Regulating the Influencer Market

PRELIMINARY, DO NOT CIRCULATE WITHOUT PERMISSION

Click for the Latest Version

Ruobing Han*

January 10, 2023

Abstract

In this paper, I study the economics of influencer markets with evidence from China's leading livestreaming platform. I begin by documenting some facts about the industry. For example, consumer attention is scarce and concentrated at the top. Modeling the demand for influencers as bundling choices from consumers, I find superstar influencers have positive net consumer traffic spillover on some competing influencers. By studying an exogenous shock where the top superstar was banned, I find influencers compete on prices but no clear evidence of competing on product variety. Also, the other major influencers could not recover from the demand loss due to losing a superstar, even when they supplied more labor time by livestreaming longer. Hence the shock to the market structure has a long-standing effect on the platform. My findings have managerial implications for platforms that run influencer markets. Using influencers to grow a platform is path-dependent, and superstar influencers may disintermediate the platform. Given the complementarity between superstars and other influencers, the platform should also be very careful when steering consumers away from the superstars. I use a structural model to distinguish how the complementarity/substitutability and correlation between unobserved consumer preferences drive consumers' choice of influencers.

*Department of Economics, Stanford University (email: robinhan@stanford.edu). I want to thank Nick Bloom, Caroline Hoxby, Han Hong, and Matthew Gentzkow for their invaluable mentorship and advice. I also thank Harikesh Nair, Wesley Hartmann, Navdeep Sahni, Xing Li, Shengmao Cao, Chuan Yu, Tianshi Mu, and Yifan Tang for their great discussion and help. Finally, we thank seminar participants at the Stanford IO workshop, Stanford GSB marketing workshop, and Luohan Academy seminar for their helpful comments. We also thank our data contacts and providers for making this project possible. They include the unnamed e-commerce platform and Luohan Academy. The ability to publish is not contingent on results. We gratefully acknowledge financial support from Stanford Fellowship. All errors are my own.

1 Introduction

Multi-sided platforms often need content creators on the supply side. It is a common business strategy to contract with superstar creators to attract demand and differentiate from competing platforms (Rochet and Tirole, 2003). In this influencer marketing era, this strategy is even more popular. However, brands and platforms may underestimate the cost and risk of relying on superstars. The dominance of superstars can be a likely byproduct of contracting with them. In fact, economists have warned that *“demand are concentrated on a group of best sellers, though there exists a large number of very good and highly substitutable alternatives" (Rosen, 1981).*

Many influencer platforms are highly concentrated, and many have superstars on them. For example, Viya, "the live stream queen", and Li Jiaqi, "the Lipstick King", jointly account for 60% measured either by Gross Merchandise Value (GMV) or viewership on Taobao Live, the livestreaming branch of the leading e-commerce platform in China. “Xinba" on the Kuaishou platform accounts for one-third of GMV in 2019. Worldwide, Kim Kardashian on Instagram, Malcolm Gladwell on New York Times, and Donald Trump and Elon Musk on Twitter are examples of superstars and their platforms. The concentration phenomenon is quite common, as can be seen in Figure 1, which shows the right-skewed market structure of influencer platforms worldwide.

Yet, the concentration of the superstars comes with a cost. Superstar influencers have overwhelming bargaining power with the platform. Instagram rolled back its version change to more videos after Kardashians slammed the app for being TikTok like[1]. Twitter’s most active (ranked 6th in followers) creator Elon Musk even bought the platform. In China, "The Lipstick King" intervened in the pricing decision of internationally well-known brands Loreal[2] forcing it to withdraw promotional listings from the brand’s own flagship stores, when he found that the promotional price of the same item was lower than the price the two sides negotiated for sale in the King’s show[3]. Practitioners have noticed the downside of relying on superstars; some have chosen business models and strategies over which the platform has more control. For example, more decentralized algorithm[4], or relying on grass-roots influencers.

“The China model is heavily dependent on very high-profile influencers who can accumulate a big

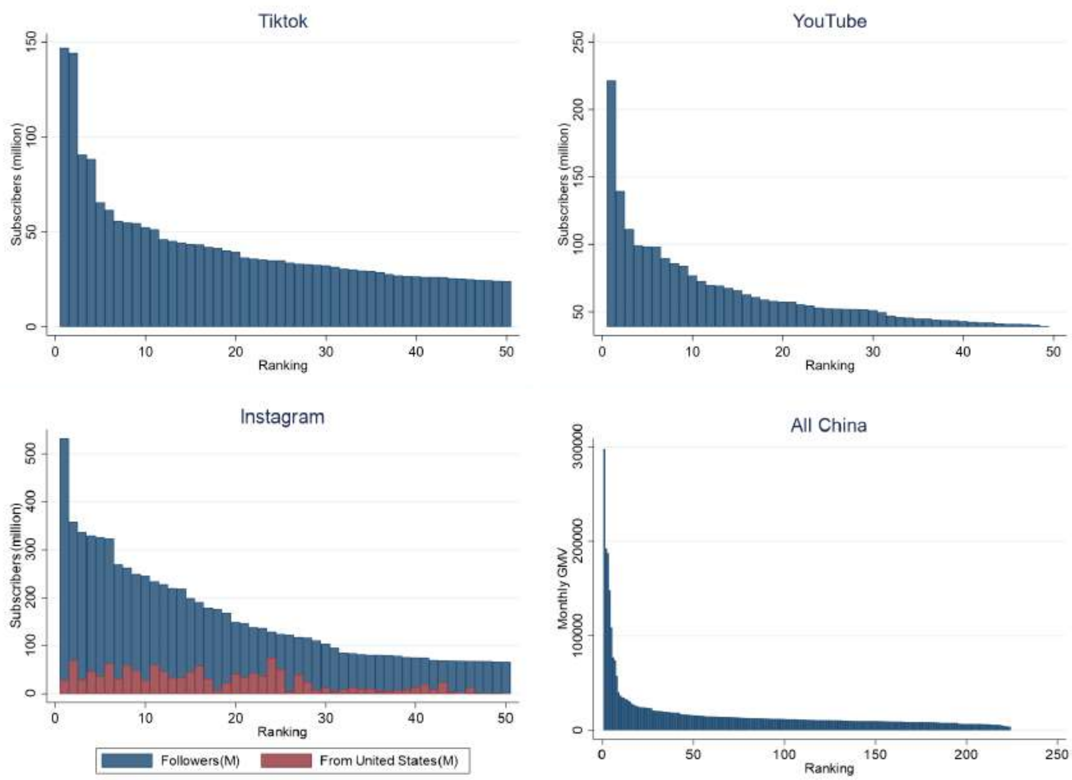
[1]https://www.cnbc.com/2022/07/28/instagram-rolling-back-changes-after-kardashians-tiktok-imitations-.htmlutm_content=Main&utm_medium=Social&utm_source=Facebook&fbclid=IwAR0weKxRfumLXRlBoGRYLpm5q5d74Wg2Wi-LThQ\FwWHMQ49WfPGnFBdRM4k#Echobox=1659036730

[2]https://www.businessoffashion.com/news/china/chinas-top-livestreaming-sales-stars-in-dispute-with-loreal/, https://theconversation.com/western-luxury-brands-are-entering-a-risky-pact-with-chinas-influencers-162055

[3]Both Viya, "the live stream queen" and Li Jiaqi, "the Lipstick King" claim they have the lowest price guarantee nationwide.

[4]https://www.nasdaq.com/articles/how-tiktok-is-democratizing-content-creation-again https://blogs.cornell.edu/info2040/2021/10/31/tiktok-and-information-cascades/.

Figure 1: High right-skewness is observed across many platforms



The bottom right panel shows the total distribution of Chinese influencers from three major platforms: Taobao, Tiktok (Chinese version), and Kuaishou. The other three panels show the number of subscribers of the top 50 influencers on three international platforms (Tiktok, YouTube, and Instagram). Source: `https://hypeauditor.com` *audience through their names ... it costs a lot of money to get traffic to your platform, whether that's money that is going into marketing or going into having a high profile influencer to create news."*

— *Mike George, president and CEO of QVC's parent company Qurate Retail.*

For consumers of products or content, are superstar influencers welfare-enhancing or reducing? The overwhelming presence of superstar influencers can make it harder for consumers to find their best match (with smaller influencers or with the content/products that the superstar influencers do not post or sponsor) and deter new influencers' entry. On the other hand, consumers' search costs can be endogenously reduced with superstar influencers offering better quality content/products, as they are more likely to be found with more followers. Alternatively, they could bargain with the brands/suppliers for the lowest prices by taking advantage of their huge fan base (network effects for the consumers).

Hence, for both consumers and the platform itself, the cost and benefits of adopting the superstar strategy are complicated and ex-ante unclear. Concentration in the advertising and retailing market always concerns regulators. But no one has studied whether and how we should think of the concentration on the market of influencers in this influencer marketing era. This paper has three research objectives using administrative data from the leading e-commerce platform in China, which also pioneered livestreaming influencer marketing. First, using a quasi-experiment, we study the welfare effects of having superstar influencers on consumers. It also helps us understand how livestreaming influencers compete with each other. Second, we model consumers' demand for influencers using the framework of bundling choice. Third, we explore the tension and interaction between superstar influencers and the platform. We try to provide a solution to the platform about how it may regulate/re-design its market by steering consumers. Those parts are interrelated: a concentrated market structure is the result of competition, and it creates subtle tension between the superstars and the platform. The exogenous event that we studied provides clear evidence of how influencers compete and a clear picture of an extreme counterfactual that would be too costly for the platform to implement. The clearly worse platform-wide performance after banning a superstar suggests the platform should try a more mild intervention, hence motivating the consumer steering intervention.

We documented the following stylized facts. First, like many industries and platforms, the platform is dominated by a few superstars. Second, we find assortative matching in the sense that stronger influencers are selling higher-quality products. Although consumers show addiction and loyalty to the superstar influencers, the platform still has intermediation power. The recent natural experiment where the top influencer was banned due to tax evasion helps us analyze how important superstar influencers are to the platform and what is their effects on competitors and consumers. We find that influencers indeed compete on pricing, but we did not find clear evidence on product variety.

To capture the pattern I observed from the event study and ultimately do counterfactuals, I build a structural model of consumer demand. On the demand side, we model the consumers' decision of which influencer(s) to watch, how long to watch, and whether to buy products they saw. We then identify each influencer's attractiveness and the influencer's quality and price dispersion in each category. The (net) effects of superstar influencers on the rest of the platform depend on the net outcome of two countervailing forces: a market expansion effect and a cannibalization effect. We capture this central trade-off of the platform superstar strategy by modeling consumers choosing a bundle of influencers from the platform. If a smaller influencer is often chosen together with the superstar(s), they are seen as a bundle, and if this bundle has a higher probability of being chosen than the probability of the smaller influencer being chosen alone, that means the superstar(s) has positive spillover effect (market expansion effect dominates the cannibalization effect) onto the influencer. The platform, in many senses, is selling a bundle of influencers in competing with the other platforms. I estimated that superstars could have a negative or positive net effect on smaller influencers.

In the next step of this paper, I would simulate counterfactuals where the platform intervenes in consumer searching and matching influencers by promoting smaller influencers, reducing the concentration of the superstar influencers (the queen and the king). Finally, we aim to draw some general implications in the influencer marketing era for the platform as an internal regulator. To some extent, the tension between the platform and its superstar resembles that of many platforms and their top content contributors. For example, the case between Epic and Apple[5] centers around how much Apple can intervene in an independent IOS app's contents and payment options for its own interests. Hence the study may also have broader antitrust implications. The platform's dilemma resembles those facing government or regulatory agencies in many senses. It is mostly about market concentration: income inequality created by superstar effects and market concentration created by super-firms.

These results contribute to several streams of literature. First, it documents interesting stylized facts that help us understand the industry; Second, it applies the demand estimation framework on influencers, helping us understand the platform strategy, the influencers as a retail channel, and how

[5] `https://time.com/6097170/apple-app-store-epic-lawsuit/`. There is another case signals the tension could be especially high when the complementor has characteristics of the platform: `https://appleinsider.com/articles/22/05/06/altstore-allows-limited-sideloading-of-iphone-apps-apple-doesnt-approve`. In that although the 3rd party application does not have many users, it is itself an app store where users can "sideload" many applications that the Apple store does not offer.

consumers shop (including view and purchase) from them. The next step intends to provide a counterfactual analysis of how the platform can design its market when dealing with its superstars by steering consumers.

The paper proceeds as follows. Section 2 reviews the literature. Section 3.2 describes the empirical background and the data. Section 4 provides some facts about livestreaming influencers and the platform. Section 5 studies an exogenous shock where the top influencer disappeared. Section 6 outlines the model. Section 7 describes demand estimation and results. Section 8 considers possible simulations. Section 9 summarizes and concludes the paper.

2 Literature

Our project touches on many threads of the literature. Firstly, consider the marketing literature on hiring superstars to grow platform business (Binken and Stremersch, 2009; Corts and Lederman, 2009). For firms, the demand creation effect by the superstars is well-documented (Krueger, 2005). There is also work on the effects of superstar CEOs on competitors (Ammann et al., 2016). Koenig (2021) looks at the impact of the introduction of television on entertainers' wages in the UK and finds that such new technological change can lead to superstar wages for some and low wages for most others. The phenomena we described can reflect a historical pattern that repeats itself, this time in the online labor market of influencers. In traditional labor markets, salespeople would never have such a large income inequality among themselves. But prior works mostly focus on once-off, single product/brand endorsement by celebrities. They did not study the interaction of stars with multi-sided platforms, whose objectives are more complex, and network effects matter. Those studying platforms do not use a structural model to capture the equilibrium effects of superstars on their competitors. We also contribute to the empirical modeling of influencers and platform competitions when consumers indeed choose a bundle of products. Examples of earlier studies on complementary/ bundling choice include Gentzkow (2007) on print and online newspapers and Cao and Chatterjee (2022) on pharmaceuticals. We apply the framework in this setting to formalize the logic of superstar strategy and empirically capture its key trade-off.

Second, there are a few theory papers on influencers in recent years, Fainmesser and Galeotti (2021) game-theoretically model how influencers act as matchmakers and compete with each other. However, they do not discuss the platform's behavior or regulator of the market. Prat and Valletti (2021) model advertisers or creators in the media markets capturing a sizeable and stable pie of consumers' time hence acting as oligopolies of the attention market. Cong and Li (2021) explain the working of the business model of influencers. They model how brands use influencers for product differentiation. Pei and Mayzlin (2021) discuss brands' optimal affiliation with influencers and the impact of this affiliation on consumer welfare. While theoretical papers help us fix ideas, we are also interested in how empirical facts can test the validity of existing theories. There are few empirical papers on the economics of influencers, probably due to data limitation[6], but also because researchers have not pinned down the theories and their economics in the first place. We attempt to provide empirical findings and theoretical understandings of influencers in this paper. In a similar setting, Förderer and Gutt (2021) document how competing influencers react to the events when a superstar gamer leaves and returns to a gaming platform. In essence, they are also studying the substitution and cannibalization among gamers.

Third, how influencers compete in the first place is understudied. I suggest that influencers have two significant roles: advertising, as described in Bagwell (2007), and as a retailing channel. I am trying to connect the literature on influencers and that on retail competition. And we empirically model these two roles of influencers, borrowing ideas from modeling the competition of multi-category supermarkets (Thomassen et al., 2017). To my knowledge, we are the first to model the competition of product-endorsing influencers using the framework of multi-category retailing.

Fourth, our project also touches on the study of consumer search, summarised in Honka et al. (2017). Switching costs that can make the market less competitive is a common concern (Klemperer, 1987), yet it is debatable as opposite evidence also exists (Dubé et al., 2009). Closely related to consumer search is consumer steering. As we described earlier, one way the platform may try to change the market structure is to change consumers' consideration sets. Barach et al. (2020) studies an online labor platform using money-back guarantees to steer consumers to certified sellers. McManus et al. (2020) studies why and how a content platform steers consumers between its own and competing for content and how this incentive interacts with its investment in providing quality services. Tudón (2022) studies how Amazon's Twitch.tv platform could steer consumers to prioritize certain content over others in a congested market, where the platform's trade-off is between entry and congestion. Although there are indeed many homogeneous influencers, we are not concerned about congestion as in Bai et al. (2020), instead, because the platform's short-run incentives and self-reinforcing algorithmic recommendation have made consumers see and choose from too few influencers, not too many. We are more interested in the superstar influencers being the search default.

[6]There are many empirical marketing papers on the effects of influencer endorsement, live-streaming effects, etc., but they are not related to the objectives of this project.

Finally, it is possible that consumer steering is just a tool of platform governance. A bigger picture problem here is platform regulation and governance/design. Large platform companies, such as the one we studied here, are actually as complex as a micro-economy (Parker and Van Alstyne, 2018). Amaldoss et al. (2021) find the degree of competition between content providers also has different implications for the media platform and consumers, and they investigate different platform design strategies given the landscape of competition. We hope to draw some antitrust implications. We want to argue that the platform's internal governance problem resembles that of an industry regulator's. The advantage here is that we have consumer-level data from all influencers, while usually, a regulator only has aggregate information on the consumers of each firm in its industry. The concentration at the top problem for the platform is a novel example of economic concentration. Rosen (1981) shows in his seminal work on superstar theory that scale-related technological change may drive higher income inequality, especially within the top earners. He predicts technology advancements would enable "many of the top practitioners to operate at a national or even international scale". Livestreaming, as a channel in e-commerce, enabled many celebrity online shopping guides to reach millions of consumers and endorse products at a much faster pace. Moreover, economists are always concerned about advertisements creating product market concentration, and recent anti-trust claims against dominant search engines such as Google are partially based on the fear that it controls people's search results, hence the results of many economic activities. For similar reasons, superstars who are very efficient advertising channels and grab the most attention can create market power for themselves and the brands that afford to hire them. Understanding these high-profile influencers, what their market power is, whether they have market power and whether they exercise it would determine how the platform, brands, and consumers should embrace them.

3 Empirical Setting and Data

3.1 The Platform and Livestreaming

The platform under focus is a leading e-commerce platform in China that hosts millions of active sellers and consumers and thousands of influencers as well. Total sales on the platform represent a sizable share of all domestic e-commerce sales in China, and its share in influencer-involved sales is even larger. Sellers on the platform offer various types of products. While influencers cover most of the categories, they especially focus on the most horizontally differentiated categories, such as clothing, cosmetics, home appliances, consumer electronics, and food. The majority, if not all, are brands/manufacturers rather than third-party retailers who contract with professional star influencers. The platform earns revenue from sellers and influencers by offering advertisements, charging commissions, and selling supplementary services[7]. Becoming a registered influencer on the platform is easy.

The arrival of livestreaming e-commerce in May 2016 marked the opening of a new chapter in retailing. The Chinese retail giant pioneered a powerful new approach: linking online livestreaming influencers with e-commerce brands to allow viewers to watch and shop simultaneously. As a result, livestream e-commerce quickly became a reliable digital tool for boosting customer engagement and sales. In 2021, the first 30 minutes of the platform's Singles' Day (11th of November) sales campaign on its livestreaming channel generated an impressive $7.5 billion GMV.

Livestreaming commerce combines instant purchasing of featured products and audience participation through a chat function or reaction buttons. In China, livestreaming marketing has transformed the retail industry and established itself as a significant sales channel in less than five years. In a 2020 survey, two-thirds of Chinese consumers said they had bought products via livestreaming in the past year.[8]

To set up a livestreaming event, brands typically contract with a key opinion leader (KOL) or influencer to host the show, introduce the product, and interact with the audience to trigger sales. We study the market of such influencers. Top influencers themselves can function as two-sided markets. They attract consumers' attention through their performance, interact with viewers and persuade them to buy products supplied by the brands that contract with them. The prices paid by the brands to influencers are typically in the form of a two-part tariff: a fixed base payment independent of sales plus commission fees proportional to the overall gross merchandise value sold during the show. The ratio could be as higher as 20%. Depending on the complementary services, including advertisements, exposure, and consumer traffic distributed, 10% to 30% of this commission fee is paid to the platform as the services fee.

[7]Platform's self-brand products similar to Amazon labels consist of a very small portion of its business.

[8]For a more detailed introduction, see https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/its-showtime-how-live-commerce-is-transforming-the-shopping-experience, and https://www.nngroup.com/articles/livestream-ecommerce-china/

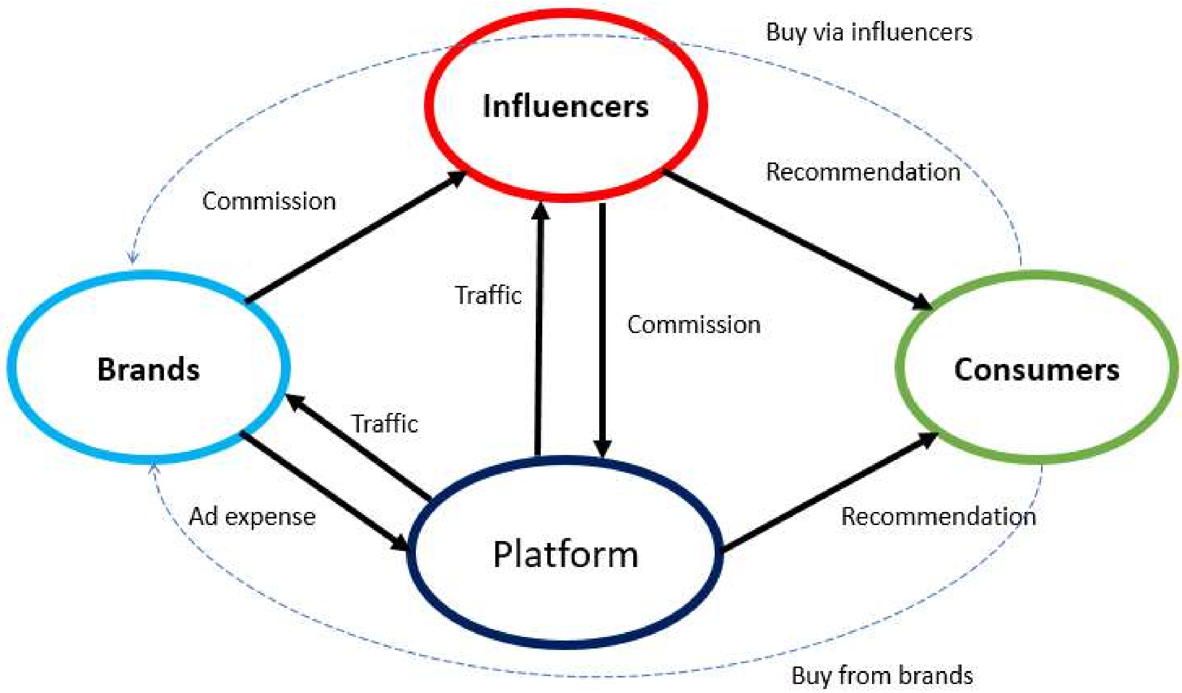


Figure 2: Platform Structure

The above Figure 2 shows a simplified version of the ecosystem on the livestreaming platform. "Traffic" flows are the consumer visits the platform directs to brands or influencers via recommendations or search results. It can be either organic (automatically determined by algorithm) or sponsored (purchased from the platform or freely given by the platform as a form of intentional consumer steering). The "recommendation" on the bottom-right is the traditional e-commerce recommendation (rank-listed recommendation on the front page or research result). The "recommendation" on the top-right refers to livestreaming. The "commission" in the middle is the service fee paid by the influencer to the platform.

Over 90% of consumers browsing the platform and almost all consumers browsing the livestreaming function do so from mobile devices rather than from the web[9]. The limited space on screens on mobile devices intensifies competition for ranking. Hence, page rank, default recommendation, and default search results for influencers are even more important in determining what consumers view and buy.

Like many platforms, the e-commerce platform employs a business strategy to use influencers as matching technology to better match brands/sellers and consumers on the platform. Yet this strategy is highly debatable even inside the company. Beyond the competition among influencers themselves, the top influencers bring trade-offs to the platform: although they can contribute the most GMV (hence the commission fee to the platform), they also crowd out platform ad revenue as brands/sellers turn

[9]Most Chinese consumers browse and shop online with mobile devices.

to influencers for promotion, so there is a substitution of their marketing expenses on the page-rank advertisement to the payment to influencers. Maximizing the platform's profit or revenue is complicated due to this subtle trade-off. Also, as an internal regulator, the platform cares about its market structure, which affects its competition with other e-commerce and livestreaming platforms, such as JD.com and Tiktok. Besides, the platform may desire a less concentrated market structure out of fairness and long-term concerns. Hence, the natural question is, how can the platform make the influencer market less concentrated? In fact, the competing platform Kuaishou brought its only superstar to court for steering consumers away from him[10].

3.2 Data

We use the internal dataset from the leading Chinese e-commerce platform. The sample period we picked for the paper is from 1st July 2020 to 8th May 2021.

There are thousands of influencers. We select the top 25 influencers in our sample for the reduced-form analysis, and in the structural modeling part, we only focus on the top 6 for the feasibility of estimation. There are two criteria for selecting those major players: first, we rank all influencers by the times they enter the weekly top-50 GMV ranking during the sample period of two years. Second, we rank all influencers by how many consumers have ever watched them. The top 25 influencers selected from the first criteria coincide with those who would be selected by the second criteria. In addition, those who made the most time on the weekly ranking list are at least "known" by 1% of the consumers, as the 25th influencers ranked by the second criteria are watched by more than 1% of the consumers while the 26th influencer is not. It is also a common practice to focus on major players, such as Illanes and Moshary (2020) and Thomassen et al. (2017), where they focus on mass merchandiser. For every live-streaming event on the platform conducted by each influencer, we can observe a list of items introduced and their prices and sales during the event. In addition, commission fees for each live-streaming event can also be observed. We can link items sold during the live-streaming event with prices and sales outside the live-streaming room and competitors' prices and sales.

For each influencer, we know the following characteristics: the probability of recommending items in each category and the price distributions, market shares in terms of the number of viewers, the distribution of watching length, and the distribution of purchasing probability within each category in each show. We drop all visits to shows less or equal to 10 seconds, to get rid of noisy visits.

[10]https://sg.news.yahoo.com/why-kuaishou-wants-pivot-away-110329666.html

We sampled 19625 representative consumers, we know their demographics, such as age, income bracket, gender, and city. We also have rich consumer behavioral data, including viewership, exposure, click, and purchase. For each viewer, we can observe what influencers are exposed to them when they start to watch a living-streaming event, what items are exposed to them, what they click on the app/screen, and whether they purchased the item and when he exits the live-streaming room. We can also observe the purchasing behavior on the platform outside the live-streaming room.

From now on, I refer to the two superstars of the platform, the King and the Queen (who were banned), for simplicity.

3.3 Descriptive Statistics

Below are the summary statistics of some key variables. All variables suggest large consumer heterogeneity.

Table 1: Summary Stats

	(1)	(2)	(3)
Variable	N	Mean	S.D.
Shows viewed per week	19625	2.14	4.67
Spending per week (RMB)	19625	61.3	743.4
Purchased item per week	19625	0.65	5.37

The table below shows the average consumer watching time per show by each influencer. There are a few influencers who have longer average visit times than the Queen. There are two possibilities to explain this. First, the Queen is not the most entertaining influencer. Second, she is very efficient at getting the deal done, including conveying product information and persuading consumers to buy.

Table 2: Watch Time (min) Per Show

Influencer	Mean	S.D.
	1.8804154	5.2140573
	1.1412888	3.7055984
	3.2125336	9.7294317
	1.3923689	4.0692488
	2.4436986	6.4147551
	2.1415293	6.2366201
	3.0951229	8.409288
	1.6678172	4.5753134
	1.7837677	5.3495078
	3.2286173	9.6509661
	3.6997843	10.165502
	1.8649982	5.1443992
	1.1867118	3.210017
	2.3561105	3.8671382
	2.159825	5.7863362
King	3.2625159	8.2420549
	1.4900249	3.0439204
	2.0646547	6.0888695
	3.1772044	9.3266662
	1.9171925	5.4487257
	1.8309265	5.4233965
	1.2257467	2.5968374
Queen	2.9509154	7.9820533
	2.479847	7.1432353
	1.3688128	4.0964129
	2.4104	6.849

4 Stylized Facts

In this section, I begin by documenting a set of new stylized facts about the platform and its influencers. These facts provide suggestive evidence of the roles and effects of influencers, especially superstars, and inform our modeling choice.

4.1 Dominance of superstars

As can be seen from the following plots, the top two influencers combined account for 60% market share in both views and sales in all livestreaming influencers.

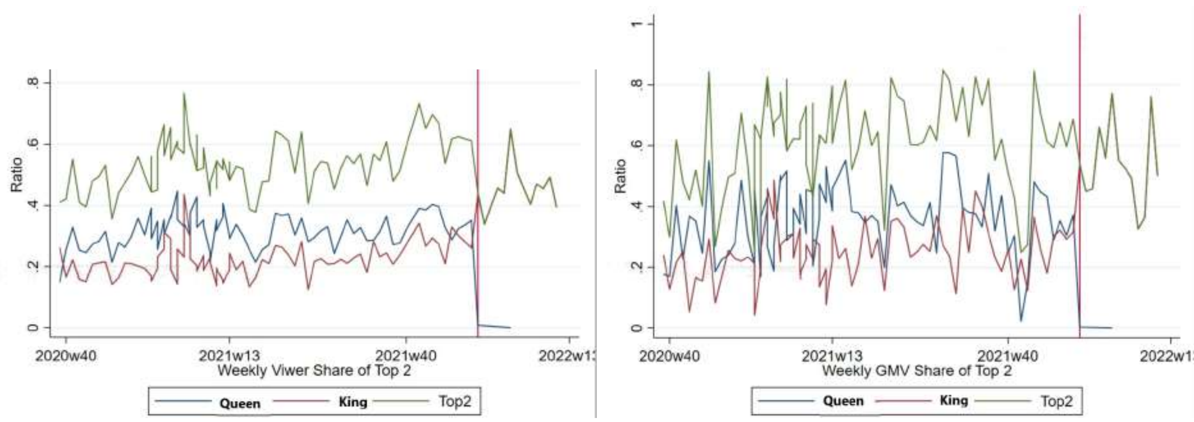


Figure 3: GMV and attention concentrated on Top 2

One feature of livestreaming is that consumers' attention is very limited. On average, they only view about 2.5 different influencers per week, very likely to be the top two superstars.

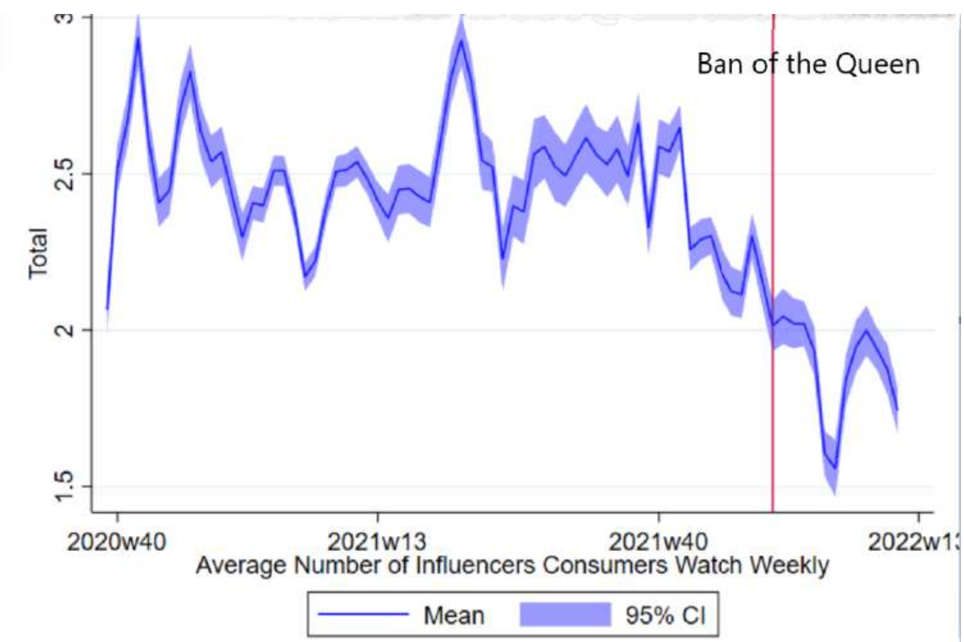


Figure 4: Scarce Attention

4.2 Driving forces for superstar dominance

First, it is crucial to understand why we have the current market structure on the influencer platform. There are at least the following reasons for it. First, there is the tipping (demand-created network effect) effect commonly seen on two-sided platforms: star influencers have a huge fan base hence they can bargain the lowest price from upstream suppliers (brands). The lowest price, in turn, attracts more consumers, creating positive feedback loops (The consumers' utility increases in the number of products indirectly). Second, the superstar effect is studied in labor economics by Rosen (1981). Third, word of mouth/herding/reputation interacts with the first point. Some literature studies how those channels lead to popular and unpopular influencers, but those are not the focus of the paper. Last but not least, the influencers' innate ability/quality/fame are different, but these parameters are treated as exogenous. In other words, we are not trying to model the dynamics of influencers' learning and growth.

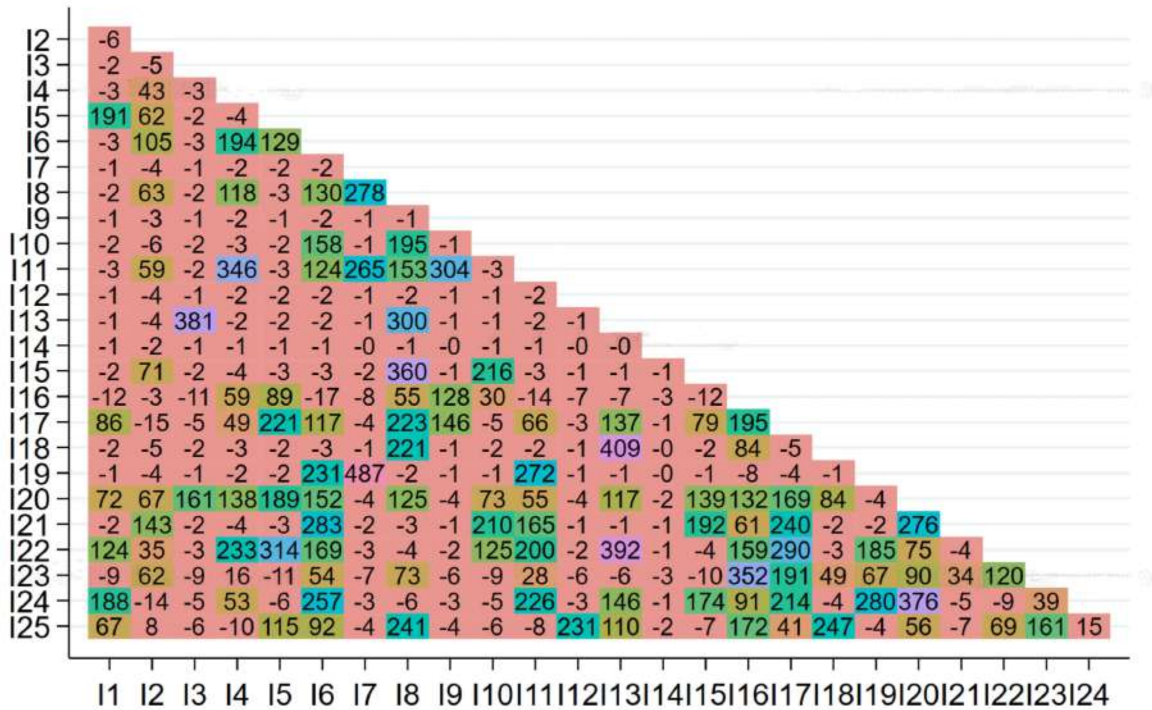
In our story, an influencer can gain market power in two ways. First, they may introduce products with better quality and lower prices. Second, their performances are more attractive (influencer fixed effects). We will consider all those when we model the competition of influencers (more generally, creators of any type who need to monetize their influence in the end).

4.3 Cannibalization, Market creation, and Spillover

The rationale for the platform's superstar strategy is its market creation effect, in that they bring a lot of consumers to the platform. The more competition from the competing platforms, the more important this effect is. There is a cannibalization effect from superstar influencers to other influencers in terms of consumer attention and sales. The platform and other influencers care about the net spillover effect, or the overall value the superstar influencers bring to them[11] if we define the spillover effect as the market creation effect net of the cannibalization effect. In Figure 5 and 6, I plot all partial correlation coefficients for each chosen pair of influencers, controlling consumer characteristics.

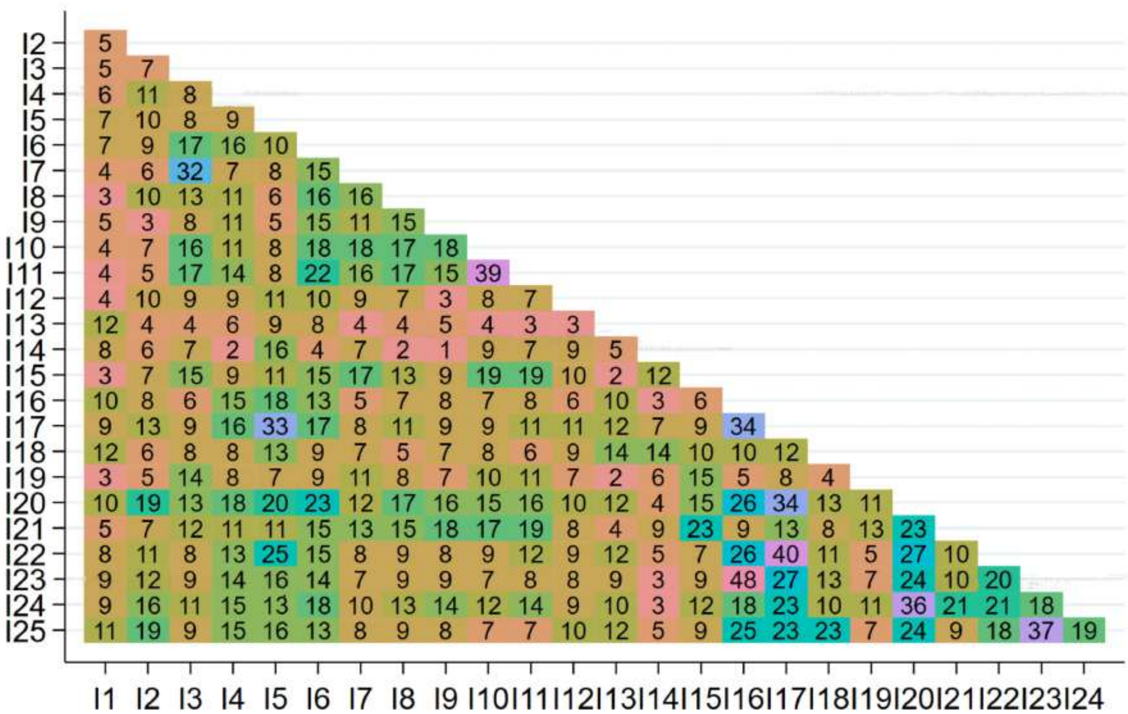
[11] Even for the platform, if the market creation effect of an influencer is positive but small, less than the cannibalization effect, that means the influencer is getting more from the platform than contributing to it in the influencer-platform contract.

Figure 5: Daily Partial Correlation Coefficients



Notes: All coefficients are multiplied by 100. Standard errors are omitted to save some space as all coefficients are significant at 1% level.

Figure 6: Weekly Partial Correlation Coefficients



Notes: All coefficients are multiplied by 100. Standard errors are omitted to save some space as all coefficients are significant at 1% level.

In the frequency of days, influencers can be substitutes or complements pairwise. Note that I23 is the Queen and I16 is the King, so there could be both winners and losers from co-existing on the same platform as the superstars.

The above pattern suggests that star influencers may have net positive spillover to the platform. Hence the overall superstar strategy is successful. But to identify the net effects of the superstar on the rest of the platform causally, we need a model, as the pattern here could be driven by common demand shock or unobservables such as consumer preferences.

4.4 Tension and Platform's Trade-off

Given that the superstars contribute substantially to the platform and may have net positive spillover to the rest of the platform, why should the platform be concerned? This is because the platform could not be monopolistic, given the superstars could be platforms themselves even if currently there are no competing platforms. And also, the superstars may have market power, so the platform could not fully internalize the growth of superstars or extract all surplus they created. We use a simple model to formalize and illustrate this point in Appendix A2. Anecdotally, employees in the platform company told me[12] superstar influencers contribute most commission fees to the platform, but sellers turn to them for marketing. Hence platform traditional ad (page-rank) revenues decrease. There could also be a short-term versus long-term trade-off: if the platform is myopic, the short-term revenue-maximizing strategy is to have all consumers watching the most productive (in terms of GMV contribution) influencers. But the platform does not want its functions to be replaced/disintermediated by the superstars. There is anecdotal evidence that brands and the platform company are concerned that the bargaining powers of these top 2 influencers are too large[13]. Also, for longer terms or platform competition concerns, the platform may care about smaller influencers' surplus, their entries and exits. Moreover, the superstar influencers could leave the platform and become competing platforms themselves. There is another risk diversification reason: the event study in section 5 shows the unfavorable consequences of being too reliant on superstars. Motivated by the above points, we do not analyze about contract design problem as in Bhargava (2021).

4.5 Assortative Matching

We find a positive assortative matching pattern between sellers and influencers in the following sense: in the matching between sellers and influencers, the more famous or larger (measured by recent GMV records and fans number) the influencers, the higher the quality (measured by the price percentile within the category, and the up-to-date consumer rating of the item) of the products the endorsed/provided by the sellers, controlling for the size of the brands (there are many measurements of "size", here we use the level of the brands recognized on the platform to proxy its size; Calculation of "level" mostly depends on cumulative, past sales records). In other words, brands with higher-quality goods can afford to hire a more prestigious influencer. Note this is a two-way selection (and we are not claiming anything causal here) in that the influencers also pick brands they think are trustworthy, products that they think are of high quality and can be sold a lot. The higher the level of the influencers, the more selective they are. There is a concern that brands do not need to hire expensive influencers to advocate their high-quality products, or similarly, brands need to "bribe" influencers to borrow their reputation for low-rating

[12]Can supply with data.

[13]https://www.businessoffashion.com/news/china/chinas-top-livestreaming-sales-stars-in-dispute-with-loreal https://theconversation.com/western-luxury-brands-are-entering-a-risky-pact-with-chinas-influencers-162055 https://sg.news.yahoo.com/why-kuaishou-wants-pivot-away-110329666.html

products. It turns out this is not the case in equilibrium. Note in the first 4 columns of Table 3, when we control for category fixed effect, we are comparing within each category. Item price is equivalent to the price percentile in measuring quality. This matching pattern may increase the inequality among brands[14].

Table 3: Matching

	log(gmv)	log(fans)	log(gmv)	log(fans)	log(gmv)	log(fans)	log(gmv)	log(fans)
Price*100			4.127*10e-2***	1.534*10e-2***				
			(2.015*10e-4)	(1.861*10e-4)				
log(Price)	0.208***	0.110***						
	(0.0008)	(0.0008)						
Category FE	Y	Y	Y	Y				
Brands Level	Y	Y	Y	Y				
Rating*100					9.72***	68.6***		
					(0.549)	(0.533)		
log(Rating)							0.431***	3.172***
							(0.0257)	(.0249)
Brands Level					Y	Y	Y	Y

Standard errors in parentheses, clustered at individual level

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: The sample here consists of 15k brands 1500 influencers 25k items 130k shows. This is the only sample that differs from the one we describe in Section 3.2 because the sample with 19625 consumers and top 25 influencers may lack power for the purpose here.

4.6 Watch and Purchasing Pattern

As we have seen in Table 2, watching time is very short, making us wonder how important watching decisions are alongside purchasing decisions. How are watching and purchasing behavior correlated for consumers? Regressing purchase decision (conversion) on time watched in the show, we find that watching time is important in determining whether to purchase, and it also shows marginal decreasing returns.

Table 4

	(1)	(2)	(3)
Variable	Buy	#Orders	Per minute Orders
Time per visit	0.0146***		
	(0.0000761)		
Time per show		0.0125***	-0.0130***
		(0.00024)	(0.000207)
N	710556	26472	26472

[14] Even if their joint utility is supermodular in their respective quality, in the framework of non-transferable utility, there can be multiple Pareto efficient equilibriums.

4.7 Platform Intermediation Power

Is platform steering a promising remedy to concentration? First of all, we check whether more exposure leads to more clicks to watch. Figure 7 plots the probability of viewing the influencer by a consumer (Y-axis) against the times the influencer is exposed to the consumer (X-axis) on each week.

Figure 7: Exposure Effectiveness

Notes: Sample from 1st July 2020 to 8th May 2022. The unit of observation is consumer-week-influencer.

Below we run the corresponding regressions.

DV: Viewed	Exposure	Log(Exposure)	IHS(Exposure)
Exposure	0.0665***	0.405***	0.32***
	(0.00695)	(0.005)	(0.0037)
N (consumer-influencer-week)	1922382	1922382	1922382

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The significant upward slope here suggests the platform has intermediation power, i.e., the ability to influence market outcomes by steering consumers. In Appendix A3, we characterize the platform's exposure decision.

5 Event Study

First of all, just how loyal and addicted are consumers to the superstars? Do consumers' subscriptions and following influencers matter? Formally, if the intermediation power[15] of the influencers and switch cost of consumers are not large, the platform would worry less about superstar influencers. In this section, we are going to exploit an exogenous shock that changed the market structure from a duopoly to a monopoly. The top influencer, "the Queen", was suddenly banned due to tax evasion on Dec 20th, 2021[16]. This natural experiment creates a counterfactual scenario that the platform did not experiment with. 42.48% consumers were fans of the Queen[17] before the ban. Most such studies of shocks to superstars such as Förderer and Gutt (2021), or in general, studies as a major player entering or exiting the market, focus on the supply side change due to lack of demand-side consumer-level data. Here, however, we can really track how followers/fans of a superstar turn to other influencers. The difference between the behaviors of fans and non-fans of the superstar has implications for the platform, which is still eager to know what happens if they let another superstar (the King) leave.

5.1 Trends of Fans and non-fans

It is clear the Queen's fans are more valuable than non-fans to the platform. The red vertical line marks the 51st week of 2021 when the ban occurred. In Figure 8, we can see that before the event, fans are much more active than the non-fans in the four major measures. Read from Table A2 in the appendix: every week, they are buying 0.469 more items, spending 39.75 more RMB[18], watching 7.427 more minutes, and 0.643 more shows. Note there is no seasonality concern here as the dates back to mid-2020, one and a half years ago, and the large gap is persistent over time. Although the gaps are much smaller after the ban, fans are still more active than non-fans.

Looking into the details on the demand side, we can see how consumers turn to other influencers or become inactive on the platform. This would also determine whether we need to model consumers' decisions on whether to watch livestreaming shows at all. Figure 9 shows the ratio of consumers watching N shows, grouped by fans and non-fans again, in the following 23 weeks (to the end of our sample period). The minimal total number of shows watched is two for the fans and three for the non-

[15]Intermediation power defined similarly as in Lee and Musolff (2021), in that consumers only choose the superstar when making choices.

[16]`https://www.bbc.com/news/world-asia-china-59732499`.

[17]Defined as watching at least one show of her per week before the event.

[18]The conversion rate between $USD and RMB is about 6.7, and 40 RMB is about 5.97 $USD.

Figure 8: Group Means of Fans and non-fans

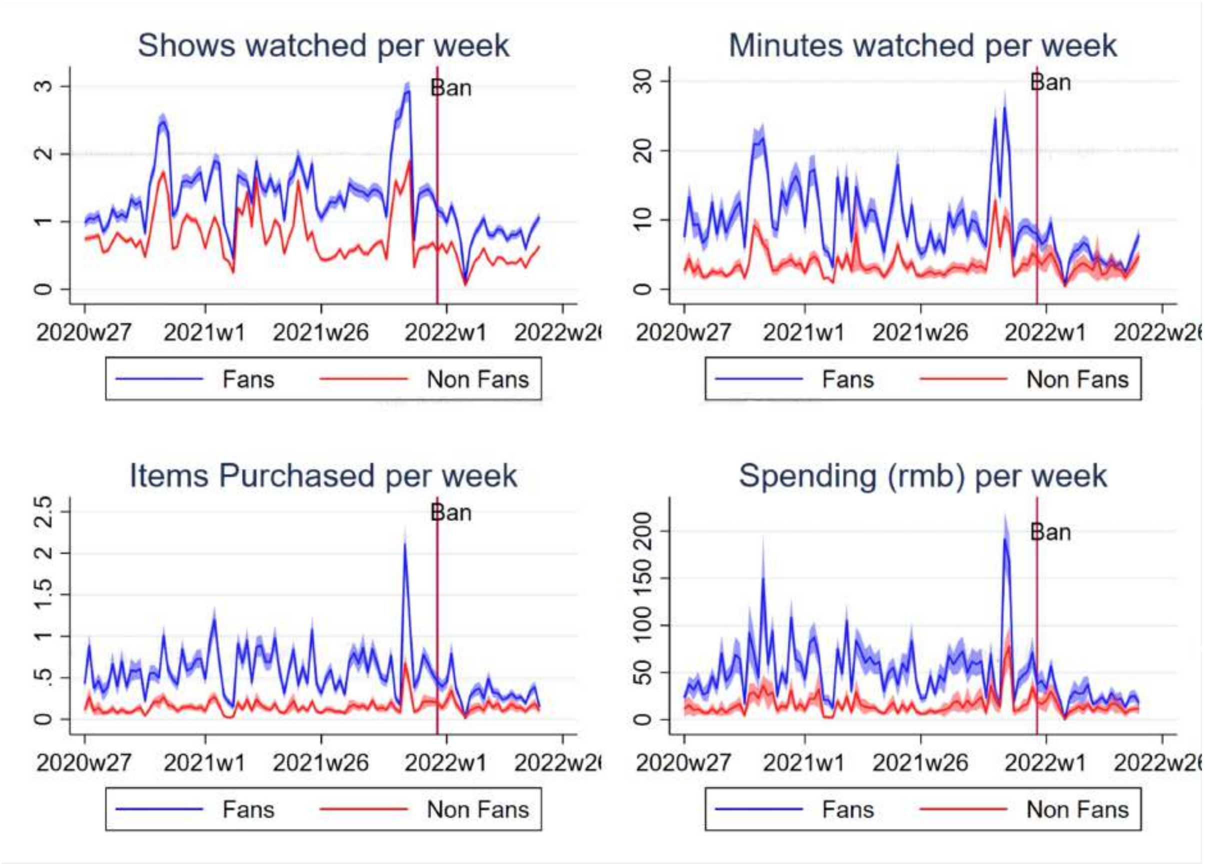
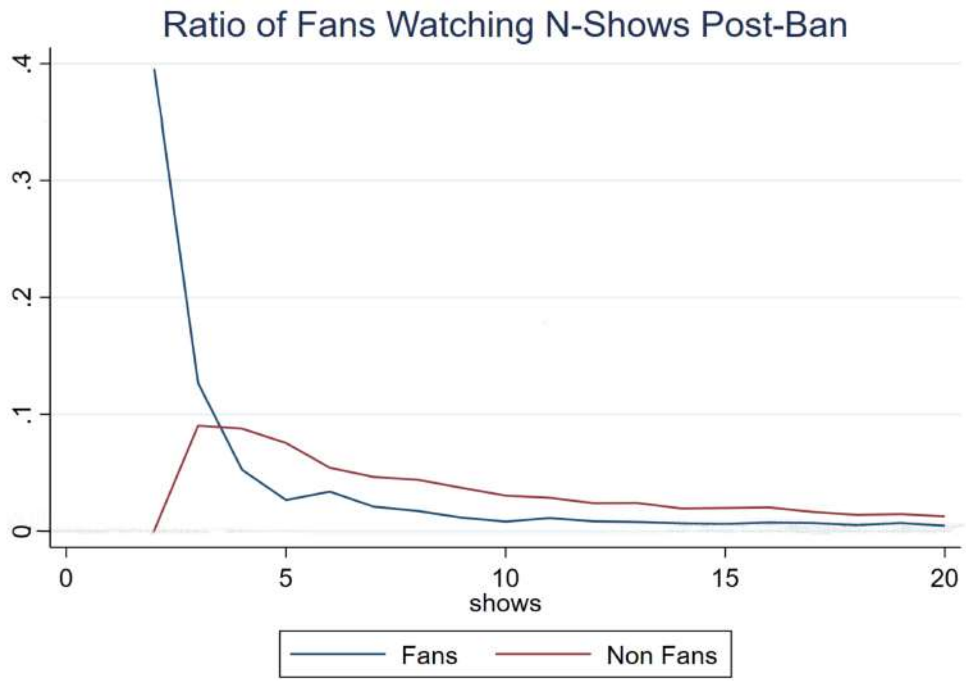


Figure 9



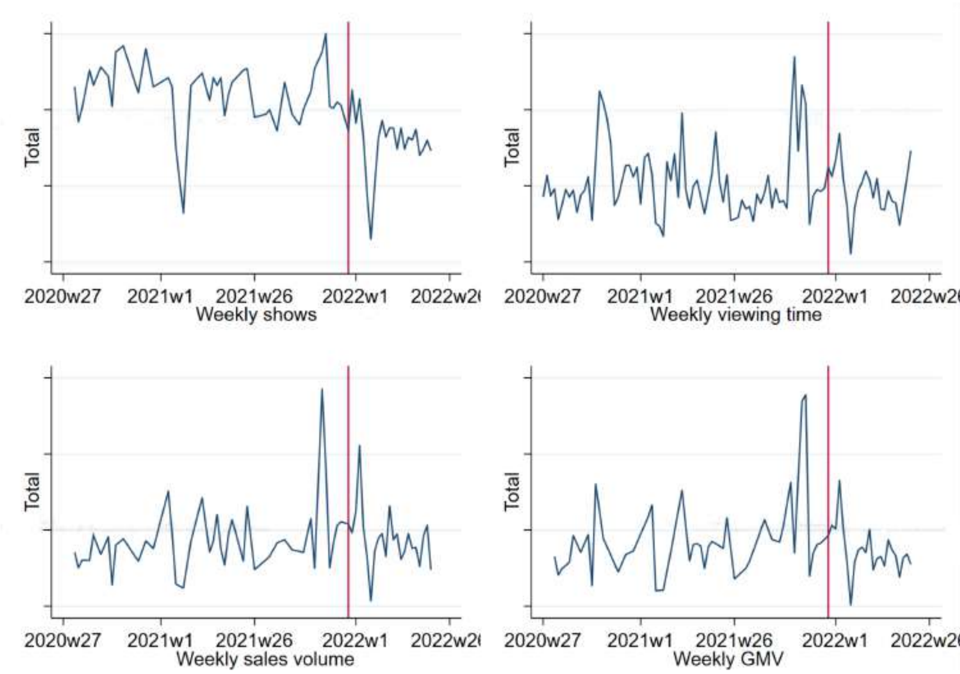
fans. If we define inactive consumers as those who watched less than 5 shows in the 23 post-ban weeks, 57.51% of her fans switched to inactive, and 17.86% of the non-fans platform consumers switched to inactive. On average, 34.70% of consumers became inactive post-event. This suggests that superstars have intermediation power/market power. This is why the risk-averse platform wants to decentralize its influencer markets, steering consumers away from superstars. The above facts suggest a large switching cost of consumers from superstar influencers to other influencers.

5.2 Platform Trends

It is not surprising that the platform's overall performance, which is just a weighted sum of fans and non-fans groups, trends down, shown in Figure 10. Event Study-DID plots are in Appendix **??**. While the total platform traffic decreases, what matters for the other influencers is whether their viewers increase due to less competition. We will formally check this in subsection 5.6.

Theoretical models make different predictions about the effect of market structure on factors such as prices (Bresnahan and Reiss, 1991) and on product variety (Stole, 2007). Hence, it is an ex-ante unclear empirical question of how a change in the market structure affects those factors. How the participants react to market structure change would suggest how they compete (Illanes and Moshary, 2020). Of

Figure 10: Platform Trends



course, every margin could be important, but the margins that respond significantly would be those that matter.

If the removal of a dominant influencer does not reduce product variety and does not increase prices (controlling for the same items), then very likely it does not reduce consumer welfare.

5.3 Empirical Strategy

To check the change of those margins in response to the event of a ban, we may just check the first difference[19]. Then the largest difficulty in interpreting the change after the event study comes from seasonality: the event happens at the end of the year 2021, two months ahead of the Chinese lunar new year. There is concern that this end-of-year timing confounds the exogenous event of banning the Queen. Since new year shopping usually boosts demand, indicating we may be underestimating the magnitude of the negative shock. To control for this confounding factor of seasonality, we use the data from the previous year as the control group for a difference-in-differences style regression. Other than the "control year", it can also be thought of as a tool to detrend the data in the current year.

The Chinese lunar new year holiday in 2022 is the 5th week, with new year's eve on the 12th of

[19]Equivalently, one may consider this as a regression discontinuity with respect to time.

February. In the previous year, the Chinese lunar new year holiday was the 7th week, with the eve on 1st February 2021. The Figure 11 timeline shows how we define the "control year" from the 26th week of 2020 to the 20th week of 2021 as the control year. Because the event happens on the 20th of December, which falls into the 51st week of 2021, seven weeks before the new year holiday week. We define the hypothetical event that would have happened in the 1st week of 2021, seven weeks before the previous new year holiday week. In this way, both Chinese lunar new year weeks would be in relative week 7 of each (treatment and control) year. And the "treated year" is from the 26th week of 2021 to the 18th week of 2022. We could add in more data from previous years as control years. However, this moves us away from the cut-off date when the treatment of the ban begins. Another concern is that the year 2020 will be affected by Covid, as can be seen in Figure A3. Since the second half of 2021, China has had an extremely low Covid case rate, livestreaming business was boosted by the pandemic, especially the lockdowns in the first quarter of 2020. But the situation is flat across our sample period as the pandemic stabilized. The impact Covid may have on online shopping would not affect our results here.

The Event Study DID regression I ran is:

$$Y_{it} = \sum_{k \neq -1} I\{t = k\} \times \text{Year}_i + \phi_s + \gamma_k + \epsilon_{it} \tag{5.3.1}$$

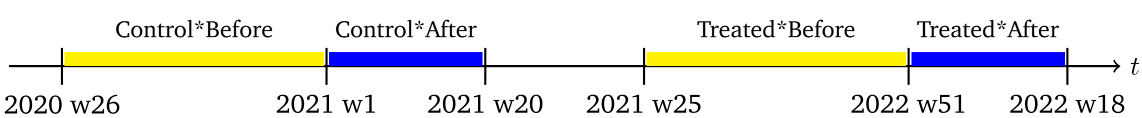
ϕ_s is the ("control" or "treated") year fixed effect. γ_k is the relative week fixed effect.

The DID style regression we run are

$$Y_{ik} = \text{Year}_i + I\{k > 0\} + \text{Year}_i \times I\{k > 0\} + K + \epsilon_{ik} \tag{5.3.2}$$

where k denotes relative week, Y_{ik} is the variable of interest of the individual (consumer or influencer or item) i in a relative week k. K is the linear weekly time trend in each year, assuming the linear trend in the first year of the data is the same as that of the second year.

Figure 11: Timeline of Event Study-DID



5.4 Price Effects

It is common knowledge that competition lowers prices. Whether the superstars use the network effects to form a “natural monopoly" matters for how the platform should regulate them. Similarly, whether suddenly reduced competition puts upward pressure on prices is important to consumer welfare and the platform as the internal regulator. The effect of market structure change on prices can be decomposed into at least two mechanisms: product selection (composition effect) and product pricing (pricing effect). To inspect these effects, we regress the log of the listing price of items on the event dummy[20], indicating whether the price is listed pre or post-event, controlling the linear time trends. We also control for category-fixed, influencer, and year-fixed effects. In the regression shown in column (2), we control for item fixed effect, while in column (1), we did not. Note, the items listed and sold by the Queen is not included in the regression because we are investigating how others would react to the event, so the difference here would not be driven by the decisions of the Queen before her being banned, similar to the other subsections.

Table 5: Price Effects

	(1)	(2)	(3)	(4)
Log(Prices)	listing		transacted	
After*Year	18.0***	2.97***	0.802	4.17***
	(1.07)	(6.56)	(2.82)	(0.969)
After	-35.1***	-1.61***	-2.61	0.494
	(1.11)	(0.532)	(2.72)	(0.8112)
Year	-0.569	-5.08***	-26.1***	-1.92
	(0.716)	(0.735)	(4.81)	(1.83)
Time trend	0.986***	-0.0172***	0.327***	0.0254
	(0.0329)	(0.0196)	(0.00935)	(0.0332)
Item FE	N	Y	N	Y
Category FE	Y	Y	Y	Y
Influencer FE	Y	Y	Y	Y
Consumer FE	N	N	Y	Y
N	244995	104874	171194	167802

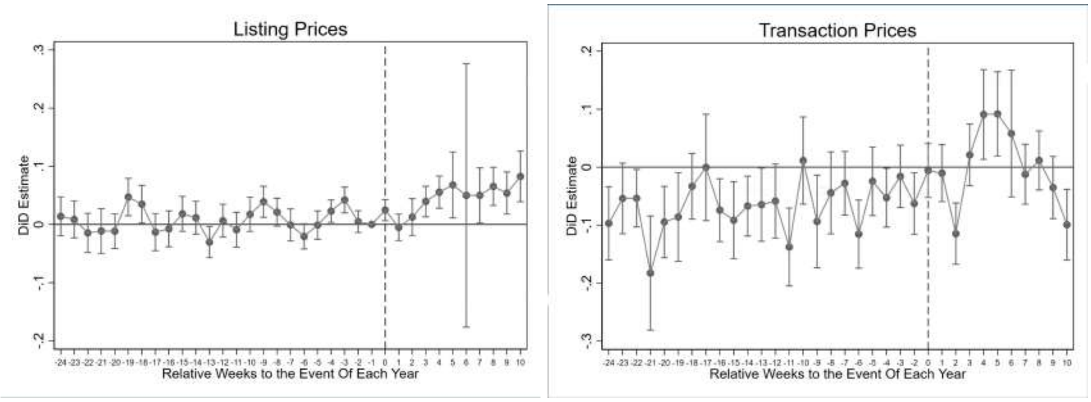
Standard errors in parentheses; all coefficients are multiplied by 100

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Sample from 1st July 2020 to 8th May 2022. The unit of observation of the first two columns is item-show-influencer (the same item could be priced differently in each show), and that of the last two columns is item-show-consumer. The number of observations in the first column is higher than that of the last two because not all item-show pairs are ever shopped by our sampled consumers. It is higher than the second column because only 34462 out of 174938 items are sold repeatedly. Standard errors clustered at the item level in (1) and (2) and at the individual consumer level in (3) and (4).

[20]As the distribution of prices is very skewed.

Figure 12: How the Relative Product Prices Change



The comparison shows that, after the ban, the prices of livestreamed items from other influencers are significantly higher by 18.0%, meaning that the other influencers are picking more expensive items within each category. Note the time trend is significantly positive but in a very small magnitude. This probably just shows the inflation level. This is a compositional change, but if we track the same items, the prices are significantly higher by about 2.97%, and the time trend is slightly negative. As for each item, gradually lowering prices is a common price discrimination strategy. In the 3rd and 4th columns, we look at equilibrium prices paid by the consumers. That is, we allow items with more sales to have more weights; either looking at the same items (by controlling item fixed effects) or the basket as a whole, the prices consumers paid increased significantly.

The overall evidence suggests that after the exogenous event, the overall item-wise pricing level increases, and other influencers are offering more expensive products. One would expect the bargaining power of other influencers to increase now, and they can bargain for a lower price so they can sell more and please their fans. But such motives are not necessary. Also, the rising prices and product selection are not decisions solely made by the influencers but the equilibrium results of influencers-brands contracting. The brands may offer fewer discounts if there is no strong endorsement and/or low price from the Queen. Since we have not modeled this bargaining yet, we could not determine much about which side drives this change in the bargaining equilibrium[21]. The following figure 12 shows how the “DID" coefficient changes over time in an Event-study style for the specification of columns (2) and (4) in Table 5. The story is consistent with the traditional competition lowers price prediction.

Currently, I am adding one linear weekly trend. I could add year-specific linear weekly trends or add influencer-specific linear trends for robustness.

[21] Do other influencers have more bargaining power with brands now? Can infer from the influencer commission fee data?

5.5 Product Variety Effects

Firms may compete in product space with or without price competition (e.g., Hotelling on a line, Salop on a circle, etc.). Influencers as retailing channels could also compete on the margin of variety by providing rich and distinct choices (Berry and Waldfogel, 2001). Traditional models such as Champsaur and Rochet (1989) predict that competition expands the product space. But would it also be possible that when there is less competition, influencers increase offerings (some may just be the items that would be endorsed by the Queen, were she still there) to capture the forfeited market share? To answer this, we regress the log of unique items (UPCs) listed and transacted per week on the event dummy, controlling for linear time trends and individual fixed effects. The results are shown in Table 6.

Table 6: “DID" Product Variety Effects

Log(#UPCs)	(1) listing	(2) transacted
After*Year	7.68	7.63***
	(6.39)	(1.43)
After	-16.34***	3.92***
	(4.83)	(1.24)
Year	-13.99	12.63***
	(10.38)	(2.54)
Time trend	0.544***	-0.138***
	(0.138)	(0.0494)
Influencer FE	Y	Y
Consumer FE	N	Y
N	1628	175701

Standard errors in parentheses; all coefficients are multiplied by 100.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Sample from 1st July 2020 to 8th May 2022. The unit of observations in the first column is item-influencer pairs, and that of the second column is item-consumer pairs. Standard errors clustered at the influencer level in (1) and at the individual consumer level in (2). The formal econometric model is in Appendix.

We find that it is not clear whether the remaining influencers are changing their product assortments, probably because they are just intermediaries between brands and consumers. We do find the different products consumers end up purchasing increased significantly, maybe because the consumers are searching for new matches (trying to buy new products from other influencers)[22].

[22] We can also check whether the composition of goods (by category) other influencers picked has changed.

5.6 Viewers and Viewing Time

The first column shows how much less traffic the platform has after the removal of the Queen, which is about 0.73%. In the second column, I find that consumers coming to the platform (this means without completing the consumer-day panel) are watching 0.128 fewer other influencers, which is about 0.5% (0.128/24). Since there are fewer consumers coming to the platform, if we run the balanced panel regression for column (2), the magnitude would be larger[23].

Table 7: Viewership

	(1) Total platform user	(2) Other influencers watched
After	-63.872***	-0.128***
	(39.31)	(0.0102)
Consumer FE	N	Y
N	311	287168

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Sample from 1st July 2020 to 8th May 2022. The observation unit in the first column is platform-day, and that in the second column is consumer-day. The standard errors clustered at the consumer level in (2). The formal econometric model is in Appendix.

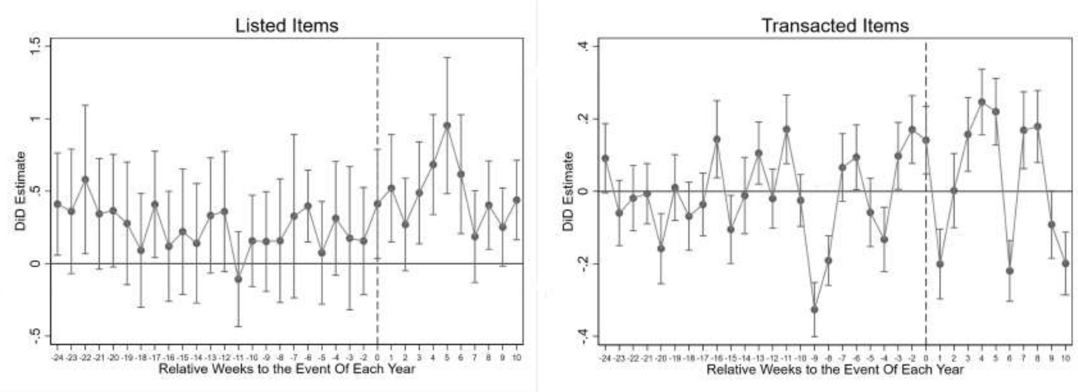
5.7 Sales and Revenue Effects

5.8 Effects on Consumer Welfare

This natural experiment can be leveraged to quantify how the removal of a dominant influencer affects prices, product variety, and purchases/consumption so that we need not take a stand on objects such

[23] There are other facts to check: for the consumers kept active on the platform, who do they watch instead? Extensively, influencers they do not watch previously are now watched. Intensively, influencers they watched previously now watch more. Also, where do the brands go?

Figure 13: How the Relative Product Variety Change



as the distribution of unobservables, the choice set facing consumers, or the strategic behavior of other influencers. We leave it to the model section to quantify any welfare effects.

5.9 Labor Supply (Shows and livestreaming time)

We can also observe on the supply side the reaction of competing influencers. We have seen how competing top influencers change their product assortment and pricing after the queen is banned. We can also see they put more effort (unobserved in most market structure studies) into the frequency, length of shows, and endorsed items. In Appendix Figure A4, we plot the weekly average livestreaming show lengths, number of shows, and total show time of the other top 24 influencers, and on the bottom-right panel, we plot the number of influencers joining the platform, where joining is defined by their first show.

Table 8: “DID" Regression on Labor Supply Trend

	(1)	(2)	(3)	(4)
	Show Length	#Shows	Total Time	Entrants
After*Year	0.460	0.564	4.35*	-16281***
	(0.407)	(0.467)	(2.47)	(2613)
After	-0.852***	-1.12***	-9.22***	-12165***
	(0.295)	(0.301)	(2.68)	(1552)
Year	0.031	0.535	2.72	-18719***
	(0.594)	(0.318)	(2.60)	(1552)
Time trend	0.0259***	0.0244***	0.237***	***
	(0.0058)	(0.0078)	(0.0498)	(2196)
Influencer FE	Y	Y	Y	N
N	6243	1628	1628	96

Standard errors in parentheses, clustered at influencer level

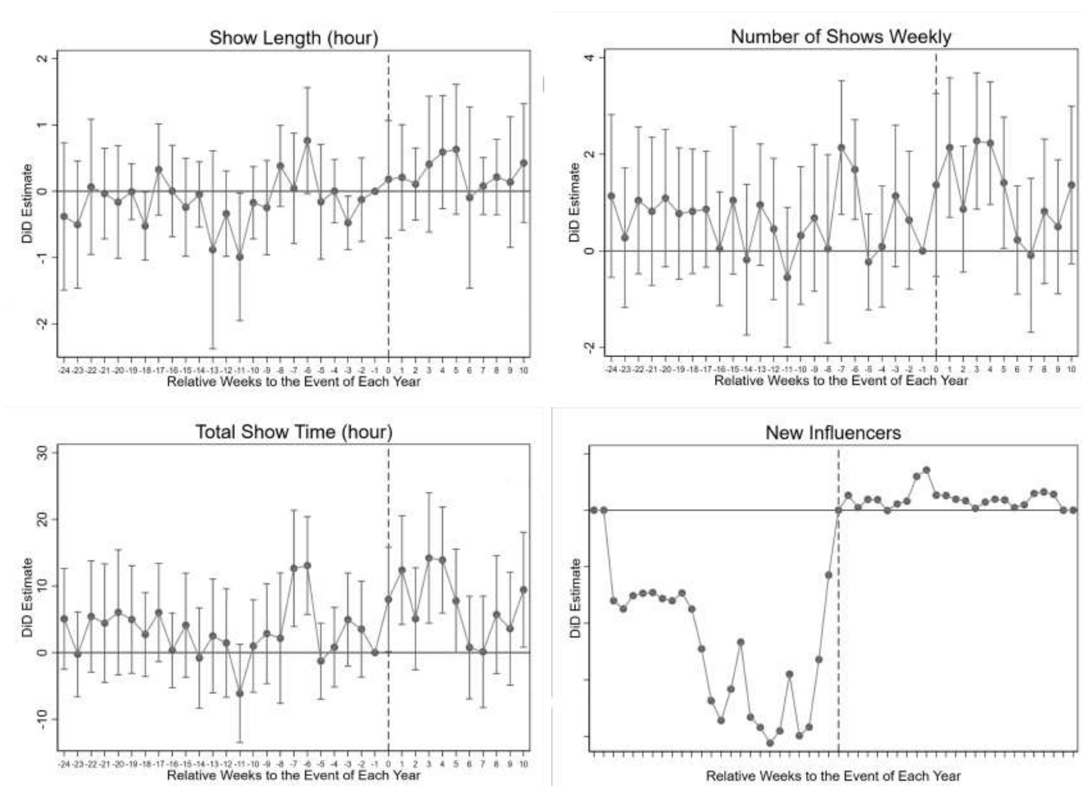
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Sample from 1st July 2020 to 8th May 2022. The unit of observations in columns (1) is show-influencer-week; that of (2) and (3) are influencer-week. We use a complete panel for (2) and (3). Column (4) is at the platform weekly level. The formal econometric models are in Appendix.

Combining the results in the first three columns in Table 8, we do not find clear evidence that the remaining influencers show more or less effort in livestreaming. This is probably because the influencers already worked at a very intense level (on average, more than four shows per week), so it is hard for them to livestream more frequently. In other words, the labor supply is inelastic with respect to the level of competition. The negatively significant coefficient in the second row shows the livestreaming business is less active at the end of the year and the beginning of the new year period (Chinese new year in early February). We can also check whether there are many entrants (extensive labor supply) to the market of influencers as evidence of entry blocking by the superstars.

After the removal of the top 1 influencer, as we can see from the bottom-right panel of Figure Appendix A4, there are no more entries of new influencers, probably because people are worried about the future of the platform, given that the Queen indeed has a positive spillover effect on the whole platform. But if we use the same months last year to control seasonality, the DID coefficient is positively significant. This means that compared to the previous year, there are more new influencers entering due to the event. The mixed evidence of negative first-difference before and after the event and positive DID coefficient suggests it is hard to say whether the superstar would block entry or the removal of them would incentivize innovation and competition. Again, banning the Queen does not leave her business to the others, she could be demand-creating herself, and both consumers and even her competitors could show some “herding effect" in that her status signals the health and potential of the platform or even the livestreaming industry. This possibility could be neglected by the platform, which may think more participation from other influencers, intensively (effort) or extensively (entry), would make the platform more diverse and energetic, hence the consumers would not leave. It turns out that even the

Figure 14: Relative Labor Supply Change



Notes: We could not calculate the standard error for the last panel.

most established, popular e-commerce platform would suffer, at least in the short run, when losing its top influencer.

So far, we are not aware of any celebrity/star influencers joining the platform after the ban[24].

5.10 Implication

Effects on prices and assortments suggest influencers indeed compete on the margins of prices, variety, and show time (labor supply). Brown (2011) documents the adverse effect of the presence of the superstar Tiger Woods on other competitors in golf tournaments. The platform may hope the other influencers could seize the opportunity when the top 1 superstar gives up a large share of the market. However, this does not happen.

This adverse effect on the platform's overall performance makes it clear that the platform should try a more conservative intervention, such as steering a fraction of consumers away from superstars to those smaller but high-quality influencers.

6 Model

6.0.1 Why we need a model

We need to model demand, as we want to know in the counterfactual, when the traffic, quality, or price levels of each influencer change, how demand responds. And we can estimate consumer welfare from the existence and associated change of every single influencer. The bargaining and contracting between the influencer and brands are complicated and may take another paper to fully understand. For now, we take the product offering by the influencers as given and assume they set prices to maximize profits under the Nash-Bertrand competition.

6.0.2 Timeline of Consumer Demand

1. Consumers choose which influencers to watch simultaneously [25].

[24] We can also show the new entrants growing up into major players. That is to say, the survival of the mature rate may be higher for new entrants.

[25] In a more general setting, consumers first form a consideration set through platform recommendation and direct search, which depends on the experience and awareness of influencers. We are not modeling the consideration set so far, assuming all consumers choosing from all influencers may underestimate the correlation and overestimate the substitutability among influencers.

- **Assumption 1**: For simplicity and estimation constraint, we restrict the choice set to be the top 6 influencers, assuming all consumers know all of them.
- **Assumption 2** Consumers have rational expectations of the influencer's product offering, prices, quality, and personal attractiveness at this stage.

2. Then, the consumer watches the live-streaming shows of these influencers and decides whether they want to purchase each item exposed/being recommended to them during the show.

Discussion: If we are interested in how large the effects of superstars on other influencers and how the market structure could be changed, focusing on the sizeable influencers make sense[26]. It is also practically more relevant, as it is also natural for the platform to try to help smaller but already sizeable influencers who could be productive, but not the marginal ones whose quality is more unsecured and easily replaceable if they left for another platform. There are two features in the data that support the assumptions I made above. First, we drop all visits less or equal to 10 seconds to get rid of noisy visits, which should not be assumed to be rational consumer behavior. Second, the top 6 are those ever visited by at least 30% of consumers, so it is reasonable to assume they are known to all consumers, and for the same reason, the other influencers are not famous enough. Instead, they serve specific consumer groups or are visible for shorter periods.

Also, we order the steps just for fixing ideas and estimation, as long as we assume the order of watching influencers does not matter, and information learned and action taken inside shows do not change the choice set of influencers, these two steps are actually simultaneously, in that consumers click into the show of one influencer, watch and making purchasing decisions, log out, and decide whether to watch the show of other influencers. We did not incorporate consumers' limited time as a constraint yet, because Figure 4 in Section 4 shows the total watching time is at a low level.

Now, we discuss the steps in more detail.

6.1 Step 1: Choose who to watch

From Figure 10 and Table 6, we find that when the Queen disappeared, the other influencers got fewer rather than more viewers. This pattern suggests influencers could be more than pure substitutes. After all, one reason for platforms to host superstars is to take advantage of their market expansion/positive spillover effect, especially when facing competing platforms. It is essential for the platform to know

[26] I have not found a trackable model where the interaction between the superstars and marginal influencers could be captured.

the driving forces for the potential mechanism underlying market expansion/stealing effect on others. If a smaller influencer gets a positive/negative spillover effect from the superstar, that is, consumers prefer (not) to watch them on the same day to watch the smaller influencer alone. It could be that the influencer and the superstar complement (substitute) each other. Or they share some traits valued by the consumers[27]. The two mechanisms that drive the sign of the estimated substitution patterns have different implications for the platform: without complementarity, suppressing the demand (by constraining the exposure or price discrimination and hence their sizes in the longer term) of the superstars will not have any adverse effect on the demand for the other influencers. The estimated substitution pattern would have a causal interpretation if driven by complementarity but is simply a correlation if it is driven by unobserved preferences. How should we empirically separate the true substitutability or complementarity of the influencers from the correlation in consumer preference? Following Gentzkow (2007), we introduce the correlation between tastes and complementarity/substitutability in this step.

Define the utility to consumer i of consuming a single influencer j on the day t to be

$$\bar{u}_{ijt} = \bar{u}_{jt} + \boldsymbol{x}_i \boldsymbol{\beta_j} + \nu_{ij} + \tau_{it}$$

Where $\bar{u}_{jt} = E_i u_{ijt}(T, d)$ is the expected utility from the next step (watching and purchasing)[28], $\boldsymbol{x}_i$ is a vector of observable consumer characteristics (age, gender, city level, income brackets), $\boldsymbol{\beta_j}$ is a vector of influencer-specific coefficients, the unobservable ν_{ij} captures the time-invariant, consumer idiosyncratic preferences over influencers (j-specific), it is assumed to be J-dimensional multivariate normal distributed. And τ_{it} is the daily demand shock of the consumer for any influencers[29] assumed to be i.i.d across consumers. It also captures seasonal demand and promotion. For simplicity, assume it is binary (high and low)[30]:

$$\tau_{it} = \begin{cases} 0 & \text{with probability} \quad 1-\gamma \\ \tau & \text{with probability} \quad \gamma \end{cases}$$

Consider the set/combination of influencers watched together and the utility of bundle r is defined

[27]Again, some traits may be controlled, but there are always unobservable consumer preferences (omitted variable).

[28]So here we are implicitly assuming the decision in that step 2 does not affect the choice set in this step 2. Otherwise, we have to take the parameters in step 2 into the likelihood function as well and estimate the parameters using SML together.

[29]Sales and promotion dates in the data could be tracked and controlled, but consumer individual shocks are unobservable to econometrician. This term is meaningful in that it captures some co-movement of demand, or bundling choice, not due to demand complementarity.

[30]We can relax this to be normally distributed to be more realistic.

as

$$u_{irt} = \sum_{j \in r} \bar{u}_{ijt} + \Gamma_r + \varepsilon_{irt} \quad \text{if } r > 0$$

where $j \in r$ means goods j is in bundle r. The interaction term Γ_r is the difference between the base utility of bundle r and the sum of all single utilities $\bar{u}_{ijt}$ the bundle consists. Assume that the interaction term for the n-good bundle is the sum of all possible two-good interaction terms in the bundle. ε_{itj} is i.i.d. type-1 extreme value error, with a scale parameter that measures how consumers trade off utils against price.

Denote the model parameters to estimate $\boldsymbol{\theta} = \{\tau, \gamma, \bar{u}_{jt}, \boldsymbol{\beta_j}\}$, integrate over ε_{itj} to get the probabilities of choosing r on each day.

$$Q_r\left(\boldsymbol{x}_i, \boldsymbol{\nu}_i, \tau_{it}; \boldsymbol{\theta}\right) = \frac{\exp\left[\sum_{j \in r}\left(\bar{u}_{jt} + \boldsymbol{x}_i \boldsymbol{\beta}_j + \nu_{ij} + \tau_{it}\right) + \Gamma_r\right]}{1 + \sum_{r=1}^{2^J} \exp\left[\sum_{j \in r}\left(\bar{u}_{jt} + \boldsymbol{x}_i \boldsymbol{\beta}_j + \nu_{ij} + \tau_{it}\right) + \Gamma_r\right]}.$$

Let $\tilde{q} \in \{0, 1, ..., 2^J\}$ be the vector of the chosen bundle on each day, q be the sequence of bundle chosen, then the probability of observing consumers i chooses q is:

$$P_{q_i}\left(\boldsymbol{x_i}, \boldsymbol{\nu_i}; \boldsymbol{\theta}\right) = \prod_{t=1}^{n}\left[\gamma Q_{\tilde{q}_t}\left(\boldsymbol{x}_i, \boldsymbol{\nu}_i, \hat{\tau}; \boldsymbol{\theta}\right) + (1-\gamma) Q_{\tilde{q}_t}\left(\boldsymbol{x}_i, \boldsymbol{\nu}_i, 0; \boldsymbol{\theta}\right)\right]$$

Discussion: where is the watching complementarity coming from? It could be interpreted as, or the mechanism could be:

- Products they bought from the superstar are complements to products from other influencers[31].
- Addiction to livestreaming as a shopping method as well as a form of entertainment.
- The positive spillover (of consumer traffic) to other influencers. This is not the extensive spillover from getting more consumers to the platform (which should be determined in the first step), but the visits to other influencers by scrolling up and down[32]. This is what Berry et al. (2014) referred to as "provider-driven complementarity". In many senses, this is the agglomeration effect in an online setting.
- There is a fixed search cost of launching the platform, so once a consumer opens the platform majorly because he/she wants to see the show of them, the marginal search cost of watching additional influencers should be really low.

[31]As the probit demand we are currently using does not capture purchasing substitutability itself, the effect of purchasing substitutability/complementarity is included in the watching complementarity.

[32]We have dropped the pass-through visit less than 10s to reduce noise.

- Algorithmic recommendation.

6.1.1 Identification

The first source of identification comes from the panel data itself. The larger the variance of the random effects ν_i, the more consistent the choices of influencers are: either consumed together or not consumed together. But the large magnitude of Γ would make the day-to-day variation strongly positively correlated: if some influencers are complements to a consumer, then the consumer might consume them together on one day and neither on another day but would be unlikely to consume either one alone. If, on the other hand, some influencers are substitutes for a consumer, they are unlikely to be consumed together. So the repeated observations (panel data) of consumers will pin down the covariance matrix of the ν_i random effects and separate it from the combined effect of Γ and τ. However, the panel data alone does not distinguish Γ and τ since the demand shock τ is not choice-specific.

The exclusion restriction helps disentangle the effects of Γ and τ by revealing the role of Γ in the following way: the moment $z_i\left[d_{ijt} - P_{ijt}(\theta)\right]$ where j is an influencer other than the Queen, z_i is the consumer attributes associated with the excluded variable. Where d_{ijt} is whether i chooses j in time t, and $P_{ijt}(\theta)$ is the model prediction for this choice. This moment is informative of Γ as z_i is excluded from the utility of another influencer j, and its effect on i choosing j must go through Γ. Intuitively, if the demand of other influencers moves in the same direction as the superstars (bundle is chosen) and is driven by complementarity, the exclusion would be effective, while if it is driven by a common demand shock realization, the exclusion will not make any difference. Specific to my example, if after the ban of the queen, the demand of other influencers dropped/increased, the substitution pattern we observed is driven by complementarity/substitutability at least partly (could not rule out the effects of consumer preference); if there is no change, the previous substitution pattern we observed is driven by taste/consumer preferences.

Since my exclusion restriction is only about the Queen, we are more confident in disentangling the effect of Γ_{Qj} from the τ associated, that is, the interaction between the queen and the others. While since we do not impose exclusion restrictions on other influencers, the identification of the substitutability among them would be weaker. In a future extension, we could also exploit another exogenous event where the King also disappeared from June to September 2022 and returned to the platform, but this period is not in my current sample, and focusing on the effect of the Queen on others is sufficient for now.

6.2 Step 2: watching and purchasing

In this step, we assume consumers have discrete-time simultaneous search for products. Consider a consumer i who has decided to watch the shows of influencer j. Then consumer i learn the attractiveness of influencer j, θ_j as well as the joint distribution of product prices p and qualities q she would like to recommend for category k.

The utility consumer i has from watching a livestreaming show s hosted by j is given by

$$u_{ij}(T_{ij}, dli) = \max_{T_{ij}, d_{l_i}} E_{ij} \left[\underbrace{\theta_{jt} T_{ij} + \beta T_{ij}^2}_{\text{Watching Utility}} + \underbrace{\phi_j \left\{ \sum_1^{l_i} d_{l_i} v_{ijl} - \lambda_1 \left(\sum_1^{l_i} d_l \right) - \lambda_2 \left(\sum_1^{l_i} d_l \right)^2 \right\}}_{\text{Purchasing Utility}} \right] \tag{6.2.1}$$

Where θ_{jt} is the attractiveness of j's show at time t, β measures the cost of time. λ_1 is a marginal utility from purchasing, and λ_2 captures decreasing marginal utility from purchasing[33]. That information is only realized in this step when consumers click into the show.

The longer a consumer watches the livestreaming events, the more items he is exposed to and hence may purchase, so we use ϕ to link the utility from watching to utility from purchasing. T does not enter the purchasing utility term directly, but indirectly via affecting d_l where d_{li} is whether i bought item l_i that was exposed to them[34]. We suppress the subscript s for now.

Inside equation 6.2.1, assume for each item l in category k introduced by an influencer j, we assume that the utility form purchasing good l is given by a probit model.

$$v_{ijl(k)} = q_{jk} - \alpha p_{jl(k)} + \varepsilon_{ijl(k)}$$

Where ε follows a normal distribution. Here q_{jk} is just the influencer-category FE, another interpretation would be the average quality of j's products in category k. The story here is that each influencer's product pool forms distributions in quality and price. Each time when j screens for a product, j picks product l of category k from his/her pool of this category and consumers seeing the price of product l in category k and decide whether to buy based on the expected quality of j's product in category k. The reason we estimate category level q_{jk} rather than item-level $q_{jl(k)}$ is that we do not see repeated observations for

[33] So we only care about the aggregate number of items purchased, not by category

[34] T implicitly determines how many items consumers watched, so dl_i, but it does not show up in the optimal condition of T, is this a problem for the model specification? This would underestimate θ_j and overestimate ϕ_j Could this capture the positive relationship of T and dl that we documented in **??**? I could check in data whether consumers quit after making a purchase.

all items with price variation. When we estimated the price coefficient, we used the subset of the sample where items have appeared with different prices. Averaging over all consumers, the estimates $\hat{\Theta}$ from T and d_{li} should rationalize the expected utility we estimated from the previous step when choosing influencers.

6.3 Limitations of the Model

1. Currently, we are assuming probit demand for each individual product. We may need nested logit where the nests are categories to capture the substitution between products.

2. Lack-of off-platform data[35], Since the within-platform distribution is correlated with the (cross-) platform competition, we also need to incorporate that consideration by collecting aggregate data on other competing platforms and making assumptions such as all consumers know and use the focal platform more or less regardless of which superstars are on it. Consumers are multi-homing, and we do not know whether they become more active on competing platforms, but it is reasonable to assume no absolute quit from the platform[36], which is still the “to-go" e-commerce platform in China without the Queen.

3. No consumer learning so far.

4. In the second stage, we did not explicitly model the fact that consumers who intend to purchase are more likely to watch longer.

7 Estimation

7.1 Step 1: Choose who to watch

The log-likelihood of observing the choices $\boldsymbol{q}$ of consumers given θ and $\boldsymbol{x_i}(\boldsymbol{X})$

$$L(\mathbf{X}, \mathbf{q}, \boldsymbol{\theta}) = \sum_i \ln \int_{\nu} P_{q_i}\left(\boldsymbol{x}_i, \boldsymbol{v}; \boldsymbol{\theta}\right) dF(\boldsymbol{\nu}; \boldsymbol{\theta})$$

where $F(\boldsymbol{\nu}; \boldsymbol{\theta})$ is the multivariate normal distribution of $\boldsymbol{\nu}$ conditional on $\boldsymbol{\theta}$. As it does not have a closed-form solution, the estimation method is simulated maximum likelihood (SML), where the simulation

[35]Tiktok total users available.

[36]In Section 5, we can see that even the fans of the queen are still watching the shows.

draws $\boldsymbol{v}$ from its multivariate normal distribution to form consistent estimates of the probability. The SML estimator is

$$\hat{\boldsymbol{\theta}}^{SML} = \arg\max_{\boldsymbol{\theta}} \left\{ \sum_i \ln \left[\frac{1}{S} \sum_s P_{q_i} \left(\boldsymbol{x}_i, \boldsymbol{\nu_{is}}; \boldsymbol{\theta} \right) \right] \right\}$$

Standard errors should be calculated using the robust asymptotic approximation proposed by McFadden and Train (2000). But currently, they are calculated as the inverse of the outer product of gradients now as Hessian matrix calculation is slow. The robust covariance matrix estimation is calculated using the average of the outer product of the gradients, which is a consistent estimator of the information matrix.

Table 9: Parameter estimates from the full model: observable characteristics

	E	King	K	L	C	Queen
Age	-0.2633	0.1964	2.0319**	1.7660***	-1.0053***	0.0835
	(0.4107)	(0.2495)	(0.8226)	(0.5064)	(0.2962)	(0.2528)
Female	0.3261**	0.2218***	0.1814*	0.0958	0.2507**	0.0328
	(0.1157)	(0.0507)	(0.1283)	(0.1066)	(0.0930)	(0.0534)
City Level	0.0082	-0.0581*	-0.1105	0.1140*	-0.1132*	0.0264
	(0.0824)	(0.0374)	(0.0931)	(0.0715)	(0.0766)	(0.0419)
Income Bracket	-0.1689*	-0.1052*	-0.0027	-0.7818**	-0.0800	-0.6584**
	(0.0956)	(0.0685)	(0.1133)	(0.2922)	(0.1079)	(0.2686)
_cons	-1.9522	0.2752	-5.7104*	-5.4856*	0.1584	1.6869
	(2.9886)	(2.1502)	(2.7071)	(2.7587)	(2.2742)	(2.1426)
N	15000	15000	15000	15000	15000	15000

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

$N = 1500$ because I sampled 3000 consumers out of 19625, each for 5 days, for the purpose of estimation time.
So the unit of observation is person-day. I can re-run with the larger sample if time permits.

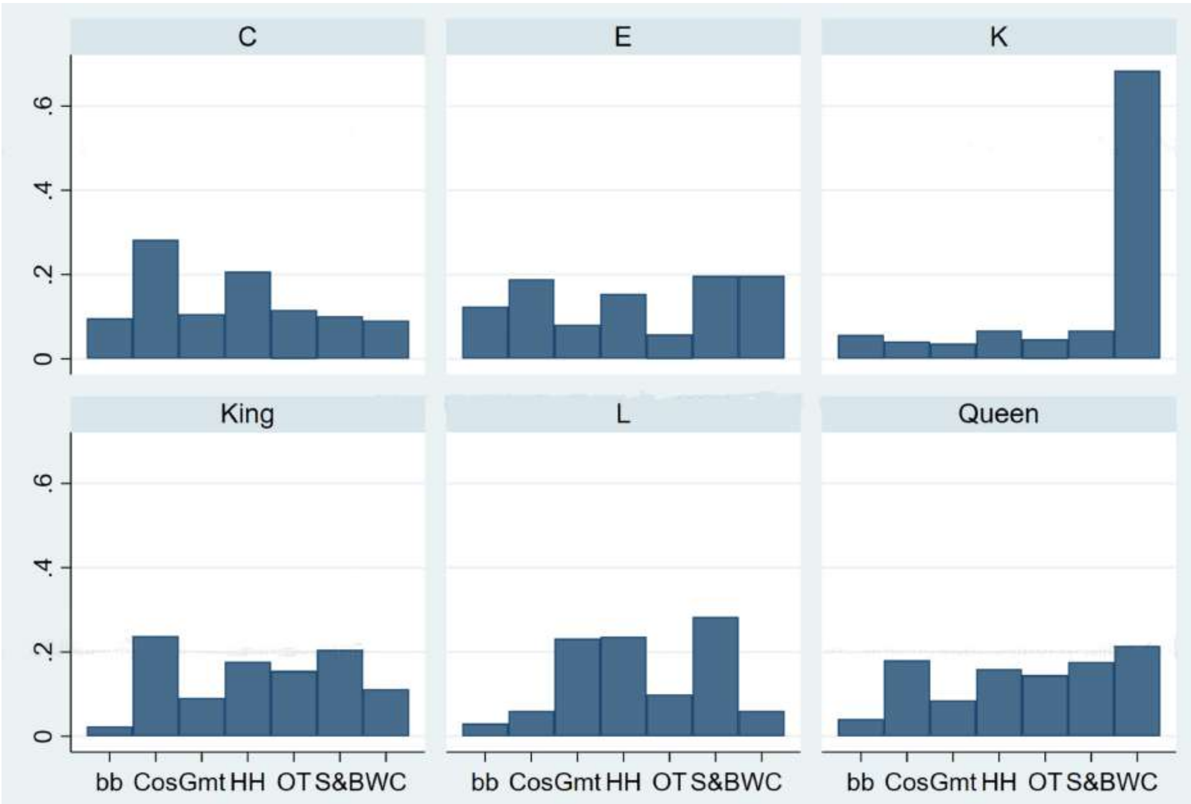
Table 9 shows the SML estimates of the coefficients on observable characteristics. Combing with the influencers' characteristics defined by items the advocates, shown in Figure 15 below, we have the following observations:

1. Some influencers are more attractive to elder consumers, while some are more attractive to young consumers. But age does not have much effect on the superstars, suggesting they are popular to all age groups.

2. Females are more likely to watch shows, especially from those influencers who specialize in women's products (baby products, cosmetics, women's clothing), as we can see from Figure **?**. In fact, the Queen covers all-category well, and there is less reason for female consumers to be addicted to her than male consumers.

3. Income effects are negative, people's time is valued more, and support the hypothesis that livestreaming influencers have the role of discount retail channel, where the more price-sensitive consumers are more likely to adopt.

Figure 15: Share of products in major categories



Notes: bb=baby products; Cos=Cosmetics; Gmt=Gourmet; HH=household; OT=others; SB= snacks and beverages; WC=Women's Clothing. Ratios add up to 100% here.

Table 10: Parameter estimates: interaction terms

	King	K	L	C	Queen
E	-0.1817	-0.0744	2.7352	0.9929	1.6332
	(0.0254)	(0.0104)	(0.3830)	(0.1390)	(0.2287)
King		0.0953	-1.4121	0.2541	2.2402
		(0.0133)	(0.1977)	(0.0356)	(0.3137)
K			2.0547	1.0822	-1.4893
			(0.2877)	(0.1515)	(0.2085)
L				-0.6981	3.2671
				(0.0978)	(0.4575)
C					0.3931
					(0.0550)

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 10 shows the value of the interactions Γ. Compare with Figure 6 and Figure 5 in section 4. Table 11 shows the superstars indeed have the highest mean (expected) utilities from consumers. Table 12 is the sample covariance of the estimated utility from the observables. Positive/negative utility components in Table 12 means that from the observables, viewerships positively/negatively correlated between the two influencers. Or, this pair of influencers should have positively/negatively correlated consumer groups. Table 13 is the estimated covariance matrix of the unobservables ν_i. The variance of the unobservable component for the Queen utility is quite large, suggesting the consumption is more consistent over days than the small probabilities from the model without unobservables would allow. A negative unobserved correlation means consumers' tastes of the two influencers are quite divided. Large diagonal values mean consumers' tastes toward that influencer are divided. Some favor him/her a lot, while some strongly dislike him/her. The variance of ν_i is large relative to the observables.

Table 11: Mean Utilities

	L	King	K	L	C	Queen
$\bar{u_{jt}}$	-3.0908	-0.9670	-3.5892	-5.0929	-2.4013	0.0500

Table 12: Covariance of observable utility

	E	King	K	L	C	Queen
E	0.2175	0.0694	-0.1092	-0.3501	0.1119	-0.2173
King		0.1747	0.0500	0.3726	0.1176	0.2996
K			0.4703	0.4160	-0.1793	0.0795
L				2.4430	-0.0341	1.6855
C					0.2268	0.1391
Queen						1.3216

Table 13: Covariance of ν unobservables

	E	King	K	L	C	Queen
E	0.3536	-0.1552	0.6081	-4.6231	-0.5208	-3.1104
King		0.0681	-0.2669	2.0293	0.2286	1.3653
K			1.0459	-7.9511	-0.8957	-5.3495
L				60.4464	6.8095	40.6681
C					0.7671	4.5814
Queen						27.3613

Table 14: Non-linear Parameters

τ	γ	Var(τ)
3.1215*	0.5000	2.308
(1.9353)	(1.6403)	

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 14 shows the non-linear parameters, which suggests half of the time, demand is overall higher by 3.12% than the other time. Variance due to τ_i unobservables is small.

7.2 Step 2: watch and purchase

The parameters to estimate in 6.2.1 are $\Theta = \{\theta_{jt}, \beta, \phi_j, \lambda_1, \lambda_2, q_{jk}\}$. Since we have individual-level data, this can be efficiently estimated using MLE. For now, we are only going to estimate q_{jk} using the purchase decision d_{li}. Before that, first estimate the price coefficient.

7.2.1 Price Coefficient

We leverage the price variations in consumer panel data to identify the consumer's price sensitivity. We run the following equation

$$Y_{ijl(k)} = \beta + \alpha p_{jl(k)} + FE_{l(k)} + FE_j + FE_i$$

Where $Y = 0, 1$. For now, although we do know the purchased quantity, restrict the demand to be bought or not for simplicity. Controlling influencer, consumer, and item FE, fix the same q.

And we estimate a point estimate $\alpha = 0.0354$ without considering consumer heterogeneity. One may worry about the price endogeneity, we could use the prices in the brands' own store (usually the MSRP) as the price instrument, but it also suffers from common demand shock as well, but this is not the main point of the paper now.

With the above price coefficient, we estimate the q_{jk} for the top 6 influencers and major 6 categories.

Price and Quality

Overall equilibrium prices and quality are positively correlated, meaning that the higher an influencer j's product pricing in category k, the higher quality j's product in k is. As this quality term is just a conditional conversion rate. This result agrees with the reduced-form, assortative matching results in Section 4. We can see the top 2 has the highest q_{jk} in women's clothing, household products, baby products, and gourmet. The top 2 are not dominating the snacks and beverages category, but everyone is close to each other, suggesting the product picking and persuasiveness for this category is hard to be differentiated, as it is overall the cheapest category, with a high overall conversion rate. And the King has the second highest quality in Cosmetics, which supports his nickname "the lipstick king".

Table 15: Women's Clothing

names	obs	p	q
King	3001	5.282	1.870
		(.865)	(.398)
L	1341	4.483	1.587
		(1.000)	(0.505)
E	12358	4.784	1.694
		(.836)	(.319)
Queen	2557	5.158	1.826
		(.889)	(.300)
K	25755	4.610	1.632
		(.779)	(.288)
C	8384	4.867	1.723
		(.915)	(.318)

Table 16: Snacks and beverages

names	obs	p	q
King	3610	3.657	1.295
		(.772)	(.275)
L	5457	3.750	1.328
		(.910)	(0.318)
E	2987	3.723	1.319
		(.918)	(.313)
Queen	2919	3.576	1.266
		(.836)	(.299)
K	1566	3.683	1.266
		(.877)	(.296)
C	1716	3.748	1.327
		(1.037)	(.396)

Table 17: Household

names	obs	p	q
King	2310	4.420	1.565
		(.945)	(.402)
L	3113	3.719	1.317
		(1.027)	(.356)
E	2393	3.773	1.336
		(.956)	(.343)
Queen	1782	4.231	1.498
		(.929)	(.312)
K	1840	3.810	1.349
		(.992)	(.349)
C	1629	3.798	1.345
		(1.011)	(.339)

Table 18: Baby Products

names	obs	p	q
King	828	4.851	1.717
		(.875)	(.297)
L	964	4.009	1.419
		(.813)	(.409)
E	3210	4.208	1.490
		(.720)	(.270)
Queen	693	4.528	1.603
		(.944)	(.316)
K	2003	4.185	1.482
		(.615)	(.258)
C	3113	4.211	1.491
		(.843)	(.293)

Table 19: Cosmetics

names	obs	p	q
King	5991	5.077	1.798
		(1.085)	(.380)
L	1713	4.680	1.657
		(1.312)	(.575)
E	2839	4.806	1.702
		(1.144)	(.660)
Queen	2879	5.173	1.032
		(1.122)	(.378)
K	1223	4.921	1.742
		(1.043)	(.365)
C	3379	5.144	1.821
		(1.115)	(.394)

Table 20: Gourmet

names	obs	p	q
King	285	6.177	2.187
		(.854)	(.306)
L	92	5.296	1.875
		(1.031)	(2.357)
E	330	5.377	1.904
		(.794)	(1.870)
Queen	220	6.213	2.200
		(.989)	(.334)
K	489	5.116	1.812
		(.586)	(.747)
C	133	5.716	2.024
		(1.224)	(.450)

8 Counterfactuals

Could the platform make the influencer market more competitive by exploiting consumer search friction (consideration set) and making it easier for consumers to find smaller influencers? For instance, by steering consumers away from superstar influencers to smaller ones (tweak the demand side). Or it may subsidize the smaller influencers so that they can recommend products with lower prices and higher quality. Also, the platform may subsidize manufacturers to encourage them to cooperate with smaller influencers (tweak the supply side). Since the natural experiment reveals what would happen when a superstar is gone, we do not need to experiment with this extreme counterfactual. A more realistic and modest counterfactual exercise is, how many consumers should the platform steer (taking the superstars out of the consumers' choice sets) away from the superstar if they were still there? The complementarity complicates things, as consumers who did not watch the superstars also tend to watch others less. Hence, the platform should target which group of consumers they want to steer smartly.

Notably, there would be secondary effects when the platform intervenes in consumers' consideration sets. The platform's recommendation algorithm would also affect influencers' pricing and how many and which firms enter.

9 Conclusion

To conclude, we have the following findings. First, pricing and product variety matter significantly in the competition of influencers. Second, I develop a structural demand model to empirically investigate the tradeoff between the market expansion effect and the cannibalization effect from the superstar onto the other influencers. The preliminary results from the consumers' bundling choice model suggest that influencers have non-negligible interactions. I find superstar influencers have overall net positive spillovers to some competitors while net adverse effects on others, boosting consumer welfare considerably. Furthermore, equilibrium prices and quality are positively correlated. Third, losing a superstar has a long-term impact on the platform, which could not recover the demand created by the superstar, even when the other influencers supply more livestreaming time. Given the potential complementarity between influencers, the platform should be more careful in steering consumers away from superstar influencers. Next, we will do a counterfactual exercise where the platform drives some consumers.

References

W. Amaldoss, J. Du, and W. Shin. Media platforms' content provision strategies and sources of profits. *Marketing Science*, 2021.

M. Ammann, P. Horsch, and D. Oesch. Competing with superstars. *Management Science*, 62(10):2842–2858, 2016.

K. Bagwell. The economic analysis of advertising. *Handbook of industrial organization*, 3:1701–1844, 2007.

J. Bai, M. Chen, J. Liu, X. Mu, and D. Y. Xu. Search and information frictions on global e-commerce platforms: Evidence from aliexpress. Technical report, National Bureau of Economic Research, 2020.

M. A. Barach, J. M. Golden, and J. J. Horton. Steering in online markets: the role of platform incentives and credibility. *Management Science*, 66(9):4047–4070, 2020.

S. Berry, A. Khwaja, V. Kumar, A. Musalem, K. C. Wilbur, G. Allenby, B. Anand, P. Chintagunta, W. M. Hanemann, P. Jeziorski, et al. Structural models of complementary choices. *Marketing Letters*, 25(3): 245–256, 2014.

S. T. Berry and J. Waldfogel. Do mergers increase product variety? evidence from radio broadcasting. *The Quarterly Journal of Economics*, 116(3):1009–1025, 2001. ISSN 00335533, 15314650. URL `http://www.jstor.org/stable/2696424`.

H. K. Bhargava. The creator economy: Managing ecosystem supply, revenue sharing, and platform design. *Management Science*, 2021.

J. L. Binken and S. Stremersch. The effect of superstar software on hardware sales in system markets. *Journal of Marketing*, 73(2):88–104, 2009.

T. F. Bresnahan and P. C. Reiss. Entry and competition in concentrated markets. *Journal of political economy*, 99(5):977–1009, 1991.

J. Brown. Quitters never win: The (adverse) incentive effects of competing with superstars. *Journal of Political Economy*, 119(5):982–1013, 2011.

S. Cao and C. Chatterjee. Equilibrium effects of pharmaceutical bundling: Evidence from india. *Available at SSRN*, 2022.

P. Champsaur and J.-C. Rochet. Multiproduct duopolists. *Econometrica: Journal of the Econometric Society*, pages 533–557, 1989.

L. W. Cong and S. Li. A model of influencer economy. *Available at SSRN 3974661*, 2021.

K. S. Corts and M. Lederman. Software exclusivity and the scope of indirect network effects in the us home video game market. *international Journal of industrial Organization*, 27(2):121–136, 2009.

J.-P. Dubé, G. J. Hitsch, and P. E. Rossi. Do switching costs make markets less competitive? *Journal of Marketing research*, 46(4):435–445, 2009.

I. P. Fainmesser and A. Galeotti. The market for online influence. *American Economic Journal: Microeconomics*, 13(4):332–72, 2021.

J. Förderer and D. Gutt. The effects of platform superstars on content production: Evidence from ninja. 2021.

M. Gentzkow. Valuing new goods in a model with complementarity: Online newspapers. *American Economic Review*, 97(3):713–744, June 2007. doi: 10.1257/aer.97.3.713. URL `https://www.aeaweb.org/articles?id=10.1257/aer.97.3.713`.

E. Honka, A. Hortaçsu, and M. A. Vitorino. Advertising, consumer awareness, and choice: Evidence from the us banking industry. *The RAND Journal of Economics*, 48(3):611–646, 2017.

G. Illanes and S. Moshary. Market structure and product assortment: Evidence from a natural experiment in liquor licensure. Working Paper 27016, National Bureau of Economic Research, April 2020. URL `http://www.nber.org/papers/w27016`.

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

F. Koenig. Technical change and superstar effects: Evidence from the rollout of television. 2021.

A. B. Krueger. The economics of real superstars: The market for rock concerts in the material world. *Journal of Labor Economics*, 23(1):1–30, 2005.

K. H. Lee and L. Musolff. Entry into two-sided markets shaped by platform-guided search. Technical report, mimeo, Princeton University, 2021.

D. McFadden and K. Train. Mixed mnl models for discrete response. *Journal of Applied Econometrics*, 15 (5):447–470, 2000. doi: https://doi.org/10.1002/1099-1255(200009/10)15:5<447::AID-JAE570>3.0.CO;2-1. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/1099-1255%28200009/10%2915%3A5%3C447%3A%3AAID-JAE570%3E3.0.CO%3B2-1.

B. McManus, A. Nevo, Z. Nolan, and J. W. Williams. Steering incentives of platforms: Evidence from the telecommunications industry. Technical report, National Bureau of Economic Research, 2020.

G. Parker and M. Van Alstyne. Innovation, openness, and platform control. *Management Science*, 64(7): 3015–3032, 2018.

A. Pei and D. Mayzlin. Influencing the influencers. *Available at SSRN 3376904*, 2021.

A. Prat and T. Valletti. Attention oligopoly. *American Economic Journal: Microeconomics*, 2021. URL http://hdl.handle.net/10044/1/90318.

J.-C. Rochet and J. Tirole. Platform competition in two-sided markets. *Journal of the European Economic Association*, 1(4):990–1029, 2003. ISSN 15424766, 15424774. URL http://www.jstor.org/stable/40005175.

S. Rosen. The economics of superstars. *The American economic review*, 71(5):845–858, 1981.

L. A. Stole. Price discrimination and competition. *Handbook of industrial organization*, 3:2221–2299, 2007.

Thomassen, H. Smith, S. Seiler, and P. Schiraldi. Multi-category competition and market power: A model of supermarket pricing. *American Economic Review*, 107(8):2308–51, August 2017. doi: 10.1257/aer.20160055. URL https://www.aeaweb.org/articles?id=10.1257/aer.20160055.

J. Tudón. Prioritization vs. congestion on platforms: evidence from amazon's twitch. tv. *The RAND Journal of Economics*, 2022.

Appendix

Figure A1: DID using last year as control to control for seasonality

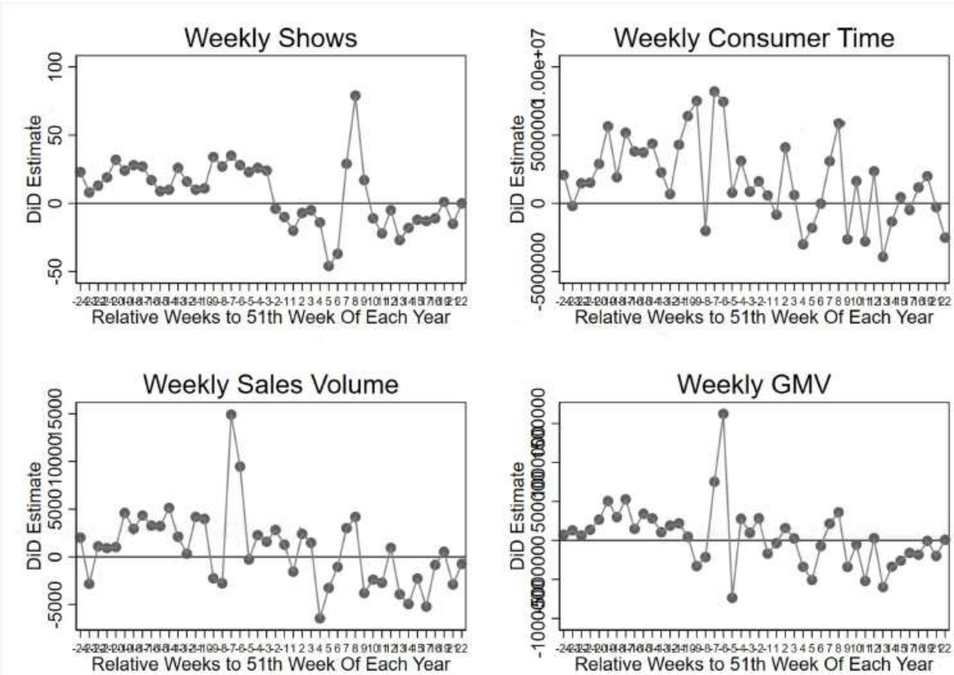


Figure A2: DID using last year as control to control for seasonality: Log values

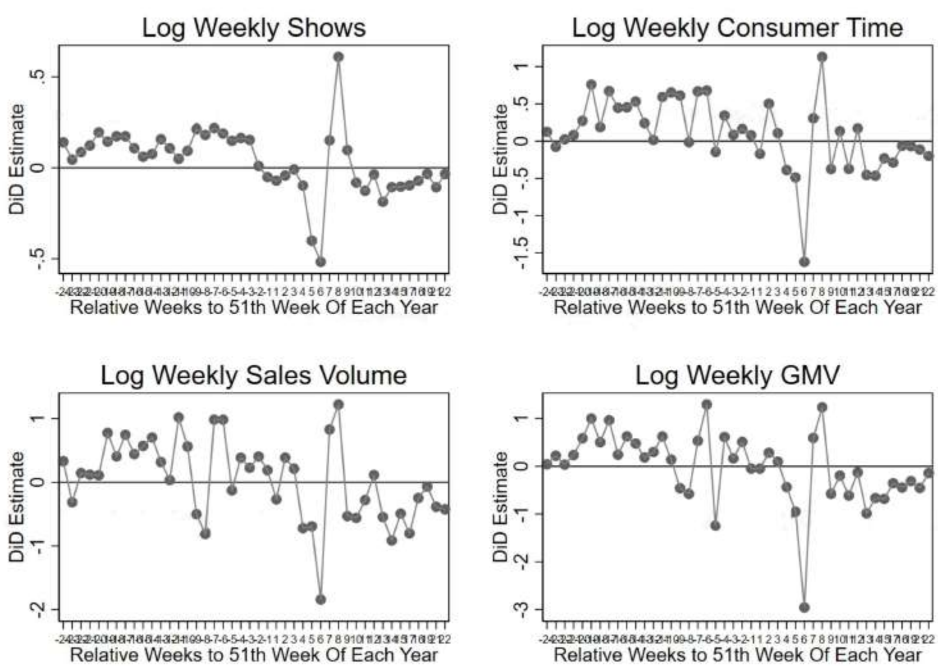
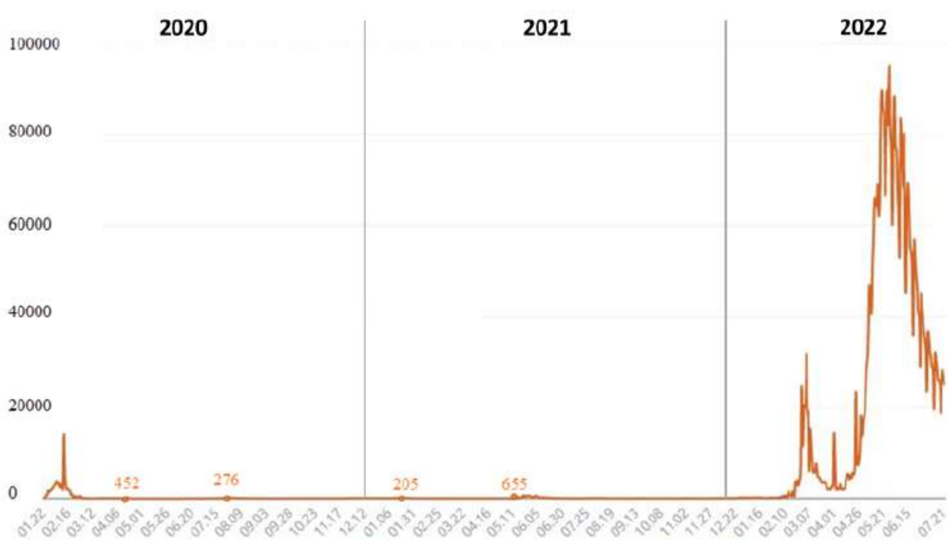


Table 21: Prices

	Cosmetics	Women's Clothing	Household	Baby products	Gourmet	Snacks & beverages
King	343.43	254.93	132.13	192.26	90.85	48.63
	(577.63)	(509.01)	(296.66)	(405.14)	(192.73)	(45.88)
L	301.09	158.63	74.71	83.46	56.03	36
	(663.23)	(308.11)	(137.11)	(124.93)	(176.09)	(84.79)
E	268.12	191.87	71.34	85.96	33.98	61.45
	(569.91)	(379.87)	(124.19)	(83.55)	(68.68)	(86.2)
Queen	354.66	251.87	99.66	124.03	79.23	46.98
	(588.81)	(562.71)	(182.17)	(242.46)	(137.96)	(58.5)
K	252.64	144.22	78.52	77.69	21.64	55.05
	(447.19)	(195.59)	(128.55)	(69.87)	(29.43)	(70.68)
C	334.25	239.26	77.28	99.78	106.97	65.42
	(533.32)	(536.27)	(143.97)	(181.51)	(177.9)	(3.63)

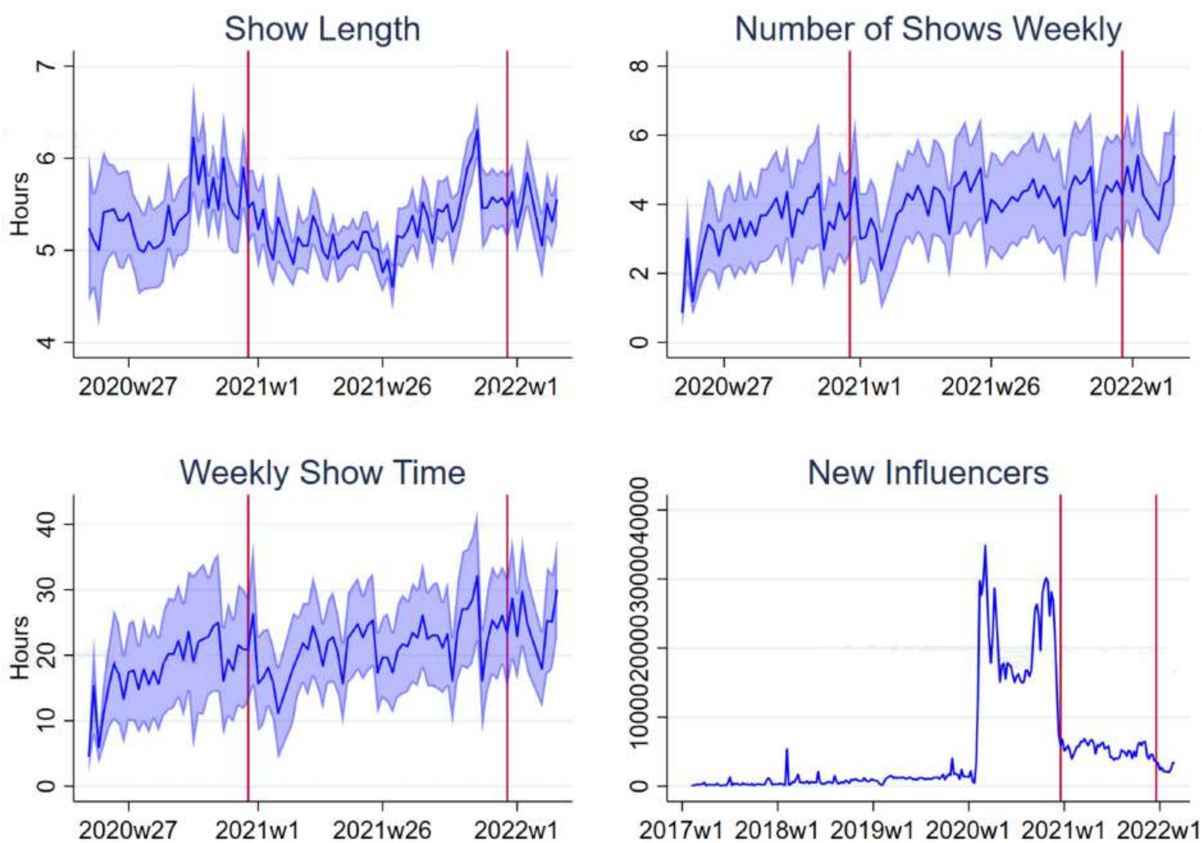
Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Figure A3: Daily new Covid cases in China



Source: `https://covid19.who.int/region/wpro/country/cn`

Figure A4: Labor Supply Trend



Notes: Sample from 1st July 2020 to 8th May 2022. The first column is at the show level, the unit of observations of the second column is the number of shows for each influencer-week, that of the third column is the total livestreaming time (hours) for each influencer-week, and that of the last column is platform-week.

Table A1: Regression Coefficients on the net Spillover Effect

Top 2 combined \ The other influencers	Viewers	Log(Viewers)	GMV	Log(GMV)
Viewers	0.508***			
	(0.0510)			
Log(Viewers)		0.626 ***		
		(0.0816)		
GMV			0.203***	
			(0.447)	
Log(GMV)				0.324***
				(0.0666)

Standard errors in parentheses.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Sample from 1st July 2020 to 8th May 2022, so number of observations are 102 (weeks).

Table A2: Fans and non-fans

	Items	Spending	Time (min)	Shows
Fans	0.469 ***	39.75***	7.427***	0.643***
	(0.0258)	(2.787)	(0.476)	(0.0438)
Fans*Ban	-0.296***	-27.19 ***	-5.467***	-0.233***
	(0.0251)	(2.905)	(0.498)	(0.275)
_cons	0.150***	15.76***	3.628***	0.749***

A2: Platform Cultivation Decision

A Model of Relation between Platforms and Influencers

December 10, 2022

1 Baseline Model

Time is discrete and consists only three periods $t = 1, 2, 3$. There is a platform from which consumers can purchase products. In the first period, consumers can only purchase from the platform, or from the outside option. The utility consumers have from purchasing on the platform is given by

$$u_i^P = v^P + \varepsilon_i^P$$

and the utility from the outside option is given by

$$u_i^O = \varepsilon_i^O$$

where v^P is the mean utility from the platform, and the idiosyncratic shock $\varepsilon_i = \{\varepsilon_i^P, \varepsilon_i^O\}$ are iid type I extreme values. Therefore in the first period, the market share of the platform is

$$s_1^P = \frac{\exp\{v^P\}}{1 + \exp\{v^P\}}$$

We assume that product prices are $p = 1$, and the cost of producing is 0, therefore, the profits of the platform is given by $\Pi_1^P = s_1^P M$ where M is the population size. We normalize $M = 1$.

In the first period, the platform can pay a cost c to cultivate a large influencer from where consumers can purchase goods too starting from the next period. We assume that consumers pay same price p when purchasing from an influencer as from the platform, and the utility consumers have from the influencer is given by

$$u_i^I = v^I + \varepsilon_i^I$$

where $\varepsilon_i = \{\varepsilon_i^P, \varepsilon_i^I, \varepsilon_i^O\}$ are independent type I extreme value random variables. Therefore, the market market shares of the platform and the influencer are

$$s_2^P = \frac{\exp\{v^P\}}{1 + \exp\{v^P\} + \exp\{v^I\}}$$

and

$$s_2^I = \frac{\exp\{v^I\}}{1 + \exp\{v^P\} + \exp\{v^I\}}$$

Further assume that the platform will charge τ per sales from the influencer, therefore, the total profits of the platform in the second period is given by

$$\Pi_2^P = s_2^P + \tau s_2^I$$

Notice that when $\tau = 0$, an influencer is a substitute of the platform and therefore, it only has business stealing effects becaue $s_2^P < s_2^P$. When $\tau = 1$, the platform can enjoy all the benefits from the influencer and it generates a business expansion effects because $s_2^P + s_2^P > s_1^P$.

In the third period, there is a probability κ that an influencer can become independent from the platform and the platform can no longer benefit from the influencer's sales. Otherwise, the economy is the same as in the second period. Therefore, the expected profits of the platform is given by

$$\Pi_3^P = \kappa s_2^P + (1 - \kappa)(s_2^P + \tau s_2^I)$$

Therefore, even if $\tau = 1$, the platform cannot fully internalize the benefits from cultivating an influencer because of the possibility of being independent.

Therefore, in the first-period, the platform chooses whether to generate an influencer by solving maximizing its expected profits in the three periods

$$\max\{W^{P,I}, W^{P,nI}\}$$

where assuming no discount,

$$W^{P,I} = \frac{\exp\{v^P\}}{1 + \exp\{v^P\}} - c + \frac{\exp\{v^P\} + \tau \exp\{v^I\}}{1 + \exp\{v^P\} + \exp\{v^I\}} + \frac{\exp\{v^P\} + (1-\kappa)\tau \exp\{v^I\}}{1 + \exp\{v^P\} + \exp\{v^I\}}$$

$$W^{P,nI} = \frac{3\exp\{v^P\}}{1 + \exp\{v^P\}}$$

Therefore, the decision depends on $\{v^P, v^I, c, \tau, \kappa\}$.

2 Extension

- Endogenize the commission fee τ and the independence probabity κ through a bargaining model.
- We could allow for v^P and v^I changing depending on whether an influencer is associated with the platform or not.
- Allowing for v^I depending on p and allowing for a lower p from the influencer is straightforward.

A3: Platform Exposure Decision

Live-Streaming E-commerce

December 29, 2022

1 Consumer Decisions

See previous notes. Recall that there is a mass of M consumers, each can choose to purchase from either the platform T, or purchase from an influencer $i \in \{1, 2, \cdots, N\}$. We assume that the mean utilities a consumer can derive when purchasing from the platform, and from each influencer is given by $v = (v^T, v_1^I, \cdots, v_N^I)$. We assume that the indirect utility of a consumer choosing influencer i or the platform T is given by

$$u_{ji} = v_i^I + \varepsilon_{ji}, i \in \{1, 2, \cdots, N\}$$

$$u_{jT} = v^T + \varepsilon_{jT}$$

and there is an outside option whose mean utility is normalized as 0. Therefore, without the platform modifying exposure, the choice probability of an influencer i is given by

$$P(i) = \frac{\exp\{v_i^I\}}{\sum_i \exp\{v_i^I\} + \exp\{v_i^T\} + 1}$$

2 Contract between the Platform and Influencers

Consider two periods. In the first period, the platform makes a take-it-or-leave-it (TIOLI) offer to each influencer $i \in \mathcal{N} = \{1, 2, \cdots, N\}$, and each influencer decides to accept or reject independently. In the second period, consumers start to purchase either from the platform or from the influencers. The platform and the influencers negotiate a two-part tariff $(\tau_i, s_i + f(x_i; v_i))$ where τ_i is a per-sales charge of the platform from the influencer, and s_i is a lump-sum transfer from the influencer to the platform. On top of the commission fees, the platform also negotiate with each influencer the exposure intensity $x_i \in [0, 1]$ with a charge $f(x_i; v_i)$ depending on not only the exposure intensity, but also the attractiveness of an influencer v_i. We assume that the advertisement fee has a form

$$f(x_i; v_i) = \frac{\theta_1 x_i}{v_i^I}$$

We assume that if a influencer rejects the offer, its attractiveness for consumers will degrade by $\delta \in (0,1)$. We assume that exposure changes consumer's choice probability in a particular way. In particular, denote the set of influencers who accept offers as $\mathcal{D}$, and who reject offers as $\mathcal{R}$. Therefore, the consumer's choice probability of an influencer i is given by

$$P(i|x,v) = \frac{(1+\theta_2 x_i)\exp\{v_i^I\}}{\sum_{j\in\mathcal{D}}(1+\theta_2 x_j)\exp\{v_j^I\} + \sum_{j\in\mathcal{S}}\exp\{\delta v_j^I\} + \exp\{v^T\} + 1}$$

In equilibrium, without loss of generality, we can focus on equilibrium where all influencers accept the offer. We assume influencers hold passive beliefs (so that when i rejects the offer, i believes that other influencers will not change their accept/reject decisions), and therefore, the platform solves

$$\max_{\tau,s,x} \Pi_P(x;v) + \sum_i (\tau_i \Pi_i(x;v) + s_i + f(x_i;v_i))$$

$$\text{s.t.}(1-\tau_i)\Pi_i(x;v) - s_i - f(x_i;v_i) \geq \Pi_i^O(x;v)$$

where

$$\Pi_i^O(x;v) = \frac{\exp\{\delta v_i^I\}}{\sum_{j\in\mathcal{N}_{-i}}(1+\theta_2 x_j)\exp\{v_j^I\} + \exp\{\delta v_i^I\} + \exp\{v^P\} + 1}$$

Equivalently,

$$\max_{\tau,s,x} \Pi_P(x;v) + \sum_i (\Pi_i(x;v) - \Pi_i^O(x;v))$$

Since

$$\begin{aligned}\frac{\partial \Pi_i(x;v)}{\partial x_i} &= \frac{\theta_2 \exp\{v_i^I\}(\sum_{j\in\mathcal{D}}(1+\theta_2 x_j)\exp\{v_j^I\} + \sum_{j\in\mathcal{S}}\exp\{\delta v_j^I\} + \exp\{v^T\} + 1) - \theta_2 \exp\{v_i^I\}(1+\theta_2 x_i)}{(\sum_{j\in\mathcal{D}}(1+\theta_2 x_j)\exp\{v_j^I\} + \sum_{j\in\mathcal{S}}\exp\{\delta v_j^I\} + \exp\{v^T\} + 1)^2} \\ &= \frac{\theta_2}{1+\theta_2 x_i} s_i(1-s_i)\end{aligned}$$

$$\begin{aligned}\frac{\partial \Pi_i(x;v)}{\partial x_j} &= \frac{-\theta_2 \exp\{v_j^I\}(1+\theta_2 x_i)\exp\{v_i^I\}}{(\sum_{j\in\mathcal{D}}(1+\theta_2 x_j)\exp\{v_j^I\} + \sum_{j\in\mathcal{S}}\exp\{\delta v_j^I\} + \exp\{v^T\} + 1)^2} \\ &= \frac{\theta_2}{1+\theta_2 x_j} s_i s_j\end{aligned}$$

$$\begin{aligned}\frac{\partial \Pi_P(x;v)}{\partial x_i} &= \frac{-\exp\{v^T\}\theta_2 \exp\{v_i^I\}}{(\sum_{j\in\mathcal{N}_{-i}}(1+\theta_2 x_j)\exp\{v_j^I\} + \exp\{\delta v_i^I\} + \exp\{v^P\} + 1)^2} \\ &= -\frac{\theta_2}{1+\theta_2 x_i} s_T s_i\end{aligned}$$

and

$$\frac{\partial \Pi_i^O(x;v)}{\partial x_i} = 0$$

The FOC is characterized by

$$-\frac{\theta_2}{1+\theta_2 x_i} s_T s_i + \frac{\theta_2}{1+\theta_2 x_i} s_i(1-s_i) - \sum_j \frac{\theta_2}{1+\theta_2 x_i} s_i s_j = 0$$

This turns out to be an identity, suggesting that without cost of tilting exposures, any allocation of exposure is optimal. We therefore need to introduce cost of exposure intensity. There are two ways to do that.

- linear cost $\theta_3 x_i$
- opportunity cost $\sum_i x_i = 1$