# Monetizing Serialized Content:

# How "Wait for Free" Impacts Paid and Free Consumption

Peter S. Lee, Vineet Kumar, K. Sudhir

April 16, 2024

**Abstract**

Entertainment such as novels and audio/video content are increasingly published as serialized chapters/episodes on digital media platforms. As a monetization strategy, these platforms offer a "wait-for-free" (WFF) option where consumers can consume the next episode in a sequence immediately after paying a per-episode cost or for free after waiting for a specified wait-period. Despite the growing interest in leveraging time as a discrimination device for serialized media, there is limited research on the link between time and consumption decisions, while accounting for the foundational characteristics–we do not account for this in modeling/analysis, but use it to generate hypothesis... so delete it here, but just keep it later of serialized content. Hence, we pose the following simple but critical question: Can making it easier to consume for free–does not show link to time increase revenues? While it is easy to see that a shorter wait to access free material will attract new consumers, a natural concern is that it will also cannibalize paid consumption from existing consumers. In this paper, we argue that unique characteristics of serialized content—complementarity in utility from consuming sequential episodes, the diminishing value of complementarity over time and willingness-to-pay dependent on varying time sensitivity–can actually lead to an increase in paid consumption among existing customers. By leveraging a natural experiment from a serial fiction platform, we estimate the impact of reducing wait-times on downstream consumption behaviors using a difference-in-difference framework, while addressing potential selection issues using a matching-based approach. Our results indicate a significant increase in paid consumption from existing consumers. Inclusive of the impact from new consumers, the reduction in wait-times increases daily aggregate paid consumption by 14%.

# 1 Introduction

> "In a highly competitive environment – fighting to occupy users' free time – producers and distributors of content depend more than ever on the engagement of the services they offer. The sustained, consistent and recurring consumption of narrative series has demonstrated its effectiveness above any other format in building up that engagement."
>
> (*Storytel 2018 report*, Link)

Serialized media refer to information goods such as books, TV shows or educational courses that consist of multiple episodes with a continuous plot under a single title. Modern advancements in digital media, such as the emergence of ebook and streaming platforms, have paved the way for serialization as a popular mode of publication across product types. Partitioning a unified content into short episodes has become common as it fits consumer trends such as growing media consumption on mobile devices and the diminishing attention span of the users (Marketing Charts, 2019; Speaking of Psychology, 2023). The largest platform for serialized fiction novels, Wattpad, is reported to now have over 80 million readers; the leading serialized comics platform, Webtoon, boasts over 85 million users. The video streaming giants Netflix, Amazon Prime Video and Disney+ together serve more than 600 million subscribers.

Many platforms with serialized content across a variety of product domains offer consumers a "wait-for-free" (WFF) option. For example, the WFF option is offered by Webtoon (comics), Radish Fiction (books), ReelShort (videos) and Real Racing (games). With WFF, consumers make consumption decisions for each episode in an à la carte fashion: they can access an episode immediately for a fixed price or for free after waiting for a pre-specified wait-time since accessing the prior episode of the same series. Essentially, the WFF option is a form of versioning where consumer can choose between two vertically differentiated episodes—one with zero waiting times for a price, and another with a non-zero waiting time for free. Consumers who are impatient or have high willingness-to-pay (WTP) for an episode may pay for immediate consumption, while those who have lower time sensitivity or lower WTP can choose to have free but delayed consumption.

The cross-sectional impact of versioning across different types of consumers on profits/revenues through the benefits of new user acquisition and the cost of cannibalization has received significant attention in the literature. For example, the literature on freemiums products discuss a variety of

Though the WFF option for serialized media is now widely used, there is limited research on how WFF impacts content consumption. A notable exception is Choi et al. (2022); the paper studies how WFF impacted free and paid viewing on an online platform for serialized comics. Using an aggregate log-linear regression model that explains the number of free and paid episode views in the weeks before and finds that despite the cannibalization effect of WFF, it led to an aggregate increase in the number of paid episodes consumed. The paper suggests that the increase in paid episodes arise from the

WFF generates incremental demand for new cartoons from its user base and the resulting paid viewing exceeds the effect of cannibalization through free viewing due to WFF. Using an episode-level viewership model, the paper finds that despite cannibalization, the introduction of WFF led to an aggregate increase in the number of paid episodes consumed because .

WFF led to an increase in the number of paid episodes consumed consumption on an episode-level, supporting the viability of the policy. Nonetheless, questions still remain around the driving mechanism such as implications on downstream consumption decisions at a consumer-level, as well as why the policy has predominantly been adopted for serialized content. Importantly, platforms monetizing content through the WFF policy are left uninformed about how to think about the optimal length of the wait-times–we don't answer this to maximize revenues.

To that end, this paper seeks to understand whether and how it would be possible to increase paid consumption from reducing wait-times. While it is easy to see that a shorter wait to access free content will increase free consumption, a natural concern is that it will also cannibalize paid consumption. Existing literature on vertically differentiated product design explore various mechanisms through which the relative positioning of the products affect consumption and revenues. Examples include product adoption through referrals (Lee et al., 2019; Kamada and Öry, 2020), network externalities (Shi et al., 2019), advertising revenues (Chiou and Tucker, 2013; Lambrecht and Misra, 2017) and consumer learning (Li et al., 2019; Deng et al., 2022). At the crux of these studies is the trade-off between new user acquisition and cannibalization. Increasing the value of the free option dissuades existing consumers from purchasing, but at the same time attracts new consumers who may choose to purchase. We argue that characteristics unique to serialized media can actually lead to an increase in paid consumption even *among existing consumers*.

We formalize three unique structural features of serialized media that affect consumption decisions by synthesizing existing research across literary studies, marketing and economics. First, episodes exhibit directed complementarities: consuming one episode enhances the value of the next. Each episode in a series strikes a balance between repetition and variation, diluting the idea of the ending – while adjacent episodes share structural and content similarities, they also introduce new elements that expand on the earlier episode (Eco, 1990; Kermode, 2000). This interconnectedness, in conjunction with narrative devices such as cliff-hangers, cutting-off techniques and analepses, allows the audience to slowly build their understanding of and become hooked to the storyline that gradually unfolds (Mittell, 2006). Hence, the consumption of each episode creates a stronger incentive to proceed sequentially in the series (Linkis, 2021).

Second, the value of these complementarities diminishes over the time since the last consumption, aligning with the notion that the subsequent consumption utility decreases as consumption capital dissipates over time (Becker and Murphy, 1988; Heather and Vuchinich, 2003). The internal momentum towards the series gradually decays and consumers forget about the previous episode, which may affect their consumption decisions of the next episode. This time-sensitive aspect of valuation underscores the importance of timing in release strategies for serialized media (Zhao et al., 2022; Godinho de Matos et al., 2023) and the phenomenon of binge consumption (Schweidel and Moe, 2016; Lu et al., 2019, 2023; Godinho de Matos and Ferreira, 2020), where the desire to maintain continuity leads consumers to consume episodes in close succession.

Third, the consumer's willingness-to-pay for an episode depends not only on the content and her price sensitivity, but also on her varying time sensitivity. Unlike other products, consumers must spend time to consume media products. Naturally, the urgency to consume depends on time sensitivity or the value of her time. The marketing literature on time-based discrimination (sequential release) assume a static time sensitivity within an individual (August et al., 2015; Luan and Sudhir, 2022). However, time sensitivity exogenously varies within an individual based on the value of her outside option (e.g., while taking a break vs. working), which can lead to varying willingness-to-pay for the same episode for a given consumer.

Hence, the properties of serialized media lead to a richer set of consumption dynamics that may result in unique mechanisms that increase monetization of existing consumers. Under shorter wait-times, the consumer's valuation for waited consumption increases, which may cause her to switch from purchasing to waiting at certain episodes (cannibalization effect). However, the complementarity properties give rise to positive across-episode spillovers within a consumer that counteract cannibalization through two channels. First, on episodes where complementarities are realized from consuming the previous episode, the consumer may switch from no consumption (outside option) to waiting or purchasing as both the realized and expected complementarity values from future episodes are higher. Second, this in turn allows her to benefit from complementarity on subsequent episodes, inducing her to make purchases at episodes where the realized value is sufficiently high. Moreover, by reducing the interval between consumption decisions, the platform has a greater chance of targeting consumers in states where her time sensitivity is low. In essence, by making it easier to consume for free, the firm is able to retain consumers over a broader product set and "harvest the acquisition" at episodes when the realized complementarity value is high or her time sensitivity is low.

To empirical validate our argument, we leverage data from a major U.S.-based serialized fiction novel platform that offers series under the WFF policy. We use a rich consumption panel data of over a million users and 20,000 series that covers a 15-month period, ranging from October 2020 to December 2021. The dataset details when and how the user accessed an episode (i.e., waiting or purchasing). We augment this dataset with various metadata on series and episodes such as publication date, required wait-time and promotional activities. We also use a panel data of in-app currency purchases to explore consumer heterogeneity based on historical platform spending.

We identify the causal effects of changes in wait-time on user consumption and purchase behaviors by exploiting a natural experiment where the platform unilaterally changed the wait-times for a subset of the series during our sample period. Specifically, the platform selected 191 series whose wait-times ranged from 3 to 72 hours and reduced them to 1 to 24 hours. The wait-times pre- and post-reduction as well as the magnitude of reduction vary across the selected series. The change was implemented without prior announcement and aimed to increase user engagement by making it easier for new and existing readers alike to consume episodes for free by waiting. We estimate the average treatment effect of wait-time reduction via a difference-in-differences (DiD) approach by comparing outcomes of interest within a tight 30-day time frame around the reduction.

Two empirical challenges remain: the series for which wait-times were reduced were decided by the platform, which may lead to selection bias, and the reduction was implemented in a staggered manner, which may cause a standard two-way fixed effects model to yield biased estimates. We first address selection bias using a panel matching approach (Imai et al., 2021). We compute the likelihood of a series receiving treatment (i.e., wait-time reduction) at a given time based on observed pre-treatment characteristics. By constructing a control group of series whose wait-times remain unchanged but have comparable probability of being selected for reduced wait-times, we allow the treatment to be random conditional on observable characteristics, satisfying conditional ignorability. We then mitigate potential concerns around staggered adoption by using a stacked DiD approach. The stacked DiD creates event-specific datasets for each of the treated series and its matched control series. By stacking the datasets based on relative periods around the reduction timing and saturating the model with cohort specific fixed effects, we circumvent the issues of biased estimates under staggered treatment adoption raised in Goodman-Bacon (2021).

Our empirical analysis proceeds in the following sequence. First, we examine the impact of changing wait-times on the existing consumers by examining the change in total number of episodes consumed (i.e., waited and purchased) and purchased in a series. The estimation results show a 37% increase in the total number of episodes consumed per consumer. The key estimate of interest is the impact of wait-time reduction on purchases, as the increased consumption would only hurt platform revenues if it came at the expense of lower purchases. Surprisingly, we find a 12% increase in the total number of episodes purchased per consumer, indicating that the positive effect from across-episode spillovers and varying time sensitivity dominates the negative cannibalization effect. Moreover, allowing for heterogeneous treatment effects based on historical spending on the platform reveals an even greater 17% increase for those that are amenable to paying for content. The results highlight that the same consumer who switches from purchasing to waiting for an episode may actually be monetized more in aggregate.

Second, to further validate our theoretical framework, we measure the change in consumption pace in response to reduced wait-times. Specifically, we analyze the impact on excess wait-time, or how much consumers waited in excess of what is required. Given the time-dependent complementarity value, shorter wait-times may endogenously increase the consumer's incentives to access the waited episode shortly after the wait-time elapses. Furthermore, if shorter wait-times make it more likely that the platform captures consumers at low time sensitivity, the reduction should lead to a decrease in excess wait-time. The analysis reveals a 21% decrease in excess wait-time for the existing consumers. This acceleration effect allows the consumers to progress through the series at a faster rate, leading to quicker decisions to purchase subsequent episodes.

Third, we measure the impact of wait-time reduction on the inflow of new consumers into the series. While our results thus far show the platform can extract more revenues from the existing consumers by making it easier to consume for free, the revenue contribution from new consumers still remain vital. The a priori value of a given series increases under reduced wait-times, incentivizing more consumers to start consuming the series for the first time. Focusing on the 30-day timeframe around the reduction, we find a 28% increase in the inflow of new readers.

Finally, to understand the net effect on aggregate consumption and revenues of the series, we examine the impact of the wait-time reduction on daily aggregate consumption and purchases. We find that the reduction on average leads to an 79% and a 14% increase, respectively. This shows that despite the risk of cannibalization, the shorter wait-times actually uplift platform revenues by stimulating paid consumption from existing and new users. We also report the elasticity of consumption and purchases with respect to wait-times at 0.23% and 0.05%, respectively.

We then conduct a battery of robustness checks. To show that our estimated impact of wait-time reduction is indeed causal and unbiased, we conduct formal tests to confirm the parallel trends and SUTVA assumptions. We also supplement the tests with a sensitivity analysis that explains how strongly potential unobserved confounders would need to be associated with both the outcome and the treatment in order to explain away the estimated treatment effect. Moreover, to rule out spurious correlations that might drive our results, we conduct falsification tests using pseudo treatment series and dates. In addition, analyses using different subsamples of the data and model specifications show broad agreement with our main results.

The rest of the article is organized as follows. First, we discuss how this research is related to previous literature. Second, we describe institutional details and data. Third, we describe the empirical strategy and the econometric model. Fourth, we present the results and discuss the key findings. Finally, we conclude and provide future research directions.

## 2 Relationship to the Literature

Our paper contributes to the well-established literature on versioning. Versioning is a widely practiced and studied price discrimination strategy where the firm offers vertically differentiated products at different prices such that consumers with heterogeneous preferences self-select in to the the version-price pair that is targeted to them. Building on the theoretical works on product differentiation (Mussa and Rosen, 1978; Deneckere and McAfee, 1996), Shapiro and Varian (1998) and Varian (2000) laid the groundwork for versioning information goods, pointing out the economic feasibility of manipulating product quality at negligible marginal production costs.

A set of studies have investigated the economic viability of versioning information goods with an emphasis on freemium strategy. Kamada and Öry (2020) models consumers' referral behaviors to show that referral rewards and freemium contracts can be adopted to encourage word-of-mouth. Shi et al. (2019) shows that in the absence of such diffusion dynamics, the freemium model can be optimal when the two products provide asymmetric network externalities. Existing empirical research probes into the impact of introducing the free version on demand for the premium version (Gu et al., 2018; Li et al., 2019; Deng et al., 2022), when firms should charge for content (Lambrecht and Misra, 2017) and how much should be provided for free (Lee et al., 2019). In many of these cases, the key determinant of the firm's decision is the trade-off between extensive margin and cannibalization: an attractive free offering expands the consumer base, but does so at the cost of cannibalizing existing paid consumption. Our study enriches this dynamic by focusing on the intensive margin that has not been examined in the literature. Complementarity between episodes lead to increased consumption in a series per consumer, and the additional purchases made on subsequent episodes offset or even dominate the negative cannibalization effect, leading to greater monetization per consumer.

Moreover, there is limited research that exploits exogenous changes in version quality to empirically investigate the causal implications. Product quality is hard to quantify and empirical settings involving a discrete change in version quality are uncommon. An exception is Li et al. (2019), where they exogenously vary the resolution of free ebook samples. The authors distinguish between "sample quality" and "functional equivalence," where the former is the degree to which the free sample reveals the quality of the premium product (e.g., textual content, image resolution) and the latter is the degree to which the utility derived from the premium product can be obtained from the free product (e.g., limited usage time, hardcover vs. softcover). By varying sample quality, the authors show that under low functional equivalence, it may be profitable for the firm to provide high quality free samples as they serve as poor substitutes. In our study, we exploit an exogenous change in functional equivalence – consumers get full access to the exact same episode by waiting but receive lower utility from delayed consumption. Our results demonstrate that owing to the unique features of serialized media, closing the gap between the two versions in terms of functional equivalence can increase firm revenues.

Our work also relates to the literature on sequential product release or using time as a discrimination device. Firms often start with limited distribution through their primary channel and after some time release a secondary channel for mass distribution that sell at a lower margin (e.g., movie theater vs. DVDs, hardcover vs. softcover). If the inter-release timing is too short, forward-looking consumers might hold off on their purchases through the first channel, and if too long, consumers gradually lose interest (buzz decay) and decide to exit the market by the time the second channel opens (Calzada and Valletti, 2012; August et al., 2015; Luan and Sudhir, 2022). While the wait-for-free policy shares commonalities, the release timing is personalized based on the user's consumption timing and is applied to a series of products, giving way to distinct consumer dynamics.

With the proliferation of serialized media content, a growing stream of literature is giving attention to their monetization strategies. Zhao et al. (2022) investigates the connection between episode release timing and a rich set of behaviors such as binge consumption, rationing and platform visits to study the platform's optimal release schedule. By estimating a structural model of purchase and consumption of chapters within a book, the authors find evidence of both within-period binge consumption (i.e., conditional on availability, the consumer will consume a large number of chapters of a book within a period) and intertemporal binge consumption (i.e., consumption of a book in the last period begets consumption in the present period). These findings are consistent with our proposed complementarity properties of serialized media. The authors conclude that a hybrid strategy of simultaneous and sequential release strategies yield highest platform profits. Similarly, Godinho de Matos et al. (2023) devise an analytical model of consumer search and consumption on a TV show streaming platform, and conduct a field experiment to show that a drip-style content release schedule leads to higher platform usage. Using data from a comics platform that allows early access for a fee, Choi et al. (2023) finds that habit formation gradually increases consumers' valuation and prompts them to pay for early access. To our knowledge, there are no papers that study the WFF policy with the exception of Choi et al. (2022) due to its novelty. The paper explores a setting where a comics platform adopted WFF policy on a subset of its comics for the first time and finds that the introduction resulted in a boost in free and paid viewership. Our work complements these results by exploiting an exogenous variation in wait-times in a WFF policy to show how the policy can effectively monetize the platform customer base.

Finally, another related stream focuses on consumption behaviors of serialized media. Zhang et al. (2022) provides evidence of time-inconsistent preferences, where consumers intentionally choose to overpay for content in order to curb future consumption (strategic self-control). Several works study the phenomenon of binge consumption, exploring the implications on downstream behaviors such as responsiveness to advertisements, series completion and spillovers to other content on the platform (Schweidel and Moe, 2016; Lu et al., 2019, 2023; Godinho de Matos and Ferreira, 2020). These papers focus on providing empirical evidence of binge consumption and abstract away from the mechanism that drives binge consumption. We add to this stream of literature by shedding light on the role of complementarity properties of serialized media that affect consumption decisions.

# 3 Institutional Details and Data

In this section, we describe the institutional details of our empirical setting and subsequently explain the data used in our empirical analysis.

## 3.1 Institutional Details

The serialized fiction market consists of three players: authors, readers and the two-sided platform. Independent authors publish their series that are comprised of multiple episodes on the platform, and the readers access each episode through a mobile application following various payment schemes. The serial fiction market has seen rapid growth globally, with notable platforms such as Wattpad and Kindle Vella.

We leverage data from a leading U.S.-based serial fiction platform specializing in the romance genre that hosts over 20,000 series and has over a million active users. The platform generates revenues through users' episode purchases. Specifically, each series belongs to one of three categories depending on the monetization type: free, premium and "wait-for-free" (WFF). Free series allow immediate access to all episodes at no cost. Premium series follow a pay-per-episode model where first several episodes are free, and users must pay using an in-app currency ("Coins") to "unlock" each subsequent episode. The only difference between WFF and premium series is that WFF allows users to unlock an episode for free once a pre-specified wait-time has elapsed after the last episode of the same series was unlocked. Like premium series, readers may alternatively pay to unlock the episode immediately. Coins can be bought with real money, and each episode costs 3 Coins regardless of wait-times, roughly equivalent to 50 cents. Users can earn Coins through other ways such as referrals, watching ads or giveaway events hosted by the platform, but those make up a negligible portion compared to direct purchases.

The wait-time varies across series, ranging from 1 to 72 hours, and the same wait-time applies to all episodes and readers within a series. If the consumer becomes eligible to unlock an episode after the wait-time has elapsed, she must actively unlock an episode to "reset the clock" for the next free episode. Take for example a series that requires a 3-hour wait-time. A user may consume the entire series for free as long as she is willing to wait at least three hours *between each episode*. The consumer returning in 12 hours will only have a single free episode available rather than four. Hence, the frequency of visits matters, and one cannot "wait-and-binge," a behavior often observed for series with fixed release schedules. Also, note in this context that there is no incentive for a user to purchase an episode now to read later (i.e., stockpile), since it will eventually become free after waiting. To be clear, although firms have previously discriminated using time in contexts such as hardcover versus softcover books, the application to serialized media, where there are complementarities across episodes, has not been previously observed. Moreover, the release timing in WFF is personalized based on the user's consumption of the previous episode and is applied separately for each episode.

Figure 1 illustrates the distribution of series and consumption across the three sales types. Although WFF series constitute a third of all series on the platform, more than 85% of episode consumption in our dataset is generated by the WFF series. Given this pattern and our research objective, we focus only on the WFF series within the data.

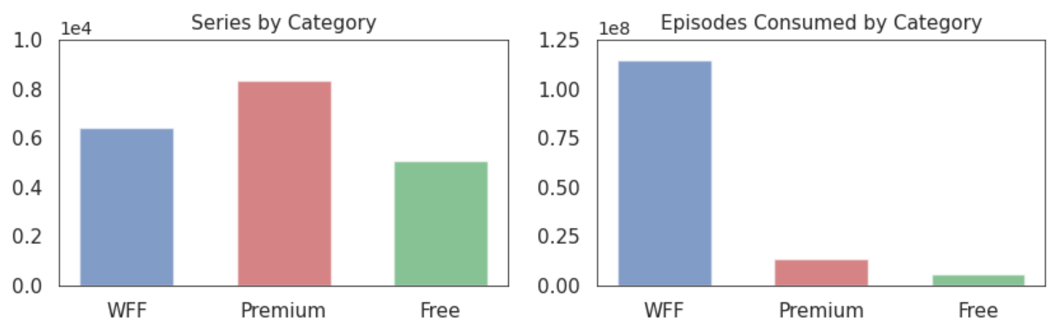


Figure 1: Distribution of series and episodes consumed across categories

The user can scroll through series available on the platform, and once she clicks on a series, additional relevant information is displayed, such as the wait-time, genre and a short description. In this example, "God of Wolves" is a paranormal romance series that requires a 3-hour wait-time and offers the first five episodes for free. The hourglass icon and "3 Coins" indicate that beginning with the sixth episode, the user may either wait 3 hours or pay 3 Coins to unlock.

On the left panel of Figure 2 is a sample episode. A typical episode is around 1,500 words, and the vast majority of the readers finish an episode within 15 minutes. At the end of the last unlocked episode, the user is presented with an option to pay to read now or wait to read for free, as shown in the right panel. The pink text on the bottom shows the wait clock ticking down.

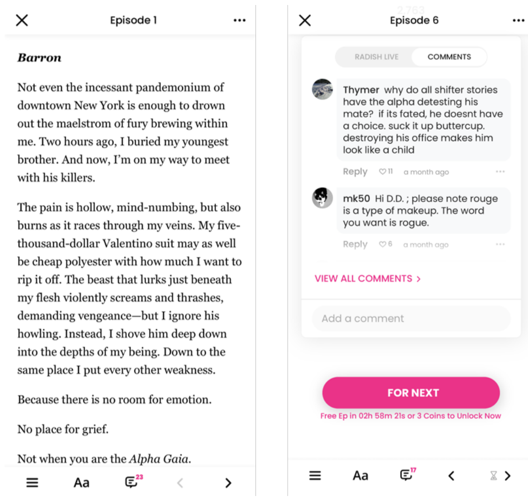


Figure 2: Wait-clock starts ticking down upon accessing the last available episode

## 3.2 Data

We leverage multiple datasets that cover user consumption as well as series and episode metadata. The consumption panel data, which covers 15 months from October 1, 2020 to December 31, 2021, matches every user with every episode that she has accessed, along with the exact time and whether she waited or purchased the episode. Series metadata include title, genre, author, sales type, date of first publication and the required wait-time for an episode. Episode metadata include series ID, sequence in the series, episode publication date and word count. The dataset also contains information on promotional activities where the platform offered coupons for specific series that can be used to unlock an episode, including the promotion dates and how many coupons were used. Finally, we have access to panel data on Coin purchases from January 1, 2019 to October 30, 2022, which we leverage to explore heterogeneity across users based on historical spending behavior. Our expansive dataset comprised of detailed access timing and method over an extended time window presents a unique opportunity to delve into the consumption dynamics of serialized media.

To isolate the effect of wait-time reduction on the *existing reader base* of the platform, we filter the panel data to the readers that joined the platform before October 1, 2020, the beginning of our consumption panel data. This ensures that there are no compositional changes to the platform users that might otherwise introduce a form of selection bias. This also means that our analysis results are conservative estimates that exclude the contribution from users that join the platform at a later date. Moreover, to reduce noise from tail end series that are rarely read, we filter for series with at least 1,000 episode accesses during over the entire observation period. Our resulting dataset covers 1,940 WFF series and 308,681 users, basic summary statistics provided in Table 1. The median series contains 44 episodes from a single season, and the median user has read two series and 44 episodes during our observation period.

| | Mean | SD | 25% | 50% | 75% |
|---|---|---|---|---|---|
| Episodes per series | 79.1 | 163.2 | 31.0 | 44.0 | 83.0 |
| Series consumed per user | 14.1 | 40.2 | 1.0 | 2.0 | 8.0 |
| Episodes consumed per user | 367.0 | 1016.0 | 7.0 | 44.0 | 233.0 |
| Episodes waited per user | 312.8 | 1233.7 | 2.0 | 18.0 | 141.0 |
| Episodes purchased per user | 132.4 | 392.3 | 1.0 | 16.0 | 87.0 |

Table 1: Summary statistics for the main dataset

We next provide a set of descriptive statistics. The left panel of Figure 3 illustrates the distribution of series by the size of their reader base, with the x-axis indicating unique reader count (log-transformed) and the y-axis indicating the number of series. The log-normal histogram shows a heavy concentration of readers on the most popular series. The right panel of the figure illustrates the distribution of users by their total spending on the platform to purchase Coins. The most notable pattern is that there is a segment of users that do not spend any money on the platform (i.e., "non-spenders") that make up 53% of the user base. These users do not make any direct contributions to platform revenues and mostly resort to free consumption. We note that although the non-spenders do not purchase Coins, they may still purchase episodes using Coins earned through alternative methods mentioned above (e.g., referrals, ads, giveaway events). The remaining 47% of the user base show a log-normal distribution, indicating a long tail of heavy spenders. One could speculate that wait-times might affect consumers differently based on inclination to spend money, and the platform's greater concern is on the consumption decisions of the spenders who directly contribute to revenues, which calls the need to examine heterogeneity in the effect of wait-time changes across consumers based on historical spending in the empirical analysis.

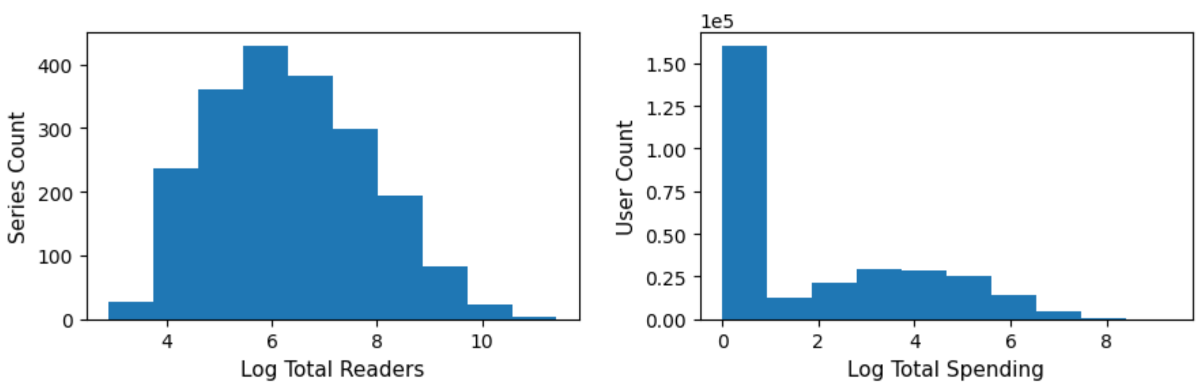


Figure 3: Distribution of readers across series and purchase propensity

Next, we explore how consuming an episode impacts the likelihood of the consuming the next episode in the series. To do so, we sample the consumption panel data for 10,000 randomly selected users. We find that 99% of the episodes are read along with the immediately preceding episode, and 90% of the episodes are read in sequential order. A logistic regression analysis reveals that reading an episode significantly increases the likelihood of reading the immediately subsequent episode ($p < 0.001$). In other words, the vast majority of readers read the series in the specified order of episodes and read an episode only if they have read the preceding episode.

We conclude this section by exploring consumption patterns across reading sessions. Because the dataset captures when and for how long each user reads an episode, we are able to define reading sessions at a granular level. For each user, we define a reading session to be one where the interval between finishing an episode and starting the next episode is less than 30 minutes. Table 2 shows the proportion of reading sessions by the number of episodes read and the number of unique series read during the session. Given that 60% of reading sessions consist of a single episode and a median user reads four episodes per day (Table 1), we can infer that users visit the platform multiple times throughout the day mostly just to access the waited episodes and occasionally end up purchasing another episode. Moreover, the table also suggests patterns of binge consumption. 40% of reading sessions involve two or more episodes, and irrespective of session length, at least 25% of sessions involved reading episodes from a single series. These patterns lend empirical support to the aforementioned features of serialized media – i.e., directed complementarity between episodes that diminishes over time.

| **Session Length (eps)** | **% Sessions** | **mean** | **25%** | **75%** |
|---|---|---|---|---|
| 1 | 58.6 | 1.0 | 1 | 1 |
| 2 | 16.2 | 1.5 | 1 | 2 |
| 3 | 7.9 | 1.8 | 1 | 3 |
| 4 | 4.6 | 2.0 | 1 | 3 |
| 5 | 2.9 | 2.2 | 1 | 3 |
| 6 | 2.0 | 2.3 | 1 | 3 |
| 7 | 1.5 | 2.4 | 1 | 3 |
| 8 | 1.1 | 2.4 | 1 | 3 |
| 9 | 0.9 | 2.5 | 1 | 3 |
| 10+ | 4.3 | 2.7 | 1 | 3 |

Table 2: Unique series read, conditional on the number of episodes consumed in the session

# 4 Empirical Strategy and Model

## 4.1 Empirical Strategy

In this section, we discuss our empirical strategy to identify the causal effect of wait-time reduction on reader consumption and purchase decisions. An ideal experiment would randomly assign different readers to different wait-times for a given series and compare outcomes between experimental designs. However, our setting features the same wait-time for all readers for a given series. Thus, we instead leverage exogenous policy changes implemented by the platform. Specifically, the platform unilaterally reduced wait-times for a certain set of series in varying degrees in a staggered manner within our observation period. For example, users that had to wait 24 hours to unlock an episode would now be able to unlock an episode every hour after the reduction. The platform indicated that the objective of the policy change was to increase overall reader engagement, but the process was rather informal without analytical evidence that called for the change or a systematic criteria. The wait-times before and after the reduction, as well as its timing varies across series, but importantly, there were no prior announcements by the platform regarding the reduction. Hence, readers could not have expected any changes to the wait-time in advance, making it exogenous to them. Similar instances can be found on other platforms where they unexpectedly implemented changes on the terms of the WFF policy such as raising the price of purchased episodes, extending the wait-time for free episodes, or making the waited episode accessible only for a limited duration (Webtoon 2022; Tapas Forum 2022).

We identified 191 series that had wait-times reduced in our dataset, which we call the *treated series*. The rest of the 1,749 series did not have any changes to their wait-times, which we call *non-treated series*. Figure 4 shows the distribution of the WFF series before and after treatment. As the wait-times were reduced, we see an increase in the proportion of series with 1-hour wait-times post-treatment and a corresponding decrease in the proportion of series with longer wait-times. A detailed breakdown of wait-times pre- and post-treatment is presented in Table 3 lists the number of series from our dataset based on pre- and post-wait-time changes: the diagonal figures represent the series that did not experience changes, and the off-diagonal figures represent the treated series. Figure 5 provides additional information about the policy change, including the distribution of treated series by adoption timing (left panel) and magnitude of wait-time reduction in hours (right panel).

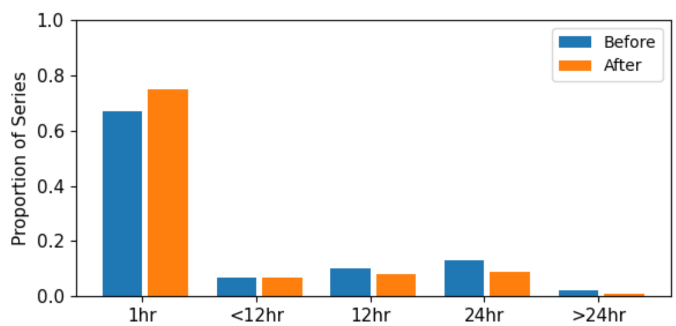


Figure 4: Distribution of series by required wait-time before and after treatment

We provide a set of model-free evidence illustrating the impact of wait-time reduction on consumption patterns of existing and new consumers. The left panel of Figure 6 compares the average number of episodes purchased by an existing consumer (log-transformed) before and after the reduction. To maintain com-

| pre/post | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 10 | 12 | 24 | 48 | 72 | All |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1304 | - | - | - | - | - | - | - | - | - | - | - | - | 1304 |
| 2 | - | 40 | - | - | - | - | - | - | - | - | - | - | - | 40 |
| 3 | 12 | - | 20 | - | - | - | - | - | - | - | - | - | - | 32 |
| 4 | 6 | - | - | 38 | - | - | - | - | - | - | - | - | - | 44 |
| 5 | 1 | - | - | - | 4 | - | - | - | - | - | - | - | - | 5 |
| 6 | 6 | - | - | - | - | 7 | - | - | - | - | - | - | - | 13 |
| 7 | 1 | - | - | - | - | - | 1 | - | - | - | - | - | - | 2 |
| 8 | 1 | - | - | - | - | - | - | 6 | - | - | - | - | - | 7 |
| 10 | 1 | - | - | - | - | - | - | - | 1 | - | - | - | - | 2 |
| 12 | 45 | - | 3 | 1 | - | - | - | - | - | 139 | - | - | - | 188 |
| 24 | 62 | - | 21 | - | - | - | - | 2 | - | 8 | 165 | - | - | 258 |
| 48 | 16 | - | - | - | - | - | - | - | - | 1 | 3 | 23 | - | 43 |
| 72 | 1 | - | - | - | - | - | - | - | - | - | - | - | 1 | 2 |
| All | 1456 | 40 | 44 | 39 | 4 | 7 | 1 | 8 | 1 | 148 | 168 | 23 | 1 | 1940 |

Table 3: Number of series by wait-time (hrs) for pre- and post- change

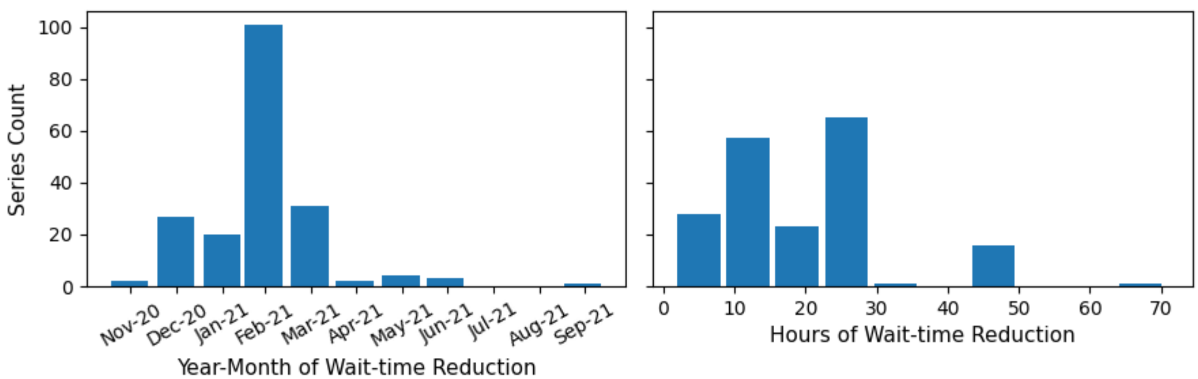


Figure 5: Distribution of treated series by treatment timing and reduction magnitude

parability, we randomly sampled ten series that were completed prior to the first observation date of the consumption panel data (i.e., no new episodes published). We see a noticeable increase in paid consumption under reduced wait-times. A one-sided paired $t$-test at the consumer-level using the entire set of treated series indicate that purchase per consumer is significantly higher post-reduction ($p < 0.01$). This evidence lends itself to the possibility that existing consumers may actually purchase more episodes net of cannibalization. The right panel illustrates the number of unique consumers (log-transformed) for a given series before and after the reduction for the same set of randomly selected series. The positive slopes indicate an expanding consumer base under shorter wait-times, potentially generating incremental revenues for the platform. Again, a one-sided paired $t$-test at the series level indicate a significantly higher consumer base for the series post-reduction ($p < 0.01$).

Make it clear in figure below caption, what subset of selected series were used. Alternatively, we visualize how retention across episodes in a given series changes upon the wait-time reduction. Figure 7 illustrates the

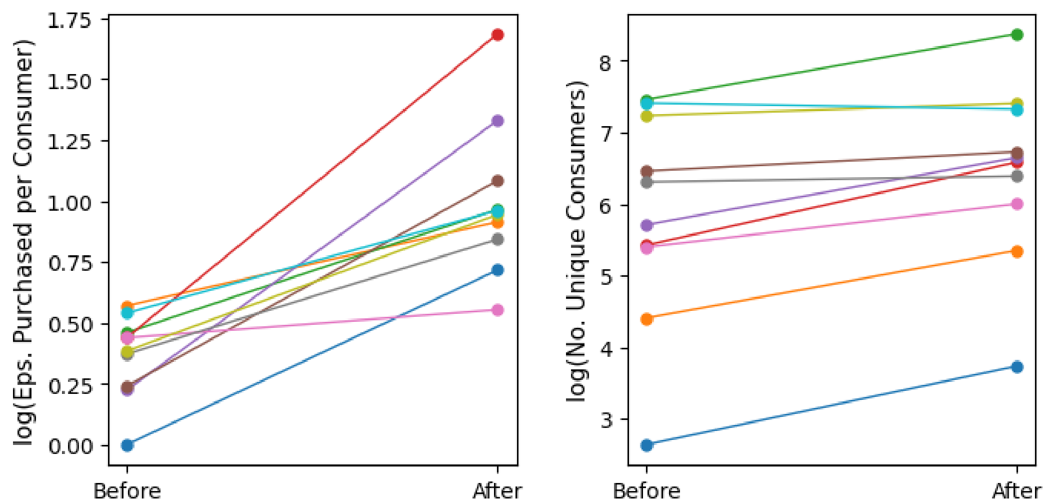


Figure 6: Comparison of consumption patters for treated series before and after wait-time reduction

average proportion of readers that consume the first episode that proceed to consume subsequent episodes across the treated series, separately for before and after the reduction. The plots overall indicate high churn especially in the early episodes, and less than 40% of the readers remain by the fiftieth episode. Importantly, we see a noticeable increase in retention under reduced wait-times. This lends empirical support to our theoretical framework: as the consumer waits less for free access, she receives a higher complementarity value and thus decides to consume the episode when her time sensitivity is low rather than churn, which in turn allows her to benefit from complementarity with the next episode. Such increase in retention may result in increased purchases despite the higher incentives to wait for free. Overall, Figures 6 and 7 provide evidence that necessitates the need to construct a robust causal model that considers the impact of wait-time reduction on consumption while controlling for a host of observable and unobservable features.

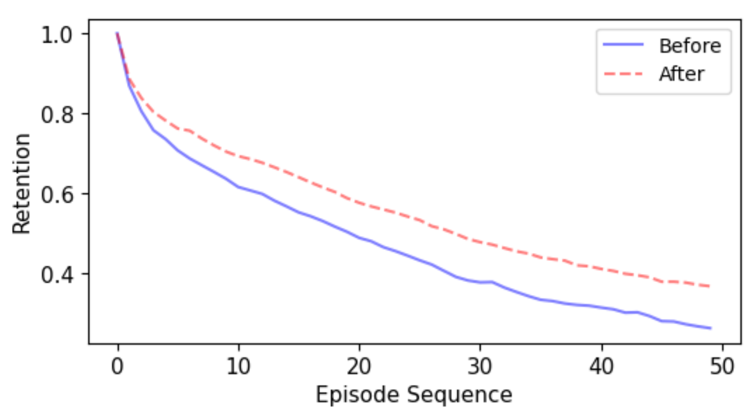


Figure 7: Retention of readers across episodes

### 4.1.1 Empirical Challenges

To identify the effect of shortening the wait-time on individual consumption and aggregate demand, we focus our analysis on a tight window around the reduction. The assumption is that any changes within this brief time period can be attributed only to the wait-time reduction, controlling for a comprehensive set of features. Comparing the treated series to a set of appropriately constructed set of control series with no wait-time changes, we can estimate the average effect of wait-time reduction using a difference-in-difference (DiD) framework.

Our empirical context poses two main challenges. The first challenge is the selection into treatment. Although the platform confirmed that they did not have a specific selection criteria for the treated series, the selection of treated series could be endogenous, which can potentially lead to biased results when naively comparing series that are systematically different. For example, if the treated series have previously been more widely read, then the estimated treatment effect may be downward biased, as the wait-time reduction will have no effect on users that have already read the series. If the platform selected longer series, then the estimate may be upward biased, as more episodes are affected by the reduction.

The second challenge is that we have an unbalanced panel data with variation in treatment timing. Since series are published on or removed from the platform at different points in time, the observed time window varies across series (only about 6% of the series are removed during the observation period). The missing observations can lead to differences in trends before treatment, making the parallel trends assumption difficult to assess and justify. Moreover, recent econometrics literature has shown that variation in treatment timing can lead to biased average treatment effect (ATE) estimates in a two-way fixed effects (TWFE) model, especially in the presence of heterogeneous treatment effects (Borusyak and Jaravel, 2018; de Chaisemartin and D'Haultfœuille, 2020; Goodman-Bacon, 2021). Specifically, the "forbidden comparison" of later treated units to already treated units may assign negative weights to certain sample treatment effects, thereby making the estimated ATE markedly different from the rest of the sample treatment effects.

We address these challenges by using panel-matching approach (Imai et al., 2021) and a stacked DiD model (Cengiz et al., 2019; Deshpande and Li, 2019; Baker et al., 2022; Deng et al., 2022). We first match each of the 191 treated series to a *matched control set* that consists of non-treated series that are fully observed around the treatment timing and have similar propensity score, which is the conditional probability of receiving the treatment given the pre-treatment covariate histories. Hence, treatment assignment is independent of potential outcomes conditional on potential outcomes, satisfying the conditional independence assumption. Any difference in observed outcomes between the matched treated and control units can be attributed to the treatment effect, thereby approximating a randomized experimental design.

We then estimate the treatment effect using the stacked DiD method, which focuses on a fixed time window around the treatment event for each treated series, effectively creating a series of "mini" DiD analyses centered on the point of treatment adoption. This approach stacks these fixed time windows to form a consolidated dataset, within which the treatment effect is estimated using a DiD model that incorporates group-specific fixed effects. By doing so, the stacked DiD model ensures that the estimation of treatment effects is grounded in a comparison of treated and control units within narrowly defined temporal contexts, thereby restoring the validity of the parallel trends assumption and reducing the risk of biased estimates arising from heterogeneous treatment effects over time. This methodological refinement allows for a more precise estimation of the treatment effect, accounting for the nuanced dynamics of staggered treatment adoption. We provide additional details on how we address the empirical challenges in the following section.

## 4.2 Constructing a Matched Control Set via Panel Matching

In order to address the potential systematic differences between the treated and the non-treated series, we create a control group for each treated series by matching it with non-treated series that have similar probability of being treated. By making treatment independent of observed potential confounders (i.e., conditional independence assumption: $Y(0), Y(1)) \perp\!\!\!\perp T|X$), we are able to draw causal conclusions about the impact of reduced wait-time by comparing the two groups. We utilize the propensity score matching procedure for time-series cross-section data (panel-matching) developed in Imai et al. (2021). Despite the popularity of matching methods, almost all of the existing methods assume a cross sectional dataset using static features measured at a point in time (Abadie and Imbens, 2011; Diamond and Sekhon, 2013; Hansen, 2004). Studies involving a panel dataset compute the average of time-varying covariates over a static time-frame (Datta et al., 2018; Deng et al., 2022; Narang and Shankar, 2019), which can miss out on important time-varying factors such as demand trends leading up to treatment that affect selection into treatment.

In our setting however, the potential demand-related confounders (e.g., waited/purchased consumption, number of readers) are time-varying, and the variation in treatment timing makes it difficult to define a single pre-treatment period for the non-treated series. Moreover, matching on the average of time-varying covariates might match series whose covariates are similar on average but exhibit very different temporal trajectories. For example, a series that is gaining traction among readers and one that is becoming increasingly unpopular prior to treatment will clearly experience different effects from reduction in wait-time. Furthermore, since we have an unbalanced panel data with staggered treatment adoption, we must match each treated series to non-treated series that are observed in the same time window. As an illustrative example, Figure 8 is a treatment variation heatmap from a random sample of series. Each row represents a series, and each column represents a week from our dataset. The red (blue) areas represent treated (non-treated) series-week observations, and white areas indicate no observation (weeks when the series was not on the platform). We want to match each treated series to non-treated series that are fully observed (blue areas) around the treatment timing and are comparable in covariate values leading up to treatment.

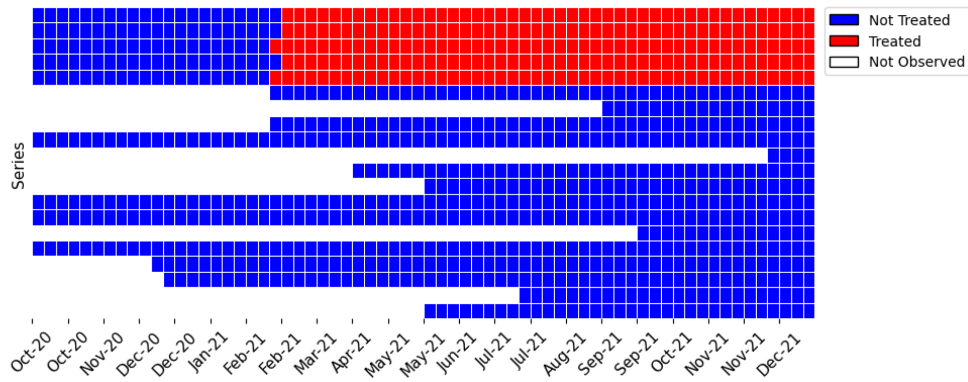


Figure 8: Treatment Variation Plot

**Constructing the Matched Control Set** We now describe the matching procedure in detail. Let us denote a treated series $s$ that receives treatment for the first time in period $t$ as observation $(s, t)$. For each treated observation $(s, t)$, we construct the matched set of never-treated units that are fully observed from time $t - L$ to $t - 1$. Imai et al. (2021) allows the matched set to include not-yet treated units, but we only allow for never treated units for a more robust comparison. Figure 9 illustrates an example of how matched sets are created when $L = 2$, indicated by the color of the boxes around the observations. In this example, treated series $s = 0$ is matched to non-treated series $s \in \{3, 4\}$ over weeks $t \in \{0, 1\}$ (blue color). Note that those non-treated series are fully observed in the two weeks prior to the treatment timing of series $s = 0$. Series $s \in \{1, 2, 5\}$ are not included in the matched set because they are either eventually treated at a later date or are not fully observed. Similarly, series $s = 1$ is matched to series $s \in \{4, 5\}$ (red color), and series $s = 2$ is matched to series $s \in \{3, 4, 5\}$ (green color), each in their respective time window. In our case, we set $L = 4$, which assumes that adjusting for covariate trends up to previous four weeks removes most of the possible confounding. Formally, the matched set for observation $(s, t)$ is defined as

$$M_{st} = \{s' : s' \neq s,\ D_{s't'} = 0\ \forall\ t' = t, t-1, \ldots, t-L\} \tag{1}$$

where $D_{st}$ is an indicator equal to 1 if series $s$ is treated at time $t$ and 0 if not.


Figure 9: Illustrative example of constructing the matched control set. The color of the boxes indicate the treated and matched control units included in the same set. For example, treated series $s = 0$ is matched to non-treated series $s \in \{3, 4\}$ over weeks $t \in \{0, 1\}$ (blue color).

**Refining the Matched Sets** The previous matched sets only account for treatment and observation history. Next, we refine the matched sets based on propensity scores, the conditional probability of treatment assignment given observed pre-treatment covariates (Rosenbaum and Rubin, 1983). The propensity score is computed using a logistic regression based on a rich set of observed covariate histories prior to treatment that can reasonably discriminate the treated and non-treated series:

$$e_{st}(\{V_{s,t-l}\}_{l=1}^{L}) = Pr(D_{st} = 1 | V_{s,t-1}, ..., V_{s,t-L}) = \frac{1}{1 + exp(-\sum_{l=1}^{L} \beta_l^T V_{s,t-l})} \tag{2}$$

where $V_{s,t}$ is a matrix of observed static and time-varying covariates for series $s$ in week $t$. The covariates used in the logistic regression include weekly count of waited/purchased episodes, series length, weeks since the series was first published and the required wait-time. The use of endogenous pre-treatment variables (i.e., waited and purchased episodes) to compute propensity score is consistent with the existing research that utilize covariates such as lagged outcomes, consumer spending and income (Heckman et al., 1998; Dehejia and Wahba, 2002). These covariates serve as critical proxies for latent variables that might influence both the selection into treatment and the post-treatment outcomes of interest. By incorporating these variables, we are able to indirectly adjust for unobservable confounders and satisfy the parallel trends assumption.

Given the fitted model, we compute the estimated propensity score $\hat{e}_{st}$ for all treated observations and their matched sets. Among the series in the matched set whose propensity score distance to the treated unit is less than a defined caliper ($C = 0.1$), we select up to $N$ series (or all units if fewer than $N$ satisfy the criterion) with replacement. Because the treatment timing varies across the treated series, potential concerns about over-reliance on specific control units from matching with replacement are mitigated. Formally, the refined matched set for the treated observation $(s, t)$ is given by:

$$M_{st}^* = \{s' : s' \in M_{st},\ |\hat{e}_{st} - \hat{e}_{s't}| < C,\ |\hat{e}_{st} - \hat{e}_{s't}| \leq (|\hat{e}_{st} - \hat{e}_{s''t}|)^{(N)}\} \quad (3)$$

where $(|\hat{e}_{st} - \hat{e}_{s''t}|)^{(N)}$ is the $N^{th}$ order statistic of the propensity score distance to the treated unit among the units in the original matched set.

**Covariate Balance Diagnostics** The number of matches, $N$, is set such that the best covariate balance between the treated series and their matched counterpart is achieved. Figure 10 compares the covariate balance across $N = 1, \ldots, 10$ (x-axis) based on $p$-value (left y-axis) and Kullback–Leibler divergence (right y-axis). The blue line indicates the number of covariates for which the $p$-values is greater than 0.01. The plot shows good balance for $N \leq 7$, beyond which the matched series become increasingly different from the treated series in terms of covariate means. The green dotted line indicates KL divergence of the propensity scores. KLD closer to zero indicate greater similarity between two distributions; we see that the balance across the two groups are stable across $N$.

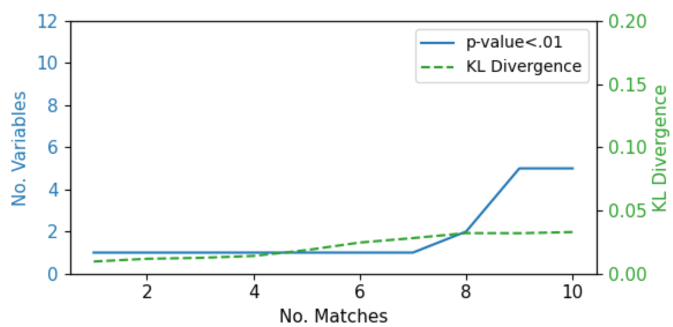


Figure 10: Assessment of match quality by the number of matches

Matching a treated unit to a single or multiple control units are both widely used in practice, each with its tradeoffs. First is the precision-bias tradeoff. One-to-one matching tends to produce less biased estimates because each treated unit is matched with its closest match, but it may yield fewer matches, resulting in less precision. On the other hand, 1:N matching could increase precision by utilizing more data (i.e., more comparisons), but may introduce bias if the matched units are not as similar to the treated unit. Since we set the difference in propensity scores of matched units to be less than a caliper $C = 0.1$, potential bias issues are limited. In terms of sample size, 1:N matching can potentially utilize all available control units, thereby maintaining a larger sample size and consequently reducing variance. At the same time, 1:N matching runs a risk of overfitting if one uses too many matches for each treated unit. In order to secure a large enough sample while against potential overfitting, we proceed with $N = 5$, beyond which some treated series start to no longer have eligible matched series within the caliper. We conduct robustness checks to show that the results of the analysis remain unchanged for $N \leq 5$.

Table 4 evaluates the balance of the covariates and propensity score distribution before and after matching at $N = 5$ (Caliendo and Kopeinig, 2008; Haviland et al., 2007). The results show that the treated series and their matched control series are not significantly different in key variables, with the exception of wait-times. We note that while it is challenging to achieve perfect balance in wait-times given the concentration of series with 1-hour wait-times, the matching process markedly reduces the gap compared to before matching. Figure 11 is a density plot of propensity scores before and after matching. Before matching, we see a greater density of control units with low probability of treatment as expected. After matching, treated and control groups are indistinguishable in terms of their treatment propensities, indicating a strong match.

| | Treated | Control (Before Matching) | | Control (After Matching) | |
|---|---|---|---|---|---|
| | Mean | Mean | p-value | Mean | p-value |
| Propensity Score | 0.697 | 0.307 | 0.000 | 0.696 | 0.918 |
| log(T1 waited + 1) | 5.304 | 4.989 | 0.231 | 5.075 | 0.123 |
| log(T2 waited + 1) | 5.263 | 5.003 | 0.577 | 5.043 | 0.150 |
| log(T3 waited + 1) | 5.252 | 5.000 | 0.605 | 5.038 | 0.162 |
| log(T4 waited + 1) | 5.217 | 4.987 | 0.702 | 4.959 | 0.093 |
| log(T1 purchased + 1) | 4.595 | 4.006 | 0.000 | 4.499 | 0.524 |
| log(T2 purchased + 1) | 4.654 | 4.020 | 0.000 | 4.604 | 0.728 |
| log(T3 purchased + 1) | 4.508 | 4.015 | 0.000 | 4.440 | 0.666 |
| log(T4 purchased + 1) | 4.456 | 4.020 | 0.000 | 4.386 | 0.658 |
| log(No. Episodes) | 4.038 | 4.035 | 0.217 | 4.004 | 0.576 |
| log(Weeks since published) | 4.253 | 3.785 | 0.000 | 4.148 | 0.122 |
| Wait-time (Standardized) | 0.954 | -0.012 | 0.000 | 1.547 | 0.000 |

Table 4: Covariate balance across treated and matched control series before and after matching

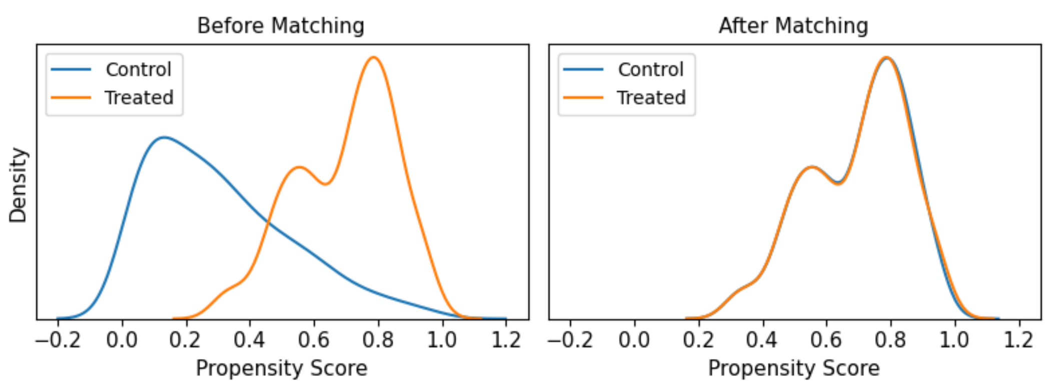


Figure 11: Propensity score distribution before and after matching

## 4.3 Econometric Model

We now turn to the econometric model to estimate the causal effect of the reduction in wait-times. To control for unobservable time-trends, we estimate the treatment effect in a difference-in-difference framework controlling for a set of fixed effects and covariates. To retrieve an unbiased measure of ATE, we must address the issue of variation in treatment timing, which can be problematic for standard TWFE models. When treatment effects vary across both time and units, then the TWFE estimand of ATE may correspond to a non-convex weighted average of individual treatment effects. Hence, we utilize a stacked difference-in-difference model. Stacked DiD has been suggested and widely used in the marketing and economics literature as a way to analyze data from a staggered treatment adoption design (Cengiz et al., 2019; Deshpande and Li, 2019; Baker et al., 2022; Deng et al., 2022).

We start by constructing event-specific datasets of equal length for each of the 191 treated series. The dataset includes the outcome and control variables of the treated series and its six matched control series, which we denote as a *series group* (or cohort) consistent with the notation from Deng et al. (2022). Note that although a non-treated series may appear in multiple series groups, the corresponding data will vary depending on the reduction timing of respective series groups. We then stack these datasets together and estimate a TWFE DiD regression with series group-specific series and period fixed effects, which fully controls for self-selection on unobserved time-invariant factors. Gardner (2022) shows that this approach estimates a convex weighted average of the individual treatment effects under parallel trends and no anticipation.

The stacked DiD model takes the following form:

$$Y_{sgp} = \beta^{DD}(after_p \times treated_s) + X_{sgp}\gamma + \delta_{sg} + \nu_{gp} + \epsilon_{sgp} \tag{4}$$

where $s$ denotes series, $g$ denotes series group, $p$ denotes period (day), and $Y_{sgp}$ denotes the main dependent variable measured for series $s$ of series group $g$ in period $p$. $treated_s$ is a binary treatment indicator for series $s$, and $after_p$ is a post-treatment dummy for period $p$. The main coefficient of interest is $\beta^{DD}$, the average treatment effect of wait-time reduction. $X_{sgp}$ is a matrix of observable control covariates; $\delta_{sg}$ is a fixed effect specific to series $s$ in series group $g$ that captures time-invariant unobservable characteristics (referred to as *Group-Series FE*); $\nu_{gp}$ is a fixed effect specific to group $g$ in period $p$, which captures unobservable time trends such as day-of-week (referred to as *Group-Period FE*). By including group specific fixed effects, the model essentially estimates the DiD from each series group and then applies variance weighting to aggregate the treatment effects (Baker et al., 2022). Finally, $\epsilon_{sgp}$ is the error term, which are clustered at the series level.

# 5 Main Analysis and Results

The main analysis aims to measure the impact of changing wait-times on consumption decisions and monetization of serialized media. Our theoretical framework leverages the unique properties of serialized media: directed complementarities across episodes that diminish over time and consumer willingness-to-pay subject to varying time sensitivity. We posit that under reduced wait-times, the across-episode complementarity value decays over a shorter period, and the platform can better capture consumers with low time sensitivity. These together lead to a higher willingness-to-pay and ultimately increased purchases from existing consumers. Whether the resulting additional purchases sufficiently counteract the negative cannibalization effect is an empirical question. At the same time, we expect to see a greater inflow of new consumers in to the series as the expected value of starting the series becomes higher, serving as new sources of revenue.

The empirical analysis proceeds in four steps. First, we examine the effect of wait-time reduction on consumption and purchases of the *existing consumers*. This enables us to test for the existence of positive across-episode spillover and consumers' varying time sensitivity and empirically measure the impact net of the negative cannibalization effect. Second, to provide further empirical support for our theoretical framework, we also examine the change in the consumption pace of existing consumers in response to the reduction. Third, we examine the effect on the *new consumers* by measuring how many consumers start the series for the first time. Finally, we measure the overall effect of reduction on aggregate consumption and revenues, complemented with a battery of robustness checks.

## 5.1 Existing Consumers

The consumer's decision on whether and how to consume an episode depends on her net utility (valuation for the episode minus cost of money, time, effort, etc.). For serialized media, the complementarity value from having consumed the previous episode represents a significant portion of the valuation. Data patterns in Section 3.2 showed that most users typically read an episode only if they have read the preceding episode, providing supporting evidence. Conversely, if a consumer decides not to consume an episode and thus is unable to benefit from complementarity, she will likely not consume any of the subsequent episodes, including episodes that she would have otherwise purchased. In aggregate, the platform loses out on the sales of not only that episode, but also all following episodes. Reducing wait-times runs the risk of cannibalizing purchases because the increase in the value of free consumption is more pronounced relative to paid consumption, as widely documented in the existing freemium literature (Lee et al., 2019; Li et al., 2019; Cao et al., 2023). Nonetheless, it may serve as a solution for the platform to increase revenues from existing consumers.

Figure 12 is an illustrative example of consumption decisions under different wait-times that may manifest from our theoretical framework. Under a 24-hour wait-time, the consumer purchases Episode 1, but waits for Episodes 2 and 3 because she has higher priorities than consuming another episode at during the day (high time sensitivity). She does not consume Episode 4 because the complementarity value has decayed too much over the 24 hours. Naturally, she also does not consume Episode 5 as she cannot benefit from complementarities. If the wait-time is reduced to 3 hours, she switches to waiting for Episode 1 rather than purchasing (cannibalization effect). Again, she waits for Episode 2 since her time sensitivity is high, but when she is prompted to make a decision 3 hours later, she has the bandwidth to consume another episode *and* the complementarity is higher, inducing her to purchase Episode 3. For the same reason, she switches from no consumption to waiting for Episode 4. This, in turn, allows her to receive complementarity on Episode 5, which she chooses to purchase. From the firm's perspective, it gains greater option value by retaining consumers over a longer period that can potentially generate future cash flow and can price discriminate within consumers across varying time sensitivity.

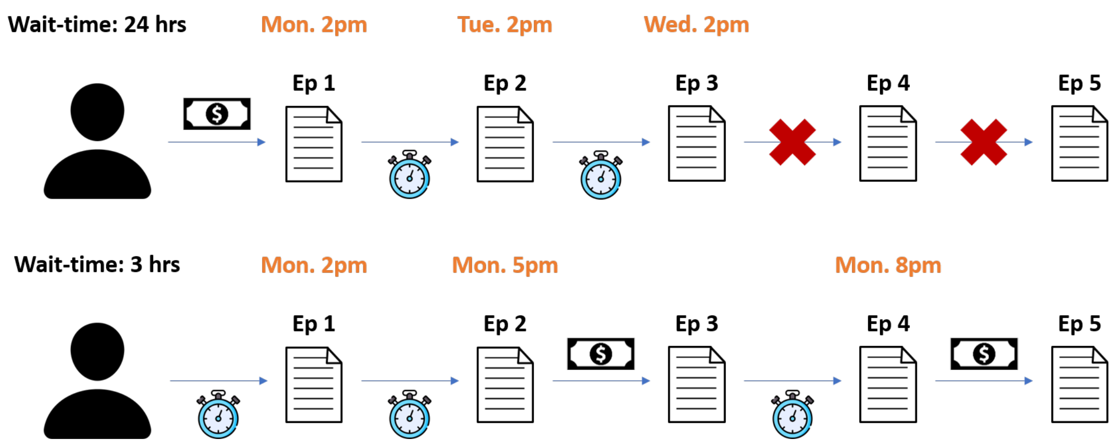


Figure 12: Illustrative example of consumption decisions under varying wait-times

We begin the empirical analysis by sampling consumers that consume episodes both within 15 days before and after treatment for every treated series and its matched control series. The assumption is that given the tight 30-day time frame, any differences in consumption patterns within the existing consumer for the treated series relative to the control series can be attributed to the wait-time reduction. We measure the impact of wait-time reduction on the total number of episodes consumed (i.e., waited and purchased) and separately for episodes purchased by running the stacked DiD regression from Equation 4 with the two-period consumer panel data. The outcome variables are log-transformed to address the skewed distribution of the data. The model controls for observable characteristics that may affect consumption, including series length, days since first and last episode publication and the presence of promotions. In addition to the series group specific series- and period-fixed effects, we also include individual-fixed effects based on pre-treatment spending behavior. Specifically, we categorize the consumers into two types based on whether they have ever purchased Coins on the platform prior to the earliest treatment date. The stacked dataset consists of 13,405 spenders (44%) and 16,742 non-spenders (56%).

Estimation results are presented in column (1) of Table 5. The coefficient on the individual-fixed effect, *non-spender*, is negative and significant, meaning consumers inclined to spend money on the platform consume more episodes. The treatment effect of interest is the coefficient of *after* $\times$ *treated*. We find that the reduction of wait-times lead to a significant increase in episodes consumed, meaning consumers on average progress further in the series. The direction of the result is expected as shorter wait-times encourage consumers to consume episodes who would have otherwise chosen not to. To assess the magnitude of the treatment effect, denote the consumption count before and after the reduction as $N_0$ and $N_1$, respectively. Using the estimated coefficient $\hat{\beta}$, we can compute $N_1 = e^{\hat{\beta}}(N_0 + 1)$ and the percentage change in the dependent variable as $(N_1 - N_0)/N_0$. The estimate 0.289 in column (1) suggests that if the episodes consumed per series per consumer before the reduction is at the mean (9.1), it would increase to 12.5, a 37% increase.

Next, we estimate the impact of wait-time reduction on the total number of episodes purchased made by a consumer for a given series. Increased consumption would only be detrimental to revenues if it came at the expense of less purchases. In the absence of across-episode complementarity and varying time sensitivity that affect the consumer's willingness-to-pay, we would not expect any increase in purchases as the reduced wait-times would only yield cannibalization effect. Results are presented in column (2) of Table 5. The coefficient on the individual-fixed effect, *non-spender*, is again negative and significant, meaning consumers inclined to spend money on the platform purchase more episodes. Note that non-spenders have non-zero purchases because there are alternative ways of obtaining Coins other than spending real money, such as referrals, watching ads or giveaway events hosted by the platform. The estimated coefficient of *after* $\times$ *treated* is positive and significant at 0.073, indicating a 12% increase in paid consumption per consumer. This surprising result shows that by making it easier to consume for free, the platform can actually monetize the same consumer that may switch from purchasing to waiting for an episode to a greater extent.

In Table 6, we explore heterogeneity in treatment effect by allowing the effect to vary based on pre-treatment spending behavior. As seen in Figure 3, there is a segment of consumers that do not spend any money on the platform regardless of the wait-time, and hence, we would not expect to find a strategic shift in their purchase behavior from the wait-time reduction. Moreover, the platform is primarily concerned with the impact on spenders, because their baseline purchases are significantly higher than that of the non-spenders as denoted by the negative coefficient on *non-spender* in Table 5, and any change in purchases of the non-spenders would have a minimal affect on revenues. The coefficient on *after* $\times$ *treated* represent the impact of wait-time reduction on spenders, and the coefficient on *after* $\times$ *treated* $\times$ *non-spender* separates out the

| | (1) log(Consumed+1) | (2) log(Purchased+1) |
|---|---|---|
| non-spender | -0.028*** | -0.179*** |
| | (0.010) | (0.023) |
| after × treated | 0.289*** | 0.073** |
| | (0.031) | (0.029) |
| log(no. episodes) | -0.329 | -0.882*** |
| | (0.341) | (0.336) |
| log(days since first pub.) | -0.982*** | 0.302** |
| | (0.173) | (0.126) |
| log(days since last pub.) | -0.074*** | -0.006 |
| | (0.023) | (0.012) |
| **1**(promotion) | -0.044 | 0.074** |
| | (0.038) | (0.034) |
| Group-Series FE | Y | Y |
| Group-Period FE | Y | Y |
| Individual FE | Y | Y |
| N Obs | 405634 | 405634 |
| N Series Groups | 191 | 191 |
| R-squared Adj. | 0.004 | 0.009 |

Note: Robust standard errors clustered at series level in parentheses.
Individual fixed effect is an indicator for historical spending on the platform.
$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table 5: Treatment effect on consumption and purchases per series per consumer

incremental effect on the non-spenders. While we find no heterogeneity in the effect on total consumption, the effect on episodes purchased is magnified to 0.107, representing a 17% increase relative to the pre-treatment levels. Meanwhile, the effect on non-spenders is positive but no longer significant, meaning their purchase behavior does not significantly change, consistent with our expectations. In other words, while both segments of consumers consume more episodes, the increase consists of increased waited episodes for non-spenders, whereas the those inclined to spend money on the platform also purchase more episodes.

For robustness, we conduct the above analyses analysis using an extended window of 30 days before and after the reduction, and our results remain qualitatively unchanged. In addition, we estimate the effect using an alternative model specification using the magnitude of wait-time reduction rather than a treatment indicator. That is, we replace *treated* with *magnitude*, computed as the difference in wait-times pre- and post-treatment (log-transformed). Under our theoretical framework, the increase in complementarity value and the platform's ability to take advantage of varying time sensitivity, and hence the treatment effect, should be proportional to the magnitude of the wait-time reduction. Under this specification, the estimated coefficients on *after* × *magnitude* are positive and significant ($p < 0.01$), in line with our expectations. In sum, we find that providing more for free may be a transient, rather than an unequivocal, cannibalization even for the existing consumers.

| | (1) log(Consumed+1) | (2) log(Purchased+1) |
|---|---|---|
| non-spender | -0.027*** | -0.174*** |
| | (0.010) | (0.025) |
| after × treated | 0.295*** | 0.107*** |
| | (0.035) | (0.035) |
| after × treated × non-spender | -0.014 | -0.071** |
| | (0.023) | (0.032) |
| log(no. episodes) | -0.331 | -0.890*** |
| | (0.341) | (0.336) |
| log(days since first pub.) | -0.982*** | 0.302** |
| | (0.173) | (0.126) |
| log(days since last pub.) | -0.074*** | -0.007 |
| | (0.023) | (0.012) |
| **1**(promotion) | -0.044 | 0.073** |
| | (0.039) | (0.034) |
| Group-Series FE | Y | Y |
| Group-Period FE | Y | Y |
| Individual FE | Y | Y |
| N Obs | 405634 | 405634 |
| N Series Groups | 191 | 191 |
| R-squared Adj. | 0.004 | 0.010 |

Note: Robust standard errors clustered at series level in parentheses.
Individual fixed effect is an indicator for historical spending on the platform.
$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table 6: Heterogeneous treatment effect based on historical platform spending

### 5.1.1 Existing Consumers: Excess Wait

Platforms focus on maintaining regular and frequent user engagement with the product as it is a critical driver of customer lifetime value (Fader et al., 2005). The WFF policy encourages such behavior by setting wait-times conditional on the user's consumption timing of the last waited episode of the series. This unique feature prevents irregular binge behavior and ensures that the consumer periodically visits the series to (1) claim the free episode and (2) reset the clock for another free episode by reading. In order to establish frequent engagement, it is important that the consumer accesses her free episode shortly after she becomes eligible.

The unique features of serialized media suggest that reducing wait-times can serve as an effective solution to accelerate the consumption pace of the existing users. First, because the complementarity value diminishes over a shorter interval, the consumer's valuation of waited consumption increases. This endogenously strengthens the consumer's incentive to quickly access the free episode (i.e., higher value of the current episode and higher expected value of subsequent episodes). Second, the platform can target instances when the consumer's time sensitivity is low and thus has the ability to quickly access the free episode. This acceleration (time-shifting) effect prompts consumers to make consumption decisions at a faster pace, leading to greater purchases. Note that in the absence of across-episode complementarity and varying time sensitivity, we would not expect any change in consumption pace beyond the mechanical shift due to changes in wait-time.

To empirically explore this dynamic, we investigate how wait-times affect *excess wait-time*. We define excess wait-time to be the time interval between when an episode becomes eligible to be consumed for free for a given consumer and when she actually consumes it. Low excess wait-time would indicate that the consumer consumed the free episode as soon as the free version became available. As before, we take the set of existing consumers within 15 days before and after the treatment date for each treated and its matched control series, and estimate the stacked DiD regression from Equation 4 with the two-period consumer panel data. The dependent variable is excess wait-time in hours (log-transformed). The model controls for observable characteristics that may affect consumption pace, including series length, relative position of the episode (episode number divided by series length), days since first and last publication and the presence of promotions. Also included are series group specific series- and period-fixed effects, as well as individual-fixed effects based on pre-treatment spending behavior.

Table 7 reports the estimates. The coefficient on *after* $\times$ *treated* in column (1) reports a significant decrease in excess wait-time, confirming our expectations. If the excess wait-time before the reduction is at the mean (3.4 hours), it would decrease to 2.7 hours, a 21% decrease. The heterogeneous treatment effects estimated in column (2) report that the acceleration effect is even greater for non-spenders. The heterogeneity driven by the fact that non-spenders are more likely to wait than purchase compared to the spenders, and hence, they have greater incentives to access the free episodes sooner to reset the wait clock for the next episode. To confirm robustness of our findings, we use a wider window of 30 days before and after the treatment date and the alternative model specification using magnitude of wait-time reduction. In both cases, treatment effect estimates remain qualitatively unchanged ($p < 0.01$).

## 5.2 New Users

Next, we investigate the impact of wait-times on the new consumers by examining the change in the inflow of new consumers in to the series. The expansion of the consumer base from increased value of the free option is well-understood from the freemium literature. In the case of serialized media, a user decides to start a series if the aggregate expected utility from the episodes outweighs the start-up cost. The aggregate expected utility increases under reduced wait-times due to higher complementarity, and more consumers will now choose to start the series, expanding the total size of the pie. Analogous to retailers incentivizing store visits through loss leaders (Hess and Gerstner, 1987), the series lures in a larger traffic of new consumers in the hopes that they will purchase episodes conditional on starting the series.

| | log(Excess Wait-time + 1) | |
|---|---|---|
| | (1) | (2) |
| non-spender | -0.029*** | -0.019* |
| | (0.010) | (0.011) |
| after × treated | -0.181*** | -0.149*** |
| | (0.046) | (0.046) |
| after × treated × non-spender | | -0.061*** |
| | | (0.023) |
| episode position | 0.582 | 0.583 |
| | (0.434) | (0.434) |
| log(no. episodes) | 2.762*** | 2.761*** |
| | (1.043) | (1.042) |
| log(days since first pub.) | 1.474*** | 1.473*** |
| | (0.433) | (0.433) |
| log(days since last pub.) | -0.210*** | -0.209*** |
| | (0.057) | (0.057) |
| **1**(promotion) | -0.072 | -0.073 |
| | (0.082) | (0.082) |
| Group-Series FE | Y | Y |
| Group-Period FE | Y | Y |
| Individual FE | Y | Y |
| N Obs | 3091134 | 3091134 |
| N Series Groups | 187 | 187 |
| R-squared Adj. | 0.018 | 0.018 |

Note: Robust standard errors clustered at series level in parentheses.
$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table 7: Treatment Effect on Excess Wait-time

To measure the impact on new consumer acquisition at the series level, we estimate Equation 4 with a two-period series level panel data during the 15 days before and after the reduction. The dependent variable is the number of new consumers (log-transformed) for a given series. The control variables include series length, days since first and last publication, and the presence of promotions. As the analysis is on a series level, we include series group specific series- and period-fixed effects.

Results in column (1) of Table 8 demonstrate that the wait-time reduction leads to a significant increase in new consumers. At the pre-treatment mean of 96, the wait-time reduction increases the number of new consumers to 123, a 28% increase. Columns (2) and (3) report regression results separately for spenders and non-spenders. The treatment effects are positive and significant for both groups – at the pre-treatment mean, the number of new spenders increase from 47 to 63 (34% increase), and the number of new non-spenders increase from 49 to 60 (23% increase). The results are robust to a wider time window of 30 days before and after treatment and alternative model specification using magnitude of wait-time reduction ($p < 0.01$).

| | log(New Consumers + 1) | | |
|---|---|---|---|
| | (1) All Consumers | (2) Spenders | (3) Non-spenders |
| after × treated | 0.245*** | 0.285*** | 0.210** |
| | (0.082) | (0.084) | (0.092) |
| log(no. episodes) | -0.473 | -0.204 | -0.767 |
| | (0.881) | (0.914) | (0.983) |
| log(days since first pub.) | -0.097 | -0.105 | -0.098 |
| | (0.866) | (0.823) | (0.946) |
| log(days since last pub.) | -0.161 | -0.189 | -0.112 |
| | (0.145) | (0.161) | (0.131) |
| **1**(promotion) | 1.341*** | 1.331*** | 1.333*** |
| | (0.335) | (0.385) | (0.296) |
| Group-Series FE | Y | Y | Y |
| Group-Period FE | Y | Y | Y |
| N Obs | 2292 | 2292 | 2292 |
| N Series Groups | 191 | 191 | 191 |
| R-squared Adj. | 0.122 | 0.127 | 0.092 |

Note: Robust standard errors clustered at series level in parentheses.
$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table 8: Treatment effect on inflow of new consumers

### 5.3 Aggregate Consumption and Revenues

Thus far, we have found empirical evidence of rich consumption dynamics separately for existing and new consumers in connection with changing wait-times, which supports our theoretical framework. Under shorter wait-times, consumers are more likely to make consumption decisions under low time sensitivity and higher complementarity value, and as a result, consume more episodes in a series. Although they face lower incentives to purchase each episode, the potential cannibalization is offset by additional purchases from the incremental consumption. The lower wait-times also accelerate the pace of waited consumption, allowing potential purchase decisions to be made quicker. Because consumers perceive higher a priori value of the series, it prompts them to start consuming the series, leading to an expansion of the consumer base. However, the central question from the perspective of the firm remains: What is the net effect of the wait-time reduction on *aggregate* consumption and platform revenues?

To answer this, we analyze the change in aggregate consumption and purchases at the series-day level over the 15 days before and after the wait-time reduction. We use four dependent variables: the daily number of episodes consumed and purchased for a given series (log-transformed), and both on a per consumer level (i.e., aggregate count divided by the number of daily unique consumers). We control for a host of observable characteristics that may affect aggregate demand, including series length, days since first and last publication, and the presence of promotion as well as its 1-week lag to account for potential carry-over effects. Again, we include series group specific series- and period-fixed effects.

The results of our main analysis presented in Table 9 demonstrate a significant positive effect of the wait-time reduction on daily aggregate consumption and purchases. Column (1) shows the impact on total daily consumption, $\log(Consumed + 1)$. The coefficient of *after* × *treated* in column (1) suggests that if the daily episodes consumed before the reduction is at the mean (174), holding all else equal, it would increase to 312, a 79% increase. Given our earlier findings of increased purchases from existing consumers as well as a greater inflow of new consumers, we expect to find an increase in aggregate daily purchases, despite the risk of the negative cannibalization effect. The estimated treatment effect in Column (2) is indeed positive and significant. It suggests that if the daily purchased episodes before the reduction is at the mean (58), it would increase to 66, a 14% increase. The estimates of the effect on daily consumption and purchases per consumer shown in columns (3) and (4) are also positive and significant, indicating an increase of 1.59 and 0.15 episodes, respectively.

To measure the elasticity of consumption and purchases with respect to wait-times, we re-estimate the model replacing *treated* with *magnitude* – the estimated treatment effect of the log-log specification yields the elasticity. The signs and significance remain unchanged, showing robustness of our results to using the magnitude of treatment rather than a binary indicator. We find that a 1% reduction in wait-times lead to 0.23% ($p < .01$) and 0.05% ($p < .05$) increase in daily aggregate consumption and purchases, respectively.

| | (1)<br>log(Consumed+1) | (2)<br>log(Purchased+1) | (3)<br>Consumed PC | (4)<br>Purchased PC |
|---|---|---|---|---|
| after × treated | 0.584*** | 0.130** | 1.589*** | 0.153*** |
| | (0.048) | (0.053) | (0.069) | (0.041) |
| log(no. episodes) | 0.776** | 0.239 | 0.363 | -0.187 |
| | (0.339) | (0.413) | (0.291) | (0.201) |
| log(days since first pub.) | -0.096 | -0.134 | -1.066** | -0.334** |
| | (0.562) | (0.563) | (0.505) | (0.170) |
| log(days since last pub.) | -0.119*** | -0.172*** | 0.123*** | 0.007 |
| | (0.039) | (0.046) | (0.032) | (0.015) |
| **1**(promotion) | 0.333*** | 0.280** | 0.003 | -0.070 |
| | (0.080) | (0.127) | (0.132) | (0.098) |
| **1**(T7 promotion) | 0.236*** | 0.127 | 0.053 | -0.073 |
| | (0.080) | (0.114) | (0.095) | (0.072) |
| Group-Series FE | Y | Y | Y | Y |
| Group-Period FE | Y | Y | Y | Y |
| N Obs | 34380 | 34380 | 34380 | 34380 |
| N Series Groups | 191 | 191 | 191 | 191 |
| R-squared Adj. | 0.064 | 0.011 | 0.121 | 0.002 |

Note: Robust standard errors clustered at series level in parentheses.
$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table 9: Treatment Effect on Daily Aggregate Consumption and Revenues

## 5.4 Robustness Checks

### 5.4.1 Identifying Assumptions

Causal identification of the DiD estimate holds under the assumptions of parallel trends, no anticipation and the stable unit treatment value assumption (SUTVA). The parallel trends assumption requires that the treatment group would have had an identical trend to the control group had the treatment not been implemented, and the no anticipation assumption requires that in periods prior to treatment, the outcome variable was not affected by the upcoming treatment. If these two assumptions hold, then any time-varying unobservables will be absorbed by the trends in the control group and the model would yield unbiased estimates. In our empirical setting, the platform unilaterally implemented the reduction without notifying the users in advance, which prevents any strategic action from the users such as delaying consumption or purchase. To formally test the parallel trends assumption, we follow the approach commonly used in the literature that exploits the pre-treatment time series in the panel data (Angrist and Krueger, 1999; Bronnenberg et al., 2020). Using only the pre-treatment periods, we run the analysis from Section 5.3 replacing $after \times treated$ with $period \times treated$. As shown in Table 10, the deviation from the common trend for the treatment series is very small and not statistically significant. Thus, we fail to reject the null hypothesis that the trend of the treated series is not significantly different from the control series, providing support of the parallel trends assumption.

| | (1)<br>log(Consumed+1) | (2)<br>log(Purchased+1) | (3)<br>Consumed PR | (4)<br>Purchased PR |
|---|---|---|---|---|
| $period \times treated$ | -0.002 | -0.010 | 0.000 | -0.006 |
| | (0.005) | (0.007) | (0.005) | (0.005) |
| Other control variables | Y | Y | Y | Y |
| Fixed Effects | Y | Y | Y | Y |
| N Obs | 20025 | 20025 | 20025 | 20025 |
| N Series Groups | 191 | 191 | 191 | 191 |
| R-squared Adj. | 0.080 | 0.032 | 0.009 | 0.000 |

Note: Robust standard errors clustered at series level in parentheses.
$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table 10: Parallel Trends

We provide further evidence that parallel trends and no anticipation assumptions are likely to hold by estimating a dynamic specification of Equation 4 using a period indicator $I_p$ that allows the treatment effect to vary by period. Figure 13 plots the estimated treatment effect by period. The estimated treatment effects are not significant prior to treatment, further indicating that the common trends assumption is likely to hold.

$$Y_{sgp} = \sum \beta_p^{DD}(I_p \times treated_s) + X_{sgp}\gamma + \delta_{sg} + \nu_{gp} + \epsilon_{sgp} \tag{5}$$

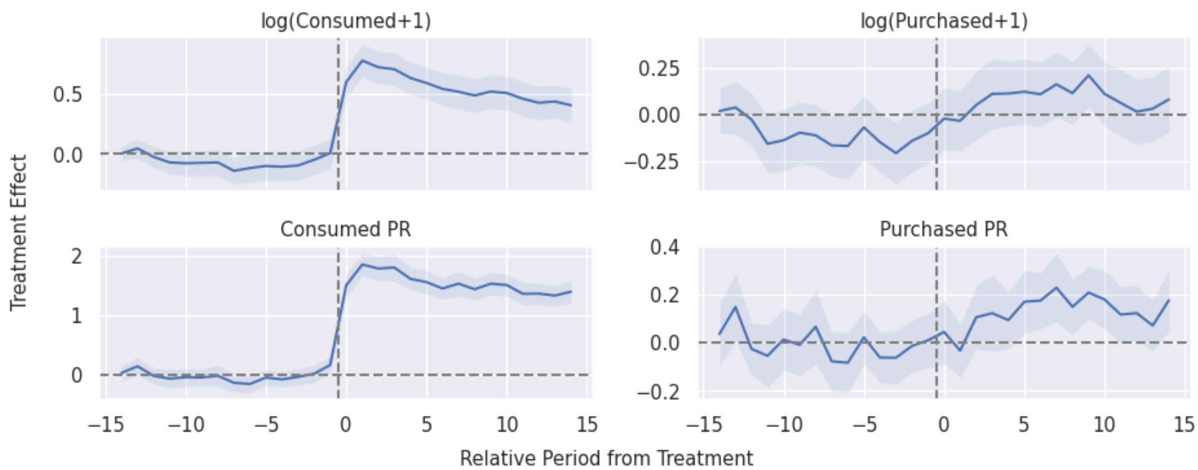


Figure 13: Estimated Treatement Effect by Period

In addition, we conduct a sensitivity test following Cinelli and Hazlett (2020) and Gibson and Zimmerman (2021) to understand how much deviation from parallel trends would be required to undermine the conclusions of our analysis, which assumes the parallel trends assumption is true. Cinelli and Hazlett (2020) reformulates the classical omitted variables framework to develop a sensitivity analysis that provides, relative to an observed covariate benchmark, how strongly unobserved confounders would need to be associated with both the outcome and treatment variables (in terms of partial $R^2$) to explain away the estimated treatment effect. The key advantage of using $R^2$ is that it is scale-free and does not require distributional assumptions of unobserved confounders as well as on the treatment assignment mechanism. As the benchmark observed covariate, we do not rely on a single variable, but rather include all observed covariates from Table 9 to be more conservative. The results presented in Figure 14 show that even if the unobserved confounders are twice as strong as the combined explanatory power of the benchmark covariates, the effects remain consistent.

SUTVA states that the potential outcomes of each unit are not influenced by the treatment assignment or outcomes of other units (i.e., no spillover effect between units). For example, there may be potential substitution where readers move away from non-treated series to spend more time reading stories with reduced wait-times. There also may be complementary effects where readers consume episodes of the treated series more frequently and in doing so also read the untreated series more frequently. In our empirical context, the platform hosts over 10,000 series on the platform, and treatment was implemented at various times across only 191 series. Hence, it is unlikely that a reduction in wait-time for any series had a meaningful spillover effect on other series on the platform.

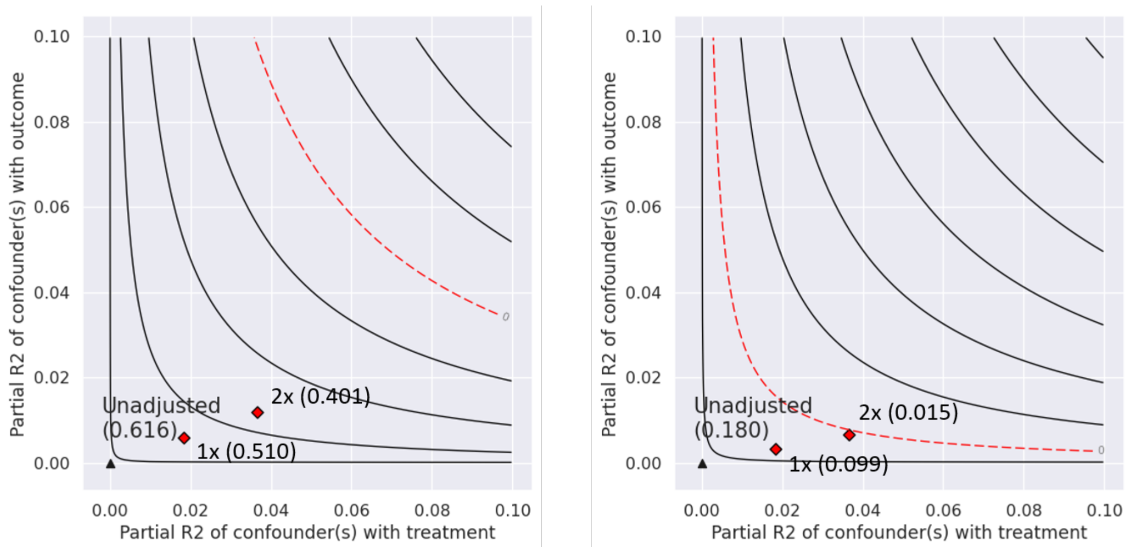


Figure 14: Sensitivity contour plots of estimated treatment effect on daily consumption (left) and purchases (right)

Nonetheless, to mitigate the bias from the potential violation of SUTVA, we run a subsample analysis using control series with minimal overlap in the reader base with the treated series. The idea is that if there are any substitution or complementary spillover effects, the untreated series who share more readers in common with the treated series should be affected more. Specifically, we define $overlap_{sg}$ as the proportion of readers of series $s$ who have also read an episode of the treated series in group $g$ during the 15 days prior to treatment. Then, we conduct the analysis from Section 5.3 on a subsample dropping all control series with an overlap greater than 10%. Alternatively, we estimate the treatment effects by explicitly controlling for potential interference between series, motivated by the approach from Clarke (2017) and Jo et al. (2020). By including an interaction term $after \times overlap$, we run the following regression which ensures that the treatment effect is isolated from any spillover effects:

$$Y_{sgp} = \beta^{DD}(after_p \times treated_s) + \rho(overlap_{sg} \times treated_p) + X_{sgp}\gamma + \delta_{sg} + \nu_{gp} + \epsilon_{sgp} \tag{6}$$

where $\rho$ represents the spillover effect from the treated to the untreated series. The results shown in Table 11 provides empirical support for SUTVA. The treatment effect estimates from the subsample analysis (top-half) and explicitly controlling for influence of reader base overlap (bottom-half) remain similar to the main analysis.

| | (1) log(Consumed+1) | (2) log(Purchased+1) | (3) Consumed PR | (4) Purchased PR |
|---|---|---|---|---|
| $after \times treated$ | 0.617*** | 0.188*** | 1.540*** | 0.130*** |
| (Subsample Analysis) | (0.053) | (0.054) | (0.073) | (0.034) |
| Other control variables | Y | Y | Y | Y |
| Fixed Effects | Y | Y | Y | Y |
| N Obs | 37020 | 37020 | 37020 | 37020 |
| N Series Groups | 188 | 188 | 188 | 188 |
| R-squared Adj. | 0.096 | 0.031 | 0.093 | 0.002 |
| $after \times treated$ | 0.605*** | 0.173*** | 1.537*** | 0.119*** |
| | (0.054) | (0.055) | (0.073) | (0.034) |
| $after \times overlap$ | -0.430 | -0.243 | -0.859** | -0.445* |
| | (0.360) | (0.323) | (0.384) | (0.227) |
| Other control variables | Y | Y | Y | Y |
| Fixed Effects | Y | Y | Y | Y |
| N Obs | 40050 | 40050 | 40050 | 40050 |
| N Series Groups | 191 | 191 | 191 | 191 |
| R-squared Adj. | 0.093 | 0.030 | 0.091 | 0.002 |

Note: Robust standard errors clustered at series level in parentheses.
$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table 11: Check for potential violation of SUTVA using reader base overlap

### 5.4.2 Falsification Tests

We test the possibility that the estimates in Table 9 are coincidentally picking up spurious effects by using pseudo treatment indicators and dates. For pseudo treatment indicators, we randomly assign a control series as treated for each matched series group and estimate the same model without the actual treated series. Under this falsification test, since the pseudo treatment indicator does not reflect the true information of whether the wait-time of the series is reduced, the estimated treatment effects should be insignificant (Ghose and Todri-Adamopoulos, 2016; Jo et al., 2020). For the pseudo treatment dates, we manipulate the treatment date to be 15 days prior to the actual date of reduction. Since the modified time frame does not include the actual treatment date, the estimates should again be insignificant. Table 12 shows that the coefficients on $after \times treated$ are indeed statistically insignificant for both specifications, indicating that our findings are not a statistical artifact of our specification.

### 5.4.3 Alternative Explanations and Model Specifications

Next, we explore whether the estimated treatment effect could arise from the platform strategically timing the wait-time reduction. One alternative explanation for the positive treatment effect is that the platform sends episode gifts in anticipation of or simultaneously with the wait-time reduction to draw reader attention. If this were the case, our estimates might reflect the effect of those marketing activities rather than the sole effect from shorter wait-times. Another explanation is that the timing of the reduction is coordinated with

| | (1) log(Consumed+1) | (2) log(Purchased+1) | (3) Consumed PR | (4) Purchased PR |
|---|---|---|---|---|
| Pseudo Treatment Indicator | | | | |
| $after \times treated$ | -0.026 | 0.018 | -0.055 | -0.018 |
| | (0.037) | (0.037) | (0.036) | (0.023) |
| Other control variables | Y | Y | Y | Y |
| Fixed Effects | Y | Y | Y | Y |
| N Obs | 34320 | 34320 | 34320 | 34320 |
| N Series Groups | 191 | 191 | 191 | 191 |
| R-squared Adj. | 0.070 | 0.030 | 0.010 | 0.001 |
| Pseudo Treatment Date | | | | |
| $after \times treated$ | -0.044 | -0.017 | -0.038 | -0.023 |
| | (0.055) | (0.064) | (0.031) | (0.035) |
| Other control variables | Y | Y | Y | Y |
| Fixed Effects | Y | Y | Y | Y |
| N Obs | 39870 | 39870 | 39870 | 39870 |
| N Series Groups | 191 | 191 | 191 | 191 |
| R-squared Adj. | 0.056 | 0.026 | 0.017 | 0.000 |

Note: Robust standard errors clustered at series level in parentheses.
$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table 12: Results of Falsification Tests using Pseudo Treatment Indicators and Dates

new episode releases, in which case the estimated effect would again be confounded by renewed reader interest from the new episode. Although we control for these factors in our main analysis, we re-estimate the model by entirely removing treated and non-treated series that sent any gifts or had any new episode releases within our 30-day timeframe. Results are presented in Table 13, and for both specifications, the effects remain qualitatively unchanged.

| | (1) log(Consumed+1) | (2) log(Purchased+1) | (3) Consumed PR | (4) Purchased PR |
|---|---|---|---|---|
| $after \times treated$ | 0.597*** | 0.179*** | 1.570*** | 0.158*** |
| (excl. gifted episodes) | (0.054) | (0.057) | (0.077) | (0.036) |
| Other control variables | Y | Y | Y | Y |
| Fixed Effects | Y | Y | Y | Y |
| N Obs | 34080 | 34080 | 34080 | 34080 |
| N Series Groups | 175 | 175 | 175 | 175 |
| R-squared Adj. | 0.044 | 0.014 | 0.086 | 0.002 |
| $after \times treated$ | 0.613*** | 0.173*** | 1.628*** | 0.154*** |
| (excl. new episodes) | (0.060) | (0.063) | (0.082) | (0.041) |
| Other control variables | Y | Y | Y | Y |
| Fixed Effects | Y | Y | Y | Y |
| N Obs | 26850 | 26850 | 26850 | 26850 |
| N Series Groups | 146 | 146 | 146 | 146 |
| R-squared Adj. | 0.076 | 0.018 | 0.097 | 0.001 |

Note: Robust standard errors clustered at series level in parentheses.
$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table 13: Robustness Checks on Marketing Activities and New Episode Releases

We also confirm the robustness of our results using a subsample of the treated series with the same post-reduction wait-time. As show in Table 3, there is variation in wait-times before and after the reduction among the 191 treated series. Our main analysis investigates the effect of the reduction irrespective of the starting and ending points, as our goal is to explore the consumption dynamics that arise due to the unique complementarity properties rather than pinpoint an optimal wait-time. Nevertheless, there may be non-uniform effects based on the absolute length of wait-times that may bias our results. Our findings are robust to a specification using treatment intensity as shown in 5.3, but to further alleviate the concerns, we select a subsample of 152 treated series for which wait-times were reduced to one hour. We confirm in Table 14 that the previous results of the effect on daily consumption and purchases hold for the subsample.

| | (1) log(Consumed+1) | (2) log(Purchased+1) | (3) Consumed PR | (4) Purchased PR |
|---|---|---|---|---|
| $after \times treated$ | 0.688*** | 0.202*** | 1.760*** | 0.145*** |
| | (0.059) | (0.062) | (0.079) | (0.036) |
| Other control variables | Y | Y | Y | Y |
| Fixed Effects | Y | Y | Y | Y |
| N Obs | 31860 | 31860 | 31860 | 31860 |
| N Series Groups | 152 | 152 | 152 | 152 |
| R-squared Adj. | 0.109 | 0.033 | 0.115 | 0.002 |

Note: Robust standard errors clustered at series level in parentheses.
$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table 14: Subsample analysis using treated series with 1-hour wait-time post-reduction

We also estimate the model with alternative dependent variable transformations (Box-Cox), time periods (14 and 60 days around treatment) and the number of matched controls on propensity scores (1-10 matches). The results from these models replicate our findings.

# 6 Conclusion

Digital content platforms today are not only competing among peers of the same medium but also across medium – content produced in text, video and audio forms are all fighting to occupy consumers' free time such that they will be the first in line when consumers open up their smartphones. In such a competitive environment, serialization has become the prevalent mode of publication due to its effectiveness in encouraging consistent and recurring consumption. Moreover, the ability of digital platforms to track individual consumption has allowed them to exploit time as a monetization lever, leading to the adoption of "wait-for-free" (WFF). In this research, we delve into the impact of wait-times on individual consumption decisions and monetization of serialized media. We begin by highlighting the conceptual characteristics unique to a series, namely the defined sequence of consumption and directed complementarities that diminish over the consumption interval. These unique characteristics motivate us to consider the role of intensive margins –

the degree to which a given consumer is monetized – a dimension that has been overlooked in the versioning literature. Changing wait-times affect whether consumers can realize the complementarity as well as the magnitude of its ex ante expected value, which together determine the shift in intensive margins. Moreover, we expect changing wait-times to also affect extenive margins – the consumers' decision to participate in the series.

Using data from a platform serving serialized books, we leverage a natural shift in policy where the platform reduced the wait-times for a set of series. We estimate using a difference-in-difference framework how the reduction impacted user consumption and purchases. We provide evidence that the positive across-episode spillovers from complementarity counteract the negative cannibalization effect. In addition, we find evidence of consumption acceleration and the expansion of extensive margins. The net impact is an 86% and 20% increase in daily aggregate consumption and purchases, respectively. We conduct a battery of robustness checks to support the identification assumptions and rule out any spurious correlations.

By illustrating the multifaceted consumption dynamics of the WFF policy that is gaining popularity, our research equips firms with a comprehensive understanding of strategic levers for policy design. This goes beyond the traditional focus on the acquisition-cannibalization trade-off, emphasizing the importance of evaluating the extent of consumer monetization. We demonstrate that firms may exploit complementarities across episodes to potentially boost long-term consumer spending. Additionally, our research underscores the significance of recognizing consumer heterogeneity. For instance, a segment of price-sensitive consumers may not respond positively to more lenient policies on free consumption. This suggests that platforms could gain more from implementing targeted strategies, tailoring policies to different consumer groups to maximize overall revenues.

Although the present study is one of the first to investigate the novel WFF policy and the economics of serialized media, it is not without its limits. First, our research relies on non-experimental variation in the data. Although we leverage various identification strategies to support causality, a randomized field experiment that manipulates wait-times across consumers for a given series could further strengthen our findings. Second, our analysis demonstrates the causal effect of changing wait-times, but cannot comment on the optimal wait-time. This would require estimating a structural model of consumers' episode consumption, which would also allow one to investigate insightful counterfactual policies such as charging a positive price on waited episodes or targeted wait-times. Estimating a rich state-dependent utility model incorporating inherent individual characteristics, consumption context and episode content to explore optimal policies would be a very interesting avenue. Given the pivotal role of episode complementarities, leveraging recent advancements in text analysis to understand how the episode content such as the strength of cliffhangers, level of suspense and sentiment affects consumption decisions may be a fruitful area of future research. Third, although we provide ample evidence that correlated marketing activities are unlikely the driver of our results, we cannot fully rule out that in some instances firms may still have tried to support the treated series with alternative methods such as curation or recommendation algorithms, which might have had some effect on our estimates. Fourth, we focus on the short-term effect of the wait-time reduction, as identification of a long-term effect is more difficult. It would be interesting to explore the long-term effect of varying wait-times on platform-wide consumption and series that they later consume.

## References


Abadie, A. and Imbens, G. W. (2011). Bias-corrected matching estimators for average treatment effects. *Journal of Business & Economic Statistics*, 29(1):1–11.

Angrist, J. D. and Krueger, A. B. (1999). Empirical strategies in labor economics. In *Handbook of labor economics*, volume 3, pages 1277–1366. Elsevier.

August, T., Dao, D., and Shin, H. (2015). Optimal timing of sequential distribution: The impact of congestion externalities and day-and-date strategies. *Marketing Science*, 34(5):755–774.

Baker, A. C., Larcker, D. F., and Wang, C. C. (2022). How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics*, 144(2):370–395.

Becker, G. S. and Murphy, K. M. (1988). A theory of rational addiction. *Journal of political Economy*, 96(4):675–700.

Borusyak, K. and Jaravel, X. (2018). *Revisiting event study designs*. SSRN.

Bronnenberg, B. J., Dubé, J.-P., and Sanders, R. E. (2020). Consumer misinformation and the brand premium: A private label blind taste test. *Marketing Science*, 39(2):382–406.

Caliendo, M. and Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of economic surveys*, 22(1):31–72.

Calzada, J. and Valletti, T. M. (2012). Intertemporal movie distribution: Versioning when customers can buy both versions. *Marketing Science*, 31(4):649–667.

Cao, J., Chintagunta, P., and Li, S. (2023). From free to paid: Monetizing a non-advertising-based app. *Journal of Marketing Research*, 60(4):707–727.

Cengiz, D., Dube, A., Lindner, A., and Zipperer, B. (2019). The effect of minimum wages on low-wage jobs. *The Quarterly Journal of Economics*, 134(3):1405–1454.

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

Choi, A. A., Rhee, K.-E., Yoon, C., and Oh, W. (2023). The cost of free: The effects of "wait-for-free" pricing schemes on the monetization of serialized digital content. *MIS Quarterly*, 47(3).

Choi, J., Chae, I., and Feinberg, F. M. (2022). Wait for free: A consumption-decelerating promotion for serialized digital media. *Available at SSRN: https://ssrn.com/abstract=4098036*.

Cinelli, C. and Hazlett, C. (2020). Making sense of sensitivity: Extending omitted variable bias. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82(1):39–67.

Clarke, D. (2017). Estimating difference-in-differences in the presence of spillovers. *Working Paper*.

Datta, H., Knox, G., and Bronnenberg, B. J. (2018). Changing their tune: How consumers' adoption of online streaming affects music consumption and discovery. *Marketing Science*, 37(1):5–21.

de Chaisemartin, C. and D'Haultfœuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–96.

Dehejia, R. H. and Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and statistics*, 84(1):151–161.

Deneckere, R. J. and McAfee, P. R. (1996). Damaged goods. *Journal of Economics & Management Strategy*, 5(2):149–174.

Deng, Y., Lambrecht, A., and Liu, Y. (2022). Spillover effects and freemium strategy in the mobile app market. *Management Science*.

Deshpande, M. and Li, Y. (2019). Who is screened out? application costs and the targeting of disability programs. *American Economic Journal: Economic Policy*, 11(4):213–248.

Diamond, A. and Sekhon, J. S. (2013). Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies. *Review of Economics and Statistics*, 95(3):932–945.

Eco, U. (1990). *The Limits of Interpretation*. Indiana University Press.

Fader, P. S., Hardie, B. G., and Lee, K. L. (2005). Rfm and clv: Using iso-value curves for customer base analysis. *Journal of marketing research*, 42(4):415–430.

Gardner, J. (2022). Two-stage differences in differences. *arXiv preprint arXiv:2207.05943*.

Ghose, A. and Todri-Adamopoulos, V. (2016). Toward a digital attribution model. *MIS quarterly*, 40(4):889–910.

Gibson, L. and Zimmerman, F. (2021). Measuring the sensitivity of difference-in-difference estimates to the parallel trends assumption. *Research Methods in Medicine & Health Sciences*, 2(4):148–156.

Godinho de Matos, M. and Ferreira, P. (2020). The effect of binge-watching on the subscription of video on demand: Results from randomized experiments. *Information Systems Research*, 31(4):1337–1360.

Godinho de Matos, M., Mamadehussene, S., and Ferreira, P. (2023). When less is more: Content strategies for subscription video on demand. *Available at SSRN 4352446*.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277.

Gu, X., Kannan, P., and Ma, L. (2018). Selling the premium in freemium. *Journal of Marketing*, 82(6):10–27.

Hansen, B. B. (2004). Full matching in an observational study of coaching for the sat. *Journal of the American Statistical Association*, 99(467):609–618.

Haviland, A., Nagin, D. S., and Rosenbaum, P. R. (2007). Combining propensity score matching and group-based trajectory analysis in an observational study. *Psychological methods*, 12(3):247.

Heather, N. and Vuchinich, R. E. (2003). *Choice, behavioural economics and addiction*. Elsevier.

Heckman, J. J., Ichimura, H., Smith, J. A., and Todd, P. E. (1998). Characterizing selection bias using experimental data. *Econometrica*, 66(5):1017–1098.

Hess, J. D. and Gerstner, E. (1987). Loss leader pricing and rain check policy. *Marketing Science*, 6(4):358–374.

Imai, K., Kim, I. S., and Wang, E. H. (2021). Matching methods for causal inference with time-series cross-sectional data. *American Journal of Political Science*.

Jo, W., Sunder, S., Choi, J., and Trivedi, M. (2020). Protecting consumers from themselves: Assessing consequences of usage restriction laws on online game usage and spending. *Marketing Science*, 39(1):117–133.

Kamada, Y. and Öry, A. (2020). Contracting with word-of-mouth management. *Management Science*, 66(11):5094–5107.

Kermode, F. (2000). *The sense of an ending: Studies in the theory of fiction with a new epilogue*. Oxford University Press.

Lambrecht, A. and Misra, K. (2017). Fee or free: When should firms charge for online content? *Management Science*, 63(4):1150–1165.

Lee, C., Kumar, V., and Gupta, S. (2019). Designing freemium: Strategic balancing growth and monetization. *Available at Github: https://vineetkumars.github.io/Papers/DesigningFreemiumS2019.pdf*.

Li, H., Jain, S., and Kannan, P. (2019). Optimal design of free samples for digital products and services. *Journal of Marketing Research*, 56(3):419–438.

Linkis, S. T. (2021). *Serialization in Literature Across Media and Markets*. Routledge.

Lu, J., Bradlow, E., and Hutchinson, J. (2019). Multiple dimensions of binging: The hidden costs and benefits. *Available at SSRN: https://ssrn.com/abstract=3493759*.

Lu, J., Karmarkar, U. R., and Venkatraman, V. (2023). Planning-to-binge: Time allocation for future media consumption. *Journal of Experimental Psychology: Applied*.

Luan, J. Y. and Sudhir, K. (2022). Optimal inter-release timing for sequentially released products. *Customer Needs and Solutions*, 9(1-2):25–46.

Mittell, J. (2006). Narrative complexity in contemporary american television. *The velvet light trap*, 58(1):29–40.

Mussa, M. and Rosen, S. (1978). Monopoly and product quality. *Journal of Economic theory*, 18(2):301–317.

Narang, U. and Shankar, V. (2019). Mobile app introduction and online and offline purchases and product returns. *Marketing Science*, 38(5):756–772.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

Schweidel, D. A. and Moe, W. W. (2016). Binge watching and advertising. *Journal of Marketing*, 80(5):1–19.

Shapiro, C. and Varian, H. R. (1998). Versioning: the smart way to. *Harvard business review*, 107(6):107.

Shi, Z., Zhang, K., and Srinivasan, K. (2019). Freemium as an optimal strategy for market dominant firms. *Marketing Science*, 38(1):150–169.

Varian, H. R. (2000). Versioning information goods. *Internet publishing and beyond: The economics of digital information and intellectual property*, pages 190–2002.

Zhang, S., Chan, T. Y., Luo, X., and Wang, X. (2022). Time-inconsistent preferences and strategic self-control in digital content consumption. *Marketing Science*, 41(3):616–636.

Zhao, C., Mehta, N., and Shi, M. (2022). The consumption of serial media products and optimal release strategy. *Working Paper*.