The Impact of Recommender Systems on Content Consumption and Production: Evidence from Field Experiments and Structural Modeling

Zhiyu Zeng,[1] Zhiqi Zhang,[1], Dennis J. Zhang,[1], Tat Chan[1]
1. Olin Business School, Washington University in St. Louis

Online content-sharing platforms such as TikTok and Facebook have become integral to daily life, leveraging complex algorithms to recommend user-generated content (UGC) to other users. While prior research and industry efforts have primarily focused on designing recommender systems to enhance users' content consumption, the impact of recommender systems on content production remains understudied. To address this gap, we conducted a randomized field experiment on one of the world's largest video-sharing platforms. We manipulated the algorithm's recommendation of creators based on their popularity, excluding a subset of highly popular creators' content from being recommended to the treatment group. Our experimental results indicate that recommending content from less popular creators led to a significant 1.34% decrease in video-watching time but a significant 2.71% increase in the number of videos uploaded by treated users. This highlights a critical trade-off in designing recommender systems: popular creator recommendations boost consumption but reduce production. To optimize recommendations, we developed a structural model wherein users' choices between content consumption and production are inversely affected by recommended creators' popularity. Counterfactual analyses based on our structural estimation reveal that the optimal strategy often involves recommending less popular content to enhance production, challenging current industry practices. Thus, a balanced approach in designing recommender systems is essential to simultaneously foster content consumption and production.

Key words: User-Generated Content, Recommender System, Field Experiment, Structural Model, Platform Economics

---

1. Introduction

Online content-sharing platforms like Facebook, YouTube, and TikTok are increasingly pivotal in society.[1] As of February 2024, Meta Inc., Facebook's parent company, connects over 3 billion people worldwide, or about 38% of the global population, through various social products, with many users spending hours each day on these products.[2] These platforms have drastically changed how we express ourselves, learn, and connect with people everywhere (Luca 2015, Baym 2021). They're

[1] https://www.pewresearch.org/internet/fact-sheet/social-media/

[2] https://investor.fb.com/home/default.aspx

not just for personal use; they also play a crucial role in how businesses and brands interact with their customers (Chatterjee and Zhou 2021, Zhang and Luo 2023), underscoring the significant impact these digital spaces have for the global economy.

These platforms, unlike traditional media, have two distinct features. First, most of the content consumed is generated by users themselves, known as user-generated content (UGC). This makes these platforms function as two-sided markets, where users can simultaneously be both content consumers and content creators (also known as content producers). Second, these platforms handle the daily upload of hundreds of millions of content pieces, while each user can only consume up to hundreds of content pieces daily. This means they need to use algorithmic recommender systems to pick a limited number of content pieces to show to each user. For example, on YouTube, the largest video-sharing platform, around 3.7 million new videos are uploaded daily in 2024.[3] YouTube used a sophisticated recommender system to determine which videos to display to each user, and 70% of the platform's video watch time was attributed to recommended videos.[4]

Therefore, recommender systems are crucial for sorting through vast amounts of content to tailor user experiences and significantly influence user behavior on these platforms. They serve as a critical tool for these platforms and are the subject of extensive research (Davidson et al. 2010, Wang et al. 2019). Both industry and academic research have focused on refining recommender systems to enhance user engagement and content consumption (Davidson et al. 2010, Wang et al. 2019). For example, YouTube has significantly improved its recommender system's machine learning architecture to increase the duration users spend watching content (Davidson et al. 2010). In academia, extensive research has been dedicated to enhancing user engagement and consumption behavior by optimizing various metrics, including click-through rates (Liu et al. 2010, Wang et al. 2019, Rafieian 2023), watch time (Covington et al. 2016, Zheng et al. 2022), dwell time (Zou et al. 2019), and visiting frequency (Xue et al. 2023). To achieve these goals, recommender systems primarily promote content that is popular and high-quality while also accommodating individual preferences when interacting with each user. This pattern of *prioritizing highly popular content* is also evident on the platform we collaborated with, a large online video-sharing platform referred to as "Platform V." We provide detailed evidence of this pattern in Online Appendix A.

While understanding the influence of recommender systems on users' content engagement and consumption is crucial, examining their impact on content creation is equally, if not more, important for two main reasons.[5] First, content production is vital for the long-term prosperity of a platform. Sustainable growth requires sufficient content production, as platforms rely on users to generate

[3] https://www.wyzowl.com/youtube-stats/

[4] https://blog.hootsuite.com/how-the-youtube-algorithm-works/

[5] Throughout the paper, we will use the terms content production and content creation interchangeably.

new content for others to consume (Zhang et al. 2012). Moreover, content production and sharing can help users develop friendships and foster a sense of community and social belonging (Daugherty et al. 2008, Pedroni et al. 2014), which are important for their long-term retention on the platform. Second and more importantly, recommending highly popular content, while helpful for content consumption, could potentially backfire on content production. Recommending highly popular creators can lead users to believe that the content-sharing marketplace is highly competitive, which may discourage them from producing content.

Overall, the impact of recommender systems on content creation is an important yet understudied question. This impact can lead to a notable trade-off between content consumption and creation when designing recommender systems. This naturally leads to our main research questions. First, how does recommending popular (or less popular) content creators simultaneously change users' content consumption and creation? Second, and more importantly, if the recommender system simultaneously impacts content consumption and creation, how can we design an optimal level of recommended popularity to balance users' content consumption and creation?

To study the impact of recommender systems on both user content consumption and production, we conducted a randomized field experiment with Platform V from January 8 to January 15, 2021. Platform V was using cutting-edge deep-learning-based recommender systems to recommend personalized content to each user (similar to the one used by YouTube (Covington et al. 2016)), and it primarily uses a creator's number of followers to measure that creator' popularity. In this experiment, we selected a small subset of highly popular creators (less than 0.05% of all creators), referred to as "Blocked Creators," based on creators' data prior to the experiment. We then randomly assigned 10% of Platform V's users to a treatment group and another 10% to a control group. While users in both the treatment and control groups were served by the same unchanged cutting-edge recommender system, the system did not include blocked creators' content when recommending to treated users. The goal of this treatment was to lower the average popularity of recommended creators for the treatment users compared to the control users. We observed a 25.64% reduction in the average popularity of recommended creators in the treatment group compared to the control group, confirming the efficacy of our intervention. Moreover, we conducted extensive randomization checks to ensure that the treatment group was comparable to the control group, indicating that the reduction in the average popularity of recommended creators was due to our intervention.

Our large-scale randomized field experiment on Platform V provided key insights into the impact of recommender systems on users' content consumption and production. Treated users, who received recommendations with a reduced average popularity of creators, significantly decreased their content consumption, measured by total time spent watching videos, by 1.34% compared to control users ($p<0.0001$). Interestingly, in contrast, treated users significantly increased their content production, measured by the daily number of videos uploaded, by 2.71% ($p<0.001$). While our intervention of recommending less popular creators increased treated users' content production, it did not affect their own popularity or the subsequent engagement levels of their uploaded videos.

The findings from our experiment highlight a critical trade-off between video consumption and production, prompting us to answer our second main research question—how recommender systems can be designed to maximize platform benefits considering both aspects. To achieve this, we developed a structural model that allows users to allocate their time between watching and creating videos, linking their marginal utility for content consumption and production to the popularity of recommended creators and their own popularity. Specifically, in our model, users' consumption is affected by the popularity of recommended creators, highlighting the effect of recommendations on user consumption. Additionally, the model allows a user's incentive to create videos to be influenced by the relative popularity of recommended creators and the focal user, emphasizing the effect of recommendations on user creation. Recognizing the diversity in user preferences over time, our structural model also categorizes users into multiple latent types and includes time-dependent random effects. This model is estimated through simulated maximum likelihood estimation with the Geweke-Hajivassiliou-Keane (GHK) sampling method (Geweke et al. 1994, Hajivassiliou and Ruud 1994, Keane 1994) to decompose the time-dependent random effects, using data from all eight days of our experimental period.

The estimation results from our structural model suggest that users' marginal utilities of consumption are positively impacted by the popularity of recommended creators, while their marginal utilities of creation are negatively impacted by the ratio of the popularity of recommended creators to their own popularity as creators. This is consistent with our field experiment results. In fact, using our structural model to simulate the treatment effect in this experiment yields both qualitatively and quantitatively similar results, increasing confidence in our model's ability to capture users' reactions to recommendation changes. Utilizing the model estimation, we conducted various counterfactual analyses by changing the average popularity of recommended creators to maximize the combined value of content consumption and production from users. Our counterfactual analysis indicates that, contrary to the industry practice, simply recommending the most popular creators may not always yield the best outcomes for the platform. For several user segments, it is often optimal to recommend relatively less popular creators so that the benefits from increased content production outweigh the costs of decreased content consumption. This demonstrates that by tailoring recommendations to trade off content consumption with content production, the platform can further increase overall value.

In summary, our study addresses a critical yet overlooked question in both academic and industry contexts—the impact of recommender systems on content creation. Our study makes several important contributions to both academic research and industry practice regarding recommender systems on content-sharing platforms. First, to the best of our knowledge, we are the first to provide rigorous and causal evidence on how content creators' popularity can simultaneously affect users' consumption and production in a real-world setting with cutting-edge recommender systems. We are also the first to document this important trade-off between content consumption and production through recommender systems. Second, we developed a new structural model that can accurately capture how recommender systems influence content consumption and production simultaneously. Third, our model and corresponding counterfactual analyses can be easily replicated on other content-sharing platforms, allowing them to adjust their recommendation strategies to increase the combined value of content consumption and production.

The rest of the paper proceeds as follows. Section 2 reviews the relevant literature. Section 3 introduces our field setting, experimental design, data, and randomization checks. Section 4 documents the main empirical findings from our field experiment. Section 5 describes the formulation and estimation methods of our structural model. Sections 6 illustrates the estimation results of our model and counterfactual analyses. In Section 7, we conclude by discussing practical implications of our research and directions for future research.

2. Literature Review

Our work relates primarily to three main streams of research: recommender systems, network effects, and platform operations,

Recommender systems. Recommender systems have become a pivotal area of research, with studies primarily aimed at enhancing user engagement and consumption behaviors. These efforts strive to optimize various metrics, including click-through rates (Liu et al. 2010, Wang et al. 2019, Rafieian 2023), watch time (Covington et al. 2016, Zheng et al. 2022), dwell time (Zou et al. 2019) and visiting frequency (Xue et al. 2023). Although these efforts have significantly enhanced user engagement and consumption, the impact of recommender systems on user content production remains underexplored, despite its critical importance for the long-term development of online content-sharing platforms.

To contribute to this question, we conducted a large-scale randomized field experiment on Platform V to investigate the impact of recommender systems on both content consumption and production. This study is pioneering in its focus and methods. To our best knowledge, this is the first to causally examine the dual impact of recommender systems through a large-scale randomized filed experiment. Our findings revealed a significant trade-off between content consumption and production, prompting us to develop a novel structural model that provides new methods and insights on how platforms can design recommendation algorithms to enhance overall value from both content consumption and production. Our work not only broadens the understanding of recommender systems multifaceted impact but also offers actionable guidance for platforms seeking to optimize these algorithms for comprehensive platform benefits.

Network Effects. Our study speaks to the body of literature on network effects (Eisenmann et al. 2006), a field that investigates the impact of networks on user behavior on both one-sided and two-sided markets. This research has explored the "same-side network effect," where the behavior of users on one side influences the-same-category behavior of the users in the same side. For instance, studies have shown that in the video game market (Shankar and Bayus 2003) and the telecommunications industry (Hu et al. 2019), consumers' consumption behavior is influenced by the consumption patterns of other consumers. Similarly, user choices on online dating platforms have been found to be affected by the selections of the users in the same side (Fong 2024). In addition to the same-side network effect, the literature also delves into the "cross-side network effect" within two-sided markets, examining how the presence of users on one side influences the utility of users on the opposite side. Examples include the dynamics between buyer size and seller size on e-commerce platforms (Chu and Manchanda 2016) and the interactions between two user groups to be matched on online dating services (Fong 2024, Halaburda et al. 2018).

Our research introduces a novel perspective to the discussion of network effects by recognizing that users on two-sided platforms often perform dual roles, engaging both in content production and consumption. Our research reveals that a user's production and consumption activities can be significantly influenced by the production behaviors of their peers. Notably, the phenomenon where a user's consumption is influenced by the production activities of same-side users introduces a new aspect to network effects, which we term "the cross-side effect within the same group." This insight extends the traditional understanding of network effects, providing a more nuanced view of user interactions and their impacts on platform dynamics.

Platform operations. Our research also intersects with the growing field focused on operational challenges within online platforms. This domain has explored strategies for enhancing platform performance through various levers such as pricing (Bimpikis et al. 2019, Zhang et al. 2020), advertising (Mookerjee et al. 2017), recommending or matching (Banerjee et al. 2016), and the stimulation of supply provision (Cabral and Li 2015, Burtch et al. 2018). Our study introduces a fresh perspective on the matching strategies of online platforms. Unlike previous studies, which predominantly aim to enhance consumption behaviors by matching, our analysis highlights the importance of valuing both supply provision and consumption in the design of matching strategies.

Specifically, our contribution aligns closely with literature examining content consumption and production on online content-sharing platforms. Past research has largely focused on one of two areas: enhancing content consumption through recommendation optimization (Wang et al. 2012), or incentivizing content production through financial rewards (Cabral and Li 2015, Burtch et al. 2018) or exploring psychological motivations (Wang et al. 2012, Zeng et al. 2022). However, there has been limited investigation on both content consumption and production simultaneously. A few studies have noted the negative correlation between content consumption and production using observational data (Ghose and Han 2011, Huang et al. 2015), yet a comprehensive understanding of these dual aspects remains underexplored. Our work contributes to this question by providing new insights into the intricate balance between encouraging content production and consumption on online content-sharing platforms, thereby optimizing overall platform value.

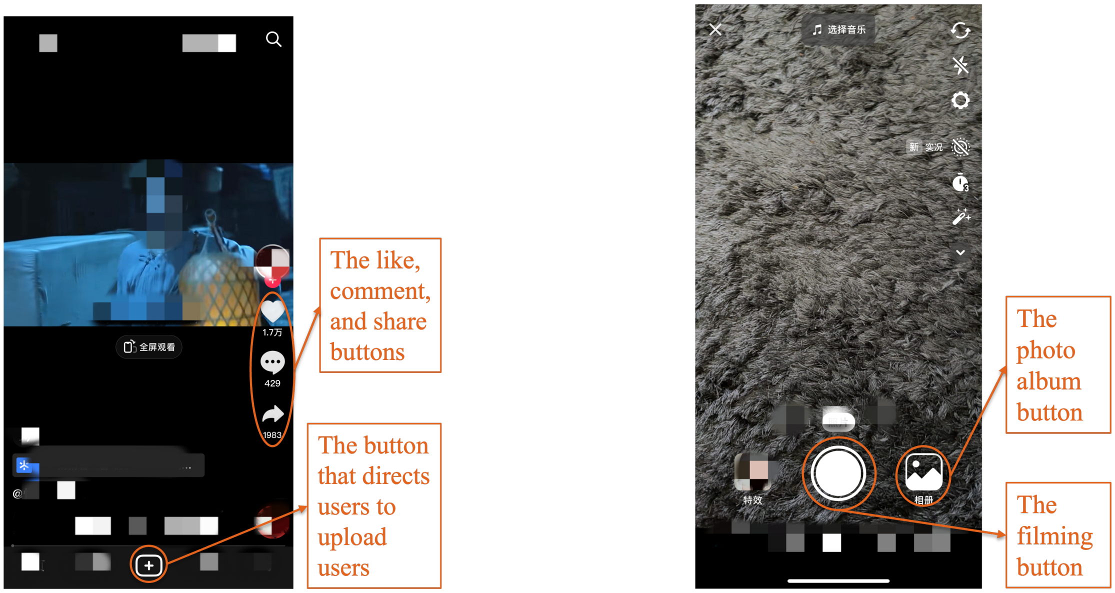
3. Setting, Experiment, Data and Randomization Checks

3.1. Empirical Setting

To conduct a randomized field experiment, we collaborated with Platform V, a large-scale video-sharing social network platform. As of 2023, the average number of active users per day on Platform V is more than 380 million. The users on Platform V can create and upload their own videos as content creators, and also watch videos created by other users as content viewers, or consumers. Similar to Facebook and Instagram, Platform V neither pays users for creating and uploading videos nor charges users for viewing videos. Instead, Platform V's revenue primarily comes from advertising. Videos on Platform V are typically a few seconds to a few minutes long, covering popular topics such as film and series re-edits, beauty, gaming, and daily life.

When a user opens the mobile application of Platform V, they are immediately shown a video covering almost the entire mobile phone screen, referred to as the "viewing page" (see Figure 1 (a)). The video starts playing automatically without requiring any clicks. If the user is not interested in the video, they can scroll up to view the next one. Consequently, a video being recommended is equivalent to a video being watched, at least being watched for a few seconds. If the user is interested in the video, they can continue watching it. While watching, users have several interaction options: clicking the "like" button to upvote the video, the "comment" button to leave comments, and the "share" button to share the video with others. Platform V utilizes a cutting-edge personalized recommender system to select a subset of videos that match the user's preferences from the supply pool of vast videos available on the platform.

At the bottom of the viewing page, there is a button that directs users to the "uploading page" (see Figure 1 (b)), where they can create and upload videos. On the uploading page, users play the role of creators in this two-sided network. To create a video, users have two options: they can either



(a) The Viewing Page (b) The Uploading Page

Figure 1 How Users Watch and Create Videos on Platform V[6]

film a video live by clicking the "filming" button located at the center bottom of the screen or select existing videos or photos from their phone's memory by clicking the "photo album" button in the bottom right corner.

3.2. Experimental Design

We conducted a randomized field experiment from January 8 to January 15, 2021, on Platform V, targeting 20% of their users. Participants were evenly and randomly divided into treatment and control groups. Platform V defined highly popular creators as those with a number of followers exceeding a specific threshold.[7] These highly popular creators comprised roughly 0.3% of all creators on the platform. Before initiating the experiment, platform managers identified a random 15% of these highly popular creators as "blocked creators," an important concept for defining the treatment in the experiment.

Upon each user login, regardless of group assignment, Platform V used the same recommender system to generate a personalized list of videos that would be recommended to the user. This list was selected mainly based on the creators' popularity on the platform and alignment with individual user preferences. This personalization recommender system is cutting-edge and commonly used in the industry (refer to Chen et al. (2024) for a detailed illustration of such a system). For users in the treatment group, videos from Blocked Creators were excluded from their personalized

[6] In order to protect the sensitive information of Platform V, we used a typical interface of short-video-sharing platforms and did some minor modification on the elements that can potentially identify Platform V.

[7] The exact threshold and details about the Blocked Creators remain confidential due to the non-disclosure agreement.

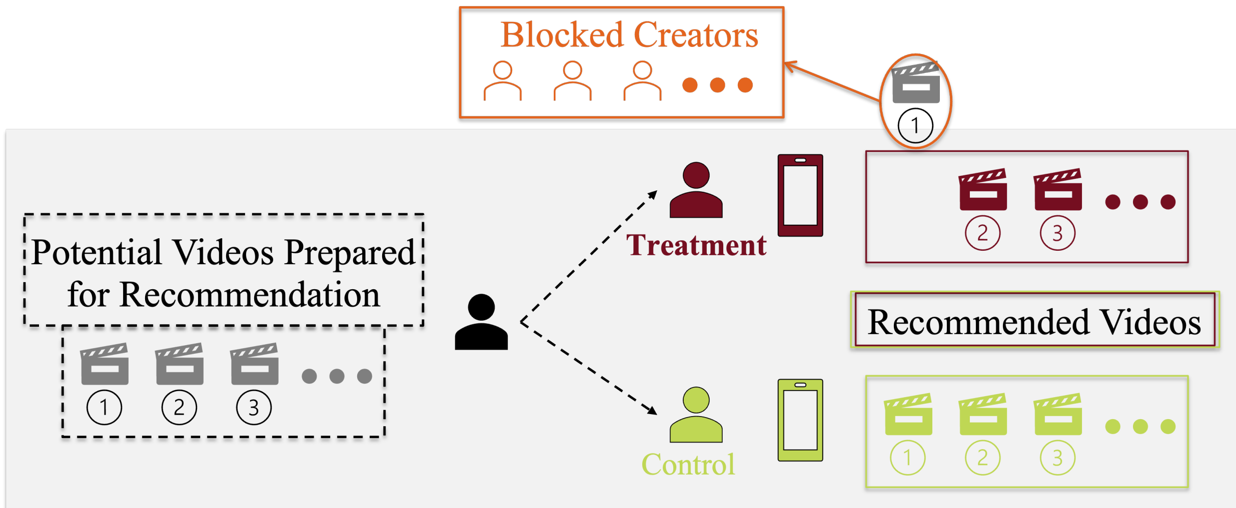


Figure 2 **Illustration of Our Randomized Field Experiment**[9]

recommendation list before being displayed, while users in the control group did not have such exclusions and followed the original recommender system. This manipulation aimed to reduce the average popularity levels of the creators recommended to treatment users, since the videos from these highly popular blocked users could be replaced by videos from less popular creators for treated users. This design of our experiment is detailed in Figure 2. We want to note that, while the treatment could potentially decrease the overall popularity of creators recommended to treatment users, the recommendation still remains highly personalized across each user. In other words, this treatment aimed to induce vertical differentiation of recommendations between the treatment and control groups while still maintaining the horizontal differentiation of recommendations across users facilitated by the original recommender system.[8]

3.3. Data and Variables

Due to data confidentiality reasons with Platform V, we did not have access to all users' data from the experiment. Instead, we were able to to gather a random sample of $100,000$ active users, $50,118$ users in the treatment group and $49,882$ in the control group. A user was active in our setting if she had both uploaded at least one video and watched at least one video during the 30 days prior to the start of our experiment. This selection criterion was highly representative of Platform V, as users meeting these conditions accounted for roughly 90% of the video supply (in terms of number

[8] It may be theoretically easier to manipulate video popularity rather than creator popularity in such an experiment. However, the number of videos is often too vast to be effectively targeted by manipulations in a real-world recommender system, which is why we focus on creator-level treatments.

[9] In this example where the highest-ranked video in a user's recommendation feed was created by Blocked Creators, it was removed from the recommendations. Videos from Blocked Creators could appear at various ranks, except the first rank. In addition, in other instances, users' potential recommendation pools might not include videos from Blocked Creators.

Table 1 Name and Description for Variables

Variable	Description
Creator Popularity$_i$	Median number of followers for user i over the week before the experiment.
Popularity of Recommendations$_{it}$	Median Creator Popularity across all videos watched by user i on day t.
Video Watch Time$_{it}$	Time spent by user i watching videos on day t.
Number of Videos Uploaded$_{it}$	Number of videos user i uploaded on day t.
Engagement per Video Uploaded$_{it}$	Engagement (views, likes, shares) per video uploaded by user i on day t.

Notes: Creator Popularity$_i$ is user level measurement, and all other three variables are user-day level measurements.

of videos) and 30% of the video demand (in terms of watching time). For each user in our sample, we collected their demographic information including gender and age, and tracked their decisions on watching and creating videos over the experiment period. We organized our data mainly on a daily basis, consistent with Platform V's data norms. Users and days are indexed by i and t respectively.

In Table 1, we outlined the primary variables in our study. First, we focused on how we measured the popularity of creators and the popularity of recommendations. To measure the popularity of user i as a creator, denoted as Creator Popularity$_i$, we used the median number of followers each user had during the week prior to our experiment.[10] This metric was chosen because Platform V primarily assesses creator popularity based on follower counts, and the treatment in our experiment was constructed based on this measure too. Additionally, using the number of followers as a popularity measure is well-established in prior research (Garcia et al. 2017, Pittman and Abell 2021), providing a reliable basis for comparison. After defining how Creator Popularity$_i$ is measured, Popularity of Recommendations$_{it}$, the average creator popularity that user i experienced on day t, can be defined. Specifically, for each user i on any day t during the experimental period, if user i watched videos, the median creator popularity across all videos watched by user i on that day is defined as Popularity of Recommendations$_{it}$.[11]

Second, for each user i on each day t, we calculated the Video Watch Time$_{it}$, which represents the total time this user spent watching videos during that day. This variable equals zero if the user did not log into the platform. We also measured the Number of Videos Uploaded$_{it}$, representing the count of videos uploaded by the user on that day; this variable also equals zero if no videos were uploaded. Furthermore, Platform V helped us construct the variable Engagement per Video Uploaded$_{it}$, which represents the average views, likes, and shares generated over the next two years across all videos that user i created on day t.

[10] Besides using the number of followers as our primary metric for creator popularity, we explored alternative metrics based on engagement statistics for robustness checks. We obtained qualitatively and quantitatively similar results with these alternative metrics. Detailed analyses are available in Online Appendix B.

[11] We do not have access to the data for all creators recommended in our sample due to data security. Instead, we rely on our collaborators at Platform V to construct Popularity of Recommendations$_{it}$ for each user per day.

Table 2 Randomization Checks

		Treatment Users (1)	Control Users (2)	P-Value of Two-Sample Proportion Test or T-Test (3)
Statistics on the Day Prior to the Experiment	Proportion of Females	65.23%	65.29%	0.93
	Age	2.29	2.28	0.38
	Pre-Experiment Number of Followers	0.05	0.05	0.98
	Pre-Experiment Number of Following	0.78	0.78	0.54
Statistics During 7 Days Prior to the Experiment	Pre-Experiment Average Number of Uploaded Videos	0.76	0.76	0.94
	Pre-Experiment Average Video Watch Time	1.33	1.33	0.71

Notes: All variables, except Proportion of Females, are standardized to have a unit standard deviation across all users.

To address the potential skewness and volatility in our data—a common challenge in video-sharing platform analytics (Li et al. 2012)—we requested Platform V to generate winsorized versions of our metrics (Dixon 1960). Specifically, for each of the raw variables, the winsorized variable was capped at the 99th percentile value of the user's corresponding measures recorded in the week preceding the experiment. These variables, denoted as Winsorized Creator Popularity$_i$, Winsorized Popularity of Recommendations$_{it}$, Winsorized Video Watch Time$_{it}$, Winsorized Number of Videos Uploaded$_{it}$, and Winsorized Engagement per Video Uploaded$_{it}$, were used to document our main estimation results, while raw data were included as a robustness check. Our analyses detailed in Section 4 confirmed that our main results are qualitatively and quantitatively robust, irrespective of whether they were based on raw or winsorized data. To protect Platform V's sensitive information, we standardized all continuous variables used in our empirical analyses to have a unit standard deviation.

3.4. Randomization Checks

Last, after defining our key variables, in order to verify the efficacy of our randomization process, we compared demographic and behavioral attributes across the treatment and control groups 7 days prior to the experiment. As detailed in Table 2, both groups demonstrated similar characteristics in terms of the proportion of females and age, as well as basic social network metrics such as the number of followers (Pre-Experiment Number of Followers) and the number of users they were following (Pre-Experiment Number of Following) on the day before the experiment. Additionally, we tracked video-related activities over the 7 days prior to the experiment, including the average number of videos uploaded per day per user (Pre-Experiment Average Number of Videos Uploaded) and the average total video watch time per day per user (Pre-Experiment Average Video Watch Time). These metrics across both groups show no significant differences, ensuring that any differences observed post-intervention can be reliably attributed to the intervention itself.

4. Main Empirical Findings

In this section, we reported the main empirical analyses from our randomized field experiment. Our main analyses were conducted at the user-day level and employed the ordinary least squares (OLS) regression using the following specification:

$$\text{Outcome Variable}_{it} = \alpha_0 + \alpha_1 \text{Treatment}_i + \epsilon_{it} \tag{1}$$

In this regression specification, Treatment_i is a binary variable indicating whether user i was in the treatment group as opposed to the control group, and $\text{Outcome Variable}_{it}$ represents the daily measurement of user i's outcome on day t. The specifics of the outcome variables would be discussed later. Standard errors are clustered at the user level to account for the correlation across observations within the same user.

4.1. Manipulation Checks

In our experiment, we excluded videos from blocked creators that would have otherwise been recommended to the treatment group. We want to note that while this treatment could lead to a decline in the average popularity of creators recommended to the treatment group compared to the control group, it may not always happen since it is possible that all other creators on users' recommended lists could have similar or higher popularity compared to blocked creators. Therefore, to test whether our treatment indeed affects the popularity of creators recommended to the treatment group, we conducted a manipulation check by comparing the average popularity of creators recommended, i.e., $\text{Popularity of Recommendations}_{it}$, to the treatment group with those recommended to the control group.

Specifically, using regression specification (1), we tested how the treatment causally impacted the $\text{Popularity of Recommendations}_{it}$ during the experiment. The results, detailed in Panel A of Table 3, show a significant reduction in the Popularity of Recommendationsit for the treatment group—26.80% lower, or 0.0797 standard deviations ($p < 0.0001$, column (1))—compared to the control group. Similarly, the Winsorized $\text{Popularity of Recommendations}_{it}$ decreased by 25.64% or 0.1418 standard deviations ($p < 0.0001$, column (2)). These findings demonstrate the effectiveness of the experimental manipulations in altering the popularity of creators recommended to the treatment users, reflecting a substantial decrease in the exposure to highly popular content among treated users.

4.2. The Impact of Our Treatment on Content Consumption

After validating the effectiveness of our intervention in decreasing the popularity of recommendations, we proceeded to examine its impact on user behavior, specifically focusing on how recommending less popular creators influenced the video-watching outcomes of treated users. Our

Table 3 The Impact of Our Treatment on Users' Watching Experiences and Outcomes

Panel A: Manipulation Checks—Our Experiment Reduced Creator Popularity Recommended to Treated Users		
Outcome Variable	Popularity of Recommendations$_{it}$	Winsorized Popularity of Recommendations$_{it}$
	(1)	(2)
Treatment$_i$	-0.0797****	-0.1418****
	(0.0023)	(0.0023)
Relative Effect Size	-26.80%	-25.64%
Observations	$749,290$	$749,290$

Panel B: Treatment Effect on Video Consumption		
Outcome Variable	Video Watch Time$_{it}$	Winsorized Video Watch Time$_{it}$
	(1)	(2)
Treatment$_i$	-0.0134****	-0.0178****
	(0.0022)	(0.0030)
Relative Effect Size	-1.38%	-1.34%
Observations	$800,000$	$800,000$

Notes: Panel A includes user-day observations during the experimental period in which the user watched at least one video, and panel B includes the whole user-day observations during the experimental period. The outcome variables are standardized to have a unit standard deviation across all observations in the experiment. Standard errors are clustered at the user level. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001.

specification follows Equation (1) with outcome variables being Video Watch Time$_{it}$ and Winsorized Video Watch Time$_{it}$. As shown in Panel B of Table 3, compared to control users, treatment users who were recommended creators with lower popularity spent less time watching videos by 1.34%-1.38% (or 0.0134-0.0178 standard deviations, p<0.0001, columns (1) and (2)), depending on whether we winsorized outcome variables. This demonstrates that recommending videos from less popular creators significantly decreases the content consumption of end users, which is well-known in both the literature and the industry.

4.3. The Impact of Our Treatment on Content Production

Shifting our focus from the consumer role to the creator role, we examined the impact of our treatment on users' content creation using Equation (1) as the specification and (Winsorized) Number of Videos Uploaded$_{it}$ as outcome variables. The results, detailed in Panel A of Table 4, demonstrated that our treatment, which lowered the popularity of recommendations, led to a significant increase in content production. Specifically, the number of videos uploaded by users in the treatment group increased by 2.39% (or 0.0065 standard deviations, p=0.0035, column (1)) using the raw outcome variable, and by 2.71% (or 0.0091 standard deviations, p=0.0001, column (2)) when using the winsorized outcome variable.

This interesting result shows that while recommending videos from less popular creators can significantly decrease content consumption, it can also significantly increase content production. We interpret the driving force of this result to be two-fold. First, there could be a time trade-off effect between content production and consumption, where less time spent on content consumption

Table 4 The Impact of Our Treatment on Users' Production Outcomes

Panel A: Treatment Effects on Video Production		
Outcome Variable	Number of Videos Uploaded$_{it}$	Winsorized Number of Videos Uploaded$_{it}$
	(1)	(2)
Treatment$_i$	0.0065**	0.0091***
	(0.0022)	(0.0022)
Relative Effect Size	2.39%	2.71%
Observations	800,000	800,000

Panel B: Our Treatment Did not Change Engagement Metrics for Uploaded Videos						
Outcome Variable	Engagement per Video Uploaded$_{it}$			Winsorized Engagement per Video Uploaded$_{it}$		
	(Likes)	(Comments)	(Shares)	(Likes)	(Comments)	(Shares)
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment$_i$	0.0022	0.0101	0.0061	0.0066	0.0091	0.0080
	(0.0058)	(0.0058)	(0.0057)	(0.0058)	(0.0058)	(0.0058)
Observations	118,097	118,097	118,097	118,097	118,097	118,097

Panel C: Our Treatment Did not Change Users' Own Creator Popularity		
Outcome Variable	After-Experiment Number of Followers$_i$	Winsorized After-Experiment Number of Followers$_i$
	(1)	(2)
Treatment$_i$	0.0003	0.0055
	(0.0063)	(0.0055)
Observations	100,000	100,000

Notes: Panel A includes all user-day observations during the whole experimental period. Panel B includes user-day observations in which the user uploaded at least one video. Panel C includes user-level observations across all users. The outcome variables are standardized to have a unit standard deviation across all observations during the experiment. Standard errors in Panel B are clustered at the user level. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001.

naturally leads to more time spent on content creation. Second, it is possible that the intrinsic motivation of users to produce videos increases by seeing videos from less popular creators. These users may compare their own popularity or video quality with the recommended videos and believe they also have a decent chance of being recommended to others. Unfortunately, as documented by Gentzkow (2007), we do not have an endogenous shock that only affects one of these mechanisms, making it difficult to separately identify these two mechanisms.

Motivated by this interesting result on content production, we first explored whether our intervention affects not only the quantity of content production but also the quality of content production. Using a similar specification, we tested the impact of our treatment on (Winsorized) Engagement per Video Uploaded$_{it}$ for users who created videos on day t. The results, presented in Table 4 Panel B, indicate that our intervention did not significantly affect the long-term engagement levels of videos uploaded by treated users compared to control users. We want to note that this analysis does not represent a clean causal estimate of the treatment's impact on created content quality since users' participation in creating videos is also affected by the treatment. To provide more robust evidence, we further calculated After-Experiment Number of Followers$_i$, representing the number of followers at the end of our experimental period for each user. Panel C of Table 4 demonstrates that the treatment does not significantly alter users' own creator popularity between the treatment and control groups after the experiment. This is consistent with the interpretation that the treatment does not change the quality of content production.

In summary, our experiment demonstrates that while decreasing the average popularity of recommended creators negatively impacts the content consumption of users, it could positively impact the content production of users. Furthermore, it affects only the quantity of content production rather than the quality.

5. The Structural Model

Our experiment revealed that a reduction in the popularity of recommended creators led to a significant trade-off between video consumption and production. This finding prompted us to explore how recommender systems could be designed to maximize platform benefits by considering its impact on both content consumption and content production. To address this, we developed a structural model, which is detailed in this section.

5.1. Model Formulation

Drawing on prior research in the domain of time and resource allocation (e.g., Ribar (1995) and Pellegrini et al. (2021)), our model assumes that users derive utility from leisure activities but are constrained by a finite amount of time to conduct such leisure activities. These leisure activities include consuming and creating videos on Platform V, as well as other leisure activities that serve as the outside option. By strategically allocating their time across various leisure activities within this time budget, users seek to maximize their own utility.

In our model, each user i makes decisions to maximize her own utility on day t by changing her choices of x_{it} and y_{it}. x_{it} is a continuous variable representing the time user i spends consuming videos on day t, while y_{it} is an integer representing the number of videos that user i uploads on day t. We assume that the time it costs for user i to upload y_{it} videos on day t is denoted by the cost function $c(y_{it})$, which takes the following quadratic form:

$$c(y_{it}) = c_1 y_{it} + c_2 y_{it}^2, \tag{2}$$

where both marginal costs c_1 and c_2 will be estimated. Note that we assume the production function is convex, which is consistent with the past literature (Lazear 2000, Moldovanu and Sela 2008, Charness et al. 2018), and this particular quadratic function form is also widely used in the past (e.g., Ederer (2010), Sarafopoulos (2015a,b)). Moreover, we assume that each user i on any day t will spend K units of time on leisure consumption, and we use o_{it} to denote user i's time spent on leisure activities other than content consumption and production on Platform V on day t. In other words, our model's time constraint can be represented as

$$x_{it} + c(y_{it}) + o_{it} = K. \tag{3}$$

After introducing the time constraints of our decision variables, let us move on to how these decisions affect users' utilities. To model the utility derived from the video consumption decision x_{it}, following the past literature (Sun et al. 2015, Sarafopoulos 2015a, Huang and Bronnenberg 2018, Kim et al. 2023), we assume that user i's utility from consuming content for x_{it} amount of time on day t is a concave function with a logarithmic form:

$$\psi_{x,it}\ln(x_{it}+1)=(\bar{\psi}_{x,it}+\delta_{x,it})\ln(x_{it}+1),$$

where $\psi_{x,it}=\bar{\psi}_{x,it}+\delta_{x,it}$ represents the marginal utility from consuming content and consists of $\bar{\psi}_{x,it}$, the base marginal utility of consumption and a random consumption marginal utility error $\delta_{x,it}$. Motivated by our experimental result and past literature (Dobrian et al. 2011, Kim et al. 2012), we assume the base marginal utility of consumption is a function of popularity of recommended creators. Specifically,

$$\bar{\psi}_{x,it}=a_0+a_1\log(\tilde{q}_{it}), \tag{4}$$

where a_0 and a_1 are estimated parameters and $\tilde{q}_{it}$ represents Popularity of Recommendations$_{it}$. We will formally define the consumption marginal utility error later along with the production marginal utility error.

Similarly, following the past literature (Lazear 2000, Goldsmith et al. 2000, Charness and Kuhn 2005), we model the utility derived from creating y_{it} videos for user i on day t as:

$$\psi_{y,it}y_{it}=(\bar{\psi}_{y,it}+\delta_{y,it})y_{it},$$

where $\psi_{y,it}=\bar{\psi}_{y,it}+\delta_{y,it}$ represents the marginal utility from producing content and consists of $\bar{\psi}_{y,it}$, the base marginal utility of production, and a random marginal utility error $\delta_{y,it}$ for production. Again, motivated by our experimental results and past literature on social comparison in production (Huang et al. 2019), we model the base marginal utility of production $\bar{\psi}_{y,it}$ as being influenced by the ratio of the popularities of recommended creators and the user herself:

$$\bar{\psi}_{y,it}=b_0+b_1\frac{\tilde{q}_{it}}{q_i}, \tag{5}$$

where b_0 and b_1 are estimated parameters and q_i is Creator Popularity$_i$ for user i prior to the experiment.

Let us now formally define the marginal utility random error for both consumption and production $\delta_{x,it}$ and $\delta_{y,it}$. Specifically, we assumed that both error terms can be decomposed into an individual-specific random effect and an idiosyncratic error shock: $\delta_{x,it}=\epsilon_{x,i}+e_{x,it}$ and $\delta_{y,it}=\epsilon_{y,i}+e_{y,it}$. The individual-specific random effects $\epsilon_{x,i}$ and $\epsilon_{y,i}$ were shared across multiple periods by the same user i. These time-invariant terms indicated consistent individual patterns over time.

They were assumed to follow a bivariate normal distribution with zero means and a covariance matrix: $\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$. The parameter ρ facilitated correlation between the consumption and production utilities, capturing the interplay between these two dimensions of user behavior. Additionally, the idiosyncratic error terms $e_{x,it}$ and $e_{y,it}$, which captured day-to-day variability in video consumption and production, were assumed to be follow a bivariate normal distribution with zero means and a covariance matrix for each day t: $\begin{bmatrix} 1 & \tau \\ \tau & 1 \end{bmatrix}$. Both ρ and τ will be estimated by our model.

Last, we could introduce the users' utility maximization problem where the overall utility for user i at day t is a function of x_{it} and y_{it} with a time allocation constraint. To ensure that the model is identifiable, we normalized the marginal utility of consuming outside leisure activities to 1, and in turn the utility from outside leisure activity for user i at day t is o_{it}. As a result, the user's maximization problem can be written as:

$$\begin{aligned} \max_{x\geq 0, y\in\{0,1,\cdots\}} U_{it}(x_{it}, y_{it}) = & \max_{x\geq 0, y\in\{0,1,\cdots\}} \psi_{x,it} \ln(x_{it}+1) + \psi_{y,it} y_{it} + o_{it}, \\ \text{s.t.} \qquad & x_{it} + c(y_{it}) + o_{it} = K. \end{aligned} \tag{6}$$

Use x^*_{it}, y^*_{it} to denote the optimal solutions. Given that our model involves mixed-integer optimization with both equality and inequality constraints, deriving a closed-form solution is generally infeasible (Chong and Żak 2013). Without loss of generality, the formulation for the optimal solution when y_{it} can attain a maximum value of Y is presented as follows:

$$\begin{aligned} x^*_{it}(\Theta), y^*_{it}(\Theta) = \underset{x\geq 0, y\in\{0,1,\cdots,Y\}}{argmax} \; & \psi_{x,it} * ln\left(\min\left\{\max\{\psi_{x,it}-1, 0\}, K-c(y)\right\}+1\right) + \psi_{y,it} * y \\ & + K - c(y) - \min\left\{\max\{\psi_{x,it}-1, 0\}, K-c(y)\right\}, \end{aligned}$$

where $\Theta = \{a_0, a_1, b_0, b_1, c_1, c_2, \tau, \rho\}$ represents estimable parameters in our model.

5.2. Model Estimation

The model in Section 5.1 describes the user time allocation problem for one class of users. We further addressed heterogeneity in user preferences by implementing latent class estimation, which assumes multiple latent user segments within our dataset. Each segment, indexed by k, is characterized by a distinct set of estimable model parameters, Θ_k. The probability that user i belongs to class k is denoted as P_k, where k ranges from 1 to K. We employed a maximum likelihood approach to estimate the parameters of our model. The likelihood function for an individual user, assuming their class is known and fixed, was derived and is detailed in Section 5.2.1. The algorithms used for latent class estimation are further elaborated in Section 5.2.2.

5.2.1. The Likelihood Function. In this section, we derived the likelihood function for observing the decisions of user i across periods conditional on them belonging to segment k. For clarity, we omitted the notation indicating conditioning on segment k in the following mathematical expressions.

Given that users maximize their utility, the observed consumption and production decisions of user i represent the optimal solutions. The probability of observing a specific vector of decisions for user i over all time periods $t=1,2,\ldots,T$, denoted as $[x_{i1}^*, x_{i2}^*, \ldots, x_{iT}^*, y_{i1}^*, y_{i2}^*, \ldots, y_{iT}^*]'$, is characterized by the joint probability of these components of optimal consumption and production decisions:

- Consumption decisions: Conditional on any given value of production decision y_{it}, the user chose to consume a continuous variable x_{it} to obtain the highest utility U_{it}, and thus the optimal decision, x_{it}^*, will fall into two possible solutions: (1) an interior solution ($x_{it}^* > 0$) satisfying the first-order condition: $\frac{\partial U_{it}}{\partial x_{it}} = \frac{\psi_{x,it}}{x_{it}+1} - 1$ such that $x_{it}^* = \psi_{x,it} - 1$; (2) a corner solution ($x_{it}^* = 0$) when $\psi_{x,it} \leq 1$. Altogether, we have

$$\begin{aligned} \delta_{x,it} &= x_{it}^* + 1 - \bar{\psi}_{x,it} \quad \text{if} \quad x_{it}^* > 0, \\ \delta_{x,it} &\leq 1 - \bar{\psi}_{x,it} \quad \text{if} \quad x_{it}^* = 0. \end{aligned} \tag{7}$$

- Production decisions: The user chose a discrete value of y_{it} as the number of videos to create to maximize utility.

$$U_{it}(x_{it}^*, y_{it}^*) \geq U_{it}(x_{it}^*, y_{it}^* + 1) \quad \text{for} \quad y_{it}^* \geq 0 \quad \text{and} \quad U_{it}(x_{it}^*, y_{it}^*) \geq U_{it}(x_{it}^*, y_{it}^* - 1) \quad \text{for} \quad y_{it}^* \geq 1.$$

  This condition is equivalent to

$$\begin{aligned} c_1 + c_2(2y_{it} - 1) - \bar{\psi}_{y,it} \leq \delta_{y,it} \leq c_1 + c_2(2y_{it} + 1) - \bar{\psi}_{y,it} \quad &\text{if} \quad y_{it}^* = 1, 2, \cdots, \\ \delta_{y,it} \leq c_1 + c_2 - \bar{\psi}_{y,it} \quad &\text{if} \quad y_{it}^* = 0. \end{aligned} \tag{8}$$

Let $L_i(\Theta_k)$ represent the likelihood of all decisions made by user i across periods when user i belonged to class k, which is the joint probability such that conditions defined in Equations (7) and (8) are satisfied over all time periods $t=1,2,\ldots,T$:

$$\begin{aligned} L_i(\Theta_k) = \int \Bigg\{ & \begin{cases} \delta_{x,it} = x_{it}^* + 1 - \bar{\psi}_{x,it}, & \text{if} \quad x_{it}^* > 0 \\ \delta_{x,it} \leq 1 - \bar{\psi}_{x,it}, & \text{if} \quad x_{it}^* = 0 \end{cases} \quad \text{for} \quad t = 1, 2, \ldots, T, \\ & \begin{cases} c_1 + c_2(2y_{it} - 1) - \bar{\psi}_{y,it} \leq \delta_{y,it} \leq c_1 + c_2(2y_{it} + 1) - \bar{\psi}_{y,it} & \text{if} \quad y_{it}^* > 0 \\ \delta_{y,it} \leq c_1 + c_2 - \bar{\psi}_{y,it} & \text{if} \quad y_{it}^* = 0 \end{cases} \quad \text{for} \quad t = 1, 2, \ldots, T \Bigg\} \\ & dF(\delta_{x,i1}, \delta_{y,i1}, \ldots, \delta_{x,iT}, \delta_{y,iT}). \end{aligned} \tag{9}$$

Computing $L_i(\Theta_k)$ from the joint distribution of a $2 \times T$-dimensional vector of correlated error terms presented a significant computational challenge. To effectively manage this complexity, we took advantage of the GHK simulation method for likelihood calculations (Geweke et al. 1994, Hajivassiliou and Ruud 1994, Keane 1994, Jiang et al. 2021). This simulation technique is highly regarded for its ability to handle truncated multivariate variables with correlations, as demonstrated in its application to multinomial probit models (Geweke et al. 1994). In such models, error terms are bound by a series of inequality constraints driven by the observed choices. By utilizing the GHK method, we drew error terms from truncated distributions rather than their original forms, incorporating the use of importance sampling weights to enhance the accuracy and efficiency of our estimates. We provided a detailed sampling procedure in the Online Appendix C. We used this method to simulate the likelihood L_i for each user i across all periods.

In addition, recommender systems are tailored to match video selections with individual user tastes and preferences, potentially introducing an endogeneity issue. The relationships observed between user outcomes and the popularity of recommendations might not accurately depict causal effects. This is why we conducted a randomized field experiment to implement an exogenous shock to recommended creators when we analyzed empirical findings. To address endogeneity in this nonlinear structural model, we employed the widely-used control function method (Agarwal 2015, Wooldridge 2015). This method functions as an instrumental variable approach but is particularly adept at managing complex models that are nonlinear in endogenous variables. Specifically, we regressed $\log(\tilde{q}_{it})$ on Treatment_{it} to obtain the residual, denoted by $r_{x,it}$. We then included $r_{x,it}$ as a control in predicting $\psi_{x,it}$. Similarly, for we regressed $\tilde{q}_{it}/q_i$ on Treatment_{it} to obtain the residual, denoted by $r_{y,it}$. We incorporated $r_{y,it}$ as a control in predicting $\psi_{y,it}$.

5.2.2. Likelihood Maximization and Latent Class Estimation. After deriving the likelihood of all decisions made by user i across periods when user i belongs to class k, given P_k as the probability of users in type k, where $k = 1, 2, \ldots, K$, the log-likelihood function for all users is represented as:

$$L = \sum_i \log\left(\sum_k P_k L_i(\Theta_k)\right) \tag{10}$$

To estimate the parameters $\Theta_1, \Theta_2, \ldots, \Theta_K$ in our latent class model, we employed the Expectation-Maximization (EM) algorithm (Moon 1996). The EM algorithm is a robust statistical technique designed to find maximum likelihood estimates in probabilistic models that include unobserved latent variables. It alternates between two main steps: the Expectation (E) step and the Maximization (M) step, iteratively repeated until convergence criteria were met. In the E step, we updated the conditional distribution of the latent variables (P_k) while fixing the current estimates of the parameters (Θ_k). The M step then updated the parameters (Θ_k) to maximize the expected

Table 5 Data Sources for Model Estimation

Notation in the Model	Data from the Experiment
x_{it}	Winsorized Video Watch Time$_{it}$
y_{it}	Winsorized Number of Videos Uploaded$_{it}$
q_i	Winsorized Creator Popularity$_i$ (+1)
$\tilde{q}_{it}$	Winsorized Popularity of Recommendations$_{it}$ (+1)

Notes: Regarding $\tilde{q}_{it}$, we imputed missing values based on the average popularity of recommendations from previous days when the user was active during the experimental period. The (+1) adjustment is applied to Winsorized Creator Popularity$_i$ and Winsorized Popularity of Recommendations$_{it}$ to avoid undefined calculations in logarithmic functions and denominators.

log-likelihood, fixing the latent variables (P_k) as found in the E step. These steps are detailed in Online Appendix D.

5.3. Model Identification

The data used to estimate the model are detailed in Table 5. The definition of variables used for our structural estimation is very similar to those discussed in Section 3.3 with some minor changes. Specifically, Winsorized Video Watch Time$_{it}$ and Winsorized Number of Videos Uploaded$_{it}$ were used as direct inputs for x_{it} and y_{it}, respectively. Both measures were recorded daily throughout the experimental period for each user, with values set to zero on days when users did not log onto Platform V. We utilized Winsorized Creator Popularity$_i$, the median number of followers for user i over the week before the experiment, as the input for q_i. For Popularity of Recommendations$_{it}$, we included observations at the user-day level throughout the experiment. On days when a user did not log in, missing observations were imputed using the average value from previous days' observations during the experimental period. In this balanced dataset prepared for structural estimation, 0.02% of Winsorized Popularity of Recommendations$_{it}$ and 0.17% of Winsorized Creator Popularity$_i$ entries are zeros. Since these variables are used in logarithmic functions and as denominators, where zero values would lead to undefined results, we incremented each by one to facilitate calculations of $\tilde{q}_{it}$ and q_i. Given that the magnitude of the maximum $\tilde{q}_{it}$ value reaches 10^6, the scaling factors for coefficients a_1 and b_1 were adjusted to 10 and 10^6, respectively, to enhance sensitivity during the parameter optimization process.

Last, we want to briefly discuss how each of the model parameters is identified from various moments in our data source. First, parameters a_0 and a_1, as defined in Equation (4), are identified by the exogenous variations in the popularity of recommendations introduced by our treatment and users' decisions regarding video consumption. Second, b_0 and b_1 defined in Equation (5) as well as c_1 and c_2 defined in Equation (2) are identified by users' choices in content production under various levels of recommended creators' popularities and the users' own popularity. We want to note that b_0 and c_1 are perfectly colinear in our model and cannot be identified simultaneously, so we normalize b_0 to 0. Last, the error correlation parameters ρ is identified through the correlation

Table 6 Estimation Results of the Structural Model

			Segment 1	Segment 2	Segment 3	Segment 4
(1)	*Marginal Utility of Consumption*	a_0	1.59	1.45	1.47	0.95
			(0.0008)	(0.0008)	(0.0006)	(0.0004)
(2)		a_1	0.55	0.64	0.61	3.28
			(0.0009)	(0.0009)	(0.0007)	(0.0004)
(3)	*Marginal Utility of Creation*	b_1	-70.04	-3.85	-2.83	-12.52
			(0.0001)	(0.0001)	(0.0003)	(<0.0001)
(6)	*Creation Time Cost*	c_1	0.73	1.32	0.91	0.85
			(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)
(7)		c_2	0.61	0.18	0.32	0.33
			(<0.0001)	(0.0001)	(<0.0001)	(<0.0001)
(4)	*Correlation Between Two Utilities*	ρ	0.33	0.34	0.43	0.27
			(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)
(5)		τ	0.21	0.22	0.21	0.17
			(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)
(8)	*Segment Probability*		33.41%	31.43%	26.86%	8.30%
			(<0.0001)	(0.0003)	(0.0003)	(<0.0001)
	Observations		800,000			

Standard errors are presented in parentheses.

between production and consumption decisions for each individual user across time while the error correlation parameter τ is identified through the overall correlation between production and consumption across each user at each day.

6. Estimation Results and Counterfactuals

6.1. Estimation Results

We employed the L-BFGS-G method[12] to estimate our model. To robustly estimate standard errors, we used bootstrapping techniques, randomly sampling with replacement at the user level 100 times and re-estimating the model for each sample. We utilized the Bayesian Information Criterion (BIC) to determine the optimal number of latent classes. After incrementally increasing the number of latent classes from one to five, we observed the lowest BIC with four classes[13], suggesting the best model fit at this level.

Table 6 presents estimates from our structural model, identifying four distinct user segments with varying preferences. These segments represent 33.41%, 31.43%, 26.86%, and 8.30% of the total users in our sample, respectively (see row (8)). We denote these four segments as Segments 1, 2, 3, and 4.

The parameter a_0 denotes the baseline marginal utility of consuming videos, irrespective of the popularity of recommendations provided to users. The values of a_0 are 1.59, 1.45, 1.47, and 0.95

[12] For details on the L-BFGS-G optimization method, see: https://docs.scipy.org/doc/scipy/reference/optimize.minimize-lbfgsb.html

[13] The BIC values for models with one to five classes were: 3004481.35, 2990920.13, 2983414.04, 2982237.32, and 2982323.47, respectively.

for the four segments, as shown in row (1). When the popularity of recommendations is set to zero, the utility for consuming x_{it} hours of videos for each segment is $(1.59 + \delta_{x,it})\ln(x_{it}+1)$, $(1.45+\delta_{x,it})\ln(x_{it}+1)$, $(1.47+\delta_{x,it})\ln(x_{it}+1)$, and $(0.95+\delta_{x,it})\ln(x_{it}+1)$, respectively. The value of a_0 qualitatively indicates users' consumption choices when the popularity of recommendations is zero. Accordingly, the consumption choices under zero popularity of recommendations in Segment 4 are the lowest, followed by Segment 2, Segment 3, and Segment 1, consistent with the sequence of a_0 values.

The parameter a_1 measures the rate of change in the marginal utility of consuming videos in response to changes in the logarithm of the median popularity of recommended creators $\log(\tilde{q}_{it})$. The parameter a_1 across all segments has statistically significant positive values of 0.55, 0.64, 0.61, and 3.28, respectively (row (2)). These positive a_1 values indicate that as $\tilde{q}_{it}$ increases, the marginal utility of consuming videos among all four segments increases. For instance, a 1% increase in $\tilde{q}_{it}$ corresponds to a $\log(1.01)$ increase in $\log(\tilde{q}_{it})$. This change, inversely scaled by the factor of 10 that was used to ensure the sensitivity of a_1 in the estimation process, results in an increase of $a_1 \times \log(1.01) \times 10^{-1}$ in the marginal utility of consumption. It is worth noting that Segments 1-3 have similar values of a_1, while Segment 4 has the highest value of a_1, more than five times that of Segments 1-3, indicating a much higher sensitivity to the popularity of recommendations in consumption for Segment 4.

The parameter b_1 measures the rate of change in the marginal utility of creating videos in response to changes in the median popularity of recommended creators $\tilde{q}_{it}$. The parameter b_1 shows a statistically significant negative impact of $\tilde{q}_{it}$ on video production, with values of -70.04, -3.85, -2.83, and -12.52, as noted in row (3). This indicates that an increase in $\tilde{q}_{it}$ reduces the utility derived from creating videos. For example, if $\tilde{q}_{it}$ increases by one thousand, inversely scaled by the factor of 10^6 that was used to ensure the sensitivity of b_1 in the estimation process, yielding $b_1 \times 10^{-3}$ in the marginal utility of production, respectively. It is worth noting that Segments 2-3 have similar values of b_1, being the lowest in absolute value, Segment 4 has a medium absolute value of b_1, roughly 4 times that of Segments 2-3. Segment 1 has the largest absolute value of b_1, roughly 5 times that of Segment 4. These indicate a high sensitivity to the popularity of recommendations in production for Segment 1 and a medium sensitivity for Segment 4, while Segments 2-3 have low sensitivity.

The parameters c_1 and c_2, outlined in rows (4) and (5) of the table, dictate the time costs for video production. Here, c_1 serves as the linear coefficient and c_2 as the quadratic term. The time required to create one video varies across the segments, averaging 1.14, 1.50, 1.23, and 1.18 hours for segments 1 through 4, respectively $(c_1 + c_2)$. Similarly, the time needed to produce two videos

Table 7 Simulated Average Treatment Effects on Video Consumption and Production

Panel A: Effects on Video Consumption		
Outcome Variable	Simulated Video Watch Time	Simulated Number of Videos Uploaded
	(1)	(2)
Treatment	-0.0486****	0.0691****
	(0.0022)	(0.0022)
Relative Effect Size	-1.89%	3.54%
Observations	800,000	800,000

We simulated each user's decision on video watching time and number of videos uploaded each day in the experimental period. All outcome variables are standardized to have unit deviation before entering the regressions. Standard errors are clustered at the user level. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001.

is 3.17, 2.20, 2.19, and 2.17 hours for each respective segment ($c_1 + c_2 \times 4$). The time for producing other numbers of videos can be calculated using the same logic.

The parameters ρ and τ, representing the correlations in random error terms, are statistically significant and positive across all segments with slight variations, as shown in rows (6) and (7). The parameter ρ depicts the time-invariant correlation between consumption and production utility for an individual. A positive ρ indicates that users who derive high utility from consumption (production) also consistently and stably derive high utility from production (consumption), suggesting an inherent attitude toward spending time on the platform, either consuming or creating. The parameter τ depicts the within-same-period correlation between consumption and production. A positive τ means that on a specific day, when a user derives high utility from production (consumption), they also derive high utility from consumption (production) on that day, which may suggest that the user has sufficient time to spend on the platform on that day.

In summary, all the parameter estimates across the four segments are significant and consistently signed, aligning with their interpretations. Except for the parameters for error correlation terms, which are quite similar across all four segments, the other parameters differentiate each segment. The parameters a_1 and b_1 are particularly important since they represent the sensitivity to changes in the median popularity of recommended creators, which was the focus of our experimental manipulation. From the aforementioned analyses, we note a much higher sensitivity to the popularity of recommendations in consumption for Segment 4 compared to Segments 1-3. Additionally, there is a high sensitivity to the popularity of recommendations in production for Segment 1 and a medium sensitivity for Segment 4, while Segments 2-3 exhibit low sensitivity.

6.2. Model Validation

The core purpose of our structural model is to understand the trade-off between content production and content consumption when recommending creators with different popularity levels. An important way to validate our model is to compare the simulated treatment effects of decreasing the median popularity of recommended creators with the observations from our experiments in

Section 4. Therefore, we recalculated the main treatment effects based on our structural model. As indicated in Table 7, our simulation shows that Video Watch Time decreased by 1.98% and the Number of Videos Uploaded increased by 3.54%, highlighting the existence of such trade-offs. These simulated effects closely align with the empirical observations—1.34% decrease in Video Watch Time and 2.71% increase in the Number of Videos Uploaded (as detailed in Sections 4.2 and 4.3)—thereby suggesting the validity of our model.

6.3. Counterfactuals

Our field experiment confirmed that recommending highly popular creators led to a trade-off between content consumption and production (see Section 4), and our model also recovered this trade-off (see Section 6.2). Given the importance of both activities for online content-sharing platforms, this raises a crucial question: How should platforms design their recommender systems to optimize the overall value derived from both consumption and production? To address this, we conducted several counterfactual analyses based on our model, providing strategic insights into enhancing recommender systems to maximize this overall value.

After incorporating our empirical findings and discussing with the managers of Platform V, we modeled Platform V's overall value as follows:

$$V(\tilde{q}_{it}) = \sum_{i=1}^{N} x_{it}^*(\tilde{q}_{it}) + w_i \sum_{i=1}^{N} y_{it}^*(\tilde{q}_{it}). \tag{11}$$

Here, $\tilde{q}_{it}$, defined in Section 5.1, represents the median popularity of the recommended creators shown to user i on day t, as a result of Platform V strategically adjusting their recommender system. The terms x_{it}^* and y_{it}^* denote functions that yield the optimal video watching time of user i on day t and the number of videos uploaded by user i on day t, respectively, based on the optimization model outlined in Equation (6), given the input of $\tilde{q}_{it}$. Let w_i denote the relative value of one unit of content production from user i to one unit of consumption of user i.

Online content-sharing platforms can employ various strategies to assess the value of content production compared to consumption. One simple and straightforward approach used by Platform V is the constant production value, referred to as the "Fixed Production Value": for each user i, w_i equals a fixed constant w_0.[14]

Platforms can also determine production value based on a creator's popularity, assigning more weight to more influential creators. In Online Appendix E, we provided an example of a piecewise linear method that segmented creators into quartiles based on their popularity—from the least popular (0-25%) to the most popular (75-100%). The weights w_1, w_2, w_3, and w_4 were assigned in

[14] We are not allowed to reveal the specific value of w_0 for Platform V.

ascending order ($w_1 \leq w_2 \leq w_3 \leq w_4$), reflecting the tiered influence of creators. The insights from this method are similar to the case of the "Fixed Production Value."

Next, we discussed three distinct recommendation strategies and summarized the counterfactual results from these strategies in Table 11. Detailed explanations can be found in Sections 6.3.1–6.3.3.

Table 8 Counterfactual Analyses

	Recommendation Strategy	Recommended Creator Popularity Percentile Segment 1 (1)	Segment 2 (2)	Segment 3 (3)	Segment 4 (4)	Overall Value (5)
(1)	Consumption-Only Uniform	99th Percentile				1.00
(2)	Balanced Uniform	94th Percentile				1.05
(3)	Balanced Segment-Targeted	71st Percentile	98th Percentile	98th Percentile	99th Percentile	1.07

Notes: The range of $\tilde{q}$ was set from the 1st to the 99th percentile of creator popularity in video supply. The overall value in the Consumption-Only and Uniform Recommendation Strategy is used as the baseline, with relative overall values of other strategies compared to it.

6.3.1. Consumption-Only Uniform Recommendation Strategy. Prevailing research and practices from prominent online content-sharing platforms such as YouTube and TikTok historically have focused their recommender systems on enhancing content consumption, as detailed in Sections 1 and 2. Insights from industry experience suggested that this focus did not necessarily stem from undervaluing content production but rather from organizational structures that compartmentalized production and consumption into separate departments. As a result, those tasked with boosting consumption, despite acknowledging the importance of production, prioritized consumption to meet their specific departmental objectives. In our overall value function framework, this emphasis on consumption led to the assignment of a zero value to w_i for all users.

We considered applying a uniform median popularity of recommendations, denoted as $\tilde{q}$, to each user. We assumed the feasible limits of recommendation adjustments as the 1st and 99th percentile of creator popularity across all videos in supply, which defined the range of varying $\tilde{q}$. The optimal solution, $\tilde{q}^*$, which maximized overall value, was defined mathematically as:

$$\tilde{q}^* = \underset{\tilde{q}}{argmax} \sum_i x_{it}^*(\tilde{q}). \tag{12}$$

This approach was termed the "Consumption-Only Uniform Recommendation Strategy."

Since consumption utility is monotonically positive to the popularity of recommendations, as suggested by the positive values of a_1 in our model, the outcome of this strategy would yield the optimal solution being the maximum of $\tilde{q}$ within the manipulation range, i.e., the 99th percentile of the creator popularity in video supply, which was also confirmed by our simulation based on the model. That aligned with the common recommendation approach discussed in Section 1, which

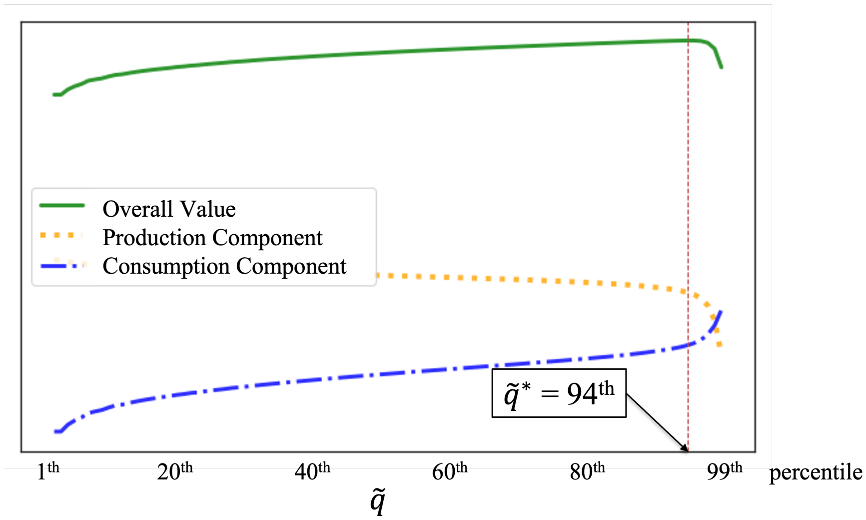


Figure 3 **Illustration of Deriving $\tilde{q}^*$ Under Balanced Uniform Recommendation Strategy**

favored the promotion of highly popular content. Additionally, our data analysis on Platform V revealed that the median popularity level of recommendations matched the 99th percentile of creator popularity in the overall video supply, suggesting that the current recommendation strategy predominantly selected content that had already achieved high popularity (see details in Online Appendix A).

6.3.2. Balanced Uniform Recommendation Strategy. Next, we uniformly applied a single median popularity level, $\tilde{q}$, to all users while recognizing the value of both content production and consumption. We determined the optimal $\tilde{q}^*$ that maximized the overall value of the platform, accounting for production:

$$\tilde{q}^* = \underset{\tilde{q}}{argmax} \left(\sum_{i=1}^{N} x_{it}^*(\tilde{q}) + w_i \sum_{i=1}^{N} y_{it}^*(\tilde{q}) \right), \tag{13}$$

This approach was termed the "Balanced and Uniform Recommendation Strategy."

When we applied the fixed production value across all user segments, unlike the Consumption-Only Uniform Recommendation Strategy, which identified the 99th percentile as optimal, the Balanced Uniform Recommendation Strategy did not always seek the highest levels of popularity. Simulation results showed that the optimal median popularity of recommendations, $\tilde{q}^*$, was at the 94th percentile, as shown in Figure 3. This suggests that this strategy recommends the less popular 94% of creators with higher frequency and more popular creators with lower frequency compared to the Consumption-Only Uniform Recommendation Strategy.

These strategic adjustments directly influenced the platform's overall value. Balanced Uniform Recommendation Strategy achieved a 5% increase in overall value compared to the baseline set by the Consumption-Only Uniform Recommendation Strategy.

6.3.3. Balanced Segment-Targeted Recommendation Strategy. Moreover, beyond including the value from creation, we adopted a more segmented approach, wherein the platform customized the popularity of recommended videos, $\tilde{q}_k$, for each user segment k. This approach was captured in the following mathematical model:

$$\tilde{q}_k^* = \underset{\tilde{q}_k}{argmax} \left(\sum_{i \in k} x_{it}^*(\tilde{q}_k) + w_i \sum_{i \in k} y_{it}^*(\tilde{q}_k) \right), \forall k = 1, 2, \cdots, K, \tag{14}$$

which determined the optimal recommendation level, $\tilde{q}_k^*$, for each segment, thus customizing the platform's recommendation strategy. We referred to this method as the "Balanced and Segment-Targeted Recommendation Strategy."

When we applied the fixed production value across all user segments, this strategy resulted in varying optimal recommendation levels: (1) Segment 1, particularly sensitive to the popularity of recommendations on production, had an optimal recommendation at the 71st percentile, recommending the less popular 71% of creators with higher frequency and more popular creators with lower frequency; (2) Segments 2 and 3, with lower sensitivity to the popularity of recommendations on both consumption and production, reaches the optimal at the 98th percentile, slightly including more less popular creators; (3) Segment 4, highly the popularity of recommendations on consumption, maintained the highest recommendation level at the 99th percentile.

Regarding the overall value increase, the Balanced Segment-Targeted Recommendation Strategy demonstrated significant improvements over the Balanced Uniform Recommendation Strategy. The overall value increased by 7%, relative to the baseline set by the Consumption-Only Uniform Recommendation Strategy. These results underscored the effectiveness of a segmented approach in enhancing platform performance by tailoring recommendations to meet the diverse needs and sensitivities of different user groups.

7. Conclusion and Discussion

This study investigated the dual roles of users on online content-sharing platforms, where they serve both as consumers and creators. Recommender systems, crucial to these platforms, theoretically influence user behavior in both consuming and producing content. Our field experiment causally examined how these systems impact not only what users consume but also their motivation to create. The results revealed that treatment users, who were recommended creators with lower popularity, decreased their content consumption but increased their content production. This suggests that while less popular content may attract fewer views, it may encourage users to contribute their own content.

To further examine these dynamics, we developed a structural model informed by our experiment's results. Utilizing this model, we conducted counterfactual analyses to explore optimal recommendation strategies that balance the effects on content consumption and production. Counterfactual analyses based on our structural estimation revealed that the optimal strategy often involved recommending less popular content to enhance production, challenging current industry practices. Thus, a balanced approach in designing recommender systems is essential to simultaneously foster content consumption and production.

Our research offered several practical implications for platform management. First, adjusting recommender systems impacted both consumption and production, which are critical for the platform's long-term health. Contrary to the intuitive strategy of promoting highly popular videos to maximize viewer engagement, our analysis indicated that this approach might not always be the best strategy.

Second, customizing the popularity of creators in recommendations according to distinct user segments could add substantial value. For instance, users in segment 4 might benefit from recommendations featuring more popular creators. In contrast, users in segment 1 might engage more with recommendations featuring less popular creators. This tailored approach helps the platform enhance overall value by adapting recommendation strategies to the diverse preferences of its user base.

However, our study was not without limitations. One significant constraint was the biased sample; the users in our experiment were active in both consuming and producing content. Most platform users are less active, particularly in content production, and would likely benefit from recommendations featuring the most popular videos. Additionally, our analysis did not account for long-term equilibrium effects. Changes in recommender systems can alter creator popularity over time. For example, a uniform strategy that boosts lesser-known creators may change the content landscape significantly over time. Since the predominant strategy for engaging less active users involves recommending popular videos, the long-term impact on overall platform dynamics may not be substantial. These areas provide fertile ground for further research to better understand the broader implications of recommender systems on user engagement and content production.

References

Agarwal, Nikhil. 2015. An empirical model of the medical match. *American Economic Review* **105**(7) 1939–1978.

Banerjee, Siddhartha, Sujay Sanghavi, Sanjay Shakkottai. 2016. Online collaborative filtering on graphs. *Operations Research* **64**(3) 756–769.

Baym, Nancy K. 2021. *Creator culture: An introduction to global social media entertainment*. NYU Press.

Bimpikis, Kostas, Ozan Candogan, Daniela Saban. 2019. Spatial pricing in ride-sharing networks. *Operations Research* **67**(3) 744–769.

Burtch, Gordon, Yili Hong, Ravi Bapna, Vladas Griskevicius. 2018. Stimulating online reviews by combining financial incentives and social norms. *Management Science* **64**(5) 2065–2082.

Cabral, Luis, Lingfang Li. 2015. A dollar for your thoughts: Feedback-conditional rebates on ebay. *Management Science* **61**(9) 2052–2063.

Charness, Gary, Uri Gneezy, Austin Henderson. 2018. Experimental methods: Measuring effort in economics experiments. *Journal of Economic Behavior & Organization* **149** 74–87.

Charness, Gary, Peter J Kuhn. 2005. Pay inequality, pay secrecy, and effort: theory and evidence.

Chatterjee, Prabirendra, Bo Zhou. 2021. Sponsored content advertising in a two-sided market. *Management Science* **67**(12) 7560–7574.

Chen, Guangying, Tat Chan, Dennis Zhang, Senmao Liu, Yuxiang Wu. 2024. The effects of diversity in algorithmic recommendations on digital content consumption: A field experiment. *Available at SSRN 4365121* .

Chong, Edwin KP, Stanislaw H Żak. 2013. *An introduction to optimization*, vol. 75. John Wiley & Sons.

Chu, Junhong, Puneet Manchanda. 2016. Quantifying cross and direct network effects in online consumer-to-consumer platforms. *Marketing Science* **35**(6) 870–893.

Covington, Paul, Jay Adams, Emre Sargin. 2016. Deep neural networks for youtube recommendations. *Proceedings of the 10th ACM conference on recommender systems*. 191–198.

Daugherty, Terry, Matthew S Eastin, Laura Bright. 2008. Exploring consumer motivations for creating user-generated content. *Journal of interactive advertising* **8**(2) 16–25.

Davidson, James, Benjamin Liebald, Junning Liu, Palash Nandy, Taylor Van Vleet, Ullas Gargi, Sujoy Gupta, Yu He, Mike Lambert, Blake Livingston, et al. 2010. The youtube video recommendation system. *Proceedings of the fourth ACM conference on Recommender systems*. 293–296.

Dixon, Wilfrid J. 1960. Simplified estimation from censored normal samples. *The Annals of Mathematical Statistics* 385–391.

Dobrian, Florin, Vyas Sekar, Asad Awan, Ion Stoica, Dilip Joseph, Aditya Ganjam, Jibin Zhan, Hui Zhang. 2011. Understanding the impact of video quality on user engagement. *ACM SIGCOMM computer communication review* **41**(4) 362–373.

Ederer, Florian. 2010. Feedback and motivation in dynamic tournaments. *Journal of Economics & Management Strategy* **19**(3) 733–769.

Eisenmann, Thomas, Geoffrey Parker, Marshall W Van Alstyne. 2006. Strategies for two-sided markets. *Harvard business review* **84**(10) 92.

Fong, Jessica. 2024. Effects of market size and competition in two-sided markets: Evidence from online dating. *Marketing Science* .

Garcia, David, Pavlin Mavrodiev, Daniele Casati, Frank Schweitzer. 2017. Understanding popularity, reputation, and social influence in the twitter society. *Policy & Internet* **9**(3) 343–364.

Gentzkow, Matthew. 2007. Valuing new goods in a model with complementarity: Online newspapers. *American Economic Review* **97**(3) 713–744.

Geweke, John, Michael Keane, David Runkle. 1994. Alternative computational approaches to inference in the multinomial probit model. *The review of economics and statistics* 609–632.

Ghose, Anindya, Sang Pil Han. 2011. An empirical analysis of user content generation and usage behavior on the mobile internet. *Management Science* **57**(9) 1671–1691.

Goldsmith, Arthur H, Jonathan R Veum, William Darity Jr. 2000. Working hard for the money? efficiency wages and worker effort. *Journal of Economic Psychology* **21**(4) 351–385.

Hajivassiliou, Vassilis A, Paul A Ruud. 1994. Classical estimation methods for ldv models using simulation. *Handbook of econometrics* **4** 2383–2441.

Halaburda, Hanna, Mikołaj Jan Piskorski, Pınar Yıldırım. 2018. Competing by restricting choice: The case of matching platforms. *Management Science* **64**(8) 3574–3594.

Hu, Mandy Mantian, Sha Yang, Daniel Yi Xu. 2019. Understanding the social learning effect in contagious switching behavior. *Management Science* **65**(10) 4771–4794.

Huang, Ni, Gordon Burtch, Bin Gu, Yili Hong, Chen Liang, Kanliang Wang, Dongpu Fu, Bo Yang. 2019. Motivating user-generated content with performance feedback: Evidence from randomized field experiments. *Management Science* **65**(1) 327–345.

Huang, Yan, Param Vir Singh, Anindya Ghose. 2015. A structural model of employee behavioral dynamics in enterprise social media. *Management Science* **61**(12) 2825–2844.

Huang, Yufeng, Bart J Bronnenberg. 2018. Pennies for your thoughts: Costly product consideration and purchase quantity thresholds. *Marketing Science* **37**(6) 1009–1028.

Jiang, Zhenling, Tat Chan, Hai Che, Youwei Wang. 2021. Consumer search and purchase: An empirical investigation of retargeting based on consumer online behaviors. *Marketing Science* **40**(2) 219–240.

Keane, Michael P. 1994. A computationally practical simulation estimator for panel data. *Econometrica: Journal of the Econometric Society* 95–116.

Kim, Changsu, Ming-Hua Jin, Jongheon Kim, Namchul Shin. 2012. User perception of the quality, value, and utility of user-generated content. *Journal of Electronic Commerce Research* **13**(4) 305.

Kim, Dong Soo, Sanghak Lee, Taegyu Hur, Jaehwan Kim, Greg M Allenby. 2023. A direct utility model for access costs and economies of scope. *Management Science* .

Lazear, Edward P. 2000. The power of incentives. *American Economic Review* **90**(2) 410–414.

Li, Haitao, Jiangchuan Liu, Ke Xu, Song Wen. 2012. Understanding video propagation in online social networks. *2012 IEEE 20th International Workshop on Quality of Service*. IEEE, 1–9.

Liu, Jiahui, Peter Dolan, Elin Rønby Pedersen. 2010. Personalized news recommendation based on click behavior. *Proceedings of the 15th international conference on Intelligent user interfaces*. 31–40.

Luca, Michael. 2015. User-generated content and social media. *Handbook of media Economics*, vol. 1. Elsevier, 563–592.

Moldovanu, Benny, Aner Sela. 2008. The optimal allocation of prizes in contests. *40 Years of Research on Rent Seeking 1*. Springer, 615–631.

Mookerjee, Radha, Subodha Kumar, Vijay S Mookerjee. 2017. Optimizing performance-based internet advertisement campaigns. *Operations Research* **65**(1) 38–54.

Moon, Todd K. 1996. The expectation-maximization algorithm. *IEEE Signal processing magazine* **13**(6) 47–60.

Pedroni, Marco, Francesca Pasquali, Simone Carlo. 2014. ≪my friends are my audience≫: Mass-mediation of personal content and relations in facebook. *Observatorio (OBS*)* **8**(3).

Pellegrini, Andrea, Abdul Rawoof Pinjari, Rico Maggi. 2021. A multiple discrete continuous model of time use that accommodates non-additively separable utility functions along with time and monetary budget constraints. *Transportation Research Part A: Policy and Practice* **144** 37–53.

Pittman, Matthew, Annika Abell. 2021. More trust in fewer followers: Diverging effects of popularity metrics and green orientation social media influencers. *Journal of Interactive Marketing* **56**(1) 70–82.

Rafieian, Omid. 2023. Optimizing user engagement through adaptive ad sequencing. *Marketing Science* **42**(5) 910–933.

Ribar, David C. 1995. A structural model of child care and the labor supply of married women. *Journal of labor Economics* **13**(3) 558–597.

Sarafopoulos, Georges. 2015a. Complexity in a duopoly game with homogeneous players, convex, log-linear demand and quadratic cost functions. *Procedia Economics and Finance* **33** 358–366.

Sarafopoulos, Georges. 2015b. Complexity in a monopoly market with a general demand and quadratic cost function. *Procedia Economics and Finance* **19** 122–128.

Shankar, Venkatesh, Barry L Bayus. 2003. Network effects and competition: An empirical analysis of the home video game industry. *Strategic management journal* **24**(4) 375–384.

Sun, Yacheng, Shibo Li, Baohong Sun. 2015. An empirical analysis of consumer purchase decisions under bucket-based price discrimination. *Marketing Science* **34**(5) 646–668.

Wang, Peng, Yunsheng Jiang, Chunxu Xu, Xiaohui Xie. 2019. Overview of content-based click-through rate prediction challenge for video recommendation. *Proceedings of the 27th ACM international conference on multimedia*. 2593–2596.

Wang, Zhi, Lifeng Sun, Wenwu Zhu, Shiqiang Yang, Hongzhi Li, Dapeng Wu. 2012. Joint social and content recommendation for user-generated videos in online social network. *IEEE Transactions on Multimedia* **15**(3) 698–709.

Wooldridge, Jeffrey M. 2015. Control function methods in applied econometrics. *Journal of Human Resources* **50**(2) 420–445.

Xue, Wanqi, Qingpeng Cai, Zhenghai Xue, Shuo Sun, Shuchang Liu, Dong Zheng, Peng Jiang, Kun Gai, Bo An. 2023. Prefrec: Recommender systems with human preferences for reinforcing long-term user engagement .

Zeng, Zhiyu, Hengchen Dai, Dennis J Zhang, Heng Zhang, Renyu Zhang, Zhiwei Xu, Zuo-Jun Max Shen. 2022. The impact of social nudges on user-generated content for social network platforms. *Management Science* .

Zhang, Dennis J, Hengchen Dai, Lingxiu Dong, Fangfang Qi, Nannan Zhang, Xiaofei Liu, Zhongyi Liu, Jiang Yang. 2020. The long-term and spillover effects of price promotions on retailing platforms: Evidence from a large randomized experiment on alibaba. *Management Science* **66**(6) 2589–2609.

Zhang, Kaifu, Theodoros Evgeniou, Vineet Padmanabhan, Emile Richard. 2012. Content contributor management and network effects in a ugc environment. *Marketing Science* **31**(3) 433–447.

Zhang, Mengxia, Lan Luo. 2023. Can consumer-posted photos serve as a leading indicator of restaurant survival? evidence from yelp. *Management Science* **69**(1) 25–50.

Zheng, Yu, Chen Gao, Jingtao Ding, Lingling Yi, Depeng Jin, Yong Li, Meng Wang. 2022. Dvr: micro-video recommendation optimizing watch-time-gain under duration bias. *Proceedings of the 30th ACM International Conference on Multimedia*. 334–345.

Zou, Lixin, Long Xia, Zhuoye Ding, Jiaxing Song, Weidong Liu, Dawei Yin. 2019. Reinforcement learning to optimize long-term user engagement in recommender systems. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2810–2818.

Online Appendices

A. Evidence of Current Recommendation Strategy

We generated a quantile-quantile (Q-Q) plot, Figure 4, to compare the distributions of video popularity in recommendations during our experiment and the overall video supply on Platform V. The popularity from recommendations is displayed along the x-axis, whereas the overall video supply's popularity is mapped to the y-axis. Each coordinate (x, y) on the plot corresponds to a quantile from the recommendation distribution aligned against the same quantile from the video supply distribution. Analysis of this plot reveals that the median (50th percentile) of content popularity in recommendations equates to the 99th percentile of content popularity in production. This figure also indicates that the recommendation algorithms tend to select content that is highly popular.

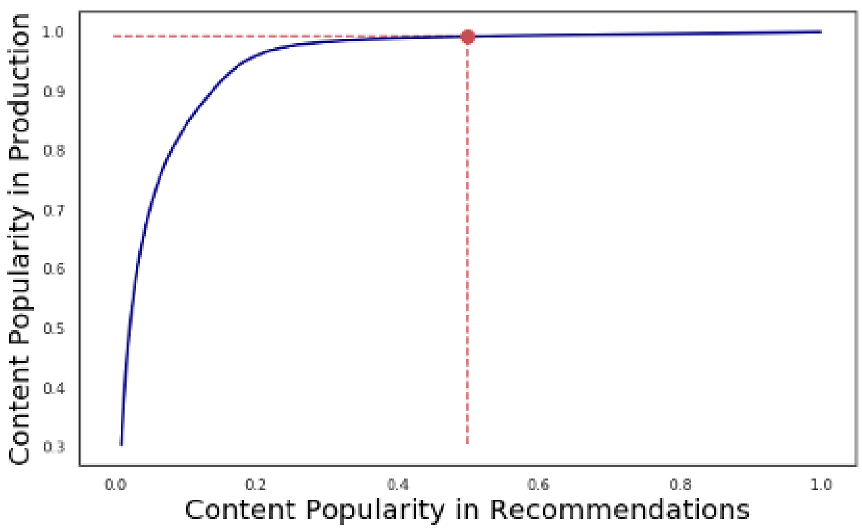


Figure 4 **Q-Q Comparison of Popularity in Recommendations Versus Video Supply on Platform V**

B. Analyses Based on Alternative Metrics for Measuring Creator Popularity

As discussed in Section 3.3, beyond the primary Creator Popularity metric based on the number of followers a user has, we also examined Creator Popularity by analyzing engagement metrics—views, likes, and shares—of videos they uploaded during the seven days leading up to each day, counting only days when at least one video was uploaded. Most engagement on short-video platforms accumulates within the first week of posting. Platform V provided cumulative user engagement data for these videos from their upload date until the end of 2023. All videos included in our study were uploaded before or during January 2020, providing us with a monitoring period of over two years. This long-term tracking allowed for a thorough assessment of their enduring popularity. We calculated the median values of views, likes, and shares per video to obtain a robust measure of popularity, which we also used to calculate the Popularity of Recommendations$_{it}$. Both our primary and alternative popularity metrics are highly positively correlated.

We explored the correlation between the Popularity of Recommendations$_{it}$ metric based on the number of followers and the Popularity of Recommended Videos$_{it}$ metric based on median likes, comments, and shares per video. The correlations were 0.71, 0.75, and 0.68 (p-values<0.0001), respectively.

Additionally, we presented regression results to predict the Popularity of Recommended Videos$_{it}$ based on median likes, comments, and shares per video as shown in Table 9. The regression results confirm that our treatment significantly reduced the Popularity of Recommendations$_{it}$, as indicated by the negative coefficients across all measures—likes, comments, and shares. Statistically significant reductions were observed for likes and shares (p<0.0001), with relative effect sizes of -30.41% for likes, -22.26% for comments, and -27.61% for shares using raw data. For winsorized data, the effect sizes were -26.66% for likes, -20.32% for comments, and -24.21% for shares. These effect sizes indicate substantial decreases in the Popularity of Recommendations$_{it}$, aligning with our treatment's intended effect of reducing exposure to high-popularity content among treated users.

Table 9 Manipulation Check for Our Treatment Using Alternative Popularity Metric

Outcome Variable	Popularity of Recommendations$_{it}$ Based on			Winsorized Popularity of Recommendations$_{it}$ Based on		
	Likes	Comments	Shares	Likes	Comments	Shares
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-0.0471****	-0.0448****	-0.0504****	-0.0977****	-0.1093****	-0.1140****
	(0.0023)	(0.0023)	(0.0023)	(0.0023)	(0.0023)	(0.0023)
Relative Effect Size	-30.41%	-22.26%	-27.61%	-26.66%	-20.32%	-24.21%
Observations	749,290	749,290	749,290	749,290	749,290	749,290

This table includes user-day observations in which the user watched at least one video. The outcome variables are standardized to have a unit standard deviation across all observations during the experiment. Standard errors are clustered at the user level. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001.

C. GHK Sampling Procedure

Denote the variance-covariance matrix of $[\delta_{x,i1}, \delta_{x,i2}, \cdots, \delta_{y,i1}, \delta_{y,i2}, \cdots]'$ as Σ. This vector of random error follows follow a multivariate normal distribution with a mean vector of zero and a covariance matrix Σ as follows:

$$\begin{pmatrix} \delta_{x,i1} = \epsilon_{x,i} + e_{x,i1} \\ \delta_{x,i2} = \epsilon_{x,i} + e_{x,i2} \\ \delta_{x,i3} = \epsilon_{x,i} + e_{x,i3} \\ \vdots \\ \delta_{y,i1} = \epsilon_{y,i} + e_{y,i1} \\ \delta_{y,i2} = \epsilon_{y,i} + e_{y,i2} \\ \delta_{y,i3} = \epsilon_{y,i} + e_{y,i3} \\ \vdots \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ \vdots \end{pmatrix}, \left(\begin{array}{cccc|cccc} 2 & 1 & 1 & \cdots & \rho+\tau & \rho & \rho & \cdots \\ 1 & 2 & 1 & \cdots & \rho & \rho+\tau & \rho & \cdots \\ 1 & 1 & 2 & & \rho & \rho & \rho+\tau & \cdots \\ \vdots & \vdots & & \ddots & \vdots & \vdots & & \ddots \\ \hline \rho+\tau & \rho & \rho & \cdots & 2 & 1 & 1 & \cdots \\ \rho & \rho+\tau & \rho & \cdots & 1 & 2 & 1 & \cdots \\ \rho & \rho & \rho+\tau & \cdots & 1 & 1 & 2 & \\ \vdots & \vdots & & \ddots & \vdots & \vdots & & \ddots \end{array} \right) \right]$$

Denote the probability density function (PDF) and cumulative distribution function (CDF) of a standard normal random variable as ϕ and Φ, respectively. Our detailed sampling and likelihood calculation process is as follows:

1. Sample for the production behavior in the first period.
   (a) If $x_{i1} = 0$, draw $u_{x,i1}$ from the truncated area with an upper bound of $\frac{1-(a_0+a_1\log(\tilde{q}_{i1})+a_2 r_{x,i1})}{\Sigma_{1,1}^{1/2}}$
   (b) If $x_{i1} > 0$, calculate $u_{x,i1}$ as $\frac{x_{i1}+1-(a_0+a_1\log(\tilde{q}_{i1})+a_2 r_{x,i1})}{\Sigma_{1,1}^{1/2}}$
2. Recursively draw $u_{x,i2}, u_{x,i3}, \cdots, u_{x,iT}$
   (a) If $x_{it} = 0$, draw $u_{x,it}$ from the truncated area with $\frac{1-(a_0+a_1\log(\tilde{q}_{it})+a_2 r_{x,it})-\sum_{j=1}^{t-1}\Sigma_{t,j}^{1/2}u_{x,ij}}{\Sigma_{t,t}^{1/2}}$
   (b) If $x_{it} > 0$, calculate $u_{x,it}$ as $\frac{1-(a_0+a_1\log(\tilde{q}_{it})+a_2 r_{x,it})-\sum_{j=1}^{t-1}\Sigma_{t,j}^{1/2}u_{x,ij}}{\Sigma_{t,t}^{1/2}}$
3. Sample for the consumption behavior in the first period
   (a) If $y_{i1} = 0$, draw $u_{y,i1}$ from the truncated area with an upper bound of $\frac{c_1+c_2-b_1\left(\frac{\tilde{q}_{i1}}{q_i}\right)-b_2 r_{y,i1}-\sum_{j=1}^{T}\Sigma_{T+1,j}^{1/2}u_{x,ij}}{\Sigma_{T+1,T+1}^{1/2}}$
   (b) If $y_{i1} = y$, draw $u_{y,i1}$ from the truncated area with a lower bound of $\frac{c_1+c_2(2y-1)-b_1\left(\frac{\tilde{q}_{i1}}{q_i}\right)-b_2 r_{y,i1}}{\Sigma_{T+1,T+1}^{1/2}}$ and an upper bound of $\frac{c_1+c_2(2y+1)-b_1\left(\frac{\tilde{q}_{i1}}{q_i}\right)-b_2 r_{y,i1}-\sum_{j=1}^{T}\Sigma_{T+1,j}^{1/2}u_{x,ij}}{\Sigma_{T+1,T+1}^{1/2}}$
4. Recursively draw $u_{y,i2}, u_{y,i3}, \cdots, u_{y,iT}$
   (a) If $y_{it} = 0$, draw $u_{y,it}$ from the truncated area with an upper bound of $\frac{c_1+c_2-b_1\left(\frac{\tilde{q}_{it}}{q_i}\right)-b_2 r_{y,it}-\sum_{j=1}^{T}\Sigma_{T+1,j}^{1/2}u_{x,ij}-\sum_{j=1}^{t-1}\Sigma_{T+t,T+j}^{1/2}u_{y,ij}}{\Sigma_{T+t,T+t}^{1/2}}$
   (b) If $y_{it} = y$, draw $u_{y,it}$ from the truncated area with a lower bound of $\frac{c_1+c_2(2y-1)-b_1\left(\frac{\tilde{q}_{it}}{q_i}\right)-b_2 r_{y,it}-\sum_{j=1}^{T}\Sigma_{T+1,j}^{1/2}u_{x,ij}-\sum_{j=1}^{t-1}\Sigma_{T+t,T+j}^{1/2}u_{y,ij}}{\Sigma_{T+t,T+t}^{1/2}}$ and an upper bound of $\frac{c_1+c_2(2y+1)-b_1\left(\frac{\tilde{q}_{it}}{q_i}\right)-b_2 r_{y,it}-\sum_{j=1}^{T}\Sigma_{T+1,j}^{1/2}u_{x,ij}-\sum_{j=1}^{t-1}\Sigma_{T+t,T+j}^{1/2}u_{y,ij}}{\Sigma_{T+t,T+t}^{1/2}}$

D. Procedure of EM Algorithm for Latent Class Estimation

In this section, we detailed our algorithm for latent class estimation.

Initialization Step	
- Initialize P_k	Set initial probabilities for each class $k = 1, 2, \ldots, K$
- Initialize Θ_k	Set initial parameter estimates for each class
Iterative Steps	
E Step	Update P_k using current Θ values:
	$P_k = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{L_i(\Theta_k)}{\sum_j L_i(\Theta_j)}\right)$
M Step	Update Θ_k by optimizing:
	$\Theta_k = \arg\max_{\Theta_k}\sum_i \log\left(\sum_j P_j L_i(\Theta_j)\right)$

Table 10 EM Algorithm for Latent Class Estimation

E. Alternative Example of Counterfactual Analyses

We provided another approach that adjusted the production value based on a creator's popularity, assigning more weight to more influential creators. Specifically, Platform V also employed a piecewise linear method that segmented creators into quartiles based on their popularity—from the least popular (0-25%) to the most popular (75-100%). The weights w_1, w_2, w_3, and w_4 were assigned in ascending order ($w_1 \leq w_2 \leq w_3 \leq w_4$), reflecting the tiered influence of creators. This strategy is referred to as the "Piecewise Linear Production Value" and is mathematically defined as follows:

$$w_i = \begin{cases} w_1 & \text{if } q_i \text{ was in the first quartile of creator popularity,} \\ w_2 & \text{if } q_i \text{ was in the second quartile of creator popularity,} \\ w_3 & \text{if } q_i \text{ was in the third quartile of creator popularity,} \\ w_4 & \text{if } q_i \text{ was in the fourth quartile of creator popularity.} \end{cases} \quad \forall \text{ user } i.$$

Table 11 Counterfactual Results with Piecewise Linear Production Value

	Recommendation Strategy	Recommended Creator Popularity Percentile: Segment 1 (1)	Segment 2 (2)	Segment 3 (3)	Segment 4 (4)	Overall Value (5)
(1)	Consumption-Only Uniform	99th Percentile				1.00
(2)	Balanced Uniform	96th Percentile				1.02
(3)	Balanced Segment-Targeted	81st Percentile	98th Percentile	98th Percentile	99th Percentile	1.09

Notes: The range of $\tilde{q}$ was set from the 1st to the 99th percentile of creator popularity in video supply. The overall value in the consumption-focused recommendation strategy is used as the baseline, with relative overall values of other strategies compared to it.

For Balanced Uniform Recommendation Strategy, the optimal $\tilde{q}^*$ adjusted slightly to the 96th percentile, maintaining a focus on a diverse range of creators, though slightly skewed towards more popular content. With a piecewise linear production value, the increase was at 2%.

For Balanced Segment-Targeted Recommendation Strategy, the recommendation levels adjusted slightly, with only Segment 1 experiencing a change in its optimal recommendation level to the 81st percentile. The overall value saw an even more significant improvement, increasing by 9%.

In summary, the results and insights in this Piecewise Linear Production Value case are similar to the Fixed Production Value reported in Section 6.3.