Cyber Income Inequality[*]

Li Liao[§] Chen Lin[†] Lu Liu[§] Xincheng Wang[‡] Yang You[§]

THU HKU THU HKU HKU

March 21, 2023

Abstract

We study the income inequality among streamers using the administrative data of a leading Chinese live-streaming platform. The live-streaming technology enables a superstar to produce new entertainment products matched with demand and occupies a larger market share. Imagine an extreme case; the best streamer hosts live for 24 hours, earns all possible income, and leaves zero time for other streamers. Our data show that the income distribution of the highest-paid streamers follows Zipf's Law and appears to be even more concentrated than any offline business: NBA top players, Forbes celebrities, and billionaires. Income inequality increased rapidly as the platform expanded from 2018 to 2020 — for example, the income share of the platform's top 10 streamers increased from 14.82% to 45.15% as its revenue grew by 142%. To estimate inequality elasticity to the market size, we study four quasi-experimental shocks: potential market size proxied by economic development and Fintech coverage, quarter-end revenue spikes induced by the seasonal incentive regime, user surge induced by capital raising, and the Covid-19 lockdown in Wuhan. Gini coefficient elasticity ranges from 1.3% to 10.6% estimated from the cross-city variations (local economic development and Covid-19 Wuhan lockdown); the time-series variations (quarter-end and user surge before capital raising) imply an elasticity ranging from 3.6% to 25.5%.

JEL Classification: *D63, L25, L82, L86*

Keywords*: Digital Economy, Income Inequality, Superstar Effect, Market Size*

[*]We would like to thank Mingzhu Tai, Haishi Harry Li, Michael Wong, John Klopfer, Yanhui Wu for their helpful comments. Lin acknowledges the financial support of the National Natural Science Foundation of China (Grant 72192841) and the Research Grants Council of the Hong Kong Special Administration Region, China (Project No. T35/710/20R). The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of the live-streaming platform. All errors are our own. First Draft: September 15, 2022.

[§]Tsinghua University, Email:liaol@pbcsf.tsinghua.edu.cn.

[†]Department of Economics, The University of Hong Kong, Email:chenlin1@hku.hk

[§]Tsinghua University, Email:liul.19@pbcsf.tsinghua.edu.cn.

[‡]HKU Business School, The University of Hong Kong, Email:xcwang@hku.hk.

[§]HKU Business School, The University of Hong Kong, Email:yangyou@hku.hk.

1 Introduction

Creative people have numerous opportunities to make money online in the Internet era. Writers, singers, video makers, and other content creators can achieve a large potential audience to share their artistic talent. It's reported that the size of the creator economy market is estimated to reach $104 billion in 2022[1]. Additionally, more than 50 million people consider themselves as creators around the world[2]. In this context, the income distribution of these creators has become a central issue. How big is the income inequality? How does digital platform expansion affect the income inequality of those earning income on the platform? Some view the Internet as a technology that provides equal economic opportunity, whereas others argue that the Internet amplifies ability differences at minimal or zero replication cost and with a strong network effect; thus, those at the top enjoy most of the economic gains, increasing income inequality.

There are many internet markets where creators can post their work, of which live streaming is brand new but booming. To estimate income inequality online, we obtain the complete administrative data of a leading Chinese live-streaming platform, including the records of 2.44 million paying users, 51.24 million viewers, and 1.42 million registered streamers from April 2018 to April 2020. First, we calculate income inequality measures using the tipping income of the streamers, which comes directly from the platform's paying users. We compute the Gini coefficient and income share for the top 10, 100, and 1,000 streamers, and we characterize the fat-tail income distribution of the top streamers using power law (PL) exponents. The Gini coefficients are higher than any country's, even if we limit our analysis to the top 1,000 streamers. We find that the income tail distribution follows Zipf's law and that streamers' incomes are even more skewed than the wealth distribution of the wealthiest billionaires on the Forbes list, movie stars, singers, and NBA

[1]This number is calculated in the Influencer Marketing Benchmark Report 2022. See `https://influencermarketinghub.com/influencer-marketing-benchmark-report/` for detail.

[2]This number is estimated using bottom's up total addressable market analysis. See `https://signalfire.com/blog/creator-economy/` for detail.

stars.

Next, we explore dynamic income inequality over time. The income share of the top 10 streamers increased from 14.82% in April 2018 to 45.15% in April 2020. The income share of the top 100 streamers increased from 44.56% to 69.61%, and that of the top 1,000 streamers increased from 86.20% to 94.35%. More than 83.27% of the platform's streamers generate revenue of less than 1,000 RMB (144 USD) monthly. Thus, the Gini coefficient calculation may be sensitive to how we select streamers for our sample, as some streamers are not pursuing a career in this industry but are amateurs who stream for entertainment.[3] If we limit our analysis to streamers who earn more than 1,000 RMB, the Gini coefficient increases from 0.84 to 0.91 for the two years from June 2018 to April 2020, when total tipping revenue increased from 83.44 million RMB to 181.40 million RMB per month. We also find that the number of superstar streamers who make more than 100,000 RMB per month increases in the sample period as the platform grows. In contrast, the number of middle-income streamers, whose monthly pay ranges from 1,000 to 100,000 RMB, significantly shrinks in the later sample period. To characterize income inequality in the upper tail, we estimate the PL exponent of the top streamers by plotting log rank and log income: the PL exponent is 1.02, much smaller (representing greater inequality) than the 1.66 of the wealth inequality among the Forbes top 100 billionaires, 2.01 - 2.69 of the wealth inequality of traditional superstars, roughly 1.5 of the wealth inequality in the U.S. (e.g., Kleiber and Kotz, 2003; Klass et al., 2006), and 1.5–3 of the income inequality in the U.S. (e.g., Atkinson and Piketty, 2007; Atkinson, Piketty, and Saez, 2011). The PL exponents of the top 100 streamers decreased from 1.44 to 0.92 from June 2018 to April 2020.

Our aggregate time-series analysis results indicate that superstars, which refer to the top streamers, benefit more than other streamers as the platform expands over time. This rising income inequality appears to be a consequence of market expansion. We investigate the allocation of users to streamers and the formation of their loyal relationships to pro-

[3] If we use streamers with positive income, the Gini coefficients rise from 0.96 to 0.98.

vide potential mechanisms behind the size-inequality relationship. Specifically, we split the overall market size into the number of paying users and the intensity of the tipping amount supplied by those active users. We document that active users who tip a lot spend their attention and money disproportionately toward superstar streamers on an intensive margin. Moreover, newly registered users immensely allocate their spending to top streamers, while the existing users will gradually build solid ties with superstar streamers.

To further test the relationship between market size and inequality, we estimate the elasticity of the Gini coefficient with respect to market size, which is proxied by total revenue.[4] For identification purposes, we explore four quasi-experimental variations: market size proxied by local economic development and Fintech coverage, quarter-end revenue run-up in response to platform incentives, user surge during a critical capital raising event, and market shrinkage induced by the Covid-19 lockdown in Wuhan.

First, we construct city-level income inequality metrics and correlate them with the local market size. Each user is assigned to one city according to the IP address of their live streaming account login. All tipping revenue breaks down to the city level according to the user's location.[5] We use city size, GDP, and local digital development level as our instrument variables for actual market size. These cross-city variations yield estimates of approximately 0.019 to 0.023: when market size doubles, the Gini coefficient increases by 0.019 to 0.023.

Second, we explore the time-series shocks caused by the platform's incentive for streamers. Our data provider ranks streamers at the end of each quarter (March, June, September, and December). Thus streamers work particularly hard at the end of the quarter to improve

[4] The parallel estimation at the country level involves regressing the Gini coefficient on GDP. However, the empirical difficulty of this approach makes determining the exogenous shock at the GDP level difficult.

[5] In a given month, if a user logs in from more than one city, we choose the city with the longest viewing time as that user's location. The user's location may change if the login IP address migrates to a different city.

their positions.[6] We find that total tipping increases by 37% in quarter-end months.[7] Thus, we use quarter-end month dummies as instrument variables for market size to identify the elasticity of the Gini coefficients in both the market aggregate time-series data and city-level panel data. The elasticity based on aggregate market data is 0.077, and that of the city-level panel data is 0.079.

Third, we exploit the user surge before a critical capital raising event of the platform. A sequence of marketing endeavors was made for market growth, which finally contributed to the improvement in financial performance before its capital raise. We use the tipping amount change from new users to instrument the market size variations, and the estimated elasticity ranges from 0.056 - 0.095. We also construct the Gini index based on the distribution of streamers' loyal fans. Our result is not due to the outlier paying users who spend a lot but that fans are more concentrated due to the exogenous market expansion.

Finally, we estimate the elasticity of income inequality by exploiting the market shrinkage caused by the Covid-19 outbreak in Wuhan. Covid-19 attracted tremendous public attention (and fear), dampening the demand for online entertainment. Moreover, the Wuhan lockdown (January 23, 2020) led to a decrease in tipping from users in Wuhan. A regression discontinuity shows that the number of paying users in Wuhan dropped by one-third and that tipping revenue from Wuhan decreased by half during the lockdown. The elasticity can be computed as the ratio of change in the Gini coefficient to the market-size shrinkage from before to after the Covid-19 outbreak. Our event-based estimates range from 0.023 to 0.056 and have strong statistical power.

This paper relates to four strands of the literature: that on estimations of social inequality using administrative data, the impact of the digital economy, the superstar effect, and

[6] Ranking is crucial for streamers in all dimensions. The ranking is associated with revenue-sharing incentive contracts with multichannel network agencies. Top-ranked streamers can attract more businesses and make their names known to more people than other streamers.

[7] Although we do not see more paying users at the quarter ends, the total tipping amount surges. The seasonality of market size is mainly driven by the intensive margin (paying users tend to spend more on streamers) rather than the extensive margin of attracting more fans.

the relationship between market size and inequality. Studies document rising inequality using administrative data from the U.S., France, the United Kingdom, Norway, Denmark, Spain, China, Japan, India, and Russia.[8] To the best of our knowledge, this is the first paper to estimate income inequality in cyberspace using complete administrative data and to document income divergence among social media influencers, even at the upper tail.

This paper also contributes to the debate on how the Internet reshapes economic activities by highlighting the challenge of rising income inequality online. Some studies emphasize the bright side of the Internet, for example, that it creates flexible employment opportunities (e.g., the gig economy) and is inclusive (Burtch, Carnahan, and Greenwood, 2018; Huang et al., 2020; Cook et al., 2021; Bernhardt et al., 2022). In addition, the long-tail effect of the Internet drives demand away from products with mass appeal (Brynjolfsson, Hu, and Smith, 2003; Anderson, 2006; Brynjolfsson, Hu, and Simester, 2011; Zentner, Smith, and Kaya, 2013) and an increasing number of suppliers can provide a wider variety of services and thrive without directly competing with the dominant players in their markets. One natural conjecture is that cyberspace might be a land of opportunity because the Internet is equal-access, and anyone can start a business at a relatively low cost. However, our empirical results point in the opposite direction because Internet technology can create a "winner-take-all" market in which very few superstars earn enormous amounts of money and dominate the activities that they engage in.

The superstar effect is widespread in economic activities (Rosen, 1981). For CEO compensation, the PL exponent has ranged from 1 to 1.5 since 2000 (Blackwell et al., 2015),[9] and the largest firms in the world have a PL exponent of approximately one (Gabaix and

[8]The U.S. (Piketty and Saez, 2003; Kopczuk, Saez, and Song, 2010; Chetty et al., 2014; Piketty and Saez, 2014; Saez and Zucman, 2016; Blanchet, Saez, and Zucman, 2022), France (Piketty, 2003, 2011), the United Kingdom (Lindert, 1986), Norway (Black et al., 2020), Denmark (Hvidberg, Kreiner, and Stantcheva, 2020), Spain (Alvaredo and Saez, 2009), Japan (Moriguchi and Saez, 2008), China (Piketty, Yang, and Zucman, 2019), India (Chancel and Piketty, 2019), Russia (Novokmet, Piketty, and Zucman, 2018), and French and British colonies (Alvaredo, Cogneau, and Piketty, 2021).

[9]See Edmans, Gabaix, and Jenter (2017) for a detailed review, which presents stylized facts about the level of CEO and other top executives' pay over time and across firms in the U.S.

Landier, 2008; Gabaix, 2009).[10] The superstar effect is even more substantial on our live streaming platform: in April 2020, the PL exponent of the top 100 streamers was 0.92, and the PL of the top 2000 streamers was 0.74. The Internet and streaming technology help superstars, who market themselves at a relatively low cost, outperform the average participant even more than they would otherwise. This is consistent with the explanation of the rise in inequality rooted in the importance of scale and skill-biased technological change (Rosen, 1981, 1983; Kaplan and Rauh, 2013; Koenig, 2022). Nevertheless, the difference is that live streaming, which features one streamer who creates and delivers content to the audience in real time over the Internet, is a new generation of technological change compared to other media. The audience engages in the live stream not only to enjoy the talent show of the streamer but also to gain pleasure from the interaction process. This social attribute means that the best way for streamers to satisfy the audience is not to spend much time behind the scenes practicing and preparing for the show but to interact with the audience around the clock, which further breaks down the constraints on labor supply. This new technological change, combined with its pay-what-you-want business model in live streaming, enables superstars to produce infinite new entertainment materials and scale up without diminishing marginal returns.

One lingering question relates to the relationship between inequality and the market.[11] One hypothesis is that various mechanisms explain increasing income equality as a function of a broader market. Gabaix and Landier (2008), and Gabaix, Landier, and Sauvagnat (2014) argue that increasing firm size explains CEO pay. They explain why firm size can

[10] Extreme inequality also applies to conventional media, for example, music, books, and movies (Rosen, 1981; Adler, 1985; Salganik, Dodds, and Watts, 2006).

[11] The most famous debate is whether the Kuznets curve exists; that is, whether the market first increases and then decreases economic inequality. Our analysis shows that the Internet market is positively related to economic inequality.

amplify a slight talent difference into a significant income gap.[12] The urban literature has extensively documented the positive relationship between city size and income inequality with explanations based on agglomeration, technology adoption, and migration.[13] This paper contributes to the literature by exploiting four exogenous variations (local development, revenue seasonality, user surge induced by capital raising, and the Covid-19 lockdown in Wuhan) that affect the platform's market size to provide a causal estimation of how inequality responds to market size.

Several theories might explain why an increase in market size leads to a rise in inequality. First, the technology change amplifies differences in streamers' abilities. In the past, a singer or an actor could entertain only a limited number of people. Today, live streaming technologies allow audiences to access their favorite content wherever and whenever they want. Furthermore, the near-zero replication costs in the digital economy encourage the provision of free live streams (Goldfarb and Tucker, 2019). A larger market increases talented streamers' advantages. Consequently, consumption distribution patterns are skewed toward a minority of products (Rosen, 1981; Adler, 2006). Second, consumers' need to consume the same art as others indicate that the larger the market, the greater the inequality. New entrants are attracted to streamers who are already popular. As Stigler and Becker

[12]Based on an assignment model developed by Tervio (2008), Gabaix and Landier (2008) suppose that many firms of various sizes compete to hire talented CEOs. This competition provides the most efficient outcome: the largest firm matches with the best CEO, the second-largest firm matches with the second-best CEO, and so on. Consequently, even talent distribution has an upper bound. Wages can be unbounded as the best managers are paired with the largest firms, which makes their talent valuable and results in a high level of compensation. This model might be analogous to the matching process between users and streamers, in which each user might only pay for the best streamer, even though the best streamer is only slightly better than the second-best streamer.

[13]In an analysis of the cross-sectional data of 79 U.S. metropolitan areas, Long, Rasmussen, and Haworth (1977) find income inequality appears to increase with city size. Nord (1980) develops a unified model using the labor market, capital market, and monopoly advantage to derive the relationship of the size distribution of income to city size. Baum-Snow and Pavan (2013) discuss the evolution of the U.S. city-size inequality premium since 1979 and investigate the role of city size in generating growth in several measures of wage inequality. They find that agglomeration economies are crucial to understanding the link between city size and wage inequality. Chen, Liu, and Lu (2018) establish a positive correlation between city size and urban inequality in China and find that the size—inequality correlation is mainly driven by between-group inequality, not within-group inequality. When the market is large enough, firms have sufficient incentive to adopt new technologies that improve economic efficiency, e.g., Behrens, Duranton, and Robert-Nicoud (2014) and Murphy, Shleifer, and Vishny (1989).

(1977) and Adler (1985) state, the consumption of a piece of art is not a momentary experience but a dynamic process in which "the more you know, the more you enjoy." When a streamer is popular, more viewers are familiar with her, and there is more media coverage of her, which is why consumers tend to consume what others are consuming.

The rest of this paper is organized as follows. Section 2 introduces the background of the live streaming business and the unique administrative data used. Section 3 describes income inequality and its dynamics. Section 4 investigates the allocation of users to streamers and the formation process of their loyal relationships to provide potential mechanisms behind the size-inequality relationship. Section 5 explores time-series and cross-sectional shocks to live streaming revenue and estimates the income inequality response to market expansion. Section 6 concludes the paper.

2 Institutional Background

2.1 The Uniqueness of Live Streaming

The "superstar" effect in the entertainment sector is not new and has been well documented. The celebrated papers Rosen (1981, 1983) point out that the superstar phenomenon has become increasingly visible. He argues that the leading cause behind this is technological change, admitting a certain kind of duplication in which the seller simultaneously delivers services to many buyers, thus allowing for an expansion in market scale. Although extensive research has been carried out afterward Hamlen (1991); Krueger (2005); Kaplan and Rauh (2013), they still need to get a consistent conclusion revealing the behind mechanisms. A most recent paper Koenig (2022) exploits the historic rollout of television to test whether the technical change that extends market scale can generate winner-take-all dynamics. Compared with the previous studies, the live-streaming industry this paper focuses on has the following unique features.

First, live streaming, which features one streamer who creates and delivers content to the audience in real time over the internet, is a new generation of technological change.

The audience engages in the live stream not only to enjoy the talent show of the streamer but also to gain pleasure from the interaction process. This social attribute means that the best way for streamers to satisfy the audience is not to spend much time behind the scenes practicing and preparing for the show but to interact with the audience around the clock, which further breaks down the constraints on labor supply. This new technological change in live streaming enables superstars to produce infinite new entertainment materials and scale up without diminishing marginal returns. Imagine an extreme case that the best streamer is the favorite of all viewers and hosts live for 24 hours, and then she will be the one earning all possible income and leaving viewers zero free time for other streamers.

Second, different from a fixed ticket price when watching concerts or basketball games, viewers can watch live-streaming content freely and give streamers voluntary tips in the form of virtual gifts. Theoretically, tipping in live streaming is similar to the Pay-What-You-Want (PWYW) pricing strategy. In PWYW pricing, buyers can pay any amount they desire for a given good or service, including zero, which provides consumers maximum power over the price setting process (Kim, Natter, and Spann, 2009; Schmidt, Spann, and Zeithammer, 2015). Thus, the streamer's income comes not only from the increase in sales volume but also from the increase in user loyalty and payment intensity, which increases the probability of generating extreme income.

Third, in the era of the platform economy, we can obtain the digital footprint of both the supply and demand sides of live streaming, which makes it possible to investigate the assignment process of customers to sellers and study the allocation of viewers' attention and money resources.

2.2 Live Streaming and the Income of Internet Celebrities

Live streaming is an innovative leisure-enhancing technological change in which voice and audio are simultaneously recorded and broadcast in real-time. All can show their talents and creativity to the public by simply registering as a streamer on a live streaming platform.[14] The existing influencers can also use it as a new instrument to interact with their fans and generate income. Viewers gain entertainment and social interaction experience with streamers and voluntarily pay for their performances. To some extent, live streaming provides an open ground for creators to show their talents, attract fans on the internet, and make income from them.

Information technology development enables live streaming to gain popularity worldwide. In the US, the introduction of Facebook Live and YouTube Live Streaming Channels, Amazon's acquisition of Twitch, and the development of Twitter's Periscope have all pointed to the appeal of the live streaming industry in the US (Lu et al., 2021). It's reported that a total of 204.2 million hours were streamed in the 2022Q2 on Twitch[15]. In China, the market size of live streaming has enjoyed the growth rate of over 60% for three years, reaching 184.44 billion yuan in 2021. The cumulative number of streamers in China has reached 140 million by the end of 2021.[16] The number of live stream users increased from 344 million in 2016 to 703 million in 2021, which account for 68% of all Internet users in China.[17] Additionally, many Chinese live streaming service providers (such as Bilibili, Tiktok, Kuaishou, DouYu, Huya, and Momo) have grown into billion-dollar businesses that are listed in the US.

The live streaming economy has developed into an industry with internet celebrities at the core. The income of internet celebrities can be huge. Some internet celebrities make as much as movie stars or top singers. According to Forbes, MrBeast, the top YouTuber, earned $54 million in 2021; Jake Paul, in second place, earned $45 million; and Markiplier, in third place, earned $38 million.[18] The most popular tiktoker Charli D'Amelio earned $17.5 mil-

[14]The talents can be singing, dancing, answering questions, sharing opinions with the audience, or just chatting with viewers with a pretty face.

[15]https://www.statista.com/statistics/1030859/hours-streamed-twitch/

[16]Data comes from Chinese Internet Performance (Live Streaming) Industry Development Report 2021. See http://perform.capa.com.cn/1670901992643.pdf for detail.

[17]https://www.statista.com/statistics/1061708/china-online-streaming-user-number/

[18]https://www.forbes.com/sites/abrambrown/2022/01/14/the-highest-paid-youtube-stars-mrbeast-jake-paul-and-markiplier-score-massive-paydays/?sh=6d7b6c311aa7

lion in 2021; Dixie D' Amelio, in second place on TikTok, earned $10 million; and Addison Rae, in third place, earned $5 million.[19] In China, celebrities like Mengling Yi and Dantong He have tens of millions of fans on TikTok thanks to their beautiful appearance. Daxian Zhang, one of the most famous game streamers on Huya, has gained remarkable popularity among up to 29 million followers for his excellent skill in the game the Glory of King. In 2022, Daxian Zhang live-streamed 2,100 hours and earned a tipping revenue of 10.93 million RMB.[20] Moreover, one streamer called "a small dough", who won the championship of the Douyu Fan Festival in 2019, has attracted 25.75 million fans with her funny performance. She could receive virtual gifts worth millions of RMB just in one month.[21]

2.3 The Live Streaming Platform and its Business Model

We obtain administrative data from one of the Chinese most popular live streaming platforms listed on a US stock exchange. According to the 2021 annual report, the platform has over 50 million active monthly users and more than 5 million paying users. Our data contain complete administrative data for two major entertainment genres: performing arts (e.g., singing, dancing, and playing music) and outdoor activities (e.g., traveling).

Two primary types of users participate in the platform's activities: viewers and streamers. Viewers can watch live streams on the platform without registering via the website or mobile app. However, anyone who wants to live stream must register on the platform and verify their identity with government-issued identification. After doing so, they can apply to create a streaming room.

During a live stream, the viewers can interact with the streamers and each other, primarily through bullet chats. Viewers can purchase virtual gifts on the platform with digital currency (pegged to RMB) and send them to streamers. According to the annual report of our data provider, more than 90% of the company's net revenues are from the sale of virtual

[19] https://www.forbes.com/sites/abrambrown/2022/01/07/top-earning-tiktokers-charli-dixie-damelio-addison-rae-bella-poarch-josh-richards/?sh=2077f0803afa.

[20] Data collected from Huya app.

[21] Data collected from Douyu app.

gifts. Streamers can redeem virtual gifts for fiat money from the platform.

Streamers' virtual gift income does not equal their total real income. First, the platform runs a 50-50 revenue-sharing policy between the platform and its streamers, according to its Director of Investor Relations in a 2019 earnings call.[22] Second, although live streamers pay a progressive income tax, the top streamers may engage in tax evasion.[23] Third, the top streamers may have a higher ratio of sharing revenue as a result of their greater bargaining power. Fourth, the top streamers may have other types of income, such as variety-show income, endorsement fees, etc. Such extra revenue may cause us to underestimate the actual level of income inequality.

2.4 Data Description

The data contain the complete online transaction (tipping) records of 51.24 million viewers, 2.44 million paying users,[24] and 1.42 million registered streamers from two representative entertainment genres: the performing arts (e.g., singing, dancing, and playing music) and outdoor activities (e.g., traveling) from April 2018 to April 2020. The watching records are also available to us: each observation includes the user's ID, the streamer's ID, the timestamps at which the user entered and exited the live-streaming channel, and the user's login IP address.

The data include 560.62 million raw tipping records, including the anonymous ID of the user who sent the virtual gift, the anonymous ID of the streamer who received the gift, the time that the gift was sent, and the dollar value of the gift. We first aggregate the raw tipping records to the user–streamer-month level (the total tipping amount received by streamer i from a paying user u in month t), ending with 15.31 million observations. To filter out suspicious transactions, we further validate whether the paying user u watched

[22] The original text is as follows: "Regarding the revenue sharing ratio between our platform and streamers, we have been fairly stable to keep it at a 50-50 split in the past few quarters, and we don't foresee a significant change going forward."

[23] http://beijing.chinatax.gov.cn/bjswj/c104182/202206/1aab3431d9aa46e9978e1ab49e3c1de6.shtml

[24] The term "paying user" refers to a registered user who gave a virtual gift to a streamer at least once during the sample period.

live-streaming in month t. If not, we delete the tipping record for that user that month, removing 1.80 million observations from the user–streamer-month level sample.[25] In total, we identify that 2.44 million unique paying users sent 3.47 billion RMB to 131,960 streamers. We further aggregate the tipping data at the streamer level to calculate streamers' income.[26]

To explore more variation, we construct city-level income inequality metrics. We assign each user u to a unique city of residence j in the month t based on the login IP address in the watching data. We first identify the prefecture-level city based on the user's login IP address. For users with login IPs in multiple cities, we define the city with the longest watching time as the user's residence city for that month. Next, we aggregate the watching time by the city for each user every month. Using the user ID, we then assign the user's city of residence to her tipping behavior.[27]

In Appendix Figure A1, we show the value of the virtual gifts received by the streamers in the sampled during the sample period.[28] Panel A shows that the platform steadily expanded over time. The total monthly tipping amount rose from 75.02 million RMB in April 2018 to 181.40 million RMB in April 2020. We decompose the total tipping amount into the number of paying users and the tipping amount per capita. The rising average monthly tip amount per user, from 387.66 to 856.02 RMB, is the main driving force behind the platform growth, as shown in Panel C.[29]

We can also decompose the total tipping amount into the number of streamers receiving virtual gifts and the average tipping income per streamer. Appendix Figure A4 plots the

[25]This filter might be too conservative and underestimates the total tipping amount. First, fans can send virtual gifts without watching the live-streaming channel. Second, some watching records might be missing from the platform data.

[26]As shown in Appendix Table A1, streamers' income varies greatly.

[27]In the user–streamer-month tipping data, there are 763,598 observations, 5.655 percent of the total 13,502,556 records do not contain valid IP addresses.

[28]In Appendix Figure A2, the raw monthly tipping numbers exhibit strong seasonality. Thus, we use the average tipping statistics of the past three months as the baseline to adjust for seasonality.

[29]Panel B shows that the number of paying users increased steadily before July 2019 and decreased since then.

number of streamers who earn any virtual gifts has decreased over time[30], while the tipping income per streamer went up. Specifically, although the number of streamers who earn tipping income has decreased by 18.45%, their average income has nearly tripled from 5,660.11 to 16,769.76 RMB, indicating that inequality among streamers has increased and virtual gifts have become more concentrated among fewer streamers.

3 Inequality Online and Dynamics

In this section, we characterize the income inequality among live streamers on the platform — top income shares, Gini coefficients, power-law exponents, and percentile income gaps, and document rising inequality from April 2018 to April 2020.

3.1 Top Share of Income

We first compute income shares of the top-performance streamers on the platform. We rank streamers according to their aggregate income over the past three months and calculate the top groups' share of total tipping income. Figure 1 shows that the income share of the top groups has been rising over time.[31] Specifically, the share of the most well-paid streamer rose from 3.71% to 15.99%, the share of the top 10 streamers rose from 14.82% to 45.15%, and the share of the top 100 streamers rose from 44.56% to 69.61% during the same period. By the end of our sample period, the income share of the top 1,000 streamers had reached an extremely high level of 94.35%. These results demonstrate that top streamers have made a large amount of money directly from their fans, and the superstar effect among streamers is increasingly apparent.

[30] Appendix Figure A3 shows the picture of the number of streamers and their average tipping income before the seasonality adjustment.

[31] Appendix Figure A5 shows the top group's share of income before the seasonality adjustment.

3.2 Gini Coefficients

We use the Gini coefficient as the measure of inequality, as is common in the literature.[32] Figure 2 plots the seasonality-adjusted inequality trend as measured by the Gini index.[33] We find that the number of streamers who earn a moderate income has dramatically declined since June 2019, whereas the number of superstar streamers steadily increased during our sample period. The value of the virtual gifts these superstar streamers receive is exceptionally high, and 60 streamers with a three-month rolling tipping income of more than one million RMB in our sample, suggesting that tipping income is concentrated in the top streamers. Furthermore, the solid lines in these figures consistently show an upward trend in the dynamic change of the Gini coefficients, reflecting a widening income gap among streamers.

3.3 Power Law Exponents

The Power Law (PL, hence) refers to a distribution that satisfies, at least in the upper tail, $P(Size > x) = kx^{-\theta}$, where θ is the PL exponent and k is a constant.[34] The PL has two properties. First, the lower the PL exponent, the fatter the tails, and the greater the inequality among the upper-tail distribution. Second, the power law distribution implies a linear relationship between log value and log rank.

Does streamers' tipping income follow a PL distribution in the upper tail? If so, how large is the PL exponent? According to the second property, we can intuitively answer these questions by visualizing the distribution of tipping income. We order the streamers according to the total amount of their received virtual gifts. Following the standard procedure of power law, we place the log streamer's rank on the y-axis and her log income on

[32]The Gini coefficient is defined as $1 - \frac{1}{y}\int(1 - F(y))^2 dy$, where $F(y)$ is the share of the population with an income level of less than y and y is the average income. This measure can be interpreted as the area between the 45-degree curve and the Lorenz curve divided by the triangle below the 45-degree curve.

[33]Appendix Figure A6 shows that the Gini coefficients have strong seasonality. Thus, we aggregate streamers' tipping income over the past three months and calculate the Gini index using the resulting income distribution.

[34]Many economic data roughly follow PL distributions. Please see (Gabaix, 2009) for a detailed review.

the x-axis and test whether this plot exhibits a straight line. The streamer's tipping income approximately follows a PL distribution, and we obtain the PL exponent by estimating the slope of the line. Appendix Figure A7 shows a good PL fit in the upper tail of the total tipping income distribution. The PL exponent is approximately 1 (following Zipf's law with parameter 1). In Panels C and D, a linear relationship fits streamers' income perfectly when we only include streamers whose income is above 100,000 RMB. Thus, PL exponents are suitable measures for inequality in the tail.[35]

We compare our PL exponent estimated from streamers' total income with that calculated from superstars in other industries. First, in Panel A of Figure 4, we use the wealth of the world's wealthiest people in Forbes as the benchmark. We find that the PL slope of streamers is much flatter than the PL slope of Forbes billionaires[36], which indicates that a streamer relatively earns less money given the same ranking, and there is a greater probability of finding very high income. Top streamers' income inequality[37] is greater than the wealth inequality in Forbes' list. Second, in Panel B of Figure 4, the world's highest-paid celebrities selected by Forbes magazine are used as a benchmark. Specifically, this list ranks those famous actors, singers, athletes, and Internet celebrities according to their pretax earnings. Based on the slope of the two lines, we find that payments are more concentrated in streamers. Third, in Figure 5, we empirically collect data on movie stars, singers, and NBA stars to provide a more fair comparison, considering that live-streaming should be more like the entertainment industry. People may already know that Taylor Swift or Michael Jordan makes a lot of money. Still, we find something new: top streamers have even higher inequality than offline superstars. Table 1 gives the PL exponent and Gini index values for different groups.

[35]The power law tests would fail if we include more streamers with lower income — a kink occurs in Panels A and B.

[36]To make it more comparable, we scale down Forbes' wealth so that the wealthiest billionaire's wealth matches the highest-paid streamers' total income.

[37]To some extent, the total income here is part of the wealth streamers accumulated from the live-streaming platform.

Then, we provide evidence that inequality is also rising among the top streamers (PL exponent in an upward trend). We construct the monthly PL exponents as the following: First, all streamers are ranked by the aggregate value of their received virtual gifts in the past three months. Next, we estimate time-varying PL exponent β_t with eq.(1) and plot the dynamics in Figure 6.

$$\ln rank_{jt} = \alpha_t + \beta_t \ln size_{jt} + \varepsilon_{jt}, \forall t \tag{1}$$

The PL exponent of 50 most well-paid streamers raised from -1.36 to -0.95 in Panel A, from -1.42 to -0.9 among the top 100 streamers in Panel B, from -1.05 to 0.9 among the top 1000 streamers in Panel C, and from -0.9 to 0.75 among the top 2000 streamers in Panel D.[38]

3.4 Inequality Measures Using Streamers' Fans

Essentially, streamers are internet idols who provide entertainment utility online, attract fans (followers of their live-streaming channel), and ultimately make money from their fans (paying users) through virtual gifts provided by the platform. Streamers need to increase the number of fans to generate further income flow.[39] In this section, instead of the streamers' income, we use the number of fans to compute inequality and test whether users are more attracted to the top streamers over time.

How do we define "fans" in the data? Suppose there are N streamers. First, $user_i$ is only eligible to be a fan if she gives virtual gifts of value no less than 50 RMB in month t. Then, we assign $user_i$ to $streamer_j$ as a fan if $user_i$'s tipping to $streamer_j$ accounts for the highest proportion of her total tipping expenditure in month t, that is, $j = \arg\max_n \frac{Tip_{int}}{\sum_{n=1}^{N} Tip_{int}}$.[40] For any given month t, we can count the number of fans of each streamer.

[38] Even if we plot the raw coefficients before the seasonality adjustment, the upward trend of the PL exponent is clear; see Appendix Figure A8.

[39] According to the platform's annual report, more than 90% of the company's revenue comes from live-streaming virtual gifts. The idol-fan relationship is crucial for the platform's revenue.

[40] Idol-fan relationship is defined at the monthly level. Thus, user i can be a fan of different streamers in different months, but each user is uniquely assigned to one streamer in a month.

3.4.1 Top Share of Fans

In this subsection, we rank the streamers by their aggregate number of fans in the past three months adjusted for seasonality. We calculate the ratio of the number of fans of the top streamer group to the total number of fans. Appendix Figure A9 shows that fans who have the resources and are willing to tip concentrate their tipping on the top streamers. The top 10 streamers' share of total tipping income doubled, increasing from 15.62% to 31.59%. The top 100 streamers' share rose from 32.84% to 49.13%, and the top 1,000 streamers' share increased from 69.99% to 80.64% during the sample period. The implication of these results is consistent with the results in Section 3.1, that is, the top group receives more and more of the available income. Appendix Figure A10 shows the results before the seasonality adjustment.

3.4.2 Gini Coefficients with Fans

In this section, we calculate the Gini coefficients among streamers according to the distribution of their loyal fans. The definition of a loyal fan is the same as in Section 3.4.1. Appendix Figure A11 and A12 show the dynamic change in this measure calculated using the streamers with more than a certain number of loyal fans over time and the overall trend is upward, similar to that in Figure 2. Even within the top streamer group, inequality is increasing, as shown in Appendix Figure A13. These results are consistent with Section 3.2.

3.4.3 Power Law Exponent with Fans

Loyal fans are an essential resource for streamers and, to a certain extent, provide sustainable income and reputation. In parallel, we calculate the PL exponent using the streamers' number of loyal fans. As in Appendix Figure A14, the relationship between $\ln rank$ and $\ln size$ in the upper tail is approximately linear, which means that the number of loyal fans follows a PL distribution. Furthermore, Appendix Figures A15 and A16 show that the monthly PL exponent calculated using the number of fans also shows an upward trend. All of these results are pretty similar to those described in Section 3.3.

3.5 Other Inequality Measures

We construct other inequality measures (Glaeser, Resseger, and Tobio (2009)) to investigate additional income distribution parameters and decompose where the widening income gaps arise.

3.5.1 Percentile Gap

Percentile values can tell us the streamers' relative standings. The first column in Appendix Table A2 presents the dynamic change in the variance of the log income variable. From June 2018 to April 2020, the variance rose from 8.38 to 9.25, indicating that the overall income gap among streamers is widening. That observation raises the question: Which streamers have increased their income, and which group has decreased its income? These questions require further exploration. The monthly-level values of the percentiles of the log income variable are shown in the other columns in Appendix Table A2. Furthermore, in Figure 7, we show the time trend of the percentile gaps. Panel A of Figure 7 shows that the gap between the superstar group and the middle-income group is widening. In addition, according to Panel B of Figure 7, the gap line between the middle-income and lower-income groups is almost flat and decreased slightly after January 2020. Accordingly, we can surmise that the increase in overall income inequality is primarily due to the rise of superstars, and inequality in the upper tail seems to be the main driving force for the overall inequality dynamics.

3.5.2 Income Distribution

We further describe the change in the number of streamers who can earn a certain amount of money. Figure 8 Panel A plots the number of streamers by income range. And Panel B shows the number of streamers with aggregate income above 1 million RMB in the past three months. We find that the low-income streamers gradually quit while the number of superstar streamers (above 1 million RMB per quarter) rises. This figure also indicates that the superstar effect is becoming increasingly pronounced and that the possibility of streamers earning a certain income has decreased.

3.5.3 Correlation Matrix

Do these inequality indicators co-move with each other? Appendix Figure A17 and Appendix Table A3 report the correlation matrix among various income inequality metrics calculated in the previous sections. We find robust correlations among the measures over time. This confirms a consistently growing trend of income inequality on the platform. Thus, we primarily choose the Gini coefficient to measure income inequality for our analysis in the next section.

4 Mechanisms

Why would online platform expansion explain the rising cyber-income inequality? Why do the more extensive the market size, the more beneficial the superstar streamers? This section illustrates how users with different consumption levels and varying stages of the customer lifecycle allocate their attention and payment flow to streamers, which may reveal potential mechanisms behind the positive size–inequality relationship. This micro perspective, exploring the behavior of individual users that make up the whole market size, is one of the main differences between this paper and previous studies. Specifically, we split the overall market size into the number of paying users and the intensity of the tipping amount supplied by those active users, and we document two findings: first, active users who tip a lot spend their attention and money disproportionately toward superstar streamers on an intensive margin; second, newly registered users immensely allocate their spending on top streamers, and the existing users will gradually build solid ties with superstar streamers.

4.1 Intensive Margin

Panel C in Appendix Figure A1 shows that the increase in per-capita tipping payment intensity is the main driving force for the expansion of the platform market size. And it seems that in practice, streamers always fiercely compete for active users who are willing to send virtual gifts extravagantly in one live broadcast. Thus, which tier of streamers these active paying users prefer and how loyal they are would be the most critical factors influencing the inequality among streamers. In this section, we explore whether those influential active users prefer superstar streamers and have higher loyalty, which will provide intensive margin evidence for a positive size-inequality relationship. Specifically, for each month t, we divide users into eight groups based on their total monthly tipping spending — less than 100 RMB, 100∼1k RMB, 1k∼5k RMB, 5k∼10k RMB, 10k∼100k RMB, 100k∼500k RMB, 500k∼1 million RMB, and >1 million RMB. Then we explore the heterogeneity of users with different consumption levels regarding their favorite streamer's rank level and loyalty.

Rank level of streamers that users appreciate. In Section 3, we have found that the incomes of streamers are stratified, and a few streamers at the top of the pyramid capture most of the money resources of paying users. A natural question is whether those users with deep pockets will disproportionately follow the top streamers. In Figure 10, we use three methods to define which streamers users appreciate. Panel A explores which tier of streamers the users' money resources are mainly allocated to.[41] Panel B identifies users' favorite streamers based on their watching time in each live broadcast room and explores the ranking of streamers that users spend the most time with.[42] And Panel C shows which tier of streamers the users' attention flow to.[43] Specifically, we use the rank

[41]The evidence about users' money allocation comes from the data recording user i's monthly tipping amount to the streamer j, from which we can define which streamer the user i is most loyal to in month t.

[42]The evidence about users' time allocation comes from the data recording user i's monthly time spending in the live broadcast room of streamer j, from which we can define which streamer the user i is most loyal to in month t.

[43]The evidence about users' attention allocation comes from the data recording user i's *Follow* behavior to streamer j. On the live streaming platform, if you like the digital content of a certain streamer and want to be alerted before her every live broadcast, then you can click the *Follow* button and become one of her fans, which can intuitively reflect the allocation of users' attention.

of streamers to measure their tier.[44] The smaller one streamer's ranking, the higher her tier and the greater her cumulative advantage. The downward curve shows that the active users with higher consumption levels will disproportionately pay their money, time, and attention to those higher-rank streamers that already have strong cumulative advantages. And Appendix Figure A18 shows that those who spend more than 100k RMB each month are almost exclusively loyal to the top 10% streamers.

Loyalty of users. Do users with high consumption levels have exceptionally high loyalty? We can focus on users' money and time allocation when analyzing loyalty. If a user spends money to buy a lot of virtual gifts for a particular streamer or spends a lot of time watching the digital content supplied by a streamer, it will indicate that the user is a loyal fan of the streamer. In Panel A of Figure 11, we construct four measures to proxy for the loyalty of users based on users' tipping behavior and money allocation — Herfindahl – Hirschman Index *HHI*[45], *Concentration*[46], the probability of users changing the loyal streamer *Shift* [47],

[44]This variable is constructed as follows: for each streamer j and month t, we sum up her tipping income in the month $(t-1)$. Then we rank streamers based on this past income to obtain their absolute rank $Rank_{j,t-1}$. Since we observe that the total number of streamers with positive income fluctuates somewhat, we also use $Percentile\,Rank_{j,t-1} = \frac{Rank_{j,t-1}}{Number\,of\,Streamers_{j,t-1}}$, which represents the relative position of streamer j in month t. We then calculate the mean of the rank of streamers the user i newly follows or is loyal to in the month t, which is labeled as $Percentile\,Rank_{i,t}$. The confidence interval plotted in Figure 10 is based on the Month – User level clustered standard errors.

[45]This variable is constructed as follows: Suppose on month t, the user i has given virtual gifts to a total of n streamers. We first calculate her value of virtual gifts to streamer j as $Tipping\,Amount_{i,j,t}$. Then we define Herfindahl – Hirschman Index $HHI_{i,t} = \sum_{j=1}^{n}(\frac{Tipping\,Amount_{i,j,t}}{\sum_{j=1}^{n} Tipping\,Amount_{i,j,t}})^2$. And we further calculate the mean for each user group. The confidence interval plotted in Figure 11 is based on the Month – User level clustered standard errors.

[46]This variable is constructed as follows: Suppose on month t, user i has given virtual gifts to a total of n streamers. We first calculate her value of virtual gifts to streamer j as $Tipping\,Amount_{i,j,t}$. Then we define $Concentration_{i,t}$ as the gift share of user i to her most loyal streamer, that is, $Concentration_{i,t} = \max_{1\leq j\leq n} \frac{Tipping\,Amount_{i,j,t}}{\sum_{j=1}^{n} Tipping\,Amount_{i,j,t}}$. And we further calculate the mean for each user group. The confidence interval plotted is calculated based on the Month – User level clustered standard errors.

[47]This variable is constructed as follows: Suppose user i is most loyal to streamer j on month $t-1$ by definition while the user i is most loyal to streamer q on month t. Then we define $Shift = 1$ if $j \neq q$. And we further calculate the mean for each user group. The confidence interval plotted is calculated based on the Month – User level clustered standard errors.

and the probability of users shifting toward higher-tier streamers Up[48]. The higher concentration or HHI value, the higher the users' loyalty to specific streamers. The lower the $Shift$ value, the higher users' loyalty to specific streamers. And the higher the probability Up, the more likely users would experience consumption upgrading to a higher-rank streamer. In Panel B, we also use the four measures mentioned above, and the only difference is that we define loyalty based on the amount of time users stay in streamers' live broadcast rooms.

From Figure 11 and Figure A19, we derive two main findings. First, HHI or Concentration, and consumption level exhibits a U-shaped relationship due to the dynamic competing process between two effects: when the consumption level of users is relatively low, the variety-seeking effect (McAlister and Pessemier, 1982; Kahn and Louie, 1990; Kahn, 1995; Ratner, Kahn, and Kahneman, 1999; Seetharaman and Che, 2009; Sevilla, Zhang, and Kahn, 2016) will be the dominating factor, in which individuals will explore rich digital contents when income rises and switch between different streamers within the choice set on the platform to pursue freshness, change, and diversity; however, if consumption level increases beyond a critical point, the brand-loyalty effect (Jacoby and Kyner, 1973; Bowen and Chen, 2001; Gefen, 2002; Kumar and Shah, 2004) will become the dominating factor, in which individuals tend to build solid ties with a particular streamer. Those active users who spend more than 1 million RMB a month have an HHI of over 0.7 and a Concentration over 80%. Second, with the increase in consumption level, the probability of users passing their affection to another streamer decreases. And even if a user has a new loyal streamer, it's more likely for the user to follow a higher-ranking super streamer. This finding indicates that both the money and time resources of active users with deep pockets are more inclined to be spent on top streamers.

User engagement. For the analysis of user engagement, we design several measures in

[48] This variable is constructed as follows: Suppose user i is most loyal to streamer j on month $t-1$ by definition while the user i is most loyal to streamer q on month t. And we can compare the month $t-1$ income ranking of streamer j and streamer q. Then we define $Up = 1$ if $j \neq q$ and $Rank_{q,t-1} < Rank_{j,t-1}$. And we further calculate the mean for each user group. The confidence interval plotted is calculated based on the Month – User level clustered standard errors.

Figure A20. Panel A and Panel B illustrate how active users are in exploring new digital content supplied by streamers, while Panel C focuses on a more direct dimension, users' viewing time. Precisely, the term *newly follow* is used here to refer to a user who adds some new streamers to the following list in the month. And the term *viewing time* is defined as the monthly sum of the duration between the user entering one live room and exiting the live room. The upward curves reveal that there has been a steady increase in engagement as the user spends more money on the platform. Users with high spending levels are also the main contributors to platform activity. Streamers must take advantage and focus on capturing the attention of and interacting with high-engagement users to serve and satisfy them. Unfortunately, considering the previous results, most of these users are also attracted to those top streamers.

To sum up, the results in this section provide suggestive evidence that those high-engagement users with deep pockets disproportionately pay money and be loyal to those superstar streamers, which indicates that the expansion of market size contributed by the spending intensity of existing users will extend inequality from the intensive margin.

4.2 Extensive Margin

We are also interested in how users willing to spend a lot of money on the platform will behave at different stages of their customer lifecycle. Criteria for selecting the subjects were as follows: first, we sorted users' aggregated tipping amount over the sample period to pick the top 10% and 1% users that the platform value a lot; second, to observe the entire lifecycle of users and control for bias, we only select the cohorts who registered between April 2018 and December 2018. And then for each month t, we divide those users into eight groups based on how long it has been since they registered — 1∼3 months, 4∼6 months, 7∼9 months, 10∼12, 13∼15 months, 16∼18 months, 19∼21 months, and 22∼24 months. Then we investigate the heterogeneity of users with varying registered months regarding their loyalty and preferred streamers' rank status. We find that new users are attracted to the top when they first enter the platform, and their loyalty will gradually increase as the registered months increase.

Rank level of streamers that users appreciate. The same method mentioned in Section 4.1 is used here to measure streamers' rank status that users are loyal to. In Panel A of Figure 12, the absolute rank of streamers refers to the average position of streamers that users who give the most significant gift share to, while in Panel B, the percentile rank of streamers refers to the average relative status of streamers that users provide the most significant gift share to. An interesting hump-shaped curve means that: newly registered users disproportionately spend their money with superstar streamers; then, they tend to explore more streamers in the first 12 months; after the first 12 months, existing users shift from lower-ranked streamers to higher-ranked streamers again, and start to build solid ties with superstar streamers. This means that in the expansion period of the platform (such as the first half of 2019), the market expansion brought about by the entry of numerous new users is also beneficial to the top streamers.

Loyalty of users. In Figure 13, we use *Concentration* and *HHI* to assess loyalty of users in different groups. It can be seen that there has been a gradual rise in loyalty as users mature from customers to repeat customers and from repeat customers to advocates. This indicates that after users are acquired by the platform, they will naturally build up their loyalty to the streamers.

Why do the more extensive the market size, the more beneficial the superstar streamers? The expansion of market size can be divided into two parts — the increase in the tipping intensity by existing users and the growth in the number of users. Active users with deep pockets would allocate their spending disproportionately toward superstars on an intensive margin. The higher the amount of tipping, the more engaged they are on the platform, and the more concentrated their spending is on the top streamers. Further, new viewers excessively pay attention to superstars on an extensive margin, and they will naturally build up their loyalty to the streamers.

5 Bigger Market, Larger Inequality

In this section, we provide the causal evidence that a more extensive market may increase inequality and estimate inequality elasticity to the market size. We study the following four quasi-experimental shocks: potential market size proxied by economic development and Fintech coverage, quarter-end revenue spikes induced by the seasonal incentive regime, user surge in the first half of 2019 before a significant capital raising event, and the Covid-19 lockdown in Wuhan.

5.1 Cross-City Estimation

This section shows that a broader market contributes to greater inequality using cross-city variations. And we use a series of variables, such as local GDP, population, and degree of digitization, to instrument the variations in market size. We construct city-level Gini coefficients and market size as the following: First, for each month, we take the prefecture-level city with the user's longest viewing time as that user's city assignment, and all tipping revenue is broken down to the city level according to the user's residence city. Second, we construct city-level Gini coefficients using the income distribution of all streamers in a specific city. [49] When we break down paying users at the prefecture level, many streamers earn very limited income in most cities, particularly small ones. Many zeros might mechanically lead to a higher Gini coefficient. To avoid systematic bias, our baseline Gini coefficient $Gini50$ is based on streamers who receive more than 50 RMB in tipping income from the city j in our entire sample period. For robustness, we also compute $Gini0$ and $Gini500$ based on streamers with positive income and income above 500 from city j. Third, we measure the local market size of city j, including total tipping amount, number of paying users, average tipping amount per user, and number of streamers with positive income from city j.

Figure 9 provides naive cross-city correlations that a broader market positively correlates with greater inequality. From Panel A to D, we correlate Gini with market size metrics one

[49] We also replicate our Gini – size relationship with the Gini coefficients calculated based on the distribution of streamers' number of loyal fans. This robustness check indicates that our results are not driven by outliers.

by one: total tipping amount, the total number of paying users, the tipping amount per user, and the total number of streamers who receive any tipping income in a city during the sample period, respectively. Our baseline city sample consists of 92 cities in which the number of paying users accounts for more than 0.2% of the total number of paying users. We see a robust linear relationship between the Gini and market size: the Gini increases from 0.85 in small cities to 0.95 in the largest cities.

Next, we use one city's population, GDP, and digital financial inclusion index, proxies of the potential market scale, as instrumental variables for the platform's actual market size. Our identification assumption is that the population or development level (GDP and digital financial inclusion index) causes variation in the local market size, but it does not directly shape income inequality on the platform, i.e., no reason why city size affects users' preference over different streamers. Appendix Figure A21 presents the correlation between the city size and local market size, and we can observe that the scale of platform use (i.e., log value of tipping amount) in one city is strongly associated with both its population and development level.

Table 2 reports the cross-city ordinary least squares (OLS) and instrumental variable (IV) regression results. We run eq.(2) in the OLS regression. City-level market size $MarketSize_j$ is the log value of the total virtual gifts sent by users in Panel A and the log number of paying users in Panel B.

$$Gini_j = \alpha + \beta_{OLS} MarketSize_j + \epsilon_j. \tag{2}$$

In the IV regression, in Columns (2) and (3), we use the city's residential population and local GDP separately as an IV of the market size variable. In Columns (4) – (7), we use digital financial inclusion to instrument the market size. The city sample includes 92 cities whose total number of paying users accounts for more than 0.2% of the total number of paying users. The F-statistics suggest that population, local GDP, and digital financial inclusion are unlikely to be weak instrumental variables. Furthermore, these cross-city variations yield consistent estimates of approximately 0.019 to 0.095 regardless of the definition of the Gini index — when the market size doubles, the Gini coefficient increases by 0.019 to 0.095. We further limit our city sample to 18 cities[50] whose total number of paying users accounts for more than 1% of the total number of paying users. Though the F statistics are relatively small, which may be mainly due to the small sample size, we find that the values of the coefficients are fairly consistent.

5.2 Market Size Seasonality

The live-streaming platform designs an incentive scheme for streamers to race for performance at the end of each quarter (March, June, September, and December), for example, posting the ranking list of streamers by virtual gifts. Top-ranked streamers can gain more exposure to the public and benefit from their popularity on the live-streaming platform.[51] In this section, we use quarter-end dummies as time-series instrumental variables to identify the effect of market size on inequality.

Appendix Figure A2 Panel A shows that the total tipping amount spikes in the quarter-end months. In Panels B and C, we decompose the total tipping amount into the number of paying users and the tipping amount per user. The seasonality mainly comes from existing users spending more on the live-streaming platform. Similarly, in Appendix Figure A3, we break down the total tipping amount by streamers and find no evidence that more streamers were active at quarter ends. Thus, streamers earn more income from existing paying users every quarter end, suggesting the platform's incentive mechanically leads to market expansion.

[50] The 18 cities are Shanghai, Dongguan, Beijing, Nanjing, Hefei, Tianjin, Ningbo, Guangzhou, Chengdu, Wuxi, Hangzhou, Wuhan, Shenzhen, Suzhou, Xi'an, Zhengzhou, Chongqing and Changsha.

[51] Much anecdotal evidence supports this argument. For example, streamers' rankings are associated with their profit-sharing ratio with multichannel network (MCN) agencies. More popular streamers are more likely to earn income from marketing businesses or gain brand endorsements.

5.2.1 Time-series Estimation

We first test whether Gini coefficients also spike at the end of each quarter (March, June, September, and December) in the monthly aggregate Gini coefficients — 25 months from April 2018 to April 2020. Specifically, we run the following regression:

$$Gini_t = \alpha + \beta QuarterEnd_t + \eta_t + \epsilon_t, \tag{3}$$

where $Gini_t$ is the Gini coefficient; $QuarterEnd_t$ is a dummy variable that equals one if the month is March, June, September, or December, and zero otherwise; and η_t represents year-quarter fixed effects. In Columns (1) and (4), $Gini_t$ is the $Gini0_t$ calculated using the full sample with all streamers who receive any income in the month t. In Columns (2) and (5), $Gini_t$ is $Gini50_t$, which is the Gini coefficient of the sub-sample of streamers whose income is above 50 RMB in month t. In Columns (3) and (6), $Gini_t$ is $Gini500_t$, which is the Gini coefficient of the sub-sample of streamers whose income is above 500 RMB.

In Appendix Table A4, we find that the Gini coefficient significantly increases at the end of each quarter. In Column (1), β is 0.014 (*s.e.*=0.005) and significant at the 1% level, which implies an approximately 1.48% increase from the baseline Gini coefficient[52]. After adding year-quarter dummies, the β_t increases to 0.015 (*s.e.*=0.003) with stronger statistical power in Column (4). We also use different measures of the Gini coefficient. We use $Gini50_t$ of streamers with income above 50 RMB in Columns (2) and (5). In Column (2), the β is 0.027 (*s.e.*=0.007), which implies an approximately 3.00% increase from the baseline Gini coefficient. After adding year-quarter fixed effects in Column (5), the β is 0.029(*s.e.*=0.006) and still significantly positive. We use $Gini500_t$ of streamers with income above 500 RMB in Columns (3) and (6): β is 0.041 (*s.e.*=0.011), which implies an approximately 4.82% increase from the baseline Gini coefficient, and β rises to 0.044 (*s.e.*=0.008) after controlling for year-quarter dummies. Thus, the empirical results show that the magnitude of inequal-

[52] 1.48% equals 0.014 ÷0.948.

ity among streamers, especially among the top streamers, significantly increases at the end of each quarter.

Next, we offer the two-stage least squares (2SLS) estimates for the inequality elasticity to the market size. In the first stage, we formalize the seasonal market expansion by running the following regressions:

$$MarketSize_t = \alpha + \beta QuarterEnd_t + \eta_t + \epsilon_t, \quad (4)$$

$MarketSize_t$ represents four platform size metrics, as reported in Appendix Table A5. As shown in Columns (1) and (3), the total tipping amount is 37.4% (*s.e.*=8.1%) higher, and the tipping amount per user is 33.6%(*s.e.*=6.1%) higher in quarter-end months, both of which are significant at the 1% level. Unlike cross-sectional variation, we find little difference in paying users and the number of streamers, only 3.8% (*s.e.*=7.5%) and 0.4%(*s.e.*=3.9%) higher, respectively. Thus, we only focus on the total tipping amount (monetary value of virtual gifts) to measure market size.

Table 3 reports both the OLS and IV results of the market size impact on Gini coefficients. We run the following regressions:

$$Gini_t = \alpha + \beta_{OLS} MarketSize_t + \eta_t + \epsilon_t, \quad (5)$$

$$Gini_t = \alpha + \beta_{IV} \widehat{MarketSize}_t + \eta_t + \epsilon_t, \quad (6)$$

where $Gini_t$ is the Gini coefficient calculated using the income distribution of streamers in month t. Panel A reports the OLS results estimated in eq.(5). Panel B reports the 2SLS results estimated in eq.(6). In Columns (1) – (3), $MarketSize_t$ is $\ln(Tipping\ Amount_t)$, which is the log value of the total tipping income of all streamers in month t. In Columns (4) and (6), $MarketSize_t$ is $\ln(Tipping\ Amount\_PC_t)$, which is the log value of the average tipping income per streamer in month t. In Columns (1) and (4), the dependent variable is $Gini0$, calculated using the income distribution of the streamers who earn any positive income in month t. In Columns (2) and (5), $Gini_t$ is $Gini50$, which is calculated using the income distribution of the streamers who receive more than 50 RMB in tipping income in month t. In Columns (3) and (6), $Gini500_{jt}$ is calculated based on the income distribution of the streamers who receive more than 500 RMB in tipping income in month t. Standard errors are robust and reported in parentheses. The F-statistics are reported in the last row in Panel B. The sample includes the 25 months from April 2018 to April 2020.

We find that market size has a significantly positive effect on the inequality magnitude. We use $\ln(Tipping\ Amount_t)$ as the proxy variable of the market size in Columns (1) – (3). The β_{IV} ranges from 0.041 to 0.117 and is significant at the 1% level. When the total tipping amount increases by 10%, the Gini coefficient increases by approximately 0.41%[53] to 1.31%[54]. β_{IV} is larger than the OLS estimate reported in Panel A. Moreover, the F-statistic is 21.091, which suggests that $QuarterEnd$ is unlikely to be a weak instrument for $\ln(Tipping\ Amount_t)$ (Kleibergen and Paap, 2006). We use $\ln(Tipping\ Amount\_PC_t)$ as a proxy variable of the market size in Columns (4) – (6), β_{IV} ranges from 0.046 to 0.130 and is significant at the 1% level. When the average tip income per streamer increases by 10%, the Gini coefficient increases by approximately 0.46%[55] to 1.46%[56]. In addition, the K.P. F-statistic is 30.694, which suggests that $QuarterEnd$ is unlikely to be a weak instrument for $\ln(TippingAmount\_PC_t)$.

5.2.2 Panel Estimation

In this section, we use city-month panel observations to estimate the effect of market size on inequality. Our baseline city sample includes 18 cities with more than 1% of the total paying users. In the Appendix, we show parallel results with 92 cities that account

[53] 0.41% equals $0.041 \times \ln 1.1 \div 0.948$.
[54] 1.31% equals $0.117 \times \ln 1.1 \div 0.851$.
[55] 0.46% equals $0.046 \times \ln 1.1 \div 0.948$.
[56] 1.46% equals $0.130 \times \ln 1.1 \div 0.851$.

for more than 0.2% of the total paying users and 302 cities with at least one streamer who earns virtual gifts worth more than 500 RMB from that city every month.

We plot the stylized relationship between the log of the tipping amount and the Gini coefficients with data of all 302 cities in Appendix Figure A22. The slope becomes steeper (from 0.016 to 0.098) as we pick higher-income bars for streamers, as shown from Panel A to C. Furthermore, we show the density distribution of time-series correlations between market size variables (total tipping amount, number of paying users, tipping amount per user, number of streamers) and Gini coefficients in Appendix Figure A23. As an example, we pick the city sample with more than 1% of total paying users (the left panel). The average of $Corr(Gini, Tipping\,Amount)$ and $Corr(Gini, Tipping\,Amount\_PC)$ are 89% and 87%, and are significantly positive at the 1% level, whereas the averages of $Corr(Gini,\ Number\,of\,paying\,Users)$ and $Corr(Gini,\ Number\,of\,Streamers)$ are much weaker, 11% and -15% respectively. The pattern holds when we move to larger city samples. Overall, in time series per se, the dynamic inequality is more associated with how much each user spends on the platform; that is, the rising inequality is attributed to the fact that each fan (paying user) contributes more money to her idol (streamer).

Then, we formally estimate the panel data's inequality elasticity to market size. The first stage is estimated as follows:

$$MarketSize_{j,t} = \alpha + \beta QuarterEnd_t + \theta_j + \eta_t + \epsilon_{j,t}, \tag{7}$$

The only innovation is that we include in city fixed effects θ_j in all panel regressions. Appendix Table A6 shows the first-stage results. We find that the total tipping amount increases by 36.0% ($s.e.$ = 7.3%) at the end of each quarter, and the tipping amount per user increases by 28.9%($s.e.$ = 5.5%) at the end of each quarter in Columns (1) and (3), both of which are significant at the 1% level. The magnitude is quite comparable to the estimates we obtained from time-series regressions. Similarly, the numbers of paying users and streamers only modestly increase at the end of each quarter, 7.0% (*s.e.* = 6.1%) and 2.8% (*s.e.* = 3.1%) without statistical significance in Columns (2) and (4). Thus, we only use $\ln(Tipping\,Amount_{j,t})$ to measure the market size in IV regressions.[57]

Alternatively, Table 4 reports the panel estimation results of OLS and IV regression with 18 cities included in the sample. We run the following regression:

$$Gini_{j,t} = \alpha + \beta_{OLS} \ln(Tipping\,Amount_{j,t}) + \theta_j + \eta_t + \epsilon_{j,t}, \tag{8}$$

$$Gini_{j,t} = \alpha + \beta_{IV} \ln(\widehat{Tipping\,Amount}_{j,t}) + \theta_j + \eta_t + \epsilon_{j,t}, \tag{9}$$

$Gini_{j,t}$ is the Gini coefficient calculated using the income distribution of streamers who receive virtual gifts from city j in month t; $\ln(Tipping\,Amount_{j,t})$ is the log total tipping amount by paying users of city j in month t, and $\ln(\widehat{Tipping\,Amount}_{j,t})$ is the predicted value from the first stage from $QuarterEnd_t$. Columns (1), (2), and (3) report the OLS results estimated in eq. (8). Columns (4), (5), and (6) report the 2SLS results estimated in eq. (9).

IV and OLS also yield similar elasticity estimates. When market size doubles, *Gini*0 is predicted to rise by 3.6% (*s.e.* = 0.3%) with OLS and 3.7% (*s.e.* = 0.3%) with IV in Columns (1) and (4); *Gini*50 increases by approximately 8.1% (*s.e.* = 0.6%) with OLS and 7.9% (*s.e.* = 0.8%) with IV in Columns (2) and (5); *Gini*500 rises by approximately 13.5% (*s.e.* = 1.0%) with OLS and 12.7% (*s.e.* = 1.1%) with IV in Columns (3) and (6).[58] These elasticity parameters estimated from panel regressions are quite similar to ones estimated in Section 5.2.1. Furthermore, we estimate elasticity with extended city samples in Appendix Tables A10 and A11: the coefficients are 3.6% (*s.e.* = 0.4%) for *Gini*0, 8.5% (*s.e.* = 0.6%) for *Gini*50, and 12.6% (*s.e.* = 0.9%) for *Gini*500 in 92 cities; 3.6% (*s.e.* = 0.9%) for *Gini*0,

[57] Robustness results are reported in Appendix Tables A7 and A8. The regression samples include 92 cities and 302 cities, respectively. The variables' detailed definitions and summary statistics are presented in Appendix Table A9.

[58] Moreover, the F-statistic is 24.544, which suggests that *QuarterEnd* is very unlikely to be a weak instrument for $\ln(Tipping\,Amount_{j,t})$.

9.8% (*s.e.* = 0.7%) for *Gini*50, and 14.4% (*s.e.* = 0.9%) for *Gini*500 in 302 cities.[59]

5.3 User Surge before a Capital Raising Event

The live-streaming platform is planned to embrace an important capital-raising event, and the prepared timeline is as follows. It officially launched the capital raising in January 2019 and finally established the partnership with its investors in July 2019. It should be noted that the platform was still losing money due to intense competition in the industry until the end of 2018. Therefore, a sequence of marketing endeavors was made for market growth, which finally contributed to the improvement in financial performance before its capital raise.

We conclude that the endeavors of the platform to attract users and expand the market could be mainly divided into three categories after collecting all available news related to the platform during our sample period and consulting the company insiders. First, fan festivals, carnivals, and other similar grand ceremonies were held frequently, aiming to harness Internet celebrities' influence in live streaming to drive existing fans to tip more and attract new users to enter. Second, it purchased streaming rights or even exclusive streaming rights of many heavyweight international e-sports games to encourage users who were keen on games to start using the platform, which is a large group. For example, the total viewership of League of Legends Worlds was 100 million, which made per-minute average views of 21.8 million in 2019. [60] Third, the platform increased positive news exposure by taking the lead in setting industry norms and proactively embracing the regulation.

[59] Alternatively, we can estimate time-series elasticity for each city j, and compare OLS and IV estimators in Appendix Figure A24. The mean of IV estimators is very close to the mean of OLS estimators, although the IV estimator has a wider distribution (less precise than OLS estimators). In Panel A, the mean of the OLS estimator using the dependent variable *Gini* is 0.039, while the mean of the IV estimator using the dependent variable *Gini* is 0.034, both of which are significantly positive at the 1% level. In Panel B, the mean of the OLS estimator using the dependent variable *Gini*50 is 0.115, whereas the mean of the IV estimator using the dependent variable *Gini*50 is 0.109, both of which are significantly positive at the 1% level. In Panel C, the mean of the OLS estimator using the dependent variable *Gini*500 is 0.177, whereas the mean of the IV estimator using the dependent variable *Gini*500 is 0.154, both of which are significantly positive at the 1% level.

[60] https://lol-eloboosting.com/blog/lol/misc/league-of-legends-the-origin-story.

In Figure A25, we plot the average log value of the tipping amount contributed by *new users* each month.[61] To exclude the interference of the Covid-19 pandemic, we first delete the months after January 2020. Then, we divide the remaining months into three-time stages based on the nodes of the launch of capital raising (i.e., 2019m1) and its closing (i.e., 2019m7), which are marked by vertical lines in this figure. Below, we will refer to the period before 2019m1 as *Stage 1*, the period between 2019m1 and 2019m7 as *Stage 2*, and the period post 2019m7 as *Stage 3*. We use three metrics to define whether a user is new in month t. First, if a newly registered user in month t converts to a paying user that month, then the user will be defined as new in month t, which is tagged as *New Registered*. Second, if a user makes a tip for the first time in month t during our sample period, then the user will be defined as a new user in month t, which is tagged as *First Appeared*. According to this definition, all users in April 2018 will be new; thus, the value of this variable in April 2018 is defined as missing. Third, if one has tipping records in month t, but does not tip during period $[t-3, t-1]$, then the user will be defined as new in month t, which is tagged as *Return*. The values of this variable before June 2018 are illustrated as missing being more comparable.

Figure A25 provides intuitive evidence that the platform's efforts were paying off —- revenue from new users was significantly higher in the first half of 2019 in Stage 2, compared to performance in Stage 1 and Stage 3. The message from Table A12 is consistent. Specifically, we run the following regressions:

$$ln(Tipping\,Amount\_New\,Users_{j,t}) = \alpha + \beta_1\, \mathbb{1}(Stage\,2)_t + \beta_2\, \mathbb{1}(Stage\,3)_t + \epsilon_{j,t}, \tag{10}$$

where $ln(Tipping\,Amount\_New\,Users_{j,t})$ is the tipping amount contributed by new users in city j in month t; $\mathbb{1}(Stage\,2)_t$ and $\mathbb{1}(Stage\,3)_t$ are dummy variables that mark whether the month t is in Stage 2 or Stage 3. From Column (1) to Column (3), we use the tipping

[61] For each month, we calculate the average value based on the 92 cities.

amount contributed by New Registered, First Appeared, and Return users as the dependent variable. On average, new users contributed significantly more revenue in Stage 2 than in Stage 1. The growth size depends on how we define new users and the city sample used in regressions. For example, if we limit the sample to 18 cities (> 2% users), the newly registered users will increase by more than 50% in Stage 2 compared to the benchmark Stage 1. And the newly registered users will increase by 13.7% if we use a broader 302 cities.

Next, we plan to use the tipping amount change contributed by new users to instrument the market size variations. First, in Table A13, we find that the size of new-registered users can explain the surge and slowdown in the entire market size, which indicates that the instrumental variable — new users, has a strong correlation relationship with the endogenous variable — market size. On the other hand, since the magnitude of new users is mainly motivated by the platform's capital-raising incentive, it is unlikely that it will directly affect users' preference for streamers and affect the inequality among streamers.

Specifically, in Table A13, we run the following regressions:

$$\Delta\, ln(Tipping\, Amount)_{j,stage} = \alpha + \beta \Delta\, ln(Tipping\, Amount\_New\, Users)_{j,stage} + \epsilon_j, \tag{11}$$

where $\Delta\, ln(Tipping\, Amount)$ is the difference in the average monthly size of the tipping amount between two stages and $\Delta\, ln(Tipping\, Amount\_New\, Users)$ is the difference in the average monthly size of the tipping amount contributed by new users. This table shows that the change in tipping amount from new users is significantly positively correlated with the change in total tipping amount, regardless of the city sample we use and the method of defining new users. In other words, new users can partly explain the rise and down in the whole market size around the capital-raising event.

Table 5 reports the 2SLS estimation results using the growth of new users during the capital-raising period to instrument market growth. We run the regressions:

$$\Delta\, Gini_{j,stage} = \alpha + \beta\Delta\, \ln(Mark\hat{e}t\, Size_{j,stage}) + \epsilon_j, \tag{12}$$

where $\Delta\, Gini$ is the difference in the average monthly Gini coefficients of the city j between two stages, and $\Delta\, \ln(Market\, Size)$ is the difference in the average monthly size of the tipping amount or the number of paying users between the two stages. In Panel A, $Gini50$ is used to calculate $\Delta\, Gini$, based on the income distribution of streamers whose incomes are more than 50 RMB from city j. $\Delta\, \ln(Tipping\, Amount\_New\, Users)$ is used to instrument the variation of whole market size $\Delta\, \ln(Tipping\, Amount)$. In Panel B, $Gini\_fans$ is used to calculate $\Delta\, Gini$, which is calculated based on the distribution of streamers' loyal fans from city j. $\Delta\, \ln(Number\, of\, New\, Users)$ is used to instrument the variation of whole market size $\Delta\, \ln(Number\, of\, Paying\, Users)$. In this main result table, we use the *New Registered* method to define *New Users*. In Columns (1), (2), and (3), We use the value difference of the variables between *Stage 2* and *Stage 1*. And in Columns (4), (5), and (6), We use the value difference of the variables between *Stage 3* and *Stage 2*. And we also use different city samples.

The results show that the expansion and contraction of market size caused by the capital-raising shock can explain the corresponding change in inequality degree. Panel B shows that our result is not simply due to the outlier paying users who spend a lot and race to be the most outstanding paying users but that fans are more concentrated due to the exogenous market expansion. In the robustness test Table A14 and Table A15, we changed the calculation methods of the dependent variable and instrumental variable, and the results were robust and consistent.

5.4 Covid-19 and the Wuhan Lockdown

5.4.1 Background

Wuhan's Covid-19 outbreak attracted massive public attention in China and crowded out the demand for entertainment as pandemic-related news unfolded daily. In this section, we implement an event study to estimate the market size shrinkage induced by the sudden Covid-19 outbreak and the change in income inequality in response to the unexpected pandemic shock.

Three significant event dates are relevant to the Covid-19 stock. First, Dr. Wenliang Li posted a coronavirus alert on December 30, 2019, marking the onset of Covid-19. Second, Covid-19's damage became publicly known and triggered social disruption when Wuhan locked down on January 23, 2020, and transportation was cut off from the rest of China. Third, beginning on February 1, 2020, the government implemented a series of escalated measures in response to the aggravation of the epidemic. On February 1, 2020, the Hubei provincial government announced the extension of the Spring Festival holiday as more time is needed to contain the virus. On February 2, 2020, *Huoshenshan Hospital*, the first massive-scale quarantine facility, was officially put into operation, and the Hubei government announced that all suspected Covid-19 cases were commanded for mandatory isolation.

5.4.2 Covid-Shock to Entertainment Demand

First, we evaluate the live-streaming market size change in response to Covid-19 in a time window from 50 days before Wuhan lockdown to 50 days after the lockdown. The pre-treatment period is from December 4 (Day -50) to December 29 (Day -23). From Day -24 to Day 0, people began to learn about Covid-19 while the government investigated what measures would be needed to contain the virus. During this interim period, Covid-19 started to crowd out entertainment demand, particularly in Wuhan, as a large amount of Covid-19 news, fake news, and rumors began to spread on social media. On Day 0, Wuhan announced the unprecedented lockdown measure — nobody was allowed to leave Wuhan, and the restriction was not eased till Day 28.[62]

We present evidence the Covid-19 outbreak distracted users from live-streaming entertainment, and the size of the Wuhan market dramatically decreased. According to Appendix Figure A27, our preliminary analysis shows that the market size, measured as total tipping amount, the number of paying users, and the number of streamers, decreased the most in the epicenter city of Wuhan, followed by the nearby city of Changsha, whereas the distant city of Chengdu was nearly unaffected.

Figure 14 shows the discontinuity in entertainment demand in Wuhan around Day 9. From that day on, the government took more aggressive measures to flatten the curve and cause widespread panic in Wuhan. Figure 14 shows that the number of paying users, the log number of tipping amount, and the number of streamers who received any tip income, all of which are indicators of market size, experienced a dramatic drop in Wuhan. Our regression discontinuity implies that the number of paying users dropped by one-third, and the daily tipping amount also dropped by half; consequently, the number of streamers with positive income dropped by about 60%. We observe no discontinuity in Chengdu and Shenzhen, more distant cities from Changsha and also less hit by Covid-19.

To find the best counterfactual for Wuhan, we focus on nine cities with more than 2% paying users: Shanghai, Beijing, Guangzhou, Chengdu, Hangzhou, Wuhan, Shenzhen, Chongqing, and Changsha. After the Wuhan government started implementing a series of escalating measures, the number of paying users, the total tipping amount, and the number of streamers who received any virtual gifts in Wuhan decreased 49.8%, 70.4%, and 39.8%[63] respectively.

[62]According to Appendix Figure A26, Wuhan citizens' attention to the Covid-19 epidemic, as measured by the Baidu search index, increased sharply after the Wuhan lockdown (Day 0) and remained at a high level since February 1, 2020 (Day 9).

[63]We get these values from β in the regression $Y_{j,t} = \alpha + \beta Treatj \times Post_{3,t} + \lambda Post_{3,t} + \theta_j + \epsilon_{j,t}$, where $Treatj = 1$ if the city is Wuhan and $Treatj = 0$ otherwise. $Post_{3,t} = 1$ indicates days after February 1, 2020 (Day 9), the post period in our regression discontinuity plot.

5.4.3 Event-based Gini Coefficient Response

We construct the daily-level Gini coefficient $Gini50_{j,t}$ using the income distribution of streamers who receive virtual gifts worth more than 50 RMB from users in the city j in the past seven days (from Day $(t-6)$ to Day t). Lagging income for seven days enables us to minimize the impact of some outlier transactions and obtain a relatively stable inequality measure. As we study high-frequency inequality, many streamers might receive little virtual gifts or even not perform on their live-streaming channel. To avoid many zeros, we only include streamers whose income has been above 50 RMB in the past seven days in the Gini calculation.[64]

Appendix Figure A28 provides evidence that market size is positively related to inequality using the city-daily level observations around the Wuhan lockdown. In Panels A, B, and C, we correlate daily Gini with the log number of paying users, the log tipping amount, and the log number of streamers, respectively, in 79 days from 50 days before the Wuhan lockdown to 4 weeks after the Wuhan lockdown (Day -50 to Day 28). The high-frequency evidence in Panels A and B indicates that the Gini coefficient rises by 5.0% (*s.e.*=0.5%) as the tipping amount doubles, rises by 7.1% (*s.e.*=0.5%) as the number of paying users doubles, controlling for city fixed effects.

Then, we exploit the regression discontinuity in Wuhan to provide causal estimates. Table 6 shows the magnitude of change in Wuhan's Gini coefficient and market size around Day 9, which refers to the day February 1 when the government implemented a series of escalated measures to deal with the Covid-19 epidemic as the aggravation of the epidemic. The Gini coefficient and market size in Wuhan are adjusted by that of its faraway cities, Chengdu, Shenzhen, Beijing, and Shanghai, to eliminate the interference of confounding factors. As this table demonstrated, compared with the benchmark cities, both the platform use (tipping amount or the number of paying users) and the inequality ($Gini50$ or $Gini\_fans$

[64] Appendix Table A17 presents the detailed definitions and summary statistics of the variables at the daily level of nine cities with more than 2% of paying users: Shanghai, Beijing, Guangzhou, Chengdu, Hangzhou, Wuhan, Shenzhen, Chongqing, and Changsha.

[65]) decreased more in Wuhan after the shock. And the Gini elasticity can be obtained by computing the *after minus before* Gini coefficient change to the Covid-19 market size shrinkage. Our event-based estimates range from 0.023 to 0.056.

5.4.4 Difference-in-Difference: Distance to Wuhan

An alternative way to evaluate Covid-19 responses is to exploit the geographical distance to Wuhan with the hypothesis that the pandemic hit users in Wuhan and nearby cities (e.g., Changsha) harder than users in more distant cities (e.g., Chengdu or Shenzhen). We experiment with three post-event dummies: $Post_{1,t} = 1$ indicates the days after December 30, 2019 (Day -24) when Dr. Wenliang Li posted a Covid-19 alert on one of his WeChat groups, and Covid-19 first became known to the public. $Post_{2,t} = 1$ indicates the days after January 23, 2020 (Day 0), when the government imposed a lockdown in Wuhan. $Post_{3,t} = 1$ indicates the days after February 1, 2020 (Day 9), when the government started implementing escalating measures.

Table A16 reports the pooled regression results of the use of live streaming in response to the Covid-19 shock. We find that the demand for live streaming was less affected in cities further from Wuhan. We run the following regression:

$$MarketSize_{j,t} = \alpha + \beta ln(Distance_j)^{66} \times Post_t + \lambda Post_t + \theta_j + \epsilon_{j,t}. \tag{13}$$

$MarketSize_{j,t}$ is the log number of viewers in city j in day t in Columns (1) and (4), the log number of paying users in the city j in day t in Columns (2) and (5), the log number of streamers in city j in day t in Columns (3) and (6). In Columns (1), (2), and (3), $Distance_j$ is the driving time (in days) between Wuhan and city j collected from Baidu Maps. In Columns (4), (5), and (6), $Distance_j$ is the straight-line distance (in a thousand kilometers)

[65] We construct the daily-level Gini coefficient $Gini\_fans$ using the distribution of streamers' number of loyal fans in city j on day t. And we determine the loyal streamer of users on day t according to the distribution of the user's tipping expenditure among various streamers in the past seven days.

[66] This variable is actually the log value of $(1 + Distance_j)$.

calculated from the latitude and longitude of Wuhan and city j.

Suppose the Covid-19 shock caused a decrease in entertainment demand in the epicenter cities. In that case, we can hypothesize that as a city's distance from Wuhan increases, the decline in entertainment demand weakens. The results in Appendix Table A16 show that as the city's distance from Wuhan increases, the changes in both the number of participants and the total amount of virtual gifts following the shocks weaken.

To identify the causal effect, we estimate the following difference-in-differences (DID) regression and show the estimated results in Tables 7 and A18.

$$Gini_{j,t} = \alpha + \beta ln(Distance_j) \times Post_t + \lambda Post_t + \delta MarketSize_{j,t} + \theta_j + \epsilon_{j,t}. \tag{14}$$

In Table 7[67] In columns (1) and (6), we do not control any of the market size variables. Section 5.4.2 has already shown evidence that Wuhan and its nearby cities experienced a dramatic reduction in entertainment demand during the epidemic. In Columns (1) and (6) of Table 7, the βs before the interaction term in eq.(14) are all positive, which further illustrates that Gini coefficients in Wuhan and its nearby cities drop more than ones in other distant cities after Covid-19 outbreak. Specifically, the coefficient before the interaction term is 0.060 (*s.e.*=0.047) and 0.033 (*s.e.*=0.029) in Panel A; the coefficient is 0.097 (*s.e.*=0.040) and 0.062 (*s.e.*=0.024) in Panel B; coefficient is 0.092 (*s.e.*=0.044), and 0.062 (*s.e.*=0.028) in Panel C. Both Wuhan lockdown $Post_2$ and escalated restriction $Post_3$ dummies yield significant and consistent parameters. These results support that market shrinkage driven by the Covid-19 outbreak has contributed to a smaller inequality.

In Columns (2) and (7), we use the log number of paying users as the measure of the market size variable $MarketSize_{j,t}$. In Columns (3) and (8), we use the log value of the tipping amount as the measure of the market size variable $MarketSize_{j,t}$. In Columns (4) and (9), we use the log number of streamers as the measure of the market size variable

[67] In Appendix Table A18, the dependent variable is $Gini0_{j,t}$, which is calculated using income distribution of all streamers who receive any virtual gifts from citizens in the city j from Day $(t-6)$ to Day t.

$MarketSize_{j,t}$. In Columns (5) and (10), we add all three market size variables to the regression. In Panel B, the coefficients drop from 0.097 in Column (1) to 0.047(*s.e.*=0.031) in Column (6); from 0.062 in Column (6) to 0.031(*s.e.*=0.019) in Column (10). In Panel C, market size variables can even fully explain the coefficients before the interaction terms: from 0.092 in Column (1) to -0.012(*s.e.*=0.060) in Column (6); from 0.062 in Column (6) to -0.001(*s.e.*=0.039) in Column (10).[68]

Figure 15 plots the dynamic treatment effects from the day of Wenliang Li's post (Day -24) to the Wuhan lockdown (Day 0) and the four weeks after the Wuhan lockdown. For any fixed $T > -24$ and $t \in [-50, T]$, we run the following regression:

$$Gini_{j,t} = \alpha_T + \beta_T ln(Distance_j) \times Post_{1t} + \gamma_T Post_{1t} + \theta_j + \epsilon_{j,t}. \tag{15}$$

In Panel A, $Gini50_{j,t}$ is calculated using the income distribution of streamers who receive virtual gifts worth more than 50 RMB from citizens in the city j from Day $(t-6)$ to Day t. In Panel B, $Gini0_{j,t}$ is calculated using the income distribution of all streamers who receive any virtual gifts from users in the city j from Day $(t-6)$ to Day t. $Distance_j$ is the driving time (in days) between Wuhan and city j, as shown on Baidu Maps. $Post_{1t} = 1$ indicates the days after December 30, 2019 (Day -24) when Dr. Wenliang Li posted a Covid-19 alert on one of his WeChat groups, and Covid-19 became known to the public. θ_j represents city-level fixed effect.

Using the observations of the nine cities from Day -50 to Day T, we run the above regression and get the estimated β_T for $\forall T > -24$. Then we plot β_T before the interaction term as a function of time. Overall, we find no significant treatment effects before Day -5. And β starts to drift up beginning on Day -5, indicating that the Gini coefficients start to decline more in Wuhan and nearby cities, where there was more panic than in other cities. This effect tends to persistently drift until Day 9 and flatten out until Day 28.

[68] In Panel A, the coefficients shrink by two-thirds, although no coefficient is statistically significant.

Appendix Figure A29 presents the different treatment effects when adding market size variables. Previously, we estimate β_T for any fixed $T > -24$ and $t \in [-50, T]$ using eq.(15). We further estimate the new regressions using eq.(16) in this figure.

$$Gini0_{j,t} = \alpha_{new,T} + \beta_{new,T} ln(Distance_j) \times Post_{1t} + \lambda_{new,T} Post_{1t} + \delta_{new,T} MarketSize_{j,t} + \theta_j + \epsilon_{j,t} \quad (16)$$

In Figure A29 of Panel A, the log number of paying users is used to measure $MarketSize_{j,t}$, whereas, in Panel B, the log tipping amount is used to measure $MarketSize_{j,t}$. Next, we plot both of the β before the interaction term as a function of time. The solid line refers to the value of $\beta_{raw,T}$, and the dashed line refers to the value of $\beta_{new,T}$. As this figure demonstrates, the dashed line is lower than the solid line, especially after Day 9, which means that the exogenous changes in the number of paying users and tipping amount induced by the Covid-19 shock can explain part of the changes in inequality among streamers.

We extend our analysis to PL exponents of income inequality in the upper tail. Although the results using PL exponents as the dependent variable are noisier, the overall results are robust and consistent. The reduced entertainment demand in Wuhan and its surrounding cities induced by the epidemic shock caused a decrease in the magnitude of inequality in these cities, providing causal evidence that a broader market is an essential determinant of more considerable inequality. The detailed results of this analysis are presented in Appendix Table A19 and Appendix Figure A30.

6 Conclusion

Equal access does not guarantee equal outcomes. Using proprietary data, we show that income inequality on a leading Chinese live-streaming platform is larger than any known income and wealth distribution in the offline world. The Internet does not necessarily promote equality; instead, it is a "winner-takes-all" market with unprecedented inequality. Active users with deep pockets pay disproportionately toward the superstar streamers. The top streamers amplify their influence through the Internet community, and the winners obtain a larger and larger market share, whereas most streamers merely gain a little attention from their audience. More disturbing is the fact that inequality has increased since 2018, and the top 100 streamers' income share increased from 40% to 70% in two years.

We hypothesize that expanding market size is the force driving this rising inequality. To address endogeneity concerns, we estimate the Gini – market size elasticity by exploiting three types of quasi-experimental variation: population size in a city (0.019 - 0.106), quarter-end shocks to tipping revenue (0.037 - 0.127), user surge induced by the capital raising event (0.065 - 0.255), and the exogenous Covid-19 lockdown in Wuhan (0.023 – 0.056). The rapid transition toward a digital economy benefits only a small portion of influencers and exacerbates rising income inequality worldwide.

References

Adler, Moshe. 1985. "Stardom and talent." *American Economic Review* 75, 208–212.

Adler, Moshe. 2006. "Stardom and talent." *Handbook of the Economics of Art and Culture* 1, 895–906.

Alvaredo, Facundo, Denis Cogneau, and Thomas Piketty. 2021. "Income inequality under colonial rule. Evidence from French Algeria, Cameroon, Tunisia, and Vietnam and comparisons with British colonies 1920–1960." *Journal of Development Economics* 152 102680.

Alvaredo, Facundo, and Emmanuel Saez. 2009. "Income and wealth concentration in Spain from a historical and fiscal perspective." *Journal of the European Economic Association* 7, 1140–1167.

Anderson, Chris. 2006. *The long tail: Why the future of business is selling less of more*. Hachette UK.

Atkinson, Anthony B, Thomas Piketty, and Emmanuel Saez. 2011. "Top incomes in the long run of history." *Journal of Economic Literature* 49, 3–71.

Atkinson, Anthony Barnes, and Thomas Piketty. 2007. *Top incomes over the twentieth century: A contrast between continental European and English-speaking countries*. OUP Oxford.

Baum-Snow, Nathaniel, and Ronni Pavan. 2013. "Inequality and city size." *Review of Economics and Statistics* 95, 1535–1548.

Behrens, Kristian, Gilles Duranton, and Frédéric Robert-Nicoud. 2014. "Productive cities: Sorting, selection, and agglomeration." *Journal of Political Economy* 122, 507–553.

Bernhardt, Annette, Christopher Campos, Allen Prohofsky, Aparna Ramesh, and Jesse Rothstein. 2022. "Independent contracting, self-employment, and gig work: Evidence from California tax data." *Industrial and Labor Relations Review*.

Black, Sandra E, Paul J Devereux, Fanny Landaud, and Kjell G Salvanes. 2020. "Where does wealth come from?" Working Paper 28239, National Bureau of Economic Research. 10.3386/w28239.

Blackwell, Calvin, Rachel Graefe-Anderson, Frank Hefner, and Dyanne Vaught. 2015. "Power laws, CEO compensation and inequality." *Economics Letters* 126, 78–80.

Blanchet, Thomas, Emmanuel Saez, and Gabriel Zucman. 2022. "Real-time inequality." Working Paper 30229, National Bureau of Economic Research. 10.3386/w30229.

Bowen, John T, and Shiang-Lih Chen. 2001. "The relationship between customer loyalty and customer satisfaction." *International journal of contemporary hospitality management*.

Brynjolfsson, Erik, Yu Hu, and Duncan Simester. 2011. "Goodbye pareto principle, hello long tail: The effect of search costs on the concentration of product sales." *Management Science* 57, 1373–1386.

Brynjolfsson, Erik, Yu Hu, and Michael D Smith. 2003. "Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers." *Management Science* 49, 1580–1596.

Burtch, Gordon, Seth Carnahan, and Brad N Greenwood. 2018. "Can you gig it? An empirical examination of the gig economy and entrepreneurial activity." *Management Science* 64, 5497–5520.

Chancel, Lucas, and Thomas Piketty. 2019. "Indian Income Inequality, 1922-2015: From British Raj to Billionaire Raj?" *Review of Income and Wealth* 65, S33–S62.

Chen, Binkai, Dan Liu, and Ming Lu. 2018. "City size, migration and urban inequality in China." *China Economic Review* 51, 42–58.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014. "Where is the land of opportunity? The geography of intergenerational mobility in the United States." *The Quarterly Journal of Economics* 129, 1553–1623.

Cook, Cody, Rebecca Diamond, Jonathan V Hall, John A List, and Paul Oyer. 2021. "The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers." *Review of Economic Studies* 88, 2210–2238.

Edmans, Alex, Xavier Gabaix, and Dirk Jenter. 2017. "Executive compensation: A survey of theory and evidence." *The Handbook of the Economics of Corporate Governance* 1, 383–539.

Gabaix, Xavier. 2009. "Power laws in economics and finance." *Annual Review of Economics* 1, 255–294.

Gabaix, Xavier, and Augustin Landier. 2008. "Why has CEO pay increased so much?" *The Quarterly Journal of Economics* 123, 49–100.

Gabaix, Xavier, Augustin Landier, and Julien Sauvagnat. 2014. "CEO pay and firm size: An update after the crisis." *The Economic Journal* 124, F40–F59.

Gefen, David. 2002. "Customer loyalty in e-commerce." *Journal of the association for information systems* 3 2.

Glaeser, Edward L, Matt Resseger, and Kristina Tobio. 2009. "Inequality in cities." *Journal of Regional Science* 49, 617–646.

Goldfarb, Avi, and Catherine Tucker. 2019. "Digital economics." *Journal of Economic Literature* 57, 3–43.

Guo, Feng, Jingyi Wang, Fang Wang, Tao Kong, Xun Zhang, and Zhiyun Cheng. 2020. "Measuring China's digital financial inclusion: Index compilation and spatial characteristics." *China Economic Quarterly* 19, 1401–1418.

Hamlen, William A. 1991. "Superstardom in Popular Music: Empirical Evidence." *The Review of Economics and Statistics* 73, 729–733. 10.2307/2109415, Publisher: The MIT Press.

Huang, Ni, Gordon Burtch, Yili Hong, and Paul A Pavlou. 2020. "Unemployment and worker participation in the gig economy: Evidence from an online labor market." *Information Systems Research* 31, 431–448.

Hvidberg, Kristoffer B, Claus Kreiner, and Stefanie Stantcheva. 2020. "Social positions and fairness views on inequality." Working Paper 28099, National Bureau of Economic Research. 10.3386/w28099.

Jacoby, Jacob, and David B Kyner. 1973. "Brand loyalty vs. repeat purchasing behavior." *Journal of Marketing research* 10, 1–9.

Kahn, Barbara E. 1995. "Consumer variety-seeking among goods and services." *Journal of Retailing and Consumer Services* 2, 139–148. 10.1016/0969-6989(95)00038-0.

Kahn, Barbara E, and Therese A Louie. 1990. "Effects of Retraction of Price Promotions on Brand Choice Behavior for Variety-Seeking and Last-Purchase-Loyal Consumers." *JOURNAL OF MARKETING RESEARCH*.

Kaplan, Steven N., and Joshua Rauh. 2013. "It's the Market: The Broad-Based Rise in the Return to Top Talent." *Journal of Economic Perspectives* 27, 35–56. 10.1257/jep.27.3.35.

Kim, Ju-Young, Martin Natter, and Martin Spann. 2009. "Pay what you want: A new participative pricing mechanism." *Journal of Marketing* 73, 44–58.

Klass, Oren S, Ofer Biham, Moshe Levy, Ofer Malcai, and Sorin Solomon. 2006. "The Forbes 400 and the Pareto wealth distribution." *Economics Letters* 90, 290–295.

Kleiber, Christian, and Samuel Kotz. 2003. *Statistical size distributions in economics and actuarial sciences*. John Wiley & Sons.

Kleibergen, Frank, and Richard Paap. 2006. "Generalized reduced rank tests using the singular value decomposition." *Journal of Econometrics* 133, 97–126.

Koenig, Felix. 2022. "Technical Change and Superstar Effects: Evidence from the Rollout of Television." *SSRN Electronic Journal*. 10.2139/ssrn.4114505.

Kopczuk, Wojciech, Emmanuel Saez, and Jae Song. 2010. "Earnings inequality and mobility in the United States: Evidence from social security data since 1937." *The Quarterly Journal of Economics* 125, 91–128.

Krueger, Alan B. 2005. "The Economics of Real Superstars: The Market for Rock Concerts in the Material World." *Journal of Labor Economics* 23, 1–30. 10.1086/425431, Publisher: The University of Chicago Press.

Kumar, VISWANATHAN, and Denish Shah. 2004. "Building and sustaining profitable customer loyalty for the 21st century." *Journal of retailing* 80, 317–329.

Lindert, Peter H. 1986. "Unequal English wealth since 1670." *Journal of Political Economy* 94, 1127–1162.

Long, James E, David W Rasmussen, and Charles T Haworth. 1977. "Income inequality and city size." *Review of Economics and Statistics*, 244–246.

Lu, Shijie, Dai Yao, Xingyu Chen, and Rajdeep Grewal. 2021. "Do larger audiences generate greater revenues under pay what you want? evidence from a live streaming platform." *Marketing Science* 40, 964–984.

McAlister, Leigh, and Edgar Pessemier. 1982. "Variety Seeking Behavior: An Interdisciplinary Review." *Journal of Consumer Research* 9 311. 10.1086/208926.

Moriguchi, Chiaki, and Emmanuel Saez. 2008. "The evolution of income concentration in Japan, 1886–2005: Evidence from income tax statistics." *The Review of Economics and Statistics* 90, 713–734.

Murphy, Kevin M, Andrei Shleifer, and Robert Vishny. 1989. "Income distribution, market size, and industrialization." *The Quarterly Journal of Economics* 104, 537–564.

Nord, Stephen. 1980. "Income inequality and city size: An examination of alternative hypotheses for large and small cities." *Review of Economics and Statistics*, 502–508.

Novokmet, Filip, Thomas Piketty, and Gabriel Zucman. 2018. "From Soviets to oligarchs: Inequality and property in Russia 1905-2016." *The Journal of Economic Inequality* 16, 189–223.

Piketty, Thomas. 2003. "Income inequality in France, 1901–1998." *Journal of Political Economy* 111, 1004–1042.

Piketty, Thomas. 2011. "On the long-run evolution of inheritance: France 1820–2050." *The Quarterly Journal of Economics* 126, 1071–1131.

Piketty, Thomas, and Emmanuel Saez. 2003. "Income inequality in the United States, 1913–1998." *The Quarterly Journal of Economics* 118, 1–41.

Piketty, Thomas, and Emmanuel Saez. 2014. "Inequality in the long run." *Science* 344, 838–843.

Piketty, Thomas, Li Yang, and Gabriel Zucman. 2019. "Capital accumulation, private property, and rising inequality in China, 1978–2015." *American Economic Review* 109, 2469–96.

Ratner, Rebecca K., Barbara E. Kahn, and Daniel Kahneman. 1999. "Choosing Less-Preferred Experiences For the Sake of Variety." *Journal of Consumer Research* 26, 1–15. 10.1086/209547.

Rosen, Sherwin. 1981. "The economics of superstars." *American Economic Review* 71, 845–858.

Rosen, Sherwin. 1983. "The economics of superstars." *The American Scholar*, 449–460.

Saez, Emmanuel, and Gabriel Zucman. 2016. "Wealth inequality in the United States since 1913: Evidence from capitalized income tax data." *The Quarterly Journal of Economics* 131, 519–578.

Salganik, Matthew J, Peter Sheridan Dodds, and Duncan J Watts. 2006. "Experimental study of inequality and unpredictability in an artificial cultural market." *Science* 311, 854–856.

Schmidt, Klaus M, Martin Spann, and Robert Zeithammer. 2015. "Pay what you want as a marketing strategy in monopolistic and competitive markets." *Management Science* 61, 1217–1236.

Seetharaman, P. B., and Hai Che. 2009. "Price Competition in Markets with Consumer Variety Seeking." *Marketing Science* 28, 516–525. 10.1287/mksc.1080.0434.

Sevilla, Julio, Jiao Zhang, and Barbara E. Kahn. 2016. "Anticipation of Future Variety Reduces Satiation from Current Experiences." *Journal of Marketing Research* 53, 954–968. 10.1509/jmr.14.0360.

Stigler, George J, and Gary S Becker. 1977. "De gustibus non est disputandum." *American Economic Review* 67, 76–90.

Tervio, Marko. 2008. "The difference that CEOs make: An assignment model approach." *American Economic Review* 98, 642–68.

Zentner, Alejandro, Michael Smith, and Cuneyd Kaya. 2013. "How video rental patterns change as consumers move online." *Management Science* 59, 2622–2634.

Figures and Tables

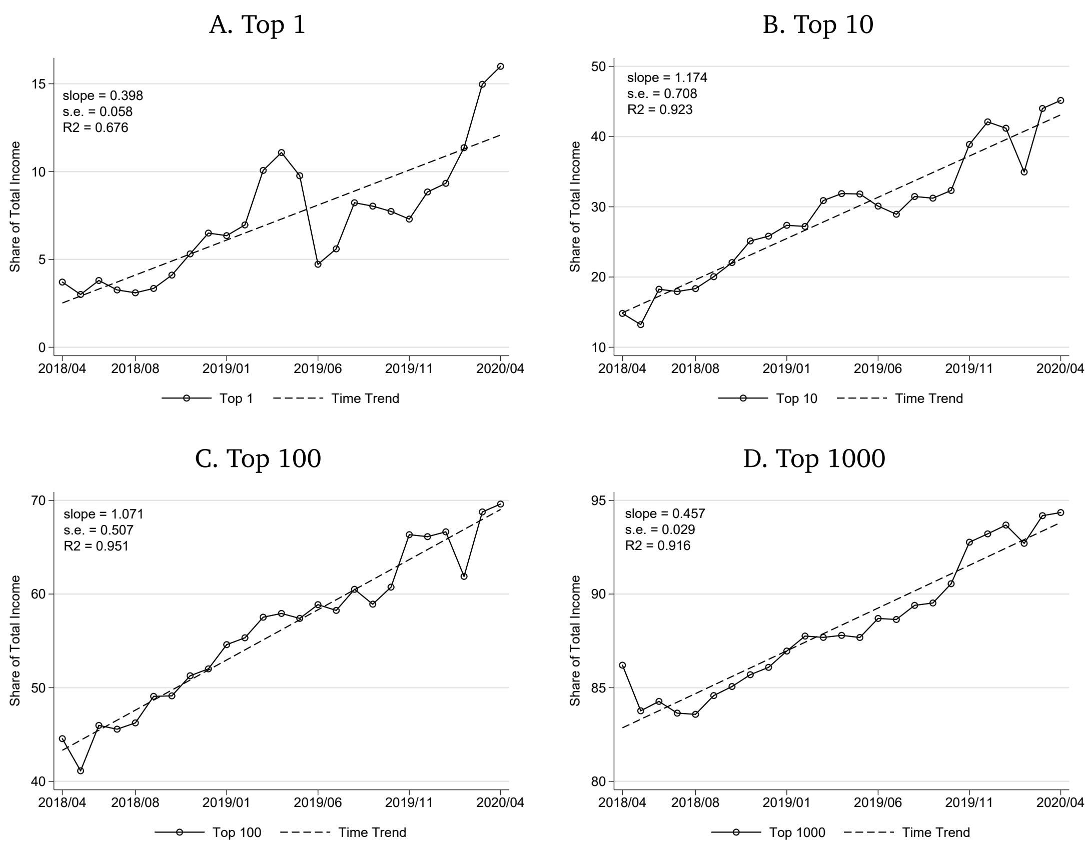


Figure 1. **Seasonality-adjusted income share of top streamers.** This figure ranks the streamers according to their aggregate income over the past three months and calculate the top groups' share of total tipping income. The upper left corner of each sub figure is marked with the slope, standard error, and R^2 of the linear fit line.

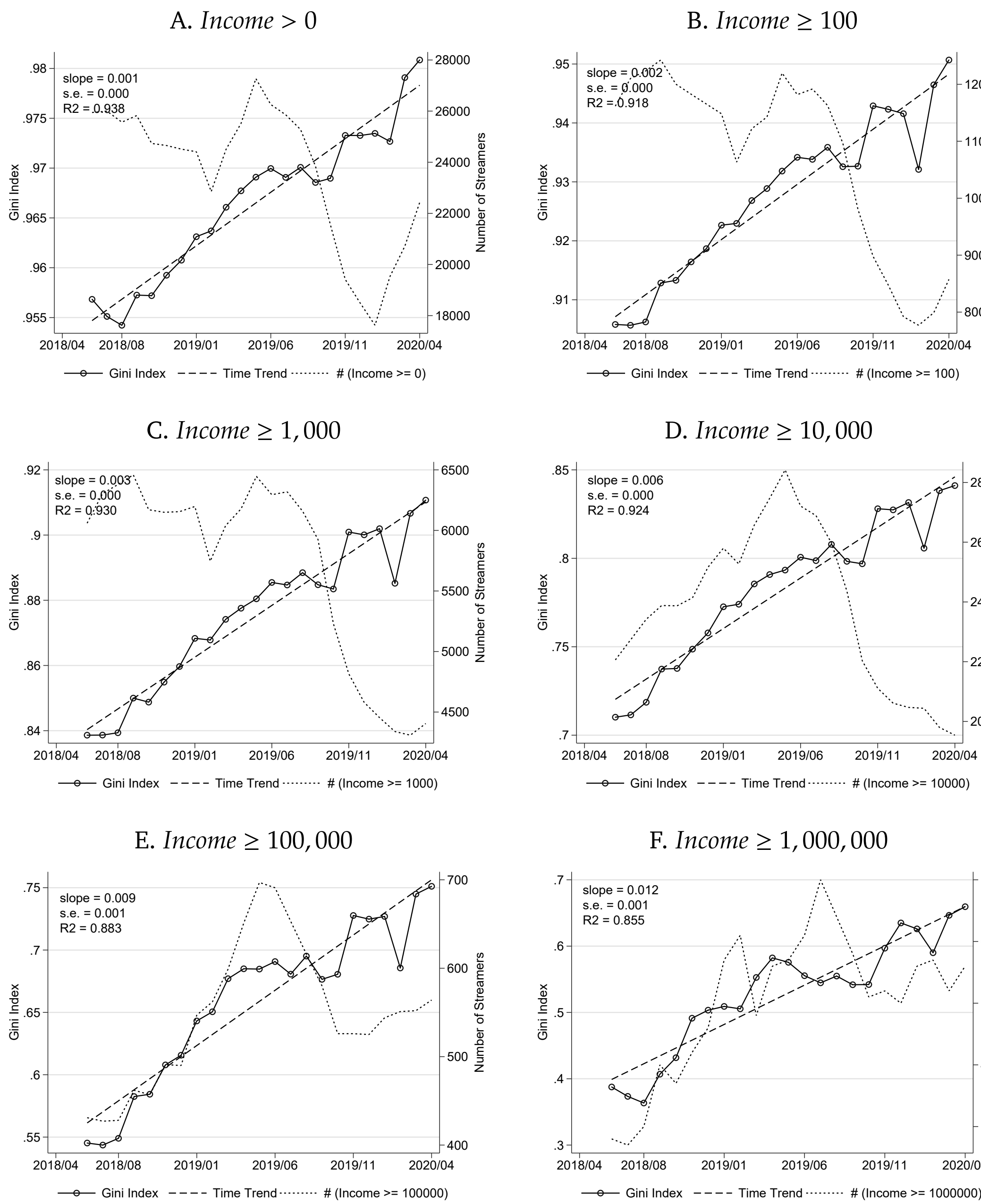


Figure 2. **Seasonality-adjusted Gini index.** This figure shows the seasonality-adjusted inequality trend as measured by the Gini index. We aggregate streamers' tipping income over the past three months and calculate the Gini index using the resulting income distribution. In Panel A, the dotted line shows the number of streamers with positive three-month tipping income, the solid line shows the Gini coefficient calculated based on this sample, and the dashed line is the linear fit time trend of the Gini index. The other five panels are similar, except that the sample is replaced by streamers who earn more than 100, 1,000, 10,000, 100,000, and 1,000,000 RMB over the past three months respectively. The upper left corner of each sub figure is marked with the slope, standard error, and R^2 of the linear fit line.

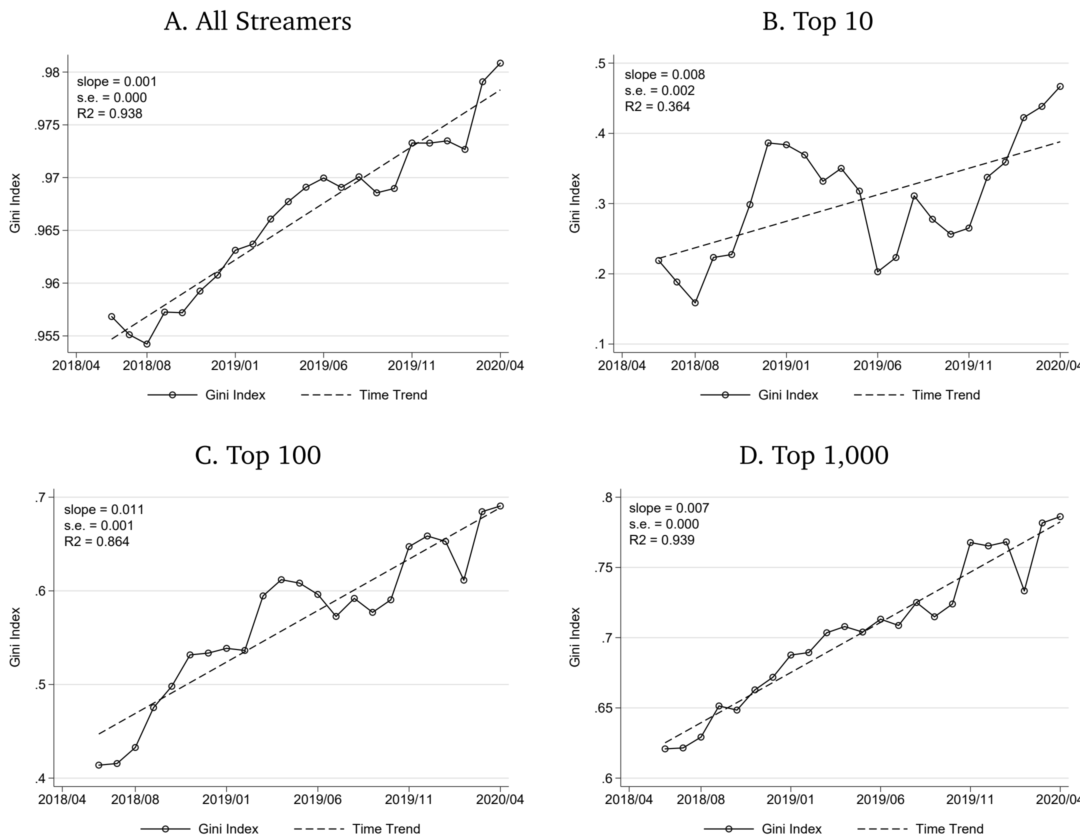


Figure 3. **Seasonality-adjusted Gini index among top streamers.** This figure shows the seasonality-adjusted inequality trend as measured by the Gini index within top streamers. We aggregate streamers' tipping income over the past three months, rank the streamers according to their aggregate income, and calculate the Gini index using the resulting income distribution of those top streamers. In Panel A, as a benchmark, the Gini index is calculated based on the whole streamer sample who has earned any virtual gifts over the past three months. From Panel B to Panel D, the Gini index is calculated based on the top 10, top 100, and top 1,000 streamers respectively. In each sub figure, the dashed line is the linear fit time trend of the Gini index, and the upper left corner is marked with the slope, standard error, and R^2 of this linear fit line.

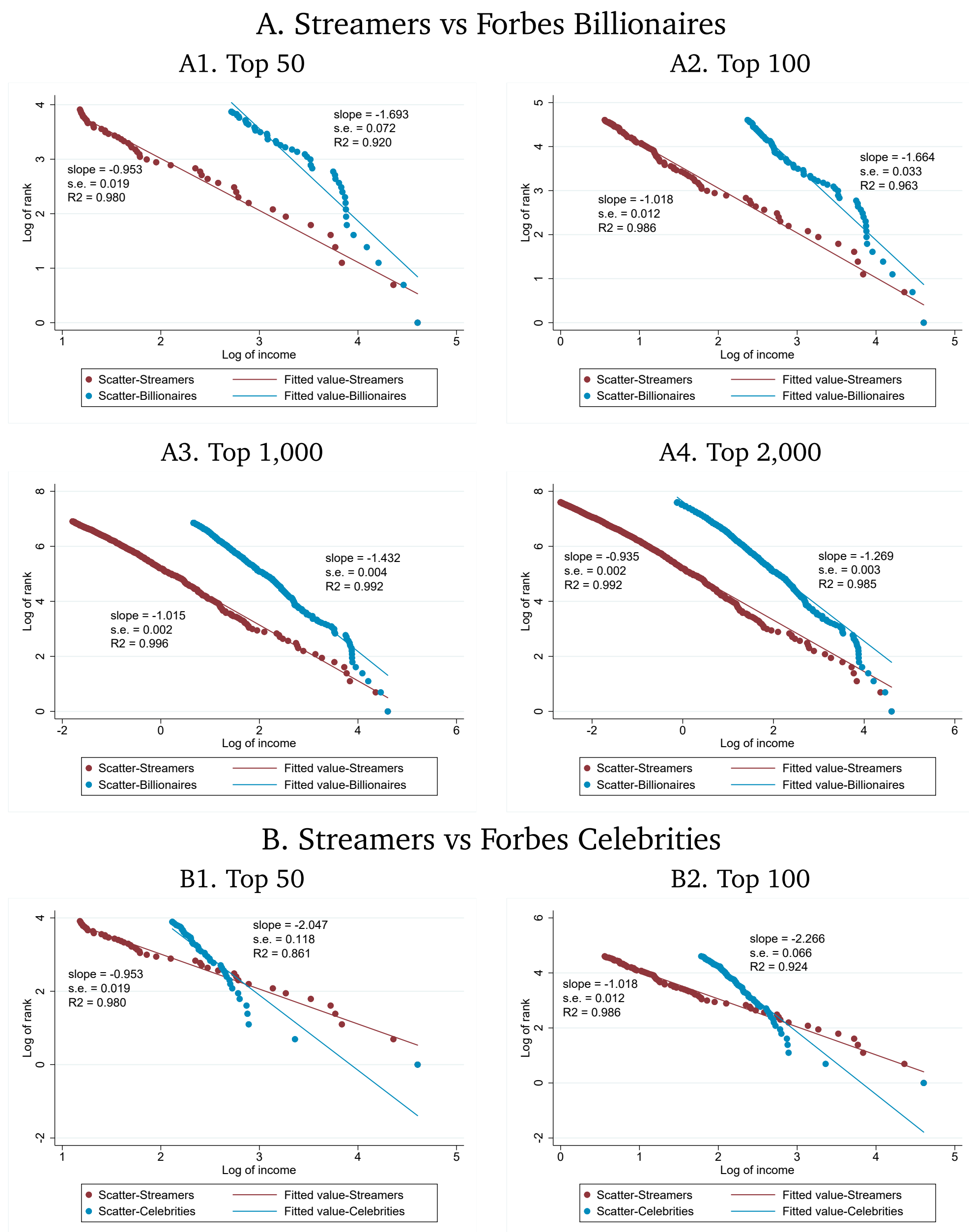
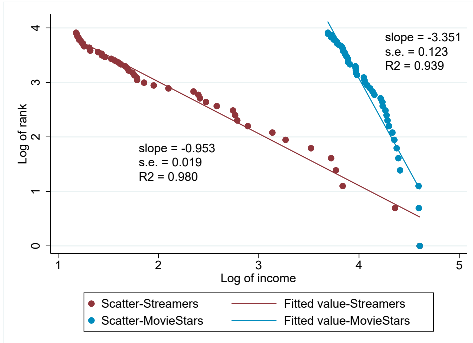


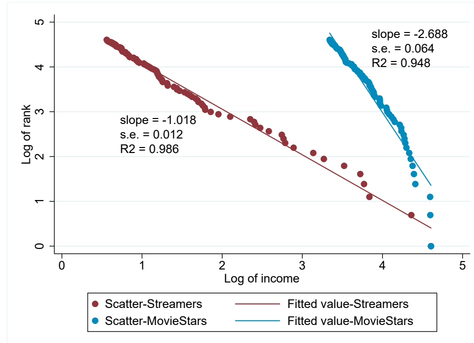
Figure 4. **PL exponent of top streamers vs PL exponent of Forbes.** Panel A compares PL exponent calculated using the streamers sample with that calculated using the 2020 Forbes' ranking of the world's richest billionaires. Panel B compares inequality among top streamers with inequality among 2020 Forbes' ranking of the world's highest-paid celebrities. Streamers are ordered according to the total value of their received virtual gifts during the whole sample period. The log of rank is placed on the y-axis, and the log of streamers' total tipping income or wealth of the rich, which we call their size, is placed on the x-axis. The size variables are normalized by their max value so that the red distribution line of Forbes and the blue distribution line of streamers can get started from the same position on the x-axis in the graph. Specifically, $AdjustedSize = \frac{size}{max(size)} \times 100$. From sub-figure A1 to sub figure A4, the sample used is top 50, top 100, top 1,000, and top 2,000, respectively. From sub-figure B1 to sub-figure B2, the sample used is top 50 and 100 respectively. The slope, standard error, and R^2 of the Forbes linear fit line are presented in the upper right corner of each sub-figure, while that of the linear fit line of streamers are presented in the bottom left corner of each sub-figure. Data source: `https://www.forbes.com/billionaires/` and `https://www.forbes.com/celebrities/list/`.

A. Streamers vs Movie Stars

A1. Top 50

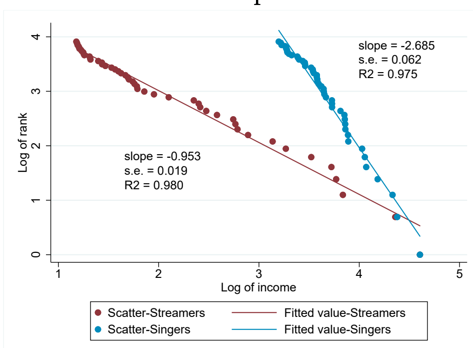


A2. Top 100

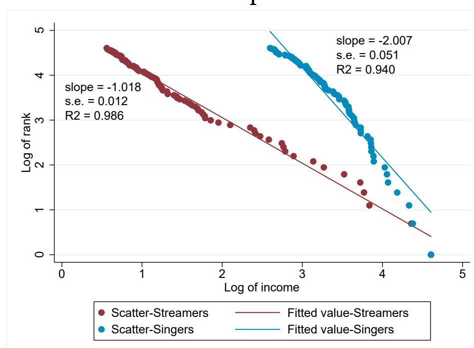


B. Streamers vs Singers

B1. Top 50

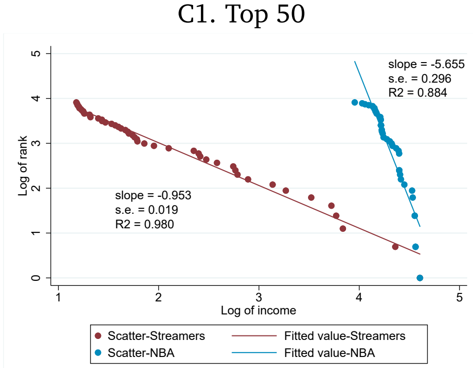


B2. Top 100



C. Streamers vs. NBA Players

C1. Top 50



C2. Top 100

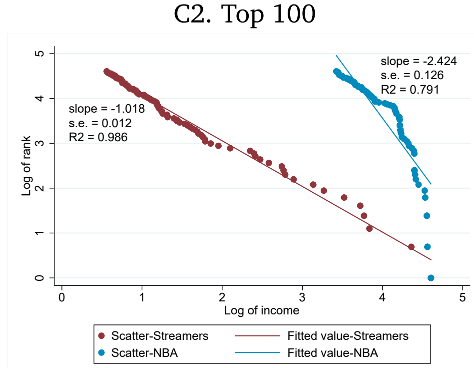


Figure 5. **PL exponent of top streamers vs. PL exponent of other superstars.** This figure compares the PL exponent calculated using the streamer's sample with that calculated using other superstars in the field of arts and sports. Streamers are ordered according to the total value of their virtual gifts during the sample period. In Panel A, movie stars are ordered by the cumulative worldwide box office of all the movies in which a star has had a leading role over their lifetime. In Panel B, singers are ordered by the compound sales, including the original album, compilations generated thanks to the album, physical singles from the album, digital singles from the album, and all the album tracks in audio or video stream. In Panel C, NBA players are ordered by their salary in the 19/20 season. The log of rank is placed on the y-axis, and the log of streamers' total tipping income or earnings of the superstars, which we call their size, is placed on the x-axis. The size variables are normalized by their max value so that the red distribution line of Forbes and the blue distribution line of streamers can get started from the same position on the x-axis in the graph. The slope, standard error, and R^2 of the Forbes linear fit line are presented in the upper right corner of each sub-figure, while that of the linear fit line of streamers are presented in the bottom left corner of each sub-figure.[69]

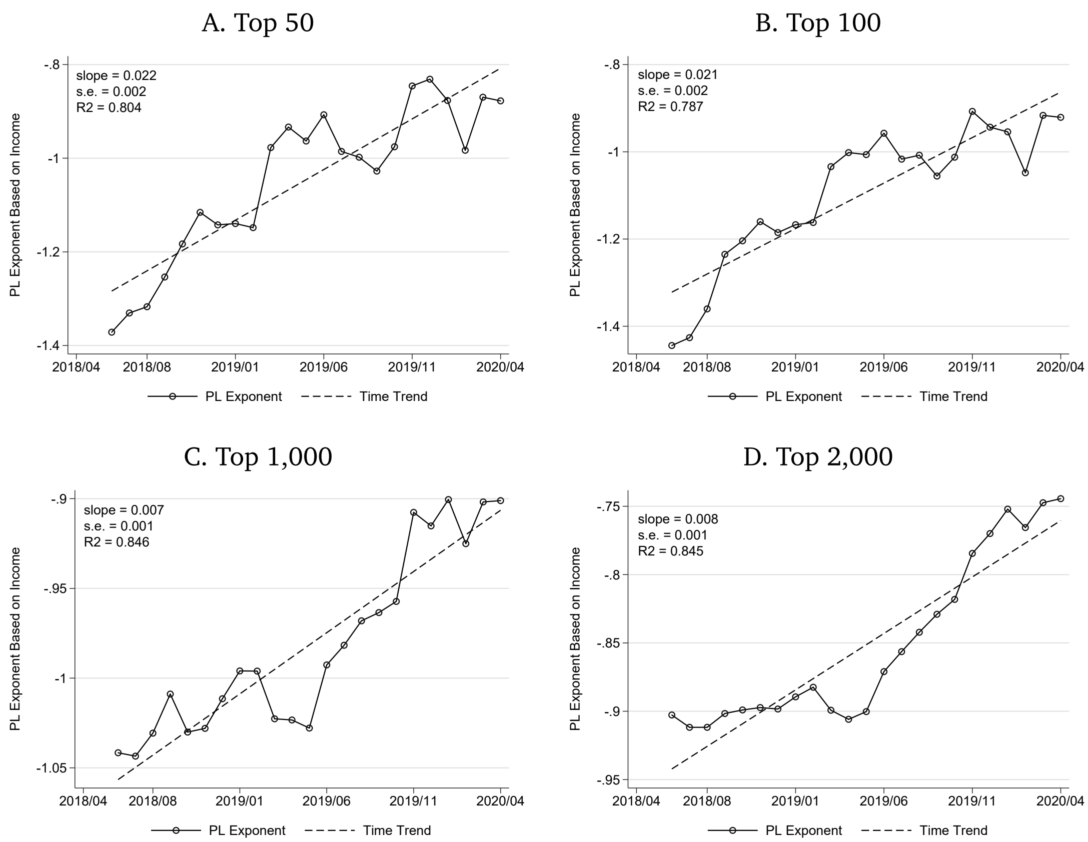


Figure 6. **Seasonality-adjusted PL exponent.** This figure shows the dynamic of inequality as measured by the seasonality-adjusted PL exponent. We aggregate streamers' tipping income over the past three months, rank them according to their aggregate income, and calculate the PL exponent using the resulting income distribution of those top streamers. Specifically, we run the linear regressions $\ln rank_{jt} = \alpha_t + \beta_t \ln size_{jt} + \varepsilon_{jt}, \forall t$, and β_t is the estimated PL exponent in month t. From Panel A to Panel D, the sample used are top 50, top 100, top 1,000, and top 2,000 streamers, respectively. The solid line plots the β_t, and the dashed line is the linear fit time trend of PL exponents. The slope, standard error, and R^2 of the linear fit time trend are presented in the top left corner of each sub-figure.

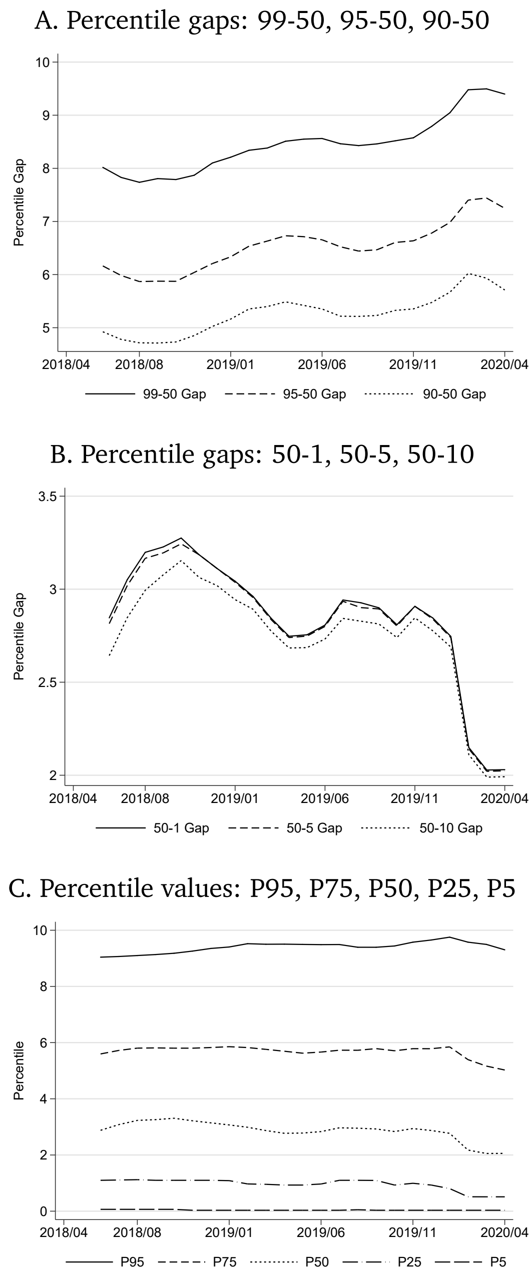
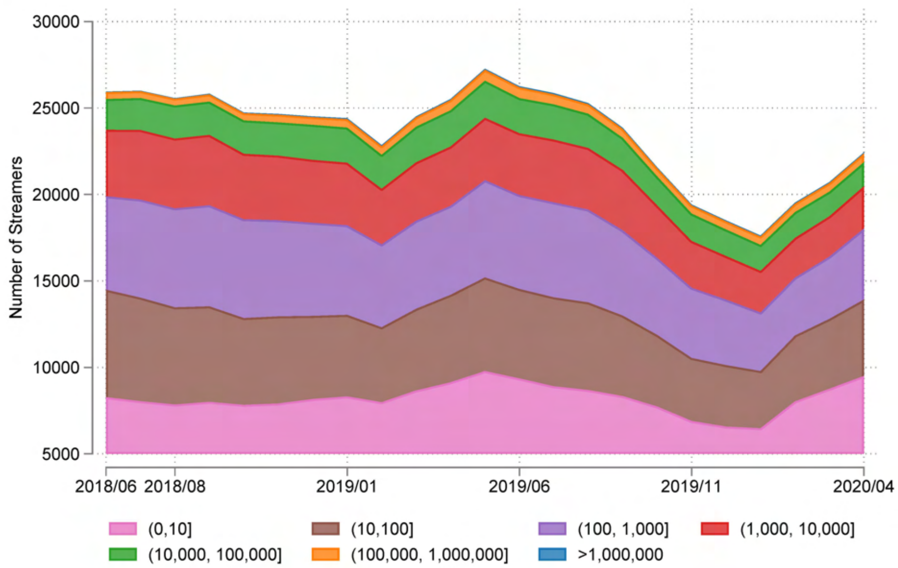


Figure 7. **Percentile gaps.** This figure shows the time trend of different percentile gaps of the log income variable. We aggregate streamers' tipping income over the past three months and calculate different percentile values using the distribution of the log income variable. The exact percentile values are presented in Table A2 of the Appendix. In this figure, Panel A plots the dynamics of the 99-50 gap ($P99 - P50$), 95-50 gap ($P95 - P50$), and 90-50 gap ($P90 - P50$), which indicates the gap between the superstar group and the middle-income group. Panel B plots the dynamics of the 50-1 gap ($P50 - P1$), 50-5 gap ($P50 - P5$), and 50-10 gap ($P50 - P10$), which indicates the gap between the middle-income and lower-income groups. Panel C presents the 95th percentile ($P95$), the 75th percentile ($P75$), the 50th percentile ($P50$), the 25th percentile ($P25$), and the 5th percentile ($P5$).

A. Income distribution of streamers



B. Number of superstar streamers

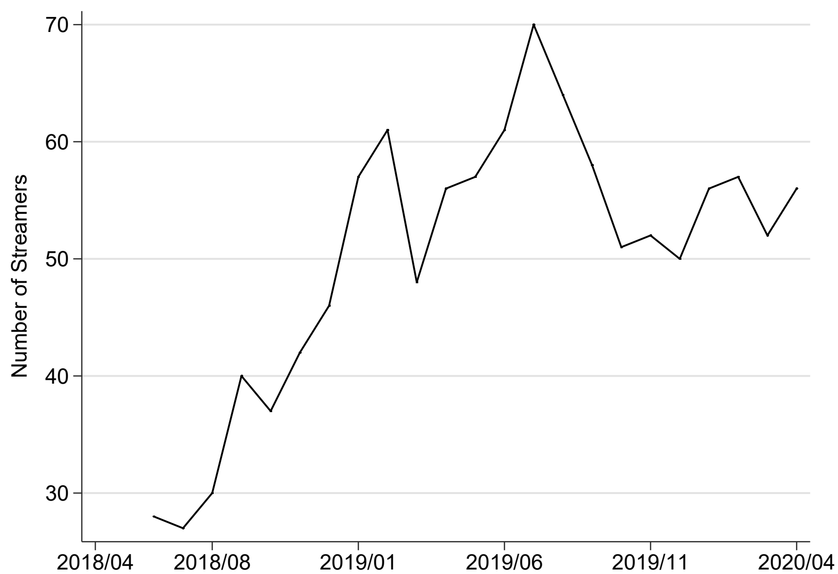


Figure 8. **Income distribution of streamers.** We aggregate streamers' tipping income over the past three months, and streamers are then divided into different groups according to their income. In Panel A, this stacked area chart not only tracks the total number of streamers who earned any positive income over the past three months (i.e., the fully-stacked height of the top line) but also helps to understand the breakdown of that total by group (e.g., the heights of the purple area refer to the number of streamers whose three-month pay range from 100 to 1,000 RMB). Since only a few superstar streamers can earn more than 1,000,000 RMB, the blue area in Panel A is unclear. So, we specifically plot the dynamics of this number in Panel B.

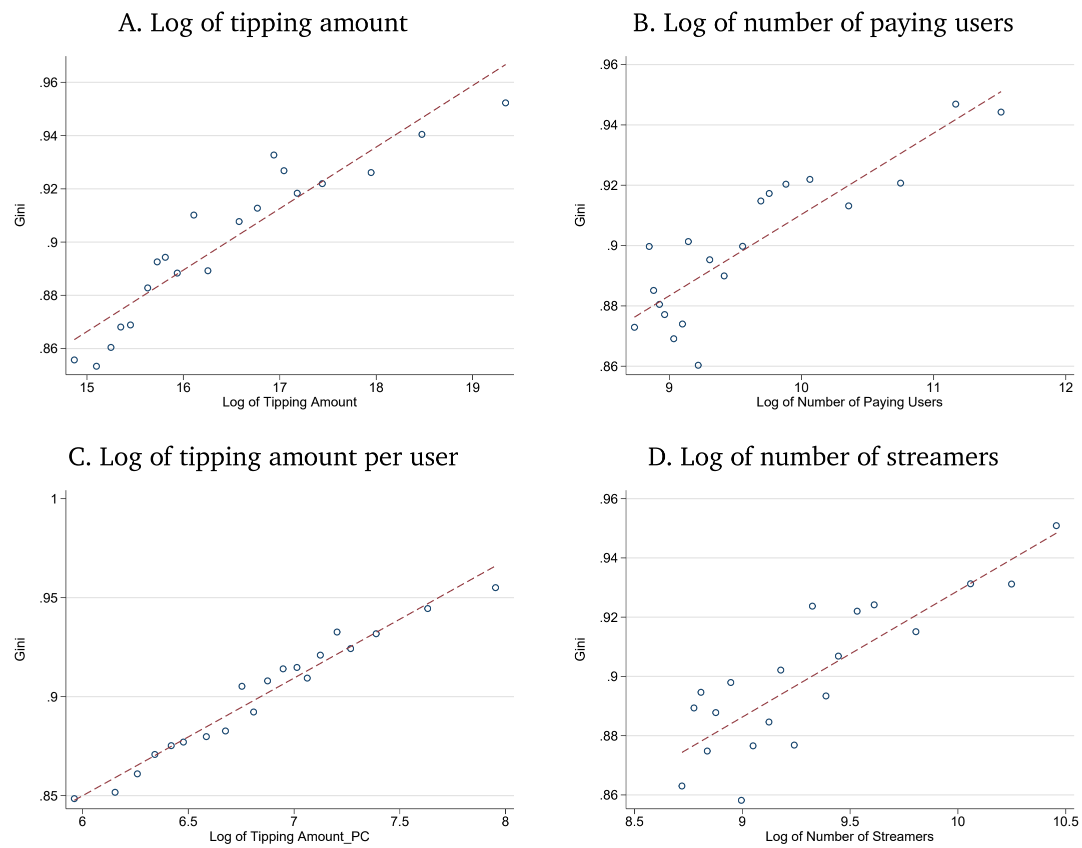
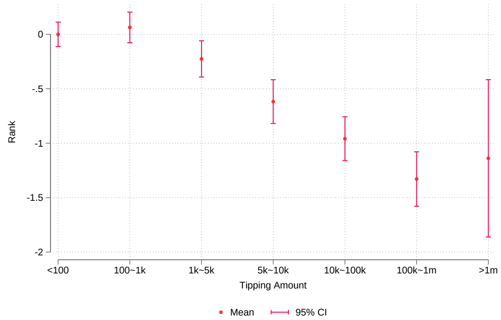


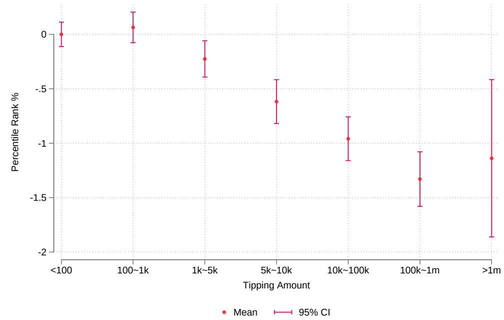
Figure 9. **Cross city relationships between market size and inequality.** This figure provides intuitive cross-city evidence that a broader market positively correlates with greater inequality. In Panels A to D, we use the total tipping amount, the total number of paying users, the tipping amount per user, and the total number of streamers who receive positive tipping income from a city during the sample period, respectively, as indicators of market size. The Gini coefficients are calculated using the income distribution of the streamers who receive more than 50 RMB in tipping income from one city during the whole sample period. The city sub-sample includes 92 cities whose number of paying users accounts for more than 0.2% of the total number of paying users.

A. The rank level of streamers that users spend the most money to

A1. Absolute rank

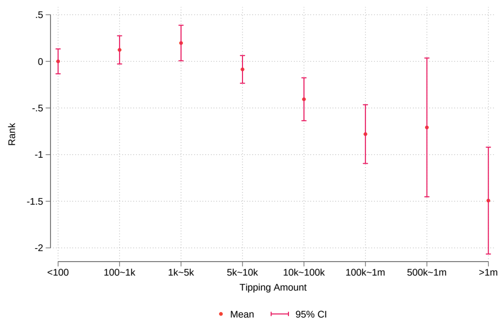


A2. Percentile rank %

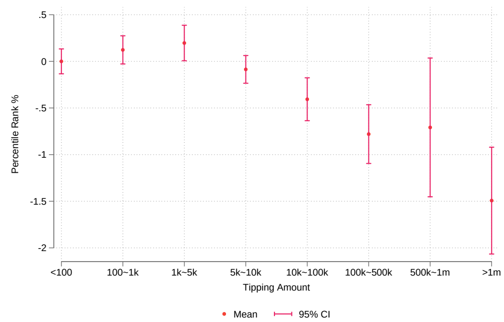


B. The rank level of streamers that users spend the most time to

B1. Absolute rank

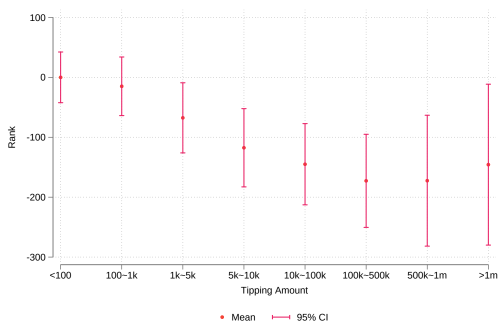


B2. Percentile rank %



C. The rank level of streamers that users newly follow

C1. Absolute rank



C2. Percentile rank %

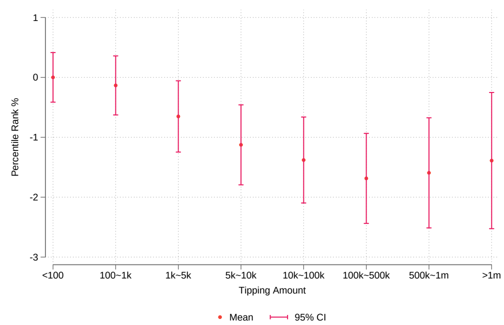


Figure 10. **The rank level of streamers that users appreciate with different consumption levels.** This figure provides evidence of whether active users who spend a lot would disproportionately appreciate the top streamers who already have cumulative advantages. For each month t, we divide users into eight groups based on their monthly tipping amount. Absolute rank is the average standing of the streamers, while percentile rank is the average relative position of the streamers that users appreciate. The smaller one streamer's ranking, the higher her tier and the greater her cumulative advantage. In Panel A, we focus on the streamer that users pay the most money to and explore which tier of streamers the users' money resources are mainly allocated. In Panel B, we identify users' favorite streamers based on their watching time in each live broadcast room and explore the ranking of streamers that users spend the most time with. In Panel C, we analyze the streamers that users newly add to their follow list and investigate which tier of streamers the users' attention flow to. The 95% confidence interval is based on the Month – User level clustered standard errors. This figure shows the results after controlling for user-level fixed effects.

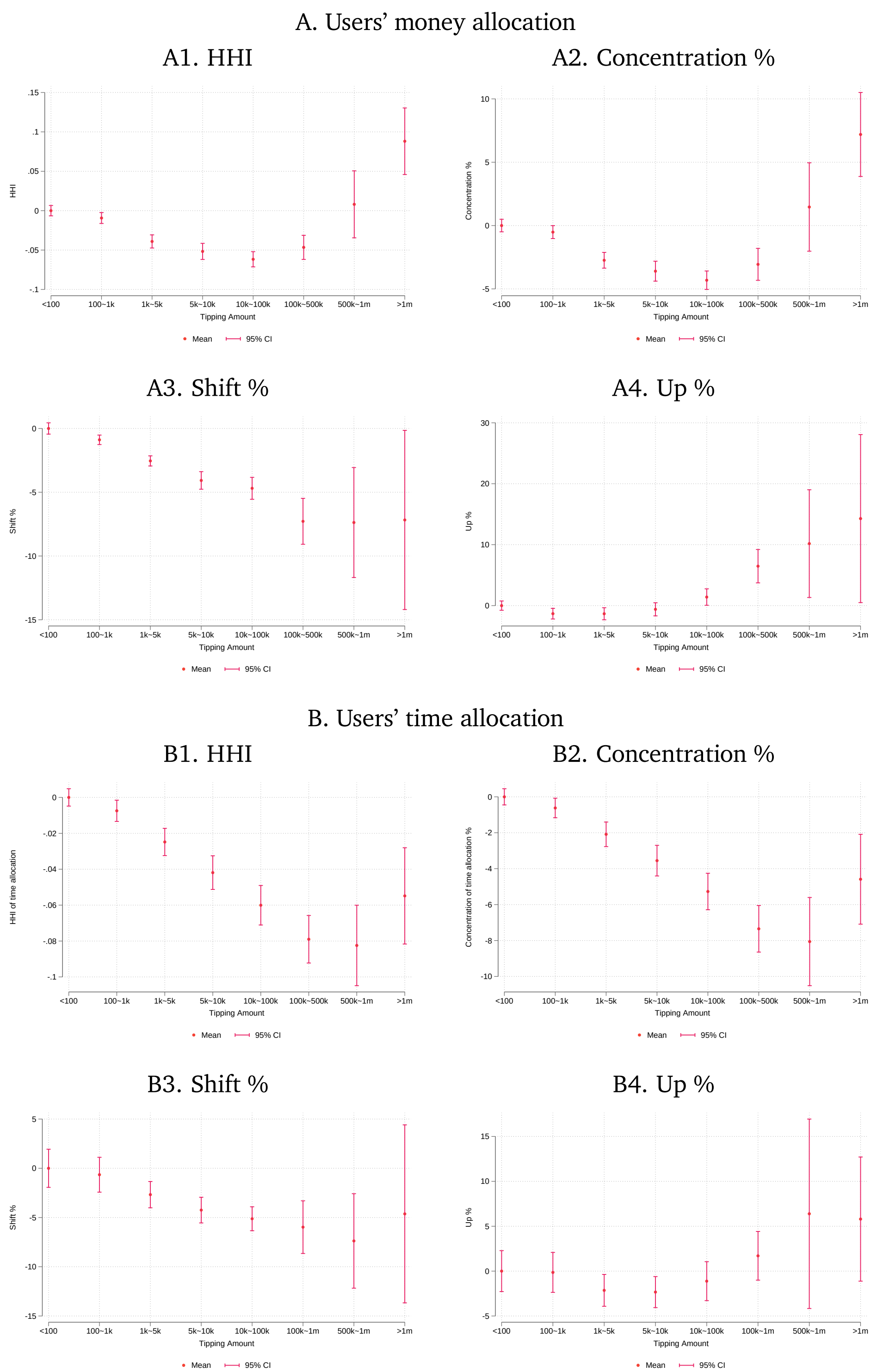


Figure 11. **Loyalty of users with different consumption levels.** This figure provides evidence of whether users with high consumption levels would have exceptionally high loyalty. When measuring loyalty, we focus on users' money allocation in Panel A, while in Panel B, we investigate users' viewing time allocation. For the x-axis, we divide users into eight groups based on their monthly tipping amount for each month t. For the y-axis, *Concentration*, *HHI*, $Shift$, and *Up* are all proxy variables of users' loyalty. The higher concentration or HHI value, the higher the users' loyalty to specific streamers. $Shift$ is the probability that the streamer that users are loyal to changes to someone else. The lower the $Shift$ value, the higher users' loyalty to specific streamers. And *Up* is the probability of users shifting toward a higher-tier streamer. The higher the probability, the more likely users would experience consumption upgrading to a higher-rank streamer. The 95% confidence interval is based on the Month – User level clustered standard errors. This figure shows the results after controlling for user-level fixed effects.

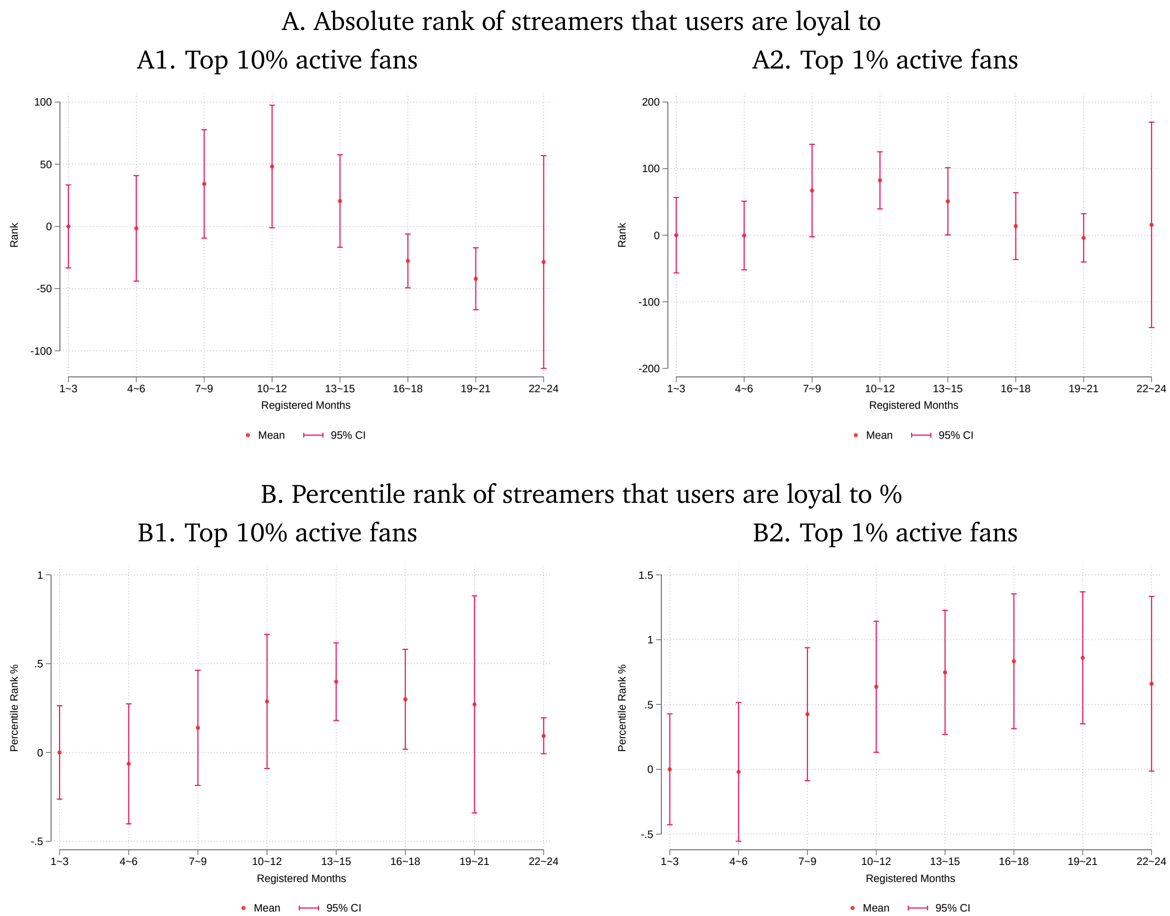


Figure 12. **The rank level of streamers to which users at different stages of customer lifecycle are loyal.** This figure explores the heterogeneity of cohorts at different stages of the customer lifecycle in terms of the rank status of streamers that users are loyal to. Criteria for selecting the subjects were as follows: first, we sorted users' aggregated tipping amount over the sample period to pick the top 10% and 1% users that the platform value a lot; second, to observe the full lifecycle of users and control for bias, we only select the cohorts who registered between April 2018 and December 2018. For each month t, we divide users into eight groups based on how long it has been since they registered. In Panel A, the absolute rank of streamers refers to the average rank of streamers to whom users give the largest gift share. While in Panel B, the percentile rank of streamers refers to the average relative status of streamers to whom users give the largest gift share. The smaller one streamer's ranking, the higher her tier and the greater her cumulative advantage. The 95% confidence interval is based on the Month – User level clustered standard errors. The results have controlled user-level fixed effects.

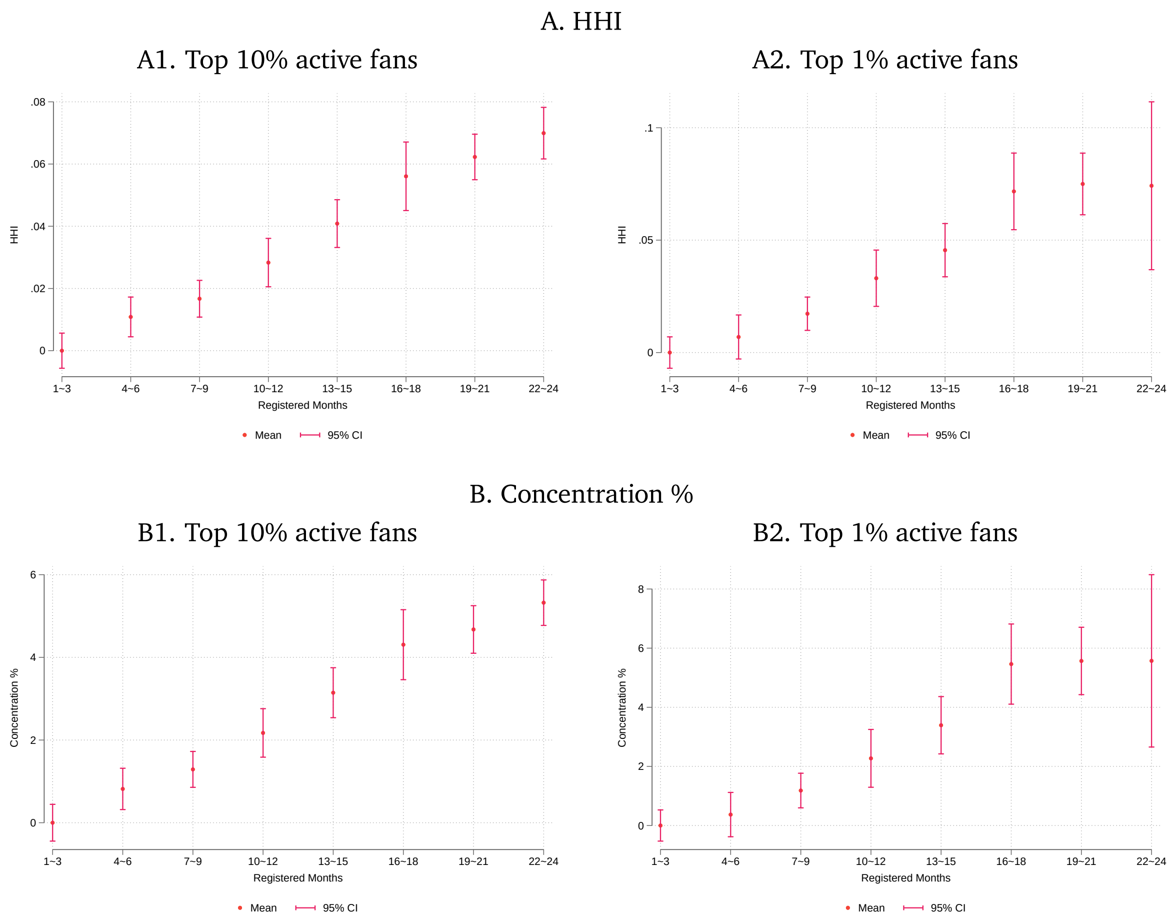


Figure 13. **Loyalty of users at different stages of the customer lifecycle.** This figure explores the heterogeneity of cohorts at different stages of the customer lifecycle in terms of their loyalty. Criteria for selecting the subjects were as follows: first, we sorted users' aggregated tipping amount over the sample period to pick the top 10% and 1% users that the platform value a lot; second, to observe the full lifecycle of users and control for bias, we only select the cohorts who registered between April 2018 and December 2018. For each month t, we divide users into eight groups based on how long it has been since they registered. *Concentration* and *HHI* are all proxy variables of users' loyalty. The higher concentration or HHI value, the higher the users' loyalty to specific streamers. The 95% confidence interval is based on the Month – User level clustered standard errors. The results have controlled user-level fixed effects.

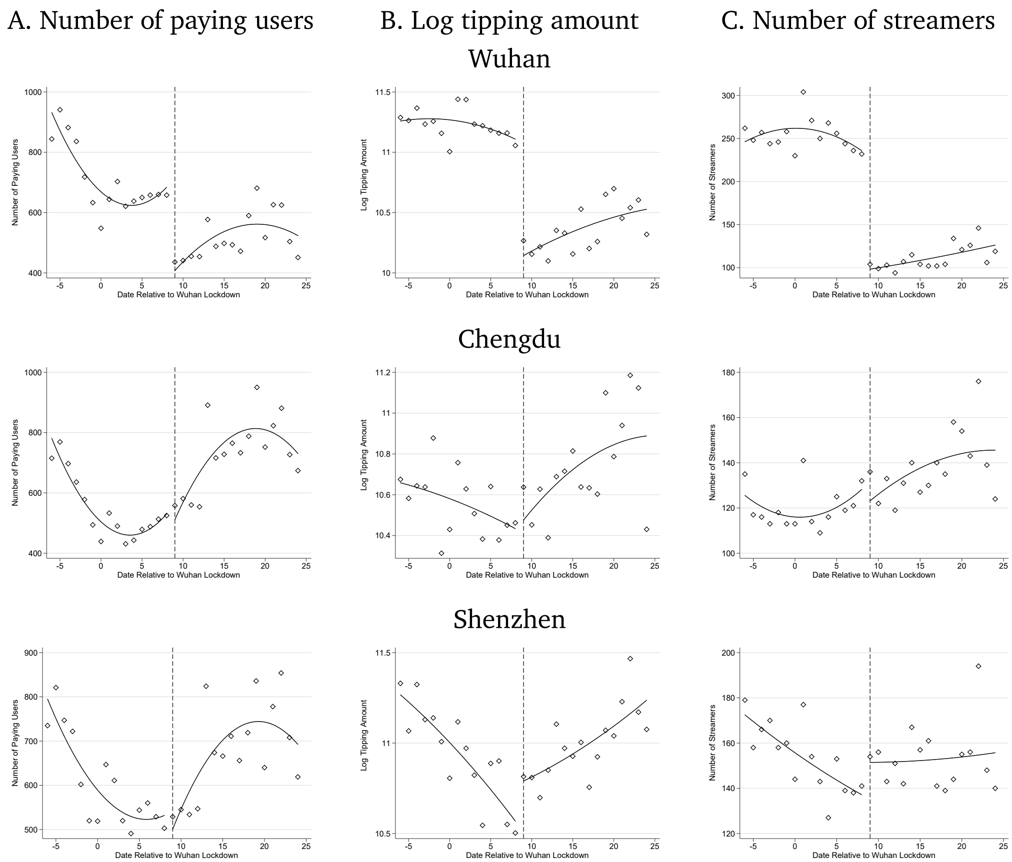


Figure 14. **Entertainment demand discontinuity in Wuhan.** The first row in this figure shows the discontinuity in entertainment demand in Wuhan around Day 9 (February 1, 2020). From that day on, the epidemic became serious (at least to an extent), the citizens of Wuhan became more concerned about it, and the government implemented a series of escalating protective measures. Panel A, Panel B, and Panel C scatter the daily value of the number of paying users, the log number of tips, and the number of streamers who received any tip income, all of which are indicators of entertainment demand, around Day 9, when the seriousness of the epidemic became clear. The solid line in each sub-figure is the quadratic fit line before and after the breakpoint, Day 9. We also plot other panels in the second and third rows in the same way, using the data from Chengdu and Shenzhen, which are all distant cities to Wuhan, for comparison.

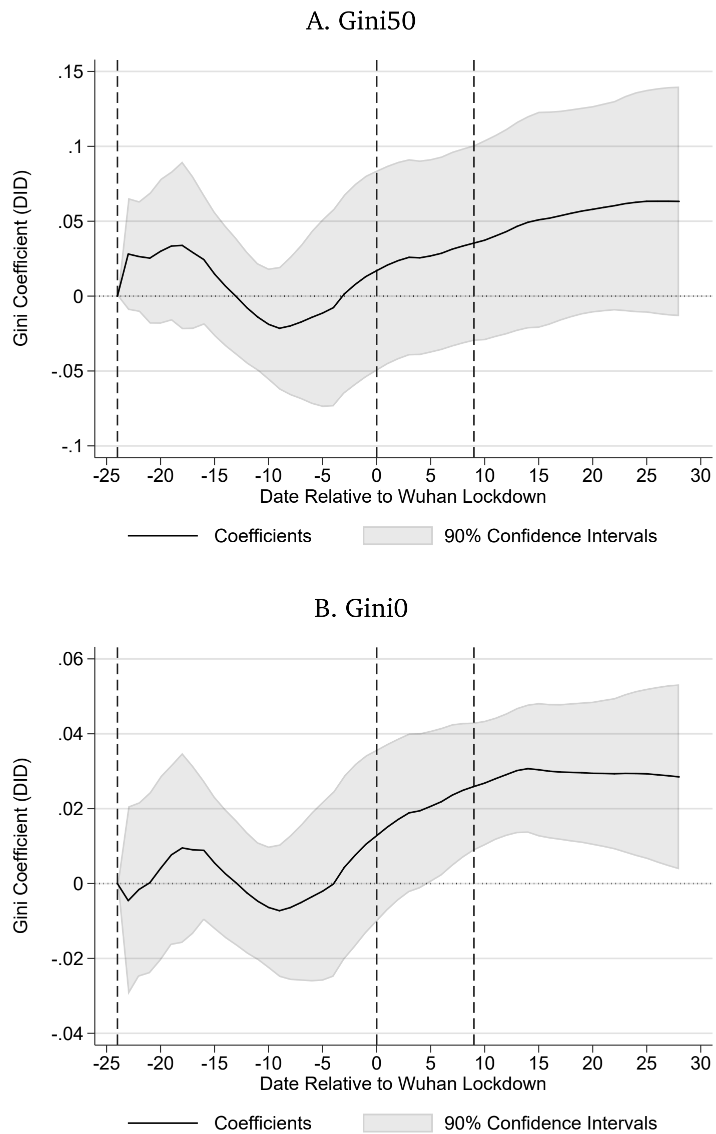


Figure 15. **The dynamic treatment effects.** This figure plots the dynamic treatment effects from the day of Wenliang Li's post (Day -24) to the Wuhan lockdown (Day 0) and the four weeks after the Wuhan lockdown. We run the regressions $Gini_{jt} = \alpha_T + \beta_T ln(Distance_j) \times Post_t + \gamma_T Post_t + \theta_j + \epsilon_{jt}$, for $t \in [-50, T]$ and $\forall T > -24$. In Panel A, $Gini50_{jt}$ is calculated using the income distribution of streamers who receive virtual gifts worth more than 50 RMB from citizens in city j from Day $(t-6)$ to Day t. In Panel B, $Gini0_{jt}$ is calculated using the income distribution of all streamers who earn positive income from users in the city j from Day $(t-6)$ to Day t. $Distance_j$ is the driving time (in days) between Wuhan and city j, as shown on Baidu Maps. $Post_{1t} = 1$ indicates the days after December 30, 2019 (Day -24) when Dr. Wenliang Li posted a Covid-19 alert on one of his WeChat groups, and Covid-19 became known to the public. θ_j represents city-level fixed effects. Using the observations of the nine cities from Day -50 to Day T, we run the above regression and get the estimated β_T for $\forall T > -24$. Next, we plot β_T before the interaction term as a function of time. The solid black line shows the absolute value of the β_T, and the gray area indicates its 90% confidence interval. The x-axis refers to the days relative to Wuhan lockdown on January 23, 2020. The third vertical dashed line indicates Day 9 (February 1, 2020), from which day a series of escalated measures were implemented in response to the aggravation of the epidemic. On February 1, 2020, the Hubei government announced the extension of the Spring Festival holiday. On February 2, 2020, *Huoshenshan Hospital* was officially put into operation, and the Hubei government announced that all suspected Covid-19 cases would be centrally isolated.

Table 1. **Inequality among top streamers vs. inequality among other superstars.** This table compares the PL exponent and Gini index calculated using the streamers sample with that calculated using other superstars. Streamers are ordered according to the total value of their virtual gifts during the sample period. Movie stars are ordered by the cumulative worldwide box office of all the movies a star has had a leading role in over their lifetime. Singers are ordered by the compound sales, including the original album, compilations generated thanks to the album, physical singles from the album, digital singles from the album, and all the album tracks in audio or video stream. NBA players are ordered by their salary in the 19/20 season. And the rankings of billionaires and celebrities are from Forbes. [70]

Group	Top	PL Exponent	Gini Index
Streamers	50	-0.953	0.563
	100	-1.018	0.601
Movie Stars	50	-3.351	0.148
	100	-2.688	0.197
Singers	50	-2.685	0.191
	100	-2.007	0.264
NBA Players	50	-5.655	0.085
	100	-2.424	0.195
Billionaires	50	-1.693	0.289
	100	-1.664	0.332
Celebrities	50	-2.047	0.253
	100	-2.266	0.250

Table 2. **Cross city estimation of the effect of market size on inequality.** This table shows the cross-city-level OLS and IV regression results. In the OLS regression, we run $Gini_j = \alpha + \beta_{OLS} MarketSize_j + \epsilon_j$. In Panel A, the *Gini* coefficients are calculated using the income distribution of streamers who receive more than 50 RMB in tipping income from city j during the sample period. $MarketSize_j$ is the log value of the total virtual gifts sent from citizens in the city j during the sample period, which measures the city's live streaming platform use. In Panel B, *Gini* is calculated using the loyal fans distribution of streamers with at least one loyal fan in city j during the sample period. $MarketSize_j$ is the log number of paying users in city j during the sample period. In the IV regression, in Columns (2) and (3), we use the city's residential population and local GDP as an IV of the market size variable, respectively. In Columns (4) – (7), we use digital financial inclusion to instrument the market size. In Column (4), the aggregate digital financial inclusion index is used, while in Columns (5) – (7), breadth of coverage, depth of use, and level of digitization are used, respectively. The city sub-sample includes 92 cities whose total number of paying users accounts for more than 0.2% of the whole platform's paying users and 18 cities whose total number of paying users accounts for more than 1% of the whole platform's paying users. The city-level population and GDP data come from China Premium Database in CEIC. And we use GDP in the year 2019 to avoid the interference of the Covid-19 pandemic shock. The data source of digital financial inclusion variables is The Peking University Digital Financial Inclusion Index of China in 2018, which utilizes Ant Financial's massive dataset on digital financial inclusion. The index calculation methodology is detailed in Guo et al. (2020). Standard errors are robust and reported in parentheses. K. P. F-statistics are reported in the last row. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

	OLS	IV					
		ln(Population)	ln(GDP)	Aggregate Index	Coverage Breadth	Usage Depth	Digitization Level
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Gini50							
92 cities							
ln(Tipping Amount)	0.023***	0.021***	0.019***	0.019***	0.019***	0.019***	0.020***
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
N	92	92	92	92	92	92	92
Adjusted R2	0.683	0.674	0.664	0.663	0.661	0.666	0.671
K.P.F-statistic		93.158	313.610	119.429	108.527	54.963	47.868
18 cities							
ln(Tipping Amount)	0.020***	0.018***	0.019***	0.014**	0.013	0.014**	0.015***
	(0.002)	(0.004)	(0.003)	(0.007)	(0.010)	(0.006)	(0.006)
N	18	18	18	18	18	18	18
Adjusted R2	0.774	0.760	0.766	0.689	0.672	0.690	0.718
K.P.F-statistic		6.868	56.695	4.913	3.861	6.196	6.330
Panel B: Gini_fans							
92 cities							
ln(Num of Paying Users)	0.071***	0.060***	0.066***	0.082***	0.075***	0.095***	0.085***
	(0.003)	(0.006)	(0.004)	(0.005)	(0.005)	(0.008)	(0.008)
N	92	92	92	92	92	92	92
Adjusted R2	0.718	0.700	0.714	0.704	0.716	0.638	0.692
K.P.F-statistic		164.788	591.956	93.492	79.33	56.828	48.97
18 cities							
ln(Num of Paying Users)	0.072***	0.069***	0.080***	0.102***	0.104***	0.100***	0.106***
	(0.010)	(0.010)	(0.011)	(0.026)	(0.033)	(0.029)	(0.030)
N	18	18	18	18	18	18	18
Adjusted R2	0.730	0.729	0.720	0.584	0.570	0.607	0.552
K.P.F-statistic		24.029	69.695	1.507	0.847	2.345	2.697

Table 3. **Time series estimation of the effect of market size on inequality.** This table reports the time series level OLS and IV results of the effect of market size on inequality measure *Gini* using *QuarterEnd* to instrument the market size. We run the following regressions: $Gini_t = \alpha + \beta_{OLS} MarketSize_t + \eta_t + \epsilon_t$, and $Gini_t = \alpha + \beta_{IV} \widehat{MarketSize}_t + \eta_t + \epsilon_t$, where $Gini_t$ is the Gini coefficient calculated using the income distribution of streamers in month t; $MarketSize_t$ is a series of variables measuring the amount of platform use and η_t represents year-quarter fixed effects. The instrumented variable $QuarterEnd_t$ is a dummy variable that equals one if the month is March, June, September, or December and zeroes otherwise. Panel A reports the OLS results. Panel B reports the 2SLS results. In Columns (1) – (3), $MarketSize_t$ is the log value of the total tipping income of all streamers in month t. In Columns (4) – (6), $MarketSize_t$ is the log value of the average tipping income per streamer in month t. In Columns (1) and (4), the dependent variable is $Gini0$, calculated using the income distribution of the streamers who earn any positive income in month t. In Columns (2) and (5), $Gini_t$ is $Gini50$, which is calculated using the income distribution of the streamers who receive more than 50 RMB in tipping income in month t. In Columns (3) and (6), $Gini500_{jt}$ is calculated based on the income distribution of the streamers who receive more than 500 RMB in tipping income in month t. Standard errors are robust and reported in parentheses. The Kleibergen and Paap F-statistics are reported in the last row in Panel B. The sample includes the 25 months from April 2018 to April 2020. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

	(1) Gini0	(2) Gini50	(3) Gini500	(4) Gini0	(5) Gini50	(6) Gini500
			Panel A: OLS			
ln (Tipping Amount)	0.038***	0.072***	0.107***			
	(0.003)	(0.005)	(0.008)			
ln (Tipping Amount_PC)				0.039***	0.074***	0.111***
				(0.006)	(0.011)	(0.016)
Adjusted R2	0.928	0.940	0.937	0.781	0.762	0.770
			Panel B: IV *QuarterEnd*			
ln (Tipping Amount)	0.041***	0.077***	0.117***			
	(0.003)	(0.006)	(0.010)			
ln (Tipping Amount_PC)				0.046***	0.086***	0.130***
				(0.007)	(0.013)	(0.021)
Adjusted R2	0.924	0.937	0.932	0.766	0.748	0.753
K. P. F-statistic	21.091	21.091	21.091	30.694	30.694	30.694
Year-Quarter FEs	Yes	Yes	Yes	Yes	Yes	Yes
N	25	25	25	25	25	25

Table 4. **Panel estimation of the effect of market size on inequality with 18 cities included in the sample.** This table reports the panel estimation results of the OLS and IV regression: $Gini_{jt} = \alpha + \beta_{OLS} \ln(TippingAmount_{jt}) + \theta_j + \eta_t + \epsilon_{jt}$, and $Gini_{jt} = \alpha + \beta_{IV} \ln(\widehat{TippingAmount}_{jt}) + \theta_j + \eta_t + \epsilon_{jt}$, where $Gini_{jt}$ is the Gini coefficients in the city j in month t; $\ln(TippingAmount_{jt})$ is the log value of the tip spending by citizens of the city j in month t measuring the amount of platform use; θ_j represents city fixed effects, and η_t represents the year-quarter fixed effects. The instrumented variable $QuarterEnd_t$ is a dummy variable that equals one when the month is March, June, September, and December and zeroes otherwise. In Columns (1) and (4), $Gini_{jt}$ is the Gini coefficient for the full sample of the streamers who receive any income from users in the city j in month t. In Columns (2) and (5), $Gini_{jt}$ is the Gini coefficient of the subsample of streamers whose incomes are more than 50 RMB from city j in month t. In Columns (3) and (6), $Gini_{jt}$ is the Gini coefficient of the subsample of streamers whose incomes are more than 500 RMB. Columns (1), (2), and (3) report the OLS results β_{OLS}. Columns (4), (5), and (6) report the 2SLS results β_{IV}. The regression observations are from the 18 cities that account for more than 1% of total paying users during the 25 months from April 2018 to April 2020. Standard errors are clustered at the city and year-quarter level and reported in parentheses. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

	OLS			IV: *QuarterEnd*		
	Gini0	Gini50	Gini500	Gini0	Gini50	Gini500
	(1)	(2)	(3)	(4)	(5)	(6)
ln (Tipping Amount)	0.036***	0.081***	0.135***	0.037***	0.079***	0.127***
	(0.003)	(0.006)	(0.010)	(0.003)	(0.008)	(0.011)
City FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter FEs	Yes	Yes	Yes	Yes	Yes	Yes
N	450	450	450	450	450	450
Adjusted R2	0.871	0.892	0.889	0.841	0.849	0.836
K. P. F-statistic					24.544	

Table 5. **Estimation results using the growth of new users during the capital raising event period to instrument market growth.** This table reports the 2SLS estimation results in Section 5.3. To exclude the interference of the Covid-19 pandemic, we first delete the months after January 2020. Then, we divide the remaining months into three-time stages based on the nodes of the launch of capital raising (i.e., 2019m1) and its closing (i.e., 2019m7). We then refer to the period before 2019m1 as *Stage 1*, the period between 2019m1 and 2019m7 as *Stage 2*, and the period post 2019m7 as *Stage 3*. We run the regressions $\Delta\, Gini_{j,stage} = \alpha + \beta \Delta\ \ln(\widehat{Market\, Size}_{j,stage}) + \epsilon_j$, where $\Delta\, Gini$ is the difference in the average monthly Gini coefficients of the city j between two stages, and $\Delta\ \ln(Market\, Size)$ is the difference in the average monthly size of the tipping amount or the number of paying users between the two stages. In Panel A, $Gini50$ is used to calculate $\Delta\, Gini$, based on the income distribution of streamers whose incomes are more than 50 RMB from city j. $\Delta\ \ln(Tipping\, Amount\_New\, Users)$ is used to instrument the variation of whole market size $\Delta\ \ln(Tipping\, Amount)$. In Panel B, $Gini\_fans$ is used to calculate $\Delta\, Gini$, which is calculated based on the distribution of streamers' loyal fans from city j. $\Delta\ \ln(Number\, of\, New\, Users)$ is used to instrument the variation of whole market size $\Delta\ \ln(Number\, of\, Paying\, Users)$. In this main result table, we use the *New Registered* method to define *New Users* — if a newly registered user in month t converts to a paying user that month, then the user will be defined as *New* in month t. In Columns (1), (2), and (3), We use the value difference of the variables between *Stage 2* and *Stage 1*. And in Columns (4), (5), and (6), We use the value difference of the variables between *Stage 3* and *Stage 2*. And we also use different city samples. Robust standard errors are reported in parentheses. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
	Stage 2 - Stage 1			Stage 3 - Stage 2		
Panel A: $\Delta\, Gini50$						
$\Delta\ \ln(Tipping\, Amount)$	0.093***	0.056***	0.077***	0.065***	0.077***	0.095***
	(0.013)	(0.014)	(0.010)	(0.011)	(0.011)	(0.013)
N	18	92	302	18	92	302
Adjusted R2	0.793	0.561	0.687	0.753	0.807	0.768
K. P. F-statistic	19.884	11.136	34.721	23.304	13.767	28.176
Panel B: $\Delta\, Gini\_fans$						
$\Delta\ \ln(Number\, of\, Paying\, Users)$	0.255**	0.125***	0.135**	0.079*	0.179*	0.129**
	(0.107)	(0.038)	(0.060)	(0.042)	(0.098)	(0.051)
N	18	92	302	18	92	302
Adjusted R-squared	-0.150	0.338	0.193	0.201	0.128	0.027
K. P. F-statistic	13.926	31.068	101.958	59.804	17.134	73.374

Table 6. **Lockdown event-based Gini coefficient response in Wuhan.** This table shows the magnitude of change of the Gini coefficient and market size of Wuhan around Day 9, which refers to the day February 1 when the government implemented a series of escalated measures to deal with the Covid-19 epidemic as the aggravation of the epidemic. The Gini coefficient and market size in Wuhan are adjusted by that of its faraway cities, Chengdu, Shenzhen, Beijing, and Shanghai, to eliminate the interference of confounding factors. In Panel A (B), we use the log value of the tipping amount (log number of paying users) to proxy market size and use $Gini50$ ($Gini\_fans$) as the inequality variable. And the Gini elasticity can be obtained by computing the *after minus before* Gini coefficient change to the Covid-19 market size shrinkage. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

	20 Days before Day 9	20 Days after Day 9	After-Before
Panel A: $Gini50$			
Gini50_Wuhan-Gini50_Chengdu	-0.024	-0.048	-0.023**
ln(Tipping Amount)_Wuhan-ln(Tipping Amount)_Chengdu	0.677	-0.343	-1.021***
Ratio			***0.023***
Gini50_Wuhan-Gini50_Shenzhen	-0.021	-0.047	-0.026***
ln(Tipping Amount)_Wuhan-ln(Tipping Amount)_Shenzhen	0.260	-0.635	-0.894***
Ratio			***0.029***
Gini50_Wuhan-Gini50_Beijing	-0.019	-0.053	-0.034**
ln(Tipping Amount)_Wuhan-ln(Tipping Amount)_Beijing	-0.188	-1.052	-0.864***
Ratio			***0.040***
Gini50_Wuhan-Gini50_Shanghai	-0.034	-0.057	-0.023***
ln(Tipping Amount)_Wuhan-ln(Tipping Amount)_Shanghai	-0.745	-1.445	-0.700***
Ratio			***0.033***
Panel B: $Gini\_fans$			
GiniFans_Wuhan-GiniFans_Chengdu	-0.023	-0.056	-0.033***
ln(Num of Paying Users)_Wuhan-ln(Num of Paying Users)_Chengdu	0.233	-0.361	-0.594***
Ratio			***0.056***
GiniFans_Wuhan-GiniFans_Shenzhen	-0.036	-0.047	-0.011*
ln(Num of Paying Users)_Wuhan-ln(Num of Paying Users)_Shenzhen	0.145	-0.280	-0.425***
Ratio			***0.026***
GiniFans_Wuhan-GiniFans_Beijing	0.007	-0.020	-0.027***
ln(Num of Paying Users)_Wuhan-ln(Num of Paying Users)_Beijing	-0.103	-0.588	-0.485***
Ratio			***0.055***
GiniFans_Wuhan-GiniFans_Shanghai	-0.019	-0.041	-0.022***
ln(Num of Paying Users)_Wuhan-ln(Num of Paying Users)_Shanghai	-0.422	-0.845	-0.423***
Ratio			***0.052***

Table 7. **The change of Gini50 to Covid-19 shock.** This table reports the change of Gini coefficients to Covid-19 shock. We run the regression: $Gini50_{jt} = \alpha + \beta ln(Distance_j) \times Post_t + \lambda Post_t + \delta X_{jt} + \theta_j + \epsilon_{jt}$. $Gini50_{jt}$ is calculated using the income distribution of streamers who receive virtual gifts worth more than 50 RMB from users in the city j from Day $(t-6)$ to Day t. In Columns (1) to (5), $Distance_j$ is the driving time (in days) between Wuhan and city j, as shown on Baidu Maps. In Columns (6) to (10), $Distance_j$ is the straight-line distance (in thousands of kilometers) calculated from the latitude and longitude of Wuhan and city j. $Post_1 = 1$ indicates the days after December 30, 2019 (Day -24) when Dr. Wenliang Li posted a Covid-19 alert on one of his WeChat groups and Covid-19 first became known to the public. $Post_2 = 1$ indicates the days after January 23, 2020 (Day 0), when the government imposed a lockdown in Wuhan. $Post_3 = 1$ indicates the days after February 1, 2020 (Day 9), when the government started implementing a series of escalating measures. We do not control any market size variables in Columns (1) and (6). In Columns (2) and (7), we use the log number of paying users as the measure of market size variable X_{jt}. In Columns (3) and (8), we use the log value of the tipping amount as the measure of market size variable X_{jt}. In Columns (4) and (9), we use the log number of streamers as the measure of the market size variable X_{jt}. In Columns (5) and (10), we add a series of market size variables to the regression. θ_j represents city-level fixed effects. This regression estimation is based on the 79 days (from Day -50 to Day 28) around the Wuhan lockdown. The city sample includes nine cities that account for more than 2% of the number of users, respectively: Shanghai, Beijing, Guangzhou, Chengdu, Hangzhou, Wuhan, Shenzhen, Chongqing, and Changsha. Standard errors clustered at the city level are reported in parentheses. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

	(1) Gini50	(2) Gini50	(3) Gini50	(4) Gini50	(5) Gini50	(6) Gini50	(7) Gini50	(8) Gini50	(9) Gini50	(10) Gini50
Panel A: Post (Wenliang Li's Alert)										
ln (Distance) * $Post_1$	0.060	0.025	0.036	0.036	0.020	0.033	0.012	0.020	0.018	0.010
	(0.047)	(0.051)	(0.025)	(0.047)	(0.032)	(0.029)	(0.031)	(0.016)	(0.029)	(0.020)
$Post_1$	-0.033	-0.008	-0.014	-0.007	-0.002	-0.029	-0.006	-0.012	-0.004	0.000
	(0.020)	(0.020)	(0.010)	(0.016)	(0.010)	(0.020)	(0.019)	(0.010)	(0.016)	(0.010)
ln (Num of Paying Users)		0.070***			0.035**		0.070***			0.035**
		(0.012)			(0.011)		(0.012)			(0.011)
ln (Tipping Amount)			0.037***		0.028***			0.038***		0.028***
			(0.004)		(0.004)			(0.004)		(0.005)
ln (Num of Streamers)				0.127***	0.024				0.128***	0.024
				(0.029)	(0.032)				(0.030)	(0.032)
Adjusted R-squared	0.263	0.415	0.460	0.365	0.504	0.260	0.414	0.459	0.364	0.503
Panel B: Post (Wuhan Lockdown)										
ln (Distance) * $Post_2$	0.097**	0.060	0.056	0.065	0.047	0.062**	0.039	0.036	0.042	0.031
	(0.040)	(0.045)	(0.033)	(0.038)	(0.031)	(0.024)	(0.028)	(0.021)	(0.023)	(0.019)
$Post_2$	-0.060***	-0.034*	-0.036**	-0.042***	-0.029**	-0.059***	-0.034*	-0.036**	-0.041***	-0.028**
	(0.015)	(0.015)	(0.012)	(0.011)	(0.010)	(0.015)	(0.015)	(0.013)	(0.012)	(0.010)
ln (Num of Paying Users)		0.049***			0.029***		0.050***			0.029***
		(0.011)			(0.008)		(0.011)			(0.008)
ln (Tipping Amount)			0.031***		0.027***			0.031***		0.027***
			(0.004)		(0.004)			(0.004)		(0.005)
ln (Num of Streamers)				0.059	-0.014				0.060	-0.014
				(0.033)	(0.032)				(0.033)	(0.032)
Adjusted R-squared	0.387	0.443	0.506	0.405	0.518	0.386	0.443	0.506	0.405	0.518
Panel C: Post (Escalated Measures)										
ln (Distance) * $Post_3$	0.092*	0.018	0.024	-0.001	-0.012	0.062*	0.017	0.019	0.005	-0.001
	(0.044)	(0.062)	(0.046)	(0.075)	(0.060)	(0.028)	(0.041)	(0.030)	(0.048)	(0.039)
$Post_3$	-0.042**	-0.007	-0.011	-0.001	0.006	-0.042**	-0.010	-0.012	-0.004	0.002
	(0.016)	(0.022)	(0.017)	(0.028)	(0.022)	(0.016)	(0.023)	(0.017)	(0.027)	(0.022)
ln (Num of Paying Users)		0.069***			0.037***		0.068***			0.037***
		(0.015)			(0.010)		(0.015)			(0.009)
ln (Tipping Amount)			0.037***		0.028***			0.037***		0.028***
			(0.003)		(0.004)			(0.003)		(0.004)
ln (Num of Streamers)				0.118***	0.017				0.116***	0.015
				(0.033)	(0.025)				(0.033)	(0.025)
Adjusted R-squared	0.266	0.414	0.458	0.359	0.500	0.267	0.415	0.459	0.359	0.500
N	711	711	711	711	711	711	711	711	711	711
City FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

For Online Publication

March 21, 2023

A Appendix

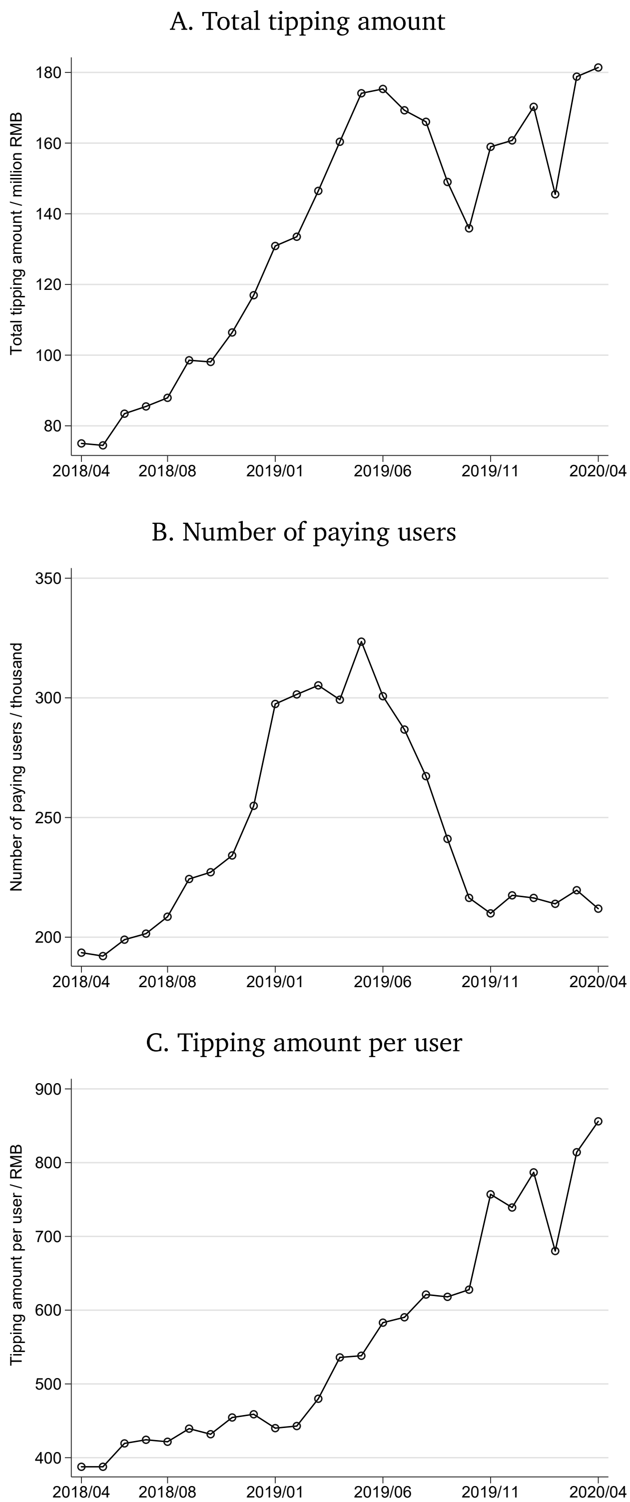


Figure A1. **Seasonality-adjusted tipping amount.** This figure shows the dynamics of seasonality-adjusted market size during the sample period. We calculate the average of the past three months to adjust for seasonality. Panel A plots the total amount of the virtual gifts received by the streamers in the sample (in a million RMB). We further decompose this tipping amount into the number of paying users (in thousand) in Panel B and the tipping amount per user (in RMB) in Panel C.

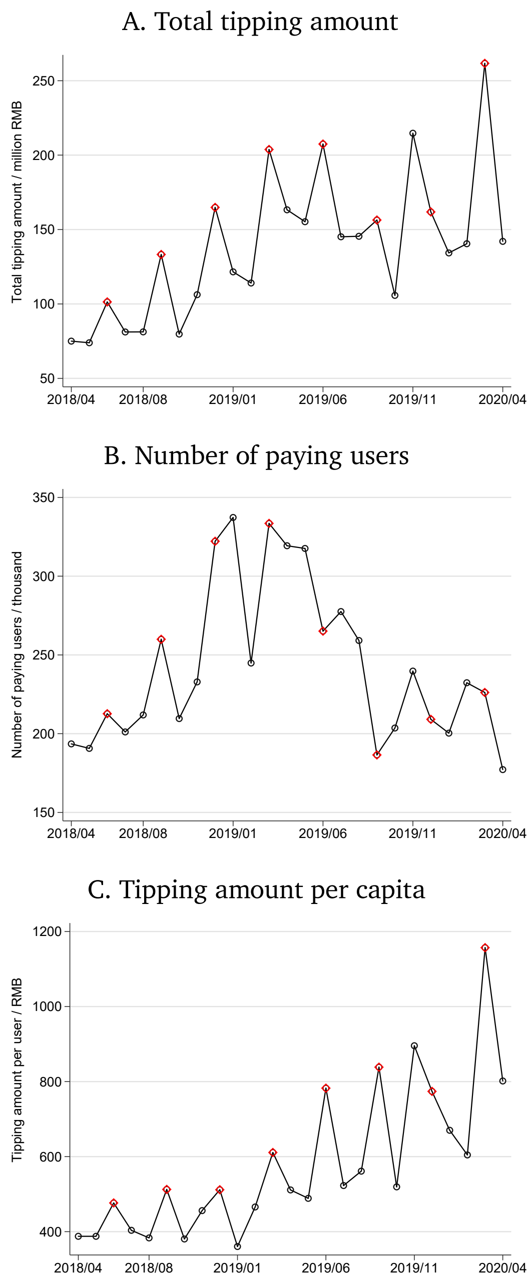


Figure A2. **Tipping amount.** This figure shows the dynamics of raw market size during the sample period. Panel A plots the total value of the virtual gifts received by the streamers in the sample (in million RMB). We further decompose this tipping amount into the number of paying users (in thousand) in Panel B and the tipping amount per user (in RMB) in Panel C. The red dots mark the months at the end of each quarter (March, June, September, and December).

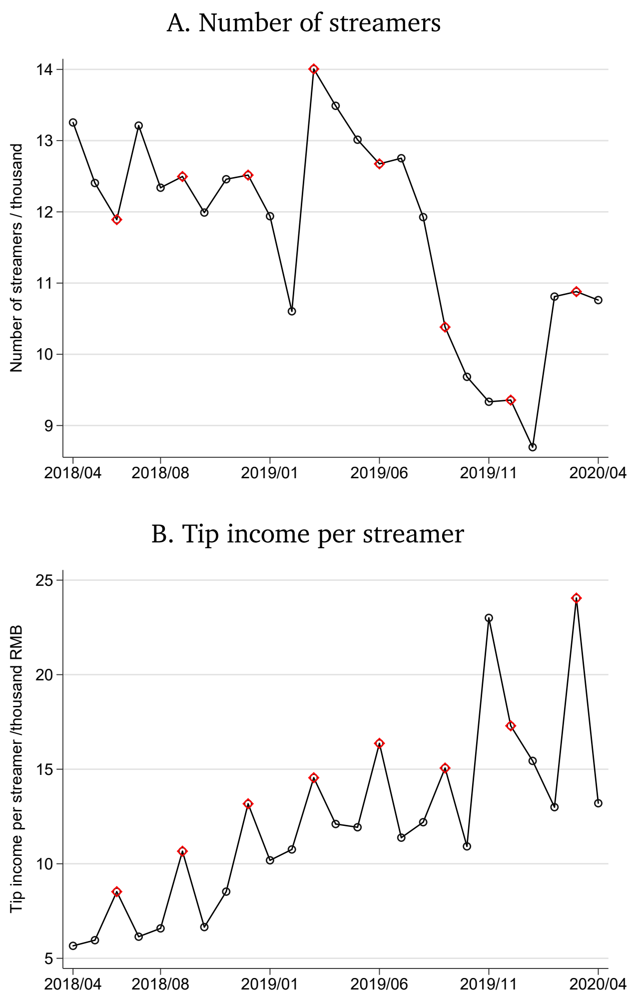


Figure A3. **Decompose raw tipping amount from the streamer side.** This figure decomposes the raw tipping amount from the streamer side. Panel A plots the number of streamers (in thousand) who can earn positive tipping income each month. Panel B shows the average tipping income (in thousand RMB) received by these streamers. The red dots mark the months at the end of each quarter (March, June, September, and December).

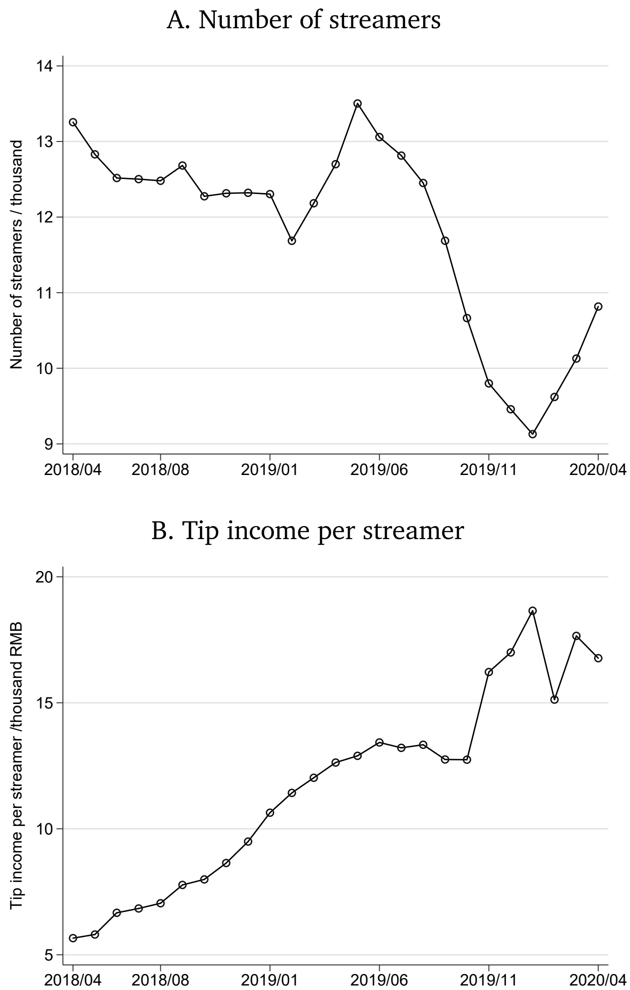


Figure A4. **Decompose seasonality-adjusted tipping amount from the streamer side.** This figure decomposes the seasonality-adjusted tipping amount from the streamer side. Panel A plots the number of streamers (in thousand) who can earn positive tipping income over the past three months. Panel B plots the average tipping income (in thousand RMB) received by these streamers.

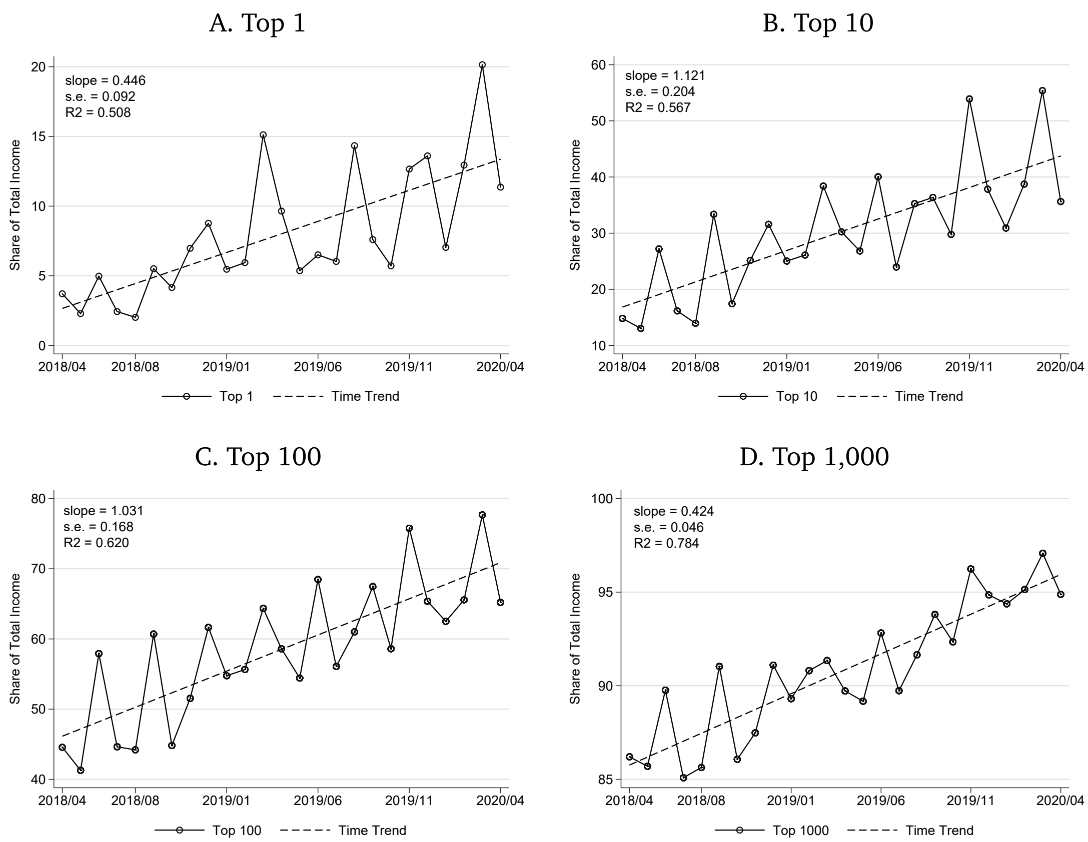


Figure A5. **Income share of top streamers.** This figure ranks the streamers according to their monthly income and calculates the top groups' share of total tipping income. The upper left corner of each sub-figure is marked with the slope, standard error, and R^2 of the linear fit line.

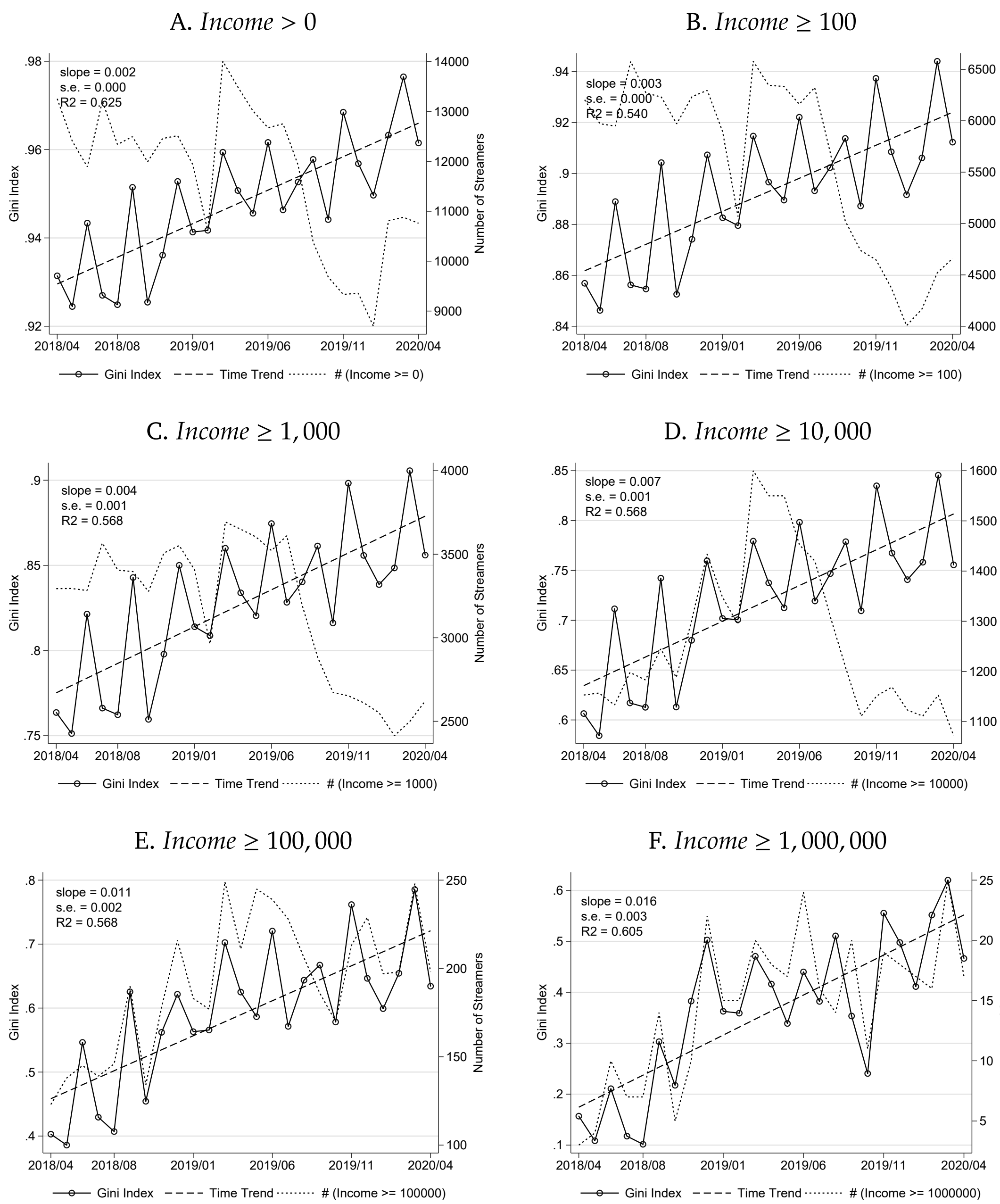


Figure A6. **Gini index.** This figure shows the inequality trend as measured by the Gini index, which is calculated using the distribution of streamers' monthly tipping income. In Panel A, the dotted line shows the number of streamers with positive monthly tipping income, the solid line shows the Gini coefficient calculated based on this sample, and the dashed line is the linear fit time trend of the Gini index. The other five panels are similar, except that the sample is replaced by streamers who earn more than 100, 1,000, 10,000, 100,000, and 1,000,000 RMB in a month, respectively. The upper left corner of each sub-figure is marked with the slope, standard error, and R^2 of the linear fit line.

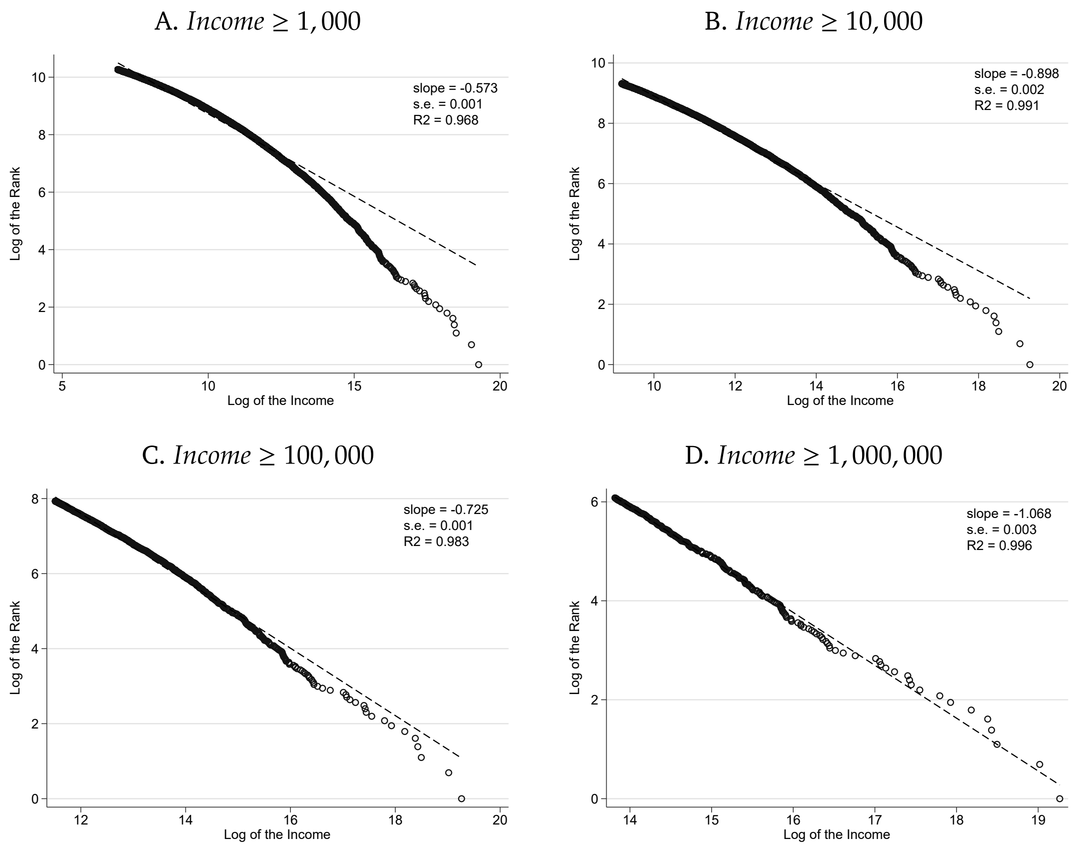


Figure A7. **Log tipping income versus log rank of the streamers.** Streamers are ordered according to the total value of their received virtual gifts during the whole sample period. The log value of rank is placed on the y-axis, and the log of streamers' total tipping income is placed on the x-axis. From Panel A to Panel D, the sample used is streamers with total tipping income greater than 1,000 and 10,000, 100,000, and 1 million RMB, respectively. The slope, standard error, and R^2 of the linear fit line are presented in the upper right corner of each sub-figure.

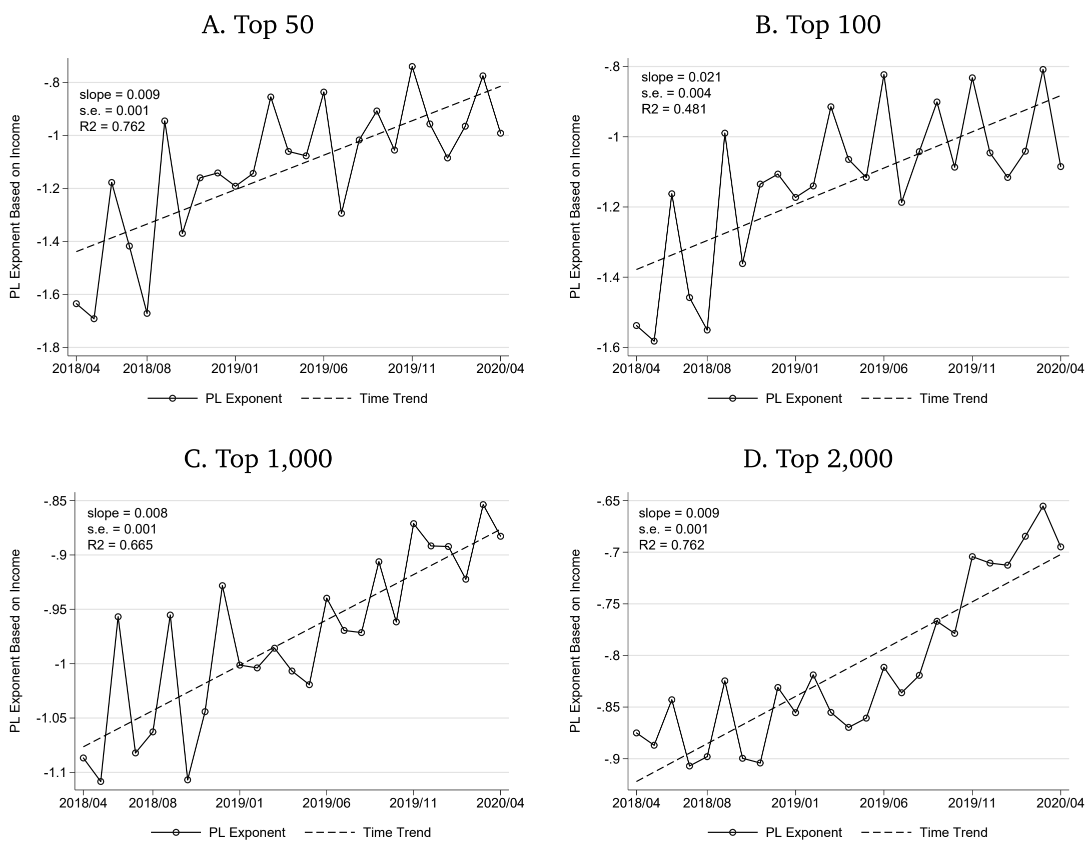


Figure A8. **PL exponent.** This figure shows the dynamic of inequality as measured by the PL exponent. We rank the streamers according to their monthly income and calculate the PL exponent using the resulting income distribution of those top streamers. Specifically, we run the linear regressions $\ln rank_{j,t} = \alpha_t + \beta_t \ln size_{j,t} + \varepsilon_{j,t}, \forall t$, and β_t is the estimated PL exponent in month t. From Panel A to Panel D, the sample used are top 50, top 100, top 1,000, and top 2,000 streamers, respectively. The solid line plots the β_t, and the dashed line is the linear fit time trend of PL exponents. The slope, standard error, and R^2 of the linear fit time trend are presented in the top left corner of each sub-figure.

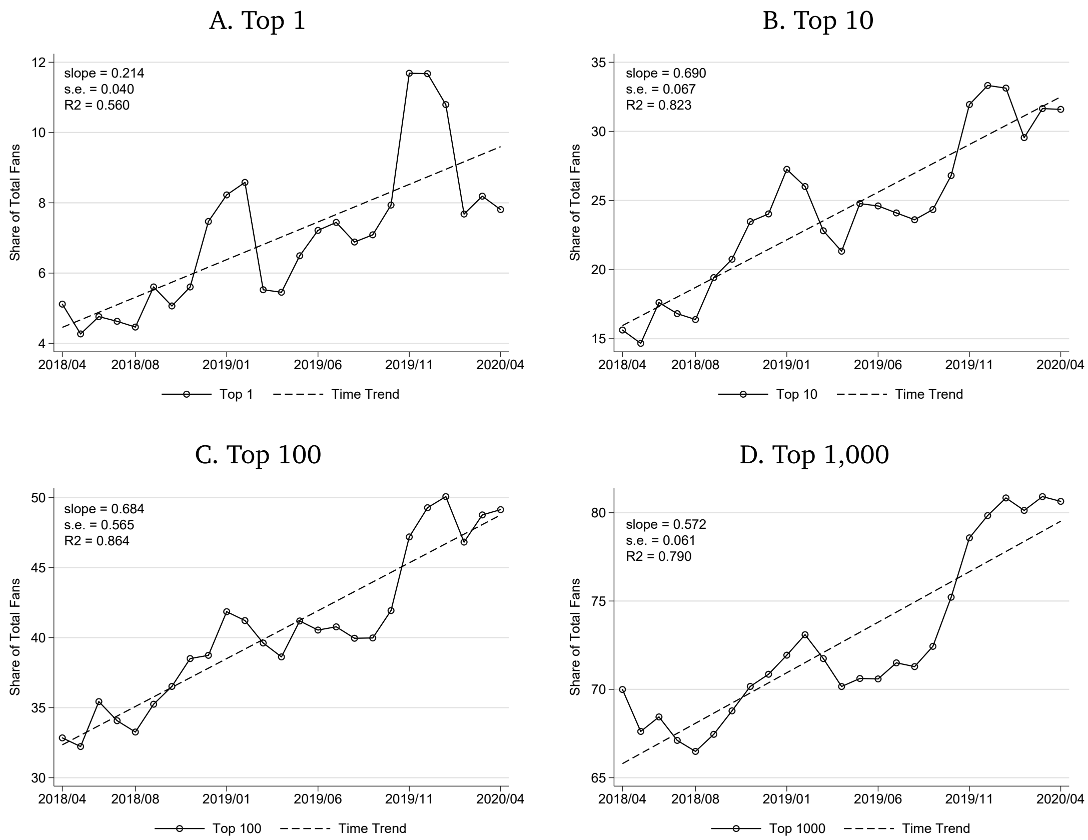


Figure A9. **Seasonality-adjusted fans share of top streamers.** In this figure, we rank the streamers by their aggregate number of loyal fans in the past three months, adjusted for seasonality, and we calculate the ratio of the number of loyal fans of the top streamer group to the total number of fans. Suppose there are N streamers. The method to determine whether $user_i$ is a loyal fan of $streamer_j$ in month t is as follows: First, $user_i$ must give virtual gifts whose value is not less than 50 RMB in month t. Second, $user_i$'s tipping to $streamer_j$ should account for the highest proportion of her total tipping expenditure in month t, that is, $j = \arg\max_n \frac{Tip_{int}}{\sum_{n=1}^{N} Tip_{int}}$. From Panel A to Panel D, fans shares of the top 1, top 10, top 100, and top 1,000 streamers are plotted, respectively. The slope, standard error, and R^2 of the linear fit time trend are presented in the top left corner of each sub-figure.

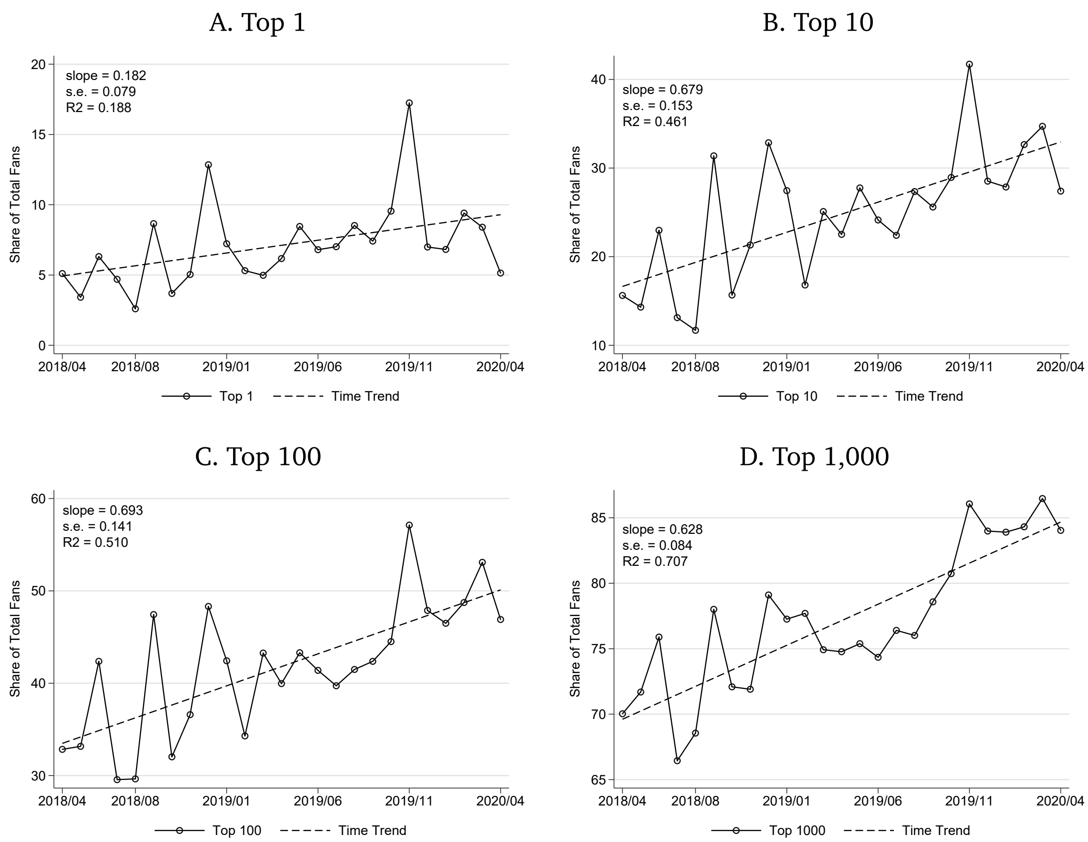
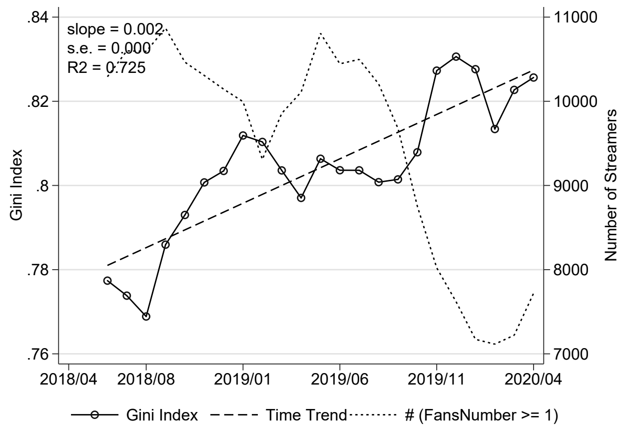
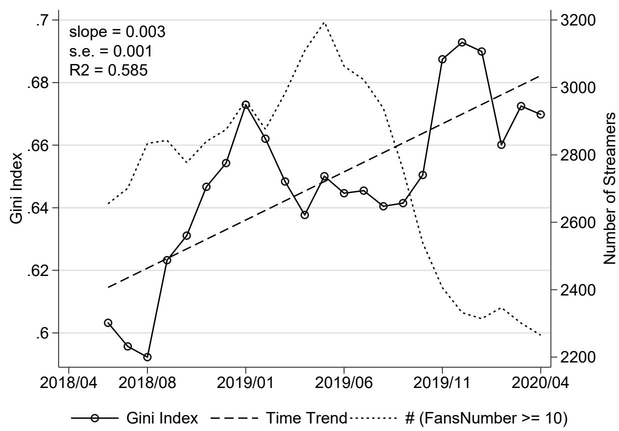


Figure A10. **Fans share of top streamers.** In this figure, we rank the streamers by their monthly number of loyal fans, and we calculate the ratio of the number of loyal fans of the top streamer group to the total number of fans. Suppose there are N streamers. The method to determine whether $user_i$ is a loyal fan of $streamer_j$ in month t is as follows: First, $user_i$ must give virtual gifts whose value is not less than 50 RMB in month t. Second, $user_i$'s tipping to $streamer_j$ should account for the highest proportion of her total tipping expenditure in month t, that is, $j = \arg\max_{n} \frac{Tip_{int}}{\sum_{n=1}^{N} Tip_{int}}$. From Panel A to Panel D, fans shares of the top 1, top 10, top 100, and top 1,000 streamers are plotted, respectively. The slope, standard error, and R^2 of the linear fit time trend are presented in the top left corner of each sub-figure.

A. $Number\ of\ fans \geq 1$



B. $Number\ of\ fans \geq 10$



C. $Number\ of\ fans \geq 100$

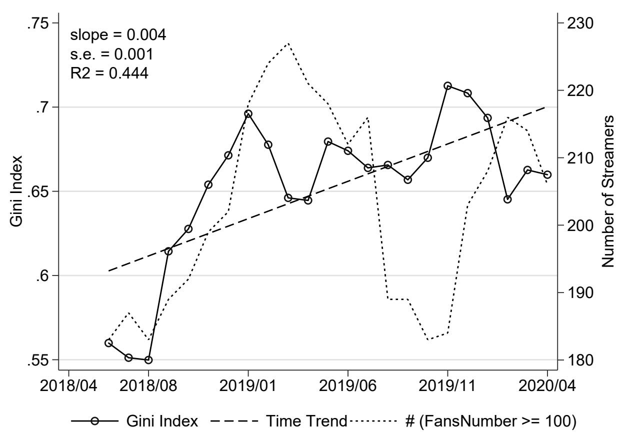
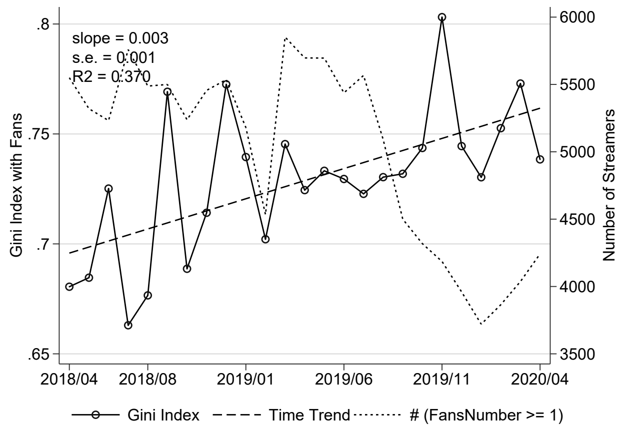
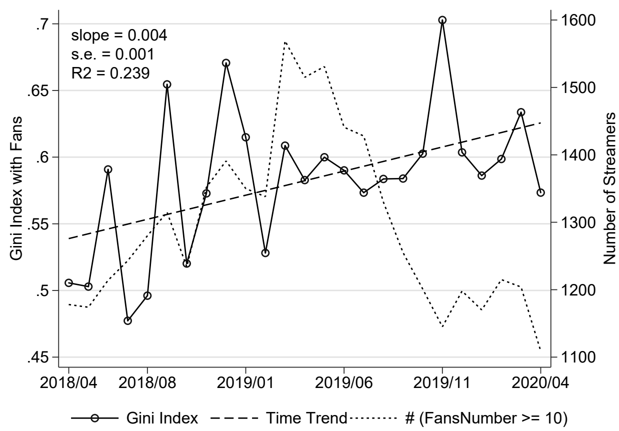


Figure A11. **Seasonality-adjusted Gini index with fans.** This figure shows the seasonality-adjusted inequality trend as measured by the Gini index. We aggregate streamers' loyal fans over the past three months and calculate the Gini index using the resulting fans distribution. In Panel A, the dotted line shows the number of streamers with any loyal fans, the solid line shows the Gini coefficient calculated based on this sample, and the dashed line is the linear fit time trend of the Gini index. The other panels are similar, except that the sample is replaced by streamers who have more than 10 and 100 loyal fans over the past three months, respectively. The upper left corner of each sub-figure is marked with the slope, standard error, and R^2 of the linear fit line.

A. $Number\ of\ fans \geq 1$



B. $Number\ of\ fans \geq 10$



C. $Number\ of\ fans \geq 100$

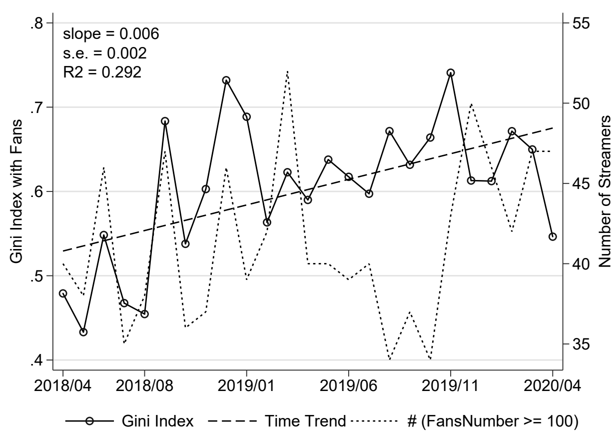


Figure A12. **Gini index with fans.** This figure shows the inequality trend as measured by the Gini index. We calculate the Gini index using streamers' monthly fans distribution. In Panel A, the dotted line shows the number of streamers with any loyal fans, the solid line shows the Gini coefficient calculated based on this sample, and the dashed line is the linear fit time trend of the Gini index. The other panels are similar, except that the sample is replaced by streamers who have more than 10 and 100 loyal fans in a month, respectively. The upper left corner of each sub-figure is marked with the slope, standard error, and R^2 of the linear fit line.

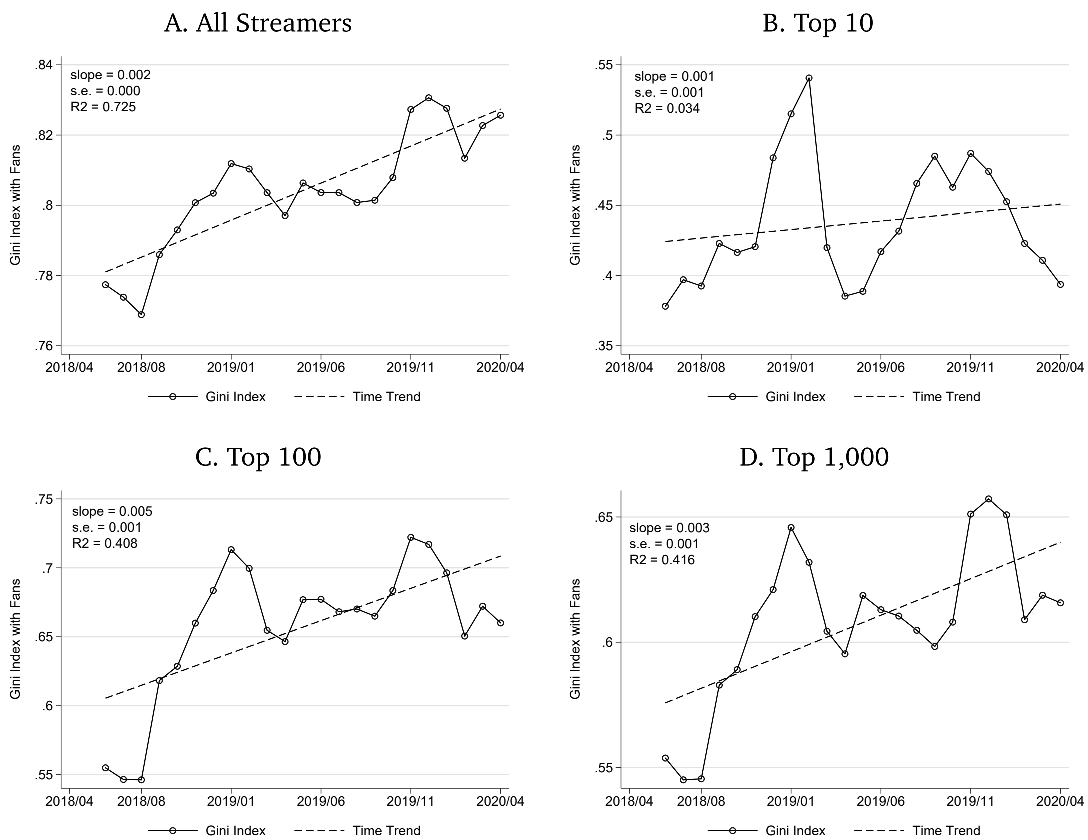


Figure A13. **Seasonality-adjusted Gini index with fans among top streamers.** This figure shows the seasonality-adjusted inequality trend as measured by the Gini index within top streamers. We aggregate streamers' loyal fans over the past three months, rank them according to their aggregate number of loyal fans, and calculate the Gini index using the resulting fan distribution of those top streamers. In Panel A, as a benchmark, the Gini index is calculated based on the streamers sample who have at least one loyal fan in the past three months. The Gini index is calculated from Panel B to Panel D based on the top 10, top 100, and top 1,000 streamers. In each sub-figure, the dashed line is the linear fit time trend of the Gini index, and the upper left corner is marked with the slope, standard error, and R^2 of this linear fit line.

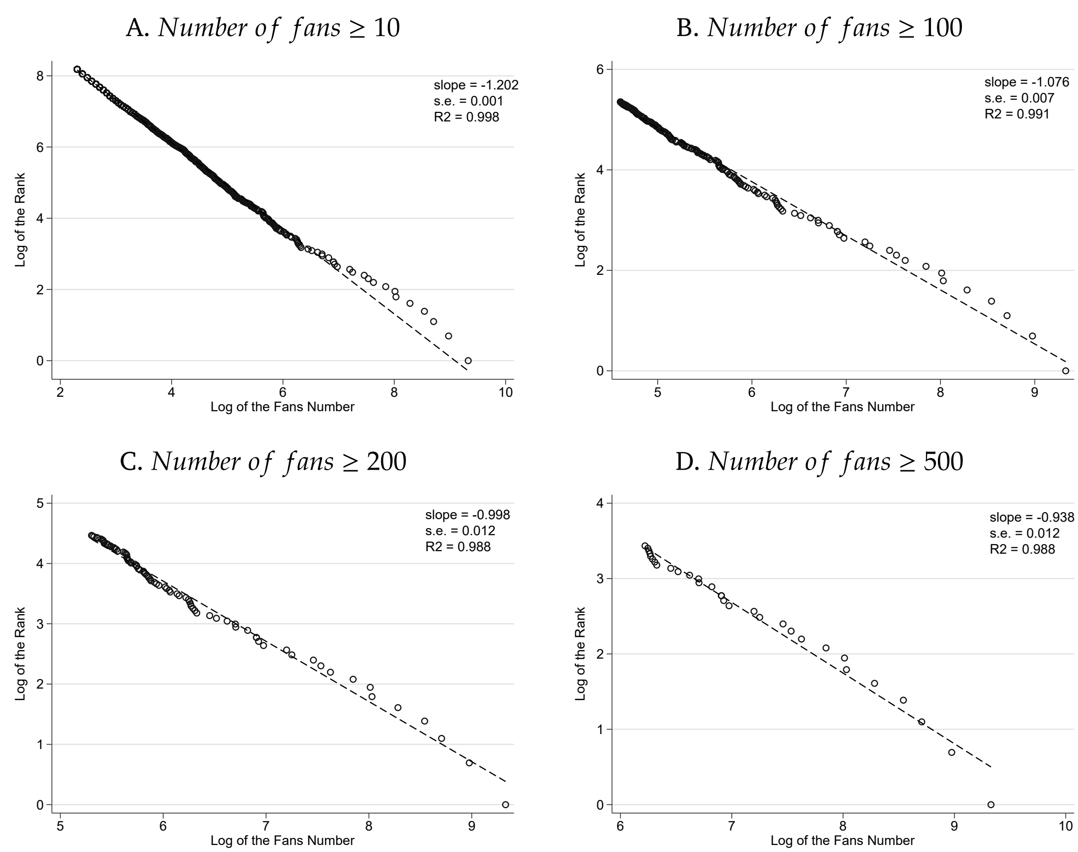


Figure A14. **Log number of fans versus log rank of the streamers.** Streamers are ordered according to the total number of their loyal fans during the whole sample period. The log value of rank is placed on the y-axis, and the log of streamers' total number of loyal fans is placed on the x-axis. From Panel A to Panel D, the sample used is streamers with loyal fans more than 10 and 100, 200, and 500, respectively. The slope, standard error, and R^2 of the linear fit line are presented in the upper right corner of each sub-figure.

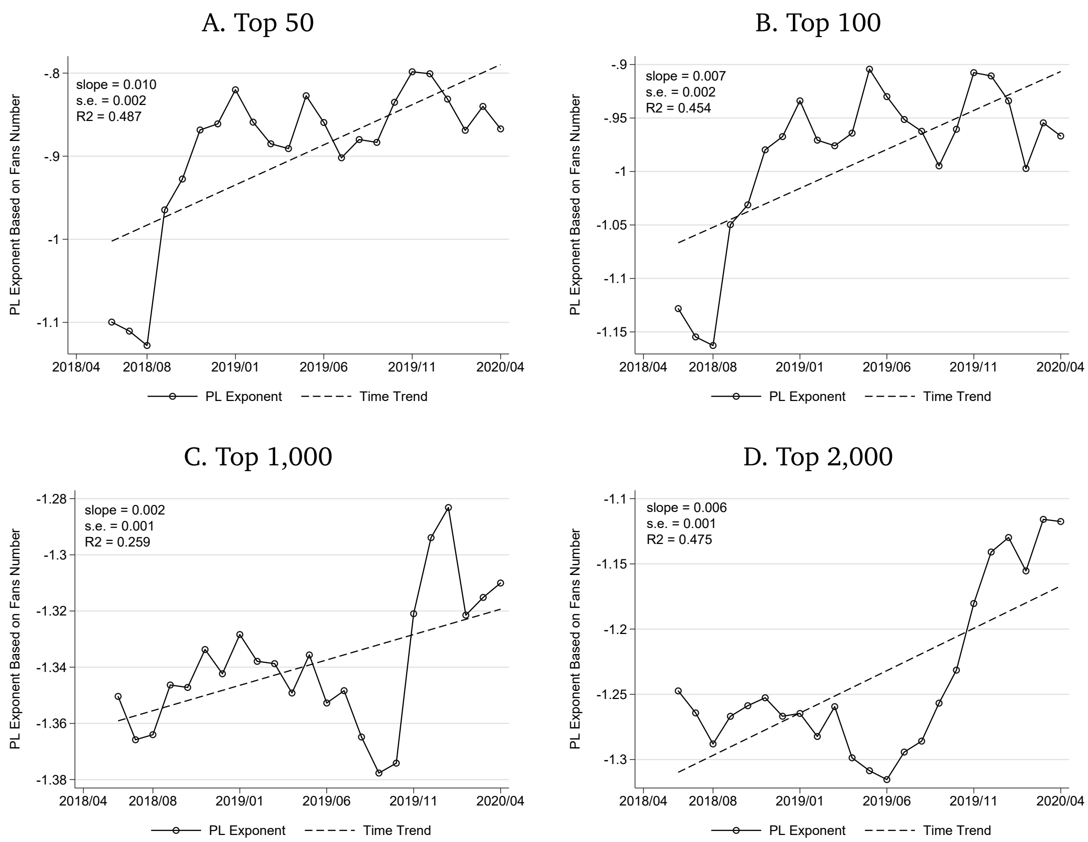


Figure A15. **Seasonally adjusted PL exponent with fans.** This figure shows the dynamic of inequality as measured by seasonality-adjusted PL exponent with fans. We aggregate streamers' number of loyal fans over the past three months, rank them according to their aggregate number of loyal fans, and calculate the PL exponent using the resulting fans distribution of those top streamers. Specifically, we run the linear regressions $\ln rank_{j,t} = \alpha_t + \beta_t \ln size_{j,t} + \varepsilon_{j,t}, \forall t$, and β_t is the estimated PL exponent in month t. From Panel A to Panel D, the sample used are top 50, top 100, top 1,000, and top 2,000 streamers, respectively. The solid line plots the β_t, and the dashed line is the linear fit time trend of PL exponents. The slope, standard error, and R^2 of the linear fit time trend are presented in the top left corner of each sub-figure.

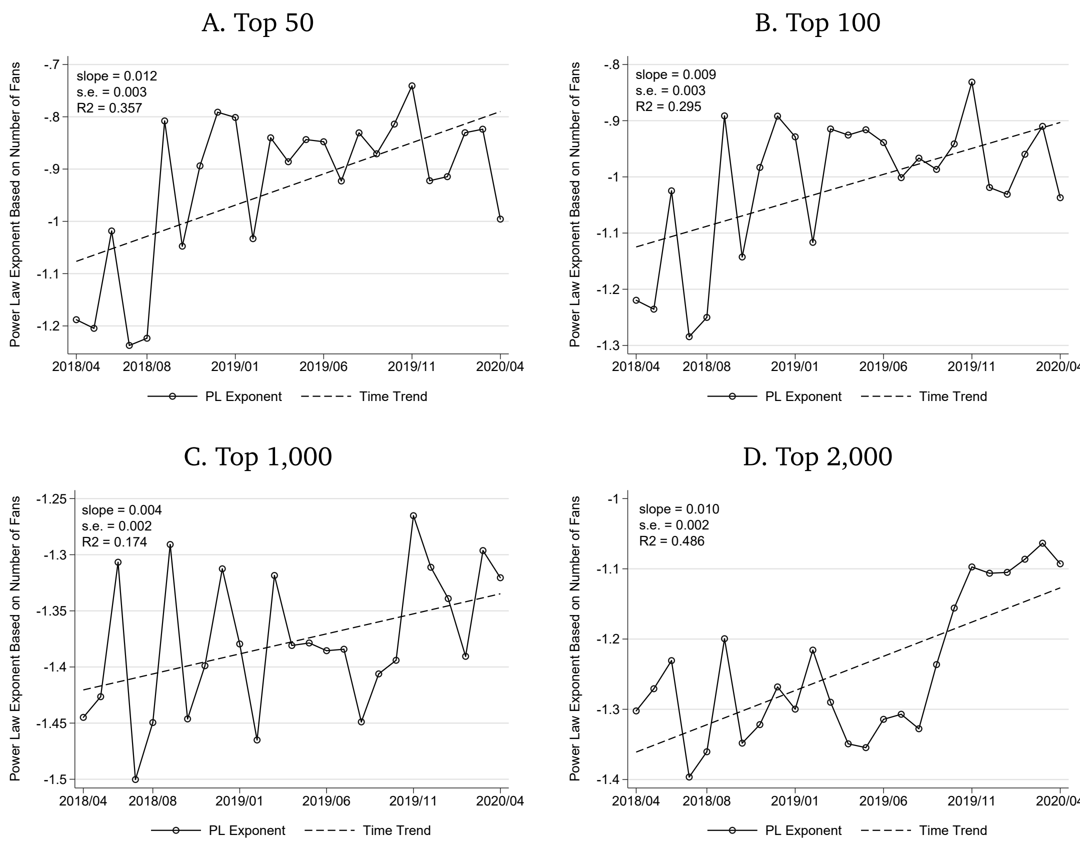


Figure A16. **PL exponent with fans.** This figure shows the dynamic of inequality as measured by the PL exponent with fans. We rank the streamers according to their monthly number of loyal fans and calculate the PL exponent using the resulting fans distribution of those top streamers. Specifically, we run the linear regressions $\ln rank_{j,t} = \alpha_t + \beta_t \ln size_{j,t} + \varepsilon_{j,t}, \forall t$, and β_t is the estimated PL exponent in month t. From Panel A to Panel D, the sample used are top 50, top 100, top 1,000, and top 2,000 streamers, respectively. The solid line plots the β_t, and the dashed line is the linear fit time trend of PL exponents. The slope, standard error, and R^2 of the linear fit time trend are presented in the top left corner of each sub-figure.

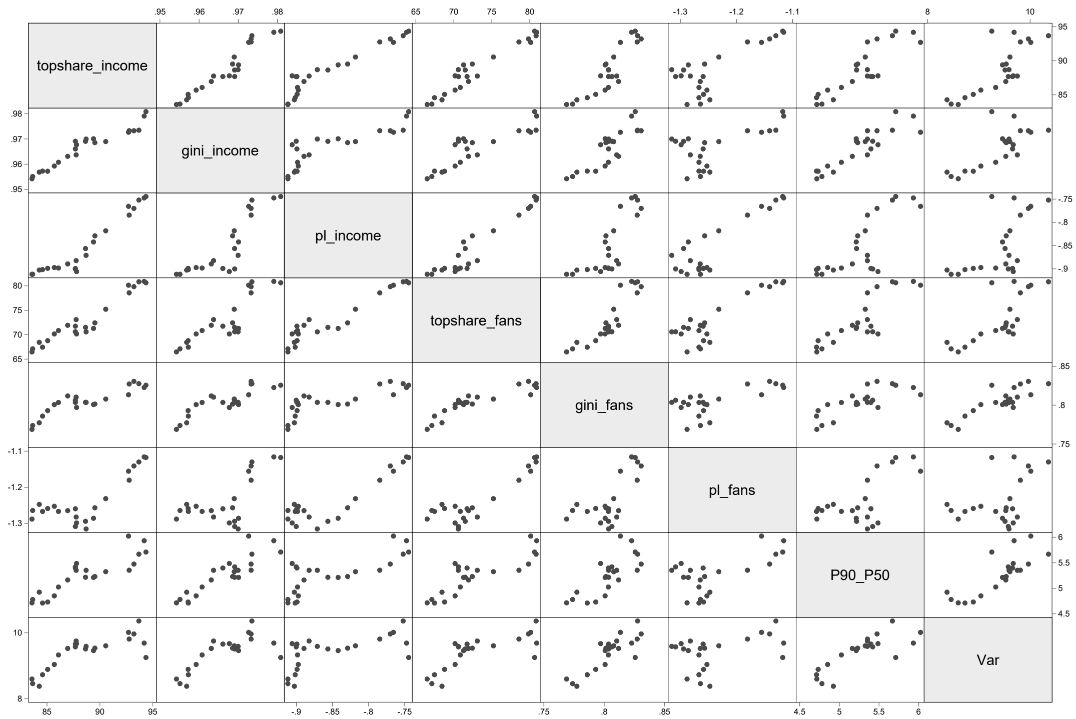
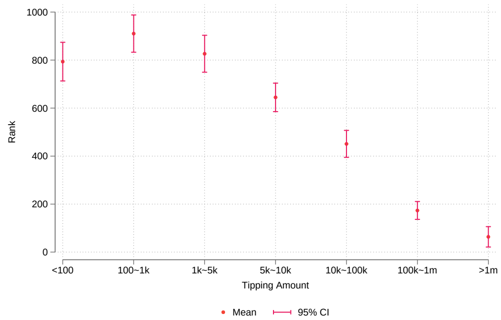


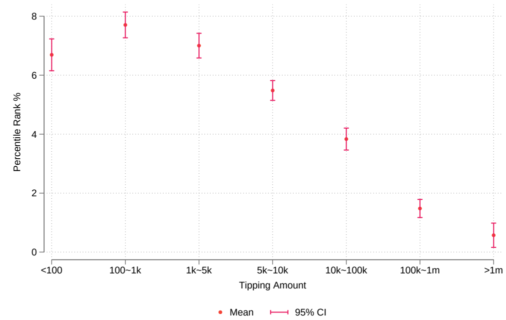
Figure A17. **Correlation matrix.** This figure reports the correlations between the various measures of inequality calculated in Section 3. The specific definitions of the variables shown in the figure are as follows: *topshare_income* refers to the income share of the top 1,000 streamers ordered by their total income in the past three months; *gini_income* is calculated using a sample of streamers who receive positive virtual gifts in the past three months; *pl_income* is calculated using the top 2,000 streamers ordered by their three-month tipping income; *topshare_fans* refers to the loyal fans share of top 1,000 streamers ordered by their total number of loyal fans in the past three months; *gini_fans* is calculated using a sample of streamers with at least one loyal fans in the past three months; *pl_fans* is calculated using a sample of the top 2,000 streamers ordered by their three-month loyal fans number; *P*90_*P*50 is the percentile gap of log tipping income of streamers in the past three months; and *Var* is the variance of log tipping income of streamers in the past three months.

A. The rank level of streamers that users spend most money to

A1. Absolute rank

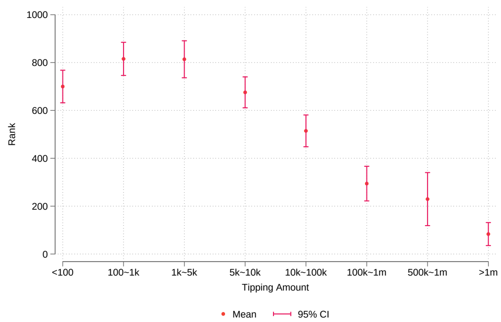


A2. Percentile rank %

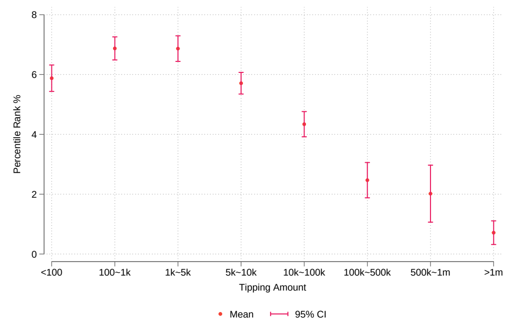


B. The rank level of streamers that users spend the most time to

B1. Absolute rank

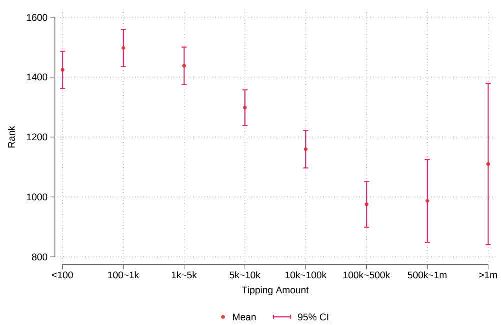


B2. Percentile rank %



C. The rank level of streamers that users newly follow

C1. Absolute rank



C2. Percentile rank %

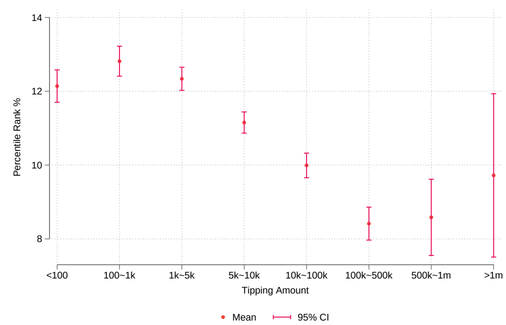


Figure A18. **The rank level of streamers that users appreciate with different consumption levels.** This figure provides evidence of whether active users who spend a lot would disproportionately appreciate the top streamers who already have cumulative advantages. For each month t, we divide users into eight groups based on their monthly tipping amount. Absolute rank is the average standing of the streamers while percentile rank is the average relative position of the streamers that users appreciate. The smaller one streamer's ranking, the higher her tier and the greater her cumulative advantage. In Panel A, we focus on the streamer that users pay most money to and explore which tier of streamers the users' money resources are mainly allocated to. In Panel B, we identify users' favorite streamers based on their watching time in each live broadcast room and explore the ranking of streamers that users spend most time to. In Panel C, we analyze the streamers that users newly add to her follow list and investigate which tier of streamers the users' attention flow to. The 95% confidence interval is based on the Month – User level clustered standard errors. This figure show the raw results without controlling for user level fixed effects.

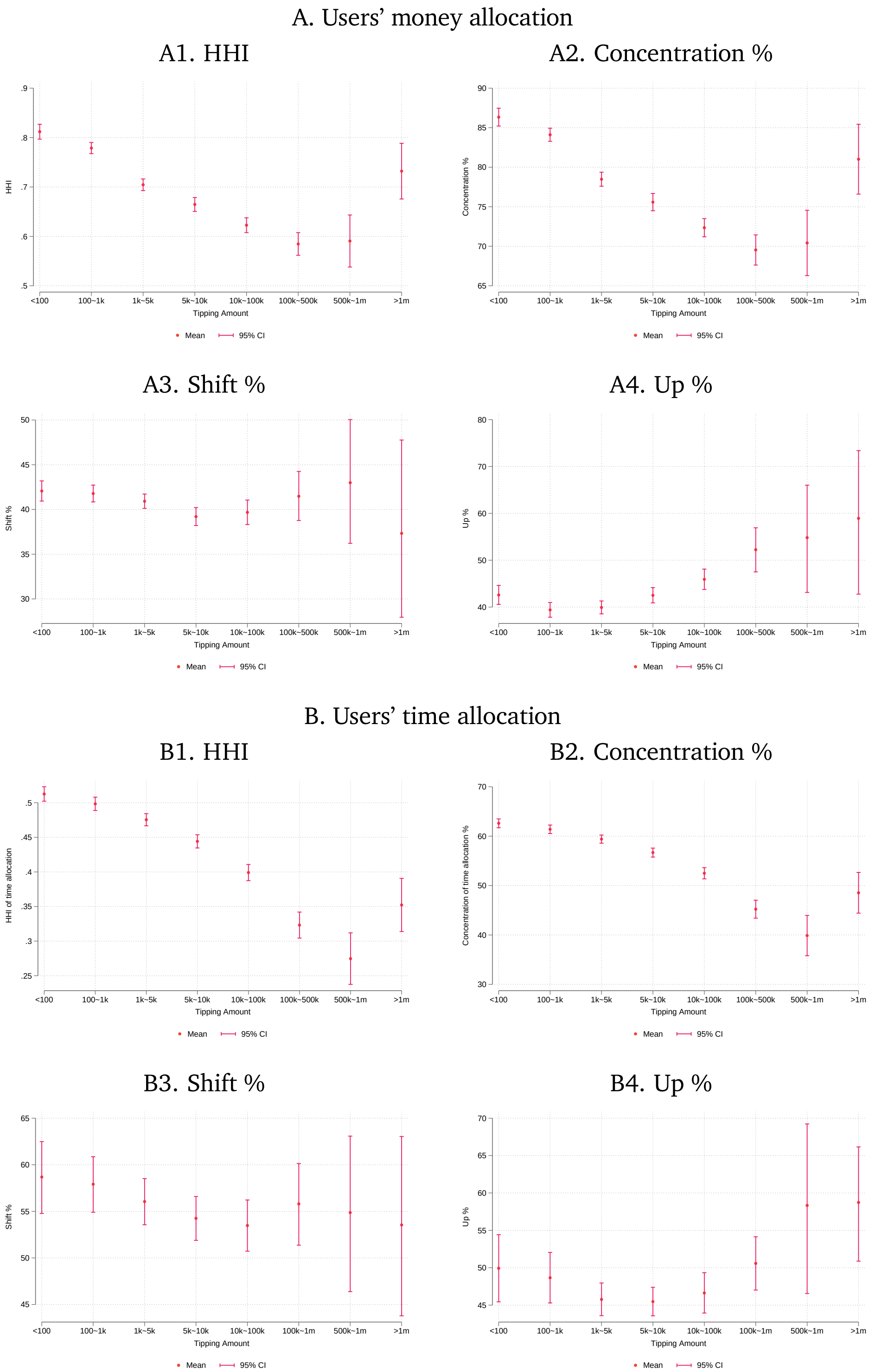


Figure A19. **Loyalty of users with different consumption levels.** This figure provides evidence whether users with high consumption levels would have exceptionally high loyalty. When measuring loyalty, we focus on users' money allocation in Panel A, while in Panel B, we investigate users' viewing time allocation. For the x-axis, we divide users into eight groups based on their monthly tipping amount for each month t. For y-axis, *Concentration*, *HHI*, $Shift$ and *Up* are all proxy variables of loyalty of users. The higher concentration or HHI value, the higher the loyalty of users to specific streamers. $Shift$ is the probability that the streamer that users are loyal to changes to someone else. The lower the $Shift$ value, the higher the loyalty of users to specific streamers. And *Up* is the probability of users shifting toward a higher-tier streamer. The higher the probability, the more likely users would experience consumption upgrading to higher-rank streamer. The 95% confidence interval is based on the Month – User level clustered standard errors. This figure shows the raw results without controlling for user level fixed effects.

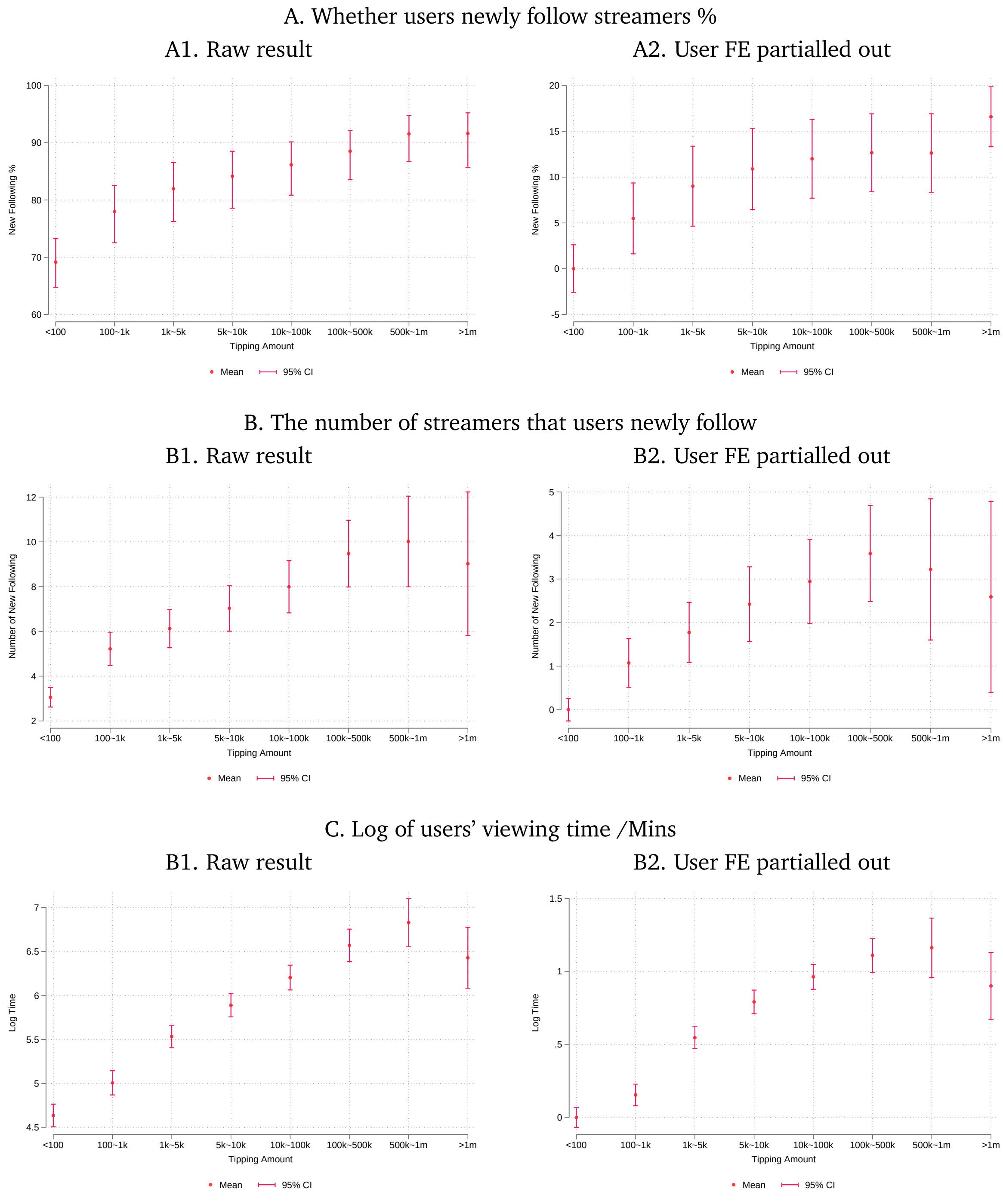


Figure A20. **Engagement of users with different consumption levels.** This figure illustrates whether users with deep pockets would also have high engagement. For each month t, we divide users into eight groups based on their monthly tipping amount. Panel A and Panel B show how active users are in exploring new digital contents supplied by streamers, while Panel C focuses on a more direct dimension, users' viewing time. The term *newly follow* is used here to refer to a user adds some new streamers to the follow list in the month. And the term *viewing time* is defined as the monthly sum of the duration between the user entering one live room and exiting the live room. The 95% confidence interval is based on the Month – User level clustered standard errors. The subfigures in A1, B1 and C1 are raw results, while the subfigures in A2, B2, and C2.

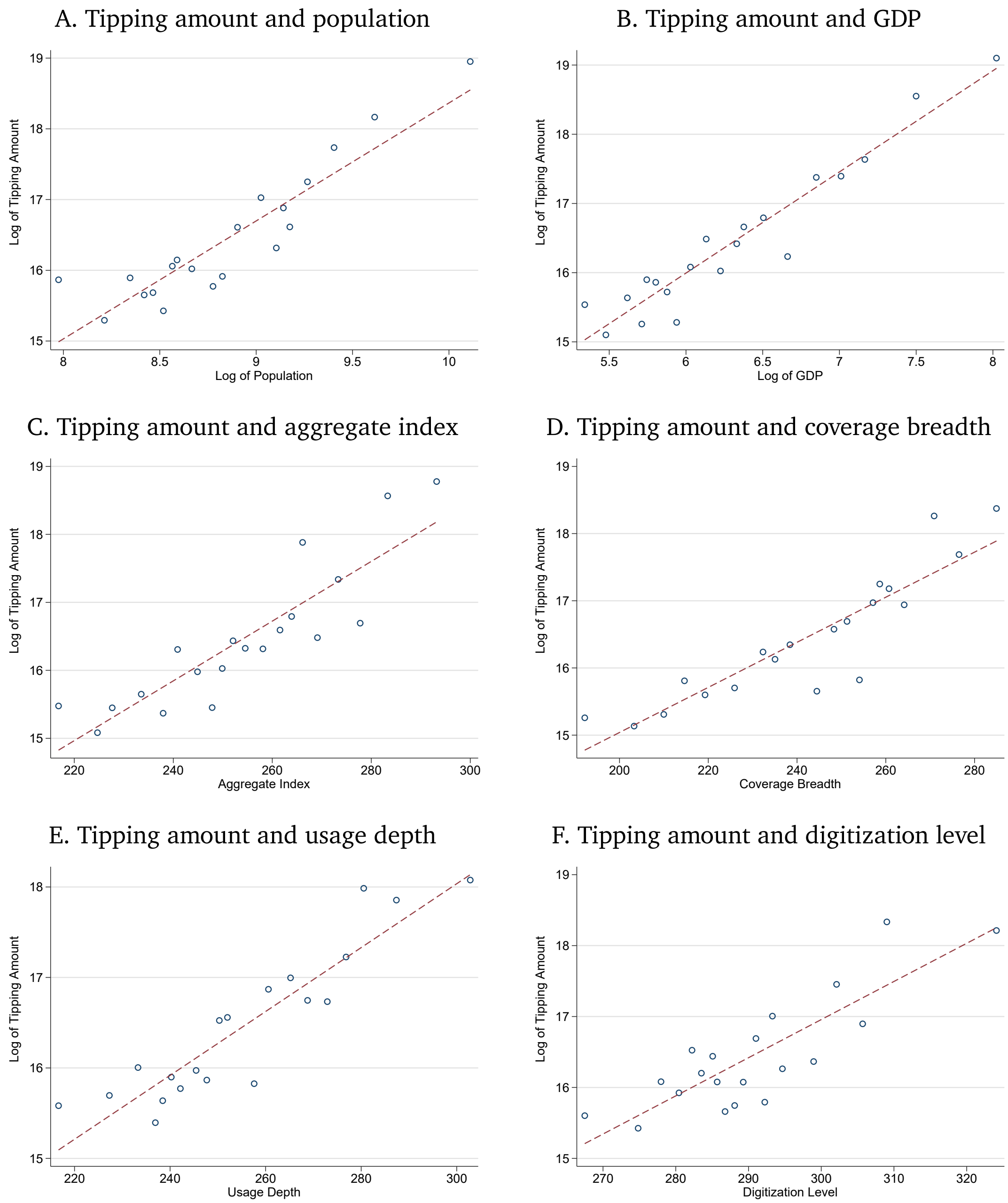
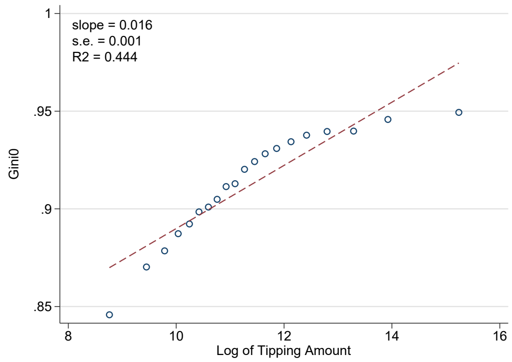
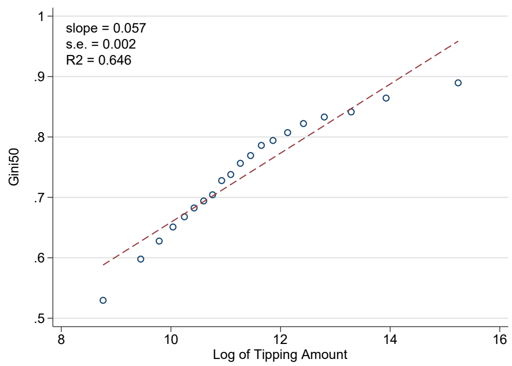


Figure A21. **Correlation between market size and city size.** In Section 5.1, we use one city's population, GDP, and digital financial inclusion index as instrumental variables for the platform's actual market size. This figure presents the correlation between the city size and local market size. We use the log value of tipping amount as an indicator of the scale of platform market size, which is placed on the y-axis and the instrument variable is placed on the x-axis. In Panel A, the instrument variable is the local residence population, and in Panel B, the instrument variable is the local GDP. From Panel C to Panel F, aggregate index, breadth of coverage, depth of use, and level of digitization are used, respectively. The city subsample includes 92 cities whose number of paying users accounts for more than 0.2% of the total number of paying users. The city-level population and GDP data come from China Premium Database in CEIC. And we use GDP in 2019 to avoid the impact of the Covid-19 pandemic shock. The data source of digital financial inclusion variables is The Peking University Digital Financial Inclusion Index of China in 2018, which utilizes Ant Financial's massive dataset on digital financial inclusion. The index calculation methodology is detailed in Guo et al. (2020).

A. Relation between tipping amount and Gini0



B. Relation between tipping amount and Gini50



C. Relation between tipping amount and Gini500

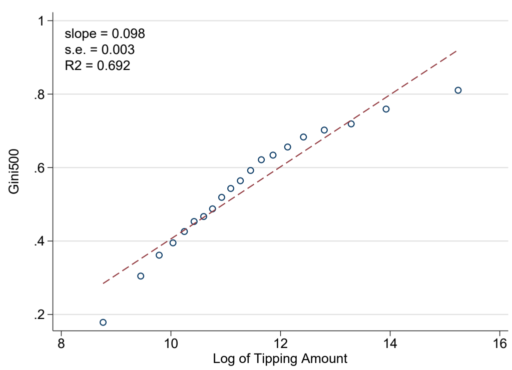


Figure A22. **City-month level relation between market size and inequality.** This figure provides intuitive city-month-level panel evidence that a broader market positively correlates with greater inequality. In this figure, we use total tipping amount as an indicator of local market size. From Panel A to Panel C, the Gini coefficients are calculated using the income distribution of all streamers, the streamers who receive more than 50 RMB in tipping income from one city in a month, and the streamers who receive more than 500 RMB, respectively. The city sample includes 302 cities and the time sample includes 25 months from April 2018 to April 2020. The red dashed line is the linear-fit line. The slope, standard error, and R^2 of the linear fit line are presented in the top left corner of each sub figure.

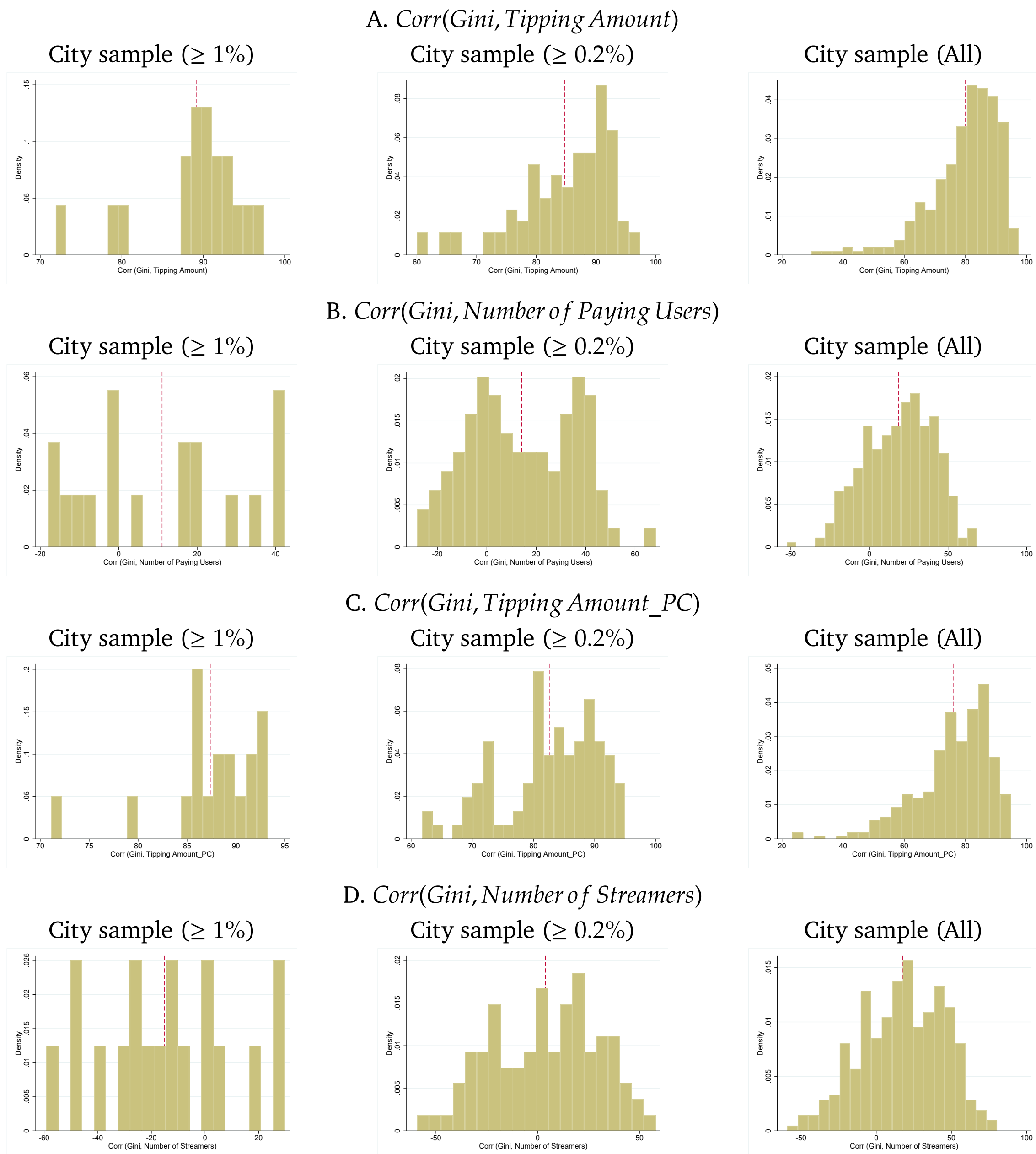


Figure A23. **Distribution of correlation between market size and Gini coefficients.** Using city-month level panel observations, we plot the density distribution of correlation between market size variables and Gini coefficients, and the red dashed line marks their mean value. The correlation coefficients are all multiplied by 100. $Gini_{j,t}$ is calculated using the income distribution of streamers who receive any virtual gifts from city j in month t. From Panel A to Panel D, we use total tipping amount, number of paying users, tipping amount per user, and the number of streamers who receive any tipping income from a city in a month, respectively, as indicators of market size. From left to right, the city samples included are different. City sample (≥ 1%) refers to the 18 cities that account for more than 1% of total paying users. City sample (≥ 0.2%) refers to the 92 cities that account for more than 0.2% of total paying users. City sample (All) refers to all 302 cities.

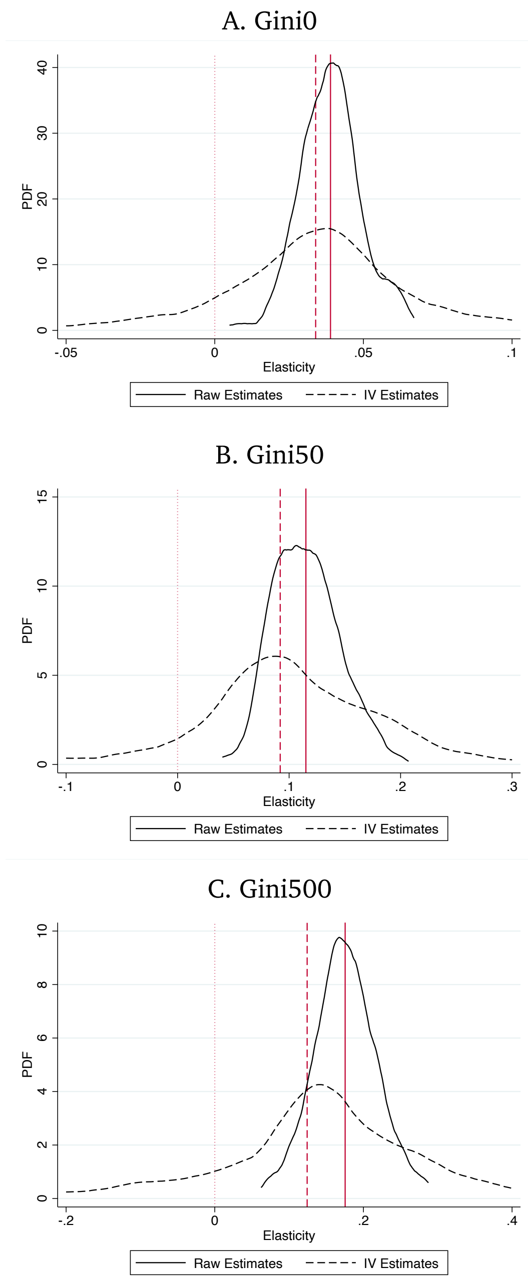


Figure A24. **OLS estimation versus IV estimation.** For each city j, we run the following OLS and IV regressions and then plot the probability distribution function of β_j: $Gini_{jt} = \alpha_j + \beta_{OLS,j}\ln(TippingAmount_{jt}) + \eta_t + \epsilon_{jt}, \forall j$ and $Gini_{jt} = \alpha_j + \beta_{IV,j}\ln(\widehat{TippingAmount_{jt}}) + \eta_t + \epsilon_{jt}, \forall j$. In Panels A to C, we measure the inequality using *Gini*0, *Gini*50, and *Gini*500, respectively. Specifically, $Gini0_{jt}$ is calculated using the income distribution of the streamers who earn any positive income from city j in month t. $Gini50_{jt}$ is calculated using the income distribution of the streamers who receive more than 50 RMB in tipping income from city j in month t. $Gini500_{jt}$ is calculated based on the income distribution of the streamers who receive more than 500 RMB in tipping income from city j in month t. The solid black line and the solid red line represent the distribution of the OLS estimators $\beta_{OLS,j}$ and their mean, respectively. The black and red dashed lines represent the distribution of IV estimators $\beta_{IV,j}$ and their mean, respectively.

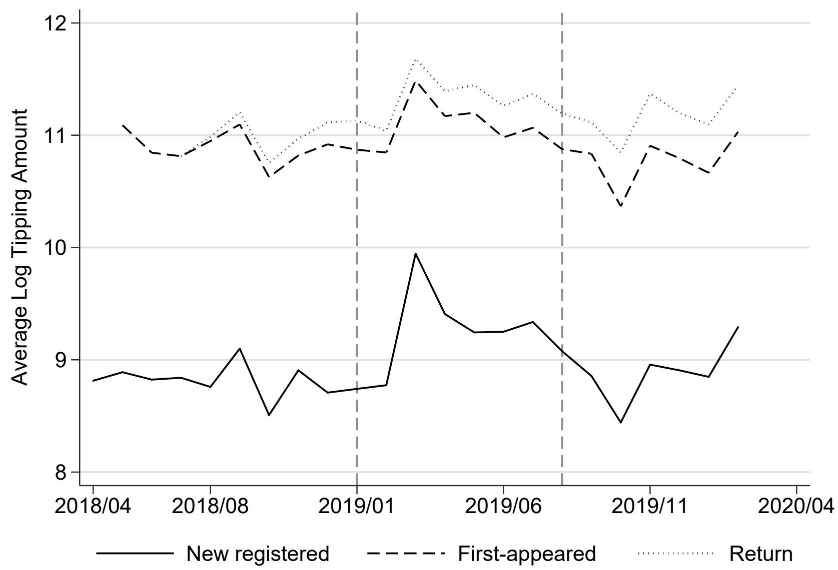


Figure A25. **The user growth rises and down around the capital raising event.** This figure plots the average log value of the tipping amount contributed by new users each month. We calculate the average value for each month based on the 302 cities. To exclude the impact of the Covid-19 pandemic, we first delete the months after January 2020. Then, we divide the remaining months into three-time stages based on the nodes of the launch of capital raising (i.e., 2019m1) and its closing (i.e., 2019m7), which are marked by vertical lines in this figure. Specifically, We use three methods to define whether a user is new or not in month *t*. First, if a newly registered user in month *t* converts to a paying user that month, then the user will be defined as new in month *t*, which is tagged as *New Registered*. Second, if a user makes a tip for the first time in month *t* during our sample period, then the user will be defined as a new user in month *t*, which is tagged as *First-Appeared*. According to this definition, all users in April 2018 are new and thus the value of this variable in April 2018 is defined as missing being more comparable. Third, if one has tipping records in month *t*, but does not have tipping records during period [t-3, t-1], then the user will be defined as new in month *t*, which is tagged as *Return*. The values of this variable before June 2018 are defined as missing, being more comparable.

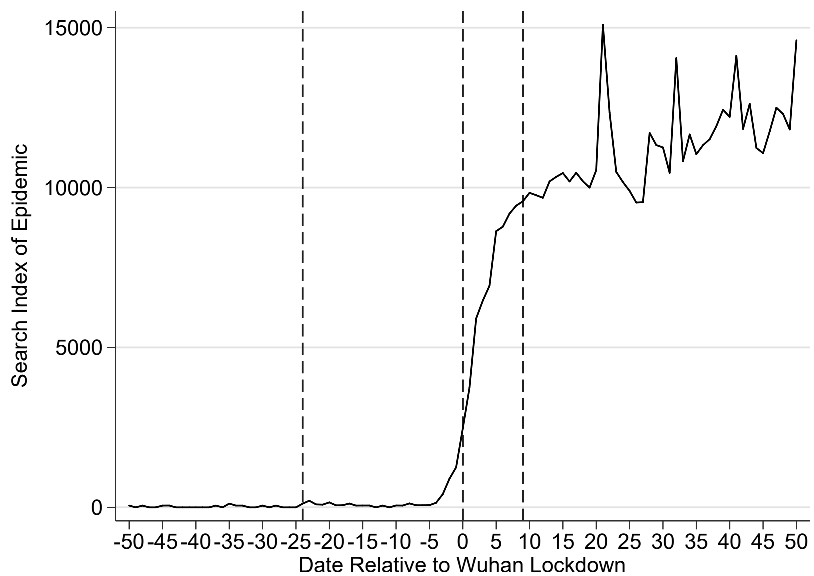


Figure A26. **Citizens' attention to the Covid-19 epidemic in Wuhan.** This figure shows Wuhan citizens' attention to the Covid-19 epidemic, as measured by the Baidu search index of the Chinese equivalent of the keyword "Epidemic". The search volume data were obtained from the website of "Baidu Index" (http://index.baidu.com/), which shows the search volume of Baidu's search engine using specific keywords at different periods and regions. The x-axis indicates the days relative to Wuhan lockdown on January 23, 2020. The first dashed line indicates Day -24 (December 30, 2019), when Dr. Wenliang Li made a Covid-19 alert post on one of his WeChat groups, and Covid-19 was first known to the public. The third dashed line indicates Day 9 (February 1, 2020), and from this day, in response to the aggravation of the epidemic, the government implemented a series of escalated measures. On February 1, 2020, the Hubei government announced the extension of the Spring Festival holiday. On February 2, 2020, *Huoshenshan Hospital* was officially put into operation, and the Hubei government announced that all suspected Covid-19 cases would be centrally isolated.

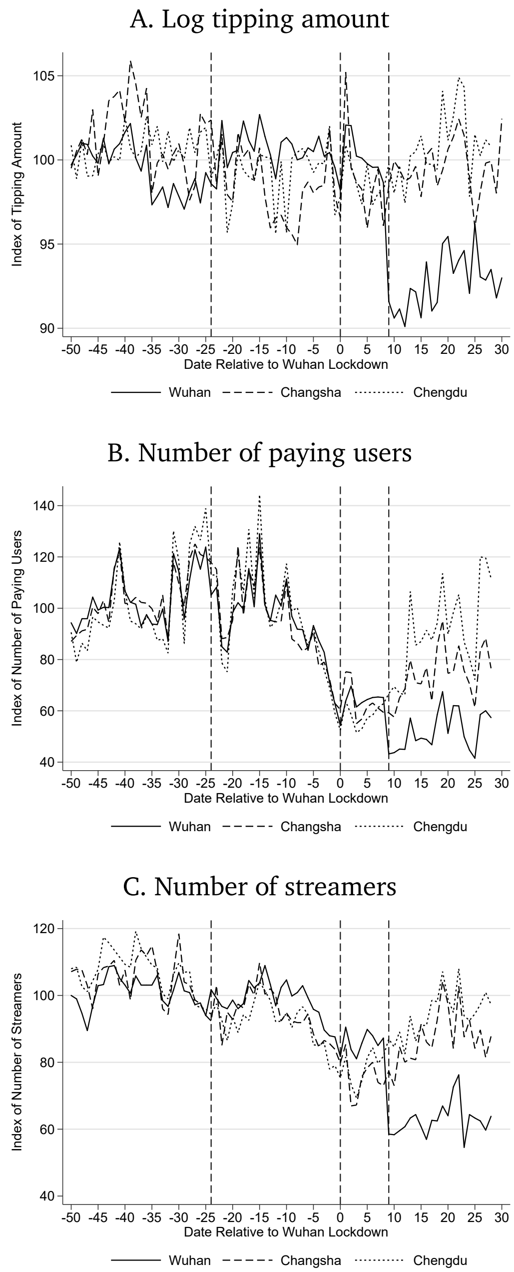


Figure A27. **Entertainment demand shock.** This figure plots the entertainment demand response to the Wuhan lockdown shock in three representative cities, the epicenter city Wuhan, its nearby city Changsha and the distant city Chengdu. The x-axis indicates the days relative to Wuhan lockdown on January 23, 2020. The first dashed line indicates Day -24 (December 30, 2019), when Dr. Wenliang Li made a Covid-19 alert post on one of his WeChat groups, and Covid-19 was first known to the public. The third dashed line indicates Day 9 (February 1, 2020), and with the aggravation of the epidemic, a series of escalated measures for Covid-19 protection were implemented from this day. On February 1, 2020, the Hubei government announced the extension of the Spring Festival holiday. On February 2, 2020, *Huoshenshan Hospital* was officially put into operation, and the Hubei government announced that all suspected Covid-19 cases would be centrally isolated. Panel A plots the index of the tipping amount within each city, which is defined as the daily log value of virtual gifts received by streamers relative to that before Day 0 (i.e., we set the mean within time intervals [-50, 0] to 100). Panel B plots the index of the number of paying users, defined as the daily number of paying users relative to that before Day 0. Panel C plots the index of the number of streamers who earn a positive tipping income within each city, which is defined as the daily number of streamers relative to that before Day 0.

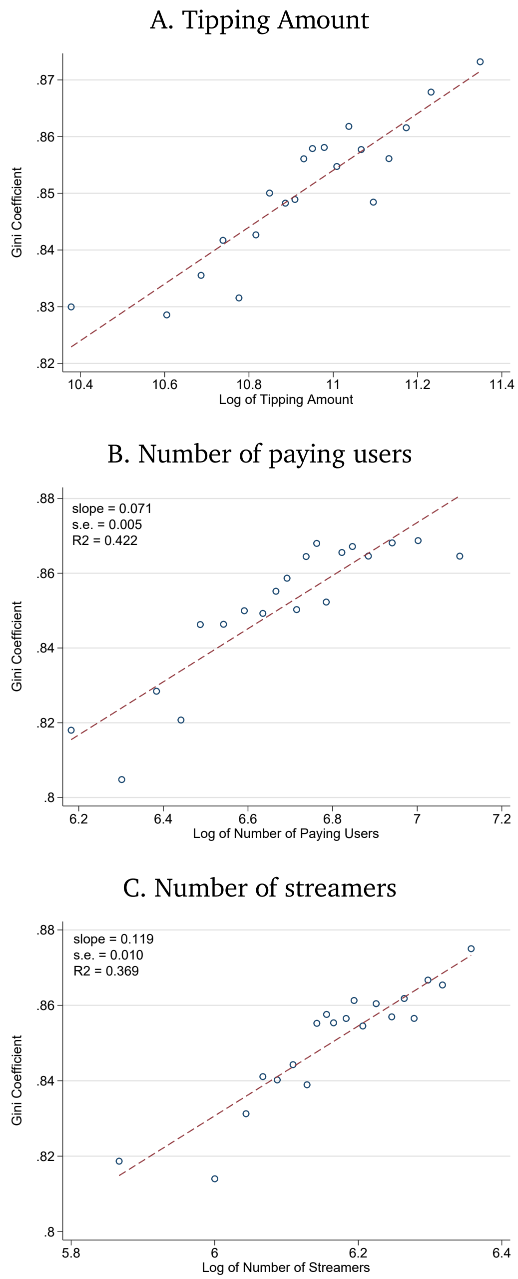


Figure A28. **City-daily level relation between market size and inequality.** This figure provides intuitive city-day-level panel evidence that a broader market positively correlates with greater inequality. We use $Gini50_{j,t}$ as a measure of inequality, which is calculated using the income distribution of streamers who receive virtual gifts worth more than 50 RMB from citizens in the city j from Day $(t-6)$ to Day t. If a user contributes more than half of the total tipping amount of a city on a day, we excluded these extreme outliers. The market size in Panels A, B, and C is measured by the log tipping amount, the log number of paying users, and the log number of streamers, respectively. The city sample includes nine cities that account for more than 2% paying users: Shanghai, Beijing, Guangzhou, Chengdu, Hangzhou, Wuhan, Shenzhen, Chongqing, and Changsha. The period is 79 days from 50 days before the Wuhan lockdown to 4 weeks after the Wuhan lockdown (Day -50 to Day 28). When graphing this scatter plot, we control for city-level fixed effects. The estimated β, standard errors and R^2 of the regression $Gini50_{j,t} = \alpha + \beta MarketSize_{j,t} + \theta_j + \epsilon_{j,t}$ are presented in the upper left corner of each subfigure.

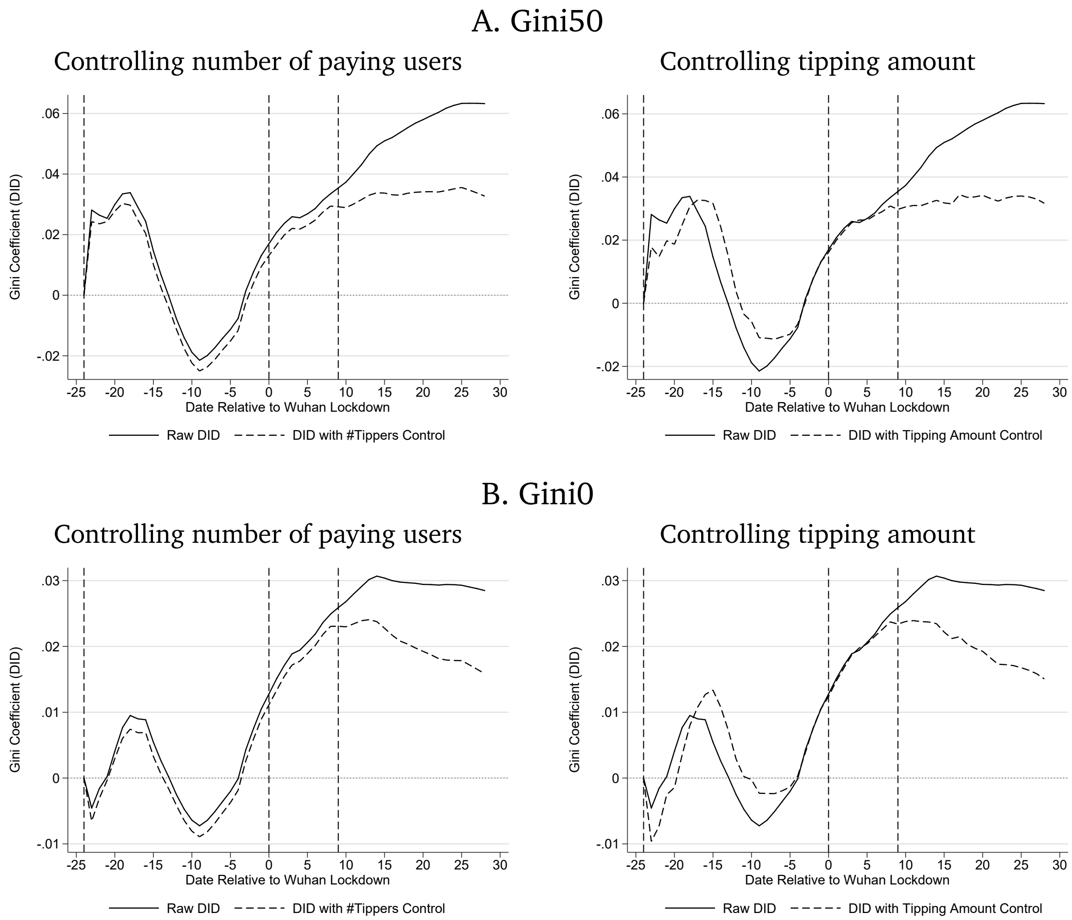
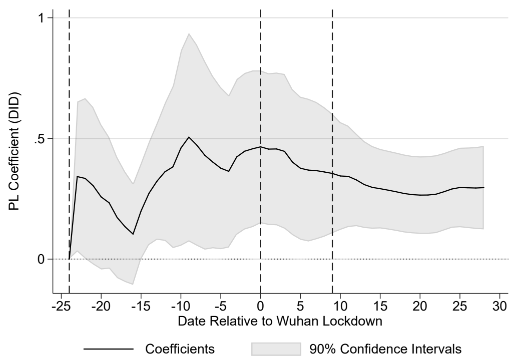
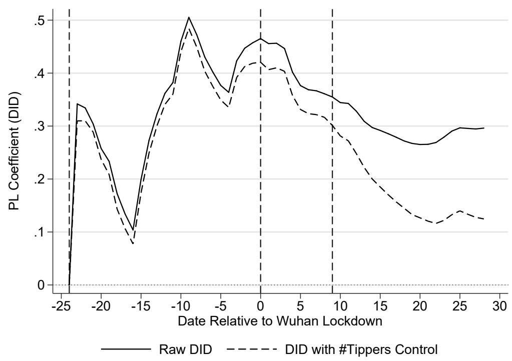


Figure A29. **The dynamic treatment effects with market size variables controlled.** In Figure 15, we plot estimated $\beta_{raw,T}$ in the following regressions $Gini_{j,t} = \alpha_{raw,T} + \beta_{raw,T} ln(Distance_j) \times Post_{1t} + \gamma_{raw,T} Post_{1t} + \theta_j + \epsilon_{j,t}$, for $t \in [-50, T]$ and $\forall T > -24$. In this figure, we additionally estimate the new regressions $Gini_{j,t} = \alpha_{new,T} + \beta_{new,T} ln(Distance_j) \times Post_t + \lambda_{new,T} Post_t + \delta_{new,T} MarketSize_{j,t} + \theta_j + \epsilon_{j,t}$, for $t \in [-50, T]$ and $\forall T > -24$. $Gini_{j,t}$ in Panel A is calculated based on the income distribution of streamers who receive virtual gifts worth more than 50 RMB from citizens in the city j from Day $(t-6)$ to Day t. $Gini_{j,t}$ in Panel B is calculated based on the income distribution of all streamers who receive any virtual gifts from citizens in the city j from Day $(t-6)$ to Day t. $Distance_j$ is the driving time (in days) between Wuhan and city j, shown on Baidu Maps. $Post_{1t} = 1$ indicates the days after December 30, 2019 (Day -24) when Dr. Wenliang Li made a Covid-19 alert post on one of his WeChat groups, and Covid-19 was first known to the public. From left to right, the log number of paying users and the log tipping amount are used to measure $MarketSize_{j,t}$, respectively. θ_j represents the city level fixed effects. Using observations of 9 cities from Day -50 to Day T, we run the above regressions and get the estimated $\beta_{raw,T}$ and $\beta_{new,T}$ for $\forall T > -24$. Then we plot both of the β before the interaction term as a function of time. The solid line refers to the value of $\beta_{raw,T}$, and the dashed line refers to the value of $\beta_{new,T}$. The third vertical dashed line indicates Day 9 (February 1, 2020), and a series of escalated measures for Covid-19 protection was implemented with the aggravation of the epidemic from this day on. On February 1, 2020, the Hubei government announced the extension of the Spring Festival holiday. On February 2, 2020, *Huoshenshan Hospital* was officially put into use, and the Hubei government announced that all suspected cases would be centrally isolated.

A. Dynamic DID coefficients



B. DID coefficients comparison: Controlling the number of paying users



C. DID coefficients comparison: Controlling tipping amount

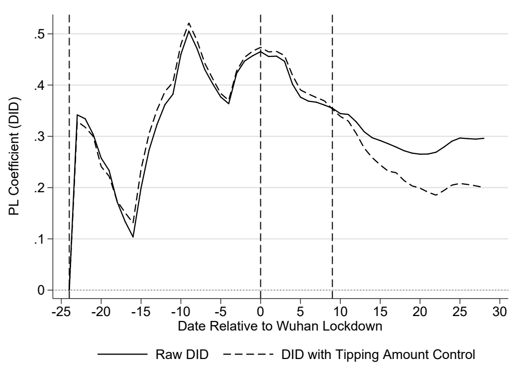


Figure A30. **The dynamic treatment effects using PL exponent as the dependent variable.** We estimate $\beta_{raw,T}$ in the DID regressions $PL_{j,t} = \alpha_{raw,T} + \beta_{raw,T} ln(Distance_j) \times Post_{1t} + \gamma_{raw,T} Post_{1t} + \theta_j + \epsilon_{j,t}$, for $t \in [-50, T]$ and $\forall T > -24$. Additionally, we also estimate the new DID regressions $PL_{j,t} = \alpha_{new,T} + \beta_{new,T} ln(Distance_j) \times Post_{1t} + \lambda_{new,T} Post_{1t} + \delta_{new,T} MarketSize_{j,t} + \theta_j + \epsilon_{j,t}$, for $t \in [-50, T]$ and $\forall T > -24$. $PL_{j,t}$ is calculated based on the top 10 streamers in city j from Day $(t-6)$ to Day t. $Distance_j$ is the driving time (in days) between Wuhan and city j, as shown on Baidu Maps. $Post_{1t} = 1$ indicates the days after December 30, 2019 (Day -24) when Dr. Wenliang Li posted a Covid-19 alert on one of his WeChat groups, and Covid-19 became known to the public. In Panel B, the log number of paying users is used to measure $MarketSize_{j,t}$, while in Panel C, the log of the tipping amount is used to measure $MarketSize_{j,t}$. θ_j represents the city level fixed effects. Using observations of 9 cities from Day -50 to Day T, we run the above regressions and get the estimated $\beta_{raw,T}$ and $\beta_{new,T}$ for $\forall T > -24$. Next, in Panel A, we plot the dynamics of $\beta_{raw,T}$ and its 90% confidence interval. In Panel B and Panel C, we plot both of the β where the solid line refers to the value of $\beta_{raw,T}$ and the dashed line refers to the value of $\beta_{new,T}$. The third vertical dashed line in the figure indicates Day 9 (February 1, 2020), from which a series of escalated measures for Covid-19 protection was implemented with the aggravation of the epidemic.

Table A1. **The income of streamers with different rankings.** This table gives a rough description of the tipping income of streamers with different rankings. The term "income" refers to the total value (in million RMB) of all virtual gifts a streamer receives during the sample period.

Rank	N	Mean	Sd	Min	P50	Max
$[1,10]$	10	99.206	62.996	37.705	87.433	232.482
$(10,50]$	40	13.853	8.084	7.560	10.936	36.930
$(50,100]$	50	5.316	0.926	4.067	5.130	7.423
$(100,500]$	400	1.727	0.807	0.864	1.479	4.059
$(500,1000]$	500	0.571	0.136	0.388	0.549	0.863
$(1000,2000]$	1,000	0.242	0.064	0.157	0.230	0.388
> 2000	129,958	0.003	0.014	0.000	0.000	0.157
Total	131,958	0.026	1.057	0.000	0.000	232.482

Table A2. **Percentiles.** This table reports the variance and the absolute value of different percentiles of seasonality-adjusted monthly log tipping income. *Var* is the variance of the log tipping income variable. Pn is the absolute value of the $n\,th$ percentile of log tipping income variable.

	Var	**P1**	**P5**	**P10**	**P25**	**P50**	**P75**	**P90**	**P95**	**P99**
Jun-18	8.378	0.033	0.065	0.236	1.099	2.879	5.597	7.803	9.042	10.898
Jul-18	8.458	0.033	0.065	0.236	1.110	3.083	5.723	7.864	9.066	10.913
Aug-18	8.597	0.033	0.065	0.236	1.121	3.231	5.804	7.947	9.100	10.968
Sep-18	8.728	0.033	0.065	0.182	1.099	3.259	5.812	7.971	9.135	11.066
Oct-18	8.891	0.033	0.065	0.154	1.099	3.308	5.803	8.040	9.181	11.097
Nov-18	9.038	0.033	0.033	0.154	1.099	3.218	5.802	8.070	9.258	11.087
Dec-18	9.328	0.033	0.033	0.125	1.099	3.144	5.825	8.169	9.352	11.245
Jan-19	9.525	0.026	0.033	0.125	1.087	3.070	5.855	8.233	9.404	11.280
Feb-19	9.751	0.026	0.033	0.095	0.970	2.987	5.824	8.340	9.522	11.328
Mar-19	9.657	0.026	0.033	0.095	0.956	2.870	5.760	8.268	9.502	11.252
Apr-19	9.671	0.026	0.033	0.089	0.930	2.773	5.693	8.260	9.505	11.284
May-19	9.574	0.026	0.033	0.095	0.930	2.781	5.624	8.200	9.495	11.331
Jun-19	9.588	0.026	0.033	0.101	0.968	2.833	5.664	8.186	9.488	11.395
Jul-19	9.511	0.026	0.033	0.125	1.099	2.968	5.730	8.184	9.492	11.431
Aug-19	9.461	0.026	0.052	0.125	1.099	2.953	5.728	8.166	9.395	11.381
Sep-19	9.533	0.026	0.033	0.113	1.094	2.927	5.786	8.157	9.394	11.386
Oct-19	9.605	0.026	0.033	0.095	0.930	2.835	5.710	8.161	9.440	11.354
Nov-19	9.810	0.033	0.033	0.095	0.993	2.941	5.785	8.295	9.576	11.515
Dec-19	9.966	0.026	0.033	0.095	0.930	2.872	5.783	8.347	9.653	11.661
Jan-20	10.359	0.026	0.033	0.083	0.804	2.772	5.848	8.441	9.754	11.819
Feb-20	10.013	0.026	0.033	0.065	0.513	2.175	5.394	8.198	9.576	11.654
Mar-20	9.687	0.026	0.033	0.065	0.511	2.054	5.165	7.987	9.496	11.550
Apr-20	9.251	0.026	0.033	0.065	0.511	2.057	5.025	7.767	9.303	11.454

Table A3. **Correlation coefficients.** This table shows the Pearson correlation coefficients between the various measures of inequality calculated in Section 3. The specific definitions of the variables shown in the figure are as follows: *topshare_income* refers to the income share of the top 1,000 streamers ordered by their total income in the past three months; *gini_income* is calculated using a sample of streamers who receive positive virtual gifts in the past three months; *pl_income* is calculated using the top 2,000 streamers ordered by their three-month tipping income; *topshare_fans* refers to the loyal fans share of top 1,000 streamers ordered by their total number of loyal fans in the past three months; *gini_fans* is calculated using a sample of streamers with at least one loyal fans in the past three months; *pl_fans* is calculated using a sample of the top 2,000 streamers ordered by their three-month loyal fans number; *P90_P50* is the percentile gap of log tipping income of streamers in the past three months; and *Var* is the variance of log tipping income of streamers in the past three months.* indicates significance at the 5% level.

	topshare_income	gini_income	pl_income	topshare_fans	gini_fans	pl_fans	P90_P50	Var
topshare_income	1.000							
gini_income	0.958*	1.000						
pl_income	0.953*	0.856*	1.000					
topshare_fans	0.963*	0.868*	0.955*	1.000				
gini_fans	0.897*	0.856*	0.794*	0.897*	1.000			
pl_fans	0.782*	0.631*	0.887*	0.885*	0.669*	1.000		
P90_P50	0.878*	0.897*	0.769*	0.851*	0.771*	0.616*	1.000	
Var	0.785*	0.761*	0.622*	0.740*	0.846*	0.396	0.802*	1.000

Table A4. **The change in Gini coefficient at the end of each quarter.** This table reports the change in the Gini coefficients at the end of each quarter. Specifically, we run the following regression: $Gini_t = \alpha + \beta QuarterEnd_t + \eta_t + \epsilon_t$, where $Gini_t$ is the Gini coefficient; $QuarterEnd_t$ is a dummy variable that equals one if the month is March, June, September or December, and zero otherwise; and η_t represents year-quarter fixed effects. In Columns (1) and (4), $Gini_t$ is the $Gini0_t$ calculated using the full sample with all streamers who receive any income in month t. In Columns (2) and (5), $Gini_t$ is $Gini50_t$, which is the Gini coefficient of the subsample of streamers whose income is more than 50 RMB in month t. In Columns (3) and (6), $Gini_t$ is $Gini500_t$, which is the Gini coefficient of the subsample of streamers whose income is more than 500 RMB. The regression includes the observations of 25 months from April 2018 to April 2020. Robust standard errors are reported in parentheses. ***, **, * indicate significance at the 1%, 5%, and 10% levels respectively.

	(1) Gini0	(2) Gini50	(3) Gini500	(4) Gini0	(5) Gini50	(6) Gini500
QuarterEnd	0.014***	0.027***	0.041***	0.015***	0.029***	0.044***
	(0.005)	(0.007)	(0.011)	(0.003)	(0.006)	(0.008)
Year-Quarter FEs	No	No	No	Yes	Yes	Yes
N	25	25	25	25	25	25
Adjusted R2	0.231	0.294	0.292	0.731	0.677	0.701

Table A5. **The first-stage regression in time series estimation.** This table reports the first-stage regression results in time series estimation. We run the regression $MarketSize_t = \alpha + \beta QuarterEnd_t + \eta_t + \epsilon_t$, where $MarketSize_t$ is a series of variables measuring the platform use; $QuarterEnd_t$ is a dummy variable that equals one if the month is March, June, September or December, and zero otherwise; and η_t represents year-quarter fixed effects. In Column (1), $MarketSize_t$ is $\ln(Tipping\,Amount_t)$, which is the log value of the total tipping income of all streamers in month t. In Column (2), $MarketSize_t$ is $\ln(Number\,of\,Paying\,Users_t)$, which is the log number of paying users in month t. In Column (3), $MarketSize_t$ is $\ln(Tipping\,Amount\_PC_t)$, which is the log value of the average tipping income per streamer in month t. In Column (4), $MarketSize_t$ is $\ln(Number\,of\,Streamers_t)$, which is the log number of streamers who receive any virtual gifts in month t. Standard errors are robust and reported in parentheses. The sample includes the 25 months from April 2018 to April 2020. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

	(1) ln(Tipping Amount)	(2) ln(Number of Paying Users)	(3) ln(Tipping Amount_PC)	(4) ln(Number of Streamers)
QuarterEnd	0.374***	0.038	0.336***	0.004
	(0.081)	(0.075)	(0.061)	(0.039)
Year-Quarter FEs	Yes	Yes	Yes	Yes
N	25	25	25	25
Adjusted R2	0.703	0.411	0.774	0.571

Table A6. **The first-stage regression in panel estimation with 18 cities included in the sample.** This table reports the estimated results in the regression $MarketSize_{j,t} = \alpha + \beta QuarterEnd_t + \theta_j + \eta_t + \epsilon_{j,t}$, where $MarketSize_{j,t}$ is a series of variables measuring the platform market size in city j in month t; $QuarterEnd_t$ is a dummy variable that equals one if the month is March, June, September or December, and zero otherwise; θ_j represents city fixed effects; and η_t represents year-quarter fixed effects. In Column (1), $MarketSize_{j,t}$ is $\ln(Tipping\,Amount_{j,t})$, which is the log value of the total tipping income of all the streamers in the city j in month t. In Column (2), $MarketSize_{j,t}$ is $\ln(Number\,of\,Paying\,Users_{j,t})$, which is the log number of paying users in city j in month t. In Column (3), $MarketSize_{j,t}$ is $\ln(Tipping\,Amount\_PC_{j,t})$, which is the log value of average tipping income per streamer in city j in month t. In Column (4), $MarketSize_{j,t}$ is $\ln(Number\,of\,Streamers_{j,t})$, which is the log number of streamers who receive any virtual gifts from city j in month t. Standard errors, clustered at the city and year-quarter levels, are reported in parentheses. The city subsample in this regression includes the 18 cities that account for more than 1% of the platform's paying users for the 25 months from April 2018 to April 2020. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

	(1) ln(Tipping Amount)	(2) ln(Number of Paying Users)	(3) ln(Tipping Amount_PC)	(4) ln(Number of Streamers)
Quarterend	0.360***	0.070	0.289***	0.028
	(0.073)	(0.061)	(0.055)	(0.030)
City FEs	Yes	Yes	Yes	Yes
Year-Quarter FEs	Yes	Yes	Yes	Yes
N	450	450	450	450
Adjusted R2	0.844	0.926	0.653	0.923

Table A7. **The first-stage regression in panel estimation with 92 cities included in the sample.** This table reports the estimated results in the regression $MarketSize_{j,t} = \alpha+\beta QuarterEnd_t+\theta_j+\eta_t+\epsilon_{j,t}$, where $MarketSize_{j,t}$ is a series of variables measuring the platform marle in city j in month t; $QuarterEnd_t$ is a dummy variable that equals one if the month is March, June, September or December, and zero otherwise; θ_j represents city fixed effects; and η_t represents year-quarter fixed effects. In Column (1), $MarketSize_{j,t}$ is $\ln(Tipping\,Amount_{j,t})$, which is the log value of the total tipping income of all the streamers in the city j in month t. In Column (2), $MarketSize_{j,t}$ is $\ln(Number\,of\,Paying\,Users_{j,t})$, which is the log number of paying users in city j in month t. In Column (3), $MarketSize_{j,t}$ is $\ln(Tipping\,Amount\_PC_{j,t})$, which is the log value of average tipping income per streamer in city j in month t. In Column (4), $MarketSize_{j,t}$ is $\ln(Number\,of\,Streamers_{j,t})$, which is the log number of streamers who receive any virtual gifts from city j in month t. Standard errors, clustered at the city and year-quarter levels, are reported in parentheses. The city subsample in this regression includes the 92 cities that account for more than 0.2% of the platform's paying users for the 25 months from April 2018 to April 2020. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

	(1) ln(Tipping Amount)	(2) ln(Number of Paying Users)	(3) ln(Tipping Amount_PC)	(4) ln(Number of Streamers)
Quarterend	0.244***	0.037	0.207***	0.030
	(0.057)	(0.060)	(0.035)	(0.032)
City FEs	Yes	Yes	Yes	Yes
Year-Quarter FEs	Yes	Yes	Yes	Yes
N	2,300	2,300	2,300	2,300
Adjusted R2	0.858	0.961	0.531	0.945

Table A8. **The first-stage regression in panel estimation with 302 cities included in the sample.** This table reports the estimated results in the regression $MarketSize_{j,t} = \alpha + \beta QuarterEnd_t + \theta_j + \eta_t + \epsilon_{j,t}$, where $MarketSize_{j,t}$ is a series of variables measuring the platform market size in city j in month t; $QuarterEnd_t$ is a dummy variable that equals one if the month is March, June, September or December, and zero otherwise; θ_j represents city fixed effects; and η_t represents year-quarter fixed effects. In Column (1), $MarketSize_{j,t}$ is $\ln(Tipping\,Amount_{j,t})$, which is the log value of the total tipping income of all the streamers in the city j in month t. In Column (2), $MarketSize_{j,t}$ is $\ln(Number\,of\,Paying\,Users_{j,t})$, which is the log number of paying users in the city j in month t. In Column (3), $MarketSize_{j,t}$ is $\ln(Tipping\,Amount\_PC_{j,t})$, which is the log value of average tipping income per streamer in city j in month t. In Column (4), $MarketSize_{j,t}$ is $\ln(Number\,of\,Streamers_{j,t})$, which is the log number of streamers who receive any virtual gifts from city j in month t. Standard errors, clustered at the city and year-quarter levels, are reported in parentheses. The city subsample in this regression includes the 302 cities with at least one streamer who can earn a tipping income worth more than 500 RMB from that city every month for the 25 months from April 2018 to April 2020. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

	(1) ln(Tipping Amount)	(2) ln(Number of Paying Users)	(3) ln(Tipping Amount_PC)	(4) ln(Number of Streamers)
Quarterend	0.193***	0.007	0.186***	0.025
	(0.050)	(0.060)	(0.031)	(0.037)
City FEs	Yes	Yes	Yes	Yes
Year-Quarter FEs	Yes	Yes	Yes	Yes
N	7,550	7,550	7,550	7,550
Adjusted R2	0.824	0.972	0.414	0.947

Table A9. **Summary statistics of monthly level data sample.** This table shows the summary statistics of variables used in the monthly level data sample. Panel A summarizes the variables used in Table 3 in Section 5.2.1. $Gini0_t$ is the Gini coefficient calculated using the income distribution of all streamers who receive any virtual gifts in month t; $Gini50_t$ is the Gini coefficient of the subsample of streamers whose incomes are more than 50 RMB in month t; $Gini500_t$ is the Gini coefficient of the subsample of streamers whose incomes are more than 500 RMB in month t; $Tipping\,Amount_t$ (in thousand RMB) is the value of the total tipping income of all streamers in month t; $Number\,of\,Paying\,Users_t$ is the number of paying users in month t; $Tipping\,Amount\_PC_t$ is the value of the average tipping income per streamer in month t; $Number\,of\,Streamers_t$ is the number of streamers who receive any virtual gifts in month t. Panel B summarizes the variables used in Table 4 in Section 5.2.2. $Gini0_{j,t}$ is the Gini coefficient for the full sample of the streamers who receive any income from users in city j in month t; $Gini50_{j,t}$ is the Gini coefficient of the subsample of streamers whose incomes are more than 50 RMB from city j in month t; $Gini500_{j,t}$ is the Gini coefficient of the subsample of streamers whose incomes are more than 500 RMB; $Tipping\,Amount_{j,t}$ (in thousand RMB) is the value of total tipping income of all streamers in city j in month t; $Number\,of\,Paying\,Users_{j,t}$ is the number of paying users in the city j in month t; $Tipping\,Amount\_PC_{j,t}$ is the value of average tipping income per streamer in city j in month t; and $Number\,of\,Streamers_{j,t}$ is the number of streamers who receive any virtual gifts from city j in month t.

Variable	N	Mean	Sd	Min	P50	Max
Panel A: Time series estimation sample						
Gini0	25	0.948	0.014	0.924	0.950	0.976
Gini50	25	0.901	0.024	0.860	0.901	0.949
Gini500	25	0.851	0.036	0.785	0.856	0.920
Tipping Amount	25	139,000.000	48,000.000	73,900.000	140,000.000	262,000.000
Number of Paying Users	25	242,556.800	49,775.530	177,208.000	232,355.000	337,275.000
Tipping Amount _PC	25	578.604	197.586	360.455	512.650	1,156.982
Number of Streamers	25	11,714.760	1,435.617	8,696.000	11,990.000	14,006.000
Panel B: Panel estimation sample with 18 cities						
Gini0	450	0.938	0.020	0.886	0.937	0.991
Gini50	450	0.861	0.047	0.729	0.861	0.977
Gini500	450	0.762	0.080	0.537	0.766	0.960
Tipping Amount	450	4,437.684	4,746.861	505.898	2,811.819	30,500.000
Number of Paying Users	450	5,589.473	2,978.586	1,621.000	4,776.000	16,358.000
Tipping Amount_PC	450	726.813	563.204	196.753	573.259	7,305.005
Number of Streamers	450	2,495.591	724.627	1,282.000	2,384.500	4,557.000

Table A10. **Panel estimation of the effect of market size on inequality, with 92 cities included in the sample.** This table reports the panel estimation results of the OLS and IV regression: $Gini_{j,t} = \alpha + \beta_{OLS}\ln(TippingAmount_{j,t}) + \theta_j + \eta_t + \epsilon_{j,t}$, and $Gini_{j,t} = \alpha + \beta_{IV}\ln(\widehat{TippingAmount}_{j,t}) + \theta_j + \eta_t + \epsilon_{j,t}$, where $Gini_{j,t}$ is the Gini coefficients in the city j in month t; $\ln(TippingAmount_{j,t})$ is the log value of the tip spending by citizens of the city j in month t measuring the amount of platform use; θ_j represents city fixed effects, and η_t represents the year-quarter fixed effects. The instrumented variable $QuarterEnd_t$ is a dummy variable that equals one when the month is March, June, September, and December and zeroes otherwise. In Columns (1) and (4), $Gini_{j,t}$ is the Gini coefficient for the full sample of the streamers who receive any income from users in the city j in month t. In Columns (2) and (5), $Gini_{j,t}$ is the Gini coefficient of the subsample of streamers whose incomes are more than 50 RMB from city j in month t. In Columns (3) and (6), $Gini_{j,t}$ is the Gini coefficient of the subsample of streamers whose incomes are more than 500 RMB. Columns (1), (2), and (3) report the OLS results β_{OLS}. Columns (4), (5), and (6) report the 2SLS results β_{IV}. The regression observations are from the 92 cities that account for more than 0.2% of total paying users during the 25 months from April 2018 to April 2020. Standard errors are clustered at the city and year-quarter level and reported in parentheses. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

	OLS			IV: *QuarterEnd*		
	Gini	Gini50	Gini500	Gini	Gini50	Gini500
	(1)	(2)	(3)	(4)	(5)	(6)
ln(Tipping Amount)	0.034***	0.095***	0.154***	0.036***	0.085***	0.126***
	(0.001)	(0.003)	(0.005)	(0.004)	(0.006)	(0.009)
City FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter FEs	Yes	Yes	Yes	Yes	Yes	Yes
N	2,300	2,300	2,300	2,300	2,300	2,300
Adjusted R2	0.799	0.855	0.837	0.751	0.772	0.72
K. P. F-statistic					18.423	

Table A11. **Panel estimation of the effect of market size on inequality, with 302 cities included in the sample.** This table reports the panel estimation results of the OLS and IV regression: $Gini_{j,t} = \alpha + \beta_{OLS}\ln(TippingAmount_{j,t}) + \theta_j + \eta_t + \epsilon_{j,t}$, and $Gini_{j,t} = \alpha + \beta_{IV}\ln(\widehat{TippingAmount}_{j,t}) + \theta_j + \eta_t + \epsilon_{j,t}$, where $Gini_{j,t}$ is the Gini coefficients in the city j in month t; $\ln(TippingAmount_{j,t})$ is the log value of the tip spending by citizens of the city j in month t measuring the amount of platform use; θ_j represents city fixed effects, and η_t represents the year-quarter fixed effects. The instrumented variable $QuarterEnd_t$ is a dummy variable that equals one when the month is March, June, September, and December and zeroes otherwise. In Columns (1) and (4), $Gini_{j,t}$ is the Gini coefficient for the full sample of the streamers who receive any income from users in the city j in month t. In Columns (2) and (5), $Gini_{j,t}$ is the Gini coefficient of the subsample of streamers whose incomes are more than 50 RMB from city j in month t. In Columns (3) and (6), $Gini_{j,t}$ is the Gini coefficient of the subsample of streamers whose incomes are more than 500 RMB. Columns (1), (2), and (3) report the OLS results β_{OLS}. Columns (4), (5), and (6) report the 2SLS results β_{IV}. The regression observations are from the 302 cities where at least one streamer can earn a tipping income worth more than 500 RMB from that city every month during the 25 months from April 2018 to April 2020. Standard errors are clustered at the city and year-quarter level and reported in parentheses. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

	OLS			IV: *QuarterEnd*		
	Gini	Gini50	Gini500	Gini	Gini50	Gini500
	(1)	(2)	(3)	(4)	(5)	(6)
ln(Tipping Amount)	0.035***	0.109***	0.172***	0.036***	0.098***	0.144***
	(0.001)	(0.002)	(0.003)	(0.009)	(0.007)	(0.012)
City FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter FEs	Yes	Yes	Yes	Yes	Yes	Yes
N	7550	7550	7550	7550	7550	7550
Adjusted R2	0.717	0.818	0.827	0.618	0.702	0.677
K. P. F-statistic					14.793	

Table A12. **The number of new users around the capital raising event.** This table shows the new user growth rise and down around the capital raising event of this platform. To exclude the interference of the Covid-19 pandemic, we first deleted the months after January 2020. Then, we divide the remaining months into three-time stages based on the nodes of the launch of capital raising (i.e., 2019m1) and its closing (i.e., 2019m7). We refer to the period before 2019m1 as *Stage 1*, the period between 2019m1 and 2019m7 as *Stage 2*, and the period post 2019m7 as *Stage 3*. Specifically, we run the following regressions: $\ln(Tipping\,Amount\_New\,Users_{j,t}) = \alpha + \beta_1 \mathbb{1}(Stage\,2)_t + \beta_2 \mathbb{1}(Stage\,3)_t + \epsilon_{j,t}$, where $\ln(Tipping\,Amount\_New\,Users_{j,t})$ is the tipping amount contributed by new users; $\mathbb{1}(Stage\,2)_t$ and $\mathbb{1}(Stage\,3)_t$ are dummy variables that mark whether the month t is in Stage 2 or Stage 3. From Column (1) to Column (3), we use the tipping amount contributed by New Registered, First Appeared, and Return users as the dependent variable. In Panel A, the regression observations are from the 18 cities that account for more than 1% of total paying users during the 25 months from April 2018 to April 2020. In Panel B, the regression observations are from the 92 cities that account for more than 0.2% of total paying users. And in Panel C, we use a broader sample of 302 cities. Standard errors are reported in parentheses. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)
	ln (Tipping Amount_New Register)	ln (Tipping Amount_First Appeared)	ln (Tipping Amount_Return)
Panel A: City Sample 18			
$\mathbb{1}(Stage\,2)$	0.502***	0.202*	0.351***
	(0.151)	(0.116)	(0.123)
$\mathbb{1}(Stage\,3)$	0.251*	-0.053	0.226*
	(0.151)	(0.116)	(0.123)
N	414	396	360
Adjusted R2	0.022	0.008	0.017
Panel B: City Sample 92			
$\mathbb{1}(Stage\,2)$	0.426***	0.194***	0.358***
	(0.098)	(0.072)	(0.076)
$\mathbb{1}(Stage\,3)$	0.095	-0.112	0.206***
	(0.098)	(0.072)	(0.076)
N	2116	2024	1840
Adjusted R2	0.008	0.007	0.011
Panel C: City Sample 302			
$\mathbb{1}(Stage\,2)$	0.137*	0.038	0.245***
	(0.072)	(0.050)	(0.053)
$\mathbb{1}(Stage\,3)$	-0.195***	-0.260***	0.126**
	(0.072)	(0.050)	(0.053)
N	6785	6644	6040
Adjusted R2	0.003	0.006	0.003

Table A13. **The first-stage regression results using user surge caused by the capital raising event as the instrument variable.** This table reports the first-stage regression results in Section 5.3. To exclude the interference of the Covid-19 pandemic, we first delete the months after January 2020. Then, we divide the remaining months into three-time stages based on the nodes of the launch of capital raising (i.e., 2019m1) and its closing (i.e., 2019m7). We then refer to the period before 2019m1 as *Stage 1*, the period between 2019m1 and 2019m7 as *Stage 2*, and the period post 2019m7 as *Stage 3*. Then we run the regressions $\Delta\ \ln(Tipping\ Amount)_{j,stage} = \alpha + \beta\Delta\ \ln(Tipping\ Amount\_New\ Users)_{j,stage} + \epsilon_j$, where $\Delta\ \ln(Tipping\ Amount)$ is the difference in the average monthly size of the tipping amount between the two stages and $\Delta\ \ln(Tipping\ Amount\_New\ Users)$ is the difference in the average monthly size of the tipping amount contributed by new users between the two stages. *New Register*, *New Register*, and *New Register* are three methods to define whether a user is new or not in the month t. In Columns (1), (2), and (3), We use the value difference of the variables between *Stage 2* and *Stage 1*. And in Columns (4), (5), and (6), We use the value difference of the variables between *Stage 3* and *Stage 2*. From Panel A to Panel C, the regression sample includes 18 cities, 92 cities, and 302 cities, respectively. Standard errors are reported in parentheses. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
	Stage 2 - Stage 1			Stage 3 - Stage 2		
	$\Delta\ \ln(Tipping\ Amount)$					
Panel A: City Sample 18						
$\Delta\ \ln(Tipping\ Amount\_New\ Register)$	0.194***			0.147***		
	(0.038)			(0.042)		
$\Delta\ \ln(Tipping\ Amount\_First\ Appeared)$		0.348***			0.205	
		(0.079)			(0.132)	
$\Delta\ \ln(Tipping\ Amount\_Return)$			0.367***			0.219
			(0.075)			(0.147)
N	18	18	18	18	18	18
Adjusted R2	0.594	0.520	0.572	0.394	0.077	0.066
Panel B: City Sample 92						
$\Delta\ \ln(Tipping\ Amount\_New\ Register)$	0.142***			0.157***		
	(0.029)			(0.038)		
$\Delta\ \ln(Tipping\ Amount\_First\ Appeared)$		0.253***			0.223***	
		(0.055)			(0.073)	
$\Delta\ \ln(Tipping\ Amount\_Return)$			0.285***			0.262***
			(0.057)			(0.088)
N	92	92	92	92	92	92
Adjusted R2	0.200	0.179	0.210	0.151	0.084	0.080
Panel C: City Sample 302						
$\Delta\ \ln(Tipping\ Amount\_New\ Register)$	0.135***			0.121***		
	(0.018)			(0.021)		
$\Delta\ \ln(Tipping\ Amount\_First\ Appeared)$		0.374***			0.349***	
		(0.031)			(0.036)	
$\Delta\ \ln(Tipping\ Amount\_Return)$			0.358***			0.411***
			(0.032)			(0.039)
N	302	302	302	302	302	302
Adjusted R2	0.153	0.323	0.294	0.098	0.240	0.268

Table A14. **Robustness test results using the growth of new users during the capital raising event period to instrument market growth.** This table is one of the robustness check of Table 5, changing the calculation method of dependent variable $\Delta\, Gini$. To exclude the interference of the Covid-19 pandemic, we first delete the months after January 2020. Then, we divide the remaining months into three-time stages based on the nodes of the launch of capital raising (i.e., 2019m1) and its closing (i.e., 2019m7). We then refer to the period before 2019m1 as *Stage 1*, the period between 2019m1 and 2019m7 as *Stage 2*, and the period post 2019m7 as *Stage 3*. We run the regressions $\Delta\, Gini_{j,stage} = \alpha + \beta\Delta \ln(\widehat{Market\, Size}_{j,stage}) + \epsilon_j$, where $\Delta\, Gini$ is the difference in the average monthly Gini coefficients of the city j between two stages, and $\Delta \ln(Market\, Size)$ is the difference in the average monthly size of the tipping amount between the two stages. In Panel A, *Gini*0 is used to calculate $\Delta\, Gini$, based on the income distribution of streamers whose incomes are positive from city j. In Panel B, *Gini*500 is used to calculate $\Delta\, Gini$, based on the income distribution of streamers whose incomes are higher than 500 RMB from city j. $\Delta \ln(Tipping\, Amount\_New\, Users)$ is used to instrument the variation of whole market size $\Delta \ln(Tipping\, Amount)$. We use the *New Registered* method to define *New Users* — if a newly registered user in month t converts to a paying user that month, then the user will be defined as *New* in month t. In Columns (1), (2), and (3), We use the value difference of the variables between *Stage 2* and *Stage 1*. And in Columns (4), (5), and (6), We use the value difference of the variables between *Stage 3* and *Stage 2*. And we also use different city samples. Robust standard errors are reported in parentheses. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
	Stage 2 - Stage 1			Stage 3 - Stage 2		
Panel A: $\Delta\, Gini0$						
$\Delta \ln(Tipping\, Amount)$	0.044***	0.025***	0.020***	0.028***	0.025***	0.028***
	(0.006)	(0.006)	(0.004)	(0.005)	(0.005)	(0.006)
Adjusted R2	0.773	0.556	0.471	0.723	0.671	0.600
Panel B: $\Delta\, Gini500$						
$\Delta \ln(Tipping\, Amount)$	0.166***	0.077***	0.116***	0.107***	0.127***	0.151***
	(0.023)	(0.026)	(0.019)	(0.025)	(0.019)	(0.018)
Adjusted R2	0.779	0.492	0.662	0.719	0.775	0.764
N	18	92	302	18	92	302
K. P. F-statistic	19.884	11.136	34.721	23.304	13.767	28.176

Table A15. **Robustness test results using the growth of new users during the capital raising event period to instrument market growth.** This table is one of the robustness check of Table 5, changing the calculation method of instrument variable Δ ln(*Tipping Amount_New Users*). To exclude the interference of the Covid-19 pandemic, we first delete the months after January 2020. Then, we divide the remaining months into three-time stages based on the nodes of the launch of capital raising (i.e., 2019m1) and its closing (i.e., 2019m7). We then refer to the period before 2019m1 as *Stage 1*, the period between 2019m1 and 2019m7 as *Stage 2*, and the period post 2019m7 as *Stage 3*. We run the regressions $\Delta\, Gini_{j,stage} = \alpha + \beta\Delta\ \ln(\widehat{Market\ Size}_{j,stage}) + \epsilon_j$, where Δ *Gini* is the difference in the average monthly *Gini*50 of the city j between two stages, and Δ ln(*Market Size*) is the difference in the average monthly size of the tipping amount between the two stages. In Panel A, we use the *First Appeared* method to define *New Users* — if a user makes a tip for the first time in month t during our sample period, then the user will be defined as a new user. In Panel B, the *Return* method is used to define *New Users* — if one has tipping records in month t, but does not tip during period $[t-3, t-1]$, then the user will be defined as new in month t. In Columns (1), (2), and (3), We use the value difference of the variables between *Stage 2* and *Stage 1*. And in Columns (4), (5), and (6), We use the value difference of the variables between *Stage 3* and *Stage 2*. And we also use different city samples. Robust standard errors are reported in parentheses. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
	Δ *Gini*50					
	Stage 2- Stage 1			Stage 3- Stage2		
Panel A: IV Δ ln(*Tipping Amount_First Appeared*)						
Δ ln(*Tipping Amount*)	0.092***	0.061***	0.096***	0.028	0.078***	0.105***
	(0.015)	(0.015)	(0.007)	(0.040)	(0.014)	(0.008)
N	18	92	302	18	92	302
Adjusted R2	0.793	0.584	0.730	0.413	0.807	0.783
K. P. F-statistic	16.221	7.969	102.757	3.31	7.169	60.992
Panel B: IV Δ ln(*Tipping Amount_Return*)						
Δ ln(*Tipping Amount*)	0.092***	0.062***	0.098***	0.037	0.090***	0.108***
	(0.015)	(0.014)	(0.007)	(0.031)	(0.012)	(0.007)
N	18	92	302	18	92	302
Adjusted R2	0.793	0.588	0.731	0.528	0.802	0.784
K. P. F-statistic	15.572	7.769	84.895	2.683	6.333	74.944

Table A16. **The response of live streaming use to Covid-19 shock.** This table reports the response of live streaming use to Covid-19 shock. We run the following regression $Y_{j,t} = \alpha + \beta ln(Distance_j) \times Post_t + \lambda Post_t + \theta_j + \epsilon_{j,t}$. In Columns (1) and (4), $Y_{j,t}$ is the log number of viewers in city j in day t. In Columns (2) and (5), $Y_{j,t}$ is the log number of paying users in the city j in Day t. In Columns (3) and (6), $Y_{j,t}$ is the log number of streamers in city j in Day t. And in Columns (1), (2), and (3), $Distance_j$ is the driving time (in days) between Wuhan and city j, as shown on Baidu Maps. In Columns (4), (5), and (6), $Distance_j$ is the straight-line distance (in thousand kilometers) calculated from the latitude and longitude of Wuhan and city j. $Post_1 = 1$ indicates the days after December 30, 2019 (Day -24) when Dr. Wenliang Li posted a Covid-19 alert on one of his WeChat groups and Covid-19 first became known to the public. $Post_2 = 1$ indicates the days after January 23, 2020 (Day 0), when the government imposed a lockdown in Wuhan. $Post_3 = 1$ indicates the days after February 1, 2020 (Day 9), when the government started implementing a series of escalating measures. θ_j represents city-level fixed effects. This regression estimation is based on the 79 days (from Day -50 to Day 28) around the Wuhan lockdown. The city sample includes nine cities that account for more than 2% of the number of users: Shanghai, Beijing, Guangzhou, Chengdu, Hangzhou, Wuhan, Shenzhen, Chongqing, and Changsha. Standard errors clustered at the city level are reported in parentheses. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

	(1) ln (Num of Paying Users)	(2) ln (Tipping Amount)	(3) ln (Num of Streamers)	(4) ln (Num of Paying Users)	(5) ln (Tipping Amount)	(6) ln (Num of Streamers)
			Panel A: Post (Wenliang Li's Alert)			
ln (Distance) * $Post_1$	0.500***	0.644	0.191**	0.299***	0.349	0.116**
	(0.123)	(0.591)	(0.075)	(0.088)	(0.362)	(0.048)
$Post_1$	-0.349***	-0.500*	-0.201***	-0.333***	-0.462	-0.196***
	(0.028)	(0.258)	(0.017)	(0.039)	(0.253)	(0.020)
Adjusted R-squared	0.568	0.585	0.805	0.567	0.584	0.805
			Panel B: Post (Wuhan Lockdown)			
ln (Distance)* $Post_2$	0.741***	1.346***	0.539***	0.455***	0.852***	0.335***
	(0.185)	(0.183)	(0.133)	(0.133)	(0.117)	(0.093)
$Post_2$	-0.519***	-0.791***	-0.316***	-0.502***	-0.772***	-0.305***
	(0.059)	(0.060)	(0.045)	(0.069)	(0.068)	(0.050)
Adjusted R-squared	0.681	0.619	0.837	0.678	0.618	0.834
			Panel C: Post (Escalated Measures)			
ln (Distance) * $Post_3$	1.084***	1.838***	0.790***	0.656***	1.148***	0.491**
	(0.238)	(0.197)	(0.232)	(0.180)	(0.192)	(0.162)
$Post_3$	-0.497***	-0.834***	-0.347***	-0.467***	-0.801***	-0.332***
	(0.078)	(0.039)	(0.089)	(0.095)	(0.068)	(0.094)
Adjusted R-squared	0.565	0.590	0.794	0.559	0.588	0.790
N	711	711	711	711	711	711
City FEs	Yes	Yes	Yes	Yes	Yes	Yes

Table A17. **Summary statistics of daily level data sample.** This table shows the summary statistics of variables used in Section 5.4. $Gini0_{j,t}$ is calculated using the income distribution of all streamers who receive any virtual gifts from users in the city j from Day $(t-6)$ to Day t; $Gini50_{j,t}$ is calculated using the income distribution of streamers who receive virtual gifts worth more than 50 RMB from citizens in the city j from Day $(t-6)$ to Day t; $PL_{j,t}$ is calculated based on the income distribution of top 10 streamers ranked using the aggregated value of virtual gifts sent from citizens in the city j from Day $t-6$ to Day t; $Number\,of\,Paying\,Users_{j,t}$ refers to the number of users located in city j who send out any virtual gifts at Day t and the location city is identified by the IP address where the user views live streaming most often in that month; $Tipping\,Amount_{j,t}$ refers to the total amount of virtual gifts (in thousand RMB) sent by citizens in city j on Day t. $Number\,of\,streamers_{j,t}$ refers to the total number of streamers who receive any virtual gifts from citizens in the city j on Day t; $Duration_j$ is the driving time (in days) between Wuhan and city j, shown on Baidu Maps; and $Distance_j$ is the straight-line distance (in thousand kilometers) calculated from the latitude and longitude of Wuhan and city j. This summary statistics are based on 79 days (from Day -50 to Day 28) around the Wuhan lockdown. The city sample includes nine cities that account for more than 2% of total users, respectively: Shanghai, Beijing, Guangzhou, Chengdu, Hangzhou, Wuhan, Shenzhen, Chongqing, and Changsha.

Variable	N	Mean	Sd	Min	P50	Max
Gini0	711	0.936	0.018	0.875	0.938	0.979
Gini50	711	0.851	0.038	0.723	0.853	0.940
PL	711	-0.775	0.276	-2.643	-0.736	-0.323
Number of Paying Users	711	834.094	269.812	362.000	800.000	1704.000
Tipping Amount	711	234.149	211.478	30.842	178.327	2415.620
Number of Streamers	711	487.195	111.033	277.000	458.000	810.000
Duration	711	0.400	0.181	0.000	0.477	0.594
Distance	711	0.676	0.322	0.000	0.759	1.058

Table A18. **The change of Gini0 to Covid-19 shock.** This table reports the change of the Gini coefficient to Covid-19 shock. We run the regression: $Gini0_{j,t} = \alpha + \beta ln(Distance_j) \times Post_t + \lambda Post_t + \delta X_{j,t} + \theta_j + \epsilon_{j,t}$. $Gini0_{j,t}$ is calculated using the income distribution of all streamers who receive any virtual gifts from citizens in the city j from Day $(t-6)$ to Day t. In Columns (1) to (5), $Distance_j$ is the driving time (in days) between Wuhan and city j, as shown on Baidu Maps. In Columns (6) to (10), $Distance_j$ is the straight-line distance (in thousands of kilometers) calculated from the latitude and longitude of Wuhan and city j. $Post_1 = 1$ indicates the days after December 30, 2019 (Day -24) when Dr. Wenliang Li posted a Covid-19 alert on one of his WeChat groups and Covid-19 first became known to the public. $Post_2 = 1$ indicates the days after January 23, 2020 (Day 0), when the government imposed a lockdown in Wuhan. $Post_3 = 1$ indicates the days after February 1, 2020 (Day 9), when the government started implementing a series of escalating measures. We do not control any market size variables in Columns (1) and (6). In Columns (2) and (7), we use the log number of paying users as the measure of market size variable $X_{j,t}$. In Columns (3) and (8), we use the log value of the tipping amount as the measure of market size variable $X_{j,t}$. In Columns (4) and (9), we use the log number of streamers as the measure of market size variable $X_{j,t}$. In Columns (5) and (10), we add a series of market size variables to the regression. θ_j represents city-level fixed effects. This regression estimation is based on the 79 days (from Day -50 to Day 28) around the Wuhan lockdown. The city sample includes nine cities that account for more than 2% of the number of users: Shanghai, Beijing, Guangzhou, Chengdu, Hangzhou, Wuhan, Shenzhen, Chongqing, and Changsha. Standard errors clustered at the city level are reported in parentheses. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

	(1) Gini0	(2) Gini0	(3) Gini0	(4) Gini0	(5) Gini0	(6) Gini0	(7) Gini0	(8) Gini0	(9) Gini0	(10) Gini0
Panel A: Post (Wenliang Li's Alert)										
ln (Distance) * $Post_1$	0.038**	0.023	0.028**	0.029**	0.022*	0.022**	0.013	0.017**	0.017	0.013
	(0.012)	(0.014)	(0.008)	(0.013)	(0.010)	(0.009)	(0.009)	(0.007)	(0.009)	(0.007)
$Post_1$	-0.016**	-0.006	-0.008**	-0.007	-0.004	-0.014**	-0.004	-0.007	-0.005	-0.003
	(0.005)	(0.005)	(0.003)	(0.005)	(0.004)	(0.006)	(0.006)	(0.004)	(0.006)	(0.005)
ln (Num of Paying Users)		0.029***			0.019**		0.029***			0.019**
		(0.005)			(0.007)		(0.005)			(0.007)
ln (Tipping Amount)			0.015***		0.011***			0.015***		0.012***
			(0.002)		(0.003)			(0.002)		(0.003)
ln (Num of Streamers)				0.044**	-0.005				0.045**	-0.005
				(0.015)	(0.021)				(0.015)	(0.021)
Adjusted R-squared	0.301	0.419	0.448	0.356	0.482	0.297	0.417	0.446	0.353	0.481
Panel B: Post (Wuhan Lockdown)										
ln (Distance) * $Post_2$	0.043***	0.027*	0.026**	0.036**	0.028**	0.028***	0.018*	0.017**	0.023**	0.018**
	(0.011)	(0.014)	(0.010)	(0.012)	(0.012)	(0.007)	(0.008)	(0.006)	(0.007)	(0.007)
$Post_2$	-0.024***	-0.013**	-0.014***	-0.020***	-0.014**	-0.024***	-0.013**	-0.014***	-0.020***	-0.014**
	(0.003)	(0.004)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)
ln (Num of Paying Users)		0.022***			0.018**		0.022***			0.018**
		(0.005)			(0.006)		(0.005)			(0.006)
ln (Tipping Amount)			0.012***		0.011***			0.012***		0.011***
			(0.003)		(0.003)			(0.003)		(0.003)
ln (Num of Streamers)				0.013	-0.024				0.014	-0.023
				(0.018)	(0.020)				(0.017)	(0.020)
Adjusted R-squared	0.376	0.424	0.465	0.379	0.486	0.376	0.425	0.465	0.379	0.486
Panel C: Post (Escalated Measures)										
ln (Distance) * $Post_3$	0.025	-0.009	-0.005	-0.012	-0.017	0.018	-0.002	-0.001	-0.005	-0.007
	(0.021)	(0.033)	(0.024)	(0.039)	(0.032)	(0.014)	(0.022)	(0.016)	(0.025)	(0.021)
$Post_3$	-0.009	0.007	0.005	0.007	0.010	-0.009	0.005	0.003	0.006	0.008
	(0.008)	(0.013)	(0.009)	(0.015)	(0.012)	(0.008)	(0.013)	(0.009)	(0.015)	(0.012)
Log Num of Paying Users		0.031***			0.021**		0.031***			0.021**
		(0.008)			(0.007)		(0.008)			(0.007)
ln (Tipping Amount)			0.016***		0.012***			0.016***		0.012***
			(0.002)		(0.003)			(0.002)		(0.003)
ln (Num of Streamers)				0.047**	-0.003				0.045**	-0.005
				(0.015)	(0.011)				(0.015)	(0.010)
Adjusted R-squared	0.281	0.417	0.442	0.346	0.483	0.283	0.416	0.442	0.345	0.482
N	711	711	711	711	711	711	711	711	711	711
City FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table A19. **The change of PL exponent to Covid-19 shock.** This table reports the change of PL exponents to Covid-19 shock. We run the regression: $PL_{j,t} = \alpha + \beta ln(Distance_j) \times Post_t + \lambda Post_t + \delta X_{j,t} + \theta_j + \epsilon_{j,t}$. $PL_{j,t}$ is calculated based on the income distribution of the top 10 streamers ranked using the aggregated value of virtual gifts sent from citizens in city j from Day $(t-6)$ to Day t. In Columns (1) to (5), $Distance_j$ is the driving time (in days) between Wuhan and city j, as shown on Baidu Maps. In Columns (6) to (10), $Distance_j$ is the straight-line distance (in thousands of kilometers) calculated from the latitude and longitude of Wuhan and city j. $Post_1 = 1$ indicates the days after December 30, 2019 (Day -24) when Dr. Wenliang Li posted a Covid-19 alert on one of his WeChat groups and Covid-19 first became known to the public. We do not control any market size variables in Columns (1) and (6). In Columns (2) and (7), we use the log number of paying users as the measure of market size variable $X_{j,t}$. In Columns (3) and (8), we use the log value of the tipping amount as the measure of market size variable $X_{j,t}$. In Columns (4) and (9), we use the log number of streamers as the measure of market size variable $X_{j,t}$. In Columns (5) and (10), we add a series of market size variables to the regression. θ_j represents city-level fixed effects. The regression estimation is based on 79 days (from Day -50 to Day 28) around the Wuhan lockdown. And the city sample includes nine cities that account for more than 2% of total users: Shanghai, Beijing, Guangzhou, Chengdu, Hangzhou, Wuhan, Shenzhen, Chongqing, and Changsha. Standard errors are clustered at the city level and reported in parentheses. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

	(1) PL	(2) PL	(3) PL	(4) PL	(5) PL	(6) PL	(7) PL	(8) PL	(9) PL	(10) PL
ln (Distance) * $Post_1$	0.296**	0.125	0.200	0.209	0.117	0.165*	0.062	0.113	0.112	0.061
	(0.105)	(0.114)	(0.133)	(0.125)	(0.137)	(0.085)	(0.079)	(0.095)	(0.095)	(0.092)
$Post_1$	-0.133**	-0.013	-0.058	-0.041	-0.013	-0.118*	-0.003	-0.049	-0.028	-0.005
	(0.043)	(0.054)	(0.063)	(0.083)	(0.092)	(0.055)	(0.059)	(0.068)	(0.092)	(0.094)
ln (Num of Paying Users)		0.343**			0.291***		0.345**			0.292***
		(0.109)			(0.073)		(0.108)			(0.073)
ln (Tipping Amount)			0.149**		0.097**			0.150**		0.098**
			(0.055)		(0.037)			(0.054)		(0.037)
ln (Num of Streamers)				0.456	-0.150				0.460	-0.150
				(0.338)	(0.311)				(0.337)	(0.311)
Adjusted R-squared	0.207	0.277	0.266	0.232	0.295	0.206	0.277	0.266	0.231	0.295
N	711	711	711	711	711	711	711	711	711	711
City FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes