92%	49%	0%	95%	72%
	% days sig (5%) positive	100%	0%	2%	93%	0%	0%

Figure 1: US Newspaper Advertising Revenue, 1950 – 2012



Source: Newspaper Association of America.

Figure 2: Screenshot of ESPN website Displaying Insider-icon

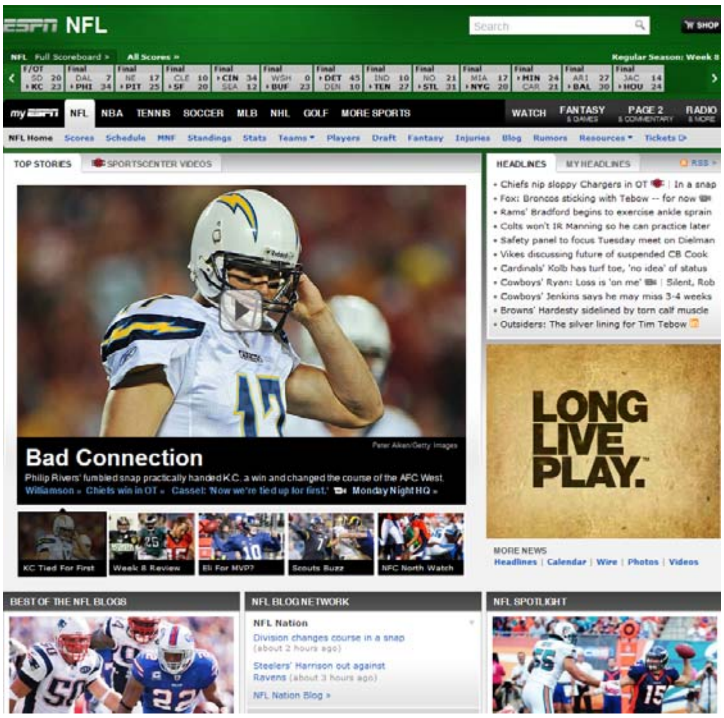


Figure 3: Relationship of Paid Articles to Unique Visitors to the Paid Section and to Total Page Views

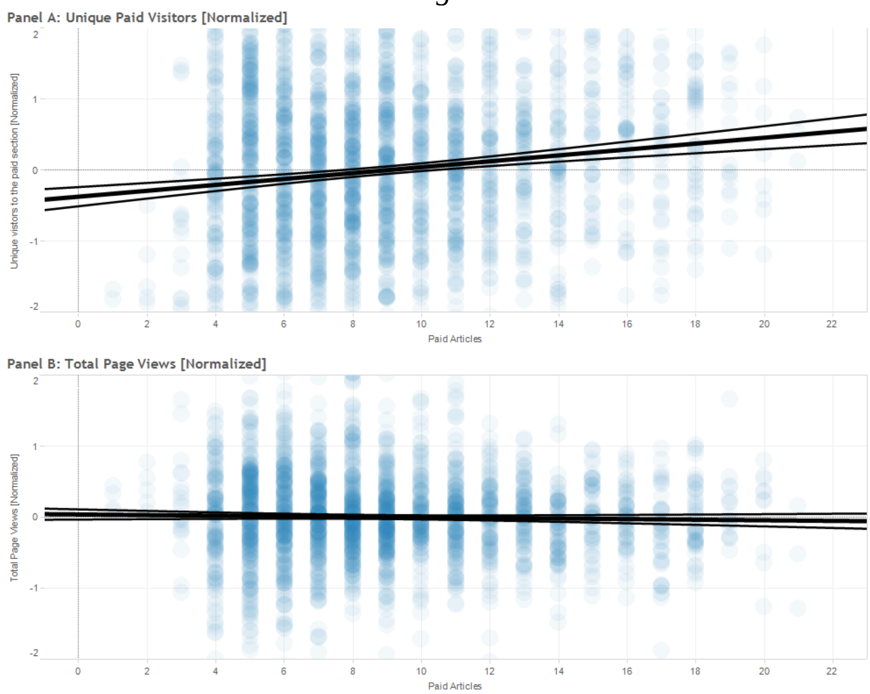


Figure 4: Relationship of Paid Articles to Unique Visitors to the Paid Section and to Total Page Views, By Season

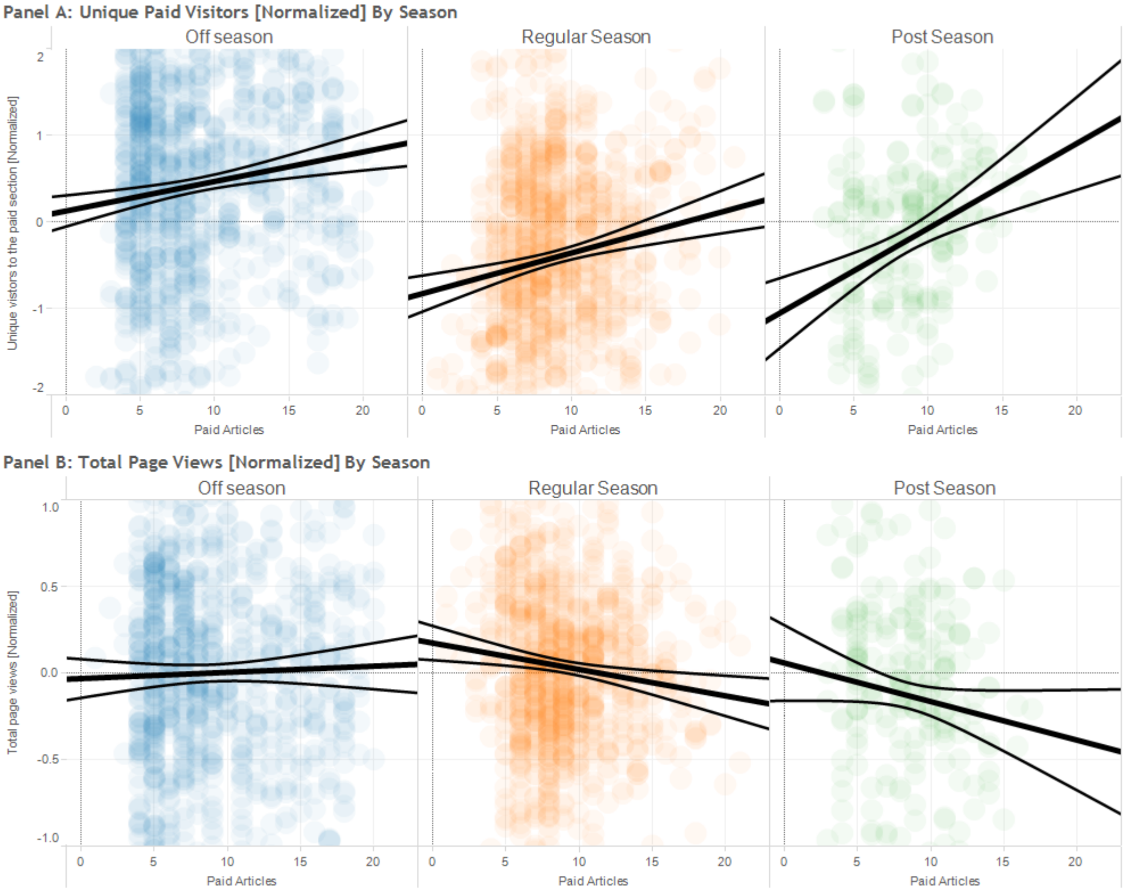


Figure 5: Dollar Effect of an Additional Paid Article

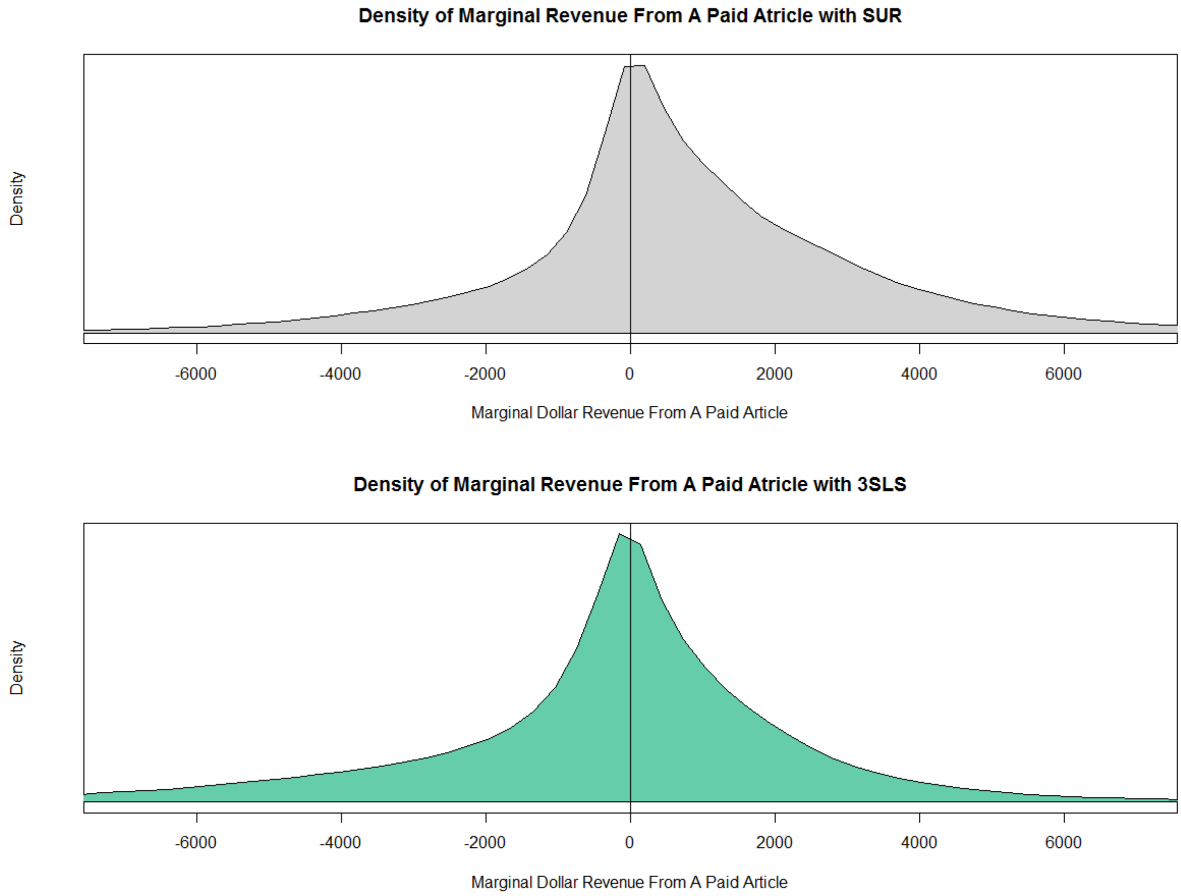


Figure 6: Dollar Effect of an Additional Paid Article by Season, SUR

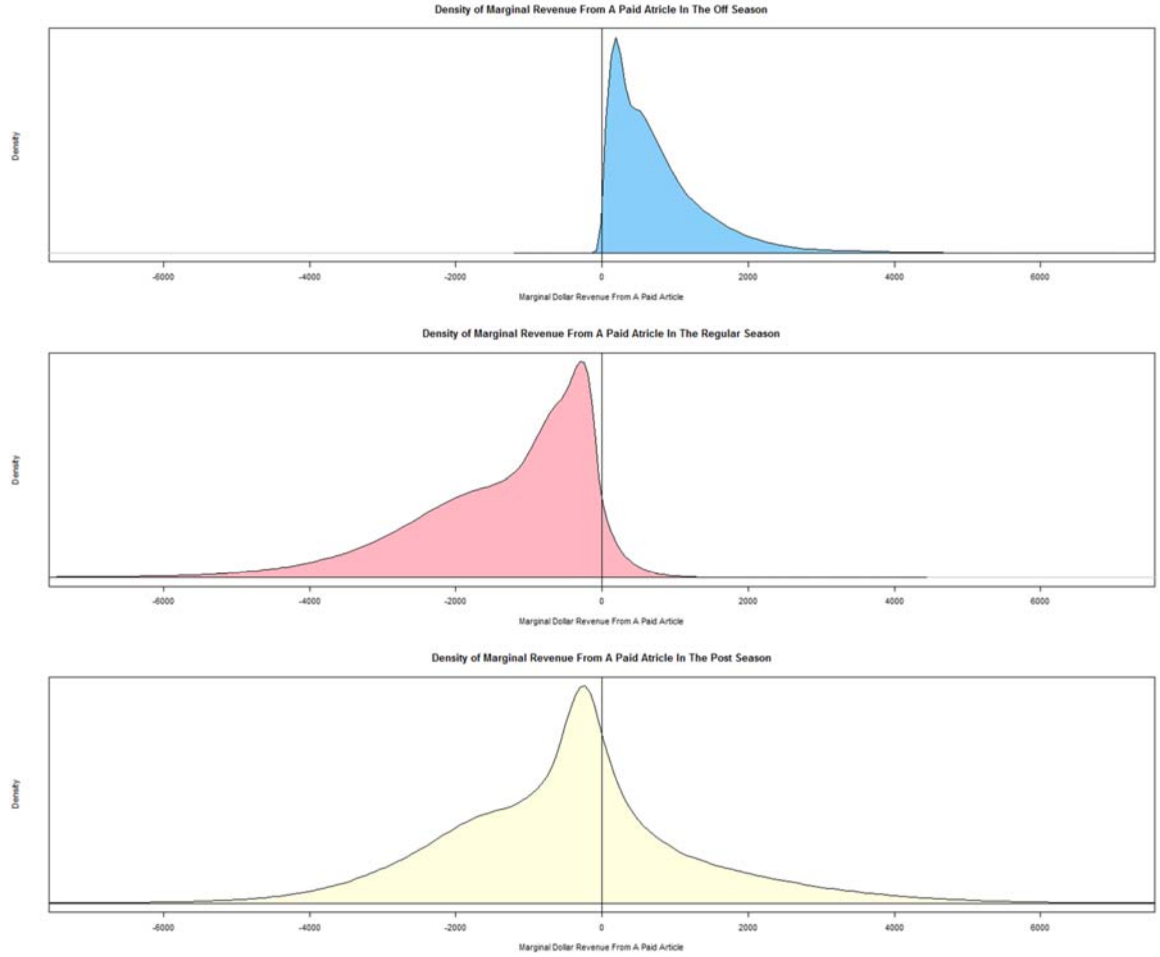


Figure 7: Dollar Effect of an Additional Paid Article by Season, 3SLS Accounting for Endogeneity

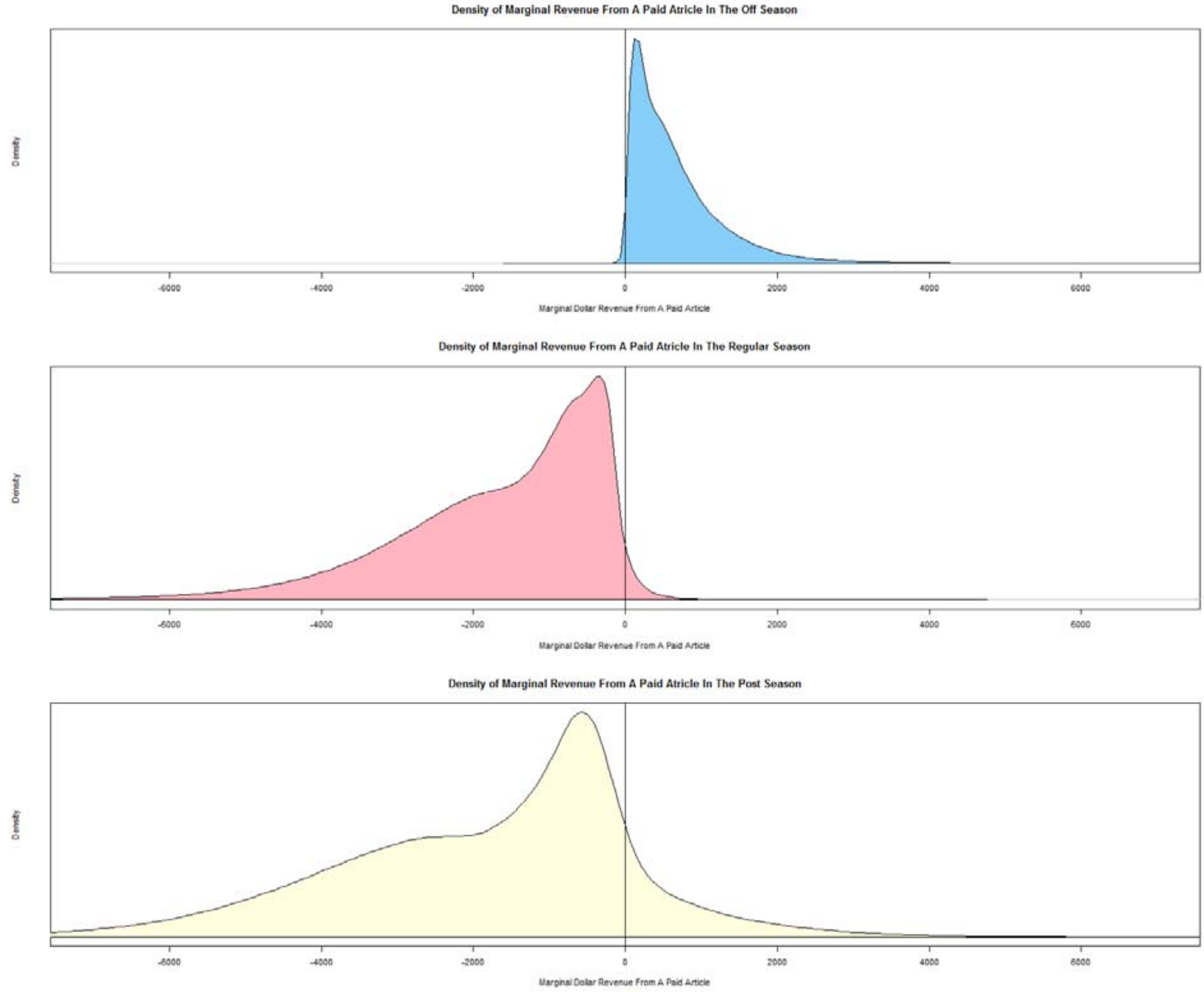


Figure 8: Dollar Effect of an Additional Paid Article by Sport for Regular Season only, 3SLS Accounting for Endogeneity

Figure 9: Dollar Effect of an Additional Paid Article by Sport for Post Season only, 3SLS Accounting for Endogeneity