Pricing and Consumption in Subscription Settings*

Pedro M. Gardete, Daniela Schmitt, Florian Stahl

June 2025

Abstract

This paper investigates how subscription pricing affects consumption intensity, a key performance driver for firms operating under subscription-based business models. We analyze data from an online news publisher, a setting in which promotional pricing is commonly employed to attract new subscribers, though its broader effects remain ambiguous. Standard economic intuition suggests that lower-paying subscribers derive lower utility and thus consume less. In contrast, we document that promotional subscribers, on average, consume substantially more than those paying regular price, even after accounting for differences in churn behavior. This empirical pattern is inconsistent with traditional demand models and points to the importance of taking unobserved heterogeneity into account. We develop and estimate an empirical model of subscription and consumption behavior, showing that, because subscription costs are sunk at the time of consumption, it is possible to recover the correlation between consumption levels and consumers' unobserved willingness to pay. We use the model to recover the underlying consumer parameters and to evaluate the impact of alternative pricing policies on both subscription revenues (via customer acquisition) and advertising revenues (via subsequent consumption). Our findings highlight the economic value of understanding how price shapes not only who subscribes, but also how much they engage with the product.

*Gardete (pedro.gardete@novasbe.pt) and Schmitt (daniela.schmitt@novasbe.pt): Nova School of Business and Economics; Stahl (florian.stahl@uni-mannheim.de): University of Mannheim. Authors contributed equally and order is alphabetical. We thank André Trindade, Arash Laghaie, Ludovic Stourm, Fan Zhang, and participants in research seminars at the London School of Economics, Tilburg University and University College London, and in the conferences European Marketing Academy Conference, Markets and Waves, Special Interest Group on Pricing of the European Marketing Academy, Statistical Challenges in E-Commerce Research (SCECR), for very valuable feedback. All errors are our own. This work was funded by Fundação para a Ciência e a Tecnologia (UIDB/00124/2020, UIDP/00124/2020 and Social Sciences DataLab - PINFRA/22209/2016), POR Lisboa and POR Norte (Social Sciences DataLab, PINFRA/22209/2016).

1 Introduction

Ask any executive how pricing policies influence the demand for a product or service, and you'll get a confident, well-reasoned reply. Ask that same executive how pricing policies affect consumption—the extent to which customers use products or services that they've paid for—and you'll get a muted response at best.

– John T. Gourville and Dilip Soman, HBR 2002

This paper examines the phenomenon of subscription pricing and its relationship with subsequent usage intensity, a key relationship for many firms. For news publishers, streaming services, or software platforms, promotional pricing campaigns are more than a means to attract new customers: they are essential tools for promoting user engagement, building customer loyalty, and driving revenue. In subscription-based business models in particular, promoting engagement and actual consumption are especially important, since firms often incur costs and/or derive value from the intensity of product consumption (e.g., via advertising).

The relationship between subscription fees and subsequent usage is nuanced and can involve various mechanisms. One is that of demand composition or selection: the fact that changes to subscription prices attract different types of customers, who in turn have different preferences for how much they would like to consume. The traditional prediction from economic theory is straightforward: Because individuals consider their consumption value before subscribing (i.e., they are forward looking), subscription prices are expected to be positively correlated with individuals' consumption intensities. When price decreases, for example, the firm is able to attract consumers who would not be willing to subscribe at the regular price, and who, ceteris paribus, value consumption less than its current customers do. This prediction is reinforced by the fact that the prices of most subscription offers are typically not significant enough to induce wealth effects (e.g., Vives, 1987, Hayashi, 2008), suggesting that consumption is more likely to be determined by preferences than by income effects induced by budget constraints and diminishing marginal utility of consumption of the outside good. In effect, lower prices can lead to lower consumption across various contexts. For example, in the contexts of a physical newspaper subscription and an online grocer, Lewis (2006) finds that discount subscribers exhibit lower repurchase rates. Just and Wansink (2011) find that a 50% price discount to an all-you-can-eat meal led participants to consume 28% less pizza and attribute this behavior to the sunk cost fallacy. In line with these findings, Datta, Foubert, and Van Heerde (2015) find that free-trial consumers of a digital television service are significantly less valuable than regular customers.

While these focal predictions are precise, they are by no means general. For instance, consumers with relatively low willingness to pay (WTP) – such as those with limited income, outstanding financial obligations, or fixed budgets (e.g., retirees) – may nonetheless exhibit high demand for news, provided that access is sufficiently affordable. Relatedly, differences in income profiles may shape how consumers allocate their consumption across sellers. In the streaming industry, for example, lower-income consumers may concentrate their viewing within a single subscription and fully exploit the available content, while higher-income consumers may diversify their viewing across multiple platforms, cherry-picking select content from each offer. A price promotion by one provider may thus be particularly effective in attracting the former group – those seeking to centralize consumption – but be less appealing to the latter, who may only respond if the promotion happens to coincide with the release of a highly anticipated title. These examples illustrate the fact that the correlation between consumers' WTP for access (extensive margin) and their subsequent consumption intensity (intensive margin) need not be necessarily positive. Given that firms and researchers often rely on single-source data for demand analyses, it is crucial to account for heterogeneous responses to price promotions in order to better understand the dynamic relationship between subscription prices and subsequent consumption levels.

In this paper, we investigate the consumption patterns of news of a leading European online news publisher offering both free and premium articles, the latter accessed via a monthly subscription. During the sample period, the publisher reduced its subscription price for new subscribers, which led to a marked increase in the number of new subscriptions. In addition, consumers who subscribed during the promotion period exhibited higher news consumption levels, i.e., read more articles than consumers who subscribed a few days before the promotion at the regular price.

Two striking behavioral patterns are present in our data. First, the higher consumption by new subscribers is partially explained by their lower churn rate, yet the higher consumption level persists even after conditioning only on active users: Conditional on remaining active, promotion subscribers consume more news than their regular-price counterparts. Second, the difference in consumption levels remains stable over time, even after a year has elapsed. Rather than inducing temporary changes in consumer behavior, it appears the promotion effectively attracted new types of subscribers to the service.

We believe there already exists evidence in the Marketing field – albeit rarely made explicit – supporting the claim that changes in subscription prices may not always have traditionally defined impacts on consumption patterns (i.e., lower prices leading to less consumption per customer). For example, Danaher (2002) manipulates subscription prices of a telecommunications service and finds that individual usage on average decreased as a result of an increase in the subscription price of the service. This result persists even after controlling for customer churn: consumers who continued using their subscription despite higher prices exhibited lower consumption levels compared to those who remained active under regular pricing. The effect is significant, both in its magnitude as well as statistically. In the context of a music streaming service, Chou and Kumar (2024) find that segments are heterogeneous in terms of their WTP and their usage. For example, women are reported to use the service less than men, despite exhibiting lower price elasticities. The authors propose that, despite their lower usage, women may face a higher valuation for leisure (or other competing activities). Basically, consumers hold preferences on one hand, but other factors also drive their WTP, which may be determinant for their final behavior. Relatedly, Albuquerque, Pavlidis, Chatow, Chen, and Jamal (2012) and Runge, Levav, and Nair (2022) find that while price promotions tend to boost subscription rates, there is little or no evidence of changes to other activities, including product usage. The diversity of these results highlights the importance of developing a flexible empirical framework. Such a framework should be capable of accounting for an arbitrary correlation between subscription decisions and subsequent consumption in order to ultimately answer our research question: how does subscription pricing impact usage intensity?

Marketing scientists often consider settings in which consumers take interdependent decisions over time. A context that is close to ours is the case of retail settings, in which consumers decide whether to buy an item and, if so, how much to buy. Although there already exists a framework relying on discrete-continuous models (Hanemann (1984), Krishnamurthi and Raj (1988), Chintagunta (1993), Kim, Allenby, and Rossi (2002), Bhat (2005), Tuchman, Nair, and Gardete (2018), Stourm, Iyengar, and Bradlow (2020)), we find it inappropriate to model subscription contexts for two reasons. First, in integrating consumers' sequential decisions, these models generally assume a positive correlation between prices and (average) consumption levels, which limits the types of situations they can be applied to. Second, discrete-continuous models deal with per-unit prices, such that the total price paid depends linearly on the quantity purchased/consumed. This contrasts with subscription settings where, crucially, the quantity subscribers are allowed to consume seldom depends linearly on the subscription price.

After documenting the patterns in the data, we develop a model to characterize consumers' subscription and consumption decisions. The model accounts for the consumption value derived by subscribers, as well as additional unobservable factors (e.g., income, competing offers, ongoing subscriptions) that may influence consumers' willingness to pay (WTP). We leverage the fact that, in subscription settings, subscription fees are sunk at the time of consumption. In line with Heckman (1979), we assume that price influences consumption through selection, allowing us, by observing price variation, to recover the overall distribution of ideal consumption levels. This feature enables us to identify the correlation between consumption values and unobservable WTP factors.[1]

Our analysis reveals that failing to account for unobservable forces would result in very biased model parameter estimates. This is significant because most empirical work relies on a single dataset of rich consumer behaviors, typically sourced from a single organization and rarely contemplates all activities that may compete, directly and/or indirectly, for consumers' time and resources.

In the counterfactual analysis section, we consider different promotional programs and find that the price promotion offered by the firm is very near the optimal promotional price if only subscription revenue is taken into account. However, once the impact of future advertising revenue from new

[1]There exists a long literature modeling the behaviors of existing subscribers in detail: see for example Ascarza and Hardie (2013). Our work is different in that we are interested in modeling how changes to the subscription price induce heterogeneous subscriber pools, which requires access to price variation in the first place. There is also a long literature on consumer expansion during stockpiling (see for example Bell, Iyer, and Padmanabhan (2002), Van Heerde, Leeflang, and Wittink (2004), Chan, Narasimhan, and Zhang (2008)), which does not apply to the context of digital subscriptions such as ours. See also Daljord, Mela, Roos, Sprigg, and Yao (2023) for insights on the design of experiments involving compliance.

subscribers is considered, we find the firm would have been better off promoting more aggressively to induce more consumption and higher advertising revenue in the future. This finding is robust even at advertising rates lower than the ones communicated to us by the management team.

We then consider a counterfactual analysis in which the firm is assumed to be able to observe preferred consumption levels by subscribers; this is an admittedly futuristic scenario in which consumers are not able to manipulate their real product consumption (e.g., implement a bot to generate fictitious consumption). In this case, firms may be able to introduce quantity discounts that effectively feature negative marginal prices, that is, subscribers who consume more may effectively pay less. We find that providing discounts for high levels of (verifiable) consumption would be extremely profitable to the firm vis-à-vis the status quo for a reasonable range of advertising elasticities, owing primarily to the subsequent advertising revenue generated by the program.

The ability to characterize the relationship between pricing and subsequent consumption is not only relevant to understand profit tradeoffs for firms, but is also essential to assess the broader societal implications of subscription pricing. Our findings contribute to understanding the extent to which news consumption can be affected by pricing strategies – a critical issue in a world where misinformation is said to be widespread and credible online news publishers face significant financial pressure. By shedding light on how pricing policies influence readership, this paper offers insights that are relevant for policymakers, media regulators, and publishers aiming to promote access to reliable and accurate information. For example, policymakers and media regulators can leverage our findings by incentivizing promotional pricing, reducing VAT on digital news, or subsidizing subscriptions to ensure a well-informed society. The negative relationship between WTP and consumption value in this market suggests that such policies could be highly effective.

In the next section, we present a selection of mechanisms related to the interplay of subscription and consumption behaviors. Section 3 describes the dataset patterns and some model-free patterns. Section 4 presents the model, its identification, and the estimation method. Sections 5 and 6 present the estimation results and counterfactual analyses, respectively, and Section 7 concludes with implications for managers, publishers and policymakers.

2 Subscriptions and Consumption

Below we illustrate that the relationship between subscription pricing and subsequent consumption is nuanced – likely shaped by multiple, interacting forces. While the goal of this paper is not to disentangle specific mechanisms, it is useful to clarify the potential pathways through which subscription pricing may influence subsequent consumption.

Positive Correlation between Subscription Prices and Consumption. A canonical structure of consumer preferences inherently predicts a positive correlation between subscription prices and subsequent consumption. For example, in purely vertically-differentiated markets served by a single seller, a price drop leads extramarginal consumers (i.e., those whose valuations fall below the regular price) to select into buying, resulting in an overall drop in average consumption. By extension, a seller holding enough market power will likely induce the same effects. We provide a formal result below (proof in the appendix):

Proposition. *(Consumers Differing Along a Single Dimension). Let* $u(\cdot)$ *be a utility function and* x_i^* *be the optimal consumption of consumer* i *conditional on a purchase at price* p *so that the consumer subscribes iff* $u(x_i^*) \geq p$*. Then, a price decrease leads simultaneously to higher demand and lower average consumption.*

Given the result above, it is not surprising that the classical prediction of economic theory is that subscription prices and consumption should be positively correlated. Along the same lines, the sunk cost fallacy predicts that costs incurred today affect future behaviors despite their payoff irrelevance. In this case, the relationship with product usage is relatively straightforward. For example, Arkes and Blumer (1985) find that consumers who obtain a price promotion for opera season tickets decrease their attendance in the future due to the lower sunk costs incurred. Along these lines, the theory predicts that a price increase will lead consumers to value their past investment more later, thus increasing consumption. Relatedly, the idea that consumers may hold separate mental accounts has received significant attention (e.g., Thaler (1985)). Consumers who appreciate certain activities are more likely to keep larger accounts for those activities. Similarly, consumers with higher wealth levels are more likely to hold larger mental accounts. It is natural to expect that mental accounts will result in higher subscription prices driving more consumption due to WTP selection of consumers.[2]

Negative Correlation between Subscription Prices and Consumption. In the introduction section, we referred to the fact that consumer heterogeneity may induce a negative correlation between consumption and WTP. One example would be consumers who simultaneously have a low WTP and a low opportunity cost of time. Similarly, in the broad entertainment category, some consumers purchase subscriptions from only a few sellers to access content while others may afford to be more opportunistic, paying for subscription fees of more sellers to access specific content offered by each one. A price promotion is less likely to sway the latter group of content-sensitive consumers; instead, it may primarily attract price-sensitive consumers looking to access a broad selection of content in one location.[3]

[2] In the online appendix, we also formalize the potential role of wealth effects and loss aversion on the correlation between subscription prices and consumption.

[3] The resulting negative correlation between subscription prices and consumption may be further reinforced by switching and multihoming costs, which make it costly for consumers to switch between or maintain multiple offers (e.g., Klemperer (1987), Hartmann and Viard (2008), Villas-Boas (2015)). Consumers are also known to face smaller but relevant costs during consumption with significant implications to competing firms (see Esteves-Sorenson and Perretti (2012)). These frictions can induce persistence and cause high-value consumers to refrain from taking advantage of price promotions. For instance, a price promotion may be effective in attracting low-type consumers who did not subscribe to competing services before, but fail to attract high-type consumers who already subscribe to a competitor offering a premium alternative. Similarly, in the context of vertically-differentiated markets, the existing literature has documented the fact that "high-types" may respond less to promotions and other offers than "low-

In our model, we assume consumers are heterogeneous in terms of their opportunity cost of time and their (unobserved) WTP. When subscription costs are sunk at the time of consumption, the latter ultimately depends on consumers' best use of their time. For individuals with more time available, who are likely to engage in higher levels of consumption, a price promotion may attract those with lower WTP. In this case, price promotions may increase the average consumption level of the subscriber pool.

WTP is typically modeled as a budget constraint that limits the overall spending on a given product or category. In the marketing field, the budget constraint has been conceptualized as the consumer's expenditure in the category (Kim, Allenby, and Rossi (2002), Allenby, Shively, Yang, and Garratt (2004)), a mental account for the type of expenditure (Thaler (1985), Kivetz (1999)), and disposable income (Pachali, Kurz, and Otter (2023)).[4] Because WTP is in effect a latent variable, it is not amenable to be measured with precision. Owing to the fact that subscription fees are sunk at the time of consumption, we show that it is nonetheless possible to identify the correlation between unobservable WTP and consumption.[5]

Given these varied and sometimes conflicting mechanisms, our empirical approach must flexibly allow for both positive and negative correlations between WTP and consumption behaviors.

3 Data Patterns

To analyze the relationship between subscription pricing and consumers' usage decisions, we utilize a dataset that originates from a leading European digital news publisher that implements a freemium business model, i.e., the publisher provides access to a combination of free and premium content on its website. Whereas visitors can read free articles at no cost, premium content is accessible only to paying subscribers. The subscription is priced at €4.99 per month for unlimited access to all content. The publisher also generates revenue by displaying advertisements on both free and premium articles, with ad revenue directly tied to the number of articles viewed. We learned in meetings with the management team that the marginal cost for the publisher to accommodate additional readership is considered irrelevant.

Our dataset focuses on consumers who subscribed between May 6^{th}, 2015, and June 15^{th}, 2015, providing detailed individual-level data on the browsing and subscription behavior of more than 10,000 new subscribers. For each subscriber, the dataset captures the timing and frequency of website visits, along with subscription initiation and cancellation dates. We examine behavior during the one-year period following each user's original subscription, offering a comprehensive view of their engagement patterns over time.

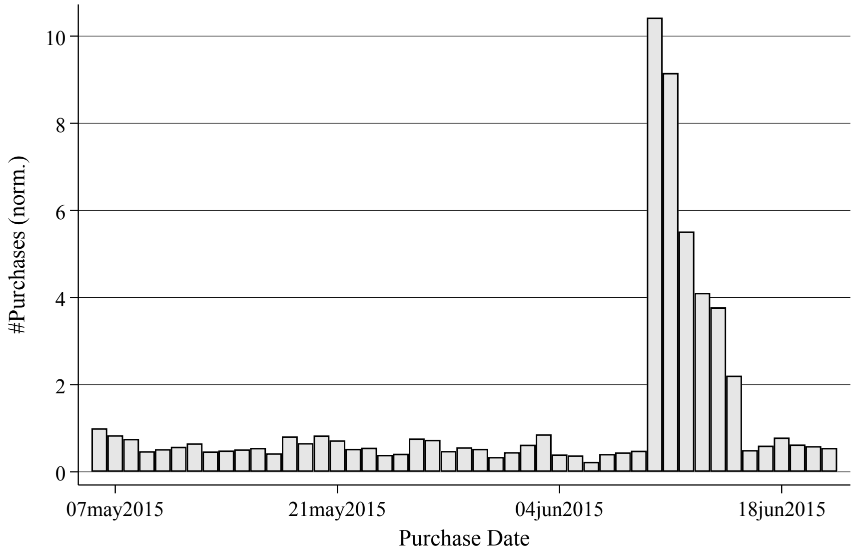
The publisher introduced its first ever price promotion in mid-June 2015. The six-day promotion ran from June 10^{th} to June 15^{th} and temporarily reduced the monthly subscription price from

types," due to the existence of higher-quality – but also more expensive – alternatives (see for example Villas-Boas and Schmidt-Mohr (1999); Stole (2007); Gardete (2013)).

[4] See also discussions in Deaton and Muellbauer (1980) and Chintagunta and Nair (2011).

[5] It suffices that at least a portion of the fees is sunk at the time of consumption.

Figure 1: New Subscriptions Per Day



Note: Number of daily subscription purchases normalized to 1 based on the May 6^{th}, 2015 level.

€4.99 to €2.00. It was not pre-announced in any channel and was available exclusively to first-time subscribers, who continued to pay the reduced price of €2.00 per month until they canceled their subscription.[6]

Figure 1 presents the number of new subscriptions during the sample period (for completeness, it also includes six days after the promotion, which are ignored in the main analysis). The data reveal a pronounced spike in new subscriptions during the six-day promotional period, prior to which subscription activity was relatively stable, with no detectable anticipation effects.[7] This momentum diminished gradually over the promotional period and is undetectable after the promotion ended, reflecting the well-defined impact on new subscriptions.

The promotional effects appear reasonable given the magnitude of the promotion: The discount generated more than ten times the number of new subscribers the publisher had attracted in any day before the promotion. However, the effects of the promotion are clearly diminishing over time. The decreasing pattern is in line with the notion of Conlisk, Gerstner, and Sobel (1984) that some types of consumers may subscribe only during promotional periods. As a result, the firm may benefit from running promotions long enough to attract them but not too long so as not to allow other types to take advantage of it.

A more enduring consequence of the price promotion is its impact on consumption patterns. To

[6]We also observe subsequent promotions offered by the news publisher, each with features that differ slightly from the focal promotion analyzed in this study. Similar to our main case, all four of these additional promotions exhibit elevated consumption levels among promotional subscribers. This repeated pattern reinforces the idea that our observed effect is not an isolated occurrence.

[7]Indeed, the management team confirmed that the promotion was not advertised anywhere other than on the provider's website.

explore these patterns, we classify subscribers into two groups: regular subscribers (who joined at the regular price) and promotional subscribers (who subscribed during the promotion period). Table 1 provides moments of the consumption behaviors during the first year following each customer's subscription decision. On average, promotional subscribers consumed almost three times as many articles as their regular counterparts. When we aggregate all articles consumed by subscribers over a maximum subscription period of 12 months, we see that promotional subscribers demonstrated significantly higher consumption levels compared to the group of regular subscribers (10,677 vs. 3,697 articles consumed in the first year, $p<0.001$).

Figure 2 presents consumption levels for promotional subscribers vs. 'regular subscribers' who purchased the subscription up to one week before the promotion started, to maximize comparability between groups.

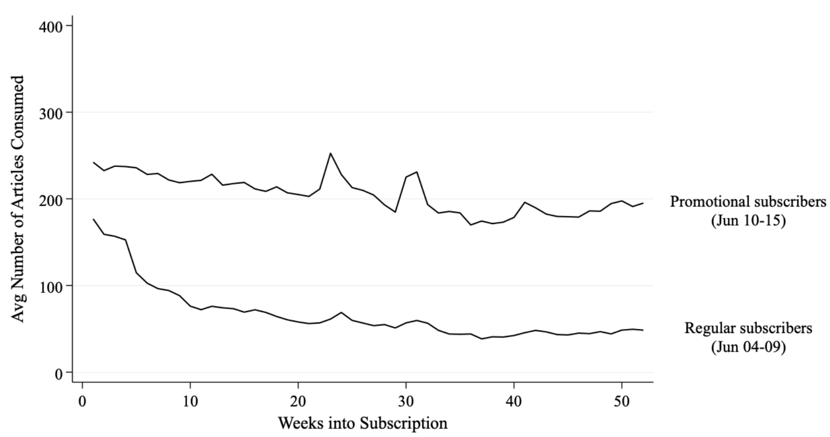
Table 1: Summary Statistics of News Articles Consumed Per Week

	Min	Mean	Median	Max	SD	N
#Articles consumed/week:						
Subscribed at regular price	.02	71.10	17.82	6,183.9	172.81	3,620
Subscribed at promotional price	.02	205.33	134.23	3,865.8	253.04	6,615
Total	.02	157.85	79.21	6,183.9	236.77	10,235

Note: Above, weekly summary statistics during the sample period, May 6^{th}, 2015, to June 15^{th}, 2015.

Across the 52-week period, promotional subscribers demonstrate consistently higher consumption levels than regular subscribers. Consumption rates begin with a large gap and this difference stabilizes, with the two groups maintaining roughly parallel trends over time. This persistent disparity suggests that promotional subscribers exhibit distinct engagement behaviors that may not be solely attributable to short-run effects of the subscription price.

Figure 2: Number of Articles Consumed up to One Year since Subscription



Note: Weekly news readership over a 12-month period after subscribing.

Table 2 highlights the differences in contract duration between promotional and regular subscriber groups. A striking result is that the median regular subscriber spends only one month with the firm. Clearly, these subscribers should not be conceptualized as a stable stock of customers consistently consuming content over time. Rather, they represent a relatively dynamic and transient flow. An inspection of the news provider's website revealed that a large part of the premium content offered is time-sensitive exclusive news – frequently involving celebrities, politicians, athletes, and actors – and is likely to capture attention and induce arousal. Hence, the median subscription duration of one month most likely reflects the fact that customers who subscribe at the regular price want to access specific news articles of immediate interest, without a clear intention of maintaining the subscription in the long term.

Table 2: Summary Statistics of Subscription Length

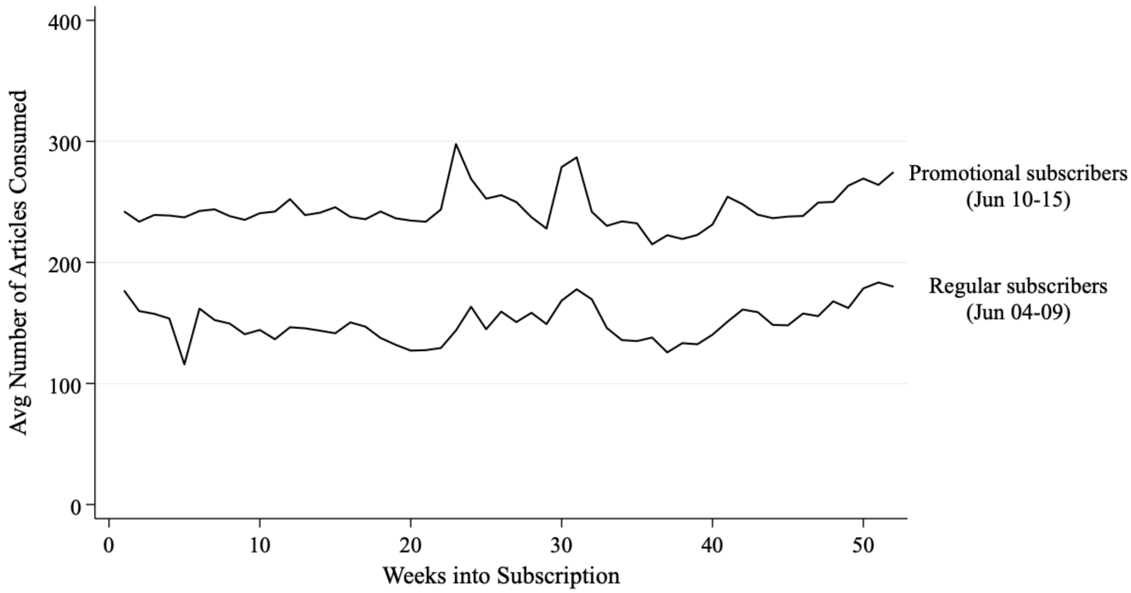
	Min	Mean	Median	Max	SD	N
Subscription length in months:						
Subscribed at regular price	1	4.89	1	12	4.83	3,620
Subscribed at promotional price	1	10.19	12	12	3.76	6,615
Total	1	8.32	12	12	4.88	10,235

Note: Length of users' subscription, measured in months, over the 12-month period following their initial subscription.

The median subscription length of consumers who subscribe at the promotional price is contrasting with the first group: twelve months instead of one month. While the different subscription durations are suggestive of the presence of heterogeneous consumers selecting into subscribing at different prices, they could also simply reflect the fact that promotional subscribers have less of an incentive to churn, as doing so would forfeit their promotional price. In other words, the lower churn rates of promotional subscribers may be enough to explain their higher consumption levels, a consequence of straightforward strategic behavior. To examine this possibility, in Figure 3 we plot consumption levels for the two subscriber groups only during periods in which each subscriber was active. Notably, conditional on not churning, promotional subscribers still display substantially higher engagement levels, consuming approximately 50% more news articles than regular subscribers.

To understand the differences between regular and promotion consumer groups further, we analyze pre-subscription consumption behaviors. We rely on a dataset shared by the news provider that comprises the consumption intensity of those customers who registered but did not subscribe immediately after. In our dataset, consumers may have registered 1, 2, up to 90 days before actually subscribing, which allows us to assess their consumption levels before paying. We analyze the behaviors of consumers who registered between Feb 8^{th}, 2015, and June 14^{th}, 2015 who did not subscribe immediately, contrasting those who subscribed at the regular price with the ones who subscribed at the promotional price. We find that customers who subscribed at the regular price consumed on average 24.85 articles per day prior to subscribing, compared with 32.87 daily articles consumed by promotional price subscribers. This difference remains significant even after removing consumers who both registered and purchased during the promotion window: customers who registered before the promotion, but ended up subscribing only during the promotion window

Figure 3: Number of Articles Consumed by Active Subscribers up to One Year since Subscription



Note: Weekly news readership over a 12-month period after subscribing, conditional on subscribers remaining active (i.e., not having churned).

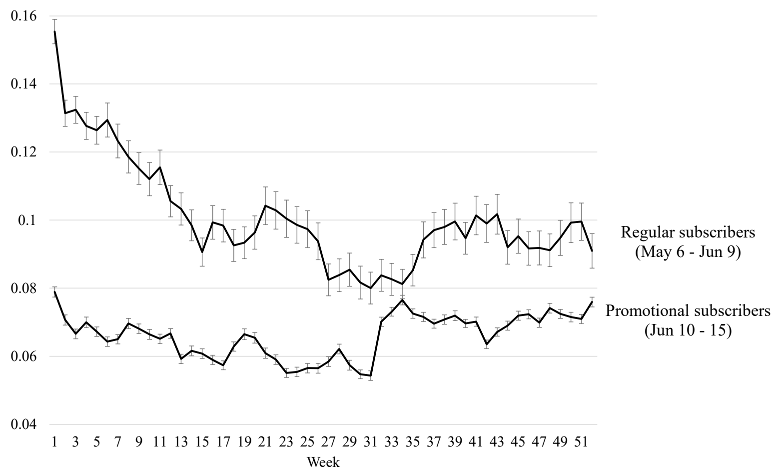
(which they could not anticipate) consumed on average 30.28 articles per day, still significantly above the consumption by regular-price subscribers (p-value less than 0.001).

Overall, the data patterns above point to the price promotion having attracted fundamentally different types of consumers into the publisher's subscriber pool. First, regular subscribers appear to behave opportunistically, willing to pay the full price to access relatively low consumption quantities. Second, the consumption levels of the promotional group are higher despite the fact that subscription costs are sunk at the time of consumption. Third, the effect of the promotion on subscription rates is significant and decays quickly, pointing to a well-defined segment of consumers who were 'not in the market to subscribe' at the regular price, but 'rushed into subscribing' once the promotion was offered. Finally, promotional subscribers consume more articles than their regular counterparts even after we control for consumer churn and after a year has elapsed since the initial subscription. While habit formation may play a role in these patterns, the parallel trends in Figure 3 suggest a steady difference in consumption levels between the two groups from the original subscription moment on, leaving little space for dynamics related to habit formation.

Overall, the combination of extended retention and increased consumption likely contributed to higher advertising revenue opportunities, as the heightened activity of promotional subscribers resulted in more frequent exposure to advertisements. These findings underscore the potential revenue advantages of using strategic promotions to target high-consumption users.

So far, we have focused on the overall consumption differences between regular and promotional subscribers, which is of primary order for the news publisher to generate advertising revenue. We now consider the specific content consumed by subscribers, which may shed light on the types of customers attracted by the regular and promotional prices. We consider 12 news categories that reflect the organization of the news provider, without being explicit about each one to preserve the anonymity of the news provider. We first consider how concentrated news consumption is across the categories, separately for regular- and promotional-price subscribers.

Figure 4: Herfindahl–Hirschman Index of Consumption Shares across Categories



Note: Average HHI per weekly active consumer over the course of a subscription year. Consumers who cancel their subscription during the first year are included until they churn.

Figure 4 presents the evolution of the Herfindahl–Hirschman index for promotional and regular subscribers over the course of a subscription year. Promotional subscribers consume more variety than regular subscribers over the course of the year. In line with the explanations discussed in section 2, it appears that regular consumers tend to subscribe in order to access a smaller subset of news, specifically in categories that are more time-sensitive. Naturally, as time progresses and consumers churn, the consumption patterns of the surviving consumers who subscribed at the regular price approximates the consumption patterns of the promotional subscribers. Given that regular consumers churn at a very fast rate, tend to procure time-sensitive news, and focus on a smaller subset of topics, it is possible that they may be able to spread their consumption for news and entertainment across more sources than promotional subscribers.

Building on these descriptive findings, we introduce an empirical model in the next section that allows us to explain the subscription and consumption patterns above, and investigate the consequences of counterfactual pricing policies by the seller.

4 Model

Consider the case of a consumer deciding whether to subscribe to an online news service. In the first stage, she decides how to allocate her budget between a composite good and, potentially, the news subscription. In the second stage, at which point the subscription cost is sunk, she decides how to allocate her time. Specifically, if she decide to subscribe to the news service, in the second stage she decides her time allocation between news consumption and other activities (including work, leisure, etc.). Let $q_{0i} \geq 0$ be the quantity of a composite good sold at price normalized to 1, $s_i \in \{0, 1\}$ the act of subscribing and $x_i \geq 0$ be the news consumption (in case consumer i subscribes). Finally, let $L_i \geq 0$ be the time allocated to activities other than news consumption. Abstracting from the sequential stages for now, consumer i solves the following problem:

$$\max_{s_i, x_i, q_{0i}, L_i} \quad u\left(s_i . x_i\right) + \alpha_i L_i + q_{0i} \tag{1}$$

$$s.t. \quad s_i . x_i + \beta_i L_i \leq T \tag{2}$$

$$s_i . p + q_{0i} \leq w_i \tag{3}$$

Above, the consumer maximizes her own utility subject to a time and a budget constraint. Function $u\left(\cdot\right)$ is the utility function of news consumption x_i, α_i is the benefit of spending time in activity L_i, and β_i is the opportunity cost of time spent consuming news (note that each unit of news consumption costs the consumer β_i units of an alternative time use). Variable s_i captures the binary subscription decision by the consumer at some subscription price p, which enables subsequent positive news consumption and thus makes x_i relevant for the optimization problem.

The first constraint represents the fact that the time spent consuming news plus the time spent on other activities is bounded above by T, the total time available to all consumers (e.g., 24 hours/day). The second constraint captures unobservable factors that affect consumer i's WTP for the subscription, as discussed in section 2. As we explain later, we allow the unobservable WTP w_i to be correlated with the optimal consumption level, x_i.

Consider first the second stage, at which point the decisions of whether to subscribe, s_i, and how much to consume of the outside good, q_{0i}, have been taken. First, we focus on the case where the consumer decided to subscribe. In this case, the consumer solves the following problem in the second stage:

$$\max_{x_i, L_i} \quad u\left(x_i\right) + \alpha_i L_i + q^*_{0i} \tag{4}$$

$$s.t. \quad x_i + \beta_i L_i \leq T \tag{5}$$

$$p + q^*_{0i} \leq w_i \tag{6}$$

Assuming $u'\left(\cdot\right) > 0$ and $u''\left(\cdot\right) < 0$, the interior solution to this problem is given by:

$$u'\left(x^*_i\right) = \frac{\alpha_i}{\beta_i} \tag{7}$$

$$L^* = \frac{T - x^*_i}{\beta_i} \tag{8}$$

Above, subscriber i consumes news until the marginal utility of news equals the benefit of other activities divided by their relative time cost. Note that the optimal news consumption level x^*_i does not depend on the subscription price p, since this cost is sunk during the second stage in which time is allocated.

In the first stage, consumer i decides whether to subscribe or spend her whole budget w_i on the composite good. Comparing the consumer's ex ante utilities, it is worth subscribing if and only if

$$u(x_i^*) + \alpha_i \frac{T - x_i^*}{\beta_i} + w_i - p \geq u(0) + \alpha_i \frac{T}{\beta_i} + w_i \tag{9}$$

$$\Leftrightarrow u(x_i^*) - x_i^* \frac{\alpha_i}{\beta_i} \geq p \tag{10}$$

where the second expression is obtained by assuming that $u(0) = 0$. It follows that the necessary and sufficient conditions for the subscription to the news service are:

$$u(x_i^*) - x_i^* \frac{\alpha_i}{\beta_i} \geq p \tag{11}$$

$$w_i \geq p \tag{12}$$

That is, the consumer must foresee enough utility from news consumption in relation to the subscription price (condition (11)), and must be willing to pay for the subscription, given her budget (condition (12)).

From here on, we assume the utility function of news consumption

$$u(x_i) = \theta \sqrt{x_i} \tag{13}$$

such that $u(\cdot)$ is increasing and exhibits negative marginal utility in the amount of news consumed. This utility specification assumes that consumers who are willing to subscribe will always consume at least some news, since $\lim_{x \to 0} u'(x) = \infty$. This assumption alleviates the need to check that condition $u'(0) > \frac{\alpha_i}{\beta_i}$ to verify positive news consumption, as long as $\frac{\alpha_i}{\beta_i}$ is finite. Another advantage of this specification is that the resulting optimal quantity x_i^* is monotonic in θ, in terms of the subscription region (11), as we discuss later. This is convenient and greatly simplifies identification, since parameter θ has a monotonic effect on the moments predicted by the model.

Taking advantage of the large variety of topics covered by the news publisher, we assume consumers hold homogeneous preferences for their top n articles. We believe this is a reasonable assumption since even if two consumers have different interests, their overall preferences for their own favorite articles should be relatively similar. Consumers' behaviors are explained by two dimensions of heterogeneity: the net time benefit of time spent in other activities, $\frac{\alpha_i}{\beta_i}$, and their WTPs, w_i.[8]

[8] Other assumptions are possible; we follow Chou and Kumar (2024), who also rely on the explanation of heterogeneous opportunity costs of time. The specific assumption of whether consumption depends on the opportunity cost of time or innate taste differences is not central for our analysis, since we opt to model the correlation between WTP and observable consumption (rather than the underlying causes for the latter). It is easy to show that assuming consumers are heterogeneous in terms of consumption utilities and outside options could also be used to rationalize the data. However, since the focus of our analysis is not on consumer welfare, modeling heterogeneous outside options would add complexity without contributing to our objectives. Throughout the analysis, we assume consumers know

In addition to allowing consumers to be heterogeneous along two dimensions, we also allow the dimensions to be correlated, that is, consumers who may be more pressed for time, for example, may tendentiously exhibit higher or lower WTP for news. From hereon, we define $\phi_i = \frac{\alpha_i}{\beta_i}$ as consumer i's net benefit of alternative activities to consuming news: a high value of ϕ_i should be interpreted as the consumer enjoying other activities more, either because they are pleasurable in themselves (i.e., high α_i), or because they are relatively efficient time-wise (i.e., low β_i). Finally, a consumer type, $\tau(i)$, is defined by the tuple $\tau(i) = \{\phi_i, w_i\}$.

Optimal news consumption is given by

$$x_i^* = u'^{-1}\left(\frac{\alpha_i}{\beta_i}\right) = \left(\frac{\theta}{2\phi_i}\right)^2 \tag{14}$$

From the expression above, it is clear that the distribution of ϕ_i across consumers induces a distribution of news consumption. Because econometricians do not observe ϕ_i in the data, but they do observe quantities x_i^*, it is useful to rewrite the subscription conditions as a function of the latter. Solving equation (14) w.r.t. ϕ_i and plugging the result into the subscription conditions yields the following result:

$$x_i^* \geq \Gamma(\theta, p) \tag{15}$$

$$w_i \geq p \tag{16}$$

where $\Gamma(\theta, p) := \left(\frac{2p}{\theta}\right)^2$. The inequalities above imply that consumers are willing to subscribe if and only if they are willing to consume enough and are simultaneously willing to pay the subscription price (w_i). The fact that the ratio $\frac{p}{\theta}$ appears in the first condition means that parameter θ can be interpreted in relation to the subscription price. For example, the effect of doubling the price on subscriptions (via condition (15)) is equivalent to the utility parameter θ halving.

As indicated by the subscripts in the model, news consumption x_i^* and individual budgets w_i are consumer specific: those facing different time or budget tradeoffs will exhibit different consumption and price thresholds to subscribe. For example, some consumers with large budgets for news (w_i) may be nonetheless unwilling to subscribe due to a high opportunity cost of time (i.e., low ϕ_i), which results in an unjustifiably low news consumption x_i^*. Analogously, some consumers may face a low opportunity cost of time and so be willing to consume a lot of news, but may decide not to subscribe due to financial reasons. We define the joint distribution of x_i^* and w_i as $F_{W,X}$, and ρ to be the correlation between the two variables.

In addition to providing the foundation for the econometric application, the model above is useful to determine that observed news consumption depends on the subscription price, despite the fact that price is sunk at the time of consumption. The reason is that the expected news consumption

x_i^* before subscribing. Given that we observe subscribers' consumption distributions, it would also be possible to explicitly model beliefs over future consumption. We assume perfect information for the sake of clarity.

in a given dataset is not equal to $E\left(x_i^*\right)$, but rather to:

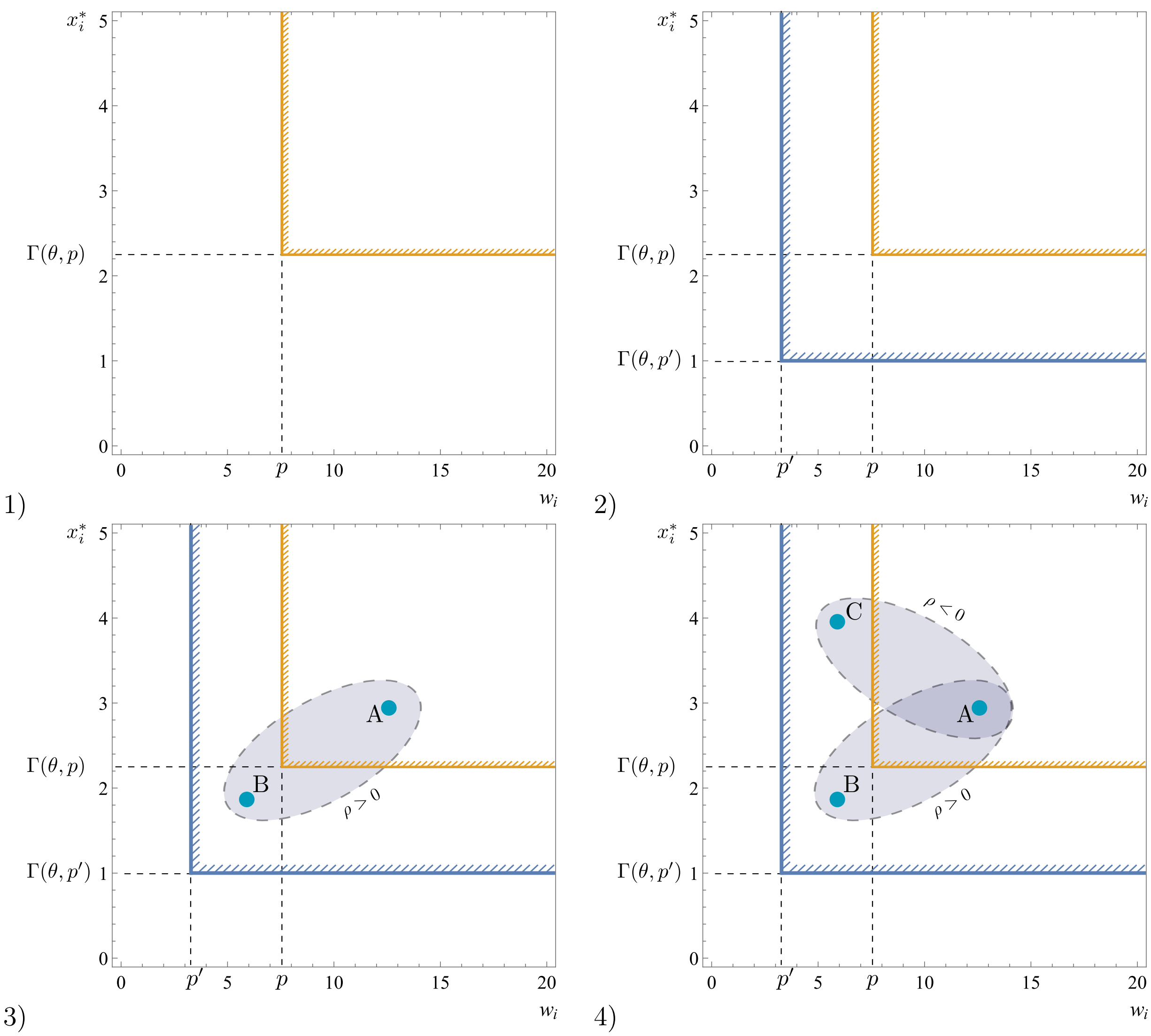
$$E\left(x_i^*\left|x_i^* \geq \Gamma\left(\theta, p\right), w_i \geq p\right.\right) \tag{17}$$

which corresponds to the consumption by subscribers. While subscription price p and the consumer's budget w_i are irrelevant at the time of news consumption, they do influence the distribution of the consumption levels x_i^* we observe in the data via selection, much in the spirit of Heckman (1979). This insight is also useful to identify which types of counterfactual analyses are valid in our context. We will mainly focus on counterfactuals surrounding changes to subscription prices. From the results above, it follows that changes to subscription prices affect the distribution of consumption only via selection and not through reoptimization of news consumption. To reiterate, the reason is that subscription prices are sunk at the time of consumption. We keep notation x_i^* throughout the paper to remind the reader that news consumption is a product of consumer i's utility optimization.

Figure 5 provides a visual representation of this simple model. Panel 1 depicts the space of consumption x_i^* and WTP w_i that justifies subscribing. The top-right area, outlined in orange, reveals that only consumers who are both interested in consuming enough news ($x_i^* \geq \Gamma\left(\theta, p\right)$) and are willing to pay enough for that service ($w_i \geq p$) will subscribe. The second panel illustrates the effect of introducing a lower price, $p' < p$. The subscription region grows downward and to the left, a combination of the preference effect and the WTP effect. Panel 3 introduces the support of a possible distribution of individuals' $\{x_i^*, w_i\}$ pairs, with a positive correlation parameter ρ. It is now possible to see that the price reduction increases the volume of subscribers, given by the intersection of the distribution's volume over its support with the new region. Consumers in this region between the blue and orange lines had not subscribed at the regular price, either because they exhibited relatively weak preferences for consumption or because they held a low WTP related to other factors. Moreover, assuming a uniform distribution over the support, it is clear that the incremental customers (in between the blue and orange regions) will consume less per capita than the ones who are willing to subscribe at the regular price (orange region). Finally, Panel 4 depicts the analogous case of a price decrease when the correlation of consumption and WTP is negative. The price reduction captures consumers who tend to take great interest in reading the news, but other factors discourage them from subscribing at the regular price. For clarity, consider the mass points A and B in Panel 3: We see that, following a price reduction, the new subscribers will pull average consumption down. This contrasts with Panel 4 of the figure, where if only mass points of consumers A and C existed, the price reduction would lead to an increase of the average consumption due to the new subscriber mass C.

The connection between the correlation of preferences and WTP with the effects of pricing on consumption is now straightforward and readily interpretable. This formulation also helps illustrate the central reason why we will be able to identify the correlation parameter ρ: the fact that price variation affects consumption levels only through selection into subscription, as we discuss later.

Figure 5: Indifference Curves for Subscription Decision



Note: Above, shaded regions represent the support of a bivariate density on w_i and x_i^*.

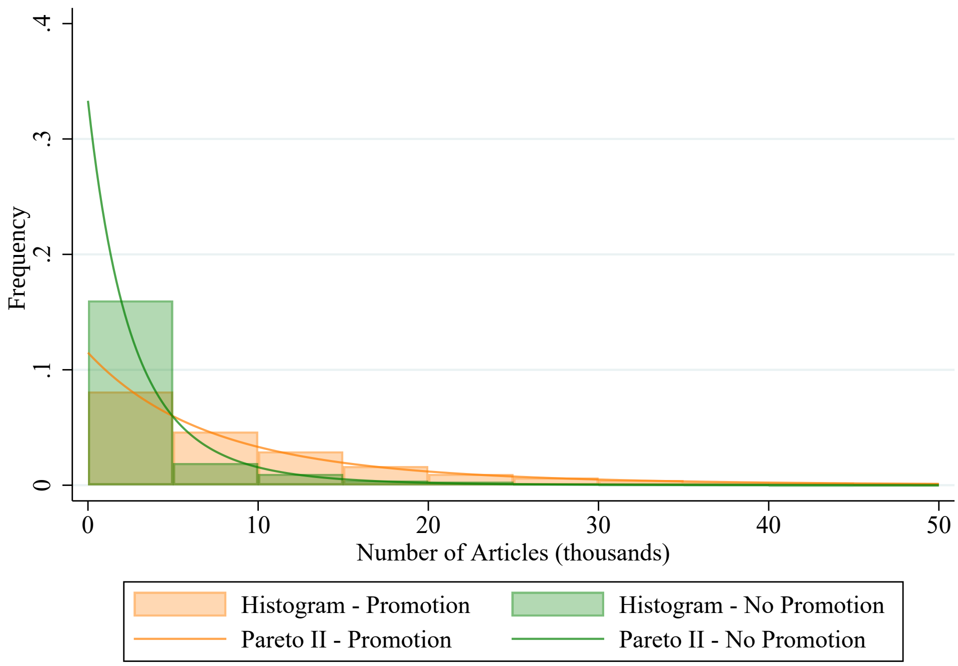
Distribution of x_i^* and w. We consider the parameterization of the joint distribution of x_i^* and w_i.[9] We start by noting that the distribution of w_i is unobservable to researchers. Indeed, being able to account for this unobserved latent variable is one of the contributions of the paper. As for the levels of consumption x_i^* , they are observable from the data, but only for subscribers. In other words, we observe the distribution

$$F_{x_i^*|subscribe_i} = F_{x_i^*|x_i^* \geq \Gamma(\theta, p_i), w_i \geq p_i} \tag{18}$$

[9]Equivalently, we could have considered the joint distribution of ϕ_i and w_i. We opt to characterize the distribution of x_i^* and w_i for clarity. This is without loss of generality, given the one-to-one relation between ϕ_i and x_i^*.

where in the data p_i belongs to one of two price levels.[10] Figure 6 presents histograms of annual readership levels for consumers who subscribed at the regular price (€4.99) and at the promotional price (€2). To be precise, we define x_i^* as the consumption level by consumer i over a 12-month period after the initial subscription decision.[11]

Figure 6: Histograms of the Number of Articles Consumed up to One Year after Subscription



Note: Histograms of news readership over a 12-month period after subscribing. The lines plotted above are fitted Pareto II densities with shape parameter $\alpha = 5$. Readership capped at 50,000 articles for readability.

Both histograms follow exponentially decaying densities, with consumers in the promotion condition consuming more news than their counterparts, in line with the results of the preliminary data analysis. Table 3 presents summary statistics of the sample used for model estimation.

[10] We later explain that each consumer is assumed to face a single price throughout the sample period, hence the use of notation p_i from here on.

[11] We focus on the sample of consumers who, within each price group, fall within three standard deviations from the mean in terms of news readership. This removed a few very high values that may be associated with non-human activities such as bots, scraping tools, etc. Note that x_i^* is the number of articles consumed rather than the time consuming them. We opt for modeling the number of articles since we believe it is a more direct metric of news consumption than the time spent on the publisher's website.

Table 3: Summary Statistics of News Articles Consumed

	Average	Standard Deviation	Min	Max	N:
Subscribed at Promotional Price	9.635	9.672	0.001	51.104	6,480
Subscribed at Regular Price	3.066	5.116	0.001	30.73	3,569

Note: Consumption values: Thousands of news articles consumed over the period of 12 months after subscribing.

The fact that the standard deviation of consumption exceeds its average is a sign of a heavy right tail, given that the readership distribution is truncated below at one. For example, the exponential distribution does not allow the standard deviation to surpass its mean. Indeed, its mean and standard deviation are equal to $\frac{1}{\lambda}$, and in cases where x_i^* is truncated from below at scalar $a > 0$, the standard deviation falls strictly below the mean ($E\left(x_i^* \middle| x_i^* > a\right) = a + \frac{1}{\lambda}$; Std.Dev.$\left(x_i^* \middle| x_i^* > a\right) = \frac{1}{\lambda}$). We address the heavy tail of news readership by assuming that x_i^* is distributed as a Pareto type II distribution (also known as Lomax distribution). This distribution has various uses across literatures, but it is mainly characterized by its heavy positive tail. As can be seen in Figure 6, it fits the observed news-consumption data quite well. The specification of the density of x_i^* is given by:

$$f_X\left(x_i^*\right) = \frac{\alpha_x}{\lambda_x}\left(1 + \frac{x_i^*}{\lambda_x}\right)^{-(1+\alpha_x)}, x_i^* \geq 0, \alpha_x > 0, \lambda_x > 0 \tag{19}$$

where α_x is the shape parameter and λ_x is the scale parameter.

Willingness to pay is unobservable to researchers. We assume it follows a normal distribution with a mass point at zero, that is,

$$w_i = \begin{cases} Normal\left(\mu_w, \sigma_w\right), & w.p.\,\gamma \\ 0, & w.p.\,1-\gamma \end{cases}, \gamma \in [0,1]$$

The interpretation of this specification is that there exists a mass of consumers of size $1-\gamma$ willing to pay nothing to access premium news.[12] This assumption is inspired by the relatively low share of subscribers to the online news publisher in comparison with the very large number of free users; it is likely that most consumers do not even consider the possibility of subscribing in the first place. The use of a mass point is also consistent with the "pain-of-paying" literature, by which even the smallest positive price will deter a large portion of consumers from buying (e.g., Reshadi and Fitzgerald (2023)). Finally, the normal specification is focal among distributions and can be justified by a central limit motivation, because WTP can be thought of as aggregating over many articles, time units and/or multiple benefits of subscribing.

The fundamental relationship in our analysis is the correlation between consumption and WTP. Because we do not observe the joint distribution of x_i^* and w_i, we employ a statistical copula to characterize their relationship. This allows us to assume a flexible bivariate density of x_i^* and w,

[12] Note that we could have set the mass point at any negative value with identical results, since consumers always face positive prices in counterfactual analyses.

approximated by a Gaussian copula with density

$$c_{W,X}\left(u_1,u_2\right)=\frac{1}{\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2}\left(\begin{array}{c}\Phi^{-1}\left(u_1\right)\\ \Phi^{-1}\left(u_2\right)\end{array}\right)'\cdot\left(\left[\begin{array}{cc}1&\rho\\ \rho&1\end{array}\right]^{-1}-I\right)\cdot\left(\begin{array}{c}\Phi^{-1}\left(u_1\right)\\ \Phi^{-1}\left(u_2\right)\end{array}\right)\right) \tag{20}$$

where

$$u_1=F_W\left(w_i\right)$$
$$u_2=F_X\left(x_i^*\right)$$

Above, $\Phi\left(\cdot\right)$ is the standard normal distribution and I is a 2×2 identity matrix. This specification allows us to have a clear definition of the correlation parameter $\rho\in(-1,1)$, which captures the correlation between the percentiles of x_i^* and w.

Model Estimation. For estimating the model parameters, we employ a method of moments estimator that matches the moments in the data with the simulated moments of the model:

$$m_1\left(\Omega\right)=\left(share|\,p^r\right)-\left(\widehat{share}\left(\Omega\right)\Big|\,p^r\right) \tag{21}$$
$$m_2\left(\Omega\right)=\left(share|\,p^d\right)-\left(\widehat{share}\left(\Omega\right)\Big|\,p^d\right) \tag{22}$$
$$m_3\left(\Omega\right)=\left(\overline{x_i^*}\big|\,p^r\right)-E\left(\widehat{x_i^*\left(\Omega\right)}\big|\,p^r\right) \tag{23}$$
$$m_4\left(\Omega\right)=\left(\overline{x_i^*}\big|\,p^d\right)-E\left(\widehat{x_i^*\left(\Omega\right)}\big|\,p^d\right) \tag{24}$$
$$m_5\left(\Omega\right)=std.dev\left(x_i^*\big|\,p^r\right)-\sigma\left(\widehat{x_i^*\left(\Omega\right)}\big|\,p^r\right) \tag{25}$$
$$m_6\left(\Omega\right)=std.dev\left(x_i^*\big|\,p^d\right)-\sigma\left(\widehat{x_i^*\left(\Omega\right)}\big|\,p^d\right) \tag{26}$$

where Ω is the set of parameters to be estimated. Moments m_1 and m_2 match the market shares observed in the data with the ones predicted by the model at the regular (p^r) and discount $\left(p^d\right)$ prices. Moments m_3 and m_4 match the average news consumption of subscribers at the different price levels, and moments m_5 and m_6 match the standard deviations. The estimator above is very convenient to compute, but each moment contemplates different numbers of observations. As such, the GMM approach is not immediately applicable. We employ a method of moments estimator with scaling (Greene, 2000, see Section 5.5 and p. 479):

$$obj\left(\Omega\right)=\sum_{i=1}^{6}m_i'\left(\Omega\right)^2 \tag{27}$$

where $m_i'=\frac{1}{\phi_i}m_i$ is a scaled version of moment m_i. We scale each moment equation by dividing it by its square. For example,

$$m_1^{'}(\Omega)^2 = \left(\frac{(share|\, p^r) - \left(\widehat{share}(\Omega)\Big|\, p^r\right)}{(share|\, p^r)}\right)^2 = \frac{1}{(share|\, p^r)^2} m_1(\Omega) \tag{28}$$

This weighting scheme ensures that moment deviations are equally penalized per percentage unit of the moment in the data. For example, a market share prediction of 4.4% versus an actual market share of 4% is penalized as much as an average consumption of 9,900 articles predicted at 9,000. This weighting method is especially useful given the very different scales of the moments to be matched in the data. Finally, standard errors of the estimated parameters are calculated via 50 bootstrap replications.

Simulation and Estimation Algorithm. The main object to recover through estimation is the correlation between the unobserved WTP levels and the partially observed (due to selection) consumption levels. This requires us to jointly simulate w_i and x_i^* for each guess of parameter ρ, and then simulate predicted moments that are matched with the ones in the data. In each iteration of the set of parameters θ, the model is used to simulate consumption levels, WTP levels, and subscription decisions. We follow this procedure:

1. Consider some guess of parameters, Ω.

2. Take K draws of $\{w_k, x_k^*\}$ pairs via the Gaussian copula, which incorporates candidate cumulative distribution functions F_W and F_X.

3. For each price level p^t, $t \in \{r, d\}$, calculate whether each simulated consumer satisfies the utility and WTP conditions, i.e., $I_{t,k}^x = x_k^* \geq \Gamma(\theta, p_t)$ and $I_{t,k}^w = w_k \geq p_t$. For example, $I_{r,k}^x = 1 \wedge I_{r,k}^w = 1$ means that simulated consumer k is willing to subscribe at the regular price, whereas $I_{r,k}^x = 0 \wedge I_{r,k}^w = 1$ means that the consumer is unwilling to subscribe at the regular price due to insufficient consumption value. For the utility condition, we include an additive logistic shock (i.e., difference of extreme value shocks) to smooth probability of subscribing, in line with discrete demand models.

4. Select the appropriate simulations based on the indices above to construct the moments. For example, the predicted average readership from consumers who subscribed at the regular price is obtained by averaging the set $\left\{x_k^* : I_{r,k}^x = 1 \wedge I_{r,k}^w = 1\right\}$. The market share at the regular price is obtained by dividing the number of simulations that satisfy conditions $I_{r,k}^x = 1 \wedge I_{r,k}^w = 1$ by the number of simulations, multiplied by the market share parameter γ.

5. Given the predicted moments, the objective function (27) is computed and the optimizer either generates a new guess for the parameters or stops at the candidate minimum.

The heavy tail of the distribution of consumption (F_X) led us to set $K = 10^7$ simulations of x_i^* and w_i, at which point different seeds for random number generation had a negligible effect on the estimates in our simulations with randomly generated data. Given the discrete nature of simulations, we employ the Nelder-Mead method (Nelder and Mead, 1965), and obtain the standard errors for the parameters via 50 bootstrap sample draws. In each bootstrap sample, we draw (with replacement) from subscribers and non-subscribers. We were provided with the number of unique visitors to the firm's website during the period at hand as well as the number of current subscribers in the beginning of the period. We use the difference of the two numbers (withheld for confidentiality) as the potential market size in the model.

Taking draws of x_i^* and w_i pairs is relatively simple, and it takes advantage of the copula correlation structure. We start by taking K independent standard normal draws of vectors Z_1 and Z_2 (we use the matrix form below for simplicity), such that

$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim N(0, I) \tag{29}$$

At each set of candidate parameters, we multiply the draws by the (lower triangular) Cholesky decomposition matrix of the Gaussian copula correlation matrix (matrix L, defined below) to obtain correlated normal draws, i.e.,

$$\begin{pmatrix} Z_W \\ Z_X \end{pmatrix} = L. \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix} . \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \tag{30}$$

Finally, we obtain draws of x_i^* and w_i by applying the standard normal c.d.f. to Z_W and Z_X, and then applying their respective inverse marginal distributions, i.e.,

$$\begin{pmatrix} W \\ X \end{pmatrix} = \begin{pmatrix} F_W^{-1}(\Phi(Z_W)) \\ F_X^{-1}(\Phi(Z_X)) \end{pmatrix} \tag{31}$$

where F_W and F_X are the marginal distributions of w and x_i^*, respectively, at the candidate parameters.

We normalize two parameters of the marginal distributions of x_i^* and w. For the distribution of x_i^*, we normalize the shape parameter $\alpha_x = 5$. This disciplines the existence of moments of the Pareto type II distribution. Specifically, moment $E\left((x_i^*)^j\right)$ (and lower) exists if and only if $\alpha_x > j$. The normalization implies that the Pareto type II distribution has well-defined first four moments, that is, mean, variance, skewness, and kurtosis. Rather than run the risk of employing an incorrect estimator, we normalize the value of α_x and utilize only the first two moments of x_i^* in the objective equation (27). The issue of potential lack of existence of moments is common when dealing with heavy-tailed distributions. Changes to the normalization of α_x yielded small effects on the predicted moments of the model.

The other parameter we normalize is the value of the mean of w_i, μ_w. We found that model fit is not especially affected by the mean of w, whose function of capturing the number of non-purchasing consumers is already accomplished by parameter γ, which has a much more direct interpretability.

Overall, we prefer to normalize these two parameters because 1) the parameters we do include already have focal roles in matching moments in the data, and 2) experimentation revealed very little effect of these parameters in terms of their ability to improve model fit. The parameters of our model are summarized in Table 4.

Table 4: Model Parameters

Parameter	Description
θ	Parameter governing utility of news consumption
γ	Proportion of consumers with WTP different from zero
λ_x	Scale parameter of x_i^*
ρ	Correlation parameter of readership and WTP percentiles
σ_w	Standard deviation of the WTP

Identification. We discuss how the parameters above are identified by the estimation moments. Market shares at the regular and discount prices are primarily matched by the utility parameter θ and the share of consumers who are willing to pay a strictly positive amount to subscribe, γ. Parameter γ has a direct effect on market share, regardless of subscribers' consumption levels, for example. As a result, it matches the overall market share across pricing regimes. Parameter θ weighs how much consumption, as captured by x_i^*, translates into utility, and thus, to the act of subscribing. It follows that parameters θ and γ directly affect market shares during the promotional and the regular price periods.

As we explained before, the correlation parameter ρ is identified by the difference in average readership across price levels. This follows from the discussion of Figure 5: The impact of price changes on consumption rates informs the correlation between WTP and consumption.

Parameters λ_x and σ_w match the means and standard deviations of consumption. A higher value of λ_x translates to a higher mean and standard deviation of x_i^* across price levels, a result that follows directly from the moments of the Pareto II distribution. Like in the case of parameters θ and γ with market shares, it does not suffice to match the moments overall; they must also match within each price condition. Parameter σ_w captures the decay of the mean and standard deviation of x_i^* as the price changes, following the assumed curvature of the normal distribution. Referring back to Figure 5, it is clear that when w_i is very spread out, the moments of x_i^* vary less across price levels than if w_i is concentrated near zero (i.e., a low value of σ_w), ceteris paribus. Hence, parameters λ_x and σ_w play a fundamental role in matching the consumption moments across price levels.

Additional Modeling Assumptions. For simplicity, the model above abstracts away from consumers' subscription renewal decisions, focusing only on the first subscription decision and the overall consumption during the same period. In reality, consumers can decide to stop their subscription during that time frame, a decision that is linked with the other ones. While it is straightforward to incorporate the option value of renewals, we believe it brings unwarranted complexity without producing an obvious benefit. In Section 6.2 we explain that consumption levels (which we model) are highly correlated with contract duration and so it suffices to assume that the conditional relationship between these two constructs is stable. We model this relationship explicitly in the counterfactual analysis section later in the paper.

Second, we define x_i^* as the total article consumption by subscribers, encompassing both premium and freely accessible articles. Limiting our counterfactual profitability analysis to advertising revenue generated solely from premium articles could be misleading, since it would ignore that the publisher also accrues advertising revenue from free articles. A related point is the fact that, in our counterfactual analyses, we focus on subscribers while ignoring the advertising revenue these consumers would have generated had they not subscribed. This omission could be problematic if subscribing caused consumers (for some unknown reason) to consume fewer articles. In this case, a seemingly favorable counterfactual scenario – where users generate additional revenue through subscriptions and engagement with premium content – could actually destroy value for the publisher, due to post-subscription reduced consumption. However, using pre-subscription data (presented in Section 3), we find that subscribing is associated with an uptake in consumption, which makes sense given the access to an additional set of articles. This empirical pattern mitigates concerns that favorable counterfactual scenarios might inadvertently reduce value for the publisher.

Third, we assume that ρ is constant across counterfactual scenarios. This assumption may not hold in cases where the scenarios differ substantially from the status quo, as consumers may find it optimal to significantly reoptimize their time and external consumption allocations. Given the relatively small financial commitments involved in subscribing, we believe that potential inaccuracies in the evaluation of counterfactual scenarios are at the very least unlikely to affect their directional validity.

Finally, the potential market is made up of the number of unique visitors (in the millions) to the news provider over the course of a month. We assume each potential consumer considers the opportunity to subscribe once during the sample period. This assumption tackles the fact that we do not observe the arrival rate of consumers to the website nor, more importantly, the number of times they contemplate subscribing. This is a common assumption in subscription and purchasing analyses (e.g., Waisman (2021)), and its effect is likely to be absorbed by the market share parameters.

We now turn to the estimation results and present measures of fit with the moments in the data.

5 Results

We start by reporting the model estimates and providing brief interpretations of their magnitudes. We then present measures of fit and document the bias that arises from ignoring the correlation parameter ρ.

Table 5 summarizes the parameter estimates of the model:

Table 5: Model Estimates

Parameter	Estimate
θ	56.836**
	(3.812)
γ	0.011**
	(0.000)
λ_x	132.001**
	(7.138)
ρ	-0.717**
	(0.016)
σ_w	2.506**
	(0.019)
N: 10,049	
Objective Function:	0.03

Note: Standard errors in parentheses. Significance levels: $^{\dagger}p \leq 0.1$, $^{*}p \leq 0.05$, $^{**}p \leq 0.01$. During estimation, parameters λ_x and σ_w were applied an exponential transformation to impose a positive support. Parameter ρ was kept between -1 and 1 through transformation $2/(1+\exp(-\rho^{'}))$-1. Standard errors obtained via 50 bootstrap samples. Figures above rounded to three decimal places with trailing zeros omitted.

All parameter signs are in line with expectation. Consumers draw positive value from news consumption ($\theta > 0$) and exhibit a standard deviation in WTP of approximately €2.5. This means that, assuming all consumers derived sufficient consumption utility from subscribing, 21.36% of active consumers in the market – those who have a strictly positive WTP – would be willing to pay the promotional price, but only 2.32% would be willing to pay the regular price. The scale parameter of consumption is positive, as expected, and about 1.1% of consumers exhibit a non-zero WTP for online news. Although this number appears small, when multiplied by the multiple millions of potential consumers in the market, it still results in hundreds of thousands of potential customers. This estimate appears reasonable; for instance, the equivalent figure for The New York Times is at least 0.5%.[13]

In order to obtain an interpretation of $\widehat{\theta}$, we consider the necessary subscription condition

$$x_i^* \geq \Gamma\left(\theta, p\right) = \left(\frac{2p}{\theta}\right)^2 \tag{32}$$

that is related with consumption utility. It follows that the estimated minimum consumption thresholds necessary to subscribe, abstracting away from random shocks and budget constraints,

[13] The New York Times specifies "**nearly** 2 Billion readers" and "**more than** 10 million paid subscribers" (https://advertising.nytimes.com/audience-and-insights/).

are given by:

$$x_i^{*reg} \geq \Gamma\left(\widehat{\theta}, 4.99\right) = 30.83\,\text{articles} \tag{33}$$

$$x_i^{*promo} \geq \Gamma\left(\widehat{\theta}, 2.99\right) = 4.9\,\text{articles} \tag{34}$$

Absent WTP concerns, consumers find it on average worthwhile to subscribe at the regular price if they intend to consume more than 31 articles over the course of a year, or at least five articles in case they face the promotional price.

The main parameter estimate of the model, ρ, can be interpreted as a relatively strong negative correlation between news consumption and WTP. The negative value of ρ is necessary but not sufficient for a price discount to induce an increase in the average news consumption level. To validate whether the model predicts such an effect, we turn to comparing the moments of the data with the ones predicted by the model.

Comparison of Model and Data Moments. Table 6 shows the data moments and the ones predicted by the model.

Table 6: Comparison of Data and Predicted Moments

	Data Moments	Model Prediction
Subscription at regular price (N: 3,569)		
Subscription rate:	0.02%	0.019%
Articles read (Mean):	3.028	3.313
Articles read (Std. Deviation):	5.031	4.442
Subscription at promotional price (N: 6,480)		
Subscription rate:	0.21%	0.211%
Articles read (Mean):	9.46	8.895
Articles read (Std. Deviation):	9.468	10.054

Note: Consumption figures in thousands. Figures above rounded to three decimal places with trailing zeros omitted.

Overall, the model fits the data moments reasonably well. The relatively low market shares – owing to the large potential market – are closely matched by the model for both segments. As for articles read, all moments are well approximated, with the largest deviations being an underestimation of the average number of articles read by the promotional segment and an overestimation of the standard deviation. Importantly, the model is able to replicate the two fundamental patterns in the data: First, both standard deviations of readership exceed the segment means and, second, the average readership is higher for the promotional segment than the regular one. Matching these two patterns is essential to capture the empirical patterns and the underlying economic forces in play. The differences in consumption levels across segments have meaningful implications for advertising revenue, as we explore in the next section.

6 Counterfactuals

We now turn to evaluating how different pricing policies interact with the demand we have recovered. All counterfactual analyses below maintain the sample period constant and the regular subscription price unless otherwise stated. In particular, the negative trend associated with subscriptions observed during the promotional period (Figure 1) motivates us to not extrapolate promotional effects to longer periods. As discussed above, the analyses treat x_i^* as the solution to a consumer time-allocation problem, conditional on having subscribed – at which point prices are sunk. Nonetheless, the seller's pricing policies still affect the equilibrium distribution of consumption, by affecting which consumers are willing to select into the subscription pool.

6.1 Effect of correlation parameter ρ

Our first goal is to understand the effect of the correlation between preferences and WTP on the demand system. We investigate this effect by ignoring/allowing for the correlation parameter ρ in estimation, namely since this parameter is typically absent from discrete-continuous model specifications.

Table 7: Model Estimates

Parameter	Original Estimate	Estimate with $\rho = 0$
θ	56.836**	256.029**
	(3.812)	(4.866)
γ	0.011**	0.012**
	(0.000)	(0.000)
λ_x	132.001**	14.622**
	(7.138)	(0.508)
ρ	-0.717**	–
	(0.016)	–
σ_w	2.506**	2.414**
	(0.019)	(0.017)
N: 10,049		
Objective Function:	0.03	0.682

Note: Standard errors in parentheses. Significance levels: † p≤0.10, * p≤0.05, ** p≤0.01. During estimation, parameters λ_x and σ_w were applied an exponential transformation to impose a positive support. Parameter ρ was kept between -1 and 1 through transformation $2/(1+\exp(-\rho^{'}))$-1. Standard errors obtained via 50 bootstrap samples. Figures above rounded to three decimal places with trailing zeros omitted.

Table 7 presents side-by-side estimates of the model parameters with and without fixing parameter ρ at zero. When preferences and WTP are assumed to be uncorrelated, we obtain a larger estimate of parameter θ, associated with the extensive margin (e.g., the act of subscribing), and

Table 8: Comparison of Data and Predicted Moments

	Data Moments	Model	Model, $\rho = 0$
Regular Subscribers (N: 3,569)			
Subscription rate:	0.02%	0.019%	0.02%
Articles read (Mean):	3.028	3.313	4.182
Articles read (Std. Deviation):	5.031	4.442	4.932
Promotional Subscribers (N: 6,480)			
Subscription rate:	0.21%	0.211%	0.209
Articles read (Mean):	9.46	8.895	4.185
Articles read (Std. Deviation):	9.468	10.054	4.974

Note: Consumption figures in thousands. Figures above rounded to three decimal places with trailing zeros omitted.

a much lower estimate of parameter λ_x, associated with the intensive margin of consuming news. When parameter ρ is normalized to zero, the model attempts to match the moments of the data by severely underpredicting the demand for news consumption. To understand this effect, it is useful to compare the moments of the constrained and unconstrained models. The new predicted moments are presented in Table 8, together with those from the data and the original model. While the market-share moments remain well matched, we observe large differences in the news-consumption predictions. These differences are relatively modest for consumers subscribing at the regular price, but the constrained model underpredicts the mean and the standard deviation of articles consumed by subscribers at the promotional price by a factor of two, producing an overall worse fit. In attempting to match readership levels across segments, the constrained model finds a middle ground in terms of mischaracterizing readership patterns: It underpredicts mean readership of the promotional segment and overpredicts readership of the regular one. Straddling both readership levels is now impossible, due to the assumption on ρ. In effect, the model can no longer predict that subscription levels increase with the price reduction, with the increase of the number or articles read from 4.182 to 4.185 sourcing exclusively from simulation error. Because the model is constrained to predict that higher prices will lead to weakly more readership, when $\rho = 0$ the optimal solution balances the fit of the first readership moment of the segments, failing to match either.[14] In summary, failing to account for the correlation between preferences and WTP would not only produce biased coefficient estimates, but it would also prevent the model from explaining the moments in the data.

6.2 Effect of Promotional Price Level on Profits

We now simulate firm profitability at different price promotion levels. In line with the data, we consider the consumers acquired during the estimation period of the data (May 6^{th} until June 15^{th} 2015), with the regular price of €4.99 until June 9^{th}, followed by a promotional price – which we vary in the counterfactual analyses – starting June 10^{th} through June 15^{th}. We first consider the

[14] The second moment of readership is actually matched better when $\rho = 0$ for the regular segment, but at a much higher mismatch for the same second moment in the case of the promotional segment.

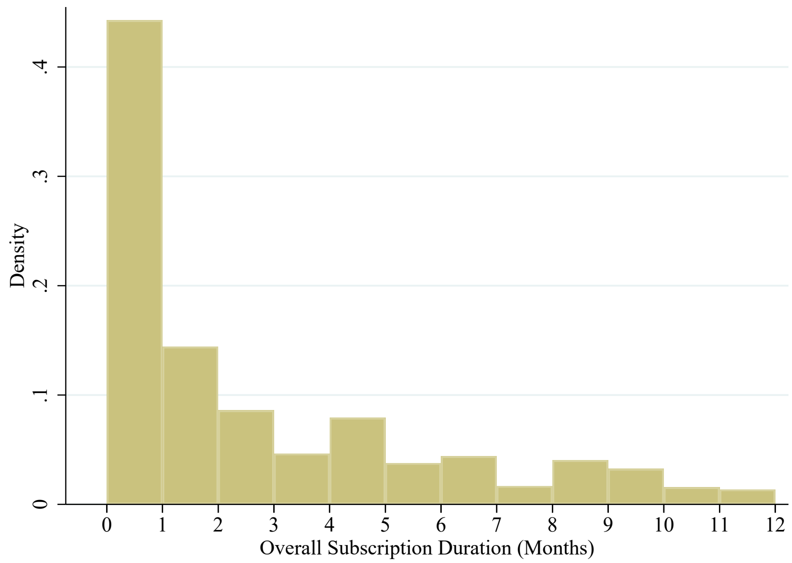
effect of different promotional prices, which are held constant over the course of a year, like in the dataset. Given that the firm's cost structure is relatively constant with the number of premium subscribers, we think of the revenues below as very good proxies for firm profits. We consider counterfactual scenarios in which we change the promotional price offered in the period from June 10^{th} to June 15^{th} and we analyze the firm-profitability effects on customers up to one year after their subscription starts.

So far, we have opted to model initial subscription decisions and consumption levels over the period of a year explicitly. To incorporate subscription revenues over the course of a year as well, we forecast consumer churn via a prediction model. The working assumption is that the relationship between consumer churn and number of articles consumed remains stable across counterfactual analyses. Indeed, the unconditional correlation between the two components in the data is 0.42, which is to be expected since consumption decisions are intimately linked with churn decisions.[15] We plot the histogram of subscription durations in Figure 7 to determine whether additional factors need to be taken into account in this prediction model.The histogram reveals a decreasing churn rate over time since the initial subscription. The decreasing trend can be described by a Cox survival model to predict how long consumers stay with the firm over the course of a year, and ultimately, how that translates to yearly revenue (consumers who churn earlier produce less subscription and advertising revenue, ceteris paribus).

We predict churn via a Cox survival model, including as predictors the total number of articles read by a consumer (i.e., the consumer type). Table 9 reports the results of the Cox regression model as well as the baseline survival probabilities. As expected, the parameter estimate associated with the number of articles read is negative, meaning that the risk of churning decreases with consumption.

[15] Note that in the data there exists heterogeneity in terms of how rapidly consumption takes place over time across individuals. We abstract from this effect to keep the counterfactual analyses parsimonious. Because we are interested in overall profitability rather than its distribution across consumers, we believe that an average analysis, conditional on the regressors, provides a good first-order approximation of the total profitability effects.

Figure 7: Histogram of Subscription Durations since Consumers' Subscription Dates



Note: Histogram above represents only consumers who churned within a 1-year period following the subscription decision, i.e., approximately 41% of subscribers.

Table 9: Cox Estimated Hazard Rates

Regressor	**Parameter Estimate**	**Standard Error**	
Number of Articles ($\times$1000)	-0.249**	0.000	
Log-likelihood	-35,046.73		
Conditional Baseline Survival/Renewal Probabilities			
1^{st} month	0.609	7^{th} month	0.266
2^{nd} month	0.487	8^{th} month	0.255
3^{rd} month	0.417	9^{th} month	0.230
4^{th} month	0.381	10^{th} month	0.211
5^{th} month	0.322	11^{th} month	0.202
6^{th} month	0.296	12^{th} month	0.197

Note: Standard errors in parentheses. Significance levels: † p≤0.10, * p≤0.05, ** p≤0.01. Figures above rounded to three decimal places with trailing zeros omitted.

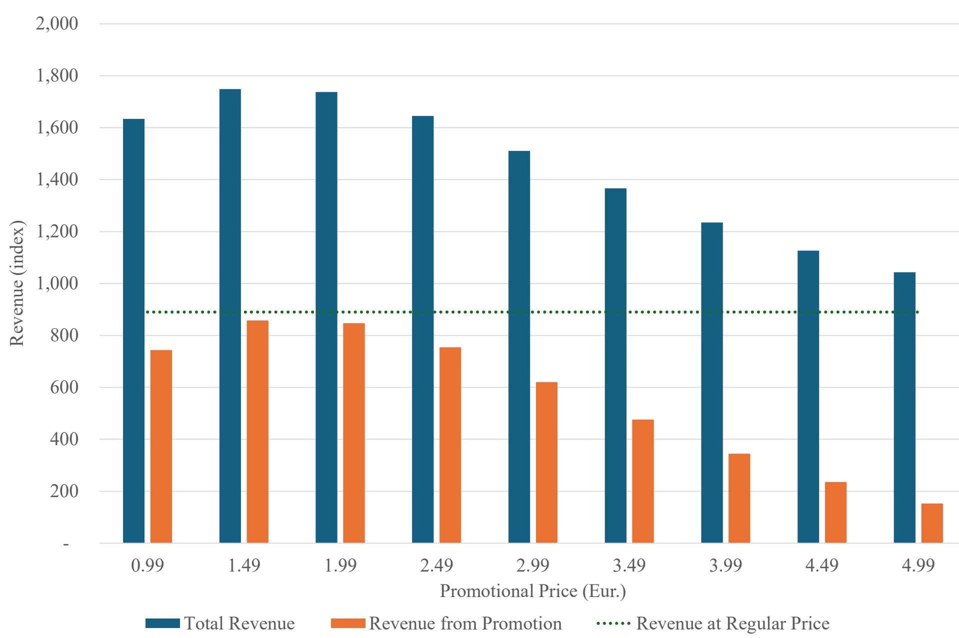
The conditional baseline survival probabilities fall over time, such that the baseline probability (before adjusting for consumption) of renewing equals approximately 61% after one month and 19.7% after 12 months. We calculate consumer i's expected number of subscription payments for the twelve renewal decisions plus the initial subscription decision as:

$$n_i = floor\left(\sum_{t=1}^{12} S\left(t\right)^{\exp\left(\beta x_i^*\right)}\right) + 2 \quad (35)$$

where each term $S(t)^{\exp\left(\beta x_i^*\right)}$ is the conditional baseline survival probability of period t, adjusted by the number of articles read by consumer i weighted by its parameter estimate, following the Cox model specification. We employ the floor function to the sum and add two months to count fractional month occurrences as a single payment and to include the initial subscription as well. Parameter n_i is used during counterfactual analyses to construct an estimate of yearly subscription revenue generated by each consumer.

Figure 8 presents counterfactual firm revenues over the course of a year at different levels of the promotional price.[16]

Figure 8: Subscription Revenue as a Function of Promotional Price

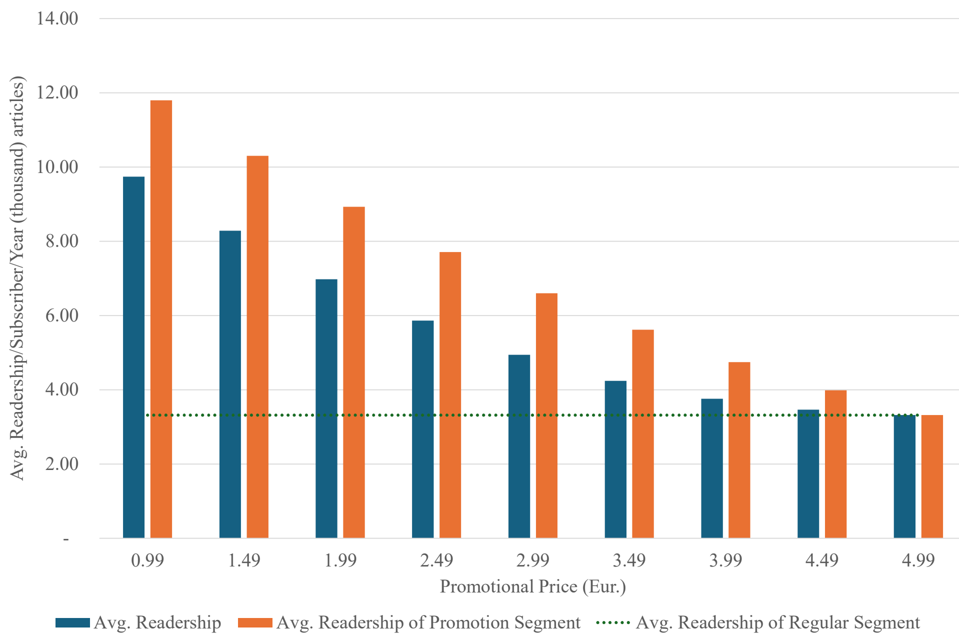


Note: The results above pertain to the sample period.

As illustrated in Figure 8, revenue is concave in price, achieving its maximum near the actual promotional price observed in the data. We remind the reader that there is no structural or theoretical reason that imposes this result. It seems that, by coincidence, the firm's promotional price was set close to the empirical revenue-maximizing level. Above, we plot a dashed line as a reference for the regular-price period revenue. In comparison, it is clear that the revenue from the promotional segment is quite significant, especially when the promotional price is close to €1.49. Note that when no promotion is introduced (price of €4.99), the difference between promotion revenue and the dashed lines is explained by the difference in regular and promotion periods alone (35 vs. 6 days).

[16] Note that, in the original promotion, consumers lock the promotional price in. We keep this feature in the counterfactual analysis: Consumers may renew their subscription at their original subscription price.

Figure 9: Average Readership as a Function of Promotional Price



Note: Above, simulated average readership levels over the course of a year after subscribing as a function of promotional price level, owing to selection into subscribing.

Figure 9 shows counterfactual average readership levels for both subscriber groups. Across the board, readership levels increase with the promotion magnitude, as a result of the higher uptake of avid readers. If consumer welfare is to be measured by consumption intensity, then there is no doubt that higher promotions lead to gains in welfare.

The figure above also reveals that the difference in readership levels increases at an increasing rate as the promotional price decreases. The reason is that deeper promotions increase the subscriber base while simultaneously attracting customers each of which consumes more, on average. Because each news article consumed is positively associated with advertising revenue, a conclusion from this analysis is that the promotional segment drives more advertising revenue than the regular one.[17] The extent to which the promotional segment generates higher ad revenues depends on the price promotion level set by the firm, in line with the results in Figure 9.

6.3 Revenue from Increased Consumption via Advertising

In contrast with traditional retail markets, many firms in digital domains accrue rents via consumption, either directly (e.g., software as a service) or indirectly (e.g., advertising income). We have already shown that price promotions alter the composition of consumers, leading to different

[17]Note that advertisers in this publication pay per impression and not per click, so the potential critique of different unobserved ad-click behaviors by different segments as a result of the price promotion has no bearing.

subscription and consumption patterns. The goal of this section is to quantify the extent to which that composition affects the firm's performance. We focus on subscription and advertising revenues generated within one year of individuals' subscription decisions in the sample period, or up to earlier dates for consumers who churn before that period ends.

Advertising rates vary wildly across countries, industries, and types of publishers. The counterfactual scenario we consider here assumes that the advertising revenue associated with the consumption of a news article remains constant across price-promotion levels. The managerial team of the news publisher indicated the incremental revenue of €0.005 per article consumed, which we use as a focal value for our analysis (we later conduct additional analyses). This value allows us to add the subscription and advertising revenues in order to understand the combined effect of price promotions.

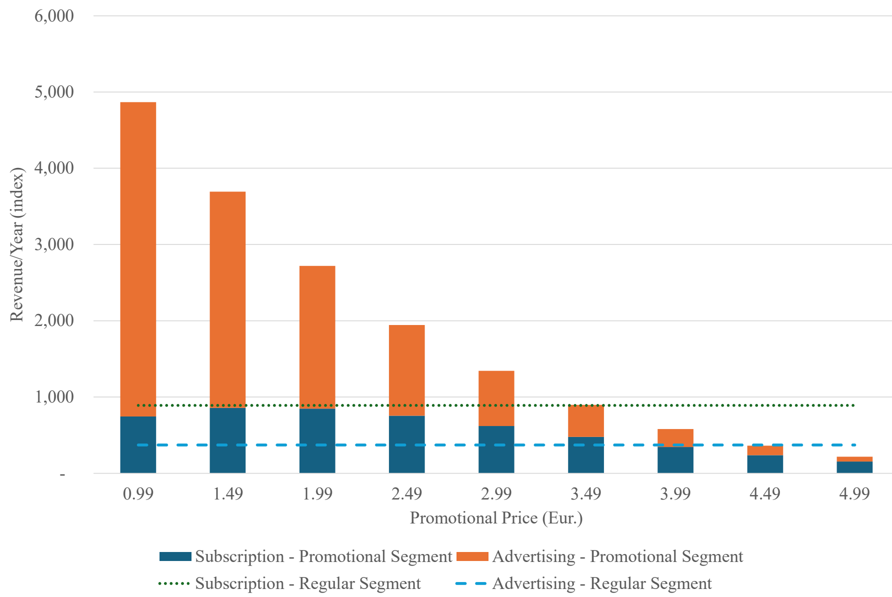
Figure 10 decomposes revenues across sources (advertising and subscription), by segment (promotional and regular). The most striking result is that advertising makes up a relatively small portion of revenue for the promotional segment when the promotional price sits above €2.99. However, at lower promotional prices it increases at an increasing rate, rapidly outweighing subscription revenue all the way to the lowest promotional price. It is clear that the promotion generated value primarily via advertising revenue.

Building on this, we consider a counterfactual analysis where advertising rates vary with the number of subscribers. First, we perform a sensitivity analysis to the unique advertising rate provided by the managerial team by, independently of the subscription base, considering situations in which the ad rate is lower than the €0.005 rate used in the previous analysis. We consider the interval [0.0005, 0.005] euros per news article read. We aim to investigate the robustness of the previous result, that the customers attracted via the price promotion are more valuable due to their advertising rents than their subscription payments.

Figure 11 plots the region in which subscription revenues dominate advertising revenue, as a function of promotional price and advertising rates. In addition, it overlays indifference curves for the firm, with curves to the left associated with higher revenues. Across the range of advertising rates considered, we find that at a low enough promotional price, advertising rents will always dominate the revenue from subscriptions.

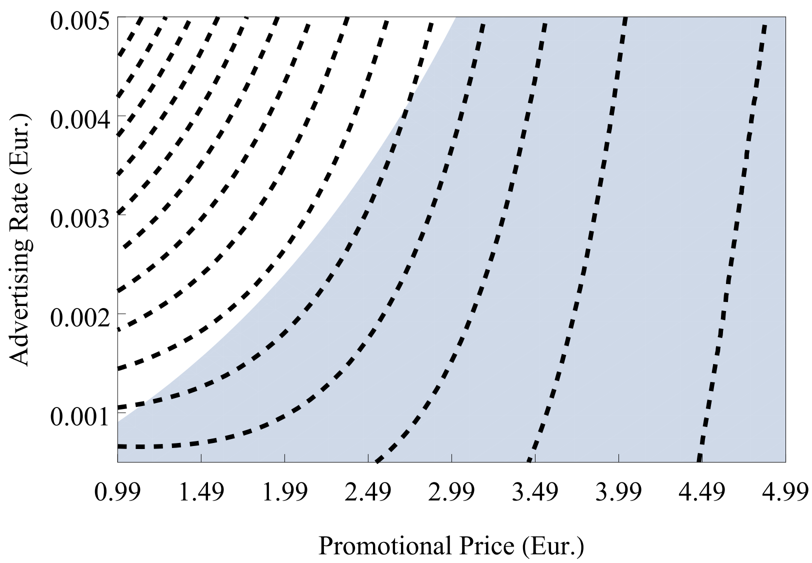
At about €0.0025 per article consumed, the gains from the promotion are approximately evenly split between subscription and advertising revenues. This means that even at half the estimate provided by the managerial team for the advertising rate, rents from advertising sourcing from the price promotion are extremely significant for the company from a revenue point of view. A central effect of the price promotion is to boost advertising revenue through increased future consumption. In the following section, we consider possible gains of price discrimination and explicitly incorporate the effect of subscription levels on advertising rates.

Figure 10: Yearly Revenue Decomposed by Segment and Source



Note: The results above pertain to the period of one year after individual subscriptions.

Figure 11: Revenue Decomposition as a Function of Promotional Price and Advertising Rate



Note: Above, the shaded area corresponds to the combinations of promotional price and advertising rate where revenues from subscriptions dominate ad revenue. Contour lines refer to total revenue (subscription plus advertising), with higher values located to the top and left. Advertising rates range from €0.0005 to €0.005 per news article read.

6.4 Charging Less for More Consumption

Subscription pricing is often complemented by second-degree pricing practices, that is, the provision of different menus with combinations of price-quantity/quality offers that consumers select into. Second-degree pricing can be profitable when a firm is able to effectively separate consumers by ensuring incentive compatibility. This strategy is less effective in the setting we analyze, because the 'best consumers' (i.e., those who would like to consume the most) are not necessarily the 'best customers' (i.e., those who are willing to pay the most) from a subscription-revenue standpoint. For example, offering a menu that includes more quantity at a higher price may fail to attract enough consumers due to a negative correlation between consumption value and WTP.[18]

It is nonetheless interesting to consider a setting in which the firm has knowledge of subscribers' ideal consumption rates. Under a perfect observability scenario, the seller could be better off discriminating heavy users positively, given the substantial growth in advertising revenue they generate and the minimal marginal costs associated with additional consumption. When consumers cannot misrepresent their types, the publisher could be better off charging, for example, €5 per month for the subscription, but then reward a select group of avid readers with a €2 reward. The rationale for such a policy lies in the fact that heavy users tend to generate disproportionately high advertising revenue, making it worthwhile for the seller to reward them accordingly.

Offering a plan that effectively charges less for more quantity can only be effective if the firm can detect real consumption (e.g., consumers actually reading articles) from fake consumption (e.g., consumers using bots to inflate article views to qualify for a discount). Given recent trends in consumer monitoring technologies (see for example Jiang, Li, Chen, and Wang (2018) and Delouya (2024)), we believe such a futuristic scenario is relevant.[19] Another reason this scenario is worth contemplating is that it serves as a useful upper bound, illustrating the potential value to the seller in a world of perfect information about individual consumption. To emphasize the uniqueness of the plan, the central idea is that consumers end up paying a lower total price for reading articles beyond a preset threshold – a structure that contrasts with classical unit-discount settings in which more quantity results in lower per-unit prices, but still entails a positive marginal cost per unit consumed.

We revisit the promotional campaign of the firm and search for the optimal minimum consumption threshold a^* that earns consumers an amount t^*. The idea is that customers who consume a number of articles beyond a^* over the span of one year earn a reward, be it through a direct monetary transfer or a discount on a renewal of their subscription, for example. For the counterfactual analysis, we maintain the promotional schedule and subscription prices of the original dataset and search for the optimal levels of a^* and t^* that maximize the firm's joint subscription and adver-

[18]In a model with two types of consumers, the fact that one of the segments is willing to pay more for less consumption than the other implies that the seller cannot take advantage of a single-crossing property to implement second-degree price discrimination. This does not necessarily apply to our case, however, since we consider a continuum of consumer types.

[19]Relatedly, in the mobile games industry apps often provide rewards for intensive and/or consecutive usage to bolster advertising revenue.

tising revenues. For consumers with consumption level $x_i^* \geq a^*$, we change the WTP from w_i to $w_i + t^*$. We keep the indirect utility fixed under the interpretation that the "windfall" t^* keeps the utilitarian component constant, but may affect consumers' WTP via a mental account, for example. This approach is conservative in that incorrectly ignoring the utility component may will lead to underestimating the total profits obtained from price discrimination.[20]

We assume a constant-elastic relationship between the advertising price and overall readership:

$$r(V) = r_0 \left(\frac{V_0}{V}\right)^{\eta} \tag{36}$$

where r_0 is the base unit advertising price (€0.005 per article read), V is the number of articles read over the duration of a year, V_0 is the number of articles read at the regular price (from the data), and η is the advertising elasticity. We consider the range $\eta \in [0.12, 0.2]$ for the elasticity parameter, which spans the estimates of the meta analyses by Sethuraman, Tellis, and Briesch (2011) and Schöndeling, Burmester, Edeling, Marchand, and Clement (2023). This specification takes into account that different levels of readership are associated with different per-unit prices of advertising. In other words, we note that large changes to readership are necessarily accompanied by changes in advertising rates.

The revenue of the firm net of the program cost is given by:

$$\begin{aligned}
\pi = {} & p\left(D_{disc}(p, a^*, t^*) + D_{regular}(p, a^*, t^*)\right) && \text{(Subscription Revenue)} \\
& + r\left(\overline{V}_{all}\right) . \left(S_{disc}(p, a^*, t^*) + S_{regular}(p, a^*, t^*)\right) && \text{(Advertising Revenue)} \\
& - t^* D_{disc}(p, a^*, t^*) && \text{(Discount Payments)}
\end{aligned}$$

where p is the regular price of €4.99, $D_{(\cdot)}$ denotes the total number of subscription payments made by the regular and discount segments, S_{disc} denotes the number of subscribers of those same segments, and parameter $\overline{V}_{all}$ is the average readership across subscribers. The discount payments above (last line of the firm revenue) are formulated so that t^* is interpreted as a per-month payment to customers. This specification provides clarity and is equivalent to a one-time payment since we abstract away from time discounting. For illustration purposes, we present the number of subscription renewals by consumers who obtain the discount, $D_{disc}(\cdot)$:

$$D_{disc}(p, a^*, t^*) = \underbrace{M\gamma}_{Potential\, Market} \underbrace{\cdot \frac{1}{R} \sum_{i=1}^{R} 1\left(x_i^* \geq a^* \wedge \alpha v_i + \varepsilon_i^1 \geq p_i + \varepsilon_i^0 \wedge w_i + t^* \geq p_i\right) .}_{Market\, Share} \underbrace{n_i}_{N.\, Payments} \tag{37}$$

where R is a preset large number of simulations. Similarly, the number of new subscribers in the

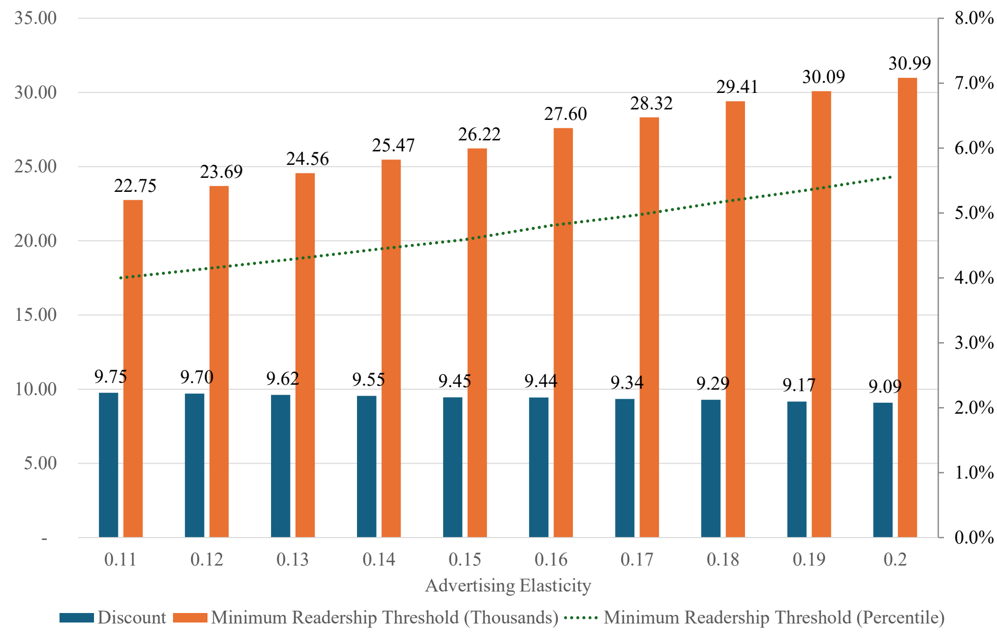
[20] The introduction of a minimum consumption threshold is related to the different problem of motivating salespeople with variable compensation schemes. In that context, the literature documents that salespeople whose performance falls just below a given compensation cutoff often exert more effort to surpass it (e.g., Misra and Nair (2011) and Chung, Steenburgh, and Sudhir (2014)). We leave the analysis of such incentives for future research efforts.

discount plan is given by

$$S_{disc}\left(p, a^{*}, t^{*}\right) = \underbrace{M\gamma}_{Potential\,Market} \cdot \underbrace{\frac{1}{R}\sum_{i=1}^{R} 1\left(x_i^{*} \geq a^{*} \wedge \alpha v_i + \varepsilon_i^{1} \geq p_i + \varepsilon_i^{0} \wedge w_i + t^{*} \geq p_i\right)}_{Market\,Share} \tag{38}$$

For each advertising elasticity level η, we maximize the publisher's profit π with respect to the minimum consumption threshold a^{*} and the discount level t^{*}. Figure 12 shows the optimal discount level t^{*} and minimum readership threshold a^{*} as a function of advertising elasticity.

Figure 12: Discount Pricing and Consumption Threshold as a Function of Advertising Elasticity



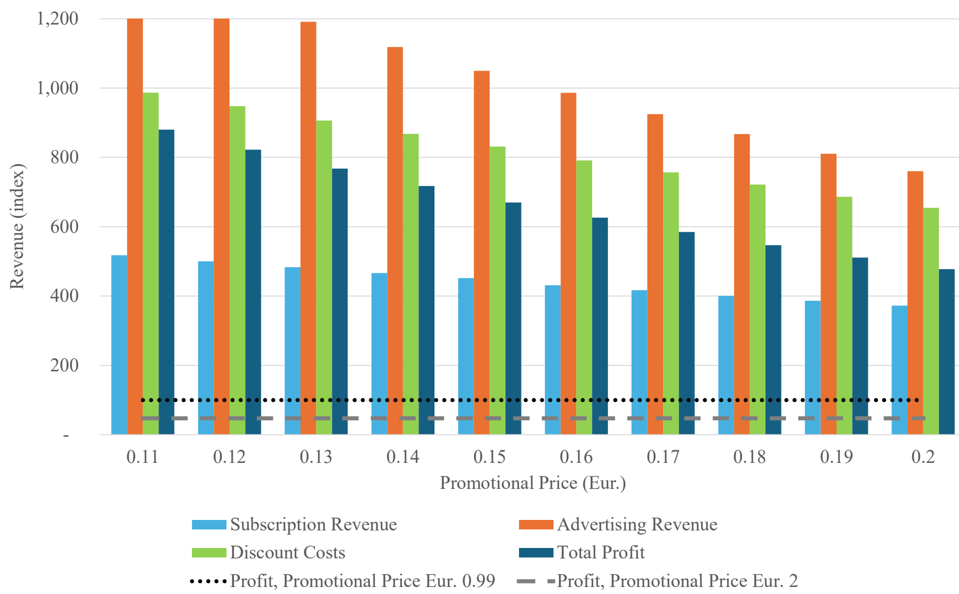
Note: The left Y axis above is used to simultaneously represent euros (short blue bars) and thousands of articles consumed (tall orange bars). The dotted line represents the same data as the orange bars in percentile terms and should be read on the right Y axis.

As the WTP of advertisers becomes more sensitive to the size of the subscriber pool, the publisher has an incentive to increase the minimum consumption threshold and decrease the consumer reward. A striking result is the fact that the per-subscription reward is higher than the regular price (€4.99); in other words, if possible, the publisher is better off paying some consumers to be part of its subscriber pool. As we show later, the loss incurred through the program cost is offset by the incremental gains in advertising revenue. Finally, the dotted line presents the minimum consumption threshold in terms of readership percentiles, to provide better interpretability about the required readership levels among the population necessary to opt into the reward. We find that the firm is never better off offering rewards to consumers below the 4^{th} consumption percentile.

Figure 13 presents normalized profits of the reward program across advertising elasticity values.

It also disaggregates profits by subscription revenue, advertising revenue, and program cost. In addition, the last bar of each cluster presents normalized profits of the program, which can be compared with the status quo represented in dotted and dashed lines. The status quo scenarios represent the firm's profit in the original promotional program of the data, with promotional prices of €0.99 (dotted line) and €2 (dashed line), as in the results presented in Figure 10.

Figure 13: Profit Decomposition of Reward Program



Note: Above, firm revenues as a function of advertising elasticity. For comparison, profits of the regular price promotion with prices of €0.99 and €2 are represented by the dotted and dashed lines.

Inspecting each cluster above, we find that subscription revenues tend to represent less than half of advertising revenues. Moreover, the cost of the reward program is very significant, surpassing the magnitude of subscription revenues. The overall effect is depicted in the last bar of each cluster. As expected, the total profit of the reward program decreases with advertising elasticity. However, it is extremely profitable throughout. The stark difference is impressive but perhaps not too surprising due to the fact that the reward program does not rely on providing information rents to enforce price discrimination. Under perfect consumption observability, the firm can effectively practice third-degree price discrimination. This counterfactual analysis shows that efforts in the direction of monitoring consumer relationships and rewarding them accordingly can be a great opportunity for firms managing subscription programs to increase profits by pricing their plans accordingly.

7 Concluding Remarks

In this paper, we investigate the relationship between subscription prices and consumption behaviors in the context of an online news publisher. We find substantial heterogeneity in consumption patterns: consumers who subscribed at the promotional price read more than three times as many articles and remained active on the platform for twice as long, on average, than those who subscribed at the regular price. These results persist even after controlling for customer churn. We develop a model that is able to capture the dynamics of this relationship. Our model takes into account substitution patterns that counter traditional economic theory, and allows us to identify the correlation between WTP and consumption intensity flexibly.

We find that failing to account for selection on unobservables can significantly distort demand estimates as well as a firm's understanding of pricing leverage. In our setting, the counterfactual analyses reveal that there is significant value from introducing price promotions, not only due to attracting additional subscription revenue but especially due to incremental advertising revenue driven by consumers who are not willing to pay the regular subscription price. Finally, we considered the case of perfect consumption monitoring, which opens up the possibility of generating significant incremental revenue through the introduction of an effective negative price on additional consumption. A firm optimizing solely for subscription revenue may underinvest in promotional pricing. Accounting for long-run advertising returns changes the calculus: deeper, shorter promotions could yield substantial net gains.

Our analysis demonstrates that even small reductions in access fees can lead to large increases in news consumption, highlighting a way to improve public access to information. This has important implications for social welfare, as lower access costs could contribute to a better-informed population. Policies designed to promote access to socially beneficial services should carefully consider the effects of pricing on consumption patterns. Understanding the relationship between pricing and consumption is not only crucial for firms evaluating revenue trade-offs, but essential for assessing the broader societal consequences of subscription pricing. This is especially relevant in an era of widespread misinformation and financial challenges for credible online news publishers. Our findings suggest that well-designed pricing interventions—whether by firms or supported by policy (e.g., VAT reductions on digital news)—can improve access to credible information and support publisher sustainability.

For future research, it is worth noting that the problem of linked decisions is pervasive in Marketing and Economics. In consumer search, consumers form expectations that link search decisions with purchase decisions. In discrete-continuous settings, consumers maximize their utilities subject to a budget constraint, which links their propensity to buy with their optimal quantity. Similarly, in subscription settings, consumers' purchase decisions correlate with their subsequent usage. We believe that investigating the extent to which classical predictions from economic theory regarding the extent to which consumers' decisions are linked, through empirical exercises, may allow researchers to advance empirical methods and also to revisit long-standing assumptions about consumer behavior.

8 Appendix

Proof of the Proposition

We start by adding a detailed mathematical structure to the proposition (omitted in the main text for readability): *Let $u(\cdot)$ be a strictly increasing differentiable utility function and x be a continuous random variable with continuous support with defined first moment and let p be a scalar. Then, $-\frac{\partial}{\partial p}P(u(x) \geq p) > 0$ and $-\frac{\partial}{\partial p}E(x|\, u(x) \geq p) < 0$.*

- **Purchase:** The effect of price on demand is given by

$$\begin{aligned}\frac{\partial}{\partial p}P(u(x) \geq p) &= \frac{\partial}{\partial p}\left(1 - F_x\left(u^{-1}(p)\right)\right)\\ &= -\underbrace{f_x\left(u^{-1}(p)\right)}_{(+)}\underbrace{\left(u^{-1}(p)\right)'}_{(+)} < 0\end{aligned}$$

  where the first positive sign follows from the fact that cumulative distribution functions are increasing and the second positive sign results from the fact that the derivative of a function is equal to the reciprocal of the derivative of the corresponding inverse function.

- **Expected Consumption:** The expected consumption is given by $E(x|\, Subscribe)$. The effect of a small price increase is given by

$$\begin{aligned}\frac{\partial}{\partial p}E(x|\, u(v) \geq p) = \frac{\partial}{\partial p}\int_{u^{-1}(p)}^{\infty} x f_{x|u(x)\geq p}(x)\, dx = \frac{\partial}{\partial p}\frac{\int_{u^{-1}(p)}^{\infty} x f_x(x)\, dx}{P(u(x) \geq p)}\\ \propto G\left(u^{-1}(p)\right) g'\left(u^{-1}(p)\right) - g\left(u^{-1}(p)\right)^2\end{aligned}$$

  where $g\left(u^{-1}(p)\right) \equiv \int_{u^{-1}(p)}^{\infty} x f_x(x)\, dx$ and $G\left(u^{-1}(p)\right) \equiv P(u(x) \geq p)$. The last proportionality relation above captures the fact that the sign of the derivative depends only on the last expression. Note that by Leibniz rule, $g'\left(u^{-1}(p)\right) = -u^{-1}(p) f_x\left(u^{-1}(p)\right)$ and $G'\left(u^{-1}(p)\right) = -f_x\left(u^{-1}(p)\right)$, such that the last expression can be rewritten as

$$\underbrace{f_x\left(u^{-1}(p)\right)}_{(+)}\left[-u^{-1}(p)\, G\left(u^{-1}(p)\right) + g\left(u^{-1}(p)\right)\right] \tag{39}$$

  Finally, note that

$$\begin{aligned}g\left(u^{-1}(p)\right) &= \int_{u^{-1}(p)}^{\infty} x f_x(x)\, dx > \int_{u^{-1}(p)}^{\infty} u^{-1}(p) f_x(x)\, dx\\ &> u^{-1}(p)\int_{u^{-1}(p)}^{\infty} f_x(x)\, dx = u^{-1}(p)\, G\left(u^{-1}(p)\right)\end{aligned}$$

  so that $g\left(u^{-1}(p)\right) > u^{-1}(p)\, G\left(u^{-1}(p)\right)$, which in turn implies that (39) is indeed positive, proving the proposition.

Funding and Competing Interests

All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript. Funding information withheld for review anonymity.

References

Albuquerque, P., P. Pavlidis, U. Chatow, K.-Y. Chen, and Z. Jamal (2012): "Evaluating promotional activities in an online two-sided market of user-generated content," *Marketing Science*, 31(3), 406–432.

Allenby, G. M., T. S. Shively, S. Yang, and M. J. Garratt (2004): "A choice model for packaged goods: Dealing with discrete quantities and quantity discounts," *Marketing Science*, 23(1), 95–108.

Arkes, H. R., and C. Blumer (1985): "The psychology of sunk cost," *Organizational behavior and human decision processes*, 35(1), 124–140.

Ascarza, E., and B. G. Hardie (2013): "A joint model of usage and churn in contractual settings," *Marketing Science*, 32(4), 570–590.

Bell, D. R., G. Iyer, and V. Padmanabhan (2002): "Price competition under stockpiling and flexible consumption," *Journal of Marketing Research*, 39(3), 292–303.

Bhat, C. R. (2005): "A multiple discrete–continuous extreme value model: formulation and application to discretionary time-use decisions," *Transportation Research Part B: Methodological*, 39(8), 679–707.

Chan, T., C. Narasimhan, and Q. Zhang (2008): "Decomposing promotional effects with a dynamic structural model of flexible consumption," *Journal of Marketing Research*, 45(4), 487–498.

Chintagunta, P. K. (1993): "Investigating purchase incidence, brand choice and purchase quantity decisions of households," *Marketing Science*, 12(2), 184–208.

Chintagunta, P. K., and H. S. Nair (2011): "Structural workshop paper-discrete-choice models of consumer demand in marketing," *Marketing Science*, 30(6), 977–996.

Chou, C., and V. Kumar (2024): "Estimating Demand for Subscription Products: Identification of Willingness to Pay Without Price Variation," *Marketing Science*.

Chung, D. J., T. Steenburgh, and K. Sudhir (2014): "Do bonuses enhance sales productivity? A dynamic structural analysis of bonus-based compensation plans," *Marketing Science*, 33(2), 165–187.

Conlisk, J., E. Gerstner, and J. Sobel (1984): "Cyclic pricing by a durable goods monopolist," *The Quarterly Journal of Economics*, 99(3), 489–505.

Daljord, Ø., C. F. Mela, J. M. Roos, J. Sprigg, and S. Yao (2023): "The design and targeting of compliance promotions," *Marketing Science*, 42(5), 866–891.

Danaher, P. J. (2002): "Optimal pricing of new subscription services: Analysis of a market experiment," *Marketing Science*, 21(2), 119–138.

Datta, H., B. Foubert, and H. J. Van Heerde (2015): "The challenge of retaining customers acquired with free trials," *Journal of Marketing Research*, 52(2), 217–234.

Deaton, A., and J. Muellbauer (1980): *Economics and consumer behavior*. Cambridge university press.

Delouya, S. (2024): "Netflix cracked down on password sharing. The result? Millions of new subscribers," *CNN-Business (Ed.), https://edition.cnn.com/2024/04/18/business/netflix-earnings-first-quarter/index.html*.

Esteves-Sorenson, C., and F. Perretti (2012): "Micro-costs: Inertia in Television Viewing," *The Economic Journal*, 122(563), 867–902.

Gardete, P. M. (2013): "Cheap-talk advertising and misrepresentation in vertically differentiated markets," *Marketing Science*, 32(4), 609–621.

Greene, W. H. (2000): *Econometric Analysis*. Prentice Hall Inc.

Hanemann, W. M. (1984): "Discrete/continuous models of consumer demand," *Econometrica: Journal of the Econometric Society*, pp. 541–561.

Hartmann, W. R., and V. B. Viard (2008): "Do frequency reward programs create switching costs? A dynamic structural analysis of demand in a reward program," *Quantitative Marketing and Economics*, 6, 109–137.

Hayashi, T. (2008): "A note on small income effects," *Journal of Economic Theory*, 139(1), 360–379.

Heckman, J. J. (1979): "Econometrica," *Econometrica*, 47(1), 153–161.

Jiang, J.-Y., C.-T. Li, Y. Chen, and W. Wang (2018): "Identifying users behind shared accounts in online streaming services," in *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pp. 65–74.

Just, D. R., and B. Wansink (2011): "The flat-rate pricing paradox: conflicting effects of "all-you-can-eat" buffet pricing," *The Review of Economics and Statistics*, 93(1), 193–200.

Kim, J., G. M. Allenby, and P. E. Rossi (2002): "Modeling consumer demand for variety," *Marketing Science*, 21(3), 229–250.

Kivetz, R. (1999): "Advances in research on mental accounting and reason-based choice," *Marketing Letters*, 10, 249–266.

KLEMPERER, P. (1987): "Markets with consumer switching costs," *The quarterly journal of economics*, 102(2), 375–394.

KRISHNAMURTHI, L., AND S. RAJ (1988): "A model of brand choice and purchase quantity price sensitivities," *Marketing Science*, 7(1), 1–20.

LEWIS, M. (2006): "Customer acquisition promotions and customer asset value," *Journal of marketing research*, 43(2), 195–203.

MISRA, S., AND H. S. NAIR (2011): "A structural model of sales-force compensation dynamics: Estimation and field implementation," *Quantitative Marketing and Economics*, 9, 211–257.

NELDER, J. A., AND R. MEAD (1965): "A simplex method for function minimization," *The computer journal*, 7(4), 308–313.

PACHALI, M. J., P. KURZ, AND T. OTTER (2023): "Omitted budget constraint bias and implications for competitive pricing," *Journal of Marketing Research*, 60(5), 968–986.

RESHADI, F., AND M. P. FITZGERALD (2023): "The pain of payment: A review and research agenda," *Psychology & Marketing*, 40(8), 1672–1688.

RUNGE, J., J. LEVAV, AND H. S. NAIR (2022): "Price promotions and "freemium" app monetization," *Quantitative marketing and economics*, 20(2), 101–139.

SCHÖNDELING, A., A. B. BURMESTER, A. EDELING, A. MARCHAND, AND M. CLEMENT (2023): "Marvelous advertising returns? A meta-analysis of advertising elasticities in the entertainment industry," *Journal of the Academy of Marketing Science*, 51(5), 1019–1045.

SETHURAMAN, R., G. J. TELLIS, AND R. A. BRIESCH (2011): "How well does advertising work? Generalizations from meta-analysis of brand advertising elasticities," *Journal of marketing research*, 48(3), 457–471.

STOLE, L. A. (2007): "Price discrimination and competition," *Handbook of industrial organization*, 3, 2221–2299.

STOURM, L., R. IYENGAR, AND E. T. BRADLOW (2020): "A flexible demand model for complements using household production theory," *Marketing Science*, 39(4), 763–787.

THALER, R. (1985): "Mental accounting and consumer choice," *Marketing science*, 4(3), 199–214.

TUCHMAN, A. E., H. S. NAIR, AND P. M. GARDETE (2018): "Television ad-skipping, consumption complementarities and the consumer demand for advertising," *Quantitative Marketing and Economics*, 16(2), 111–174.

VAN HEERDE, H. J., P. S. LEEFLANG, AND D. R. WITTINK (2004): "Decomposing the sales promotion bump with store data," *Marketing Science*, 23(3), 317–334.

Villas-Boas, J. M. (2015): "A short survey on switching costs and dynamic competition," *International Journal of Research in Marketing*, 32(2), 219–222.

Villas-Boas, J. M., and U. Schmidt-Mohr (1999): "Oligopoly with asymmetric information: Differentiation in credit markets," *The RAND Journal of Economics*, pp. 375–396.

Vives, X. (1987): "Small income effects: A Marshallian theory of consumer surplus and downward sloping demand," *The review of economic studies*, 54(1), 87–103.

Waisman, C. (2021): "Selling mechanisms for perishable goods: An empirical analysis of an online resale market for event tickets," *Quantitative Marketing and Economics*, 19(2), 127–178.