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

Qifan Han* Jasmine Yang† Andrey Simonov‡

February 27, 2025

Abstract

How effective is platform self-regulation at eliminating harmful content? We examine Twitch's ban on unlicensed online gambling livestreams implemented in October 2022. Using a novel panel dataset covering the top 6,000 Twitch streamers, we identify banned content and affected streamers by leveraging video analysis of historical clips, high-frequency stream titles, and in-stream chat data. To address key identification challenges, we apply both a two-way fixed effects difference-in-differences (DiD) estimator and a Synthetic DiD approach, and we propose a network analysis to account for potential interference effects. On the supply side, the policy led to a 63.2% reduction in weekly gambling streams and a 44.3% decrease in overall content production among streamers whose content was banned. Moreover, streamers whose gambling content was not banned reduced their gambling and overall content production by 12.2% and 17.6%, respectively, indicating a regulatory spillover effect. This reaction was more pronounced among popular streamers and those with greater reputation concerns. On the demand side, while the policy reduced total viewership and low-tier subscriptions for the affected streamers, revenue from high-tier subscriptions – reflecting more loyal viewers – remained unaffected. We discuss the implications of Twitch's policy ban and the broader practices of content self-regulation on digital platforms.

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

*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 over the necessity of content regulation. As technological advancements expand the range of regulatory challenges for governments, content regulation has increasingly shifted to platforms adopting self-regulation measures. For example, Facebook exploits automatic removal of comments that are classified as toxic content.[1] However, the optimal form of self-regulation often depends on how the platform profits (Liu et al., 2022), potentially limiting its success due to the misalignment between the platform's economic interests and regulation target (Cusumano et al., 2021). Furthermore, the success of platform self-regulation hinges on content producers' compliance with the policy and their avoidance of circumventing it by altering their content production.

In this paper, we investigate the regulatory effects and economic consequences of content self-regulation on Twitch, a major game streaming platform. The content of interest are streams of online gambling (e.g., slots, virtual casino) that have recently surged on Twitch, experiencing a 166% increase in viewership between the first quarters of 2020 and 2022 (StreamScheme, 2023; StreamHatchet, 2022b). This increase in online gambling popularity has been accompanied by several high-profile scandals, with 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 scandals related to online gambling, in October 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). Four major websites – Stake, Rollbit, Duelbits, and Roobet – were banned by this policy, whereas streaming of gambling on other websites was not restricted.[3] The four targeted websites did not differ significantly in their popularity on Twitch or in their website designs aimed at encouraging online gambling. As a result, the limited scope of the policy

[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/.

[3] The limited scope of this policy partially aligns with the findings in Liu et al. (2022), suggesting that platforms relying on advertising are less aggressive in content moderation to maintain their user base.

potentially allowed streamers to substitute to gambling content from unrestricted websites.

We examine how content producers and consumers react to Twitch's ban on livestreams of gambling content, and whether the policy ban had any unintended impact on the platform. On one hand, content producers might substitute their banned livestreams with other unbanned gambling or 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 effective in reducing harmful effects of gambling, merely shifting an exposure of viewers to gambling to another platform. 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 broader implications helps platforms and policymakers evaluate the impact 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 both a two-way fixed effect difference-in-differences (TWFE-DiD) estimator and a synthetic DiD estimator (Arkhangelsky et al., 2021) to address several identification challenges. Additionally, we show the robustness of our results by incorporating network analysis in our estimation to internalize potential interference caused by the policy at the treatment group level or the streamer community level, 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 show that the policy successfully reduced the overall supply of gambling content across the platform by completely removing banned gambling content and decreasing unbanned gambling content. We do not see any evidence of the substitution of affected streamers for other content, including games with gambling-like features. Instead, streamers of banned and unbanned content produce less content overall as a result of the policy, with a reduction of weekly streams of 44.3% and 17.6%, respectively.

We further investigate the policy impact on the supply side among various types of streamers. 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 further tested the underlying mechanism and find that streamers' concerns about their reputation help explain the heterogeneous effects of the banning policy.

We then turn to the policy impact on the demand side. 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 leverage website traffic data and find that although the policy had significant effects on both the supply- and the demand-side outcomes of Twitch, the banned online gambling websites targeted by the policy did not suffer from a traffic loss compared to their unbanned competitors after the policy implementation.

To the best of our knowledge, our research is the first to 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 shown that gambling-like content in video games provides similar psychological experiences to real gambling, such as slots (Drummond and Sauer, 2018; Larche et al., 2021; Amano and Simonov, 2023), and that consumption of this content is correlated with future gambling behavior (Zendle et al., 2019; Kristiansen and Severin, 2020; Close et al., 2021), 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 regulatory effect 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). Particularly, Liu et al. (2022) explores the relationship between a platform's monetization channel and its optimal strategy for content moderation. They find that platforms under advertising behave less aggressively than those under subscriptions, because the former relies on a larger user base, whereas the latter focuses more on consumers' willingness to pay. Our context aligns with this result because Twitch mainly profits from advertising and adopted a less aggressive banning policy. Additionally, we contribute to this literature by empirically examining the impact of content moderation on both advertisement income[4] and subscription income as Twitch makes profits through both channels. Moreover, many theoretical works have discussed the unintended consequences of self-regulation policies on digital platforms, including regulations on ad transparency (Wu et al., 2022), search self-preferencing (Zou and Zhou, 2024), and product safety certification (Iyer and Singh, 2018). We provide new evidence to this literature by showing that self-content regulation on Twitch led to an unintended reduction in the production of unregulated content, highlighting a potential misalignment between regulation policy targets and the platform's economic interests.

Second, we add to the literature of content creation and consumption on platforms that

[4]We do not directly collect data on Twitch advertisements. However, since Twitch primarily uses pre-roll and mid-roll ads, its advertisement income is significantly tied to total hours watched by viewers, which we discuss in the demand-side analysis.

rely on user-generated content (UGC). Zhao et al. (2023) document both direct and indirect spillover effects on viewer flows caused by streamers' content switching on Twitch, showing that the effect size varies with streamer popularity. 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. Our paper extends this literature by empirically examining the impact of a direct ban on a specific type of content, thereby providing evidence on the overall substitutions in content creation. In addition, we discover that both content creation and consumption can be affected disproportionately, and we explore how these non-uniform effects can be attributed to the heterogeneity among content producers and consumers on the platform.

Third, our paper is part of the rapidly growing literature on the impact of livestreaming as a marketing channel. Zhang et al. (2023) find that 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 the supply of gambling livestreams on gambling websites and find that the banning policy did not lead to a negative promotional effect on engagement in online gambling platforms.

Finally, our paper connects to a growing strand of literature that examines how the legalization of sports gambling affects household financial health (Baker et al., 2024; Hollenbeck et al., 2024), tax revenue, and irresponsible gambling behavior (Taylor et al., 2024). While our primary focus is on livestreams of online gambling through virtual casino games, viewers of these livestreams may also develop addictions to sports betting. Therefore, our findings may inform potential regulations on sports gambling livestreams to mitigate their negative financial impacts.

The remainder of the paper is structured as follows. Section 3 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 presents descriptive evidence. Section 7 discusses the main challenges over identifying the causal effect and our empirical approaches. Section 8 presents the main results in supply-side analysis, including heterogeneous treatment effects among different groups of streamers. Section 9 presents the main results in demand-side analysis. Section 10 describes the policy impact on banned online gambling websites. Section 11 concludes.

3 Conceptual Framework

In this section, we present a conceptual framework that illustrates how the banning policy may affect the supply of gambling and non-gambling content, and highlights the main empirical targets of this paper.

Consider a representative streamer i who has streamed gambling content before a banning policy was implemented on the livestreaming platform. The streamer makes a static decision on her streaming plan, which is characterized by a vector of content quantities denoted by $\boldsymbol{q}_i = (q_{1i}, \ldots, q_{ki}, \ldots)'$. For simplicity, we assume that her streaming plan consists of three types of content: gambling content targeted by the policy, gambling content not targeted by the policy, and non-gambling content, corresponding to $k = 1, 2, 3$ respectively.

We assume that the streamer faces a unit mass of viewers whose tastes for gambling livestreams are captured by a preference parameter θ, and she earns revenue from her viewers' payments for each type of content. In reality, the price menu for each type of payment (subscriptions, donations) is typically determined by the livestreaming platform, and both streamers and viewers are price takers. Therefore, we assume that the unit payment from each viewer for content k, p_k, is exogenously determined. If streamer i chooses to stream content k at a quantity q_k, she expects to receive the following revenue:

$$\pi_k(q_k) = \mathbb{E}[\phi_k(\theta, q_k)] \cdot p_k = \int \phi_{jk}(\theta_j, q_k) dF_\theta \cdot p_k$$

where $\mathbb{E}[\phi_k(\theta, q_k)]$ is the expected probability that the streamer receives the payment from her viewers, with the expectation taken with respect to the distribution F_θ among her viewers. The probability function $\phi_{jk}(\theta_j, q_k)$ has two key properties: first, because θ_j captures viewer j's (positive) taste for gambling content, ϕ_{jk} increases in θ_j for $k = 1, 2$ and decreases in θ_j for $k = 3$. Second, a streamer is more likely to receive payment for streaming a particular type of content for a longer duration by developing a loyal viewer base of that content. However, viewers may also approach a satiation point if the streamer oversupplies a certain type of content. Therefore, we assume that $\frac{\partial \pi(k)}{\partial q_k} > 0$, $\frac{\partial^2 \pi(k)}{\partial q_k^2} \leq 0$, and $\pi_k(0) = 0$.

Generating livestreams also incurs costs. We assume a linear cost function for each type of content k with cost parameter c_k, although our main results hold if a quadratic cost function is assumed, which is also common in content marketing literature. Combining streamer i's expected revenue with her cost function, she chooses her optimal streaming plan $\boldsymbol{q}_i^* = (q_{1i}^*, q_{2i}^*, q_{3i}^*)'$ by solving the following optimization problem:

$$\boldsymbol{q}^* = \arg \max_{\boldsymbol{q}=(q_1,q_2,q_3)} \sum_{k=1}^{3} \Big\{ \pi_k(q_k) - c_k q_k \Big\} \quad \text{s.t.} \quad \sum_{k=1}^{3} q_k \leq 1, \quad q_k \geq 0 \quad \forall k. \tag{1}$$

We then characterize the optimization problem following the implementation of the banning policy. Since the policy completely removes the supply of content $k = 1$ from the market, it forces $q_1 = 0$ for streamer i, which can be modeled as setting c_1 to $+\infty$. Denote the post-policy quantities by $\tilde{q}$ and the post-policy cost by $\tilde{c}$. The post-policy optimization problem is given by

$$\tilde{\boldsymbol{q}}^* = \arg \max_{\tilde{\boldsymbol{q}}} \sum_{k=2}^{3} \Big\{ \pi_k(\tilde{q}_k) - \tilde{c}_k \tilde{q}_k \Big\} \quad \text{s.t.} \quad \sum_{k=2}^{3} \tilde{q}_k \leq 1, \quad \tilde{q}_k \geq 0 \quad \forall k. \tag{2}$$

As discussed above, policymakers and the livestreaming platform may have different targets when they evaluate the policy's impact. 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, which, in our conceptual framework, is measured by $\sum_{k=1}^{2}\{\tilde{q}_k^* - q_k^*\}$. In contrast, the platform focuses on the economic consequences of the policy. Since a livestreaming platform primarily earns revenue from advertisements and shares of viewer payments, its income depends on changes in both the total supply of livestreams and streamer revenue. Therefore, we focus on how our model predicts these changes and then test them in our empirical studies. First, the following proposition summarizes the theoretical results:

Proposition 1

1. *Suppose that neither F_θ, c_2 nor c_3 changes after the policy implementation, the streamer who optimizes her profit given (1) and (2) finds that $q_2^* < \tilde{q}_2^*$, $q_3^* < \tilde{q}_3^*$ and $q_2^*+q_3^* < \tilde{q}_2^*+\tilde{q}_3^*$, i.e. the supply of all other content increases after the policy implementation. However, the total supply of gambling content will decrease;*

2. *Let* $\Delta c_k = \tilde{c}_k - c_k$, $k = 2, 3$. *There exists* Δc_k^* *such that if* $\Delta c_k > \Delta c_k^*$, $q_k^* > \tilde{q}_k^*$; *otherwise,* $q_k^* < \tilde{q}_k^*$;

3. *Denote* $r_{kj} = \frac{p_k}{p_j}$ *as the price ratio between content* k *and* j. *Then, there exists* r_{21}^* *and* r_{31}^* *such that if* $r_{21} < r_{21}^*$ *and* $r_{31} < r_{31}^*$, *then* $\sum_k \pi_k(q_k^*) < \sum_k \pi_k(\tilde{q}_k^*)$, *i.e. total revenue decreases after the policy implementation. In addition,* r_{21}^* *and* r_{31}^* *decreases in* Δc_2 *and* Δc_3.

We provide derivations of the results in Appendix A. The first part of Proposition 1 predicts that streamers who has gambling content banned by the policy will increase their supply of unbanned gambling and non-gambling livestreams if nothing else changes in the market, but the policy still leads to a reduction in total gambling livestreams. This is not surprising: since one type of content is banned, streamers are motivated to switch their time and resources (which is captured by the shallow price in the constrained maximization problem) into streaming unbanned gambling content or non-gambling content, based on how much they expect their current viewers to enjoy gambling livestreams. However, many real-world elements complicate this prediction. First, it is possible that the cost of supplying other content changes after the policy. For example, streamers who have streamed banned gambling content may expect that additional banning policy would come in the future. They may also concern about providing gambling content will hurt their reputation under a banning policy. Both will increase their mental cost of supplying unbanned gambling livestreams, leading to the second results in Proposition 1 that their supply of gambling content may decrease after the banning policy. Second, the policy may also lead to changes in the distribution of viewer preferences F_θ. A viewer who is only interested in gambling content may expect that his favorite streamer streams less gambling due to the policy, and switch to other platforms or become less likely to donate to the affected streamer, leading to further reduction in gambling livestreams. Finally, streamers may change their actual behavior in gambling and non-gambling livestreams to increase the profitability of their livestreams, which is not captured by this model.

Finally, the third part of Proposition 1 shows that the change in streamer revenue primarily depends on viewers' willingness to pay for each type of content. If the streamer's gambling content attracts a distinct audience that does not overlap with her non-gambling viewership, and she cannot attract payment from these viewers through unbanned content, then her total revenue will decrease following the ban. Additionally, if producing unbanned content requires more effort for a streamer, leading to a huge decline in total streaming activity, a revenue loss becomes even more likely.

The conceptual framework illustrates how the incentives of the platform, the regulator and the content providers (streamers) may differ. While the total supply of gambling livestreams will decrease, the platform and content creators may primarily be concerned with their loss in total profits. 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 a number of key characteristics of streamers in 8.

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 1, 2022, to December 31, 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 1, 2022 to October 26, 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 26, 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

B for details).[5] 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 their streaming content before the announcement of the banning policy on September 20, 2022. All streamers who have never streamed any gambling content before the policy announcement are classified as untreated streamers. Particularly, we note that streamers who almost never streamed online gambling content are less likely to be affected by the policy and may behave similarly to non-gambling streamers. Therefore, we have also examined our results based on an alternative treatment definition, where streamers in the lowest 25th quantile of total gambling streams (7.3 hours) before the policy announcement are classified as untreated streamers. Our empirical findings are qualitatively the same across different 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.[6] 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

[5]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.

[6]Removing these streamers means that we do not consider the policy's impact on market entry or exit in this paper. However, most of the removed accounts are official accounts of video game developers, or of the banned gambling websites Twitch permanently deactivated after the policy. Since these accounts did not generate individual user-generated content like unofficial streamers, removing them does not compromise the paper's main focus.

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 summary statistics for our sample streamers. Our final dataset comprises 158 banned streamers, 317 unbanned streamers and 4,626 untreated streamers. Streamers across all groups exhibit similar streaming patterns, especially for metrics such as average log weekly streaming hours of games with gambling-like features and average log weekly livestreaming hours. We find that banned streamers in general rely more on gambling streams than unbanned streamers, whereas they are in general more popular among viewers in terms of total hours watched and subscriptions.

	Mean	Group Mean		
		Banned	Unbanned	Untreated
Supply-Side Variables				
Gambling Hours	0.090	1.592	0.902	0.000
Lootbox Hours	1.438	1.302	1.361	1.502
Streaming Hours	2.577	2.894	2.639	2.573
Other Games Hours	1.135	0.482	0.571	1.134
Demand-Side Variables				
Total Hours Watched	8.431	9.805	8.801	8.421
Tier 1 Subscription	3.272	3.496	2.569	3.424
Tier 2 Subscription	0.376	0.207	0.168	0.429
Tier 3 Subscription	0.387	0.211	0.161	0.434

Table 1: **Summary Statistics for Streamers.** All summary statistics are presented as the logarithm of the average weekly level.

Video game data. 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.

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 clips 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.

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 9).[7] In Appendix B, 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).[8] We use the data to analyze streamer revenue changes due to the policy.

Other streamer-level data. For each streamer, we also collect from *Streams Charts* 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). We use the residence information as one source to match video clips to streams (see discussions in Appendix C.1). Additionally, we use details on prohibitions to explore the underlying mechanisms behind the observed heterogeneous treatment effects.

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.

5 Detection of Banned Content and Streamers

As discussed in Section 4, our streamer-level data does not specify whether a gambling stream was about 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 several different methods and data

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

[8]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.

sources (videos, stream titles and real-time chat logs) to classify treated streamers into banned and unbanned groups, which we detail next.

5.1 Identifying Banned Content From Video Clips and VODs

We first construct a list of streamers and define them as "treated" if they stream gambling content at least once before the policy is implemented. For each of these treated streamers, we fetch URLs of all their past gambling video clips pre-policy using Twitch API.[9] 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 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 in video clips through a three-step procedure as shown in Figure 1: First, we download all past gambling video clips of a streamer (over 4TB) based on clip URLs using *youtube-dl* package (which also supports Twitch video downloads). Then, to optimize processing, we decompose each video into frames and sample three equally spaced frames (start, middle and end) for content analysis. This sampling approach is effective as clips are under 60 seconds with minimal scene changes, thus reducing the number of frames for content processing without losing critical information. Finally, we convert sampled frames to grayscale[10] and extract texts using Google's *Tesseract-OCR*. We check the extracted texts against all banned websites' keywords. We use a restrictive criterion for classification by considering clips as containing banned content only if keywords from banned websites are detected in all sampled frames (or unbanned if absent in all sampled frames), to avoid misclassifications.

We complement content analysis on video clips with Video on Demand (VOD) - archives of video content - posted on other platforms if they are made available by any streamers (e.g., xQc, Adin Ross and ItsSliker).[11] We match all videos (i.e., video clips and VODs) with the

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

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

[11]VODs are archives of content previously streamed live on Twitch and can span several hours (if unedited). 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`. For long VODs, we manually check through the content to identify the presence (and time of presence) of the banned content.

original game streams they come from to compile a list of streams and streamers with banned content. Appendix C.1 provide further details on the matching procedure.

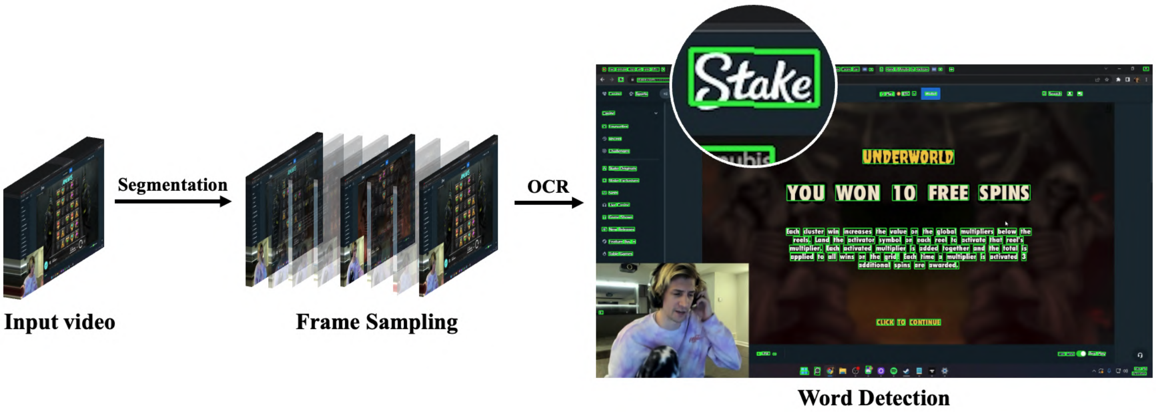


Figure 1: **Banned content detection using videos**. We sample frames from each streamer's past gambling video clips and use OCR to extract texts from images. A clip is considered to contain banned gambling content if keywords from banned websites are detected in the list of extracted texts.

5.2 Identifying Banned Content From Stream Titles

Twitch streamers often include commands (an exclamation point followed by a keyword such as "!Stake", "!Roobet") in stream titles, allowing viewers to type them in chat to trigger bots that provide information. When names of the banned websites are used as keywords following an exclamation point in gambling streams, these are commands to referral links where viewers can use to access gambling websites and streamers earn commissions in exchange.[12]

Recognizing this convention, we also complement the detection of banned streams and streamers leveraging our high-frequency data which tracks the real-time stream titles and games streamed in every stream at an interval of every 10 or 15 minutes. We define a gambling stream as containing banned content prior to the policy implementation if both (1) a banned website's referral link ("!" followed by banned website keywords) is observed in the stream title and (2) the game being streamed at that time is also an online *gambling* content. In Appendix C.2, we demonstrate the validity of this approach by showing that: first, *non-gambling* streamers almost

[12]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/`.

never include banned websites' referral links in their stream titles, while *gambling* make such usages; second, almost none *gambling* streamers include referral links to banned websites in the titles of their *non-gambling* streams; third, we verified that *banned* streamers (after our final banned classification) indeed do not have any usages of referral links in gambling stream titles. In addition, we also checked an alternative banned gambling stream detection approach, which solely bases on detecting the presence of banned website keywords in stream titles. Although the alternative approach also works well, we show that our approach, which is more conservative, has even better performance.

5.3 Identifying Banned Content From Chat Logs

While some streamers may neither make video clips accessible nor include referral links to banned websites in stream titles when streaming banned gambling content, we complement the identification of streamers' latent treatment status with a third approach by detecting the presence of banned content using in-stream chat logs.

To predict which gambling streams contain banned content based on in-stream chats, we need to first construct a ground truth chat sample for banned and unbanned gambling streams, respectively. We use multiple criteria to construct the ground truth. For the *banned* ground truth chats, we use chats from gambling streams that meet both 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. While the first source alone can be treated as the ground truth, satisfying both criteria ensures high quality of the ground truth sample. For the *unbanned* ground truth chats, we use chats from gambling streams that satisfy all three following criteria: (1) the gambling streams must have video clips with no banned website names detected in any of the 3 sampled frames of a video using OCR, (2) the streams are from streamers with no pre-policy gambling video clips detected to contain any banned content by OCR , (3) the gambling streams must have neither referral links or keywords of banned websites featured in stream titles. In Appendix C.3, we provide further justifications on our approach of constructing the ground truth samples for banned and unbanned gambling streams based on chats. Our final ground truth chats sample covers 534 banned streams and

821 unbanned streams, which serves as the basis for our stream classification.

Having constructed the ground truth chats samples, our detection strategy is based on the hypothesis that banned gambling streams will contain a higher proportion of referral links to banned websites in chats compared to unbanned gambling streams. As discussed previously, this is motivated by the industry practice that streamers often use chat bot commands to automatically respond to viewers' referral link requests. In Appendix C.4, we conduct several validation analyses to justify our detection strategy. First, we show that banned gambling streams contain 100 times more referral links of banned websites — despite having fewer total chats, and 130 times more unique viewers who posted referral links to banned websites in chats compared to unbanned gambling streams (see Appendix Table C.4.1). Second, we use the average number of referrals per hour as test statistics and demonstrate that the distribution of referral links to banned websites in unbanned streams has a huge mass at zero, whereas banned streams have a significantly larger probability of containing more than one banned websites' referral links per hour (see Appendix Figure C.4.1). The distribution of referral links in banned streams also first-order statistically dominates that in unbanned streams. Finally, we conduct a bootstrapped Kolmogorov-Smirnov test and reject the null that the number of referral links per hour follow the same distribution in banned and unbanned gambling streams (see Appendix Figure C.4.2). We check the predictive performance using the KS value as the threshold for classification and confirm that a *threshold-based* classification approach based on referral links per hour is adequate.

Since our goal is to reduce misclassifications, we propose to use grid search with 10-fold cross validation to search for the optimal threshold value that jointly minimizes Type I and Type II errors for classification.[13] To achieve this, we first correct the imbalance in our ground truth samples (chats from 534 banned versus 821 unbanned streams) by implementing a *random undersampling* (Breiman, 2017) during the training phase of each cross validation fold. We then implement a 10-fold cross validation with grid search and evaluate thresholds in-sample to identify the optimal threshold.[14] We find that a threshold of 0.3 jointly minimizes Type I and Type II errors in-sample.[15] We report out-of-sample performance, confusion matrices and ROC curve

[13] 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.

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

[15] A threshold of 0.3 also achieves good performance on most other measures. See Appendix Table C.4.2 for details.

in Appendix Table C.4.3 and Figures C.4.3 and C.4.4. The threshold-based approach can distinguish between the two samples 97.5% on average, demonstrating strong predictive performance.

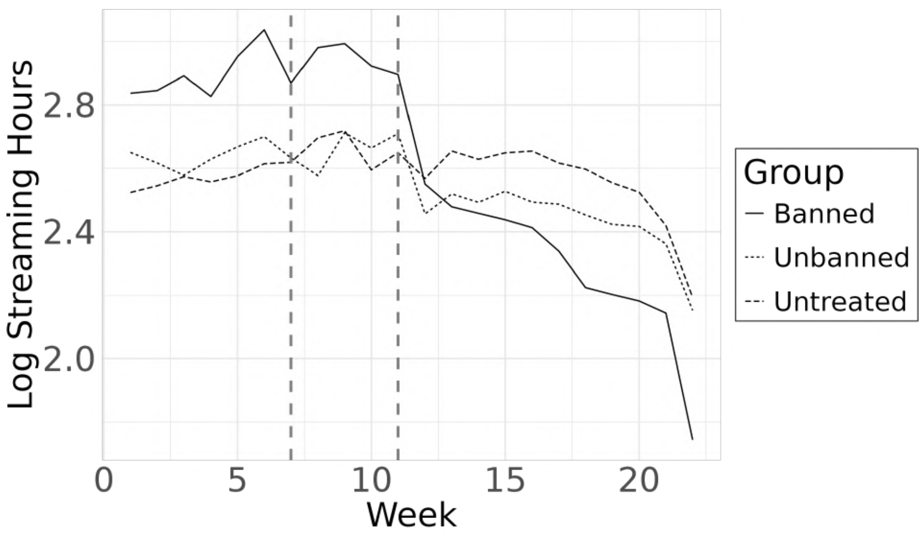
Identification of Treated Groups Using All Sources. We combine detection results from all three approaches (video clips, stream titles and chat logs) described in this section to construct a list of banned streams. Then, we use these identified banned streams to classify treated streamers into banned and unbanned groups. Appendix Figure C.4.5 illustrate the number of streamers detected by each approach alone and combined.

We note that we have used multiple data sources and a conservative approach with multiple criteria when we approach the problem of detecting banned and unbanned gambling content. While we have demonstrated the validity and extremely high classification accuracy of our approach, one may still concern about potential misclassifications which affect casual estimates. In Appendix D, we conduct a formal sensitivity analysis and prove that we have correctly identified ATT for the banned streamers.

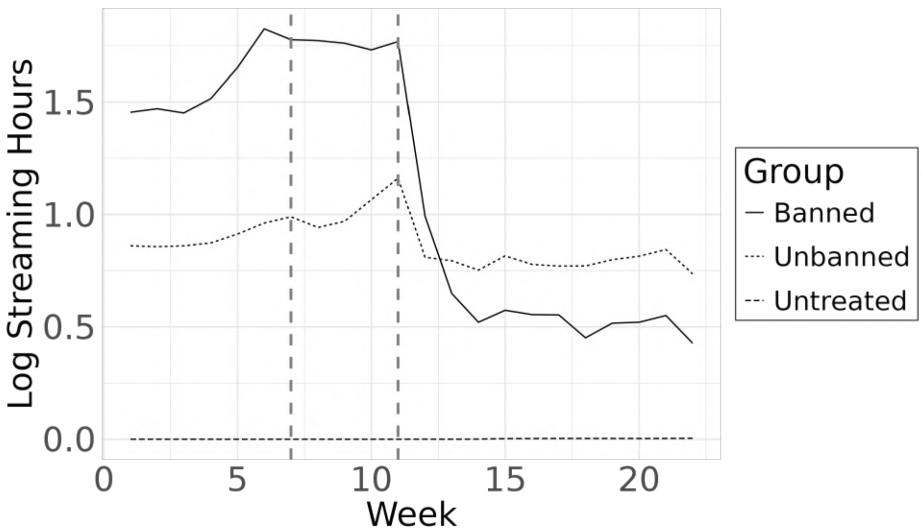
6 Descriptive Evidence

We first present some descriptive evidence on the impact of implementing banning policy. Figure 2a 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 20, 2022) and implementing the policy (October 18, 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.

Figure 2b 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



(a) All Content



(b) Online Gambling

Figure 2: **Average log weekly streaming hours**.

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.

7 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 regulatory effect of the content regulation and reveal any potential side effects that might not align with the platform's incentives. In Section 6, 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 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. In addition, 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). 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.

7.1 Two-way Fixed Effect DiD

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

$$\begin{aligned} y_{it} = \alpha_i + \gamma_t + \beta_1 \cdot \text{Banned}_i \cdot \text{Post}_t + \beta_2 \cdot \text{Banned}_i \cdot \text{Announcement}_t \\ + \beta_3 \cdot \text{Unbanned}_i \cdot \text{Post}_t + \beta_4 \cdot \text{Unbanned}_i \cdot \text{Announcement}_t + \varepsilon_{it} \end{aligned} \tag{3}$$

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. In Appendix F.1, we leverage a flexible event-study model to estimate the time-varying treatment effects and investigate the persistence of the policy's impact.

7.2 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:

$$(\hat{\beta}^{\text{SynthDiD}}, \hat{\alpha}, \hat{\gamma}) = \underset{\beta,\alpha,\gamma}{\arg\min} \left\{ \sum_i \sum_t (\tilde{y}_{it} - \alpha_i - \gamma_t - \beta_j^{\text{SynthDiD}} \cdot D_{j,it}) \hat{w}_i \hat{\lambda}_t \right\} \tag{4}$$

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 $\tilde{y}_{it}$.

Finally, even with the use of multiple causal inference estimators, 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 9.2.

8 Supply-Side Results

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

8.1 Content Creation of Gambling and Non-Gambling Livestreams

Table 2 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, unbanned streamers also reduced their creation of gambling content even though they were not directly affected by the policy. Therefore, the banning policy was successful in regulating the overall supply of online gambling content across the platform.

	log(Gambling hours + 1)	
	TWFE-DiD	SynthDiD
Banned (β_1)	-1.022***	-1.001***
	(0.089)	(0.082)
Δ%	-64.0%	-63.2%
Unbanned (β_3)	-0.116**	-0.130***
	(0.049)	(0.043)
Δ%	-11.0%	-12.2%
Streamer FE	✓	✓
Week FE	✓	✓
Observations	112,222	112,222
Mean dependent variable	1.132	1.132

Table 2: **The effect of Twitch's banning policy on streaming hours of online gambling**. 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 3. 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.

	log(LootBox Games + 1)	
	TWFE-DiD	SynthDiD
Banned (β_1)	0.012	0.021
	(0.065)	(0.054)
$\Delta\%$	-	-
Unbanned (β_3)	-0.046	-0.054
	(0.046)	(0.036)
$\Delta\%$	-	-
Streamer FE	✓	✓
Week FE	✓	✓
Observations	112,222	112,222
Mean dependent variable	1.341	1.341

Table 3: **The effect of Twitch's banning policy on streaming hours of video games with gambling-like features**. 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: $*p < 0.1$; $**p < 0.05$; $***p < 0.01$

Finally, we turn to examine the policy's impact on total content creation on the treated streamers. Estimates in the first two columns of Table 4 show that both banned and unbanned streamers reduced their total streaming hours after the policy implementation. The SynthDiD estimates imply a reduction of 44.3% for streamers of banned gambling content and 17.6% for streamers of unbanned gambling content. Given that gambling content accounts for 27.5% and 17.6% of the total content for banned and unbanned streamers, respectively, the estimates imply that streamers reduced the production of non-gambling content by 26.9% and 15.5%, respectively.

Table E.0.1 in Appendix E further investigates the sources of the decline 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 13.8% and 12.3%, respectively. However, we observe pre-trends for this content, as shown in Figure F.1.1 and Figure F.1.2 in Appendix F.1 show, and do not interpret these estimates as causal effects of the banning policy on non-gambling content creation.

For brevity, we only present the estimates of post-policy implementation effects, β_1 and β_3 in Table 2 to Table 4. In Appendix E, we present full estimation results including the policy's effect during the announcement period, β_2 and β_4. Interestingly, we find a 13 - 16%

	log(Streaming hours + 1)		log(Other games + 1)	
	TWFE-DiD	SynthDiD	TWFE-DiD	SynthDiD
Banned (β_1)	-0.583***	-0.585***	-0.146***	-0.149***
	(0.084)	(0.075)	(0.040)	(0.032)
$\Delta\%$	-44.2%	-44.3%	-13.5%	-13.8%
Unbanned (β_3)	-0.182***	-0.194***	-0.133***	-0.130***
	(0.044)	(0.040)	(0.032)	(0.026)
$\Delta\%$	-16.6%	-17.6%	-12.5%	-12.2%
Streamer FE	✓	✓	✓	✓
Week FE	✓	✓	✓	✓
Observations	112,222	112,222	112,222	112,222
Mean dependent variable	2.724	2.724	0.541	0.541

Table 4: **The effect of Twitch's banning policy on total streaming hours and streaming hours of games without gambling-like features**. 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: $*p < 0.1$; $**p < 0.05$; $***p < 0.01$

increase in the weekly supply of gambling livestreams after the policy announcement and before the policy implementation (see also Figures F.1.1 and F.1.2 in Appendix F.1). Since Twitch's announcement clarified the date to implement the banning policy, these increases may be explained by streamers' incentives to grasp more revenue from gambling livestreams before the actual implementation. Therefore, these results may alert the platform to carefully choose the gap between a policy's announcement and its implementation to avoid any unpredictable issues due to the temporary increase in the supply of the regulation target that may harm the platform.

8.2 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 impact on regulating gambling content and investigate the underlying mechanisms.

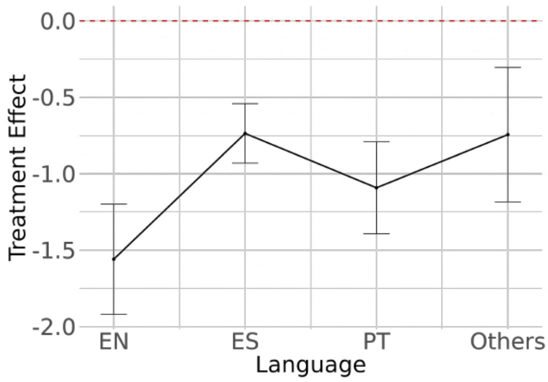
HTE on Main Language. First, we examine potential heterogeneous effect of the banning policy on streamers of different languages. This might be interesting because the banning policy only targeted websites that were unlicensed in the U.S, whereas the main language of all 4 websites are English. Moreover, English is not the dominant language in gambling livestreams on Twitch.[16] 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 3 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 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 total livestreaming hours, compared to untreated streamers in the same language groups.

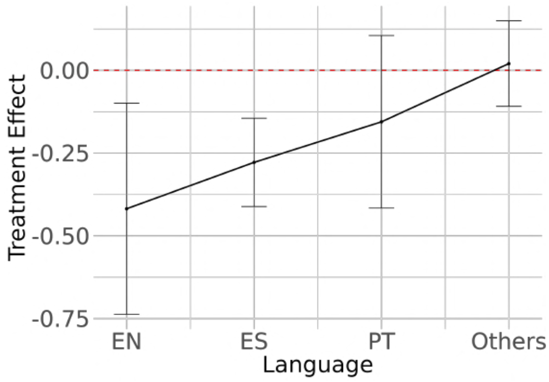
HTE on Streamer Popularity. Next, we explore whether the policy impact on content creation varies across streamers with different levels of popularity. From Figure 4, we find that in the banned group, streamers with higher popularity reduced their content creation of gambling livestreams slightly more than streamers with low popularity after the policy. More interestingly, 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 large negative impact on the total content creation across the platform.

How can we explain the finding that more popular streamers were affected more by the policy? One potential explanation is that these streamers may care more about their reputation. On one hand, Twitch regulated gambling livestreams because it was controversial and raised reputation concerns for the platform. If a highly popular streamer continues providing a large amount of gambling content but is later prohibited by the platform, they will lose more revenue due to their high number of subscriptions, in-stream donations and total hours watched. On

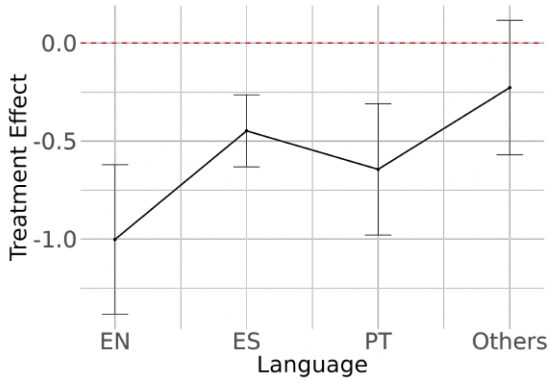
[16]Among the 475 gambling streamers in our dataset, 72 used English as the main language, 92 used Spanish, 111 used Portuguese, and 200 used other languages. Streamers tend to use the same language in both gambling and non-gambling streams.



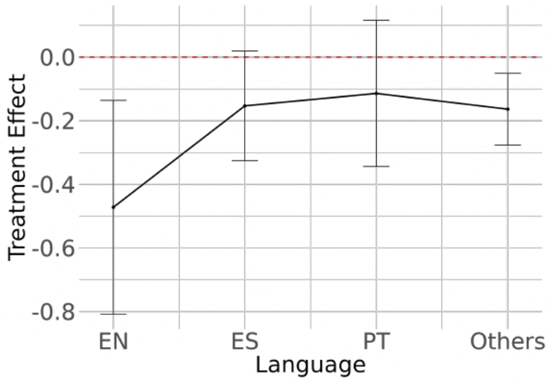
(a) Gambling, Banned



(b) Gambling, Unbanned



(c) Total Livestreams, Banned

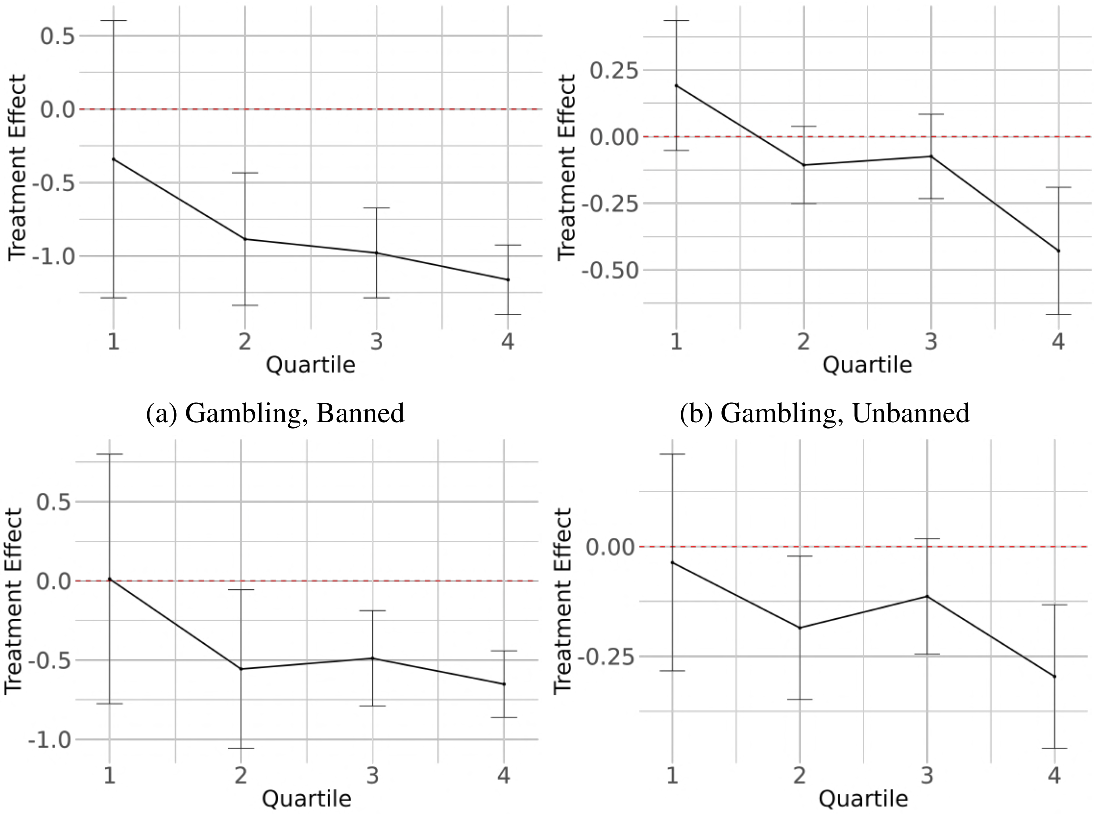


(d) Total Livestreams, Unbanned

Figure 3: **Heterogeneous treatment effects on streamers with different main languages**. 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.

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.

Therefore, 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 days of account prohibition each streamer received before the policy announcement. Column 1 and 2 in Table 5 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 days of prohibitions before the policy implementation). Specifically, a banned streamer with no prior account



(a) Gambling, Banned

(b) Gambling, Unbanned

(c) Total Livestreams, Banned

(d) Total Livestreams, Unbanned

Figure 4: **Heterogeneous treatment effects on streamers with different popularity**. 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.

prohibitions reduced gambling streams by 64.6%, and each additional day of prior prohibition resulted in a 0.4% smaller reduction compared to the baseline. Moreover, unbanned streamers are less sensitive to streamer reputation, with each additional prior prohibition resulting in a 0.6% 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.

9 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,

	× Banned Days	
	Main Effect	Interaction Effect
Banned (β_1)	-1.039***	0.004**
	(0.091)	(0.002)
Unbanned (β_3)	-0.133***	0.006***
	(0.050)	(0.002)
Streamer FE	✓	✓
Week FE	✓	✓
Observations	112,222	112,222
Mean dependent variable	1.132	1.132

Table 5: **Underlying mechanisms of heterogeneous treatment effects on gambling livestreams**. 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: $*p < 0.1$; $**p < 0.05$; $***p < 0.01$

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.

9.1 Total Content Consumption and Revenue Channels

Table 6 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 4). 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. In Appendix E, we show full estimation results and find that total hours watched did not significantly decrease for banned streamers but decreased approximately 19.6% for unbanned streamers at the marginal significance. Both groups exhibit a similar proportional decline in content consumption relative to content creation, implying consistent market responses.

	log(Hours Watched + 1)	
	TWFE-DiD	SynthDiD
Banned (β_1)	-1.694***	-1.696***
	(0.268)	(0.241)
Δ%	-81.6%	-81.6%
Unbanned (β_3)	-0.523***	-0.554***
	(0.134)	(0.122)
Δ%	-40.7%	-42.5%
Streamer FE	✓	✓
Week FE	✓	✓
Observations	112,222	112,222
Mean dependent variable	9.135	9.135

Table 6: **The effect of Twitch's banning policy on total hours watched**. 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[17]. Therefore, a higher-tier subscription can be seen as a signal of viewer loyalty and engagement in the streamer's community.

For brevity, we only show SynthDiD estimates for these outcome variables in Table 7, where we present full estimation results in Appendix E. 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

[17]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.

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. By comparing the reduction in total hours watched and subscriptions, we also find that viewership is approximately twice as elastic as subscription consumption.

	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	2.877	0.181	0.178

Table 7: **The effect of Twitch's banning policy on different tiers of subscriptions**. 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: $*p < 0.1$; $**p < 0.05$; $***p < 0.01$.

9.2 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, 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 streamers from streaming banned gambling content and thus their viewers from 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.

Our context presents unique operational challenges in using existing methods aiming at identifying causal effects under network interference (Aronow and Samii, 2017; Li and Wager, 2022) due to: (i) the complexity of the network structure, characterized by significant heterogeneity in streamer connectivity, and (ii) our treatment assignment is not random, nor can interference be assumed to arise randomly from unobserved streamer characteristics (Li and Wager, 2022). In contrast to separately identifying the policy's effect and potential spillover effects, we mitigate potential concerns of SUTVA violation by proposing a procedure to automate the construction of treated and untreated groups to internalize potential spillover effects at the treated group level or at the streamer community level.

In general, our procedure to mitigate SUTVA exploits the network structure of streamers. Recall in Section 4, our dataset captures individual registered viewers in streams across time, therefore allowing us to construct streamer networks capturing the extent of registered viewers overlap among streamers before the policy. Based on the network, we employ a community detection algorithm called the Louvain method (Blondel et al., 2008) to identify distinct communities where streamers across communities have minimal to no overlap in viewers. Given that a community may contain streamers of mixed treatment status, we then apply a filtering algorithm based on breadth-first search (Moore, 1959) to select a large enough subgroup of closely connected streamers within the community who share the same treatment status. By selecting subgroups of streamers of the same treatment status while ensuring that they come from different communities, we minimize demand spillovers at the treated group or community level, alleviating the concerns of SUTVA. We show that our results remain robust using the selected subgroups.

Next, we outline the details of the proposed approach.

9.2.1 Network Construction

To mitigate SUTVA, our procedure starts by constructing a network of shared viewership among streamers using each streamer's list of registered viewers in chats (a subset of the total viewers of a stream) before the policy is implemented.[18] To reduce network complexity 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

[18] 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 B.

streamers).[19] After excluding a few streamers with missing chats data, our final network consists of n =1,270 streamers (155 banned, 315 unbanned, 800 untreated) with $n(n-1)/2 \simeq 805{,}815$ edges.

In our network, a *node* represents a streamer, and an *edge* between two nodes indicates the presence of common viewers between 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 8 shows the summary statistics of viewer counts of the streamers from the three groups.

Group	Mean	Std	5_{quantile}	25_{quantile}	50_{quantile}	75_{quantile}	95_{quantile}
Untreated	15,395.00	29,541.50	1,338.60	3,957.00	7,933.50	16,248.75	49,197.95
Unbanned	18,184.54	43,827.04	666.10	3,792.50	8,698.00	19,213.50	52,443.20
Banned	37,134.57	54,788.90	3,782.20	10,825.00	22,637.00	35,796.50	124,309.20

Table 8: **Viewer count before the policy implementation**. This table presents the descriptive statistics of viewer counts for different groups of streamers before the policy implementation.

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 within a group, as shown in Table 8. Figure 5 visualizes the network, where each treatment group is coded with a different color. We observe that it is possible that streamers from different treatment groups are connected in the network. This implies that the regulation directly affecting banned streamers may also affect the remaining streamers who are unbanned or untreated, suggesting potential interference among streamers. This finding motivates us to propose a generic procedure for complex networks designed to reduce demand spillovers across treatment groups for valid identification.

9.2.2 Proposed Solution

Our proposed solution to alleviate SUTVA relies on using a combination of community detection and breath-first search based filtering algorithms to automate the selection of subgroup of streamers for each treatment status. The goal is to ensure that there is no significant cross-

[19]Using a different random sample or varying the sample size for untreated streamers does not qualitatively affect our results.

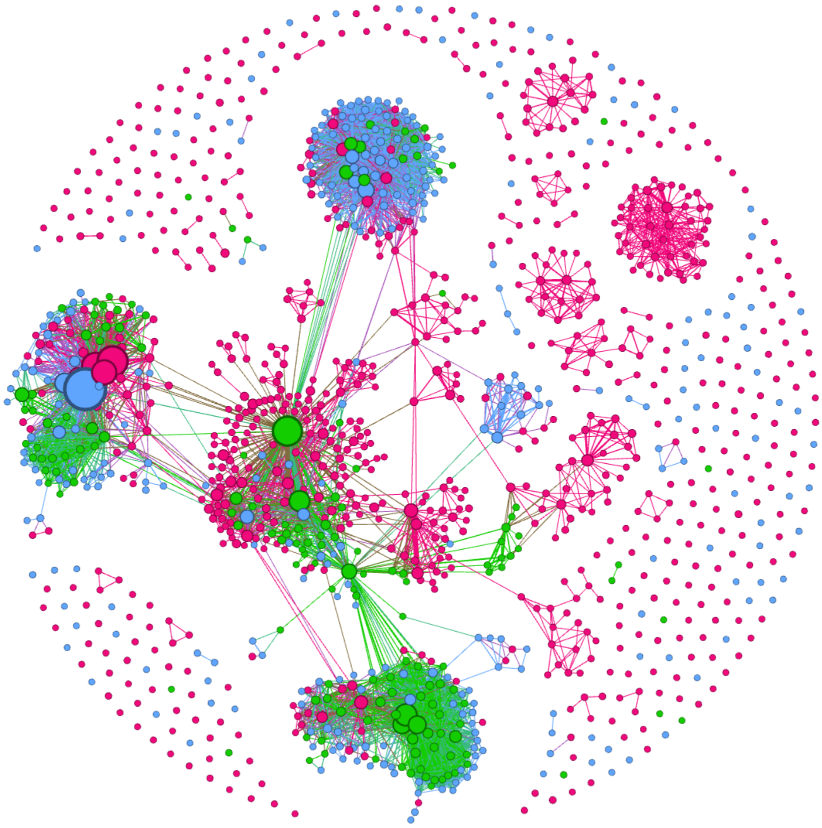


Figure 5: **Network of streamers based on common viewers**. The network is plotted by force-directed placement (Fruchterman and Reingold, 1991), which tries to place connected nodes close to each other and in doing so, helps reveal the basic structure of the network. A node denotes a streamer, and an 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.

group interference. In this part, we delve into the details of our proposed solution. To aid readers with understanding, we also provide an interactive version of the network with labeled nodes, which can be accessed at `https://twitch-gambling.github.io/streamer-network/network/`.

Community Detection. Some streamers have more viewers shared in common and thus are subject to more demand spillovers. To reduce interference among streamers, we first use the Louvain method (Blondel et al., 2008) to identify non-overlapping communities of streamers with intense overlaps in viewers from the large network.[20] The method works by

[20]We also experimented the Girvan-Newman algorithm for community detection, but found Louvain to be more effective in unfolding communities for large networks.

iteratively optimizing modularity, which measures how much more connected the nodes within communities are with respect to those outside communities.

Using the method, we identify 55 community clusters from the network. Appendix Table G.0.1 shows the share of streamers with each treatment status (banned, unbanned, untreated) within a community cluster. Figure 6 illustrates a few examples of communities identified from the network. We observe that the communities can be categorized into three general categories:

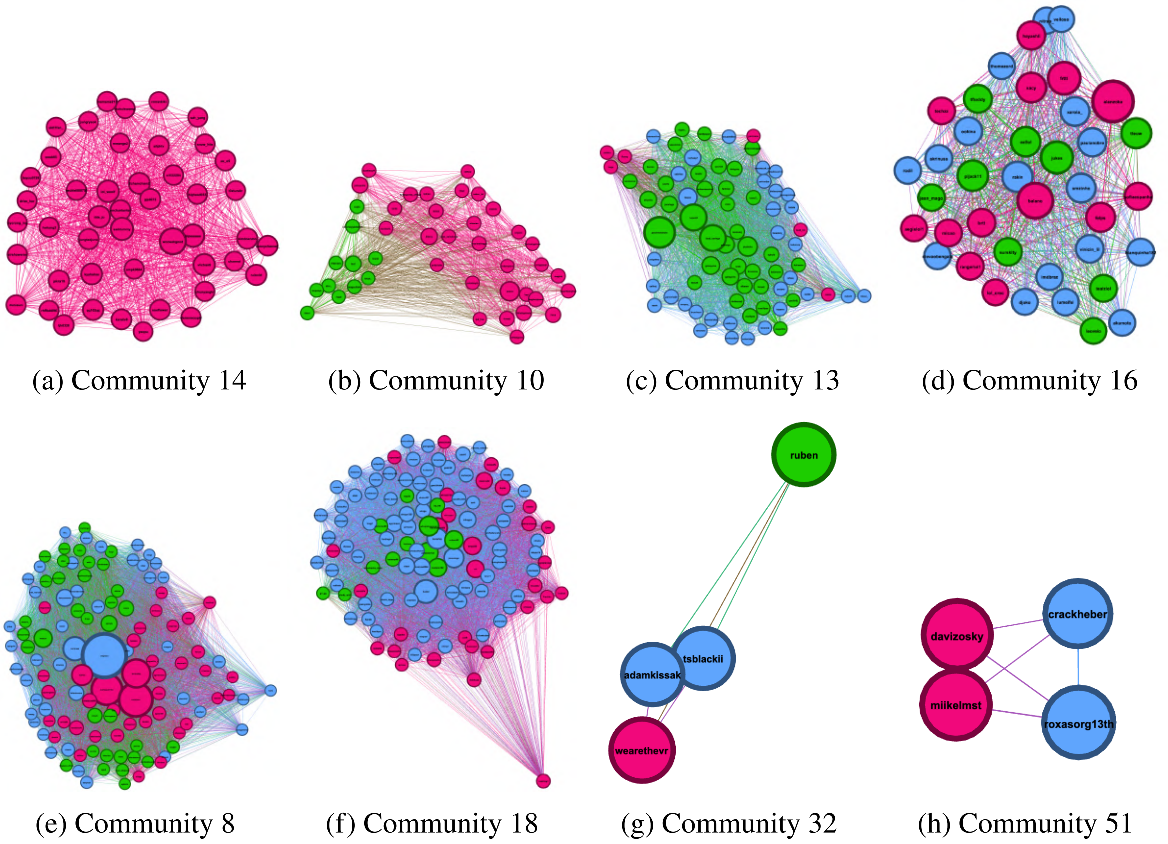
1. *Homogeneous communities*, where all streamers belong to the same treatment status (e.g., community 14).

2. *Dominantly homogeneous communities*, where one treatment status is dominant (e.g., communities 10, 13, 18).

3. *High mixed communities*, where streamers of different treatment statuses are interwoven and there is no dominant treatment status (e.g., communities 8, 16, 32, 51).

While it is straightforward to select streamers for the untreated group (i.e., there are many homogeneous communities that consist entirely of untreated streamers and they are well separated from other groups), it is operationally challenging to select streamers for the banned and unbanned groups. Communities that contain most banned and unbanned streamers are mixed in terms of treatment status. For example, community 13 which consists of predominantly unbanned streamers also contains a fraction of banned and untreated streamers. This introduces complexities in selecting subgroups of homogeneous streamers from these communities where streamers of different treatment statuses are interconnected, creating potential interference.

Breadth-First Search-Based Filtering Algorithm. To address the challenge caused by mixed-type communities where most banned and unbanned streamers belong to, we begin by selecting mixed communities from which it is operationally easy to disentangle streamers of one treatment status from others (see Appendix G for details). For each selected mixed-type community for a treatment status, we identify the streamer with the most edges to other streamers of the target treatment status within the community. We use this streamer as the start node and apply a filtering algorithm based on Breadth-First Search (BFS) (Moore, 1959). Essentially, starting the process with a well-connected node with other nodes of the same treatment status allows the construction of a large enough subset of streamers that belong to the same treatment status before encountering nodes of other status, therefore reducing the likelihood of interference.



(a) Community 14 (b) Community 10 (c) Community 13 (d) Community 16

(e) Community 8 (f) Community 18 (g) Community 32 (h) Community 51

Figure 6: **Examples of Communities**. Communities are detected using the Louvain Method (Blondel et al., 2008). Red denotes untreated, blue unbanned (treated) and green banned (treated) group.

Algorithm 1 outlines the pseudo-code of the process for filtering streamers in any community j.

We illustrate the idea of this process in Figure 7. Intuitively, for any given community, the algorithm works by traversing the network from the start node iteratively. A queue is used to keep track of the nodes that are encountered which we need to explore next in the iterative implementation of BFS. We begin from the start streamer of the community (for example, "jonvlogs" from community 13) 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 him (Panel (b)). Next, 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

Algorithm 1 BFS-Based Filtering

Input: A graph G_j and start node (streamer) $s_{center,j}$ of G_j for any community j

Initialize V := a set to store all visited streamers
Initialize T := a set to store all streamers of the same treatment status as $s_{c,j}$
Initialize Q := a queue to store all streamers encountered but not yet visited
Q.enqueue($s_{center,j}$)
V.add($s_{center,j}$)
Set target_treatment := treatment status of $s_{center,j}$
while $Q \neq \emptyset$ **do**
    current_streamer := Q.dequeue()
    if current_streamer $\notin V$ **then**
        V.add(current_streamer)
        if current_streamer's treatment == target_treatment **then**
            T.add(current_streamer)
            for neighbor $\in G_j$.neighbors(current_streamer) **do**
                if neighbor $\notin V$ **then**
                    Q.enqueue(neighbor)
                end if
            end for
        end if
    end if
end while
return A subset of streamers T with homogeneous treatment status

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 this process iteratively through the layers until all streamers are processed (i.e., queue is empty) to render a final subset of streamers of homogeneous treatment status.

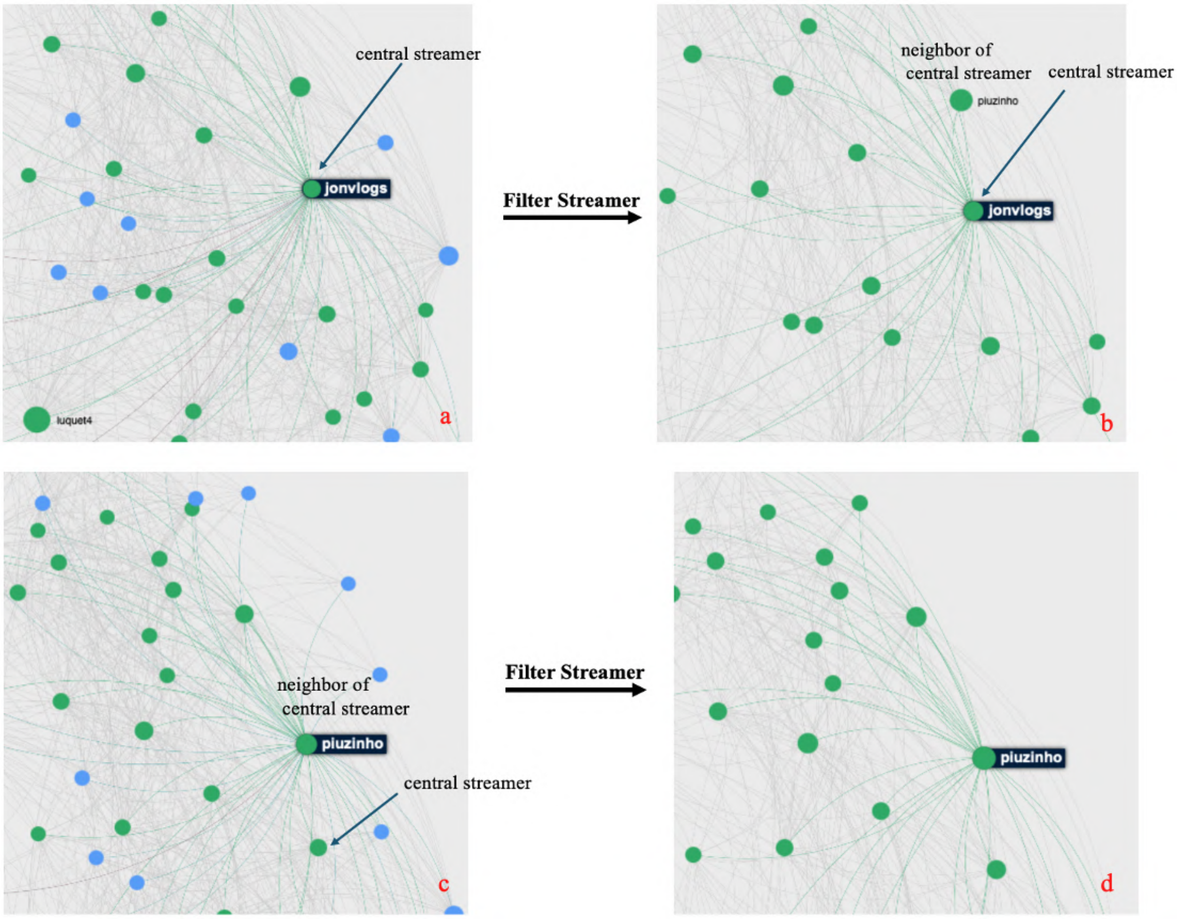


Figure 7: **Filtering streamers of different treatment status iteratively using BFS**. Panel (a) shows the initial cluster with the central streamer "jonvlogs" (green node) and its neighbors, including streamers of different statuses (blue nodes). Panel (b) depicts the cluster after filtering out streamers 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.

9.2.3 Demand-Side Results on Selected Groups

In Table 9, we summarize causal estimates for the same outcome variables as in Table 6 and 7 based on the selected , and compare them with previously reported results. 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. 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 our full streamer sample will overestimate the ATT on the total hours watched, the estimates in Table 9 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.

	log (Hours Watched + 1)	Tier 1	Tier 2	Tier 3
Banned	-2.18 (0.549)***	-0.634 (0.283)**	0.017 (0.024)	-0.021 (0.035)
Δ%	-88.7%	-46.9%	-	-
Unbanned	-0.764 (0.244)***	-0.214 (0.102)**	0.017 (0.02)	0.04 (0.018)**
Δ%	-53.4%	-19.3%	-	4.1%
Streamer FE	✓	✓	✓	✓
Week FE	✓	✓	✓	✓
Observations	10,494	10,494	10,494	10,494

Table 9: **The impact of Twitch's banning policy on demand-side outcomes, based on selected groups of streamers**. All standard errors are clustered at the community level. FE, fixed effect. Significance level: $*p < 0.1$; $**p < 0.05$; $***p < 0.01$

In Table E.0.3 of Appendix E, we present full estimation results for both demand-side and 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 8. Notably, the streaming hours for gambling content among unbanned streamers in the selected group were unaffected by the policy. This outcome is expected because the selected streamers, who are less connected to those in different treated groups, typically have lower popularity, making them less responsive to the banning policy, as we displayed in Figure 4.

Our construction of streamer communities also allows us to fully internalize potential interference within each community by aggregating observations at the community level. Hence, we report the full estimation results based on community-level observations in Table E.0.4 of Appendix E. Note that the number of observations (946) shrinks to the product of the number of distinct communities (43) and the number of weeks (22). We find that when the interference is fully internalized, we obtain slightly higher results compared to streamer-level estimates. The discrepancy between estimates for the same treatment effect may imply the presence of within-community spillovers on the demand-side. The banning policy indeed led viewers to shift to other streamers within the same community, both in terms of non-monetary content consumption (total hours watched) and monetary consumption (subscriptions).

10 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:

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

Table 10: **Types of gambling websites used for event study of website traffic**.

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 10).

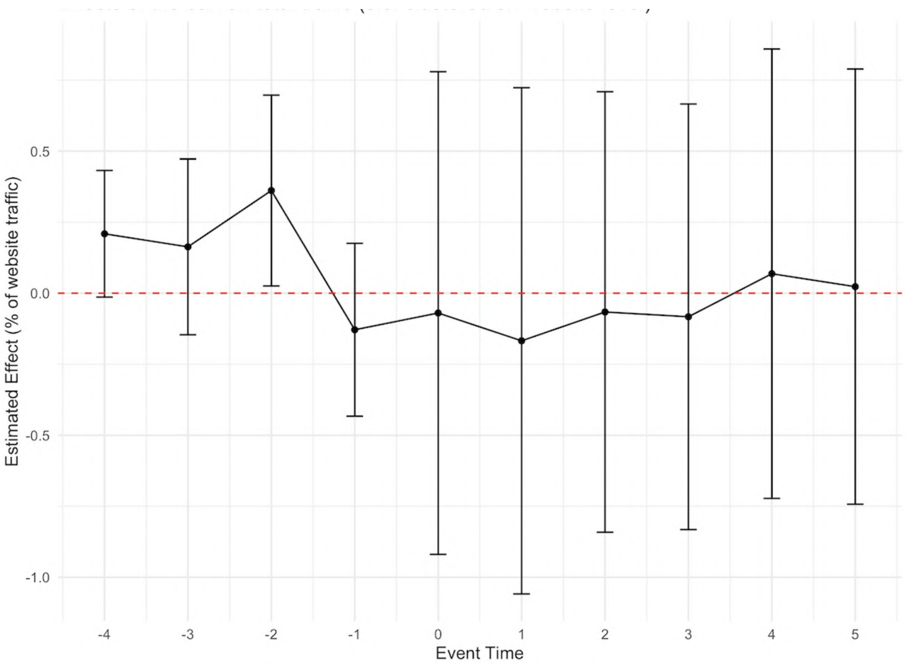


Figure 8: **Time-varying policy effect on website traffic**. 1. This figure presents the time-varying treatment effects on the total traffic for banned gambling websites, using unbanned gambling websites described in Table 10 as the control group. 2. We include linear time trends for each website and cluster the standard errors at the website level.

Figure 8 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 relatively 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 8 indicates that the policy did not have any significant impact on traffic to the gambling websites.

11 Conclusion

In this paper, we empirically study the impacts of Twitch's banning policy on content production and consumption across the platform. Based on 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.

We show that while the policy led to a persistent reduction in the supply of gambling livestreams, it also reduced the supply of non-gambling content, bringing unintended side effect on content production to the platform. Moreover, we find that the policy had more prominent impact on streamers with higher popularity, which can be explained by their low reliance on gambling content and concern over personal reputation. 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 findings provide managerial insights for influencers, platform developers and policymakers. First, we find that the policy may have a non-uniform impact on viewer demand, suggesting that content creators and platforms can use this information to take strategic actions to reduce losses from regulation policies. Second, our results show that regulation policies targeting a single content type may also reduce content production in other genres, alerting platforms to be cautious in content regulation to protect content diversity or profitability. Finally, we provide new evidence on the debate over the need for regulating gambling-like features in video games, which helps game developers refine their product design and policymakers evaluate welfare effects.

Our research is subjected to several limitation that point to potential directions for future work. First, it is plausible that streamers significantly affected by the policy may migrate to other platforms to continue streaming gambling content. While our dataset shows very few such cases among the top 6,000 streamers, it omits streamers with very low popularity and thus prevents us from fully discussing how the policy affected market exits among these streamers. Future research could benefit from datasets including an even broader range of streamers or streamer activities on other platforms, and examine how these policies influence competition with rival platforms. Second, due to the lack of viewer information, we are unable to address the policy's social impact, such as whether it lowered gambling content consumption among young viewers or improved mental health for those who heavily relied on the regulated content. Third, we focus on how the policy changed overall supply and demand of streaming content on Twitch, without deeply examining the underlying mechanisms behind substitution and competition among different types of content. Future studies could explore how content creators make decisions when facing policy interventions and investigate the generalizability of our results to other regulation policies in content markets. Fourth, content regulation may incentivize creators to distribute their content through more harmful, secret channels, as suggested by Wu (2024). In our context, this implies that regulating gambling livestreams may lead to more unlicensed gambling livestreams or prompt gambling websites to host gambling streams on their own websites. A potential direction for future research would be to explore whether content regulations on digital media platforms affect content distribution by content providers themselves, such as gambling websites or game developers.

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.

Angrist, J. D. and Pischke, J.-S. (2009), *Mostly harmless econometrics: An empiricist's companion*, Princeton university press.

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.

Aronow, P. M. and Samii, C. (2017), 'Estimating average causal effects under general interference, with application to a social network experiment', *Annals of Applied Statistics* **11**(4), 1912–1947.
URL: *https://doi.org/10.1214/16-AOAS1005*

Baker, S. R., Balthrop, J., Johnson, M. J., Kotter, J. D. and Pisciotta, K. (2024), Gambling away stability: Sports betting's impact on vulnerable households, Technical report, National Bureau of Economic Research.

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.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R. and Lefebvre, E. (2008), 'Fast unfolding of communities in large networks', *Journal of statistical mechanics: theory and experiment* **2008**(10), P10008.

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

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

Close, J., Spicer, S. G., Nicklin, L. L., Uther, M., Lloyd, J. and Lloyd, H. (2021), 'Secondary analysis of loot box data: Are high-spending "whales" wealthy gamers or problem gamblers?', *Addictive Behaviors* **117**, 106851.

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

Drummond, A. and Sauer, J. D. (2018), 'Video game loot boxes are psychologically akin to gambling', *Nature human behaviour* **2**(8), 530–532.

Fruchterman, T. M. and Reingold, E. M. (1991), 'Graph drawing by force-directed placement', *Software: Practice and experience* **21**(11), 1129–1164.

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.

Hollenbeck, B., Larsen, P. and Proserpio, D. (2024), 'The financial consequences of legalized sports gambling', *Available at SSRN* .

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* .

Iyer, G. and Singh, S. (2018), 'Voluntary product safety certification', *Management Science* **64**(2), 695–714.

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

Kristiansen, S. and Severin, M. C. (2020), ‘Loot box engagement and problem gambling among adolescent gamers: Findings from a national survey’, *Addictive Behaviors* **103**, 106254.

Larche, C. J., Chini, K., Lee, C., Dixon, M. J. and Fernandes, M. (2021), ‘Rare loot box rewards trigger larger arousal and reward responses, and greater urge to open more loot boxes’, *Journal of gambling studies* **37**, 141–163.

Li, S. and Wager, S. (2022), ‘Random graph asymptotics for treatment effect estimation under network interference’, *The Annals of Statistics* **50**(4), 2334–2358.

Liu, Y., Yildirim, P. and Zhang, Z. J. (2022), ‘Implications of revenue models and technology for content moderation strategies’, *Marketing Science* **41**(4), 831–847.

Moore, E. F. (1959), The shortest path through a maze, *in* ‘Proc. of the International Symposium on the Theory of Switching’, Harvard University Press, pp. 285–292.

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/.

Taylor, W., McCarthy, D. and Wilbur, K. C. (2024), ‘Online gambling policy effects on tax revenue and irresponsible gambling’, *Available at SSRN 4856684* .

Wu, Y. (2024), ‘Creation, consumption, and control of sensitive content’, *Marketing Science* **43**(4), 885–902.

Wu, Y., Gal-Or, E. and Geylani, T. (2022), ‘Regulating native advertising’, *Management Science* **68**(11), 8045–8061.

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

Zendle, D., Meyer, R. and Over, H. (2019), ‘Adolescents and loot boxes: Links with problem gambling and motivations for purchase’, *Royal Society open science* **6**(6), 190049.

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

Zhao, K., Lu, Y., Hu, Y. and Hong, Y. (2023), 'Direct and indirect spillovers from content providers' switching: Evidence from online livestreaming', *Information Systems Research* **34**(3), 847–866.

Zou, T. and Zhou, B. (2024), 'Self-preferencing and search neutrality in online retail platforms', *Management Science* .

Appendix

Table of Contents

A Proof of Proposition 1

The optimal quantities $\boldsymbol{q}^*$ are solved by the following Lagrangian:

$$\mathcal{L}(\boldsymbol{q}, \lambda, \boldsymbol{\mu}) = \sum_{k=1}^{3} \phi_k(q_k) - c_k(q_k) + \lambda(1 - \sum_k q_k) + \sum_k \mu_k q_k \tag{5}$$

Applying Kuhn-Tucker Conditions to (5) yield:

$$\begin{aligned}
&q_k\left(\frac{\partial \phi(q_k, \theta)}{\partial q_k} - \frac{\partial c_k(q_k)}{\partial q_k} - \lambda + \mu_k\right) = 0 \\
&\lambda\left(1 - \sum_k q_k\right) = 0 \\
&\mu_k q_k = 0
\end{aligned}$$

This system can be satisfied by the following equalities:

$$\begin{aligned}
&\frac{\partial \phi(q_1^*, \theta)}{\partial q_1} - \frac{\partial c_1(q_1^*)}{\partial q_1} = \frac{\partial \phi(q_2^*, \theta)}{\partial q_2} - \frac{\partial c_2(q_2^*)}{\partial q_2} = \frac{\partial \phi(q_3^*, \theta)}{\partial q_3} - \frac{\partial c_3(q_3^*)}{\partial q_3} = \lambda > 0 \\
&\mu_k = 0
\end{aligned}$$

as λ, the Lagrange multiplier for the total quantity constraint, is the shadow price and thus is positive. The first constraint binds and all the non-negativity constraints slack. Note that q_k^* are uniquely determined by the first equation as $\frac{\partial \phi(q_k,\theta)}{\partial q_k}$ strictly decreases in q_k and $\frac{\partial c_k(q_k)}{\partial q_k}$ weakly increases in q_k.

Similarly, the post-policy optimal quantities $\tilde{q}^*$ are solved by the following Lagrangian:

$$\mathcal{L}(\tilde{\boldsymbol{q}}, \tilde{\lambda}, \tilde{\boldsymbol{\mu}}) = \sum_{k=2}^{3} \phi_k(q_k) - c_k(q_k) + \tilde{\lambda}(1 - \sum_{k=2}^{3} q_k) + \sum_{k=2}^{3} \tilde{\mu}_k q_k \tag{6}$$

which can be solved by the following equalities:

$$\begin{aligned}
&\frac{\partial \phi(q_2^*, \theta)}{\partial q_2} - \frac{\partial c_2(q_2^*)}{\partial q_2} = \frac{\partial \phi(q_3^*, \theta)}{\partial q_3} - \frac{\partial c_3(q_3^*)}{\partial q_3} = \tilde{\mu}_k = 0 \\
&\tilde{\lambda} = 0
\end{aligned}$$

where all constraints slack. It is not possible that if (θ, c_1, c_2, c_3) do not change, the total quantity constraint still binds, otherwise $q_k^* = \tilde{q}_k^*$ for $k = 2, 3$ and $\tilde{q}_2^* + \tilde{q}_3^* < 1$, forming a contradiction.

Next, we discuss the results when $\Delta c_k = \tilde{c}_k - c_k > 0$, i.e. production cost for content k

increases after the policy. Note that $\frac{\partial^2\phi(q_k,\theta)}{\partial q_k^2} < 0$ and we have assumed a linear cost (the results hold for any cost function such that $\frac{\partial C_k(q_k)}{\partial q_k} > 0$ and $\frac{\partial^2 C_k(q_k)}{\partial q_k^2} \leq 0$). As illustrated in Figure, $q_k^* = \tilde{q}_k^*$ when $\tilde{c}_k = \tilde{c}_k^* = \frac{\partial\phi(q_k^*,\theta)}{\partial q_k}$, which also coincidence with $\Delta c_k^* = \lambda$ when we assume linear cost functions. Therefore, when $\tilde{c}_k > \tilde{c}_k^*$, cost increases too much such that $\tilde{q}_k^* < q_k^*$.

The third part of Proposition 1 is intuitive: since the revenue change for streamer i is measured by

$$\begin{aligned}
&\sum_k \pi_k(q_k^*) - \sum_k \pi_k(\tilde{q}_k^*) \\
&= (\tilde{q}_2^* p_2 + \tilde{q}_3^* p_3) - \sum_{k=1}^{3} q_k^* p_k \\
&= \sum_{k=2}^{3} (\tilde{q}_k^* - q_k^*) p_k - q_1^* p_1
\end{aligned}$$

, the sign of this term is dependent on the relative size of p_1, p_2 and p_3. Additionally, $\tilde{q}_k^*$ will decrease in Δc_k for $k = 2, 3$ as we have discussed above, making the threshold values r_{21}^* and r_{31}^* smaller.

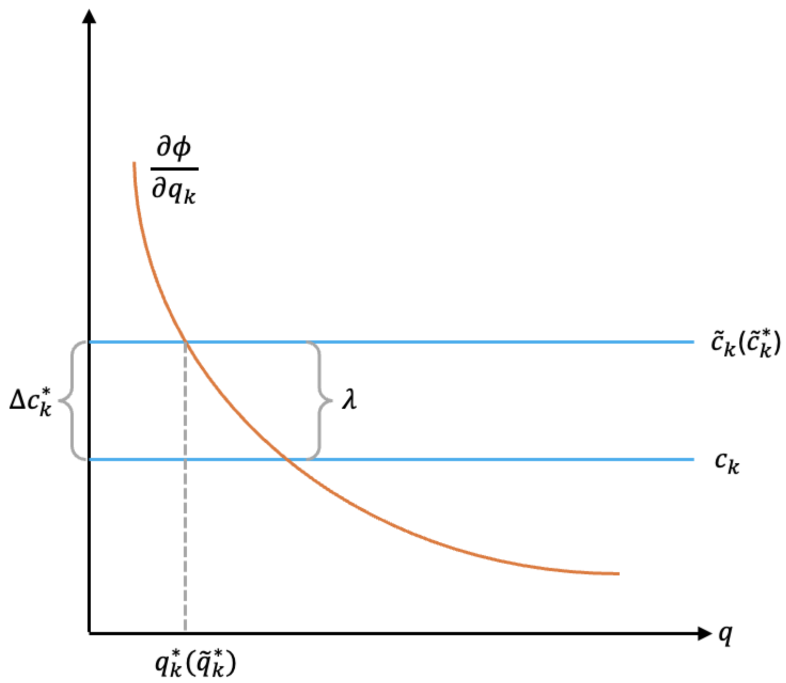


Figure A.0.1: **Figure of the optimal livestreaming choices**.

B 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 26, 2022 to August 20, 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 B.0.1 describes variables retrieved from the API at the time of each request.

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 [21]	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

Table B.0.1: **Variable definitions**.

The streamers are pre-selected from a list of 3,606,607 streamers who appear in a 1-month pilot study (September 21, 2022 and October 18, 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

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

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).

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 9 for details). Table B.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).

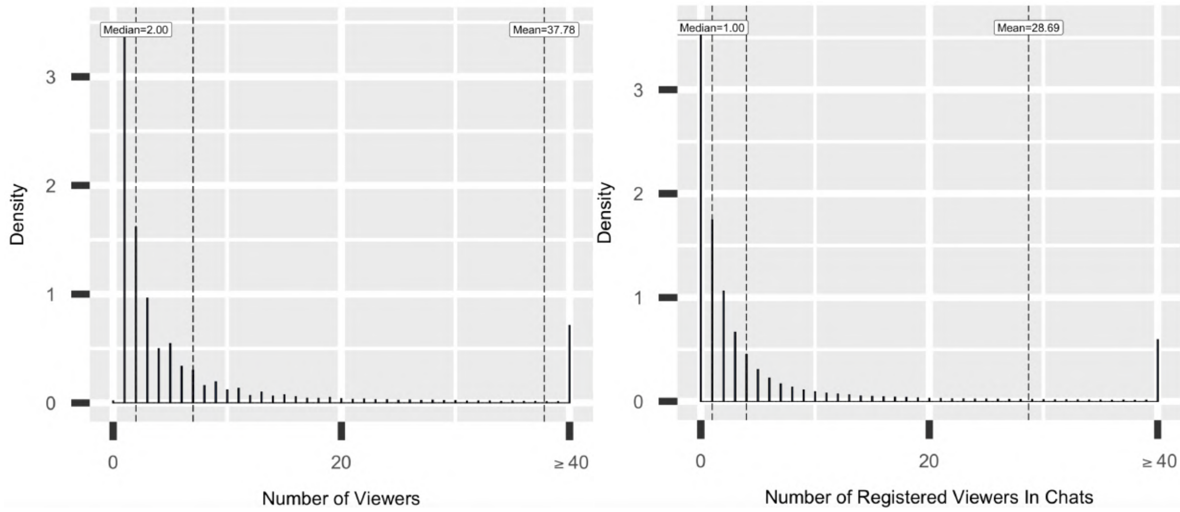


Figure B.0.1: **Distribution of Registered Viewers vs Viewers**

.

C Banned Content and Streamer Detection

C.1 Matching Video Clips to Original Streams

Since our detection of banned content is at a video clip level and video clips are clips from in-game streams, we need to match clips with their original streams to determine *the list of streams* with banned content. Since we can use our high-frequency data to back out the start and end time of a stream, we can facilitate the matching as long as we also know the exact start and end time of a clip. However, this creates a challenge since (i) not all clips have a video offset time from the Twitch API, which tells us how many seconds into a video (and thus stream) the start of the clip occurs, and (ii) a clip's "created time" extracted from the Twitch API is not necessarily the time when the actual content is broadcasted but its upload time. This can be an issue if there is a severe lag between when a viewer clips an in-stream moment and when she uploads the clip.

For clips that suffer from these issues, our strategy of matching relies on the fact that if a clip features a streamer's desktop time (see Figure C.1.1), the desktop time must reflect the *local time* the content in that clip is streamed. In that case, we can use OCR to extract the timestamp[22] from video frames and then use it to determine a clip's *local* start and end time. However, there could still be several concerns:

1. Sometimes multiple timestamps may appear across all sampled frame of a video. For example, this could happen when a frame captures a moment of a streamer with a chat log shown on the side, where chats on Twitch naturally come with timestamps. In that case, it creates difficulty in determining which timestamp is the desktop time.

2. Streamers may operate at different time zones, which can alter both the date and the hour component of a timestamp. For example, since Twitch records timestamps in UTC, a clip capturing a stream moment at 08/13/2022 19:45 in UTC (as recorded by Twitch) may appear as 08/13/2022 15:45 in EDT (as reflected by the desktop time shown in a video frame).

3. Streamers may have their desktop time displayed in different formats (e.g., DD/MM/YYYY vs MM/DD/YYYY, or 09:30 PM vs 21:30), which creates confusions in matching to their streams' timestamps.

We use several methods to address the above cases. First, we observe from OCR's extracted texts that if there are multiple timestamps extracted from a video frame, a streamer's desktop time always appears as the *last* timestamp in the list. This is because most streamers use

[22] The timestamp will be extracted by OCR as part of the texts, in a format that contains the date, hour and minute component of a timestamp

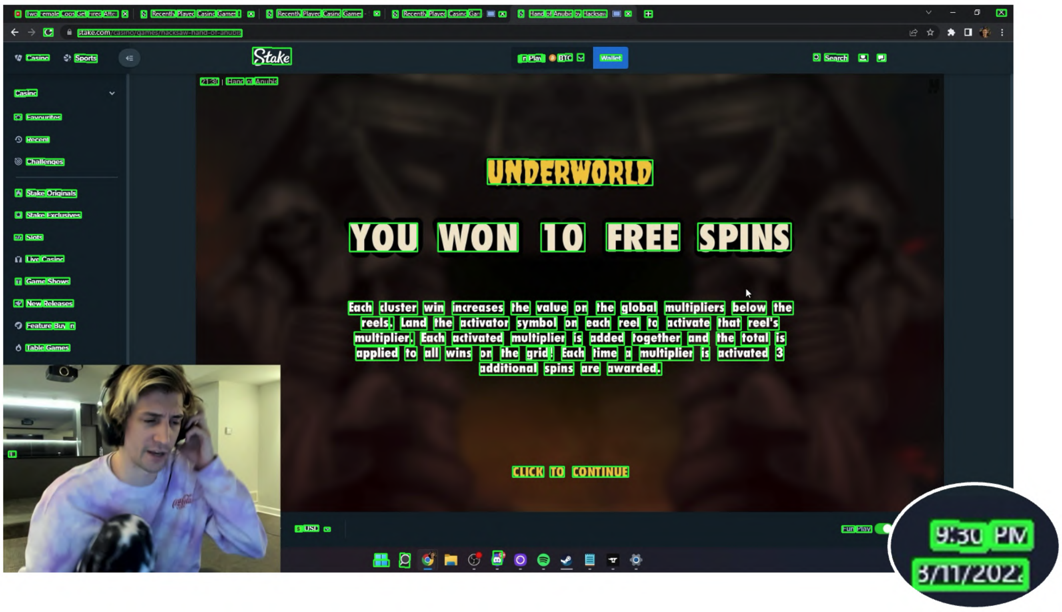


Figure C.1.1: **Desktop time during a broadcast**. This figure illustrates that OCR can detect a streamer's desktop time from video frames (if this information is displayed during the broadcast). In this example, it shows that XQC was broadcasting Stake at 9:30pm on August 11, 2022 at his local time.

windows systems, where the desktop time always appears in the lower right corner of the screen (see the zoomed-in region in Figure C.1.1). OCR detects text in reading order. Therefore, regardless of whether the frame is read by OCR from left to right or top to bottom, the desktop time always appears as the last timestamp in the list of extracted timestamps. This helps resolve the first concern.

Next, we check whether the desktop time shown in a clip matches with the clip's upload time recorded by Twitch. To achieve this, we use a 2-step matching approach. In the first step, we start by matching the two timestamps based on only the *minute* component of a timestamp, given potential concerns of time zone differences which could alter the date and hour components of a timestamp as discussed in the second concern. To facilitate matching, we convert both timestamps into the same 24-hour format (i.e., if AM/PM is present in the timestamp like 09:30PM, we convert it to 21:30). Then, we match the minute component of the two timestamps while allowing for discrepancy of up to 3 minutes between the timestamps. Since the desktop time may either capture the time at the start, middle or end of a clip, this small discrepancy is set to account for the maximum duration of a clip (i.e., 2 minutes) plus a minor time discrepancy due to transition of a minute (i.e., because desktop time does not reflect the second component, 9:30 PM can either be the start of 9:30 PM or close to 9:31PM). Our rationale here is that it is highly unlikely for the minute component of two timestamps to randomly fall within the same 3-minute range. Therefore, if there is a match, we are confident to proceed to the second step of the matching.

In the second step, we continue by verifying the *date and hour* component 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 (i) there is a match on the date, (ii) the hour difference of the two timestamps does not exceed 12 hours, and (ii) the clips fall within a streamer's local time zone. We determine a streamer's local time zone by (i) obtaining information about a streamer's country and city of residence from *Streams Charts*, and (ii) identifying the common hour difference of all clips of a streamer (based on clips' desktop time) relative to UTC and then infer a streamer's time zone based on this hour difference. All clips that satisfy the criteria and have clear start and end times are then matched to their original streams. In that case, if clips are found to contain banned content, the corresponding streams they are matched to are also indicated to contain banned content.

We note that, using the above procedure, some clips will be left out either if they do not contain a streamer's desktop time or cannot be matched based on timestamps. As discussed in Section 5, our goal is to minimize false positives and avoid misclassifications. If a banned stream is not captured based on one data source, we will use alternative sources for detection.

C.2 Banned Content Detection Using Stream Titles

As discussed in Section 5.2, our strategy of detecting streams and streamers with banned content from stream titles relies on checking if both (i) a banned website's referral link is found in the stream title, and (ii) the in-stream game featured at that time is also of online *gambling* content. To provide further support for this strategy, in Table C.2.1, we first provide side evidence by summarizing the number of streamers, with or without any history of streaming gambling content, who use banned *referral links* in stream titles under different scenarios. We have several key findings:

First, we note that by definition, "untreated" streamers should have no usage of banned website referral links in *gambling* streams, which is our used strategy. However, even for streams that do not contain gambling content, referral links of banned websites are also almost never used in stream titles by "untreated" streamers. Among 4,626 "untreated" streamers, only 2 streamers are found to include banned websites' referral links in their stream titles.[23] In contrast, among "treated" streamers, a lot more streamers (52 people) have such usages.

Second, within "treated" streamers, almost no streamers include referral links to banned

[23]We manually checked the video clips of these two streamers ("rigoncs" and "xhocii") on *Stream Charts* for the streams which they were found to use banned websites' referral links and confirmed that they were not streaming any gambling content at that time. Thus, there is no issue with our treatment group classification. We then checked when these referral links were used. We found that they only appear in a few consecutive streams in a short time period. It could be that these two streamers have a short-term contract with banned websites such that they put referral links in their titles while not needed to stream actual gambling.

websites in the titles of their non-gambling streams. Only 2 streamers ("pijack11" and "poderosobagual") are exceptions, and they are also the two exceptions that have banned websites' referral links in stream titles in post-treatment period which should not happen. We manually checked the usages of commands in streams by these two streamers and argue that these two exceptions do not weaken the validity of our strategy. We find that they consistently used banned websites' referral links in both gambling and non-gambling streams in pre-treatment periods. Additionally, they only used the referral links in streaming titles within *one day* after the policy is implemented, which could be because they did not immediately change their stream titles.

Table C.2.1 provides preliminary evidence that our strategy - aligned with the convention used by the Twitch community - could be reasonable at identifying banned streamers. To provide direct evidence, in Table C.2.2 Panel A, we validating our approach after classifying the "treated" streamers into banned and unbanned groups by showing that unbanned streamers do not include banned website referral links in their *gambling* stream titles. In general, we also do not use such commands in *any* stream titles.

We emphasize that our detection approach ("!" followed by banned website keywords) is more conservative compared to defining a stream (or streamer) as containing banned content solely based on the presence of banned website keywords in stream titles. In Table C.2.2 Panel B, we show that our proposed approach better minimizes concerns of false positives (i.e., misclassifying unbanned content as banned), though the less conservative definition also has reasonable good performance. Nevertheless, for readers concerned with how potential misclassifications (if any) may impact the results, we conduct a formal sensitivity analysis to quantify the impact in Appendix D. We also emphasize that we use multiple data sources (video clips, titles, and chats which we discuss next) for banned detection, and all banned streamers who are detected from stream titles are verified to be in the banned group by other sources (video clips or chats), thus further validating our detection approach (see Figure C.4.5 for details).

	"Treated" Streamers (Streamers with Gambling Content)	"Untreated" Streamers (Streamers without Gambling Content)
No. Streamers	475	4,626
Using banned website referral links		
Banned Referrals in Gambling Stream Titles (**Our Approach**)	50	0
Banned Referrals in Any Stream Titles	52	2
Banned Referrals in Any Stream Titles In Post-Treatment Period	2	0

Table C.2.1: **Usage of banned referral links in stream titles for treated and untreated streamers**.

	Banned Streamers	Unbanned Streamers	Untreated Streamers
No. Streamers	158	317	4,626
Panel A: Using banned website referral links			
Banned Referrals in Gambling Stream Titles (**Our Approach**)	50	0	0
Banned Referrals in Any Stream Titles	52	0	2
Banned Referrals in Any Stream Titles In Post-Treatment Period	2	0	0
Panel B: Using banned website keywords			
Banned Keywords in Gambling Stream Titles	78	7	0
Banned Keywords in Any Stream Titles	83	7	4
Banned Keywords in Any Stream Titles In Post-Treatment Period	4	0	1

Table C.2.2: **Usage of banned referral links and banned website keywords in stream titles for banned, unbanned, and untreated streamers**. Treated streamers are separated into banned and unbanned groups, where the groups reflect those after we finish classifying the streamers using all sources.

C.3 Justifications for Ground Truth Chat Sample Construction

As discussed in Section 5.3, we use multiple criteria to construct ground truth chats samples separately for banned and unbanned gambling streams (see Table C.3.1).

Banned Gambling Streams	Unbanned Gambling Streams
(i) The gambling streams must have video clips with banned website names detected at all 3 randomly sampled frames using OCR. (ii) The gambling stream titles must contain referral links to banned websites.	(i) The gambling streams must have video clips with no banned website names detected at all three randomly sampled frames using OCR. (ii) The gambling streams must result from a streamer with no banned content detected in any of his pre-policy gambling video clips using OCR. (iii) The gambling stream titles must neither contain a referral link nor a keyword of a banned website.

Table C.3.1: **Criteria used to construct ground truth chat samples for banned and unbanned gambling streams**.

While we use both sources (OCR and stream titles) to construct the ground truth chats sample for *banned* gambling streams to ensure accuracy, for *unbanned* gambling content, we emphasize that using the first source alone, i.e. OCR on gambling video clips for which criteria (i) and (ii) are based off, is not enough to guarantee that the stream has no banned content. OCR with no illegal website keywords detected in all three randomly sampled frames is only a necessary but insufficient condition for a stream to be unbanned. For example, a clip may still contain banned content, but OCR cannot detect it if a clip's frames do not feature banned website keywords. Thus, we need a second source, i.e., stream titles for which criteria (iii) is based off, to verify that our generated ground truth chat sample for unbanned is valid.

Having established previously that stream titles can be used to classify banned and unbanned content (see Appendix C.2), we now check stream titles for all gambling streams whose video clips already satisfy criteria (i) and (ii) for unbanned streams. We find that indeed almost no video clips (1/161584 = 0.0006% clips) contain referral links to banned websites (i.e., "!" followed banned website keywords). This also serves as a cross-check that using referral links is a valid approach for banned and unbanned stream classification. To ensure our ground truth sample is of high quality, we enforce an even stricter restriction as in criteria (iii) to allow neither banned websites' referral link nor keywords to appear in stream titles. In general, the multiple criteria we imposed ensure that we have ground truth chats samples of high quality for both banned and unbanned streams, which serve next as our benchmarks for predicting which gambling streams contain banned content based on in-stream chats (if these streams can neither be classified based on video clips or stream titles).

C.4 Banned Content Detection Using In-Stream Chats

C.4.1 Validation: Is it Sufficient to Use Referral Links to Banned Websites in Chats for Banned Content Classification?

We now test our detection strategy based on in-stream chats, which is built on the hypothesis that banned gambling streams will contain a higher proportion of referral links to banned websites chats compared to unbanned gambling streams.

First, we check the summary statistics of banned and unbanned ground truth chats samples (Table C.4.1). We find that banned gambling streams contain 100 times more referral links of banned websites — despite having fewer total chats. In addition, banned gambling streams have 130 times more unique viewers who posted referral links to banned websites in chats. Both statistics suggest that there is a difference in referral links of banned websites in chats between banned and unbanned streams.

	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

Table C.4.1: **Summary statistics of ground truth chats samples for banned and unbanned gambling streams**.

Given the difference, we next consider whether it is adequate to use a simple test statistic based on the number of referral links to banned websites per hour to classify gambling streams[24]. In Figure C.4.1, we first plot the probability density function and cumulative distribution function of referral links to banned websites used in chats per hour separately for banned and unbanned streams. We observe from the left panel that the distribution of referral links to banned websites in unbanned streams has a huge mass at zero, where banned streams have a significantly larger probability of containing more than one banned websites' 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 to banned websites in banned streams first-order statistically dominates that in unbanned streams. These visualizations support our hypothesis that using referral links of banned websites in chats could be a valid criteria for classifying banned and unbanned gambling streams.

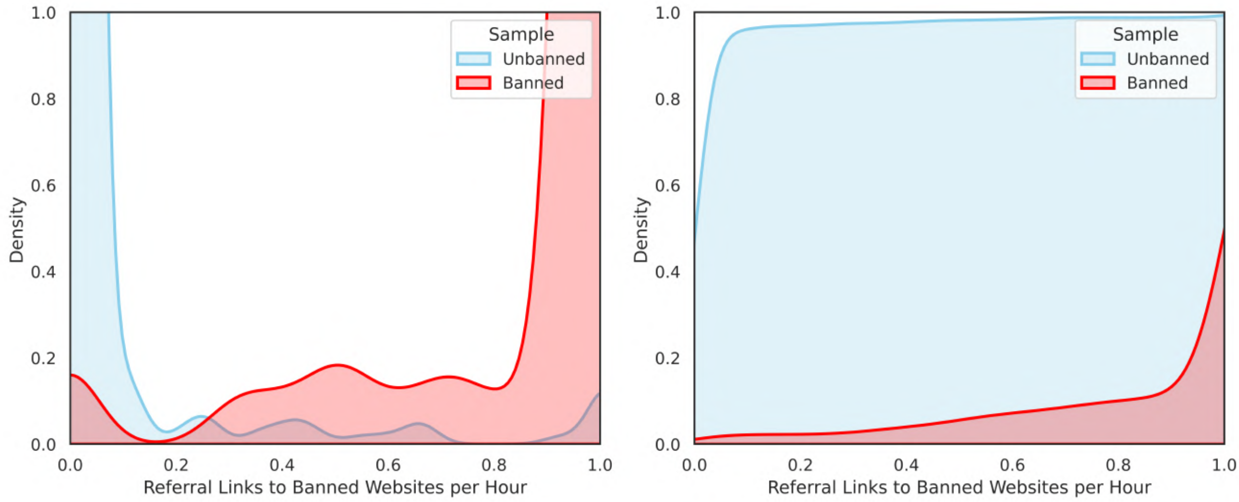


Figure C.4.1: **Referral links to banned websites per hour in chats for banned and unbanned gambling streams**. The left panel shows the density of banned referral links per hour in chats for banned (red) and unbanned (blue) streams, whereas the right panel shows the cumulative distributions of banned referral links per hour. We right-censored the number of banned referral links per hour at one in both panels.

Next, we evaluate whether there is a difference between the empirical distributions of the two samples formally by conducting a bootstrapped Kolmogorov-Smirnov (KS) test. In our case, the bootstrapped KS statistics is defined as the maximum absolute difference between two empirical distributions:

$$D_{n,m} = \sup_x |F_{\text{unbanned},n}(x) - F_{\text{banned},m}(x)| \tag{7}$$

[24]Using referral links per hour instead of total number of referral links in streams helps alleviate the concern that streams of longer duration tend to have more referral links.

$F_{\text{unbanned},n}$ and $F_{\text{banned},m}$ are the empirical distribution functions of the banned and unbanned chats samples respectively. We perform the test using the function from Bootstrapped KS2 Tester[25] by bootstrapping 5000 times with $m = 534$ observations for banned sample and $n = 821$ for unbanned sample. Figure C.4.2 shows the distribution of bootstrapped KS statistics. Since at any common levels of α,

$$D_{n,m} = 0.95 > c(\alpha)\sqrt{\frac{n+m}{nm}} = \sqrt{-\ln\left(\frac{\alpha}{2}\right) \times \frac{1}{2}}\sqrt{\frac{n+m}{nm}}$$

we always reject the null that the numbers of referral links per hour follow the same distribution in banned and unbanned streams.

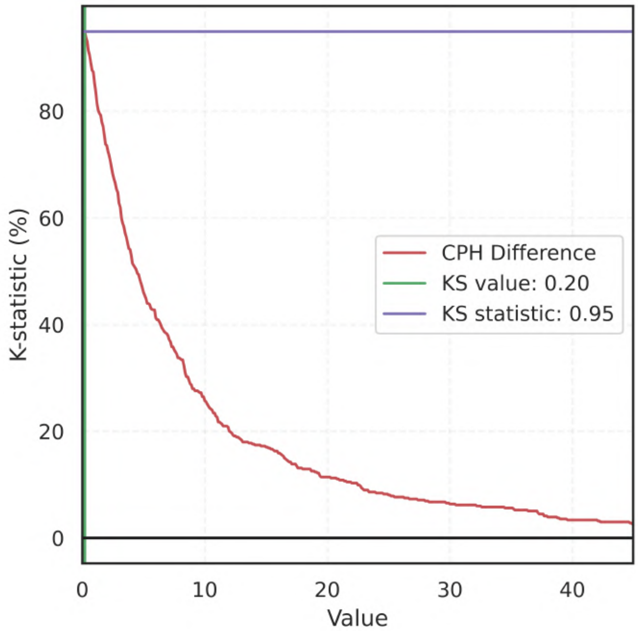


Figure C.4.2: **Distribution of bootstrapped Kolmogorov-Smirnov statistics.** The red line shows that the difference between the cumulative probability distributions of the banned and unbanned samples. The purple line shows the bootstrapped KS statistics, and the green line shows the value at which the distributional difference between the two samples is the largest.

Finally, having established the distributional difference, we conduct a preliminary test to check whether it is adequate to classify banned and unbanned streams based on the KS statistic. We use the KS value (0.2) as the threshold value and find that a simple threshold-based classification rule already achieves good out-of-sample predictive performance (i.e., Type I error =0.115, Type II error=0.010, which are important measures in our context since we want to minimize misclassifications). We also experimented with different threshold values and find that all achieve good predictive performance; the only difference is that a higher threshold

[25]https://github.com/swharden/Bootstrapped-KS2

tends to be conservative in terms of Type I errors but allows for relatively more Type II errors. With this, we conclude that a threshold-based classification is adequate for detecting banned and unbanned gambling streams, where our target is just to find an optimal threshold value to minimize the total errors.

C.4.2 Predictive Performance

Table C.4.2 summarizes the best threshold values for different measures and the corresponding in-sample and out-of-sample performances. We find that a threshold of 0.3 performs generally well across measures, in particular in terms of minimizing total errors, which is our main target. Thus, we propose to use 0.30 as the final threshold value to classify banned and unbanned gambling streams based on in-stream chats.

Measure	Best Threshold	Averaged In-Sample Results Across Folds	Averaged Out-Of-Sample Results Across Folds
Total Error	0.30	0.049	0.051
Total Squared Error	0.30	0.001	0.002
Weighted Total Squared Error	0.30	0.002	0.003
Type I Error	1.05	0.007	0.007
Type II Error	0.00	0.022	0.022
Accuracy	0.30	0.976	0.975
AUC	0.30	0.976	0.975
Precision	1.05	0.991	0.987
Recall	0.00	0.978	0.978
F1-Score	0.25	0.976	0.968

Table C.4.2: **Optimal thresholds and performance for different measures**. The total error is calculated as the sum of Type I and Type II errors. The total squared error is calculated as the sum of the squared Type I and Type II errors. A weighted version of the total squared error gives twice the weight to the squared Type I error compared to the squared Type II error. Precision is defined as the ratio of true positives to predicted positives, while recall is the ratio of true positives to actual positives. The F1 score is computed as the weighted average of precision and recall, given by the formula $F_1 = \frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}$.

Measure	Averaged Out-Of-Sample Results Across Folds
Total Error	0.051
Total Squared Error	0.002
Weighted Total Squared Error	0.003
Type I Error	0.024
Type II Error	0.026
Accuracy	0.975
AUC	0.975
Precision	0.964
Recall	0.974
F1-Score	0.968

Table C.4.3: **Predictive performance with 0.3 as the optimal threshold for all measures**. The total error is calculated as the sum of Type I and Type II errors. The total squared error is calculated as the sum of the squared Type I and Type II errors. A weighted version of the total error gives twice the weight to the squared Type I error compared to the squared Type II error. Precision is defined as the ratio of true positives to predicted positives, while recall is the ratio of true positives to actual positives. The F1 score is computed as the weighted average of precision and recall, given by the formula $F_1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$.

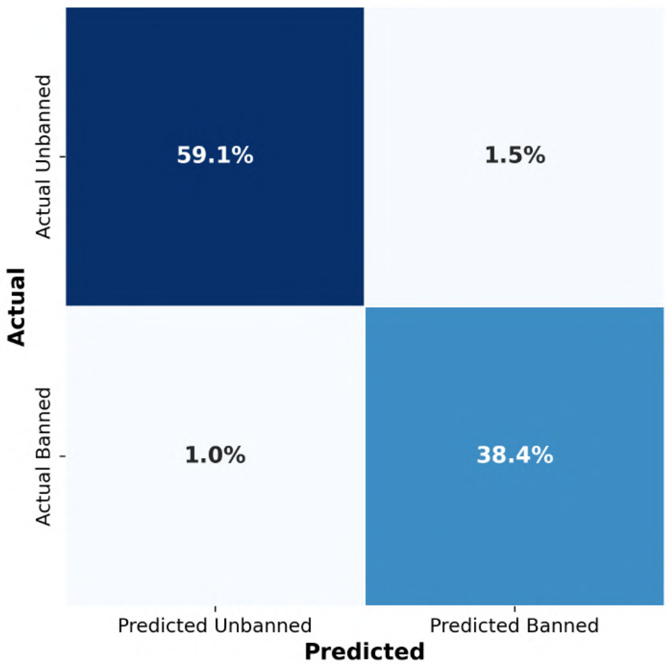


Figure C.4.3: **Aggregated confusion matrix across folds**.

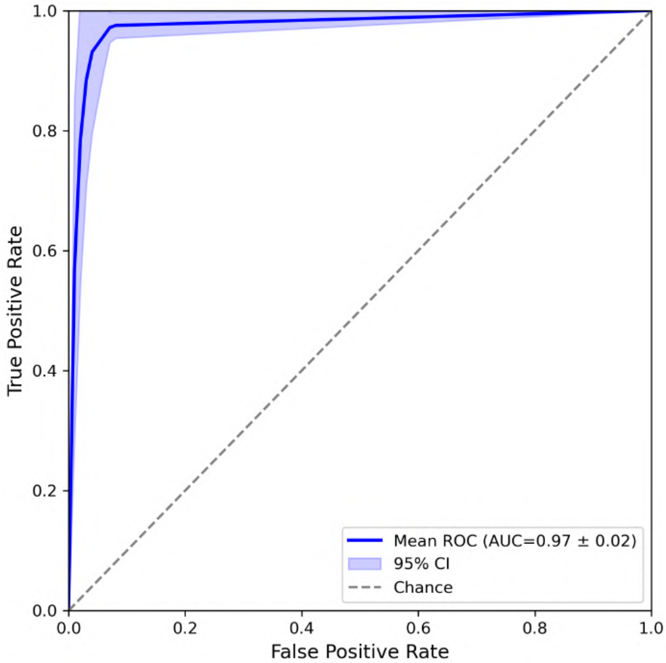


Figure C.4.4: **Average receiver operating characteristic (ROC) curve across folds**.

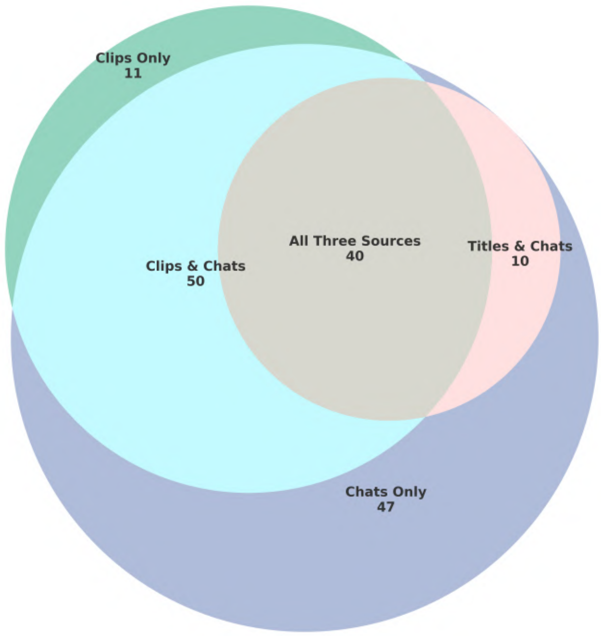


Figure C.4.5: **Detection of banned streamers by source type**. This figure shows the number of banned streamers detected by each source type alone and together. In total, we are able to detect 158 banned streamers using different sources.

D Robustness Check on Potential Misclassification in Treatment Assignments

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. To alleviate this concern, we propose a sensitivity analysis to address potential misclassification issues in our casual estimates. Since we have chosen a conservative approach to detect banned streamers, our groups of banned streamers can be viewed as a subset of streamers who actually streamed banned gambling content. Therefore, we focus on potential misclassification in the group of unbanned streamers. 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.

$$\frac{\#\,\{\text{streamers who streamed banned content} \mid \text{Classified as unbanned streamers}\}}{\#\,\{\text{streamers who streamed unbanned content} \mid \text{Classified as unbanned streamers}\}}$$

We use this sensitivity parameter to test our estimation results from for the group of unbanned 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, because more than half of streamers in the current unbanned group have shown solid evidence of not streaming banned gambling content in collected video clips, stream titles and in-stream chats. We provide a formal proof of identification with misclassification in the DiD framework and the validity of the sensitivity parameter in the following sections.

D.1 Technical 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[26], 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

[26]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.

$D_1 = D_2^* = 0$.

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

$$Y_t = \begin{cases} Y_t(1,0)D_1^*(1-D_2^*) + Y_t(0,1)(1-D_1^*)D_2^* + Y_t(0,0)(1-D_1^*)(1-D_2^*) & \text{, if } t=1 \\ Y_t(0,0) & \text{, if } t=0 \end{cases} \tag{8}$$

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^*, 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 2

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

and β_3 can be decomposed similarly.

As we have introduced in Section 5 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 3 *Suppose that Assumption 1 and Condition 1 hold. Then,*

$$
\begin{aligned}
\beta_1 =& \mathbb{E}[Y_1(1,0) - Y_1(0,0) \mid D_1^* = 1, D_2^* = 0, D_1 = 1, D_2 = 0] \\
\beta_3 =& \mathbb{E}[Y_1(1,0) - Y_1(0,0) \mid D_1^* = 1, D_2^* = 0, D_1 = 0, D_2 = 1] P(D_1^* = 1, D_2^* = 0 \mid D_1 = 0, D_2 = 1) + \\
& \mathbb{E}[Y_1(0,1) - Y_1(0,0) \mid D_1^* = 0, D_2^* = 1, D_1 = 0, D_2 = 1] P(D_1^* = 0, D_2^* = 1 \mid D_1 = 0, D_2 = 1) \\
=& ATT_1 \times P(D_1^* = 1, D_2^* = 0 \mid D_1 = 0, D_2 = 1) + ATT_2 \times P(D_1^* = 0, D_2^* = 1 \mid D_1 = 0, D_2 = 1)
\end{aligned}
\tag{9}
$$

The results stated in Proposition 3 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)

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

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 4 *Under Assumption 1, 2 and assume that Condition 1 holds, then*

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

D.2 Usage of the Sensitivity Parameter

We can then test the sensitivity of our causal estimates in Section 8 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 4, 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 4 are $(0, 2.427)$ and $(0, 1.465)$ respectively. Therefore, both of them are valid under the threat of 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 2

Note that

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

Plugging the expression of Y_t from (8), we have

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

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. □

D.3.2 Proof of Proposition 3

The results are derived directly by plugging the bounds of conditional probabilities into results of Proposition 3. □

D.3.3 Proof of Proposition 4

Note that with model (8), 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 2 yields the following bounds on the probability of correct classification conditional on actual second treatment:

$$\frac{1}{1+\Gamma} \leq P(D_1^* = 0, D_2^* = 1 \mid D_1 = 0, D_2 = 1) \leq \frac{1}{1+1/\Gamma} \tag{11}$$

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. □

E Additional Estimation Results

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

(a) Estimates for Gambling and LootBox Games

	log(Gambling hours + 1)		log(LootBox Games+ 1)	
	TWFE	SynthDiD	TWFE	SynthDiD
β_1	-1.022***	-1.001***	0.012	0.021
	(0.089)	(0.082)	(0.065)	(0.054)
β_2	0.166**	0.079	-0.004	0.011
	(0.074)	(0.070)	(0.054)	(0.050)
β_3	-0.116**	-0.130***	-0.046	-0.054
	(0.049)	(0.043)	(0.043)	(0.036)
β_4	0.133***	0.099	-0.026	-0.009
	(0.039)	(0.037)	(0.039)	(0.036)
Streamer FE	✓	✓	✓	✓
Week FE	✓	✓	✓	✓
Observations	112,222	112,222	112,222	112,222
Mean dependent variable	1.132	1.132	1.341	1.341

(b) Estimates for Streaming Hours and Other Games

	log(Streaming hours + 1)		log(Other Games+ 1)	
	TWFE	SynthDiD	TWFE	SynthDiD
β_1	-0.583***	-0.585***	-0.146***	-0.149***
	(0.084)	(0.075)	(0.04)	(0.032)
β_2	-0.038	-0.041	-0.13***	-0.098***
	(0.057)	(0.052)	(0.036)	(0.033)
β_3	-0.182***	-0.194***	-0.133**	-0.105***
	(0.044)	(0.040)	(0.032)	(0.034)
β_4	-0.065*	-0.062	-0.156***	-0.141***
	(0.039)	(0.037)	(0.028)	(0.026)
Streamer FE	✓	✓	✓	✓
Week FE	✓	✓	✓	✓
Observations	112,222	112,222	112,222	112,222
Mean dependent variable	2.724	2.724	0.541	0.541

Table E.0.1: **Full estimation results of supply-side outcomes**.

(a) Estimates for Hours Watched and Tier 1 Subscriptions

	log(Hours Watched + 1)		Tier 1	
	TWFE	SynthDiD	TWFE	SynthDiD
β_1	-1.694***	-1.696***	-0.59***	-0.584***
	(0.268)	(0.241)	(0.121)	(0.110)
β_2	-0.268	-0.272	0.046	0.040
	(0.173)	(0.160)	(0.087)	(0.078)
β_3	-0.523***	-0.554***	-0.178**	-0.185***
	(0.134)	(0.122)	(0.058)	(0.054)
β_4	-0.218*	-0.223	-0.181***	-0.171***
	(0.117)	(0.113)	(0.053)	(0.046)
Streamer FE	✓	✓	✓	✓
Week FE	✓	✓	✓	✓
Observations	112,222	112,222	112,222	112,222
Mean dependent variable	9.135	9.135	2.877	2.877

(b) Estimates for Tier 2 and Tier 3 Subscriptions

	Tier 2		Tier 3	
	TWFE	SynthDiD	TWFE	SynthDiD
β_1	0.01	0.017	-0.024	-0.024
	(0.013)	(0.011)	(0.015)	(0.014)
β_2	0.004	0.010	-0.018	-0.015
	(0.013)	(0.013)	(0.012)	(0.012)
β_3	0.012	0.013	0.013	0.011
	(0.011)	(0.009)	(0.012)	(0.011)
β_4	0.023**	0.027**	0.003	0.003
	(0.009)	(0.009)	(0.008)	(0.009)
Streamer FE	✓	✓	✓	✓
Week FE	✓	✓	✓	✓
Observations	112,222	112,222	112,222	112,222
Mean dependent variable	0.181	0.181	0.178	0.178

Table E.0.2: **Full estimation results of demand-side outcomes**.

(a) Supply-Side Outcomes

	Gambling	LootBox	Streaming Hours	Other Games
β_1	-1.362***	-0.135	-0.785***	-0.153*
	(0.164)	(0.147)	(0.181)	(0.079)
β_2	0.016	0.004	-0.05	-0.227***
	(0.132)	(0.124)	(0.136)	(0.077)
β_3	-0.065	0.008	-0.217***	-0.138**
	(0.07)	(0.081)	(0.072)	(0.055)
β_4	0.176***	0	-0.093	-0.265***
	(0.063)	(0.081)	(0.069)	(0.065)
Streamer FE	✓	✓	✓	✓
Week FE	✓	✓	✓	✓
Observation	10,494	10,494	10,494	10,494

(b) Demand-Side Outcomes

	Hours Watched	Tier 1	Tier 2	Tier 3
β_1	-2.18***	-0.634**	-0.017	-0.021
	(0.549)	(0.283)	(0.024)	(0.035)
β_2	-0.036	0.301	0.009	-0.016
	(0.402)	(0.208)	(0.024)	(0.026)
β_3	-0.764***	-0.214**	0.017	0.04**
	(0.244)	(0.102)	(0.02)	(0.018)
β_4	-0.328	-0.271***	0.043**	0.026
	(0.209)	(0.103)	(0.021)	(0.02)
Streamer FE	✓	✓	✓	✓
Week FE	✓	✓	✓	✓
Observation	10,494	10,494	10,494	10,494

Table E.0.3: **Full estimation results with selected groups of streamers**.

(a) Supply-Side Outcomes

	Gambling	LootBox	Streaming Hours	Other Games
β_1	-1.514***	0.095	-0.685***	-0.018
	(0.118)	(0.199)	(0.201)	(0.196)
β_2	-0.082	0.395	0.3	0.069
	(0.136)	(0.377)	(0.232)	(0.205)
β_3	-0.139	0.092	-0.227*	-0.128
	(0.18)	(0.103)	(0.136)	(0.091)
β_4	0.156	0.102	-0.044	-0.13
	(0.133)	(0.105)	(0.09)	(0.102)
Streamer FE	✓	✓	✓	✓
Week FE	✓	✓	✓	✓
Observation	946	946	946	946

(b) Demand-Side Outcomes

	Hours Watched	Tier 1	Tier 2	Tier 3
β_1	-1.942***	-0.763***	-0.004	-0.059
	(0.867)	(0.296)	(0.026)	(0.064)
β_2	0.554	0.147	-0.007	-0.015
	(0.507)	(0.189)	(0.027)	(0.021)
β_3	-0.961***	-0.485***	0.019	0.045*
	(0.535)	(0.282)	(0.057)	(0.027)
β_4	-0.67*	-0.689**	0.06	0.048
	(0.343)	(0.273)	(0.059)	(0.042)
Streamer FE	✓	✓	✓	✓
Week FE	✓	✓	✓	✓
Observation	946	946	946	946

Table E.0.4: **Full estimation results with selected streamer communities**.

F Additional Robustness Checks

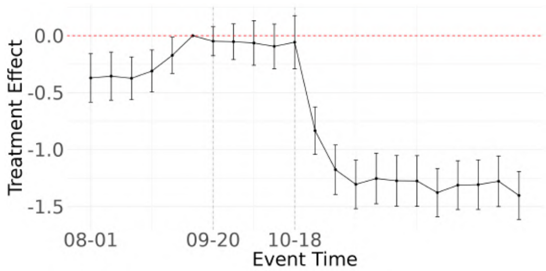
F.1 Event Studies

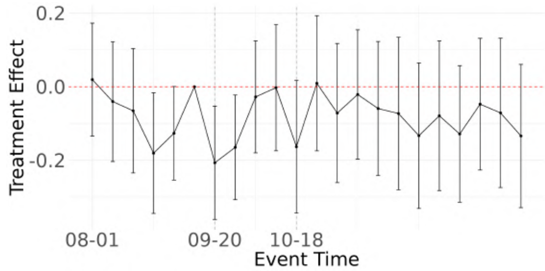
Given that the banning policy reduced the supply of gambling streams and caused spillover effects on unbanned gambling streams, we now investigate the persistence of the impact. The effect of the policy might diminish if treated streamers temporarily ceased gambling livestreams immediately after the 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. We normalize the point estimate one period before the policy announcement to zero, and report both the point-wise confidence intervals based on standard errors clustered at the streamer level.

Figure F.1.1 shows the estimates of time-varying treatment effects on the four types of content creation among banned streamers discussed in Section 8. First, we observe a small negative pre-trend in the supply of gambling content, but the magnitude is significantly smaller compared to the post-treatment estimates, thus do not invalidate our main results. Consistent with our estimates reported in Table 2, we find that the policy led to a persistent reduction in gambling livestreams after the policy implementation. Interestingly, we observe that the policy's negative impact increased in two weeks after the policy implementation and remained persistent afterwards. This result provides an additional evidence that the reduction in the 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. Second, panel (c) in Figure F.1.1 shows that the reduction in total livestreaming hours were persistent for banned streamers, while this reduction diminished for unbanned streamers in the last couple of periods in our panel. Finally, we find that the reductions in the supply of other games for both treated groups were estimated due to the disappearance of a positive pre-trend. We provide additional causal inference methods to test the robustness of these results.

F.2 Robustness Checks with Matched Sample

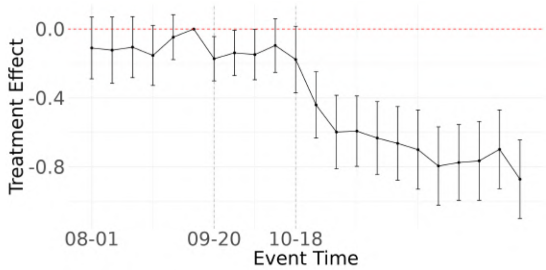
To test the robustness of our empirical results, we perform two additional checks. First, our treated and control groups differ significantly in size, and concerns may arise about selection issues due to the nonrandom choice of streaming gambling content. To mitigate these concerns, we use a matched sample to reconduct our main analysis. In marketing research, the matched samples are often generated through Propensity Score Matching (Angrist and Pischke, 2009), which allows researchers to match treated and control units based on the estimated probability

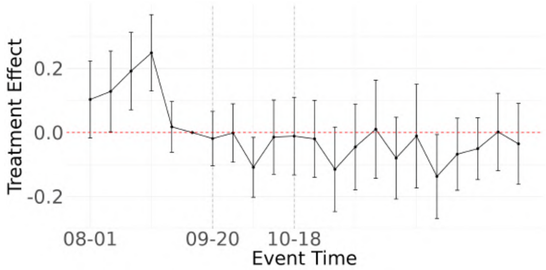




(a) Gambling Content

(b) LootBox Games

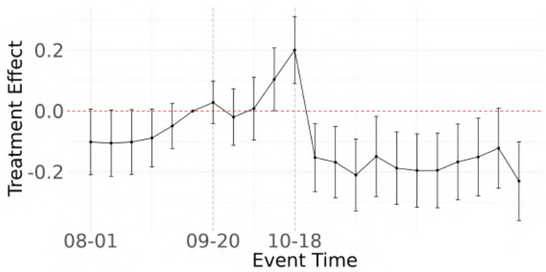


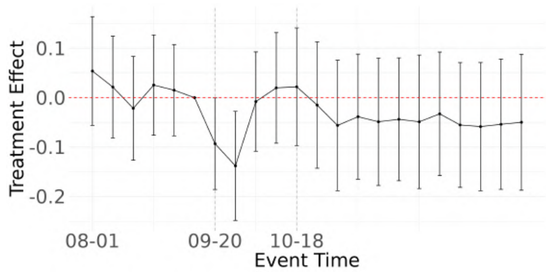


(c) Total Livestreams

(d) Other Games

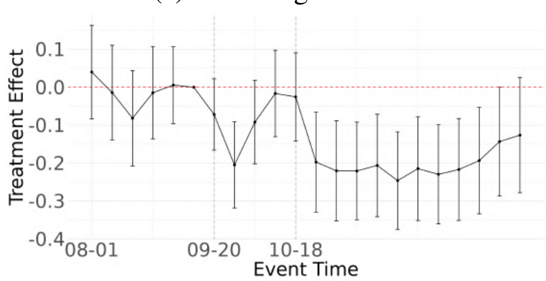
Figure F.1.1: **Time-varying treatment effects on content creation in group of banned streamers**. 1. The coefficients in each subfigure show the point estimates of β_1. 2. β_1 in one period ahead of the policy announcement is normalized to zero. 3. The 95% point-wise confidence intervals are illustrated as the inner bars, and the standard errors are clustered at the streamer level.

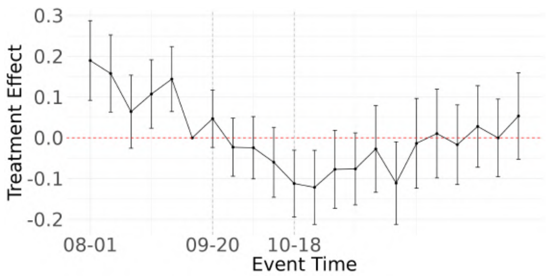




(a) Gambling Content

(b) LootBox Games

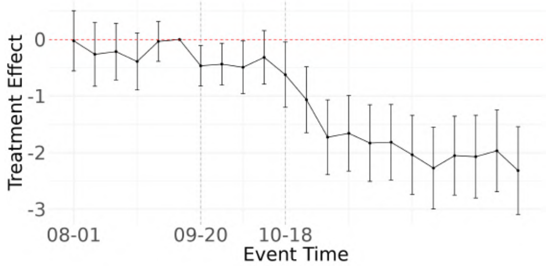




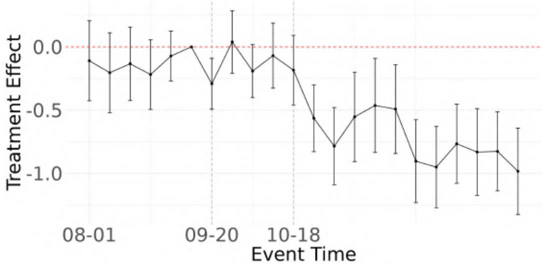
(c) Total Livestreams

(d) Other Games

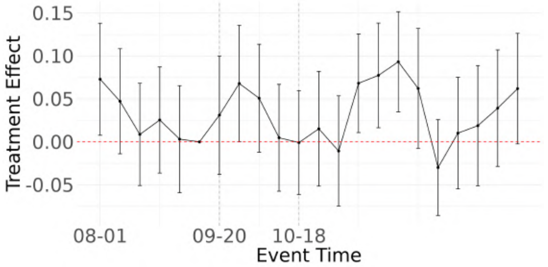
Figure F.1.2: **Time-varying treatment effects on content creation in group of unbanned streamers**. The coefficients in each subfigure show the point estimates of β_3. 2. β_3 in one period ahead of the policy announcement is normalized to zero. 3. The 95% point-wise confidence intervals are illustrated as the inner bars, and the standard errors are clustered at the streamer level.



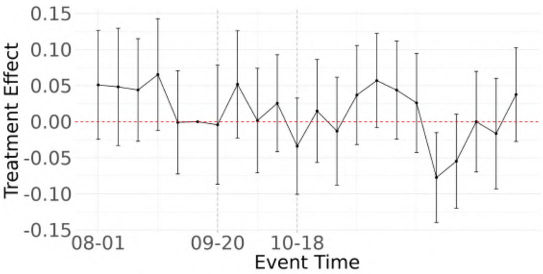
(a) Total Hours Watched



(b) Tier 1 Subscription

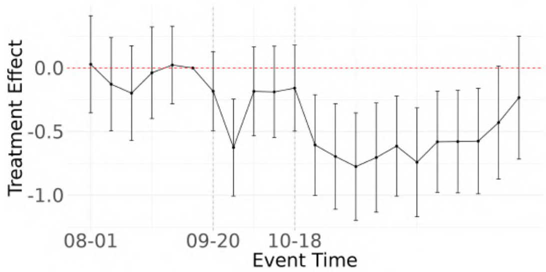


(c) Tier 2 Subscription

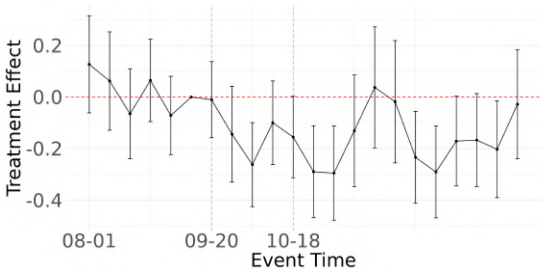


(d) Tier 3 Subscription

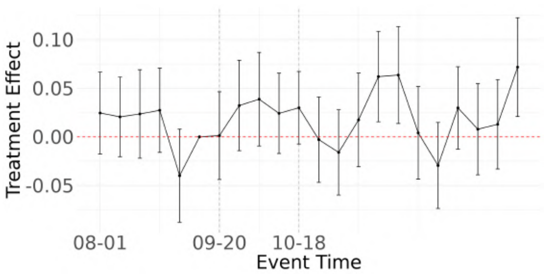
Figure F.1.3: **Time-varying treatment effects on content consumption in group of banned streamers**. 1. The coefficients in each subfigure show the point estimates of β_1. 2. β_1 in one period ahead of the policy announcement is normalized to zero. 3. The 95% point-wise confidence intervals are illustrated as the inner bars, and the standard errors are clustered at the streamer level.



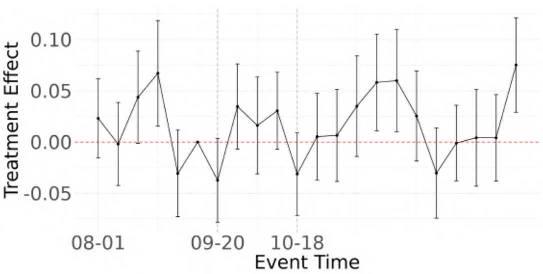
(a) Total Hours Watched



(b) Tier 1 Subscription



(c) Tier 2 Subscription



(d) Tier 3 Subscription

Figure F.1.4: **Time-varying treatment effects on content consumption in group of unbanned streamers**. The coefficients in each subfigure show the point estimates of β_3. 2. β_3 in one period ahead of the policy announcement is normalized to zero. 3. The 95% point-wise confidence intervals are illustrated as the inner bars, and the standard errors are clustered at the streamer level.

of receiving the treatment conditional on a given set of covariates. We begin by applying nearest-neighbor matching with a logit model. However, one may also argue that if a streamer had never streamed any gambling content before the policy implementation, her probability of being treated was always zero. To handle this, we also apply Coarsened Exact Matching (CEM), which matches units into stratas based on a set of coarsened covariates. In both methods, we match units using pre-treatment observations of weekly hours watched by viewers, streaming hours of games with and without loot boxes, and of non-gaming content to ensure that matched streamers have both similar streaming schedules and popularity. Table F.2.1 shows the balance check after applying each matching algorithm, and we find that each covariate is well balanced in terms of both mean and distribution.

	PSM				CEM			
	Banned		Unbanned		Banned		Unbanned	
	Mean Diff	Var Ratio	Mean Diff	Var Ratio	Mean Diff	Var Ratio	Mean Diff	Var Ratio
Hrs Watched	0.091	1.112	0.017	0.981	0.006	1.005	0.003	1.000
Loot Box Hrs	0.150	1.140	0.089	1.025	-0.000	1.005	-0.000	1.001
Other Games Hrs	-0.059	0.807	-0.061	0.841	-0.004	0.998	-0.002	0.998
Nongaming Hrs	-0.023	0.964	-0.091	0.876	-0.005	0.998	0.000	1.000

Table F.2.1: **Summary statistics of matched samples**. The table reports standardized mean differences and variance ratios of covariates between treated and control units based on each matching method. The mean differences are calculated by subtracting the mean of the matched control units from that of the matched treated units. All covariates are log-transformed.

Table F.2.2 reports the TWFE-DiD estimates based on matched units for each algorithm, using the same outcome variables as our main analyses. All findings are qualitatively unchanged, except that we now observe a small negative effect (less than 3%) on the most expensive subscription choice for banned streamers after the policy. However, given that this impact is significantly smaller than the effect on the cheapest subscription, these results still support our finding that the policy had minimal or no impact on the more loyal viewers of the treated streamers.

(a) PSM Sample

	Supply				Demand			
	Gambling	Loot Box	Streaming	Others	Hrs Watched	Tier 1	Tier 2	Tier 3
β_1	-1.019***	-0.024	-0.625***	-0.131***	-1.805***	-0.66***	0.001	-0.033**
	(0.089)	(0.067)	(0.086)	(0.042)	(0.274)	(0.125)	(0.015)	(0.016)
β_2	0.166**	-0.051	-0.066	-0.096**	-0.333*	0.023	-0.008	-0.023*
	(0.074)	(0.056)	(0.059)	(0.039)	(0.181)	(0.092)	(0.015)	(0.013)
β_3	-0.113**	-0.037	-0.143***	-0.186***	-0.513***	-0.178***	0.015	0.016
	(0.049)	(0.045)	(0.034)	(0.046)	(0.139)	(0.061)	(0.011)	(0.012)
β_4	0.133***	-0.031	-0.156***	-0.066	-0.142	-0.151***	0.03***	0.01
	(0.039)	(0.041)	(0.03)	(0.041)	(0.123)	(0.056)	(0.01)	(0.009)
Streamer FE	✓	✓	✓	✓	✓	✓	✓	✓
Week FE	✓	✓	✓	✓	✓	✓	✓	✓

(b) CEM Sample

	Supply				Demand			
	Gambling	Loot Box	Streaming	Others	Hrs Watched	Tier 1	Tier 2	Tier 3
β_1	-1.019***	0.012	-0.583***	-0.145***	-1.692***	-0.594***	0.008	-0.025*
	(0.089)	(0.065)	(0.084)	(0.04)	(0.268)	(0.121)	(0.013)	(0.015)
β_2	0.166**	-0.007	-0.129***	-0.04	-0.273	0.042	0.004	-0.018
	(0.074)	(0.054)	(0.037)	(0.057)	(0.173)	(0.087)	(0.013)	(0.012)
β_3	-0.113**	-0.048	-0.182***	-0.131***	-0.522***	-0.181***	0.011	0.012
	(0.049)	(0.043)	(0.044)	(0.032)	(0.134)	(0.058)	(0.011)	(0.012)
β_4	0.133***	-0.03	-0.155***	-0.067*	-0.221*	-0.183***	0.023**	0.003
	(0.039)	(0.039)	(0.028)	(0.039)	(0.118)	(0.053)	(0.009)	(0.008)
Streamer FE	✓	✓	✓	✓	✓	✓	✓	✓
Week FE	✓	✓	✓	✓	✓	✓	✓	✓

Table F.2.2: **Full estimation results with matched sample**.

G Network Analysis and Community Selection

As explained in Section 9.2.2, we use the Louvain method (Blondel et al., 2008) to identify 55 communities from the network as listed in Table G.0.1. For each treatment group, we want to select a subset of streamers such that there is no significant across-group interference. However, it is not straightforward to select streamers for the banned and unbanned groups since most of them are within communities that are mixed in nature in terms of treatment status. These are the communities which we need the automated approach as described in Section 9.2.2 to select subgroups of streamers of the same treatment status from.

To determine which mixed-type communities to use for each treatment group, we use a set of simple criteria. First, among these mixed-type communities, we exclude very small communities such as community 32 and 51 (i.e., communities that do not have at least 10 streamers) from consideration, since they lack the size necessary to extract meaningful subgroups of homogeneous streamers from. We also exclude communities where streamers of different treatment status are highly interwoven such as community 8 and 16 (i.e., variance of the shares of banned, unbanned and untreated less than a threshold of 0.05), as it is operationally difficult to disentangle streamers of the same status from others to ensure low interference.

Then, from the remaining mixed-type communities, we prioritize the selection of banned and unbanned streamers over untreated streamers when it comes to determine streamers of which treatment status to select from each community. More specifically, since we can only select streamers of one treatment status from each community to avoid interference across treatment status, we prioritize the selection of banned streamers from a community (and thus determine the starting node as a banned streamer and iteratively select banned streamers using the algorithm proposed in Section 9.2.2) if the community consists of a substantial share of banned streamers (at least 20%). After addressing communities with a large share of banned streamers, we apply the same logic to select unbanned streamers in the remaining communities. Finally, the rest of mixed-type communities are assigned for untreated streamers.

Table G.0.2 shows the final corresponding communities used for each treatment status. Since there are naturally fewer banned and unbanned streamers compared to untreated in our sample, this selection approach is necessary to ensure that our final subgroups contain enough banned and unbanned streamers to main a good representation of treatment statues as in our original sample.

Community ID	Community Size	Untreated Share (%)	Unbanned Share (%)	Banned Share (%)
0	3	0.00	66.67	33.33
1	204	74.51	9.31	16.18
2	7	100.00	0.00	0.00
3	4	25.00	75.00	0.00
4	3	0.00	100.00	0.00
5	2	50.00	50.00	0.00
6	39	89.74	2.56	7.69
7	3	100.00	0.00	0.00
8	118	33.05	38.14	28.81
9	2	50.00	50.00	0.00
10	42	73.81	0.00	26.19
11	4	100.00	0.00	0.00
12	8	25.00	75.00	0.00
13	75	8.00	42.67	49.33
14	46	100.00	0.00	0.00
15	13	100.00	0.00	0.00
16	39	33.33	43.59	23.08
17	3	100.00	0.00	0.00
18	116	25.86	63.79	10.34
19	17	100.00	0.00	0.00
20	2	100.00	0.00	0.00
21	20	100.00	0.00	0.00
22	5	100.00	0.00	0.00
23	2	100.00	0.00	0.00
24	2	50.00	0.00	50.00
25	3	100.00	0.00	0.00
26	16	93.75	0.00	6.25
27	2	0.00	50.00	50.00
28	2	100.00	0.00	0.00
29	19	100.00	0.00	0.00
30	2	0.00	0.00	100.00
31	2	100.00	0.00	0.00
32	4	25.00	50.00	25.00
33	2	100.00	0.00	0.00
34	4	100.00	0.00	0.00

(To be continued)

Community ID	Community Size	Untreated Share (%)	Unbanned Share (%)	Banned Share (%)
35	2	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
41	7	100.00	0.00	0.00
42	3	100.00	0.00	0.00
43	2	100.00	0.00	0.00
44	2	100.00	0.00	0.00
45	2	100.00	0.00	0.00
46	3	100.00	0.00	0.00
47	2	50.00	50.00	0.00
48	2	100.00	0.00	0.00
49	19	42.11	57.89	0.00
50	2	0.00	100.00	0.00
51	4	50.00	50.00	0.00
52	2	0.00	100.00	0.00
53	2	0.00	100.00	0.00
54	2	50.00	50.00	0.00

Table G.0.1: **Community clusters and treatment status**. This table presents the 55 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 9.2.

Panel A: Communities for Banned Group				
Community ID	Center	Community Size	Streamers Kept	Banned Share (%)
13	jonvlogs	75	37	49.33
10	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 (%)
18	zloyn	116	73	62.93
49	ilgabbrone	19	10	52.63
Others (4 clusters)	-	9	9	100.00

Panel C: Communities for Untreated Group				
Community ID	Center	Community Size	Streamers Kept	Untreated Share (%)
1	nyanners	204	98	48.04
6	trymacs	39	35	89.74
26	meduska	16	15	93.75
Others (30 clusters)	-	187	187	100.00

Table G.0.2: **Community clusters by treatment status**. 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 2, 7, 11, 14, 15, 17, 19, 20, 21, 22, 23, 25, 28, 29, 31, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 48 are solely of untreated streamers; 4,50,52,53 are solely unbanned streamers; 30 is solely banned streamers.