The Impact of Banning Online Gambling Livestreams: Evidence from Twitch.tv*

Qifan Han† Jasmine Yang‡ Andrey Simonov§

July 30, 2024

CLICK HERE FOR LATEST VERSION

Abstract

The necessity of content regulation on digital platforms, particularly concerning misinformation and harmful content, has sparked a growing debate. While many platforms have increasingly relied on self-regulation to address these issues, the effectiveness of such measures remains uncertain, as platforms may prioritize profits over consumer protection, potentially leading to misaligned incentives with regulators. We investigate the effectiveness and market outcomes of content self-regulation by studying Twitch's ban on online gambling livestreams in October 2022, using a novel high-frequency panel dataset covering the top 6,000 Twitch streamers. To identify banned content and streamers affected by the policy, we leverage video analysis on historical video clips, high-frequency stream titles, and in-stream chat analysis. To tackle key identification challenges, we use three causal estimators: two-way fixed effects DiD, Synthetic DiD, and the doubly-robust estimator of group-time average treatment effects, and propose a network analysis to construct valid treated and control groups. On the supply side, we find that the policy caused a reduction in weekly gambling streams by 63.2% for streamers whose content was banned and 12.2% for streamers whose content was not banned. However, the policy also decreased non-gambling streams as an unintended cost for the platform, resulting in an overall reduction in content production and diversity. Additionally, the more popular streamers experienced a higher content reduction, driven by two underlying mechanisms: lower reliance on gambling content and concerns for reputation. On the demand side, we find that the policy only reduced total viewership and low-tier subscriptions, with revenue from loyal viewers unaffected. We discuss the implications of Twitch's policy ban and the broader practices of content self-regulation on platforms in general.

Keywords: content regulation, self-regulation, online gambling, live streaming, unstructured data, causal inference

*All opinions represent our own and do not reflect those of Twitch. All remaining errors are our own.

†Boston University. Email: qfhan94@bu.edu
‡Columbia University. Email: yyang24@gsb.columbia.edu
§Columbia University and CEPR. Email: as5443@gsb.columbia.edu

1 Introduction

The increasing prevalence of misinformation (e.g. Beck et al., 2023; Gandhi and Hollenbeck, 2023; Ananthakrishnan et al., 2020) and harmful content (e.g. Beknazar-Yuzbashev et al., 2022; Aridor et al., 2024) on digital platforms has sparked a significant debate on the necessity of content regulation. As technology advancements present a broader set of regulatory challenges for governments, content regulation has increasingly relied on self-regulation by platforms. For example, Facebook exploits automatic removal of comments that are classified as toxic content.[1] However, the success of self-regulation is often limited by the lack of alignment between the incentives of the platform and those of the regulator (Cusumano et al., 2021). For self-regulation to be effective, it must ensure that the platform's economic interests and user base are not compromised, requiring both content producers and consumers to maintain engagement and satisfaction while remaining active on the platform.

In this paper, we investigate the effectiveness and economic consequences of content self-regulation on Twitch (a major game streaming platform globally), focusing on the recent surge in streams of gambling content (e.g., slots, virtual casino), which has seen a 132% increase since the first half of 2020 (StreamScheme, 2023; StreamHatchet, 2022b). Twitch has witnessed several high-profile scandals involving popular streamers borrowing large sums of money and defrauding followers to support gambling pursuits, along with instances of viewers falling victim to scam roulette games.[2] In response to these growing concerns, on October 18th, 2022, Twitch banned streamers from broadcasting gaming sessions from online gambling websites that include slots, roulette or dice games and lack official licenses in the U.S. or other jurisdictions with sufficient consumer protection (GameRant, 2022). This ban only targeted 4 major websites - Stake, Rollbit, Duelbits, and Roobet, but also potentially affected numerous smaller gambling sites with concerns of subsequent platform actions. Twitch's banning policy provides a lens to understand the effectiveness of content self-regulation, highlighting the challenges posed by misaligned incentives among platforms, streamers, and policymakers in balancing economic interests, user engagement, and consumer protection.

We examine how content producers and consumers react to Twitch's ban on livestreams of gambling content, and whether the policy ban had any unwanted impact on the platform. On one hand, content producers might substitute their banned livestreams with other gambling or

[1]Source: https://www.nytimes.com/2021/08/31/technology/facebook-accenture-content-moderation.html.

[2]Source: https://www.wired.com/story/twitch-streamers-crypto-gambling-boom/.

gambling-like content (e.g., gambling-like elements in video games), and viewers might continue watching this new content. In this case, the ban incurs minimal costs to the platform but risks not being wholly effective, merely shifting it to another form. On the other hand, content producers might significantly decrease their production or switch to other streaming services with more lenient gambling rules. Consumers may subsequently leave the platform, either following their favored streamers or due to a loss of content variety or unmet content needs. If this occurs, the misalignment between the policy's target and its actual economic consequences makes the self-regulation costly to the platform. Understanding the policy impacts and these broader implications helps platforms and policymakers evaluate the effectiveness and costs of such regulatory measures.

We assemble a novel high-frequency streaming dataset covering the top 6,000 livestreamers from August 2022 to December 2022 to study the impact of Twitch's policy ban. Alongside stream activity data, we also collect extensive game-level data, including the presence of gambling-like features in over 5,000 video games streamed during our study period. To tackle the key challenges of detecting banned versus unbanned content within online gambling livestreams and identifying streamers who have streamed banned content, we leverage video analysis on historical video clips, as well as text analysis on stream titles and in-stream chat logs. We model the policy impact using three causal estimators: a two-way fixed effect difference-in-differences (TWFE-DiD) estimator, a synthetic DiD estimator (Arkhangelsky et al., 2021) and a doubly-robust estimator of group-time average treatment effect (Callaway and Sant'Anna, 2021) to address identification challenges. Additionally, we incorporate network analysis in our demand-side estimation to construct valid treated and untreated groups, mitigating concerns over potential violations of the Stable Unit Treatment Values Assumption (SUTVA).

All our empirical approaches yield similar results and insights. First, we find that the banning policy led to a significant decrease of approximately 63.2% in weekly streams of gambling content among streamers who have streamed banned content ("banned streamers" hereafter), as well as a 12% reduction in weekly gambling streams among streamers who have only streamed unbanned gambling content ("unbanned streamers" hereafter). These estimates confirm the policy's effectiveness across the platform. In examining potential substitution patterns, we find that streamers did not actively switch to broadcast video games with gambling-like features, suggesting that online gambling and gambling-like elements in video games might not be as interchangeable as some policymakers have presumed. However, the policy inadvertently affected the production of non-gambling content. Weekly output of video games without gambling-like features dropped by 13.8% among banned gambling streamers and 12.2% among unbanned streamers, leading to an overall reduction of 44.3% and 17.6% in total livestreams, respectively. Our findings highlight both the intended and unintended effects of the policy on content production within the platform.

We further examine our main results on the supply side over different time periods and among various types of streamers. We find that the negative effects on both gambling livestreams and overall livestream volume persisted after the policy implementation. Although the policy specifically targeted four English-based websites, our analysis reveals a spillover effect on gambling streams in other popular languages, such as Spanish and Portuguese. More notably, we find that the policy had a more significant impact on streamers with higher popularity. We test two underlying mechanism to explain this observation: streamers' concerns about their reputation and their reliance on gambling livestreams. Both factors help explain the heterogeneous effects of the banning policy.

We then turn to the policy impact on demand-side variables. First, we find that both banned and unbanned streamers suffered from reductions in total hours watched by viewers, whereas the magnitude of reduction in content consumption was even higher compared to the reduction in content creation. Furthermore, our analysis of subscription levels reveal that only the lowest tier (cheapest) subscriptions decreased after the policy implementation for affected streamers. This finding indicates that while affected streamers saw a drop in revenue from casual viewers, they did not experience significant losses among their loyal viewers or in engagement from their core communities. Finally, we examine website traffic data and find that although the policy had significant effects on both the supply- and the demand-side outcomes of the livestreaming platform, the online gambling websites targeted by the policy were not affected in terms of traffic.

To the best of our knowledge, our research is the first to rigorously quantify the causal effects of a banning policy targeting online gambling on streaming platforms and to explore the decision-making mechanisms of streamers. While previous research has investigated the influence of gambling-like content in video games (Amano and Simonov, 2023), there remains a gap in understanding whether content producers and consumers differentiate this content from genuine online gambling. Our findings provide insights into the dynamics of substitution between online gambling and video games with gambling-like content. These insights are valuable to developers in both online gambling and video game industries, providing guidance on their product design and competition strategies, as well as pricing strategies. Moreover, our findings are valuable to both platforms and regulators, enabling them to assess the effectiveness of content self-regulation and to refine policies aimed at preventing gambling addictions among minors.

2 Related Literature

This paper contributes to several strands of existing literature. First, we contribute to the literature on content self-regulation on digital platforms, which typically focuses on misinformation (Beck et al., 2023; Gandhi and Hollenbeck, 2023; Ananthakrishnan et al., 2020) and harmful content, such as toxic posts (Beknazar-Yuzbashev et al., 2022) and hate speech (Howard, 2019). Our study extends this literature by examining the regulation of a less studied, controversial content — gambling livestreams. We demonstrate that self-regulation can lead to an unwanted negative spillover effect on the production of unregulated content, highlighting a potential misalignment between regulatory policy targets and the platform's economic interests.

Second, we add to the literature of the impact of banning policies on content creation and consumption. For example, Johnson et al. (2024) investigate the impact of prohibiting personalized advertising on child-directed content on Youtube and find that the policy led to a reduction in content production, quality and overall demand for child-directed content creators. Kircher and Foerderer (2024) study the impact of banning targeted advertising in children's gaming apps and find that the policy caused substantial app abandonment and a reduction in feature updates. Our study contributes to this literature by empirically examining the impact of a direct ban on content itself, demonstrating its influence on both the supply-side content production and various revenue channels on the demand side. Additionally, our findings indicate that the policy's effects are not uniform across the supply or demand sides. We explore how these non-uniform effects can be attributed to the heterogeneity among content producers and consumers.

Third, our paper is also part of a rapidly-growing literature on the impact of livestreaming as a marketing channel. Zhang et al. (2023) adopting livestreams significantly increases the revenue of the online store channel for the same seller, with a more pronounced effect for small-scale sellers. Huang and Morozov (2023) investigate the immediate impact of livestreaming on the popularity of promoted products and find that streams significantly boost the concurrent number of players on a leading video game platform, especially for lesser-known games. In contrast, we examine the effect of reducing content production of gambling livestreams on gambling websites and find that the banning policy does not have a negative promotional effect on engagement in online gambling platform. However, we also discover that livestreaming has a disproportionate effect on both content creation and consumption, due to the rich heterogeneity among content, content providers and consumers on the livestreaming platform.

The remainder of the paper is structured as follows. Section 6 introduces a conceptual framework to formalize our empirical targets. Section 4 describes the data sources we use to detect banned content and streamers, and to estimate the empirical targets. Section 5 describes our detection procedure. Section 6 discusses the main challenges over identifying the causal effect and our empirical approaches. Section 7 presents the main results on supply-side outcome variables, including time-varying effects and heterogeneous treatment effects among subgroup of streamers. Section 8 presents the main results on demand-side outcome variables. Section 9 describes the policy impact on online gambling websites. Section 10 concludes.

3 Conceptual Framework

In this section, we introduce a conceptual framework to illustrate the differing incentives among the platform, regulator, and streamers, and to highlight the main empirical targets of this paper.

We consider a livestreaming platform populated by a continuum of streamers, each denoted by i. Each streamer makes a streaming plan aimed at maximizing her revenue, π_i, by choosing a specific streaming hour q_{ki} for each type of content $k \in K$. Each k can represent either a distinct type of streaming content (e.g. chatting, a specific game title) or a group of content sharing similar characteristics (e.g. all role-playing games). The overall steaming plan of streamer i is represented by the vector $\boldsymbol{q}_i = (q_{1i}, q_{2i}, \ldots, q_{Ki})'$ and her revenue is a function of all streamers' streaming plans, denoted as $\pi_i(\boldsymbol{q}_i, \boldsymbol{q}_{-i})$. Consequently, the aggregated supply of each type of content k across the platform is the sum of streaming hours of all streamers on the platform, which can be expressed as

$$Q_k = \int q_{ki} dF_i$$

where F_i denotes the distribution of streamer i across the platform.

The primary goal of the regulator is to reduce the total supply of gambling content on the platform by banning some gambling content in the livestreams. Without loss of generality, we denote the supply of banned gambling content as Q_1, and any remaining unbanned gambling content as Q_2. In addition, we denote the total supply of gambling content as $Q_{\text{gambling}} = Q_1 + Q_2$. When a banning policy is implemented on the platform, some streamers are affected by the policy and must revise their streaming plans. We denote the post-policy plan of streamer

i for content type k as $\tilde{q}_{ki}$ and the total supply of content k as $\widetilde{Q}_k$. The banning policy explicitly imposes the restriction that $\tilde{q}_{1i} \equiv 0$ for all streamer i, and consequently, $\widetilde{Q}_1 = 0$. However, not all gambling content faces ban on the platform, allowing streamers to continue streaming unbanned gambling content and weakening the impact of the regulation. Therefore, the regulator assesses the effectiveness of the banning policy by monitoring the changes in total supply of gambling content on the platform:

$$\Delta\%Q_{\text{gambling}} = \frac{\widetilde{Q}_2 - Q_{\text{gambling}}}{Q_{\text{gambling}}} = \frac{\widetilde{Q}_2 - (Q_1 + Q_2)}{Q_1 + Q_2} \tag{1}$$

In contrast, the platform focuses on the changes in content production and consumption after the policy implementation. On the supply side, content production is measured by the changes in each type of content k,

$$\Delta\%Q_k = \frac{\widetilde{Q}_k - Q_k}{Q_k} \tag{2}$$

On the demand side, the platform's objective is defined by the total revenue generated by all its steamers based on their current streaming plans,

$$\pi(\boldsymbol{Q}) = \int \pi_i(\boldsymbol{q}_i, \boldsymbol{q}_{-i})dF_i$$

since the revenue is shared between the platform and the streamers. Therefore, the policy impact on content consumption can be specified as

$$\Delta\%\pi = \frac{\pi(\widetilde{\boldsymbol{Q}}) - \pi(\boldsymbol{Q})}{\pi(\boldsymbol{Q})} \tag{3}$$

where $\pi(\cdot)$ can be measured using various demand-side outcome variables, such as total hours watched by viewers, in-stream bits and live donations, gifts, or subscriptions. In this paper, we specifically focus on total hours watched by viewers and subscription revenue as the objective functions of the platform.

The stylized model illustrates the potential misalignment among the incentives of the platform, the regulator and the content providers (streamers). While the effectiveness of the content regulation is measured by Equation (1), the platform may primarily be concerned with the outcomes described in (2) and (3). In scenarios of content self-regulation, the platform might inadvertently overlook some negative economic consequences measure by these outcomes. Moreover, both targets are influenced by how streamers react to the banning policy, as streamers always prioritize maximizing their own revenue. Therefore, our empirical target is to provide accurate causal estimates that relate to the outcomes of interest as outlined in the above equations. To elaborate on the policy's impact, we also explore heterogeneous treatment effects by examining the relationship between these outcome variables and the distribution of streamers F_i in Section 7.

We consider a livestreaming platform populated by a continuum of streamers, each denoted by i. Each streamer chooses her streaming plan $\boldsymbol{q}_i^* = (q_{1i}^*, q_{2i}^*, \ldots, q_{Ki}^*)'$, whereas q_{ki}^* denotes her optimal streaming hour for each type of content $k \in K$, to maximize her revenue. Each k can represent either a distinct type of streaming content (e.g. chatting, a specific game title) or a group of content sharing similar characteristics (e.g. all role-playing games). Therefore, streamer i's streaming plan is given by solving the following optimization problem

$$\boldsymbol{q}_i^* = \max_{\boldsymbol{q}_i} \pi_i(\boldsymbol{q}_i, \boldsymbol{q}_{-i}) \Rightarrow \frac{\partial \pi_i(\boldsymbol{q}_i, \boldsymbol{q}_{-i})}{\partial q_{ki}} = 0, \ \forall \ k$$

In the market equilibrium, the aggregated supply of each type of content k across the platform is the sum of streaming hours of all streamers, which can be expressed as

$$Q_k^* = \int q_{ki}^* dF_i$$

where F_i denotes the distribution of streamer i across the platform.

The primary goal of the regulator is to reduce the total supply of gambling content on the platform by banning some gambling content in the livestreams. Without loss of generality, we denote the supply of banned gambling content as Q_1, and any remaining unbanned gambling content as Q_2. In addition, we denote the total supply of gambling content as $Q_{\text{gambling}} = Q_1 + Q_2$. When a banning policy is implemented on the platform, some streamers are affected by the policy and must revise their streaming plans. We denote the post-policy plan of streamer i for content type k as $\tilde{q}_{ki}^*$ and the equilibrium supply of content k as $\widetilde{Q}_k^*$. The banning policy explicitly imposes the restriction that $\tilde{q}_{1i}^* \equiv 0$ for all streamer i, and consequently, $\widetilde{Q}_1^* = 0$. However, not all gambling content faces ban on the platform, allowing streamers to continue streaming unbanned gambling content, potentially even increasing the total supply of gambling livestreams on the platform after the policy implementation. Therefore, the regulator assesses the effectiveness of the banning policy by monitoring the changes in total supply of gambling content on the platform:

4 Data

In this section, we describe several novel data sources which we compile to study the effect of the banning policy and to identify banned gambling content and streamers.

Video streaming data on Twitch. We created a novel high-frequency dataset of Twitch streaming activities for the top 6,000 live streamers over five months, from August 1st, 2022, to December 31st, 2022. We compiled the dataset from two sources. First, we obtained the top 6,000 streamers and their stream IDs from *sullygnome*. Based on the stream IDs, we obtained data on streamed content, viewer count and follower count, recorded every 10 minutes, from *twitchtracker* for the period from August 1st, 2022 to October 26th, 2022. Additionally, we obtained all titles used in a stream and the timestamps of each title change. These data allow us to restore each streamer's status (online or offline), the start time and end time of the stream and all activities within a stream. Second, for the period starting from October 26th, 2022, we utilized the high-frequency streaming data on the same streamers collected by Yang and Simonov (2024), through sending requests to the Twitch API every 15 minutes (See Appendix A for details).[3] The combined dataset offers us a comprehensive view of the streaming activities before and after the policy implementation.

For our main analyses, we aggregate streaming data to the weekly level for each streamer, recording streaming metrics including total streaming hours, total hours watched by viewers, weekly average viewership, and total streaming hours of each type of content within a week, including online gambling, video games, and non-video game content.

We classify streamers based on how much online gambling content they have streamed before the announcement of the banning policy on Sep 20th, 2022. Specifically, we assume that streamers who almost never streamed online gambling content are unlikely to be affected by the policy and should behave similarly to non-gambling streamers. Therefore, we set a threshold such that any streamers who streamed less than 7.3 hours in total before the policy implementation are removed from treated groups and added into the untreated group. The threshold is selected as the first quartile of the distribution of total gambling hours before the policy announcement. We have also examined our results based on alternative treatment definitions (e.g., streamers are treated if they have streamed gambling content at least once before policy announcement). our empirical findings are qualitatively the same across different

[3]We compile the first dataset from *sullygnome* and *twitchtracker*, since the second dataset does not cover data prior to the policy implementation. To ensure the quality of the first data source, we checked a few days of overlapping data from both sources and found them consistent, except that the first dataset occasionally misses stream title information and merges streams that live for a short duration into one stream.

definitions, with minor differences in the magnitude of the effects.

To identify the causal effects of the policy, we exclude streamers who only broadcasted either before or after the policy announcement. Additionally, we classify streamers into three groups: banned, unbanned and untreated. We define *banned streamers* as those who had streamed content from banned websites before the policy implementation, viewing them as directly impacted by the policy. In contrast, *unbanned streamers* are defined as those who streamed online gambling, but only from unbanned gambling websites. Since they streamed content similar to what was banned, the policy might indirectly affect these streamers. Finally, all streamers who never streamed online gambling are classified as *untreated streamers*.

However, while we have high-frequency data of all gambling livestreams during our studied period, we do not observe which streams are about banned websites, and hence we are ignorant about the list of banned streamers based on only the streaming dataset. We address this challenge by leveraging three additional data sources: historical video clips, high-frequency stream titles and in-stream chats logs. We provide further details of identifying banned content and banned streamers in Section 5.

Table 1 reports the pre-treatment summary statistics at the streamer level. Our final dataset comprises 158 banned streamers, 317 unbanned streamers and 4,626 untreated streamers. Streamers across all groups exhibit similar streaming patterns, characterized by metrics such as the total number of games played and average weekly livestreaming hours. We find that banned streamers in general rely more on gambling streams than unbanned streamers, both in terms of total hours spent on gambling streams and the proportion of gambling streams relative to total livestreams.

Table 1: Summary Statistics of Streamer Groups

	Banned	Unbanned	Untreated
Total Number of Games Played	17.574 (20.945)	20.076 (32.391)	22.950 (32.320)
Hours of Livestreams by Week	25.613 (19.301)	21.175 (18.805)	21.375 (18.571)
Hours of Gambling Streams by Week	9.881 (13.789)	5.181 (13.000)	0.000 (0.000)

Notes. The table reports the streamer-level summary statistics for three groups of streamers. All observations are aggregated at the weekly level and calculated based on data from pre-treatment periods.

Video game attributes. To examine potential substitution patterns between online gambling and video games featuring gambling-like content, we combine a rich game-level dataset with the streamer-level panel dataset described above. Our game-level dataset is obtained from the API of *IGDB*, one of the most extensive online video game databases owned by Twitch. For each game, we obtain game attributes that encompass release date, genres, supporting gaming platforms, available languages, age ratings, and the presence of remakes or expansions.

Games with gambling-like content. We classify all video games that were streamed at least once in our streamer-level dataset based on the inclusion of gambling-like features. In light of the absence of a comprehensive dataset containing information on this feature, we compile a list of games with gambling-like content by investigating three primary sources: first, we extract a roster of games featuring the concept of "Loot boxes" or "Gacha system" from *GiantBomb*. These terms encapsulate the practice of randomly obtaining in-game items with predetermined odds in video games, through the use of in-game currencies that can be purchased with real money. Consequently, we suggest that engaging with these features mimics aspects of genuine gambling experiences. Second, we employ text-mining techniques to scrutinize stream titles in our high-frequency dataset. We collect content from streams with titles containing keywords or phrases that are closely connected to gambling-like activities, such as "loot box opening", "insane pulls" or "getting shafted", and we record games played under these titles. Third, we identify games featuring gambling-like content by examining their PEGI and ESRB ratings, as some rating details disclose whether loot boxes are present in the games. We manually combine information from all three sources. By synthesizing data from all three sources, we compile a list of 464 unique video game titles within our streaming data identified as containing gambling-like features.

In addition to these games, we categorize other streaming content into three groups: online gambling, games without loot boxes, and non-gaming content such as chatting or outdoor activities. Specifically, the online gambling category encompasses all streams labeled under "Slots" (the largest gambling category on Twitch), "slots!", "Blackjack", "Casino" and "Virtual Casino". As the policy ban did not cover all online gambling content, we provide a detailed discussion in Section 5 of our methods for detecting banned versus unbanned content within online gambling streams.

Video data. We collect all video clips from online gambling streams of individual streamers before the policy implementation using Twitch API's get clips endpoint. We use these video clips to identify the presence of banned content and therefore the latent treatment status of the streamers. We discuss this process in detail in Section 5.1.

Chat logs data. We obtained in-stream chat logs from *Streams Charts*, whose data of coverage aligns with the coverage of our high-frequency streaming data. For each stream of a streamer, we observe the chat messages (texts), the account names and IDs of the chatters who posted the messages, and the timestamps of when the messages were posted. We use the chats as an additional source to identify the latent treatment status of streamers (which we detail in Section 5.3). Additionally, we use chatters (i.e., individual registered viewers on Twitch) to construct valid treated and untreated groups for demand-side analysis (which we detail in Section 8).[4] In Appendix A, we show that chatters, although being a subset of all viewers in streams, is a reliable proxy for overall viewership base in streams.

Subscriptions and revenue data. We collect from *Streams Charts* the number of subscribers gained per stream of a streamer and its breakdown, which includes all types (i.e., new subscriptions, re-subscriptions and gifted subscriptions) and tiers (i.e., Tier 1, Tier 2, Tier 3).[5] We use the data for revenue analysis in Section 8.

Other streamer-level data. For each streamer, we collect their country and city of residence, and their histories of prohibition on Twitch (i.e., the number of channel prohibitions, the dates and the duration of each prohibition) from *Streams Charts*. We use the residence information as one source to match video clips to streams (see discussions in Appendix B.1). Additionally, we use details on prohibitions to explore the underlying mechanisms behind the observed heterogeneous treatment effects in Section 8.

Website traffic data. We obtain *SimilarWeb*'s website traffic data from *Dewey*, which contains monthly total traffic on a domain and subdomain level starting from September 2018 with a breakdown by device type. We use the data to examine whether banned gambling websites have experienced sudden traffic changes due to the policy (see Section 9).

4.1 Descriptive Evidence

We first present some descriptive evidence on the impact of implementing banning policy. Figure 1 illustrates the log of weekly average streaming hours over time in our sample, whereas the vertical dashed lines show the date of announcing the policy (September 20th, 2022) and implementing the policy (October 18th, 2022). Both treated groups show downward trends in weekly streaming hours after the policy implementation, whereas the weekly streaming hours of streamers in the untreated group are relatively more stable over time. In addition, the seasonal trend has a similar effect on all three groups both before and after the policy implementation.

[4] The Twitch API only has viewer count in streams, but not the list of individual viewers in streams.

[5] This dataset is only available for streamers who enabled channel tracking. Streams Charts do not count subscriptions when a channel is offline and do not track auto-renewed subscriptions unless the viewer notifies auto-renewal by clicking appropriate button in chats. While we use this subscription data for revenue discussion, we note its limitation of possibly not being fully representative, as streamers self-select into disclosing this information.

Figure 1: Average Log Weekly Streaming Hours

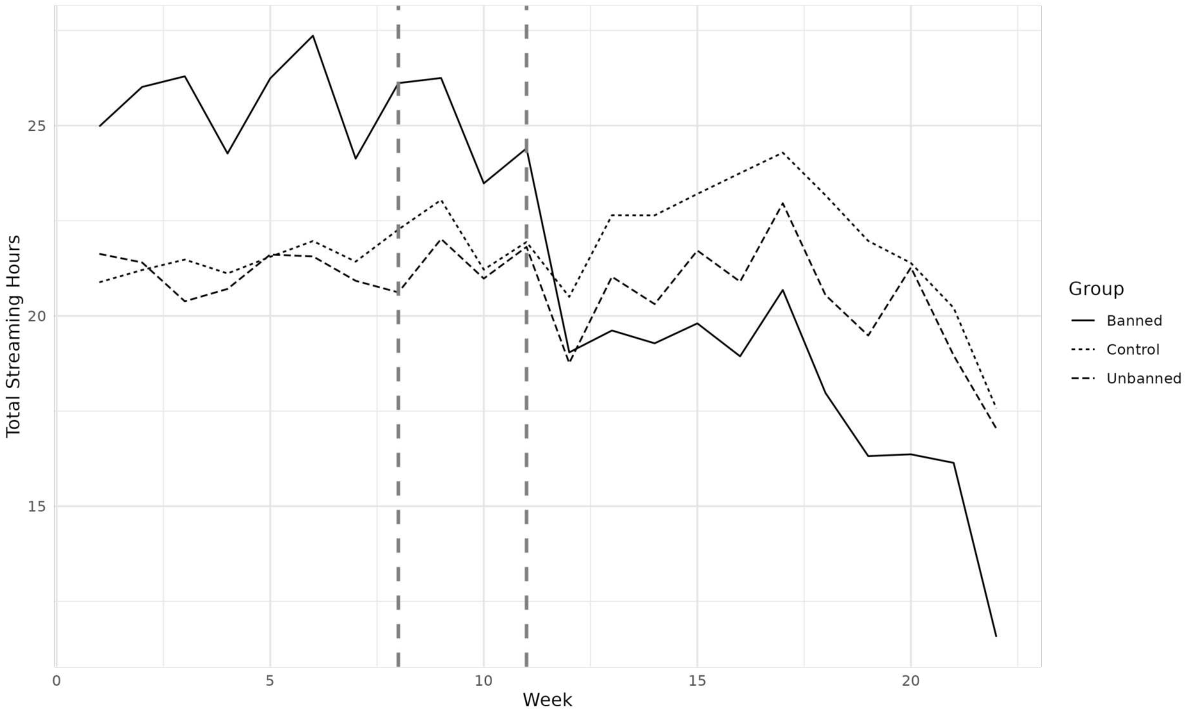
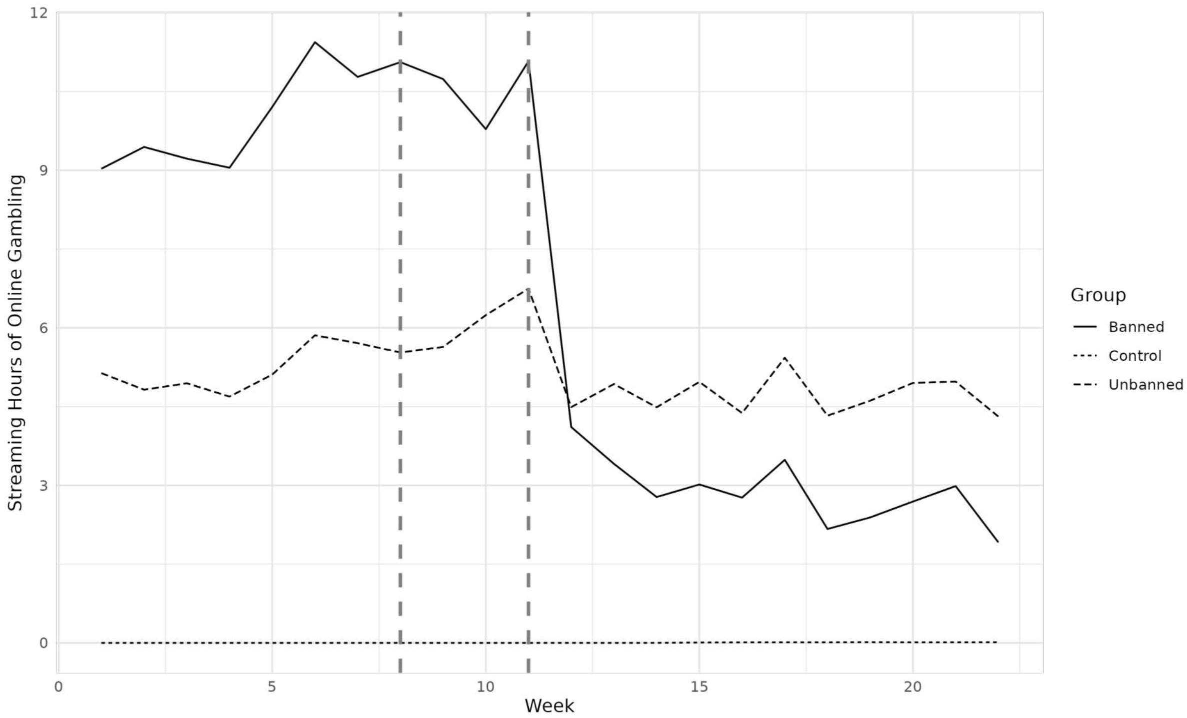


Figure 2 displays the changes in average log weekly streaming hours of online gambling over time. As expected, we observe that banned streamers significantly reduced weekly streaming hours of gambling after the policy implementation. Moreover, we find that unbanned streamers also reduced their supply of gambling livestreams, but with a much smaller magnitude. This trend provides some preliminary evidence that the policy also had impact on streamers who were not directly affected. Interestingly, both groups of treated streamers show a slight increase in supply of gambling livestreams during the weeks between the policy announcement and implementation, which may be due to that the announcement temporarily increased demand of gambling content after the policy was announced.

To summarize, our descriptive evidence suggests that streamers who had streamed gambling content before the policy announcement decreased both their total livestreams and their livestreams of gambling content, whether they were directly or indirectly affected. Nonetheless, the details of the policy's impact, such as the actual magnitudes, persistence or heterogeneous treatment effects, remain unclear.

Figure 2: Average Log Weekly Streaming Hours of Online Gambling



5 Detection of Banned Content and Streamers

As discussed in Section 4, our streamer-level data does not specify whether a gambling stream was on banned or unbanned websites, nor does it indicate whether a streamer has streamed banned gambling content. Therefore, we need to leverage additional data sources to address these issues of missing information. In this section, we describe how we systematically detect production of banned gambling content, identify treated groups and measure the intensity of banned content production of each streamer. We combine multiple methods and data sources including videos, stream titles and chat logs to detect banned streams, which we overview below and describe in more details in Appendix B.

5.1 Identifying Banned Content From Video Clips and VODs

First, we construct a list of streamers who had history of streaming gambling content at least once before the policy implementation. We fetch URLs of all past gambling video clips of a streamer pre-policy from Twitch API.[6] Clipping (up to 60 seconds) is a functionality enabled by Twitch that allows both the streamer and any logged-in viewer to capture and share moments

[6]We restrict our analysis to content that was produced between August 1st and October 17th, 2022 (inclusive) to align with the coverage of our stream-level data.

from a streamer's livestreams. Thus, we treat clips as random stream moments, for which we can use as one source to check any existence of banned content production by streamers.

We identify banned content from video clips using a procedure as shown in Figure 3. The procedure involves three key steps: In **step 1**, we download all historical video clips of a streamer (in total over 4TB in size) from clip URLs using *youtube-dl* package (which also supports twitch video downloads). In **step 2**, we decompose each video into a sequence of frames. To optimize the processing workload, we sample three frames from each video, covering the start, middle and end of each clip for content analysis. This strategy is effective since clipped videos are under 60 seconds with minimal scene changes, and thus sampling reduces the total number of frames that need to be analyzed without losing critical information. Finally, in **step 3**, we transform each sampled frame to grayscale[7] and use Google's *Tesseract-OCR* engine to extract all texts from each frame. We check the extracted texts against all banned websites' keywords and use a restrictive criterion by classifying a clip as only containing banned content if keywords are detected in all 3 frames of a video (or unbanned if absent in all frames), to ensure accuracy of the classification.

We complement our analysis on video clips with additional Video on Demand (VOD) posted on YouTube that are available for a few prominent streamers, including xQc, Adin Ross and ItsSliker. VODs are archives of content previously streamed live on Twitch and can span several hours (if unedited).[8] For these long VODs, we manually inspect them to identify the presence (and time of presence) of the banned content. We match these videos (video clips and VODs) to their original streams to compile a list of gambling streams and streamers with banned content. The detailed process of our matching procedure is provided in Appendix B.1.

We note that not all streamers have video clips or VODs accessible (e.g., some streamers may have disabled the clipping functionality on their streams). Therefore, we propose a few additional data sources and methods to help us identify the latent treatment status of streamers, which we detail next.

5.2 Identifying Banned Content From Stream Titles

Twitch streamers sometimes put an exclamation point followed by a keyword (e.g., !Stake, !Roobet) in stream titles. These are commands that viewers can type into a channel's chat

[7]Grayscale conversion is a common pre-processing step in video analysis. It helps reduce reduce distractions caused by color variance and enhances the focus on text detection.

[8]Although VODs on Twitch are only saved up to 2 months for Partners, streamers can enable VOD storage and export their videos to other platforms, such as YouTube, by linking with their Twitch accounts. For example, xQc's VODs can be found on YouTube via `https://www.youtube.com/@xqcgames3433/videos`.

Figure 3: Banned Content Detection Using Videos

Notes. This figure shows the procedure for which we detect banned gambling content from videos.

to interact with the streamer, and the commands will trigger a bot to respond with the required information. When names of the banned websites are used as keywords in gambling streams, viewers can automatically receive these websites' referral links and streamers can earn commissions from whenever a viewer signs up on the gambling websites and uses their services.[9]

Recognizing this user habit, we leverage our high-frequency streaming data, which captures stream titles and the games streamed at every 10-minute intervals prior to the policy implementation, as an additional source for identifying banned content in gambling streams. We define a stream as containing banned content when both of the following conditions are met: (1) the presence of "!" followed by banned website keywords in stream titles, and (2) the game being streamed is categorized as online *gambling* content.

We demonstrate the validity of this approach for detecting additional banned content (and streamers) in Table B.2.1. Our findings show that *non-gambling* streamers typically do not include such commands in their stream titles. In addition, since these commands are often used to enhance viewer engagement in chats, streamers have low incentives to include commands in titles that do not align with their streaming content, as it could risk reducing engagement. Furthermore, many streamers in our dataset are top streamers who treat streaming as a full-time job, thereby reducing the likelihood of of inappropriate use of commands. Overall, our approach is more conservative compared to defining a stream as containing banned content solely based on the presence of banned website names in stream titles (see Appendix B.2 for

[9]This is common practice within the community of gambling content streams. It became even more prevalent ever since Twitch banned the sharing of links to gambling sites in 2021, in response to concerns from the National Council on Problem Gambling and its allies regarding the risk of under-age gambling and gambling addiction. Source: https://www.ncpgambling.org/news/ncpg-responds-to-twitch-banning-gambling-sites-links/.

additional discussions). This approach ensures the accuracy of classification while minimizing the risk of misclassifying unbanned content as banned (i.e. minimizing false positives).

We note that, however, that not all streamers include such commands when streaming banned content. Therefore, we complement our banned content detection with a third source - chats logs, which we detail next.

Table 2: Banned Content Detection Using Stream Title

	Streamers with Gambling Content	Streamers without Gambling Content
Total Number	475	4,626
Banned Referrals in Any Stream Titles	52	2
Banned Referrals in Gambling Stream Titles	50	0

Notes: This table shows the number of streamers who were detected using referral links to banned websites in any stream titles versus only in gambling stream titles in the two groups. We detect banned referrals by checking instances of an exclamation point followed by a banned website keyword in stream titles.

5.3 Identifying Banned Content From Chat Logs

We use chats logs as the third additional source to predict the presence of banned content for streams that neither have OCR nor commands in stream titles. To achieve this, we need to construct ground truth chat samples by identifying ground truth streams that feature banned and unbanned content, respectively. We use multiple criteria to construct the samples.

For the banned ground truth chats, we use chats from streams that meet the following two criteria: (1) the gambling streams must have video clips that match to it where we have already found the presence of banned content in all three sampled frames of a video using OCR, and (2) the gambling streams must have banned websites' referral links included in stream titles. We note that while each source alone can be treated as the ground truth, combining the two criteria together ensures high quality of the ground truth sample.

For the unbanned ground truth chats, we use chats from streams that satisfy all three of the following criteria: (1) the gambling streams must have video clips with no banned website names detected in the 3 sampled frames of a video using OCR, (2) the streams that come from streamers who have no history of streaming banned content based on OCR detection on all their pre-policy gambling stream video clips, (3) the gambling streams must have neither banned referral links nor banned keywords featured in stream titles.

We note that using multiple criteria for constructing the unbanned ground truth is necessary because the absence of banned content in all 3 frames of a video as detected by OCR is necessary but not sufficient to classify a stream as unbanned. For example, a clip might show content from a banned website that OCR cannot detect if it does not feature the banned website's logo or keyword in any frame. However, if OCR detects banned content, it confirms that the streamer indeed streams banned content. We provide more discussions regarding the robustness of our ground truth sample construction approach in Appendix B.3.

Using these stringent criteria, we obtain a ground truth chats sample, covering 534 banned streams and 821 unbanned streams, which serves as the basis for our stream classification below.

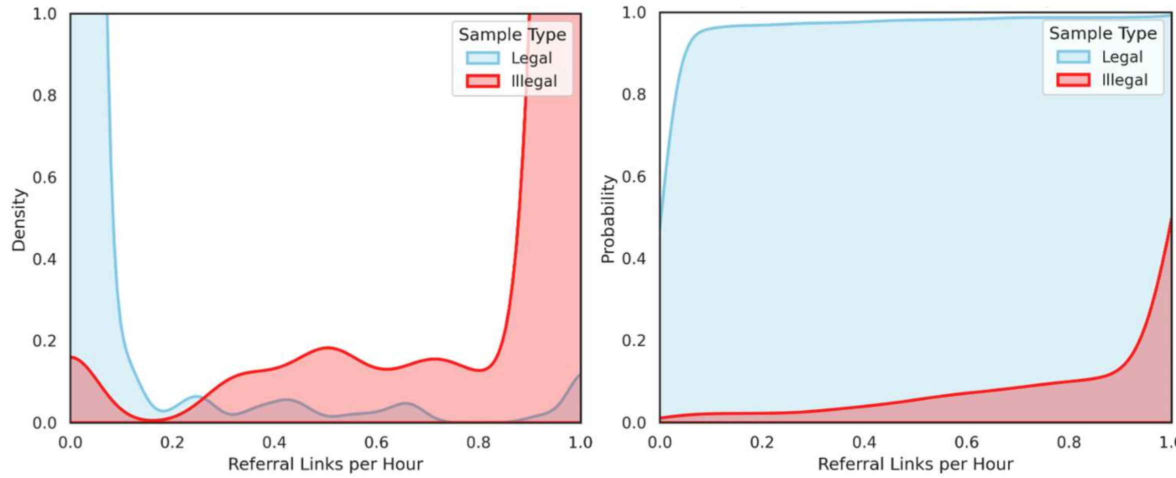
Inference Step. Based on industry knowledge and observed practices on Twitch, we hypothesize that banned streams will contain a higher proportion of referral links to banned websites. This is because streamers who engage in promoting banned content often use chat bots to automatically respond to viewers who required referral links to these websites,facilitating viewer access.

We test this hypothesis by first checking the summary statistics (Table 3) of banned and unbanned chats samples. We find that banned gambling streams contain 100 times more referral links of banned websites — despite having fewer total chats. Additionally, banned gambling streams have 130 times more unique viewers who posted banned referral links in chats. Both statistics suggest that there is a difference in referral links of banned websites in chats between banned and unbanned groups.

Based on the summary statistics, we consider a test statistic of the number of banned referral links per hour. Figure 4 shows the density and distribution of number of referral links in banned and unbanned stream samples. We observe that the distribution of referral links in unbanned streams has a huge mass at zero, where banned streams have a significantly larger probability of containing more than one referral links per hour. The difference in the distributions is further supported by the right panel, where we show that the distribution of referral links in banned streams first-order statistically dominates that in unbanned streams. This evidence supports that using referral links of banned websites in chats could be a valid criteria for identifying banned and unbanned streams.

Next, we evaluate the distributional difference formally by conducting a bootstrapped Kolmogorov-Smirnov test for 5,000 repetitions, deriving the KS statistics in each bootstrap sample using the function from Bootstrapped KS2 Tester, based on 534 streams of both banned and unbanned content. Figure 5 depicts the distribution of bootstrapped KS statistics, where this statistic is defined as the maximum absolute difference between two empirical distributions. Since the average p-value is significantly less than the usual nominal level, we always reject the null that the numbers of referral links follow the same distribution in banned and unbanned streams. Moreover, we find that the threshold-based classification with the bootstrapped mean estimator of KS statistic (0.9501) as the threshold value shows good out-of-sample prediction performance. Therefore, we conclude from the inference step that a threshold-based classification is sufficient for detecting banned streams in our research, where our target is to find an optimal threshold value for the classification method in the prediction step.

Figure 4: Distributions of Referral Links per Hour in Banned and Unbanned Gambling Streams



Notes. This figure shows the density and distribution of referral links per hour in banned (red) and unbanned (blue) gambling streams, where we right-censored the number of referral links per hour at one in both panels. The left panel displays the density of referral links per hour, highlighting that banned gambling streams have a significantly higher probability of containing a large number of referral links compared to unbanned streams. The right panel shows the cumulative distributions of referral links, indicating that the distribution of referral links in banned streams first-order statistically dominates that in unbanned streams. These visualizations support our hypothesis that banned streams has a very distribution of referral links, making the number of referral links per hours a robust criterion for distinguishing between banned and unbanned gambling streams.

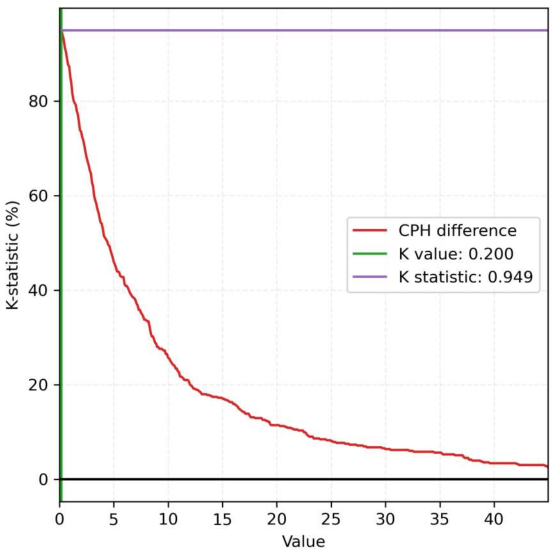
Table 3: Summary Statistics of Ground Truth Chats Samples

	Banned Gambling Streams	Unbanned Gambling Streams
Total Chats	4,031,759	4,305,627
Unique Streams	534	821
Unique Streamers	39	99
Unique Viewers	199,450	258,913
Banned Referrals in Chats	18,872	152
Viewers Posting Banned Referrals in Chats	12,155	92

Notes: This table presents the summary statistics of the ground truth banned and unbanned chat samples.

Prediction Step. To find an optimal threshold that jointly minimizes Type I and Type II errors

Figure 5: Distribution of Bootstrapped KS Statistics in Inference Step



Notes. This figure shows the distribution of Bootstrapped Kolmogorov-Smirnov statistics.

for classification, we propose to use grid search with 10-fold cross validation.[10] To achieve this, we first correct the imbalance in our ground truth samples (534 banned versus 821 unbanned streams) by implementing a *random undersampling* (Breiman, 2017) during the training phase of each cross validation fold. Next, we employ a 10-fold cross validation with grid search and evaluate thresholds in-sample to identify the optimal threshold.[11] We choose a threshold of 0.3, which jointly minimizes Type I and Type II errors in-sample.[12]

The threshold-based approach demonstrates a strong capability to distinguish between two sample groups. On average there is a 93% probability that the method is able to distinguish between banned and unbanned samples (See Figure B.4.1 in Appendix for confusion matrices across folds, and Table B.4 for detailed in-sample and out-sample performance.)

Identification of Treated Groups Using All Sources. We combine detection results from all three approaches (video clips, stream titles and chat logs) described in this section and collect a list of banned streams. Then, we use these identified banned streams to classify treated streamers into banned and unbanned groups.

[10]In our context, type I error is defined as incorrectly classifying an banned stream as banned, and type II error is defined as incorrectly classifying a banned stream as unbanned.

[11]We set the range of grid search to be from 0 to 2, with increments of 0.05.

[12]A threshold of 0.3 also achieves good performance on other measures. See Appendix B.4 for detailed discussions.

6 Empirical Strategy

In this section, we outline the three main empirical strategies for estimating the causal effects of the banning policy. Our primary goal is to assess how streamers whose content were banned changed their production decisions of livestreaming, and how this affected the demand for their livestreams and their revenue. Additionally, we are interested in whether these impacts also applied to streamers who streamed online gambling content under the threat of a future ban. These policy impacts inform us about the effectiveness of the content regulation and reveal any potential side effects that might not align with the platform's incentives. In Section 4.1, we show that all groups of streamers exhibit similar trends in both average log weekly streaming hours and average log weekly streaming hours of online gambling. Therefore, we use the classic difference-in-differences (DiD) framework to estimate the policy impact as the average treatment effect on the treated (ATT).

Nonetheless, we face several identification challenges: first, the classic two-way fixed effect DiD model usually assumes for parallel trend assumption (PTA) between streamers with different treatment status. Although our descriptive evidence suggests that different groups of streamers tend to follow similar trends over time, the assumption still remains questionable because streamers might pursue strategic changes in their streaming plan in response to the policy implementation. The potential strategic interaction may potentially affect both supply- and demand-side outcome variables in our identification. Second, the policy may induce streamers to change their streaming content, leading viewers to switch to other channels to fulfill their interests. Consequently, the banning policy may generate equilibrium spillovers to the observed demand of streamers, resulting in a upward bias in the DiD estimates of demand-side variables or even potential bias in supply-side estimates. Therefore, using untreated streamers who share viewers with treated streamers as the untreated group leads to potential violation of the Stable Unit Treatment Values Assumption (SUTVA). Finally, since there is no pure randomization for the groups of treated and untreated streamers, these streamers may have systematic differences in both observed and unobserved factors, leading to potential bias in two-way fixed effect DiD estimates.

To overcome these identification challenges, we start with the classic two-way fixed effect DiD model as the baseline, and subsequently turn to discuss two alternative empirical approaches in the main analysis. To alleviate concerns over the violation of the parallel trend assumption, we adopt the synthetic difference-in-differences (SynthDiD) estimator proposed by Arkhangelsky et al. (2021); Berman and Israeli (2022). To address potential systematic differences between different groups of streamers, we apply the doubly-robust estimator of group-time average treatment effects proposed by Callaway and Sant'Anna (2021), which utilizes generalized propensity scores to balance units between the treated and untreated groups. Finally, to ensure our estimates are unbiased under potential violations of SUTVA, we leverage network analysis on chat logs to evaluate the extent of overlapping viewership between streamers. This allows us to reconstruct the treated and untreated groups based on a selected subset of streamers and use them for additional checks of validity of our estimated treatment effects.

6.1 Two-way Fixed Effect DiD

We begin with the classic DiD model including streamer and week fixed effects:

$$
\begin{aligned}
y_{it} = \alpha_i + \gamma_t + \beta_1 \cdot \text{Banned}_i \cdot \text{Post}_t + \beta_2 \cdot \text{Banned}_i \cdot \text{Announcement}_t \\
+ \beta_3 \cdot \text{Unbanned}_i \cdot \text{Post}_t + \beta_4 \cdot \text{Unbanned}_i \cdot \text{Announcement}_t + \varepsilon_{it}
\end{aligned}
\tag{4}
$$

where Banned_i and Unbanned_i are indicators of the two treated groups, and Post_t and Announcement_t are indicators of post-treatment periods and periods between the policy announcement and the policy implementation. On the supply side, the dependent variable y_{it} includes supply-side outcomes including log of weekly streaming hours of each type of streaming content we described in Section 4 and log of total weekly streaming hours. On the demand side, we use log of total hours watched by all viewers, and three tiers of subscriptions as dependent variables. Our primary focus lies in the estimated values of β_1 and β_3, which capture the policy's effects on banned streamers and unbanned streamers after the policy implementation.

Additionally, we leverage a flexible event-study model to estimate the time-varying treatment effects. Compare to our baseline specification with 3 time periods, the event-study estimates provide more insights into the persistence of the treatment effects. Following the empirical suggestions from Freyaldenhoven et al. (2021), we normalize the causal estimate one period ahead of the policy implementation to zero, and report both the point-wise confidence intervals as well as the uniform sup-t confidence bands for the event-time path of the treatment effects.

6.2 Synthetic DiD

Arkhangelsky et al. (2021) introduced Synthetic DiD (SynthDiD) method, which combines the synthetic control method (Abadie and Gardeazabal, 2003; Abadie et al., 2010) with classic DiD regression, whereas it has been used commonly in marketing research involving panel studies, e.g. Berman and Israeli (2022). SynthDiD addresses the identification challenge related the parallel trend assumption by leveraging the idea of synthetic control to construct an "artificial" parallel trend. The method is implemented in two steps: in Step 1, we determine a set of unit weights $(\hat{w}_i)_{i=1}^N$ and time weights $(\hat{\lambda}_t)_{t=1}^T$ to align pre-treatment trends between the treated group and the untreated group. The unit weights are used to prioritize the role of untreated units that are more similar to the treated units, while the time weights are used to prioritize time periods such that the corresponding time trends closely resemble those from the pre-treatment periods. In Step 2, we estimate ATT based on the following specification including both individual and time fixed effects:

$$(\hat{\beta}^{\text{SynthDiD}}, \hat{\alpha}, \hat{\gamma}) = \underset{\beta,\alpha,\gamma}{\arg\min} \left\{ \sum_i \sum_t (\tilde{y}_{it} - \alpha_i - \gamma_t - \beta_j^{\text{SynthDiD}} \cdot D_{j,it}) \hat{w}_i \hat{\lambda}_t \right\} \tag{5}$$

whereas $\beta_j^{\text{SynthDiD}}, j = 1, 2, 3, 4$ is the SynthDiD counterpart of the DiD estimator β_j in our baseline specification, and $D_{j,it}$ denotes the corresponding interaction term of group and time indicators. To adopt this estimator under our specification with two group indicators and two time indicators, we follow the suggestion from Clarke et al. (2023) by first partialing out all interactions of indicators other than $D_{j,it}$, then performing the SynthDiD estimator based on the residual $\tilde{y}_{it}$.

6.3 Doubly-Robust Estimator of Group-Time ATT

To tackle the challenges associated with selection in our identification, we utilize the doubly-robust estimator for group-time average treatment effects introduced by Callaway and Sant'Anna (2021). This method features inverse propensity score weighting that utilizes generalized propensity score, which is defined as

$$P_{g,s}(X) = P(G_g = 1 \mid X, G_g + C = 1) \tag{6}$$

where G_g is the indicator that the unit is first treated at time g, and the second condition ensures that the unit is either in group g or in the never-treated group indicated by C. Since all units are treated at the same time and there is a group of streamers who never receive the treatment, our identification relies on the following conditional parallel trend assumption based on the "Never-Treated" Group:

$$\mathbb{E}[y_t(0) - y_{t-1}(0) \mid X, G_g = 1] = \mathbb{E}[y_t(0) - y_{t-1}(0) \mid X, C = 1] \quad \text{a.s.} \tag{7}$$

The average treatment effect of the treated for cohort g at time t is non-parametrically identified as

$$\beta(g,t) = \beta_{DR}(g,t;\tau) = \mathbb{E}\left[\left(\frac{G_g}{\mathbb{E}(G_g)} - \frac{\frac{P_g(X)C}{1-P_g(X)}}{\mathbb{E}\left(\frac{P_g(X)C}{1-P_g(X)}\right)}(y_t - y_{g-\tau-1} - m_{g,t,\tau}(X))\right)\right] \tag{8}$$

whereas $m_{g,t,\tau}(X) = \mathbb{E}[y_t - y_{g-\tau-1} \mid X, C = 1]$. Finally, the aggregate treatment effect $\hat{\beta}^{DR}$ is estimated by averaging over all $\widehat{\text{ATT}}(g,t)$ based on their groups. Through leveraging the generalized propensity scores, this estimator effectively addresses the aforementioned selection issue. Moreover, it enables us to estimate the time-varying average treatment effects and compare them with event-study estimates. We omit more technical details and refer readers to Callaway and Sant'Anna (2021).

Finally, even with the use of advanced econometric tools, the identification challenge posed by potential violation of SUTVA still remains. To address this issue, we leverage network analysis to construct comparable groups of treated and untreated streamers for our identification. We discuss this approach in detail in Section 8.2.

7 Supply-Side Results

In this section, we report DiD, SynthDiD and the doubly-robust estimators of the ATT of banning policy on supply-side outcome variables. For clarity, we interpret our results based on the SynthDiD estimates.

7.1 Content Creation of Gambling and Non-Gambling Livestreams

Table 4 reports the impact of the banning policy on gambling livestreams. We find that the supply of gambling livestreams decreased across the platform after the policy implementation, with banned streamers reducing their gambling streams by 63.2% (= $\exp(-1.001) - 1$) and unbanned streamers reducing theirs by 12.2%. These estimates have two implications: first, in addition to the banned gambling content, banned streamers reduced livestreams of unbanned gambling content as well as banned gambling content. This is because the overall share of banned streams among all gambling streams for these streamers is 55.8%, which is smaller than the 63.2% reduction caused by the policy. Second, even though Twitch only banned four websites, streamers not directly affected by the policy also reduced their creation of gambling content. Therefore, the banning policy was successful in regulating the overall supply of online gambling content across the platform.

Table 4: The Impact of Twitch's Ban Policy on Gambling Content Creation

	log(Gambling hours + 1)		
	TWFE	SynthDiD	ATTGT
Banned (β_1)	-1.022***	-1.001***	-1.023***
	(0.089)	(0.082)	(0.102)
Δ%	-64.0%	-63.2%	-64.0%
Unbanned (β_3)	-0.116**	-0.130***	-0.242***
	(0.049)	(0.043)	(0.049)
Δ%	-11.0%	-12.2%	-21.5%
Streamer FE	✓	✓	✓
Week FE	✓	✓	✓
Observations	112,222	112,222	112,222
Mean dependent variable	1.131	1.131	1.131

Notes. All standard errors are clustered at the streamer level. The mean dependent variable is calculated based on observations of all treated streamers before the policy announcement. FE, fixed effect.
Significance level: $*p < 0.1$; $**p < 0.05$;$***p < 0.01$

To answer the question that whether potential substitution between online gambling and gambling-like features in video games weakened the effect of banning policy on Twitch, we report estimated treatment effect on the supply of video games featuring gambling-like elements in Table 5. The overall non-significant estimates suggest that on the supply side, gambling streamers did not treat video games with gambling-like features as a means to pursue to maintain their viewership. Our findings provide new evidence on the similarity between traditional gambling and gambling-like activities in video games, indicating that the potential substitution between these two types of content might not be as interchangeable as some policymakers have presumed.

Finally, we turn to examine the policy's impact on total content creation on the treated streamers. Estimates in the first three columns of Table 6 show that both banned and unbanned streamers reduced their total streaming hours after the policy implementation. However, were the reductions solely due to the decrease in gambling content creation for both banned and unbanned streamers, as we observe in Table 4? We further investigate the sources of the decline

Table 5: The Impact of Twitch's Ban Policy on Content Creation of Video Games with Gambling-like Features

	log(LootBox Games + 1)		
	TWFE	SynthDiD	ATTGT
Banned (β_1)	0.012	0.021	0.089
	(0.065)	(0.054)	(0.076)
$\Delta\%$	-	-	-
Unbanned (β_3)	-0.046	-0.054	-0.096
	(0.046)	(0.036)	(0.053)
$\Delta\%$	-	-	-9.1%
Streamer FE	✓	✓	✓
Week FE	✓	✓	✓
Observations	112,222	112,222	112,222
Mean dependent variable	1.487	1.487	1.487

Notes. All standard errors are clustered at the streamer level. The mean dependent variable is calculated based on observations of all streamers before the policy announcement. FE, fixed effect.
Significance level: $*p < 0.1$; $**p < 0.05$;$***p < 0.01$

in content creation and find that, relative the untreated group streamers, both group of treated streamers reduced their livestreaming of video games without gambling-like features by similar magnitudes of 13.8% and 12.3%, respectively. These results highlight the need for attention from the platform, as the banning policy reduced non-targeted content creation among both directly and indirectly affected streamers. Since video games without gambling-like features encompass a wide variety of game genres, this implies that the policy may inadvertently harm the diversity of content on the livestreaming platform.

7.2 Supply-side Event Study

Given that the banning policy reduced the supply of gambling streams and caused spillover effects on unbanned gambling streams, we now investigate the persistence of the impact. The effectiveness of the policy might diminish if treated streamers temporarily ceased gambling livestreams immediately after the policy's implementation, but resumed them later when demand increased, such as during holidays. We use an event-study to analyze potential changes in the treatment effects over time, and to provide visual evidence on the parallel trend assumption. Following the empirical suggestions from (Freyaldenhoven et al., 2021), we normalize the point estimate one period before the policy implementation to zero, and report both the point-wise confidence intervals (illustrated by the inner bars) and the sup-t confidence bands (Montiel Olea

Table 6: The Impact of Twitch's Ban Policy on Gambling Livestreams on Total Streaming Hours and Streaming Hours of Games without Gambling Features

	log(Streaming hours + 1)			log (Other games + 1)		
	TWFE	SynthDiD	ATTGT	TWFE	SynthDiD	ATTGT
Banned (β_1)	-0.583***	-0.585***	-0.543***	-0.146***	-0.149***	-0.172***
	(0.084)	(0.075)	(0.096)	(0.040)	(0.032)	(0.050)
Δ%	-44.2%	-44.3%	-41.9%	-13.5%	-13.8%	-15.6%
Unbanned (β_3)	-0.182***	-0.194***	-0.242***	-0.133***	-0.130***	-0.075**
	(0.044)	(0.040)	(0.050)	(0.032)	(0.026)	(0.036)
Δ%	-16.6%	-17.6%	-21.5%	-12.5%	-12.2%	-7.2%
Streamer FE	✓	✓	✓	✓	✓	✓
Week FE	✓	✓	✓	✓	✓	✓
Observations	112,222	112,222	112,222	112,222	112,222	112,222
Mean dependent variable	2.587	2.587	2.587	1.079	1.079	1.079

Notes. All standard errors are clustered at the streamer level. The mean dependent variable is calculated based on observations of all streamers before the policy announcement. FE, fixed effect.
Significance level: $*p < 0.1$; $**p < 0.05$; $***p < 0.01$

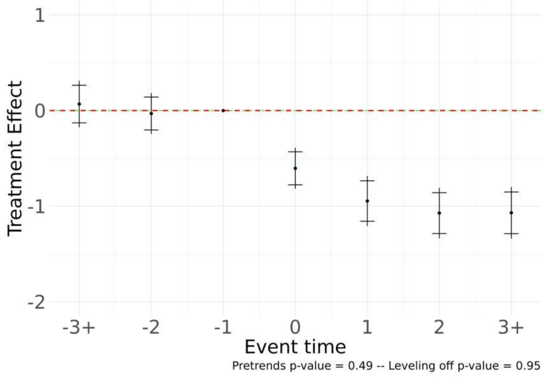
and Plagborg-Møller, 2019) (illustrated by the range of vertical lines).

Figure 6 shows the estimates of time-varying treatment effects on the four types of content creation among banned streamers discussed in Section 7.1. Consistent with our estimates reported in Table 4, we find that the policy led to a persistent reduction in gambling livestreams after the policy implementation. Interestingly, we observe that the estimate is significantly larger in magnitude in the second week after policy implementation compared to that in the first week. This result suggests that the reduction in supply of gambling content among banned streamers were not completely from the banned websites (because they had been removed from the platform from the first week after the policy implementation). Streamers complete their adjustment of gambling content creation after the second week, as the estimates are not statistically different thereafter. The other three subfigures also exhibit persistent treatment effects after the policy implementation. In Figure 7, we show plots of the same outcome variables for unbanned streamers, whereas we find a similar persistent policy impact with smaller magnitudes, which are consistent with our point estimates.

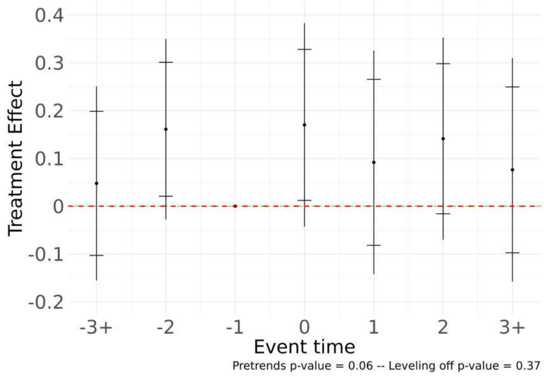
7.3 Heterogeneous Treatment Effects

We now explore whether the policy impact on content creation varies across different types of streamers, as examining the impact on subgroups of streamers help us better understand the policy's effectiveness and investigate the underlying mechanisms.

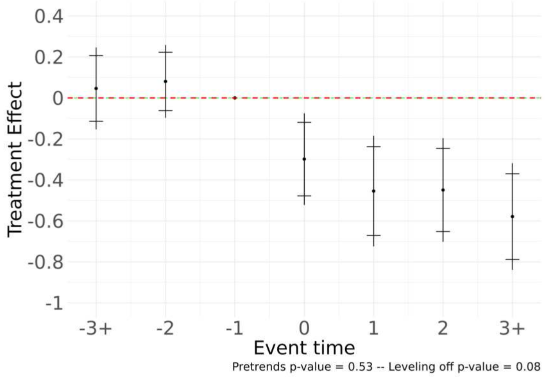
Figure 6: Time-Varying Treatment Effects on Content Creation in Group of Banned Streamers



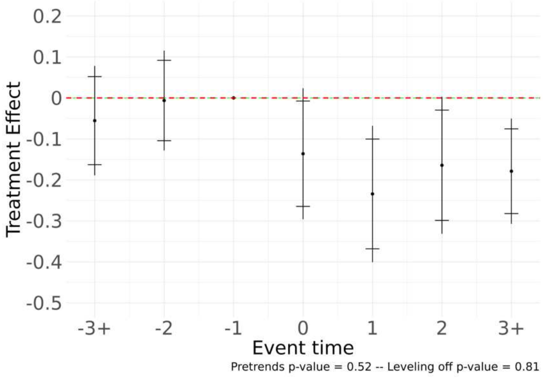
(a) Gambling Content



(b) LootBox Games



(c) Total Livestreams



(d) Other Games

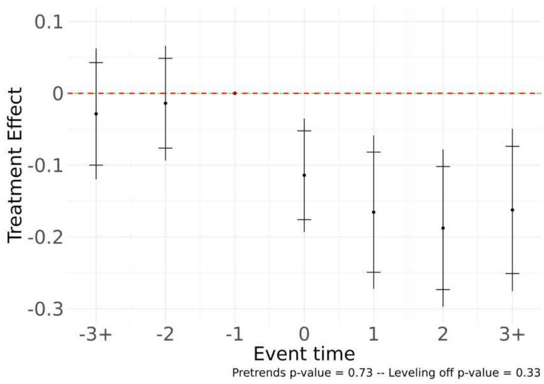
Notes. 1. The coefficients in each subfigure show the point estimates of β_1. 2. β_1 in one period ahead of the policy implementation is normalized to one. 3. The 95% point-wise confidence intervals are illustrated as the inner bars, and the sup-t confidence bands are illustrated as the vertical lines. 4. The p-values are derived from the test of pre-trends and the test of existence of dynamic effects of the policy. We omit more technical details and refer readers to (Freyaldenhoven et al., 2021).

HTE on Main Language. First, we examine potential heterogeneous effect of the banning policy on streamers of different languages. This might be interesting because the banning policy only targeted websites that were unlicensed in the U.S, whereas the main language of all 4 websites are English. Moreover, English is not the dominant language in gambling livestreams on Twitch.[13] Therefore, if the banning policy on English-based websites also decreased gambling livestreams in other languages, its influence extended beyond its primary target, making it more effective in controlling risky content across the platform.

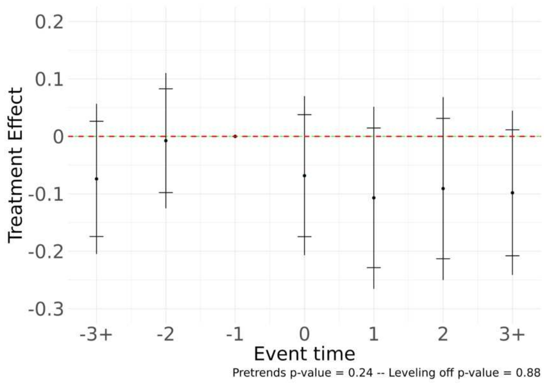
Figure 8 shows the treatment effects on gambling livestreams and total livestreams across the main language groups. We find evidence that although the banning policy only targeted

[13] Among the 475 gambling streamers in our dataset, 72 used English as the main language, 92 used Spanish, 111 used Portuguese, and 200 used other languages. Streamers tend to use the same language in both gambling and non-gambling streams.

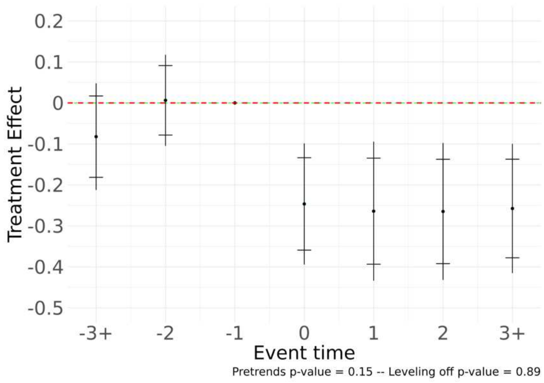
Figure 7: Time-Varying Treatment Effects on Content Creation in Group of Unbanned Streamers



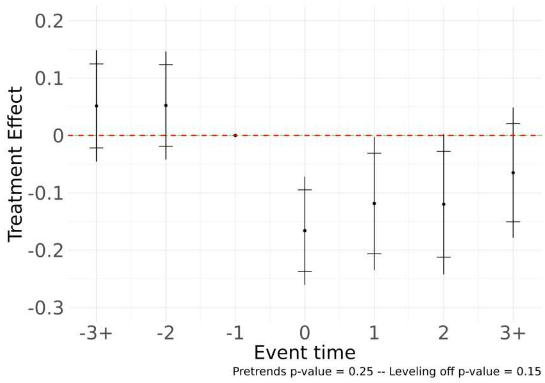
(a) Gambling Content



(b) LootBox Games



(c) Total Livestreams



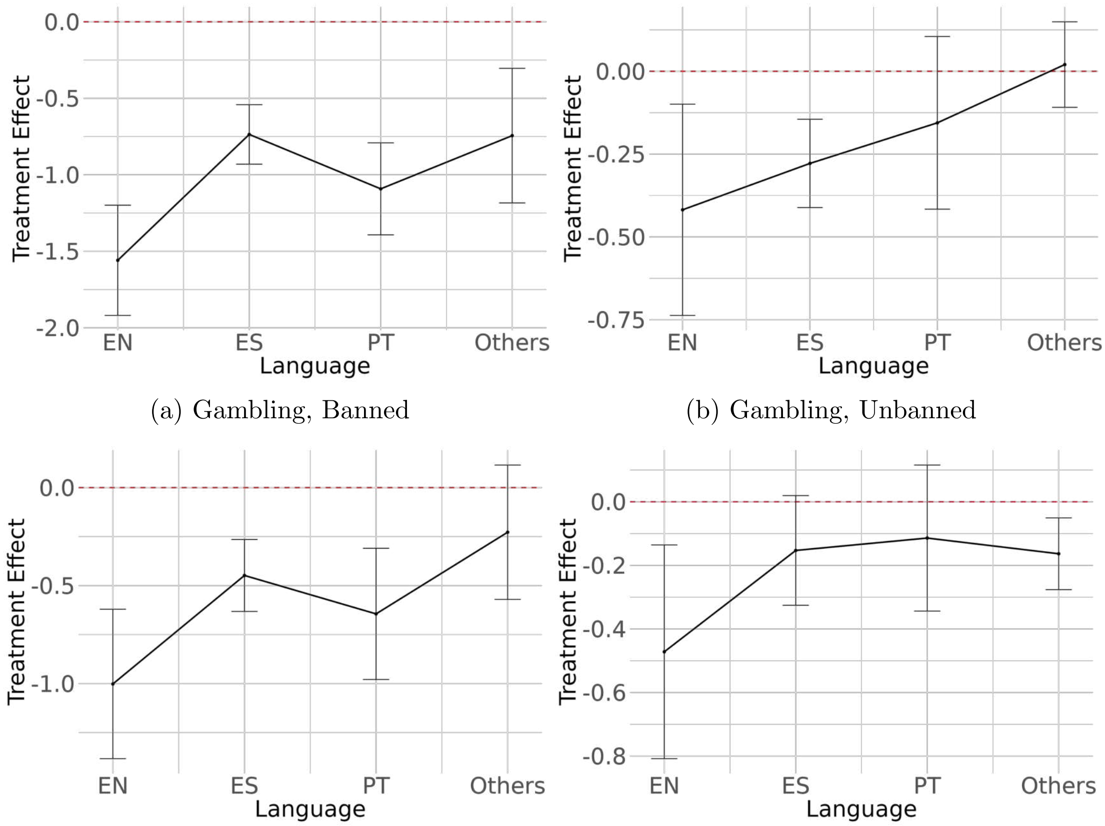
(d) Other Games

Notes. 1. The coefficients in each subfigure show the point estimates of β_3. 2. β_3 in one period ahead of the policy implementation is normalized to one. 3. The 95% point-wise confidence intervals are illustrated as the inner bars, and the sup-t confidence bands are illustrated as the vertical lines. 4. The p-values are derived from the test of pre-trends and the test of existence of dynamic effects of the policy. We omit more technical details and refer readers to (Freyaldenhoven et al., 2021).

English-based websites, it also had a spillover effect, as the supply of gambling livestreams in Spanish, Portuguese and other language channels were also significantly decreased with smaller magnitudes. However, the spillover effect on total livestreams is only prominent among banned streamers, as both Spanish streamers and Portuguese streamers did not decrease their totla livestreaming hours, compared to untreated streamers in the same language groups.

HTE on Streamer Popularity. Next, we explore whether the policy impact on content creation varies across streamers with different levels of popularity. From Figure 9, we find that in the banned group, streamers with higher popularity reduced their content creation of gambling livestreams slightly more than streamers with low popularity after the policy. More interestingly,

Figure 8: Heterogeneous Treatment Effects on Streamers with Different Main Languages



(a) Gambling, Banned

(b) Gambling, Unbanned

(c) Total Livestreams, Banned

(d) Total Livestreams, Unbanned

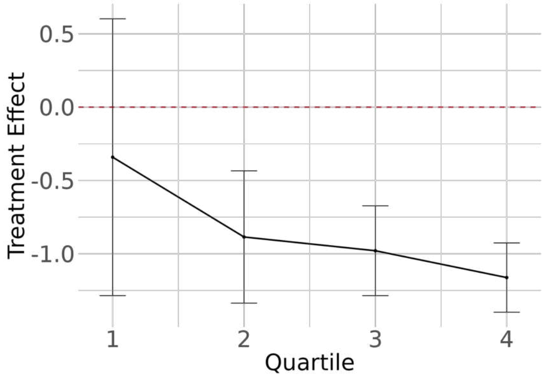
Notes. 1. The coefficients in each subfigure show the point estimates of β_1 for banned streamers and β_3 for unbanned streamers. 2. The language groups is derived based on the most frequently used language of each streamer, collected at stream-level from Twitch API.

only streamers with the highest level of popularity in the unbanned group significantly reduced their content creation of gambling livestreams, with a magnitude of approximately 34.8%. The policy impact on total livestreaming hours also appears to increase with streamer popularity: while streamers with low popularity almost did not respond to the policy, streamers with the highest popularity contributed most to the overall reduction in content creation of total gambling livestreams. Since streamers with higher popularity had higher average streaming hours on the platform, our result suggests that the policy led to a tremendous negative impact on the total content creation across the platform.

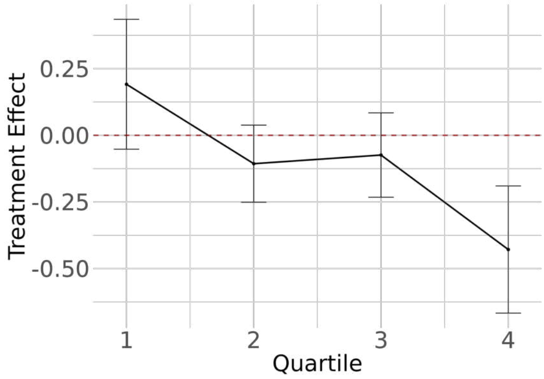
7.4 Mechanism

Our findings from heterogeneous treatment effects suggest that streamers with higher popularity were affected more by the policy. In this section, we investigate the mechanism behind the

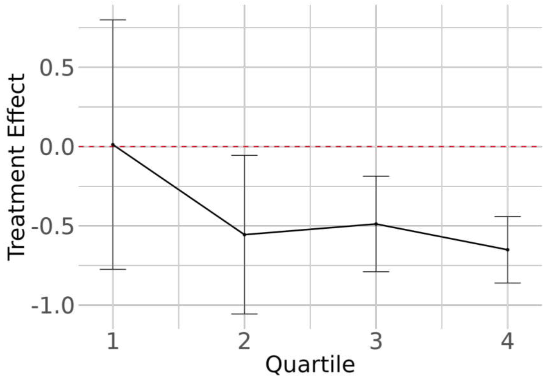
Figure 9: Heterogeneous Treatment Effects on Streamers with Different Popularity



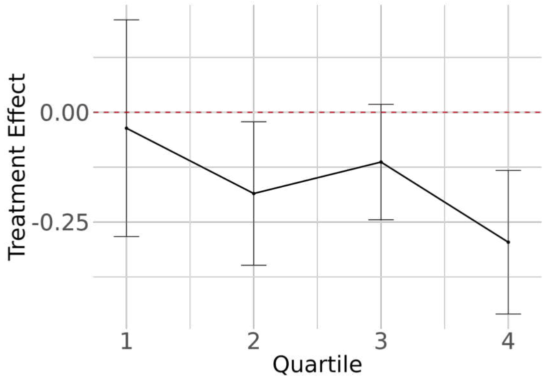
(a) Gambling, Banned



(b) Gambling, Unbanned



(c) Total Livestreams, Banned



(d) Total Livestreams, Unbanned

Notes. 1. The coefficients in each subfigure show the point estimates of β_1 for banned streamers and β_3 for unbanned streamers. 2. The quartiles are derived based on pre-policy average concurrent viewership.

observed heterogeneity, through two potential channels.

Reputation Concerns. First, streamers with higher popularity may care more about their reputation. On one hand, they stand to lose more revenue if prohibited by the platform due to their high number of subscriptions, in-stream donations and total hours watched. On the other hand, reducing gambling content and showing support to the platform's policy can enhance their reputation within the Twitch community. These factors may incentivize popular streamers to behave conservatively, resulting in a greater reduction of gambling streams in response to the policy.

If streamers value their reputation, they are more likely to carefully plan their streaming content to avoid temporary prohibitions from the platform. We test this hypothesis by running a DiD specification that includes interactions between treatment indicators and the number of account prohibitions each streamer received before the policy implementation. Column 1 and 2 in Table 7 present our estimation results. In addition to the negative policy impact on gambling streams, we find that both banned and unbanned streamers reduced gambling livestreams less if they cared less about their reputation (reflected as having more prohibitions before the policy implementation). Specifically, a banned streamer with no prior account prohibitions reduced gambling streams by 64.6%, and each additional prior prohibition resulted in a 2.8% smaller reduction compared to the baseline. Moreover, unbanned streamers are less sensitive to streamer reputation, with each additional prior prohibition resulting in a 6% smaller reduction, on top of a baseline decrease of -13.1%. We observe similar effects when we replace the number of times prohibited by the total number of days prohibited before the policy implementation.

Table 7: Underlying Mechanisms of Policy Impact on Gambling Streams

	× Banned Counts		× Share of Gambling Streams	
	Main Effect	Interaction Effect	Main Effect	Interaction Effect
Banned (β_1)	-1.038***	0.028***	-1.305***	1.610***
	(0.010)	(0.007)	(0.010)	(0.020)
Unbanned (β_3)	-0.141***	0.060***	-0.394***	1.200***
	(0.007)	(0.007)	(0.007)	(0.014)
Streamer FE	✓	✓	✓	✓
Week FE	✓	✓	✓	✓
Observations	112,222	112,222	112,222	112,222
Mean dependent variable	1.131	1.131	1.131	1.131

Notes. All standard errors are clustered at the streamer level. The mean dependent variable is calculated based on observations of all treated streamers before the policy announcement. FE, fixed effect.
Significance level: $*p < 0.1$; $**p < 0.05$;$***p < 0.01$

Reliance on Gambling Streams. Second, streamers with higher popularity may have relied less on gambling content to generate revenue. As a result, they faced lower risks when they chose to reduce their gambling livestreams. To test this hypothesis, we run a DiD specification that includes the interactions between treatment indicators and the share of gambling streams in total livestreams for each streamer. We find that streamers who relied more on gambling content reduced their gambling livestreams less, as indicated by the positive interaction effect in Table 7. Specifically, 1% increase in the share of gambling streams decreased the policy impact by 1.6% for a banned streamer and by 1.2% for an unbanned streamer. These estimates provide supportive evidence that the policy had a greater impact on those who relied less on gambling content, partially explaining the higher effectiveness on popular streamers.

7.5 Sensitivity Analysis

Although we have used multiple data sources and techniques to detect banned content and banned streamers, one may still concern that some streamers are not classified into the correct treated group due to missing data issues or measurement error from the data collection process, which we cannot control. Since we have chosen a conservative approach to detect banned streamers, our banned group of streamers can be viewed as a subset of streamers who actually streamed gambling content. Therefore, we focus on potential misclassification in the unbanned group, which may result in a potential bias in the estimated effects on unbanned streamers.

To address this concern, we conduct a sensitivity analysis to address potential misclassification issues in our casual estimates. The sensitivity parameter in our framework is the fraction of streamers who actually streamed banned content but were classified into the unbanned group, i.e.

$$\frac{\#\left\{\text{streamers who streamed banned content} \mid \text{Classified in the unbanned group}\right\}}{\#\left\{\text{streamers who streamed unbanned content} \mid \text{Classified in the unbanned group}\right\}}$$

We use this sensitivity parameter to test our estimation results from for the unbanned group of streamers. We find that the signs of all the estimated effects are robust as long as the sensitivity parameter is less than 0.999, i.e. less than half of all streamers in the unbanned group are misclassified. This condition is satisfied as we have more than half of unbanned streamers using a main language other than English, where the banned websites are all English-based. We provide a formal proof of identification with misclassification in the DiD framework and the validity of the sensitivity parameter in Appendix D.

8 Demand-Side Results

In this section, we present estimation results on key demand-side outcome variables. We first focus on two sets of demand-side treatment effects: the effect on total hours watched by viewers, one of the most crucial metrics of channel popularity and viewer engagement, and the effect on three tiers of subscriptions. Then, we discuss how we use network analysis to address potential violation of SUTVA assumption in the demand-side analysis.

8.1 Total Content Consumption and Revenue Channels

Table 8 presents estimated causal effects on total hours watched. The SynthDiD estimator yields an estimate of -1.696 for the banned streamers, suggesting that banned streamers experienced an 80.8% decrease in average weekly viewership, as a consequence of the 44.3% decrease in total streaming hours (Table 6). For the unbanned streamers, the SynthDiD estimator yields an estimate of -0.554, corresponding to a decrease of 40.2% in average weekly viewership, as a consequence of 17.6% decrease in total streaming hours. Although not completely comparable, the differences in magnitudes between demand- and supply-side estimates indicate that streamers suffered non-uniformly more in content consumption compared to their reduction in content creation, suggesting that their viewers before the policy implementation had higher preferences over gambling content.

Table 8: The Impact of Twitch's Ban Policy on Total Hours Watched

	log(Hours Watched + 1)		
	TWFE	SynthDiD	ATTGT
Banned (β_1)	-1.694***	-1.696***	-1.529***
	(0.268)	(0.241)	(0.317)
Δ%	-81.6%	-81.6%	-78.3%
Unbanned (β_3)	-0.523***	-0.554***	-0.659***
	(0.134)	(0.122)	(0.155)
Δ%	-40.7%	-42.5%	-48.3%
Streamer FE	✓	✓	✓
Week FE	✓	✓	✓
Observations	112,222	112,222	112,222
Mean dependent variable	8.487	8.487	8.487

Notes. All standard errors are clustered at the streamer level. The mean dependent variable is calculated based on observations of all streamers before the policy announcement. FE, fixed effect.
Significance level: $^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

While total hours watched reflects viewer engagement and content popularity, it does not directly indicate the revenue received by each streamer. Therefore, we also examine the policy impact on three tiers of subscriptions on Twitch. These outcome variables are informative for two reasons: first, the three tiers of subscriptions directly reflect changes in streamers' total revenue, as a Tier 1 subscription is worth $4.99, Tier 2 is worth $9.99 and Tier 3 is worth $24.99. Second, these subscriptions generally do not include exclusive livestreaming content[14].

[14] Viewers who purchase higher-tier subscriptions to a streamer enjoy benefits such as access to additional in-chat emotes, tenure-based chat badges, and subscriber-only chats.

Therefore, a higher-tier subscription can be seen as a signal of viewer loyalty and engagement in the streamer's community.

For clarity, we only show SynthDiD estimates for these outcome variables in Table 9, where we present full estimation results in Appendix C. We find that only Tier 1 subscriptions were significantly affected by the banning policy, with banned streamers experiencing a 44.2% loss in weekly Tier 1 subscriptions after the policy and unbanned streamers experiencing a loss of -16.9%, suggesting a huge loss in total revenue for the treated streamers due to the high number of Tier 1 subscriptions on average. In contrast, we find insignificant estimates of policy impact on both Tier 2 and Tier 3 subscriptions. Since viewers pay much more for higher tiers of subscriptions, only a small amount of loyal viewers are willing to support the streamer at these levels. This finding suggests that, on average, it was unlikely that either banned or unbanned streamers suffered significant losses in their loyal viewers or engagement from their core communities.

Table 9: The Impact of Twitch's Ban Policy on Different Tiers of Subscriptions

	Tier 1	Tier 2	Tier 3
Banned (β_1)	-0.584***	0.017	-0.024
	(0.110)	(0.011)	(0.014)
Δ%	-44.2%	-	-
Unbanned (β_3)	-0.185**	0.013	0.011
	(0.054)	(0.009)	(0.011)
Δ%	-16.9%	-	-
Streamer FE	✓	✓	✓
Week FE	✓	✓	✓
Observations	112,222	112,222	112,222
Mean dependent Variable	3.373	0.406	0.410

Notes. All standard errors are clustered at the streamer level. The mean dependent variable is calculated based on observations of all streamers before the policy announcement. FE, fixed effect.
Significance level: $^*p < 0.1$; $^{**}p < 0.05$;$^{***}p < 0.01$

8.2 Using Network Analysis to Address Potential Violation of SUTVA

A key assumption to identify the ATT is that the post-policy outcome of each streamer only depends on her own treatment assignment, commonly referred to as the SUTVA assumption (Rubin, 1980, 1990). In our context, if viewers focus more on certain types of content instead of particular streamers, the banning policy might induce viewers to switch to other streamers who provided the content of their interests. In that sense, the policy restricting the banned treated group from streaming unlicensed gambling content and thus their viewers for watching that content may also affect the remaining viewers, suggesting potential spillover between the demands across streamers both within and between treated and untreated groups.

To mitigate potential violations of SUTVA, we use network analysis on chat logs to evaluate the extent of overlapping viewership between streamers. Using this network, we employ *Louvain Community Detection*, a community detection algorithm, to identify distinct clusters, where streamers within the same cluster cluster share many viewers, and streamers across different clusters have minimal or no shared viewers. Finally, we use the *Breadth First Search (BFS) algorithm*, a graph traversing algorithm, to automate the selection of subsets of streamers within a cluster, and to construct treated and untreated groups for valid identification. We detail each procedure below.

Network Construction. We use the list of registered viewers in chats (a subset of the total viewers of a stream), prior to policy implementation, to construct a network of shared viewership among streamers.[15] To reduce the dimensionality of the network (which scales up with the number of streamers), we construct the network using all streamers from the two treated groups and a random sample of 800 streamers from the untreated group (out of 4626 untreated streamers). After excluding a few streamers with missing chats data, our final network consisting of 1270 streamers (155 banned, 315 unbanned, 800 untreated) with 805,815 edges.

In our network, a *node* represents a streamer, and an *edge* between two nodes indicates the presence of common viewers between the two streamers. The size of a node is proportional to the number of unique viewers of a streamer before the policy implementation, and the weight of an edge corresponds to the number of common viewers shared between two streamers. Table 10 shows the summary statistics of viewer counts of the streamers from the three groups.

Table 10: Viewer Count Before the Policy Implementation

Group	Mean	Std	5_{quantile}	25_{quantile}	50_{quantile}	75_{quantile}	95_{quantile}
Untreated	15,394.995	29,541.497	1,338.6	3,957.0	7,933.5	16,248.75	49,197.95
Unbanned	18,184.537	43,827.043	666.1	3,792.5	8,698.0	19,213.50	52,443.20
Banned	37,134.568	54,788.896	3,782.2	10,825.0	22,637.0	35,796.50	124,309.20

Notes: This table presents the descriptive statistics of viewer counts for different groups of streamers before the policy implementation.

[15] Since Twitch does not provide data on the full list of viewers within a stream, we use registered viewers in chats as a proxy for a streamer's total viewers in stream. The validity of this approach is discussed in Appendix A.

To identify streamers with significant viewer overlaps and enhance visibility and interpretability of the network , we filter out edges with fewer than 1,000 shared viewers. The threshold is selected heuristically based on the observed 5th quantile of unique viewers with a group, as shown in Table 10. The interactive version of our network can be accessed at interactive network.

Community Detection. To construct valid treated and untreated groups where streamers have minimal or no shared viewers, we use the *Louvain Community Detection* algorithm to identify distinct clusters within the network, whereas streamers in different clusters are not densely connected in terms of viewers.[16] Using the algorithm, we identify 52 community clusters from the network. Table E.0.1 in Appendix E shows the share of streamers with different treatment status (banned, unbanned, untreated) within each community cluster. Figure 10 shows an illustration of a cluster for each group.

To construct valid treated and untreated groups, we need to ensure that streamers with different treatment status are separated into distinct clusters, minimizing viewer overlap across groups. However, from Table E.0.1 and Figure 10, we observe that while it is straightforward to select clusters for the untreated group (many small clusters consist entirely of untreated streamers and are well separated from other groups), it is more challenging to select clusters for the banned and unbanned groups. Most clusters that contain many banned and unbanned streamers also contain streamers of other treatment status. For example, the cluster with predominantly unbanned streamers centered around "zloyn" (bottom left in Figure 10) also contains a large fraction of banned and untreated streamers.

To address the challenge of mixed-type clusters where most banned and unbanned streamers belong to, we first select a few mixed clusters where it is easy to disentangle streamers of the same status from others within the same cluster (see Table 11 of the selected clusters). Then, for each of these selected (but mixed) clusters, we identify the streamer with the most edges to other streamers of the same treatment status within a cluster (column 2). This central streamer becomes the focal point of the cluster. We then use the *Breadth-First Search (BFS) algorithm* to traverse the network and systematically remove the neighborhood of streamers of different status connected to the central streamer (column 6 shows the remained streamers

[16]The Louvain algorithm identifies clusters with high internal connectivity, by recursively merging communities into single nodes and then evaluating how much more densely connected the nodes within a community are, compared to how connected they would be in a random network. We also experimented the Girvan-Newman algorithm for community detection, but found Louvain to be more effective in unfolding communities for large networks.

within a cluster after this step). This approach allows us to automate the selection of subsets of streamers within a cluster, ensuring that the selected streamers do not have intense connections with streamers of other types within the same cluster. We detail our automated procedure next.

Table 11: Community Clusters By Treatment Status

Panel A: Communities for Banned Group				
Community ID	Center	Community Size	Streamers Kept	Banned Share (%)
15	jonvlogs	77	38	49.35
20	aker_	42	11	26.19
Others (1 cluster)	-	2	2	100.00

Panel B: Communities for Unbanned Group				
Community ID	Center	Community Size	Streamers Kept	Unbanned Share (%)
19	zloyn	115	82	64.35
Others (4 clusters)	-	9	9	100.00

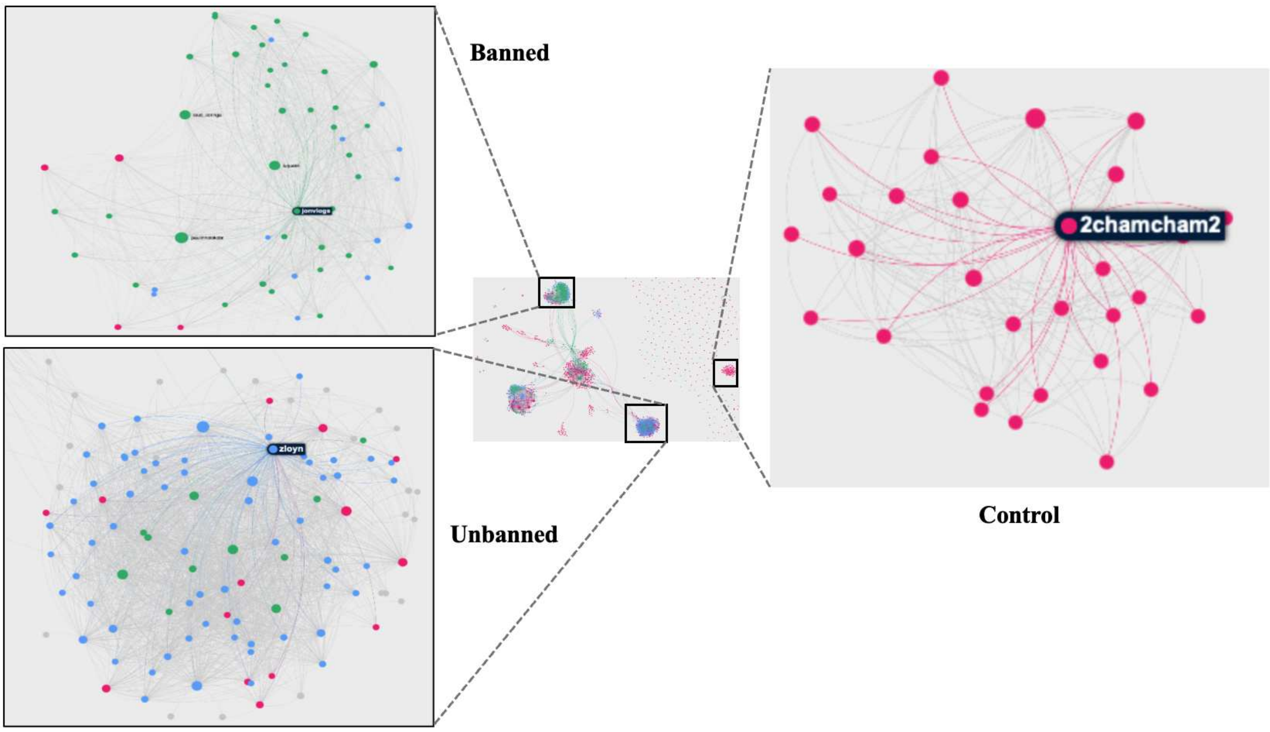
Panel C: Communities for Untreated Group				
Community ID	Center	Community Size	Streamers Kept	Untreated Share (%)
3	nyanners	209	98	75.12
Others (29 clusters)	-	183	183	100.00

Notes: This table presents the communities clusters used for each group. Panel A shows clusters used for banned streamers, Panel B shows clusters used for unbanend streamers, and Panel C shows clusters used for untreated streamers. The "others" communities in the column are communities that contain only a treatment status of only one type. Specifically, communities 1, 11,12,13, 16, 18, 21, 22, 23, 24, 26, 28, 30, 32, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 49 are solely of untreated streamers; 2,4,7, 51 are solely unbanned streamers; 31 is solely banned streamers.

BFS-Based Filtering Algorithm. We leverage Breadth-First Search (BFS), a graph traversing algorithm, to identify a subset of streamers who belong to the same treatment status within a cluster. The approach allows us to automate the selection process to include enough streamers of the same status, while effectively handling large complex networks consisting of clusters of mixed streamer status. We use this process to construct treated and untreated groups for valid identification, when treated streamers belong to clusters of mixed treatment status.

We illustrate the idea of this process in Figure 11. In **Step 1**, we begin with a central streamer within a cluster (for example, "jonvlogs" from community 15) and search the network outward from him, reaching to his closest streamers (Panel (a)). We declare his direct neighbors to be at distance 1, and exclude his direct streamers who have a different treatment status from

Figure 10: Network of Streamers Based on Common Viewers



Notes. This figure shows the network of streamers based on common viewers they share, before the policy implementation. The central panel shows the overall streamer network. A node denotes a streamer, and an edge indicates the presence of common viewers between two streamers. The size of a node is the number of viewers of a streamer, and the weight of an edge corresponds to the number of common viewers shared two streamers. The displayed network filters out edge weight that are below 1000 for visibility and identification of clusters. The nodes are colored such that red denotes the untreated group, blue denotes the unbanned (treated) group, and green denotes the banned (treated) group. The zoomed-in sections highlight specific clusters of the network which are predominantly banned (top left), unbanned (bottom left), and untreated (right) groups. The zoomed-in areas provide a clearer view of the connections and streamers' treatment status within the clusters.

him (Panel (b)). In **Step 2**, we find all neighbors of "jonvlogs"'s direct streamers (not counting streamers who are already direct neighbors of the central streamer) and declare those to be at distance 2. For example, "pluzinho" (green node) is one such streamer at distance 1 who are connected to other streamers of different status (blue node) in Panel (c). The BFS process then includes all direct neighbors of "pluzinho", again filtering out any streamers of different treatment status from him (Panel (d)). We continue the process *interactively* through the layers of the network. In our case, we stop at distance 3, which covers enough streamers within a cluster, while ensuring that the retained streamers is homogeneous in terms of treatment status.

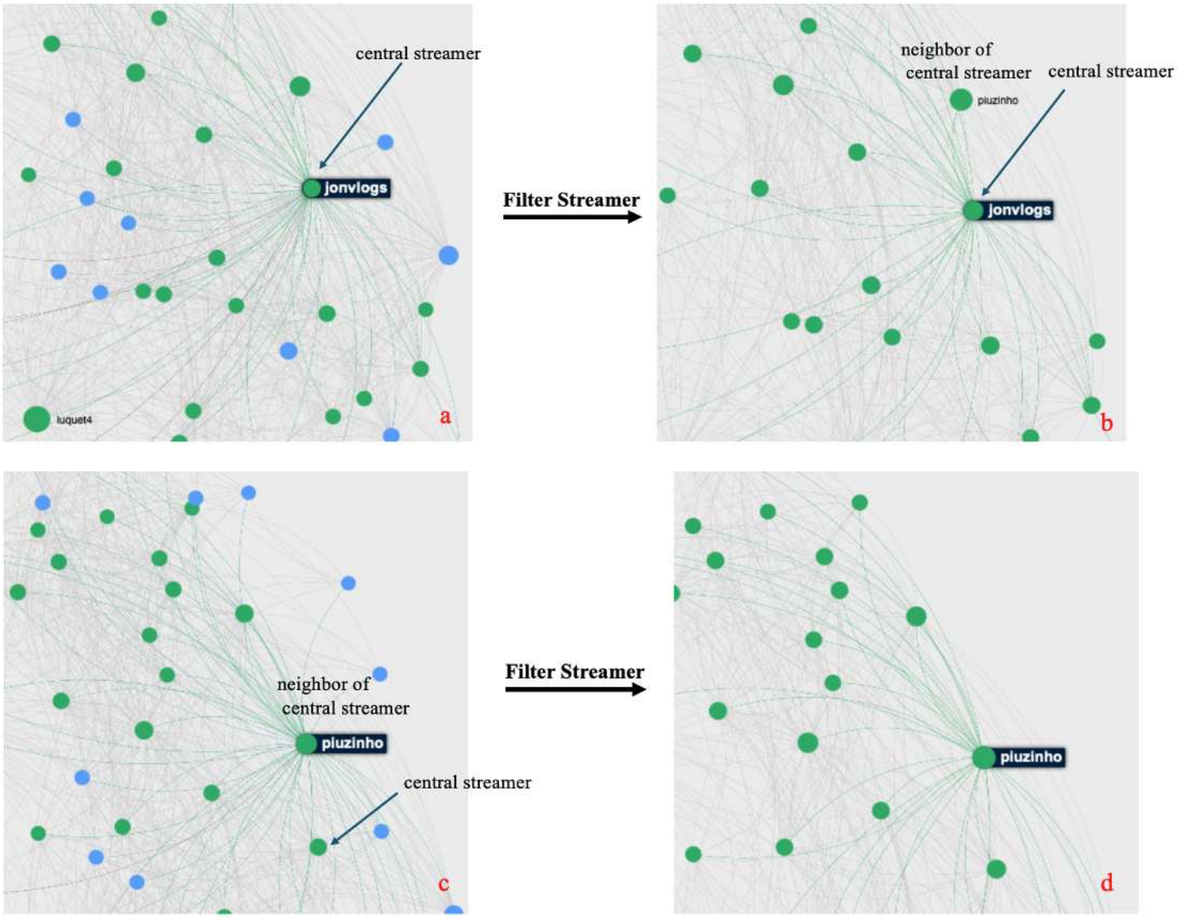
Let G be a graph with nodes (streamers) and edges. Let Q be the queue[17] for BFS traversal, and V be the set of visited streamers, T be the set of target nodes that meet the criteria (i.e. belong to the same treatment status and cluster). Algorithm 1 gives the full description of the procedure.

Algorithm 1 BFS-Based Filtering

```
Initialize graph G from the data and set the central streamer center_streamer
Initialize queue Q ← [center_streamer], visited set V ← {}, and target streamers set T ← {}
Set center_treatment ← G[center_streamer]['treatment'] and center_community ← G[center_streamer]['community']
while Q is not empty do
    streamer ← Q.pop(0)
    if streamer ∉ V then
        V.add(streamer)
        if G[streamer]['treatment'] == center_treatment and G[streamer]['community'] == center_community then
            T.add(streamer)
            for all neighbor ∈ G.neighbors(streamer) do
                if neighbor ∉ V then
                    Q.append(neighbor)
                end if
            end for
        end if
    end if
end while
return filtered streamers with homogeneous treatment status
```

[17]In BFS, the queue is used to keep track of which streamers (nodes) we need to visit next. We add streamers to the queue as we discover them, and we remove them from the queue as we visit and process them.

Figure 11: Filtering Streamers of Different Treatment Status Iteractively Using BFS



Notes. This figure illustrates the process of filtering streamers within a cluster using the Breadth-First Search (BFS) algorithm. Panel (a) shows the initial cluster with the central streamer "jonvlogs" (green node) and its neighbors, including streamers of different statuses (blue nodes). Panel (b) depicts the cluster after filtering out streamers of different statuses connected to the central streamer. Panel (c) demonstrates the next step of filtering by finding neighbors of the neighbors (e.g., neighbors of one such neighbor "piuzinho"). Panel (d) shows the final filtered cluster with only streamers of the same status connected to the central streamer "piuzinho". The process continues successively through the network.

8.3 Demand-Side Results on Selected Groups

In Table 12, we summarize causal estimates for the same outcome variables as in Table 8 and 9, and compare them with previously reported results. Since we have aggregated streamers with the same treatment status to avoid within-group spillover, it is infeasible to construct the synthetic counterfactual in the SynthDiD method, whereas the *did* package reports the error message that the treated group is too small to perform the doubly-robust estimator. Therefore, we report estimates from TWFE-DiD regressions.

Table 12: The Impact of Twitch's Ban Policy on Demand-Side Outcomes, Based on Selected Groups of Streamers

	log (Hours Watched + 1)	Tier 1	Tier 2	Tier 3
Banned	-2.023 (0.259)***	-0.619 (0.137)***	0.018 (0.027)	-0.020 (0.023)
Δ%	-86.8%	-46.1%	-	-
Unbanned	-0.802 (0.259)**	-0.256 (0.137)*	0.025 (0.027)	0.045 (0.023)*
Δ%	-55.2%	-22.6%	-	4.6%
Streamer FE	✓	✓	✓	✓
Week FE	✓	✓	✓	✓
Observations	8,694	8,694	8,694	8,694

Notes. All standard errors are clustered at the streamer level. FE, fixed effect.
Significance level: $*p < 0.1$; $**p < 0.05$;$***p < 0.01$

We find that all effects are qualitatively similar to previously reported estimates, except that the DiD estimator at the aggregated level indicates a slight increase in Tier 3 subscriptions for unbanned streamers at the marginal significance. However, this result does not change our conclusion that neither banned nor unbanned streamers suffered from losses in their core viewership, where both groups experienced significant loss in revenue from non-loyal viewers. Although we expect that the DiD estimates on all collected streamers will overestimate the ATT on the total hours watched, the estimates in Table 12 suggest that previous results indeed underestimate the negative impact on the affected streamers. One explanation to the results is that since the selected streamers share much fewer viewers with others on the platform, they are in general less popular and can be replaced by other streamers easily when they were forced not to change their streaming content, resulting in a higher effect on these streamers.

In Table C.0.3 of Appendix C, we present estimation results for the supply-side variables among the selected groups of streamers. This additional analysis addresses concerns regarding potential violations of SUTVA on the demand-side outcomes influencing streamers' decisions on content production. Overall, the findings corroborate those presented in Section 7. Notably, the streaming hours for gambling content among unbanned streamers in the selected group were unaffected by the policy. This outcome is expected as these streamers typically have lower popularity, making them less responsive to the banning policy, as we displayed in Figure 9.

9 Website Traffic Analysis

To examine whether Twitch's policy also has an effect on traffic to online gambling websites, we identify and extract a list of 3,753 websites from text detected via OCR on sampled video frames. Specifically, these websites are identified by matching patterns using regular expressions (regex) that start with optional protocols ("http://" or "https://") and include top-level domains, which could be either generic (e.g., ".com",".net") or country-specific (e.g., ".us",".uk"). We then manually checked through the website list to refine it down to a subset of 105 websites of online gambling or gambling-alike content. We define a site as an "online gambling site", if it directly offers any online slots, roulette or dice games, whereas users can register and play with real money. These websites can be broadly categorized into 5 different groups:

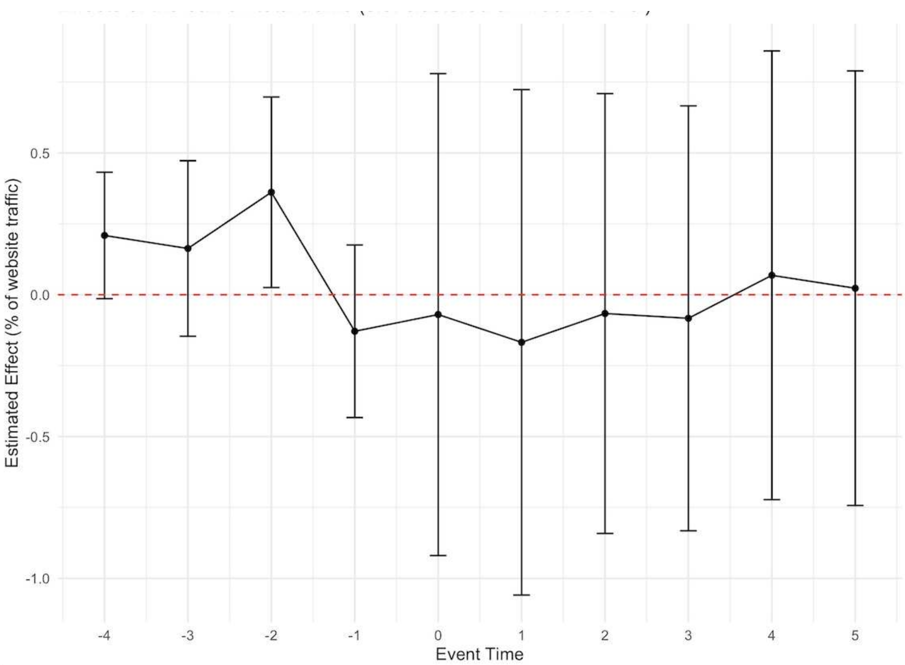
Table 13: Gambling Websites Used for Event Study of Website Traffic

Type	Example
Banned websites by the policy	stake.com
Unbanned gambling sites detected by OCR	1bet.com
Unbanned gambling sites not detected by OCR	mcluck.com
Unbanned websites of sports betting	1xbet.com
Gambling sites with different top-level domains	stake.mba

We obtained SimilarWeb's website traffic data from *Dewey*, which contains monthly total traffic on a domain and subdomain level with a breakdown by device type since September 2018. The dataset includes observations of desktop, mobile and overall visits, as well as average visit duration. We use the data to examine whether banned gambling websites have experienced sudden traffic changes due to the policy (see Section 9).

Figure 12 presents the time-varying treatment effect on the total traffic of banned gambling websites compared to the unbanned gambling websites. The estimate is conducted by including website and month fixed effects, and allowing for different linear trends per website to untreated for their differences in life cycle stages. Since the policy reduced promotions of the banned websites on Twitch, it might decrease the total traffic of banned websites and potentially increase the total traffic of unbanned website streamed on Twitch because they might be exposed rel-

Figure 12: banned Content Detection Using Videos



Notes. 1. This figure presents the time-varying treatment effects on the total traffic for banned gambling websites, using unbanned gambling websites described in Table 13 as the control group. 2. We include linear time trends for each website and cluster the standard errors at the website level.

atively more on the livestreaming platform. Therefore, even if the parallel trend assumption may fail to hold in the analysis, we should still expect a negative estimate if the policy had any impact on website traffic. However, the persistent null effect in Figure 12 indicates that the policy did not have any significant impact on traffic to the gambling websites.

10 Conclusion

In this paper, we empirically study the effects of the Twitch's banning policy on online gambling livestreams. To do so, we assemble a novel high-frequency panel dataset of livestreams from top Twitch influencers. We leverage video analysis, text analysis and a threshold-based classification approach to identify banned versus unbanned gambling content within streams, then partition streamers into groups which were directly or indirectly affected by the policy. Moreover, we employ multiple empirical approaches to overcome potential identification challenges in identifying the treatment effects on a variety of supply and demand-side outcome variables.

Our results show that the policy led to a persistent reduction in the supply of gambling livestreams for streamers reliant on gambling livestreams. However, the policy also reduced streams of non-gambling content, bringing unwanted side effect on content production to the platform. Moreover, we find that the policy had more prominent impact on streamers with higher popularity. We further examine the underlying mechanisms and show that both low reliance on gambling content and concern over personal reputation could contribute to the non-uniform effect among streamers. On content consumption, we find that the policy reduced both total hours watched and low-tier subscriptions for treated streamers. However, the policy did not reduce high-tier subscriptions, suggesting that these streamers did not suffered from losses in their core viewership.

Our research is subjected to several limitations. First, it is plausible that streamers who were significantly affected by the policy might migrate to other platforms to stream gambling content. However, we do not have access to data from other platforms supporting online gambling streams, such as YouTube or Kick. It would be interesting to examine potential spillover effects from Twitch to other livestreaming platforms due to the banning policy. Second, due to the lack of individual-level attributes of viewers and the low information density of chat logs, it is more difficult to evaluate the policy impact on gambling behavior, expenditures or mental health issues of viewers of gambling streams. With more information about individual viewers, studying the policy effect on these outcomes might be a fruitful future research area.

Our findings provide managerial insights for influencers, platform developers and policymakers. First, our results can assist streaming platforms in assessing the impact of content restriction policies and guide them in advising streamers on optimizing their content for increased profitability. Second, our results provide new evidence on the debate on the need for regulating video games featuring gambling-like content, which helps game developers improve their product design and evaluate welfare effects. Finally, our findings enable policymakers to better anticipate the outcomes of implementing restrictive regulations or laws on online gambling, facilitating more effective and informed policymaking in the future.

References

Abadie, A., Diamond, A. and Hainmueller, J. (2010), 'Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program', *Journal of the American statistical Association* **105**(490), 493–505.

Abadie, A. and Gardeazabal, J. (2003), 'The economic costs of conflict: A case study of the basque country', *American economic review* **93**(1), 113–132.

Amano, T. and Simonov, A. (2023), 'A welfare analysis of gambling in video games', *Available at SSRN 4355019* .

Ananthakrishnan, U. M., Li, B. and Smith, M. D. (2020), 'A tangled web: Should online review portals display fraudulent reviews?', *Information Systems Research* **31**(3), 950–971.

Aridor, G., Jiménez-Durán, R., Levy, R. and Song, L. (2024), 'The economics of social media'.

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W. and Wager, S. (2021), 'Synthetic difference-in-differences', *American Economic Review* **111**(12), 4088–4118.

Beck, B. B., Wuyts, S. and Jap, S. (2023), 'Guardians of trust: How review platforms can fight fakery and build consumer trust', *Journal of Marketing Research* p. 00222437231195576.

Beknazar-Yuzbashev, G., Jiménez Durán, R., McCrosky, J. and Stalinski, M. (2022), 'Toxic content and user engagement on social media: Evidence from a field experiment', *Available at SSRN 4307346* .

Berman, R. and Israeli, A. (2022), 'The value of descriptive analytics: Evidence from online retailers', *Marketing Science* **41**(6), 1074–1096.

Breiman, L. (2017), *Classification and regression trees*, Routledge.

Callaway, B. and Sant'Anna, P. H. (2021), 'Difference-in-differences with multiple time periods', *Journal of econometrics* **225**(2), 200–230.

Clarke, D., Pailañir, D., Athey, S. and Imbens, G. (2023), 'Synthetic difference in differences estimation', *arXiv preprint arXiv:2301.11859* .

Cusumano, M. A., Gawer, A. and Yoffie, D. B. (2021), 'Can self-regulation save digital platforms?', *Industrial and Corporate Change* **30**(5), 1259–1285.

Freyaldenhoven, S., Hansen, C., Pérez, J. P. and Shapiro, J. M. (2021), Visualization, identification, and estimation in the linear panel event-study design, Technical report, National Bureau of Economic Research.

GameRant (2022), https://gamerant.com/twitch-gambling-ban-ineffective-too-little-too-late/?newsletter_popup=1.

Gandhi, A. and Hollenbeck, B. (2023), Misinformation and mistrust: The equilibrium effects of fake reviews on amazon. com, Technical report, Technical Report, UCLA Anderson School.

Howard, J. W. (2019), 'Free speech and hate speech', *Annual Review of Political Science* **22**(1), 93–109.

Huang, Y. and Morozov, I. (2023), 'The promotional effects of live streams by twitch influencers', *Available at SSRN 4582182* .

Johnson, G., Lin, T., Cooper, J. C. and Zhong, L. (2024), 'Coppacalypse? the youtube settlement's impact on kids content', *Available at SSRN 4430334* .

Kircher, T. and Foerderer, J. (2024), 'Ban targeted advertising? an empirical investigation of the consequences for app development', *Management Science* **70**(2), 1070–1092.

Montiel Olea, J. L. and Plagborg-Møller, M. (2019), 'Simultaneous confidence bands: Theory, implementation, and an application to svars', *Journal of Applied Econometrics* **34**(1), 1–17.

Rubin, D. B. (1980), 'Randomization analysis of experimental data: The fisher randomization test comment', *Journal of the American statistical association* **75**(371), 591–593.

Rubin, D. B. (1990), 'Comment: Neyman (1923) and causal inference in experiments and observational studies', *Statistical Science* **5**(4), 472–480.

StreamHatchet (2022b), https://streamhatchet.com/blog/blog-online-gambling-rises-on-twitch/.

StreamScheme (2023), https://www.streamscheme.com/twitch-statistics/.

Yang, J. and Simonov, A. (2024), 'Creator content production decisions on twitch.tv', *Work in Progress* .

Zhang, W., Liu, Z., Liu, X. and Muller, E. (2023), 'Doubling revenues by adopting livestream shopping: A synthetic did approach', *Available at SSRN 4318978* .

Appendix

Table of Contents

A Streamer Data Collection **49**

B Banned Content and Streamer Detection **50**

B.1 Matching Video Clips to Original Streams 50

B.2 Alternative Detection of Banned Content From Stream Titles 52

B.3 Ground Truth Sample Justification 52

B.4 Inference and Prediction Using Chats 53

C Additional Estimation Results **54**

D Technical Details of Sensitivity Analysis on Misclassification **58**

D.1 Details of the Sensitivity Analysis 58

D.2 Robustness of Main Results 60

D.3 Proofs for Appendix D 61

E Network Analysis **63**

A Streamer Data Collection

In our main data, we utilized a subset of high-frequency streaming activity data collected by Yang and Simonov (2024). This data tracks the streaming activities of 30,000 Twitch streamers (and individual registered viewers in streams), by sending requests to the Twitch API every 15 minutes from October 26th, 2022 to August 20th, 2023. Every 15 minutes, the authors retrieved from the API the stream information (e.g., started time, streamed game, viewer count, registered viewer list) of all tracked streamers. Table A.0.1 describes variables retrieved from the API at the time of each request.

Table A.0.1: Variable definitions

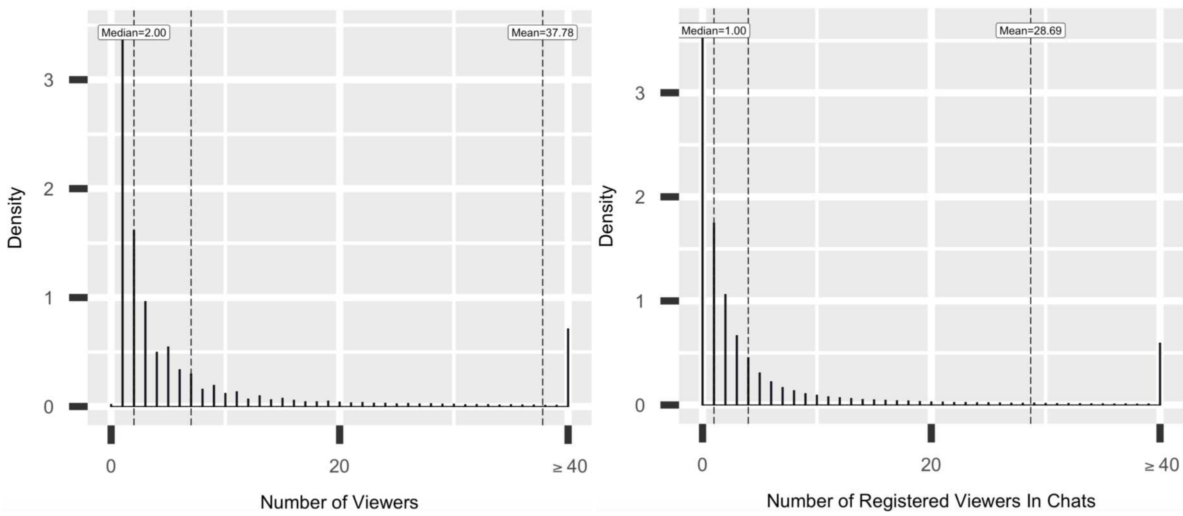
Variable	**Description**	**Object**
id	ID that identifies the stream	String
user_id	ID of the streamer	String
user_login	The login name of the streamer	String
game_login	ID of the game or category being streamed	String
type	The type of stream (e.g., live)	String
language	The language that the stream uses	String
title	The title of the stream	String
started_at	The UTC date and time of when the stream begins	Datetime object
viewer_count	Number of users watching the stream	Integer
registered_viewer_list [18]	The list of registered users in chats	List of Strings
registered_viewer_count	Number of registered users in chats	Integer
thumbnail_url	The URL to an image of a frame from the last 5 minutes of the stream	String
is_mature	True if the stream is for mature audiences	Boolean
tags	The tags applied to the stream	String
timestamp	The ET date and time of when the data is requested from the API	Datetime object

The streamers are pre-selected from a list of 3,606,607 streamers who appear in a 1-month pilot study (September 21st, 2022 and October 18th, 2022). During the pilot study, the authors sent requests to the Twitch API, at an hourly rate, to retrieve the top 100,000 most-viewed streamers that were live on Twitch at the time of the request. In total, the pilot study covers activities of 3,606,607 unique streamers broadcasting 34,704 types of content (e.g., both games and non-games such as "Just Chatting", "Pools, Hot Tubs, and Beaches", "Sports"). From this list, the authors then selected a subset of 30,000 streamers to track using weighted sampling, with weights proportional to the average viewership over the pilot study. The sampling approach ensures that the tracked streamers are representative of streamers on the Twitch platform (i.e., covering both superstars and lesser-known streamers). The sample size was selected by taking into account: (1) the daily quota for Twitch API requests, (2) the need to balance data collection efficiency and coverage (since IDs of registered viewers in streams were also tracked).

[18] Registered viewers in chats were collected from the Twitch TMI API. The data collection on registered viewers stopped on March 31st, 2023, after which Twitch permanently shut down the third-party Legacy Chatters endpoint on April 3, 2023.

Registered viewers in chats vs all viewers. Since Twitch API does not provide the list of viewers, we use registered viewers in chats to proxy for all viewers in a stream. This enables us to use individuals' watching history to construct valid treated and untreated groups to study the policy's effect on the demand side (See Section **??** for details). Table A.0.1 shows the distribution of registered viewers in chats compared to the total viewers. We show that the distribution of registered viewers in chats is generally representative of the total viewership in stream, evidenced by similar distributions and medians (except that viewer count has a higher mean, as registered viewers in chats is a subset of viewers in streams) and high correlation (correlation = 0.988).

Figure A.0.1: Distribution of Registered Viewers vs Viewers



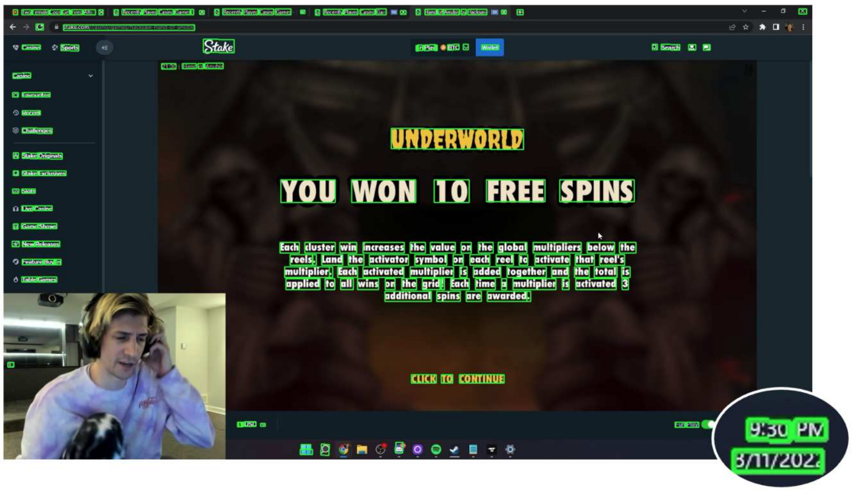
B Banned Content and Streamer Detection

B.1 Matching Video Clips to Original Streams

To match video clips to their original streams, we need to know the exact start and end times of a clip. However, (1) not all clips have a video offset time from the Twitch API, which indicates how many seconds into a video (and thus stream) the start of the clip occurred, and (2) a clip's created time from the Twitch API can sometimes reflect its upload time, but not the time of when the actual content in the clip was broadcasted. This can be an issue if a lag between the timestamps is severe.

For clips that have such issues, we employ several strategies to correctly matching them to their streams. We start by filtering clips with the streamer's desktop time displayed during the broadcast (see Figure B.1.1). A streamer's desktop time represents the actual broadcast time at his or her local time. OCR can extract these timestamps (usually containing only the date, hour and minute components of a timestamp, no seconds) as part of the texts. However, observing streamers' actual broadcast times from video clips is also not enough for matching, with the reasons as follows. First, sometimes multiple timestamps (including the streamer's desktop time) can be detected by OCR from a video frame. This happens when a video also displays chats or other information that also come with timestamps. Second, streamers may operate at different time zones, which can alter both the date and the hour component of a timestamp. Third, streamers may have their system time displayed in different formats (e.g., DD/MM/YYYY vs MM/DD/YYYY, or 09:20 PM vs 21:20), which creates confusions in matching to their streams' timestamps.

Figure B.1.1: Desktop Time During A Broadcast



Notes. This figure illustrates that OCR can detect a streamer's desktop time from video frames (if this information is displayed during the broadcast). In this example, it shows that XQC was broadcasting Stake at 9:30pm on August 11, 2022 at his local time.

We combine several different methods to address above the issues. First, we note that a streamer's desktop time always appears in an OCR's extracted texts as the last timestamp. This is because almost all gambling streamers use windows systems, where the desktop time is always displayed in the lower right corner. OCR detects text in reading order, and so regardless of whether the frame is read from left to right or top to bottom, the timestamp always appears close to the end of the extracted texts. Next, we check whether a clip's upload time corresponds to actual broadcast time as reflected by a streamer's system time detected by OCR, using the following 2-step procedure. In **step 1**, we start by matching a streamer's system time with the clip's upload time using only the *minute* component of the two timestamps. We allow a discrepancy of up to 3 minutes between the two timestamps, considering the maximum duration of a clip and minor time differences due to the transition of a minute (since desktop time does not show the seconds component of a timestamp, 9:30 PM can either be the start of 9:30 PM or close to 9:31PM). Our rationale is that it is highly unlikely for the minute component of the timestamps to randomly fall within the same 3-minute range. In **step 2**, we verify the *date and hour* components of the two timestamps. Twitch records timestamps in UTC. Given that there are 24 time zones, any local time can be up to 12 hours (in rare cases, 14 hours) ahead or behind UTC. We retain only clips where the hour difference of the two timestamps does not exceed 12 hours and, in addition, the clips fall within a streamer's local time zone. We use a combined approach to determine a streamer's local time zone. First, we obtain information about a streamer's country and city of residence from *Streams Charts*. Additionally, we identify the most common hour difference relative to UTC across all available clips of a streamer and use it to infer about the time zone the streamer is located. The remaining clips with clear start and end times are matched to their original streams for analysis.

B.2 Alternative Detection of Banned Content From Stream Titles

Table B.2.1: Detecting Banned Content From Stream Title

	Streamers with Gambling Content	Streamers without Gambling Content
No. Streamers	475	4626
Panel A: Using banned website referral links		
Banned Referrals in Any Stream Titles	52	2
Banned Referrals in Gambling Stream Titles	50	0
Banned Referrals in Any Stream Titles In Post-Treatment Period	2	0
Panel B: Using banned website keywords		
Banned Keywords in Any Stream Titles	90	4
Banned Keywords in Gambling Stream Titles	85	0
Banned Keywords in Any Stream Titles In Post-Treatment Period	4	1

Notes: This table shows the detection of banned streamers by various source types for both the treated and untreated groups.

B.3 Ground Truth Sample Justification

We note that OCR with no banned website keywords detected at all 3 frames is a **necessary but not sufficient condition** to make sure that a clip shows unbanned gambling content. For example, a clip can have content from an banned website, but OCR cannot detect it if a clip does not feature banned website logo or keyword in frames. (However, if OCR detects banned content, it means that a streamer indeed streams banned content.) Thus, we need to use multiple criteria to generate the ground truth unbanned gambling sample as indicated above.

In addition, we checked the (stream) titles for all gambling clips with no banned website keyword detected at all 3 frames of the clips (see Table B.4.1). We find (1) only 1561/161584 = 0.9% clips could be potentially "misclassified" if we just using 1 criteria (clips with no referral link in stream title); (2) almost no clips (1/161584) "misclassified" if we use 2 criteria (clips with no referral link in title and the clips are all from streamers with no history of streaming banned content based on all available clips pre-policy). To enforce an even stricter restriction, when we generate the ground truth sample, we allow neither banned referral link nor banned keyword to appear in stream title. The multiple criteria we impose ensure a ground truth banned sample (even if there is misclassification, we justify that the probability is extremely low).

Table B.3.1: Detecting Banned Content From Stream Title

	Video Clips
Total Number	475
Misclassified unbanned Title (OCR)	2
Misclassified unbanned Title (OCR + Streamer History)	2

Notes: This table shows the detection of banned streamers by various source types for both the treated and untreated groups.

B.4 Inference and Prediction Using Chats

Table B.4.1 summarizes the in-sample and out-of-sample performance.[19]

Table B.4.1: Detecting Banned Content From Stream Title

(a) Panel A: In-Sample Performance

Measure	Best Threshold	In-Sample Results Across Folds
Total Error	0.30	0.001
Weighted Total Error	0.30	0.003
Accuracy	0.30	0.976
AUC	0.30	0.976
Precision	1.05	0.991
Recall	0.00	0.978
F1-Score	0.25	0.976

(b) Panel B: Out-Of-Sample Performance

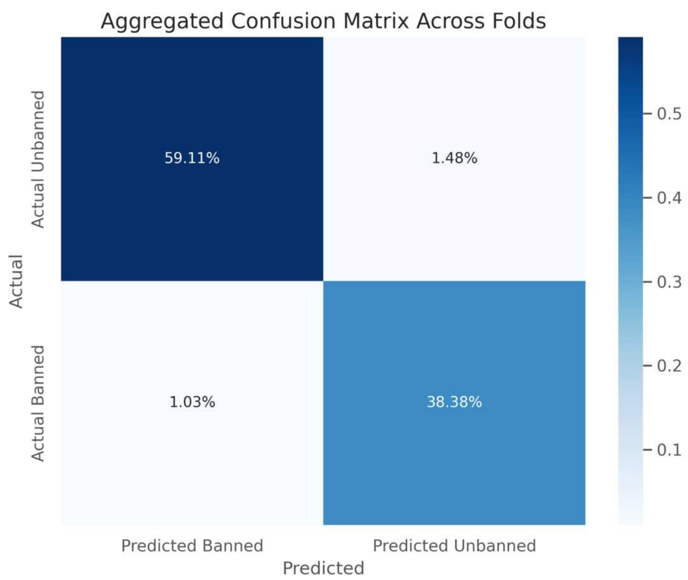
Measure	Averaged Out-Of-Sample Results Across Folds
Total Error	0.002
Weighted Total Error	0.003
Type I Error	0.003
Type II Error	0.024
Accuracy	0.871
AUC	0.975
Precision	0.964
Recall	0.974
F1-Score	0.853

[19] Total Error=Type I^2 + Type II^2, Total Error Weighted=$2 \times$ Type I^2 + Type II^2, Precision=$\frac{\text{True Positive}}{\text{Predicted Positive}}$, Recall=$\frac{\text{True Positive}}{\text{Actual Positive}}$, F1=$\frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}$.

Table B.4.2: Detecting Banned Content From Stream Title

Measure	Averaged Out-Of-Sample Results Across Folds
Total Error	0.002
Weighted Total Error	0.003
Type I Error	0.003
Type II Error	0.024
Accuracy	0.871
AUC	0.975
Precision	0.964
Recall	0.974
F1-Score	0.853

Figure B.4.1: Aggregated Confusion Matrix Across Folds



Notes. This figure shows aggregated confusion matrices across folds.

C Additional Estimation Results

In this section, we present full estimation results of supply-side and demand-side outcome variables.

Table C.0.1: Full Estimation Results of Supply-Side Outcomes

(a) Estimates for Gambling and LootBox Games

	log(Gambling hours + 1)			log(LootBox Games+ 1)		
	TWFE	SynthDiD	ATTGT	TWFE	SynthDiD	ATTGT
β_1	-1.022***	-1.001***	-1.023***	0.012	0.021	0.089
	(0.089)	(0.082)	(0.102)	(0.065)	(0.054)	(0.076)
β_2	0.166*	0.079	-0.013	-0.004	0.011	0.111
	(0.074)	(0.070)	(0.070)	(0.054)	(0.050)	(0.073)
β_3	-0.116*	-0.130***	-0.242***	-0.046	-0.054	-0.096
	(0.049)	(0.043)	(0.049)	(0.043)	(0.036)	(0.053)
β_4	0.133***	0.099	0.051	-0.026	-0.009	0.063
	(0.039)	(0.037)	(0.040)	(0.039)	(0.036)	(0.046)
Streamer FE	✓	✓	✓	✓	✓	✓
Week FE	✓	✓	✓	✓	✓	✓
Observations	112,222	112,222	112,222	112,222	112,222	112,222
Mean dependent variable	1.131	1.131	1.131	1.487	1.487	1.487

(b) Estimates for Streaming Hours and Other Games

	log(Streaming hours + 1)			log(Other Games+ 1)		
	TWFE	SynthDiD	ATTGT	TWFE	SynthDiD	ATTGT
β_1	-0.583***	-0.585***	0.543***	-0.146***	-0.149***	-0.172***
	(0.084)	(0.075)	(0.096)	(0.04)	(0.032)	(0.050)
β_2	-0.038	-0.041	0.029	-0.13***	-0.098***	-0.018
	(0.057)	(0.052)	(0.071)	(0.036)	(0.033)	(0.034)
β_3	-0.182***	-0.194***	-0.242***	-0.133***	-0.130***	-0.075**
	(0.044)	(0.040)	(0.050)	(0.032)	(0.026)	(0.036)
β_4	-0.065	-0.062	-0.015	-0.156***	-0.141***	-0.105***
	(0.039)	(0.037)	(0.051)	(0.028)	(0.026)	(0.034)
Streamer FE	✓	✓	✓	✓	✓	✓
Week FE	✓	✓	✓	✓	✓	✓
Observations	112,222	112,222	112,222	112,222	112,222	112,222
Mean dependent variable	2.587	2.587	2.587	1.079	1.079	1.079

Table C.0.2: Full Estimation Results of Demand-Side Outcomes

(a) Estimates for Hours Watched and Tier 1 Subscriptions

	log(Hours Watched + 1)			Tier 1		
	TWFE	SynthDiD	ATTGT	TWFE	SynthDiD	ATTGT
β_1	-1.694***	-1.696***	-1.529***	-0.59***	-0.584***	-0.507***
	(0.268)	(0.241)	(0.317)	(0.121)	(0.110)	(0.141)
β_2	-0.268	-0.272	-0.008	0.046	0.040	0.189
	(0.173)	(0.160)	(0.210)	(0.087)	(0.078)	(0.105)
β_3	-0.523***	-0.554***	-0.659***	-0.178**	-0.185***	-0.191**
	(0.134)	(0.122)	(0.155)	(0.058)	(0.054)	(0.074)
β_4	-0.218	-0.223	-0.115	-0.181***	-0.171***	-0.160**
	(0.117)	(0.113)	(0.147)	(0.053)	(0.046)	(0.064)
Streamer FE	✓	✓	✓	✓	✓	✓
Week FE	✓	✓	✓	✓	✓	✓
Observations	112,222	112,222	112,222	112,222	112,222	112,222
Mean dependent variable	8.487	8.487	8.487	3.373	3.373	3.373

(b) Estimates for Tier 2 and Tier 3 Subscriptions

	Tier 2			Tier 3		
	TWFE	SynthDiD	ATTGT	TWFE	SynthDiD	ATTGT
β_1	0.01	0.017	0.041	-0.024	-0.024	0.018
	(0.013)	(0.011)	(0.023)	(0.015)	(0.014)	(0.030)
β_2	0.004	0.010	-0.001	-0.018	-0.015	0.012
	(0.013)	(0.013)	(0.024)	(0.012)	(0.012)	(0.027)
β_3	0.012	0.013	0.012	0.013	0.011	0.055**
	(0.011)	(0.009)	(0.021)	(0.012)	(0.011)	(0.017)
β_4	0.023**	0.027**	0.030*	0.003	0.003	0.049**
	(0.009)	(0.009)	(0.016)	(0.008)	(0.009)	(0.015)
Streamer FE	✓	✓	✓	✓	✓	✓
Week FE	✓	✓	✓	✓	✓	✓
Observations	112,222	112,222	112,222	112,222	112,222	112,222
Mean dependent variable	0.406	0.406	0.406	0.410	0.410	0.410

Table C.0.3: Full Estimation Results of Isolated Streamers

(a) Supply-Side Outcomes

	Gambling	LootBox	Streaming Hours	Other Games
β_1	-1.332***	-0.068	-0.731***	-0.168*
	(0.094)	(0.078)	(0.082)	(0.064)
β_2	0.028	0.057	-0.028	-0.262**
	(0.121)	(0.102)	(0.106)	(0.082)
β_3	-0.056	0.037	-0.224**	-0.168*
	(0.094)	(0.078)	(0.082)	(0.064)
β_4	0.221	0.011	-0.077	-0.307***
	(0.121)	(0.102)	(0.106)	(0.082)
Streamer FE	✓	✓	✓	✓
Week FE	✓	✓	✓	✓
Observation	8,694	8,694	8,694	8,694

(b) Demand-Side Outcomes

	Hours Watched	Tier 1	Tier 2	Tier 3
β_1	-2.023***	-0.619***	0.018	-0.020
	(0.259)	(0.137)	(0.027)	(0.023)
β_2	0.027	0.309	0.010	-0.013
	(0.336)	(0.177)	(0.035)	(0.029)
β_3	-0.802**	-0.256*	0.025	0.045*
	(0.259)	(0.137)	(0.027)	(0.023)
β_4	0.349	-0.303	0.039	0.024
	(0.336)	(0.177)	(0.035)	(0.029)
Streamer FE	✓	✓	✓	✓
Week FE	✓	✓	✓	✓
Observation	8,694	8,694	8,694	8,694

D Technical Details of Sensitivity Analysis on Misclassification

D.1 Details of the Sensitivity Analysis

To show the validity of our sensitivity analysis, we first summarize our problem with a more general econometric framework. Denote the outcome variable as $Y_t(D_1^*, D_2^*)$, whereas $t \in \{0,1\}$ denotes the pre- and post-treatment period respectively[20], and D_1^*, D_2^* denote the binary treatments ("banned" group and "unbanned" group) received by each streamer. However, D_1^* and D_2^* are latent variables which are not observable, and we are only able to see the revealed treatment assignment D_1, D_2 from data, where D_1, D_2 may not be identical to D_1^*, D_2^* for all streamers. For example, if a streamer indeed received the first treatment, i.e. she should have been classified into the "banned" group, but were misclassified into the "unbanned" group in our dataset because we found no evidence of streaming banned content, then $D_1^* = D_2 = 1$ and $D_1 = D_2^* = 0$.

Using the above notations, we can rewrite our outcomes variable as follows:

$$Y_t = \begin{cases} Y_t(1,0)D_1^*(1-D_2^*) + Y_t(0,1)(1-D_1^*)D_2^* + Y_t(0,0)(1-D_1^*)(1-D_2^*) & , \text{if } t=1 \\ Y_t(0,0) & , \text{if } t=0 \end{cases} \tag{9}$$

In this framework, (Y_0, Y_1, D_1, D_2) are *observed* from the dataset, whereas (D_1^*, D_2^*) are latent variables. We do not impose any latent structures on how the true treatment assignments are mapped to observed treatment assignments. Instead, we impose identification restrictions and bounds on conditional probabilities of (mis)classifications in this framework.

As in all other DiD frameworks, the identification of ATT relies on imposing the parallel trend assumption. However, we require extra parallel trends because of potential misclassification. We state our assumptions need for identication as follows:

Assumption 1 (Parallel Trends) *.*

1. $\mathbb{E}[Y_1(0,0) - Y_0(0,0) \mid D_1^* = 1, D_2^* = 0] = \mathbb{E}[Y_1(0,0) - Y_0(0,0) \mid D_1^* = 0, D_2^* = 0]$

2. $\mathbb{E}[Y_1(0,0) - Y_0(0,0) \mid D_1^* = 1, D_2^* = 0, D_1 = 1, D_2 = 0] =$
$$\mathbb{E}[Y_1(0,0) - Y_0(0,0) \mid D_1^* = 1, D_2^* = 0, D_1 = 0, D_2 = 1],$$
$\mathbb{E}[Y_1(0,0) - Y_0(0,0) \mid D_1^* = 0, D_2^* = 1, D_1 = 1, D_2 = 0] =$
$$\mathbb{E}[Y_1(0,0) - Y_0(0,0) \mid D_1^* = 0, D_2^* = 1, D_1 = 0, D_2 = 1].$$

The first part of Assumption 1 is identical to the parallel trend assumption in standard DiD frameworks without misclassification. In addition, the second part of Assumption 1 posits that the outcome variables of both the correctly classified streamers and misclassified streamers should follow the same trend within each group on the same true treatment assignments D_1^*, D_2^*,

[20] For simplicity, we present all identification results based on a two-period setup. All results can be extended directly into a panel model with multiple pre- and post-treatment periods.

i.e. the expected outcome is mean independent of the treatment arm within each (true) treated group.

To show the relationship between our DiD estimates and the causal estimands of interest, we first state the following result which decomposes the DiD estimator under our setup:

Proposition 1

$$\begin{aligned}\beta_1 = & \mathbb{E}[Y_1(1,0) - Y_1(0,0) \mid D_1^* = 1, D_2^* = 0, D_1 = 0, D_2 = 0] P(D_1^* = 1, D_2^* = 0 \mid D_1 = 1, D_2 = 0) \\ & + \mathbb{E}[Y_1(0,1) - Y_1(0,0) \mid D_1^* = 0, D_2^* = 1, D_1 = 1, D_2 = 0] P(D_1^* = 0, D_2^* = 1 \mid D_1 = 1, D_2 = 0) \\ & - \mathbb{E}[Y_1(1,0) - Y_1(0,0) \mid D_1^* = 1, D_2^* = 0, D_1 = 0, D_2 = 0] P(D_1^* = 1, D_2^* = 0 \mid D_1 = 0, D_2 = 0) \\ & - \mathbb{E}[Y_1(0,1) - Y_1(0,0) \mid D_1^* = 0, D_2^* = 1, D_1 = 0, D_2 = 0] P(D_1^* = 0, D_2^* = 1 \mid D_1 = 0, D_2 = 0)\end{aligned}$$

and β_3 can be decomposed similarly.

As we have introduced in Section **??** of the paper, we have combined historical video clips, in-stream titles and chat logs to detect banned content as well as banned streamers in a conservative way. Therefore, we can rule out several possibilities of misclassification in our dataset. First, we have no misclassified streamers in the untreated group, as our untreated group only contains streamers with no record of streaming any gambling content in the ground-truth data extracted from Twitch API. Second, we have no misclassified streamers in the observed banned group. This is because all streamers in this group had shown sufficient evidence of streaming banned content. We summarize these restrictions as the following condition:

Condition 1 (Conditional Probabilities of Classifications)

$$\begin{aligned}P(D_1^* = 1, D_2^* = 0 \mid D_1 = 0, D_2 = 0) &= 0 \\ P(D_1^* = 0, D_2^* = 1 \mid D_1 = 0, D_2 = 0) &= 0 \\ P(D_1^* = 1, D_2^* = 0 \mid D_1 = 1, D_2 = 0) &= 1 \\ 0 \leq P(D_1^* = 1, D_2^* = 0 \mid D_1 = 0, D_2 = 1) &\leq 1\end{aligned}$$

Based on these conditions, we can prove that we have correctly identified ATT for the banned streamers, and our causal estimate of the unbanned streamers can be decomposed into a weighted average of the actual ATT for the two treated groups. We summarize these results in the following proposition:

Proposition 2 *Suppose that Assumption 1 and Condition 1 hold. Then,*

$$\begin{aligned}\beta_1 = & \mathbb{E}[Y_1(1,0) - Y_1(0,0) \mid D_1^* = 1, D_2^* = 0, D_1 = 1, D_2 = 0] \\ \beta_3 = & \mathbb{E}[Y_1(1,0) - Y_1(0,0) \mid D_1^* = 1, D_2^* = 0, D_1 = 0, D_2 = 1] P(D_1^* = 1, D_2^* = 0 \mid D_1 = 0, D_2 = 1) + \\ & \mathbb{E}[Y_1(0,1) - Y_1(0,0) \mid D_1^* = 0, D_2^* = 1, D_1 = 0, D_2 = 1] P(D_1^* = 0, D_2^* = 1 \mid D_1 = 0, D_2 = 1) \\ = & ATT_1 \times P(D_1^* = 1, D_2^* = 0 \mid D_1 = 0, D_2 = 1) + ATT_2 \times P(D_1^* = 0, D_2^* = 1 \mid D_1 = 0, D_2 = 1)\end{aligned} \tag{10}$$

The results stated in Proposition 10 is quite intuitive. On the one hand, since we do not have any misclassified streamer in the observed banned group, this group of treated streamers can be seen as a subpopulation of all streamers who indeed streamed banned content. Therefore, the second part of Assumption 1 ensures that the treatment effect is correctly identified *in the banned group*, as in any classic DiD setup. On the other hand, the observed "unbanned" group is a potential mixture of both "banned" and unbanned streamers, and the corresponding DiD estimate is therefore a weighted average of the two ATTs and is attenuated towards the direction of ATT for the "banned" group.

Since our DiD estimate for the banned group is correctly identified and is more negative compared to the DiD estimate of the unbanned group, suggesting that the actual ATT for the unbanned group might be nonnegative. Therefore, we adopt a sensitivity analysis based on the above identification results to address the concerns over potential misclassification. The sensitivity analysis is performed based on the following assumption:

Assumption 2 (Bounds on Probability of Misclassification)

$$\frac{1}{\Gamma} \le \frac{P(D_1^* = 1, D_2^* = 0 \mid D_1 = 0, D_2 = 1)}{P(D_1^* = 0, D_2^* = 1 \mid D_1 = 0, D_2 = 1)} \le \Gamma$$

We introduce sensitivity parameter Γ in Assumption 2, which bound the fraction between misclassified streamers in the observed unbanned group compared to the correctly classified streamers. It is natural to adopt this sensitivity parameter since the we can choose the value of the sensitivity parameter based on how confident we are on the classification for each streamer in the observed unbanned group. We derive the bounds of ATT as in the following proposition:

Proposition 3 *Under Assumption 1, 2, 3 and assume that (3) holds, then*

$$(1 + \frac{1}{\Gamma})\beta_3 - \Gamma\beta_1 \le ATT_2 \le (1 + \Gamma)\beta_{DiD,2} - \frac{1}{\Gamma}\beta_{DiD,1} \quad \text{if } \beta_d > 0$$
$$(1 + \Gamma)\beta_3 - \frac{1}{\Gamma}\beta_1 \le ATT_2 \le (1 + \frac{1}{\Gamma})\beta_3 - \Gamma\beta_1 \quad \text{if } \beta_d < 0$$

D.2 Robustness of Main Results

We can then test the sensitivity of our causal estimates in Section **??** based on the derived bounds. For example, the estimated treatment effects on the log weekly streaming hours of gambling contents are -1.001 and -0.130 respectively for the two treated groups. Plugging these estimates into Proposition 3, We get

$$0.130(2.001\frac{1}{\Gamma} - 1 - \Gamma) < ATT_2 < 0.130(2.001\Gamma - 1 - \frac{1}{\Gamma})$$

Based on the above result, we are confident with the result that the banning policy led to a reduction in the supply of gambling content of unbanned streamers as long as

$$0.130(2.001\Gamma - 1 - \frac{1}{\Gamma}) < 0 \Leftrightarrow \Gamma \in (0, 0.999)$$

i.e. the fraction of misclassified banned streamers in our observed unbanned group is less than approximately 50%. Similarly, the range of sensitivity parameters for the estimates in Table 6 are $(0, 2.427)$ and $(0, 1.465)$ respectively. Therefore, both of them are robust to misclassification as long as the fraction of misclassified banned streamers in our observed unbanned group is less than approximately 50%.

D.3 Proofs for Appendix D

D.3.1 Proof of Proposition 1

Note that

$$\begin{aligned}
\beta_{DiD,1} &= \mathbb{E}[Y_1 - Y_0 \mid D_1 = 1, D_2 = 0] - \mathbb{E}[Y_1 - Y_0 \mid D_1 = 0, D_2 = 0] \\
&= (\mathbb{E}[Y_1 \mid D_1 = 1, D_2 = 0] - \mathbb{E}[Y_1 \mid D_1 = 0, D_2 = 0]) \\
&- (\mathbb{E}[Y_0 \mid D_1 = 1, D_2 = 0] - \mathbb{E}[Y_0 \mid D_1 = 0, D_2 = 0])
\end{aligned}$$

Plugging the expression of Y_t from (9), we have

$$\begin{aligned}
&\mathbb{E}[Y_1 \mid D_1 = 1, D_2 = 0] - \mathbb{E}[Y_1 \mid D_1 = 0, D_2 = 0] \\
=& \mathbb{E}[Y_1(1,0)D_1^*(1-D_2^*) + Y_1(0,1)(1-D_1^*)D_2^* + Y_1(0,0)(1-D_1^*)(1-D_2^*) \mid D_1 = 1, D_2 = 0] \\
&- \ \mathbb{E}[Y_1(1,0)D_1^*(1-D_2^*) + Y_1(0,1)(1-D_1^*)D_2^* + Y_1(0,0)(1-D_1^*)(1-D_2^*) \mid D_1 = 0, D_2 = 0] \\
=& \mathbb{E}[(Y_1(1,0) - Y_1(0,0)D_1)D_1^*(1-D_2^*) + (Y_1(0,1) - Y_1(0,0))(1-D_1^*)D_2^* \\
&+ Y_1(0,0)(1 - D_1^*D_2^*) \mid D_1 = 1, D_2 = 0] \\
&- \mathbb{E}[(Y_1(1,0) - Y_1(0,0)D_1)D_1^*(1-D_2^*) + (Y_1(0,1) - Y_1(0,0))(1-D_1^*)D_2^* \\
&+ Y_1(0,0)(1 - D_1^*D_2^*) \mid D_1 = 0, D_2 = 0] \\
=& \mathbb{E}[Y_1(1,0) - Y_1(0,0) \mid D_1^* = 1, D_2^* = 0, D_1 = 0, D_2 = 0]P(D_1^* = 1, D_2^* = 0 \mid D_1 = 1, D_2 = 0) \\
&+ \ \mathbb{E}[Y_1(0,1) - Y_1(0,0) \mid D_1^* = 0, D_2^* = 1, D_1 = 0, D_2 = 0]P(D_1^* = 0, D_2^* = 1 \mid D_1 = 1, D_2 = 0) \\
&+ \ \mathbb{E}[Y_1(0,0) \mid D_1 = 0, D_2 = 0)] - \mathbb{E}[Y_1(0,0)D_1^*D_2^* \mid D_1 = 1, D_2 = 0] \\
&- \ \{\mathbb{E}[Y_1(1,0) - Y_1(0,0) \mid D_1^* = 1, D_2^* = 0, D_1 = 1, D_2 = 0]P(D_1^* = 1, D_2^* = 0 \mid D_1 = 0, D_2 = 0) \\
&+ \ \mathbb{E}[Y_1(0,1) - Y_1(0,0) \mid D_1^* = 0, D_2^* = 1, D_1 = 0, D_2 = 0]P(D_1^* = 0, D_2^* = 1 \mid D_1 = 0, D_2 = 0) \\
&+ \ \mathbb{E}[Y_1(0,0) \mid D_1 = 0, D_2 = 0)] - \mathbb{E}[Y_1(0,0)D_1^*D_2^* \mid D_1 = 1, D_2 = 0]\} \\
=& \mathbb{E}[Y_1(1,0) - Y_1(0,0) \mid D_1^* = 1, D_2^* = 0, D_1 = 0, D_2 = 0]P(D_1^* = 1, D_2^* = 0 \mid D_1 = 1, D_2 = 0) \\
&+ \ \mathbb{E}[Y_1(0,1) - Y_1(0,0) \mid D_1^* = 0, D_2^* = 1, D_1 = 1, D_2 = 0]P(D_1^* = 0, D_2^* = 1 \mid D_1 = 1, D_2 = 0) \\
&- \ \mathbb{E}[Y_1(1,0) - Y_1(0,0) \mid D_1^* = 1, D_2^* = 0, D_1 = 0, D_2 = 0]P(D_1^* = 1, D_2^* = 0 \mid D_1 = 0, D_2 = 0) \\
&- \ \mathbb{E}[Y_1(0,1) - Y_1(0,0) \mid D_1^* = 0, D_2^* = 1, D_1 = 0, D_2 = 0]P(D_1^* = 0, D_2^* = 1 \mid D_1 = 0, D_2 = 0)
\end{aligned} \tag{11}$$

whereas the last equality holds under Assumption 1 and the fact that

$$\mathbb{E}[Y_1(0,0)D_1^*D_2^* \mid D_1 = 1, D_2 = 0] \equiv 0$$

under Assumption 2. The decomposition of $\beta_{DiD,2}$ can be obtained similarly. $\square$

D.3.2 Proof of Proposition 2

The results are derived directly by plugging the bounds of conditional probabilities into results of Proposition 10. $\square$

D.3.3 Proof of Proposition 3

Note that with model (9), there will be no misclassification in the untreated group, i.e.

$$P(D_1^* = 1, D_2^* = 0 \mid D_1 = 0, D_2 = 1) + P(D_1^* = 0, D_2^* = 1 \mid D_1 = 0, D_2 = 1) = 1$$

Combining this equality with Assumption **??** yields the following bounds on the probability of correct classification conditional on actual second treatment:

$$\frac{1}{1+\Gamma} \leq P(D_1^* = 0, D_2^* = 1 \mid D_1 = 0, D_2 = 1) \leq \frac{1}{1+1/\Gamma} \tag{12}$$

Therefore,

$$\begin{aligned}
ATT_2 &= \frac{1}{P(D_1^* = 0, D_2^* = 1 \mid D_1 = 0, D_2 = 1)}\beta_{DiD,2} - ATT_1 \times \frac{P(D_1^* = 1, D_2^* = 0 \mid D_1 = 0, D_2 = 1)}{P(D_1^* = 0, D_2^* = 1 \mid D_1 = 0, D_2 = 1)} \\
&= \frac{1}{P(D_1^* = 0, D_2^* = 1 \mid D_1 = 0, D_2 = 1)}\beta_{DiD,2} - \beta_{DiD,1} \times \frac{P(D_1^* = 1, D_2^* = 0 \mid D_1 = 0, D_2 = 1)}{P(D_1^* = 0, D_2^* = 1 \mid D_1 = 0, D_2 = 1)} \\
&\leq \frac{1}{P(D_1^* = 0, D_2^* = 1 \mid D_1 = 0, D_2 = 1)}\beta_{DiD,2} - 1/\Gamma\beta_{DiD_1} \\
&\leq (1+\Gamma)\beta_{DiD,2} - 1/\Gamma\beta_{DiD,1}
\end{aligned}$$

when $\beta_{DiD,d} > 0$ for $d = 1, 2$. The lower bound of ATT_2 when $\beta_{DiD,d} > 0$ and the bounds when $\beta_{DiD,d} < 0$ can be obtained similarly. $\square$

E Network Analysis

Table E.0.1: Community Clusters and Treatment Status (Part 1)

Community ID	Community Size	Untreated Share (%)	Unbanned Share (%)	Banned Share (%)
0	3	0.00	66.67	33.33
1	6	100.00	0.00	0.00
2	2	0.00	100.00	0.00
3	209	75.12	9.09	15.79
4	2	0.00	100.00	0.00
5	2	50.00	50.00	0.00
6	4	25.00	75.00	0.00
7	3	0.00	100.00	0.00
8	2	50.00	50.00	0.00
9	2	50.00	50.00	0.00
10	122	33.61	38.52	27.87
11	4	100.00	0.00	0.00
12	46	100.00	0.00	0.00
13	13	100.00	0.00	0.00
14	8	25.00	75.00	0.00
15	77	7.79	42.86	49.35
16	3	100.00	0.00	0.00
17	39	33.33	43.59	23.08
18	17	100.00	0.00	0.00
19	115	25.22	64.35	10.43
20	42	73.81	0.00	26.19
21	2	100.00	0.00	0.00
22	20	100.00	0.00	0.00
23	5	100.00	0.00	0.00
24	2	100.00	0.00	0.00
25	2	50.00	0.00	50.00
26	3	100.00	0.00	0.00
27	16	93.75	0.00	6.25
28	2	100.00	0.00	0.00
29	39	89.74	2.56	7.69
30	19	100.00	0.00	0.00
31	2	0.00	0.00	100.00
32	2	100.00	0.00	0.00
33	4	25.00	50.00	25.00
34	2	100.00	0.00	0.00
35	4	100.00	0.00	0.00
36	2	100.00	0.00	0.00
37	2	100.00	0.00	0.00
38	2	100.00	0.00	0.00
39	2	100.00	0.00	0.00
40	2	100.00	0.00	0.00

Table E.0.2: Community Clusters and Treatment Status (Part 2)

Community ID	Community Size	Untreated Share (%)	Unbanned Share (%)	Banned Share (%)
41	2	100.00	0.00	0.00
42	7	100.00	0.00	0.00
43	3	100.00	0.00	0.00
44	3	100.00	0.00	0.00
45	2	100.00	0.00	0.00
46	2	100.00	0.00	0.00
47	2	100.00	0.00	0.00
48	2	50.00	50.00	0.00
49	2	100.00	0.00	0.00
50	19	42.11	57.89	0.00
51	2	0.00	100.00	0.00

Notes: This table presents the 52 communities detected by the Louvain Community Detection algorithm. For each community cluster, we show the number of streamers within the cluster and the share of streamers of different treatment status. Note that some clusters consist of streamers of one treatment status, while others are mixed. We discuss the streamers and communities used in our constructed treated and untreated groups for SUTVA alleviation in Section 8.2.