Income Inequality and Digital Platform Size: Evidence from Administrative Data from Chinese Live-streaming Platform[*]

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

June 1, 2026

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 and some digital platforms: NBA top players, Forbes celebrities, Netflix 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 1.9% to 5.2% estimated from the cross-

[*]We would like to thank Mingzhu Tai, Haishi Harry Li, Michael Wong, John Klopfer, Naijia Guo, Jasmine Hao, Hsin-Tien (Tiffany) Tsai, Yanhui Wu, Will Shuo Liu, Han Qiu, Grace Xing Hu, Fudong Zhang, Zhen Zhou, Xun Zhang, and also seminar participants at Cambridge Centre for Alternative Finance Annual Conference, 2023 Hong Kong Conference for Fintech, AI, and Big Data in Business, 4th International Conference on Digital, Innovation, Financing & Entrepreneurship, 2023 China Financial Research Conference 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.

[§]PBC School of Finance, Tsinghua University, Email:liaol@pbcsf.tsinghua.edu.cn.

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

[§]PBC School of Finance, 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.

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 5.6% to 8.1%.

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

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

1 Introduction

Digital economy is a game-changer, fundamentally differentiating itself from traditional economies by significantly reducing five key economic costs: search, replication, transportation, tracking, and verification, as highlighted by Goldfarb and Tucker (2019). These cost reductions revolutionize economic activity, fostering efficiency, innovation, and global reach, thereby making the digital economy critically important for modern society.

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. Over the past decade, the creator economy has seen exponential growth, now capturing scholarly attention (Jain and Qian, 2021; Mitchell, 2021; Bhargava, 2022; Cong and Li, 2023). 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]. Influencers in some cases earn seven-figure revenues. Accidentally leaked Twitch data in 2021 exposed the multi-million dollar earnings of notable streamers, including xQc, Asmongold, and Ludwig, further underscoring this industry's financial potency[3]. 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

[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.

[3]See `https://dotesports.com/streaming/news/full-list-of-all-twitch-payouts-twitch-leaks` for detail.

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 Fnbaorbes list, movie stars, singers, and NBA 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.[4] 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

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

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. On the demand side, we investigate the allocation of users to streamers and the formation of their loyal relationships to provide 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. On the supply side, we find that leading streamers often exert considerable effort to occupy their audience's leisure time in order to capture a larger market share. These popular streamers are often motivated to work harder and longer hours to deliver content that continually engages their followers. This increased investment of time can serve to cement their market dominance.

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.[5] For identification purposes, we explore four quasi-experimental variations: market

[5]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.

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.[6] 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 their positions.[7] We find that total tipping increases by 37% in quarter-end months.[8] 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

[6] 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.

[7] 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.

[8] 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.

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 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.[9] 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 digitalization 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,

[9]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).

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; Yang, 2013; Bar-Isaac, Caruana, and Cuñat, 2012) 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),[10] and the largest firms in the world have a PL exponent of approximately one (Gabaix and Landier, 2008; Gabaix, 2009).[11] 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; **?**). 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

[10] 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.

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

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.[12] 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 amplify a slight talent difference into a significant income gap.[13] The urban literature has extensively documented the positive relationship between city size and income inequality with explanations based on agglomeration, technology adoption, and migration.[14] This paper contributes to the literature by exploiting four exogenous variations (local development,

[12]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.

[13]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.

[14]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).

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. 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 (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 explores time-series and cross-sectional shocks to live streaming revenue and estimates the income inequality response to market expansion. Section 5 explores the underlying labor supply mechanism. Section 6 concludes the paper.

2 Institutional Background

2.1 Live Streaming

2.1.1 Live Streaming and the Income of Internet Celebrities

Live streaming is an innovative leisure-enhancing technical 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.[15] 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[16]. 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.[17] 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.[18] 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.

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

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

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

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

The live streaming economy has evolved 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.[19] The most popular tiktoker Charli D'Amelio earned $17.5 million in 2021; Dixie D' Amelio, in second place on TikTok, earned $10 million; and Addison Rae, in third place, earned $5 million.[20] 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. Moreover, streamer "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]

2.1.2 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 **?** exploits the historic rollout of television to test

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

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

[21]Data collected from Douyu app.

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, streamers can work harder to occupy the market. Contrary to platforms such as Netflix and TikTok, the salient feature of live-streaming is its real-time attribute which inherently offers a form of exclusivity. This unique aspect implies that users are incapable of concurrently viewing live-stream services provided by two different streamers, which compels users to make a choice between streamers. Streamers, understanding this exclusivity, can strategically monopolize the leisure time of their audience by extending the duration of their live broadcasts. This tactic not only increases their share of audience attention but also potentially enhances their economic gain in the platform's ecosystem. Moreover, 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 also 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.

2.1.3 The Live Streaming Platform and its Business Model

We obtain administrative data from one of the Chinese most popular live streaming platforms[22]. 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). Attractive streamers typically leverage leverage their personal appeal and proactively foster audience engagement. They engage in chats, performances, and games to captivate viewers. Some stream live outdoor activities or from tourist spots, providing immersive experiences. Regardless of the content, interaction patterns are consistent, with streamers responding to queries, chatting, and accepting compliments and gifts. Strict regulations are enforced globally on live content, ensuring community and legal standards are met. Violations involving inappropriate or illegal content face penalties, from content removal to account suspension, and potential legal consequences.

Two primary types of users: There are two primary types of users participate in the

[22]Our sample platform stands as a representative in the live-streaming industry for several reasons. It's one of the largest platforms in China, attracting millions of viewers and hosting a vast array of content, from gaming to lifestyle. Despite the recent advent of live-streaming services on various short video platforms, the sample platform we studied maintained a substantial market share during our sample period. It's reported that the platform held over 40% of the live-streaming market in 2018. And its business model, including monetization strategies of tipping virtual gifts, is typical for this industry. Furthermore, it's been a forerunner in technological advancements for live-streaming, enhancing the viewing experience. The platform also hosts prominent streamers, embodying the career trajectory and challenges faced by streamers industry-wide. Therefore, studying our sample platform provides comprehensive insights into the live-streaming industry.

platform's activities —— viewers and streamers. Viewers functions as the demand side in this digital live-streaming marketplace. After undergoing a verification process that includes real-name authentication, users can transition from the demand to the supply side of the market by becoming a streamer and initiating their own live-streaming channel. Viewers engage in real-time viewing of a streamer's live stream and interact with the streamers and each other, primarily through bullet chats, fostering a dynamic community within each live-stream. Viewers can also voluntarily 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 gifts. Streamers can redeem virtual gifts for fiat money from the platform.

Possible measurement errors in income: With the data limitation, we can only observe the streamer's income generated on the live-streaming platform. Although the tipping income obtained from the sample platform serves as a reasonable proxy for streamer total income due to non-compete restrictions[23], it's important to note that there are still inherent measurement errors in income estimation. And there are multiple channels which may introduce biases in our estimation of the streamer's total income. **First**, streamers do not receive the full monetary value of the gifts bestowed by the audience. 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.[24] Beyond the profit-sharing arrangement with the platform, streamers associated with talent agencies may also share a portion of their earnings with these agencies, as determined by contractual agreements. And the top streamers may have a higher ratio of sharing revenue as a result of their greater bar-

[23]Non-compete restrictions in the live streaming industry often arise from contractual terms and technological discrepancies between platforms. Live streaming platforms may stipulate non-compete restrictions in their agreements with streamers, preventing them from offering similar services on rival platforms for a defined duration. This strategy is designed to shield the platform's interests from audience diversion. Technologically, differences in streaming methodologies across platforms can limit simultaneous multi-platform broadcasting, with some platforms establishing technical barriers to this end.

[24]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."

gaining power. **Second**, our dataset may omit certain revenue that streamers derive from their live-streams. In specific instances, some streamers may establish direct contact with their fans via platforms like WeChat or Alipay, facilitating transactions that circumvent the live-streaming platform. Such arrangements may serve two purposes: they can potentially avoid taxation, [25] and prevent the live-streaming platform from obtaining a share of the earnings. **Third**, the top streamers may have other types of income. With the burgeoning popularity of live-streaming, the boundaries between top-tier streamers, and traditional stars have begun to blur. Top streamers may receive remuneration for their presence at events or endorsements, often in the form of brand ambassadorship fees. To explore the relationship between income inequality as measured by tipping income and actual total income disparity, we turn to insights provided by the Forbes list of The World's Highest-paid Athletes. This list accounts for athletes' on-field and off-field earnings. Specifically, on-field earnings encompass prize money, salaries, and bonuses, while off-field earnings are estimated to include sponsorship deals, appearance fees, memorabilia and licensing income, along with cash returns from businesses in which the athlete has a significant stake. By plotting the distribution of top athletes' on-field and total income, employing the Power Law curve, we found that the PL exponent for both categories are not significantly different. As illustrated in Figure A1, the PL exponent for on-field income stands at -2.97. When off-field income is incorporated, the PL exponent for the total income is observed to be -2.64. Consequently, we hypothesize that relying solely on tipping income as an inequality metric may provide a representative view of the actual distribution of earnings among streamers.

[25] The live streamers pay a progressive income tax in China.
See http://beijing.chinatax.gov.cn/bjswj/c104182/202206/1aab3431d9aa46e9978e1ab49e3c1de6.shtml for detail.

2.2 Data Description

The data contain the complete online transaction (tipping) records of 51.24 million viewers, 2.44 million paying users,[26] and 1.42 million registered streamers[27] 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 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.[28] 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.[29]

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

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

[27] In our empirical study, the sample size of streamers is considerably smaller. On one hand, a mere 9.29% of these streamers earn income from audience. On the other hand, some streamers demonstrate a lack of continuity in their broadcasting duration. We define 15,576 streamers as *full-time* streamers, characterized by an average live streaming duration exceeding 40 hours over a three-month period.

[28] 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.

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

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.[30]

In Appendix Figure A2, we show the value of the virtual gifts received by the streamers in the sampled during the sample period.[31] 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.[32]

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 A5 plots the number of streamers who earn any virtual gifts has decreased over time[33], 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, Theil index, and per-

[30] 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.

[31] In Appendix Figure A3, 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.

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

[33] Appendix Figure A4 shows the picture of the number of streamers and their average tipping income before the seasonality adjustment.

centile income gaps, and document rising inequality from April 2018 to April 2020. In our principal findings, we typically utilize the streamers' income as a criterion for sample selection. However, we have also imposed constraints concerning the duration of the streamers' streaming time, and the results are robust.

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.[34] 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.

3.2 Gini Coefficients

We use the Gini coefficient as the measure of inequality, as is common in the literature.[35] Figure 2 plots the seasonality-adjusted inequality trend as measured by the Gini index.[36] 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

[34] Appendix Figure A6 shows the top group's share of income before the seasonality adjustment.

[35] 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.

[36] Appendix Figure A7 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.

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. Further, Figure 3 illustrates the dynamic trend in inequality among top streamers, revealing significant disparities in income even within this group. Since some participants may treat streaming more as a leisure activity rather than a serious occupation, the calculation of the Gini coefficient employing minimal income restrictions may introduce bias. A more nuanced approach that focuses on the inequality among top-ranking streamers alleviates this problem to some degree, as those who rank higher are typically more diligent in their work.

In order to isolate the impact of non-professional streamers, we refine our sample to include only *full-time* streamers with a monthly broadcasting duration exceeding 40 hours. Subsequently, we recalculate the Gini coefficient for their tipping income. Referring to Appendix Figure A8, we discern that imposing restrictions on streamers' streaming time scarcely alters the Gini coefficient among high-income streamers. However, a reduction in inequality coefficient is observable for the lower-income streamer cohort. For instance, the Gini coefficient for streamers earning over 1,000 RMB in the past three months witnessed an escalation from 0.84 to 0.91 during the sample period, as depicted in Figure 2. In contrast, when focusing solely on full-time streamers with earnings exceeding 1,000 RMB in the same period, the Gini coefficient rose from 0.76 to 0.86. Notwithstanding these variations, the trend of rising income inequality remains robustly present.

Additionally, we compute the Gini coefficient at the MCN level. There indeed exists support from MCNs behind some streamers, establishing a mutually beneficial relationship. Streamers benefit from the resources and professional expertise offered by MCNs, while MCNs derive income from collaboration with streamers. From Appendix Figure A9, we discover a significant correlation between inequality at the MCN level and that at the streamer level. The correlation between the Gini coefficient among the top 10 entities is as high as

0.84. This relationship becomes even more pronounced for the top 100 and 1000 entities, where the correlation reaches 0.98. Additionally, since the inequality coefficient calculated at the MCN level is greater, our estimation at the individual level may underestimate the actual degree of inequality.

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.[37] 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 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 A10 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.[38]

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

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

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

streamers is much flatter than the PL slope of Forbes billionaires[39], 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[40] 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.

In the preceding analysis, it is discovered that the degree of inequality amongst streamers is significantly greater than that found among traditional entertainment and sports stars. With this in mind, one might consider a comparative analysis with other digital platforms. Do the unique characteristics of live streaming technology, such as its exclusivity and tipping mechanisms, engender a degree of inequality among streamers that surpasses that observed on video digital platforms like Netflix and TikTok? Figure 6 entails a comparison of various indicators representing different digital media platforms, which include Netflix, Douyin[41], and Kuaishou (a leading video-sharing platform in China). For Netflix, the metric pertains to the most viewed films and television series based on total hours viewed. In the case of Douyin, the focus is on the count of likes accrued by short videos uploaded

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

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

[41]Douyin is a free video players app, which is known in the international scene as Tiktok. The app, which means "Shaking Sound" in Chinese, allows users to create, edit, and share short videos.

during the period from January 1, 2019, to June 30, 2019. Meanwhile, for Kuaishou, the metric evaluated is the follower count of its top content creators. We examine the relationship between log-rank and log-size. The results show that the slope for these platforms is markedly steeper compared to our live streaming platform. This stark distinction suggests that the inherent characteristics of live streaming technology contribute to a greater level of inequality. We acknowledge that the degree of comparability in this study is not entirely precise due to data limitation. For instance, the level of inequality for live-streaming streamers is calculated based on their income, while for Netflix, it's computed based on viewing hours. Despite these discrepancies, this analysis provides a glimpse into the varied degrees of inequality across different industries and platforms.

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 7.

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

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.[42]

3.4 Theil Index

The Theil index T is a measure of economic inequality. It is one of the many measures used to gauge the dispersion of income or wealth in a particular population. $T = 0$ indicates perfect equality (everyone has the same income). And higher values of T indicate greater levels of inequality. In Appendix Figure A20, we initially impose an income constraint

[42]Even if we plot the raw coefficients before the seasonality adjustment, the upward trend of the PL exponent is clear; see Appendix Figure A11.

on the sample of streamers to investigate the dynamic changes in inequality among those earning above a certain threshold over time. Consistent with previous sections, we observe a robust increase in income dispersion among these streamers, as measured by the Theil index. In Appendix Figure A21, we further explore the inequality among the top streamers, discovering a widening disparity in their incomes. For instance, the Theil index for the top 100 streamers expanded threefold during the sample period, rising from 0.4 to 1.2. This substantial growth accentuates the trend of increasing dispersion in income among leading streamers.

3.5 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.[43] 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, $j = \arg\max_n \frac{Tip_{int}}{\sum_{n=1}^{N} Tip_{int}}$.[44] For any given month t, we can count the number of fans of each streamer.

3.5.1 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 A12 shows that fans who

[43] 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.

[44] 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.

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 A13 shows the results before the seasonality adjustment.

3.5.2 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 3.5.1. Appendix Figure A14 and A15 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 A16. These results are consistent with Section 3.2.

3.5.3 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 A17, 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 A18 and A19 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.

3.6 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.

3.6.1 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 8, we show the time trend of the percentile gaps. Panel A of Figure 8 shows that the gap between the superstar group and the middle-income group is widening. In addition, according to Panel B of Figure 8, 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.

3.6.2 Income Distribution

We further describe the change in the number of streamers who can earn a certain amount of money. Figure 9 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.

3.6.3 Correlation Matrix

Do these inequality indicators co-move with each other? Appendix Figure A22 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 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: quarter-end revenue spikes induced by the seasonal incentive regime, user surge in the first half of 2019 before a significant capital raising event, potential market size proxied by economic development and Fintech coverage, and the Covid-19 lockdown in Wuhan.

4.1 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.[45] In this section, we use quarter-end dummies as time-series instrumental variables to identify the effect of market size on inequality.

Appendix Figure A3 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 A4, 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

[45] 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.

paying users every quarter end, suggesting the platform's incentive mechanically leads to market expansion.

4.1.1 Time Series Estimation

We first test whether Gini coefficients also spike at the end of each quarter (March, June, September, and December) in the aggregated monthly data spanning 25 months from April 2018 to April.[46] Specifically, we run the following regression:

$$Gini_t = \alpha + \beta QuarterEnd_t + \eta_t + \epsilon_t, \tag{2}$$

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 the Columns (1) and (4) of Appendix Table A4, $Gini_t$ 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 (2), the β is 0.027 ($s.e.$=0.007), which implies an approximately 1.125 standard deviations increase at the quarter end.[47] After adding year-quarter fixed effects in Column (5), the β is 0.029($s.e.$=0.006) and still significantly positive. We

[46]To compute the monthly aggregate Gini coefficient, first sum the monthly tipping income for each streamer. For every month, order the tipping income amounts in ascending order. Using this ordered data, calculate the cumulative proportion of total income received by each cumulative proportion of streamers. The Gini coefficient is then derived as twice the area between the 45-degree line (line of equality) and the Lorenz curve, formed using the cumulative proportions. This process is repeated monthly. Summary statistics for the variables are presented in Panel A of Table 2. Based on streamers with monthly tipping income exceeding 50 RMB, the Gini coefficient ranges from 0.860 to 0.949, with an mean of 0.901 and a standard deviation of 0.024. For those with tipping income above 500 RMB, the Gini coefficient varies between 0.785 and 0.920, with a mean of 0.851 and a standard deviation of 0.036.

[47]The Gini coefficient increases by 0.027, equivalent to 1.125 standard deviations (0.027÷0.024).

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 1.139 standard deviations increase, and β rises to 0.044 ($s.e.$=0.008) after controlling for year-quarter fixed effects. Thus, the empirical results show that the magnitude of inequality 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:

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

$MarketSize_t$ 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. We observe that paying users and the number of streamers do not exhibit significant growth at the end of quarters. Thus, we only focus on the total tipping amount (monetary value of virtual gifts) to measure market size in the following main regressions.

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

$$Gini_t = \alpha + \beta_{OLS} MarketSize_t + \eta_t + \epsilon_t, \tag{4}$$

$$Gini_t = \alpha + \beta_{IV} \widehat{MarketSize}_t + \eta_t + \epsilon_t, \tag{5}$$

where $Gini_t$ is the Gini coefficient calculated using the income distribution of streamers in month t. $MarketSize_t$ is $\ln(Tipping\,Amount_t)$, which is the log value of the total tipping

income of all streamers in month t.[48] Panel A presents the estimated elasticities of the Gini coefficients, quantifying the responsiveness of income inequality to changes in market size. Panel B displays the outcomes following Z-score normalization, thereby allowing for a more direct comparison of the coefficients across different identification strategies. The first two columns of each panel report the OLS results estimated in eq.(4). The last two columns in each panel reports the 2SLS results estimated in eq.(5). *Gini*50 is calculated using the income distribution of the streamers who receive more than 50 RMB in tipping income in month t. *Gini*500 is calculated based on the income distribution of the streamers who receive more than 500 RMB in tipping income in month t. Robust standard errors are reported in parentheses. The F-statistics are reported in the last row. 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. The elasticity ranges from 0.072 to 0.117 and is significant at the 1% level. When the market size doubles, *Gini*50 increases by approximately 0.077, corresponding to an increase of 8.5% relative to the mean. Given that the Gini coefficients derived from different streamer samples yield varying means and standard deviations, normalizing the variables enhances comparability. The coefficients in Panel B suggest that a one standard deviation increase in the log-transformed market size leads to a 1.028 to 1.115 standard deviations increase in the Gini coefficient. β_{IV} is larger than the OLS estimates. Moreover, the F-statistic is 21.970, which suggests that *QuarterEnd* is unlikely to be a weak instrument for $\ln(Tipping\ Amount_t)$ (Kleibergen and Paap, 2006).

[48]The market size variable is skewed, making it challenging to fit a linear model. Taking the natural logarithm of market size enable a more accurate linear regression fit.

4.1.2 Panel Estimation

In this section, we use city-month panel observations to estimate the effect of market size on inequality.[49] 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 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 relationship between the log of the tipping amount and the Gini coefficients with data of all 302 cities in Appendix Figure A27. 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 A28. 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 estimate inequality elasticity to market size using the panel data. The first

[49] To compute the Gini coefficient for each city every month, first accumulate the monthly tipping amounts given to each streamer from users residing in that particular city. With these city-specific monthly tipping earnings aggregated for every streamer, arrange the income data in ascending order. Then, using this ordered data, compute the cumulative proportion of total income received by each streamer relative to the overall income of streamers in that city for the month. The Gini coefficient is derived by calculating twice the area between the 45-degree line and the Lorenz curve, which is plotted using these cumulative proportions. This process is systematically repeated for each city and each month to obtain a panel set of Gini coefficients. Summary statistics for the variables are presented in Panel B of Table 2. Based on the sample of streamers with monthly tipping income from a city exceeding 50 RMB, the Gini coefficient ranges from 0.729 to 0.977, with a mean of 0.861 and a standard deviation of 0.047. For those with tipping income above 500 RMB, the Gini coefficient varies between 0.537 and 0.960, with a mean of 0.762 and a standard deviation of 0.080.

stage is estimated as follows:

$$MarketSize_{j,t} = \alpha + \beta QuarterEnd_t + \theta_j + \eta_t + \epsilon_{j,t}, \tag{6}$$

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.[50]

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

$$Gini_{j,t} = \alpha + \beta_{OLS} \ln(Tipping\ Amount_{j,t}) + \theta_j + \eta_t + \epsilon_{j,t}, \tag{7}$$

$$Gini_{j,t} = \alpha + \beta_{IV} \ln(\widehat{Tipping\ Amount}_{j,t}) + \theta_j + \eta_t + \epsilon_{j,t}, \tag{8}$$

$Gini_{j,t}$ 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 from paying users of city j in month t, and $\ln(\widehat{Tipping\ Amount}_{j,t})$ is the predicted value from the first stage from $QuarterEnd_t$. Panel A presents the estimated coefficients of the Gini coefficients response to market size, while Panel B displays the outcomes following Z-score normalization. Columns (1), (2), (5), and (6) report the OLS results estimated in eq. (7).

[50]Robustness results are reported in Appendix Tables A7 and A8. The regression samples include 92 cities and 302 cities, respectively.

Columns (3), (4), (7), and (8) report the 2SLS results estimated in eq. (8).

The coefficients in Panel A suggests IV and OLS yield similar elasticity estimates. When market size doubles, *Gini*50 increases by approximately 9.4% with OLS and 9.2% with IV; *Gini*500 rises by approximately 17.7% with OLS and 16.7% with IV. The coefficients in Panel B suggest that a one standard deviation increase in the log-transformed market size leads to a 1.075 to 1.142 standard deviations. This magnitude is very similar to the 1.028 to 1.115 range obtained in Section **??**.[51]

4.2 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. 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. [52] Third, the platform increased positive news exposure by taking the lead in setting industry norms and proactively embracing the regulation.

[51] Moreover, the F-statistic is 101.869, which suggests that *QuarterEnd* is unlikely to be a weak instrument for $\ln(Tipping\ Amount_{j,t})$.

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

In Figure A30, we plot the average log value of the tipping amount contributed by *new users* each month.[53] To exclude the interference of the Covid-19 pandemic, we 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 A30 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:

$$ln(Tipping\ Amount_New\ Users_{j,t}) = \alpha + \beta_1\ \mathbb{1}(Stage\,2)_t + \beta_2\ \mathbb{1}(Stage\,3)_t + \epsilon_{j,t}, \tag{9}$$

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

[53] 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:

$$\Delta\, ln(Tipping\ Amount)_{j,stage} = \alpha + \beta \Delta\, ln(Tipping\ Amount_New\ Users)_{j,stage} + \epsilon_j, \qquad (10)$$

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:

$$\Delta\, Gini_{j,stage} = \alpha + \beta \Delta\, \ln(\hat{Market\ Size}_{j,stage}) + \epsilon_j, \tag{11}$$

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. $Gini50$ 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 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. Panel A reports the elasticity of Gini coefficients response to market size, while Panel B reports the estimated results after z-score normalization.

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. The coefficients of 0.496 to 0.834 imply that for every one standard deviation increase in the change of market size, we observe a 0.496 to 0.834 standard deviation increase in the change of the Gini coefficient. This highlights a robust positive relationship between shifts in market size and changes in income inequality as captured by the Gini coefficient.

4.3 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. [54] 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 , number of paying users, average tipping amount per user, and number of streamers with positive income from city j.

Figure 10 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 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 A26 presents the correlation between the city size and local market size, and we can observe that the scale of platform use (i.e.,

[54] 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.

log value of tipping amount) in one city is strongly associated with both its population and development level.

Table 6 reports the cross-city ordinary least squares (OLS) and instrumental variable (IV) regression results. We run eq.(12) 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.

$$Gini_j = \alpha + \beta_{OLS} MarketSize_j + \epsilon_j. \tag{12}$$

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. Panel A reports the elasticity of Gini coefficients response to market size, while Panel B reports the estimated results after z-score normalization. The F-statistics suggest that population, local GDP, and digital financial inclusion are unlikely to be weak instrumental variables. These cross-city variations yield consistent estimates of approximately 0.690 to 0.738 — when the market size doubles, the Gini coefficient increases by around 0.7 standard deviations.

4.4 Covid-19 and the Wuhan Lockdown

4.4.1 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.

4.4.2 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.[55]

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 A32, 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.

[55] According to Appendix Figure A31, 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).

Figure 11 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 11 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 from Wuhan users dropped by about 60%. We observe no discontinuity in Chengdu and Shenzhen, more distant cities from Changsha and also less hit by Covid-19.

4.4.3 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. 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. The Gini coefficient at the city-date level is approximately 0.849 with a standard deviation of 0.037.

Appendix Figure A33 provides evidence that market size is positively related to inequality using the city-date 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 7 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 (*Gini*50) 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 0.328 to 0.570 (in standard deviations).

4.5 Robustness checks using full-time sample

In our streamer sample, it is possible that some participants engage in streaming solely for entertainment, without a serious investment of time and effort into this occupation. To address this concern in our primary results, we set a tipping income threshold, incorporating only those streamers who received 50 or even 500 RMB or more into our Gini coefficient calculations. Given the intensely competitive nature of the streaming industry, streamers who do not produce sufficient content struggle to attract a sizable audience. Accordingly, by imposing a tipping income threshold, we effectively filters out a significant portion of non-full-time streamers from our sample.

In this section, we further refine our sample by imposing an additional constraint: only those streamers who have logged a monthly streaming duration of at least 40 hours are included in our Gini coefficient computations. We then replicate the results across four identification strategies using this revised Gini measure as the dependent variable. The results, reported in Appendix Table 8, confirm the robustness of our findings. Gini coefficient

elasticity ranges from 0.238 to 0.753 standard deviations 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 0.406 to 1.131 standard deviations.

4.6 Robustness checks using Gini$_{fans}$

In our primary analysis, we employ tipping amounts to construct a Gini coefficient for income inequality among streamers. Recognizing that the number of followers also constitutes a vital resource for streamers, this section extends the analysis by utilizing a Gini coefficient calculated based on the distribution of each streamer's loyal followers as the dependent variable. We use the total number of followers as a proxy for market size to perform a robustness check on the relationship between market size and inequality. The results are presented in Appendix Table A20. Across all four identification strategies, we find the outcomes to be robustly consistent, reinforcing the main findings.

5 Mechanisms

We hypothesize that the attention competition among live streamers significantly contributes to the observed relationship between market expansion and the escalating income inequality. The exclusive nature of live streaming only permits viewers to concentrate on one streamer at a time, thus prompting streamers to prolong their broadcast duration as a strategy to monopolize viewers' discretionary time.

5.1 Motivating Evidence

The motivation for exploring this mechanism arise from two empirical observations. First, we note that streamers ranked higher tend to exhibit a greater degree of effort, as measured by their broadcasting duration, a trend that persists even among the top-tier streamers. Specifically, we compute streamers' monthly streaming duration based on their broadcast records. We then rank the streamers according to their monthly tipping income

(or income per hour) and plot the relationship between the ranking and streaming duration. The analysis in Appendix Figure A38 reveals a negative correlation, indicating that streamers with higher tipping incomes (or income per hour) tend to stream for a longer duration time. To account for the potential mechanical relationship between streaming time and total tipping income within the same month, we examine the relationship between the ranking in month $t-1$ and the streaming time in month t. Additionally, we control for year-month fixed effects to isolate time-specific influences. The negative association holds even when we narrow the sample to the top ten streamers, demonstrating the robustness of the relationship. Leading streamers often occupy their audience's entire leisure time by broadcasting an average of 6 hours per day, enabling them to capture a larger market share and consequently, accrue higher tipping income. In addition, when exploring the relationship between streaming time and the previous month's performance ranking in Appendix Table A21, we control for both streamer-level fixed effects and year-month fixed effects. This allows us to account for inherent characteristics specific to each streamer. Our findings suggest that even for a given streamer, an improvement in ranking is associated with increased motivation to work harder.

Second, we observe high mobility rates among the top streamers. In traditional occupations where wages are predetermined, an increase in wage typically generates two effects: the substitution effect implies that higher wages make labor more attractive relative to leisure, incentivizing workers to allocate more time to work; the income effect suggests that higher wages lead to increased income, which may permit workers to enjoy more leisure time without sacrificing their standard of living. Consequently, workers at the top of their industry might opt for additional leisure time due to the income effect. However, the live-streaming industry presents a distinct dynamic. Our analysis reveals pronounced mobility rates among streamers, wherein tenure at the top is precarious. Failure to maintain sufficient streaming time and thereby capture audience attention can rapidly result in the loss of a top-tier status. From Appendix Figure A40, with only 60% of the top 100

streamers in month t remaining in the top 100 in month $t+1$, and just 30% remaining after six months. It is evident that without sustained effort, top-tier streamers face the imminent risk of being supplanted by emerging contenders in the market.

5.2 Productivity or Work Harder?

In Section 4, we have found that expanding market size would exacerbate income inequality among them. Then does this inequality emerge from top-tier streamers earning more money per unit of time, or do they command a larger market share because they are willing to work for extended periods? We decompose income into two constituent elements: income per hour and streaming hours. We examine the dispersion in both income per hour and streaming hours to understand the growing income inequality between top-tier and lower-tier streamers. In Appendix Figure A41, we present an empirical investigation into the relationship between market size and streaming hours dispersion across different groups of streamers. Streamers are categorized into four income-based rankings on a monthly basis: Group 1 (rank 1-10), Group 2 (rank 10-100), Group 3 (rank 100-500), and Group 4 (rank 500-1000). To measure dispersion in streaming hours, we calculate the monthly average for each group and then use the ratio of Group 1's average streaming hours to those of Groups 2, 3, and 4, respectively. Our results indicate a widening gap in streaming hours as market size-defined as tipping income per streamer-increases. Specifically, streamers in Group 1 display 1.5 to 3 times the average streaming hours compared to those in Group 4.

In addition, we also examine Appendix Figure A42, which focus on the income per hour disparities across the different income-based groups of streamers. Similar to the streaming hours dispersion, the income per hour also exhibits widening gaps between Group 1 and the lower-ranked groups as market size expands. Despite the top-tier streamers (Group 1) investing more effort, as indicated by longer streaming hours, they also garner substantially higher income per hour. This points to a compounding inequality where top-ranked streamers not only invest more time but also reap higher tipping returns on a per-hour basis. The

multiplicative effects of streaming hours and income per hour suggest a complex interplay of factors reinforcing income inequality in the live-streaming market. This underscores a key distinction between the live-streaming labor market and traditional industries, where wages (tipping per hour) are not predetermined.

5.3 Streaming Time Response to Market Size: Macro Perspective

Given that the intensification of income inequality, to some extent, is attributable to the elongation of live streamers' broadcasting time, we subsequently employ the identification strategy previously mentioned to examine the response of the streamers' streaming time to an expansion in market size.

We use the *Quarter End* dummy variable as an instrumental variable for streamers' tipping income to scrutinize the response of streaming duration. The findings delineated in Table 12 reveal that, after controlling for streamer-level fixed effects and quarter-level fixed effects, streamers significantly increase their broadcasting time in response to a larger market size, with the intention of capturing more of the audience's leisure time. The analysis shows the time-series aggregate labor supply elasticity with respect to total tipping income to be between 19.9% and 24.6% according to the OLS model and between 10.3% and 13.0% using the IV approach. The coefficient of 0.486 to 1.161 after z-score normalization suggests that as the market size increases by one standard deviation from its mean, streaming time rises by 0.486 to 1.161 standard deviations. The responsiveness of streaming hours to market size exhibits a similar economic magnitude to our estimated impact of the Gini coefficient for income on market size. This implies that variations in market size can predict a significant portion of the variations in streaming time. The labor supply response of streamers potentially maximizes their share of viewer attention, and thus their earning potential, amidst an expanding market landscape.

5.4 Streaming Time Response to Market Size: Micro Perspective

Upon exploiting the demand shock to Wuhan triggered by the Covid-19 pandemic, we discern that concurrent with a decrease in the city's market size, the labor supply also contracts by approximately 20% in Wuhan. However, this observed correlation, shown in Figure A37, is somewhat simplistic and does not adequately address the issue of endogeneity. One could argue that Covid-19 may be the largest omitted variable that triggers a reduction in Wuhan's labor supply. This could be due to potential diversions of streamers' attention caused by the pandemic or perhaps the possibility that they were physically incapacitated by the virus and thus unable to broadcast. Consequently, we adopt an approach akin to that of e Bartik instrument, wherein we construct an instrumental variable based on the degree of streamers' dependence on Wuhan in the three months preceding the outbreak of Covid-19. This approach helps alleviate some of the concerns related to endogeneity since we still confine our sample to streamers who are permanently resident in Wuhan. The virus indiscriminately affects streamers and would not be influenced by their prior dependence on the tipping behavior of Wuhan residents. However, it would impact the reduction in market size they face. For Streamer j and date t, we run the following 2SLS regression:

$$\log(Tipping\ Income_{j,t}) = \alpha + \beta Post_t \times Wuhan\ Dependence_j + \epsilon_{j,t}$$

$$log(Streaming\ Time_{j,t}) = \alpha + \beta \log(\widehat{Tipping\ Income}_{j,t}) + \epsilon_{j,t}$$

where the independent variable $Tipping\ Income_{j,t}$, represents the income of streamer j over the preceding 7 days, while the dependent variable, $Streaming\ Time_{j,t}$ measures the daily broadcasting duration of the broadcaster. Our sample is restricted to *full-time* streamers based on their broadcasting duration three months prior to the lockdown in Wuhan (with an average monthly broadcasting time exceeding 40 hours). Moreover, we focus only on those streamers whose primary residence is in Wuhan. The construction formula for the

instrumental variable $Wuhan\ Dependence_j$ is as follows.

$$Wuhan\ Dependence_j = \frac{Tipping\ Amount\ from\ Wuhan\ Users}{Total\ Tipping\ Amount}$$

where the income calculations are based on data from the three months preceding the lockdown. And in our regression analysis, the variable *Wuhan Dependence* is operationalized as a categorical variable, taking values from 1 to 5. This categorization is derived from the quintile ranking of the continuous measure. From results presented in Table 13, we observe that the streaming time response is more pronounced for Wuhan streamers who rely more on tipping from Wuhan local residents. This observation suggests that streamers adjust their labor supply in response to changes in the market size they face.

Our findings present a divergence from the widely-discussed relationship between wages and labor supply in existing literature. Traditional labor economic theories predict that an increase in wages heightens the appeal of work due to the substitution effect; concurrently, it also allows individuals to achieve the same income by working fewer hours, termed as the income effect (Borjas and Van Ours, 2010). However, in the context of live-streaming tips, the streamer does not have a predetermined wage. The amount of tipping received is closely tied to both the content and streaming time of the live-stream, introducing significant uncertainty.

5.5 Discussion of Other Potential Mechanisms

The expansion of market size disproportionately benefits top-tier streamers, thereby exacerbating inequality among streamers. Our primary focus in this paper elucidates the mechanism through which streamers intensify their streaming efforts to capture a larger share of the expanding market. We also acknowledge that market size may influence inequality through various other pathways, such as changes in user loyalty over time, shifts in tipping habits, or the role of algorithm.

In fact, we also investigate viewer behavior from the demand side, specifically focusing

on viewing and tipping loyalty. We find that the allocation of user's financial resources and attention appears to favor leading streamers. The asymmetric distribution of both money and attention among live stream streamers may also exacerbate the prevailing income disparities.

In addition, when investigating the mechanisms underlying the pronounced disparities in income among live stream streamers, the role of algorithms is inescapable (Fleder and Hosanagar, 2009; Tucker and Zhang, 2011; Goldfarb and Tucker, 2019). It is possible that algorithms employed in content recommendation may assign weight to a streamer's historical popularity, and thereby benefit high-profile streamers. Nevertheless, this reflects a phenomenon broadly prevalent in the real world. For instance, upon entering a shopping mall, the first displays that capture one's attention are likely to belong to well-established brands with significant name recognition. In addition, we have endeavored to address the endogeneity problem introduced by the algorithm itself through certain empirical strategies. If income inequality were solely a byproduct of the algorithm, it would be improbable to observe a decrease in the income gap among streamers in Wuhan during the Covid period due to the contraction of the market size.

An important question in our context is whether the algorithm will persistently favor these top-ranking streamers disproportionately as the market size expands. For User-Generated Content (UGC) platforms, high market concentration may impinge upon the diversity of content. This could also lead to the weaker bargaining power of the platforms themselves. In response to these challenges, several measures have actually been undertaken by internet platforms to bolster the position of small to medium-sized content creators. They have sought to broaden the income avenues available to these creators, as evidenced by Douyin's *Xingtu* initiative and Weibo's *Advertisement Sharing Program*. Furthermore, platforms are shifting the emphasis on providing traffic recommendations and financial support to focus on content quality rather than the creator's follower count. This strategic reorientation is exemplified by Douyu's *Star Support Plan* and Bilibili's *New Star*

Plan. Such shifts indicate a move towards greater equity and inclusion, potentially mitigating the impact of algorithmically-induced disparities. The income dispersion continues, even with the platform's support of small and medium-sized content creators.

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 - 0.021), quarter-end shocks to tipping revenue (0.079 - 0.081), user surge induced by the capital raising event (0.056 - 0.077), and the exogenous Covid-19 lockdown in Wuhan (0.023 – 0.052). 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.

Bar-Isaac, Heski, Guillermo Caruana, and Vicente Cuñat. 2012. "Search, design, and market structure." *American Economic Review* 102, 1140–1160.

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*.

Bhargava, Hemant K. 2022. "The creator economy: Managing ecosystem supply, revenue sharing, and platform design." *Management Science* 68, 5233–5251.

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.

Borjas, George J, and Jan C Van Ours. 2010. *Labor economics*. McGraw-Hill/Irwin Boston.

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.

Cong, Lin William, and Siguang Li. 2023. "A model of influencer economy."Technical report, National Bureau of Economic Research.

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.

Fleder, Daniel, and Kartik Hosanagar. 2009. "Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity." *Management science* 55, 697–712.

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.

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.

Jain, Sanjay, and Kun Qian. 2021. "Compensating online content producers: A theoretical analysis." *Management Science* 67, 7075–7090.

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.

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.

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.

Mitchell, Matthew. 2021. "Free ad (vice): Internet influencers and disclosure regulation." *The RAND Journal of Economics* 52, 3–21.

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.

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.

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.

Tucker, Catherine, and Juanjuan Zhang. 2011. "How does popularity information affect choices? A field experiment." *Management Science* 57, 828–842.

Yang, Huanxing. 2013. "Targeted search and the long tail effect." *The RAND Journal of Economics* 44, 733–756.

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

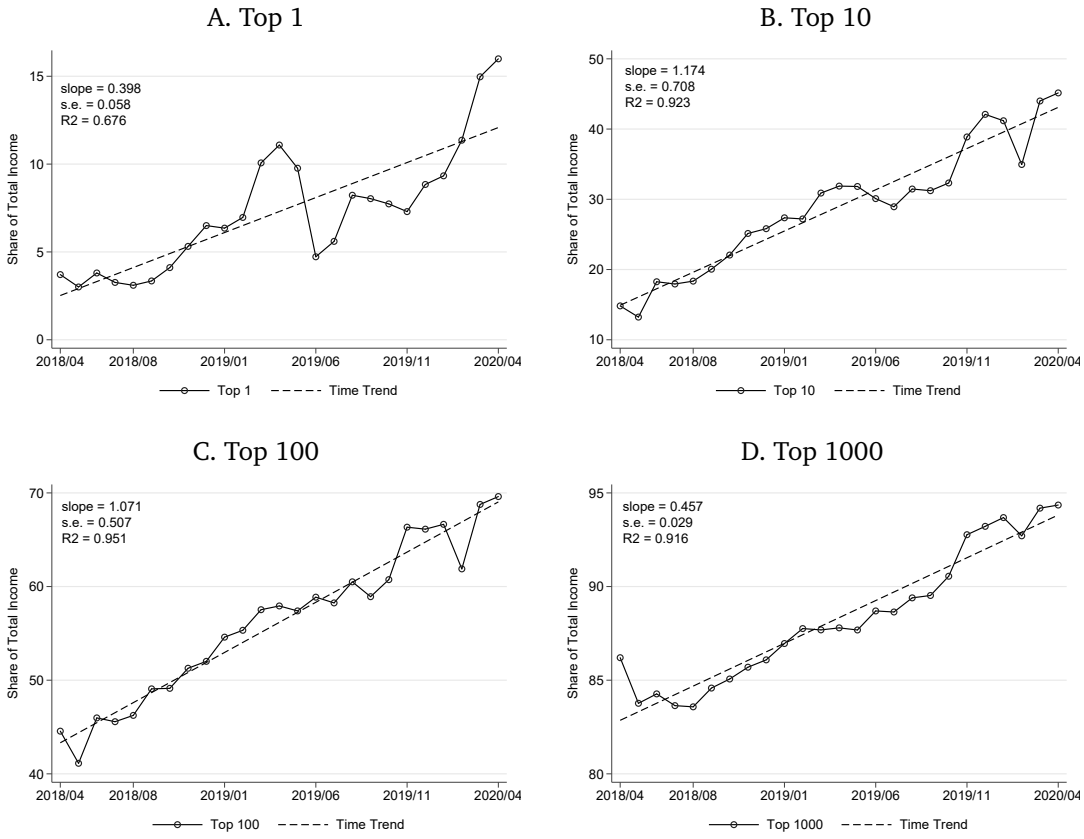


Figure 1. **Seasonality-adjusted income share of top streamers.** This figure ranks the streamers according to their aggregate income over the past three months and calculate the top groups' share of total tipping income. The upper left corner of each sub figure is marked with the slope, standard error,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

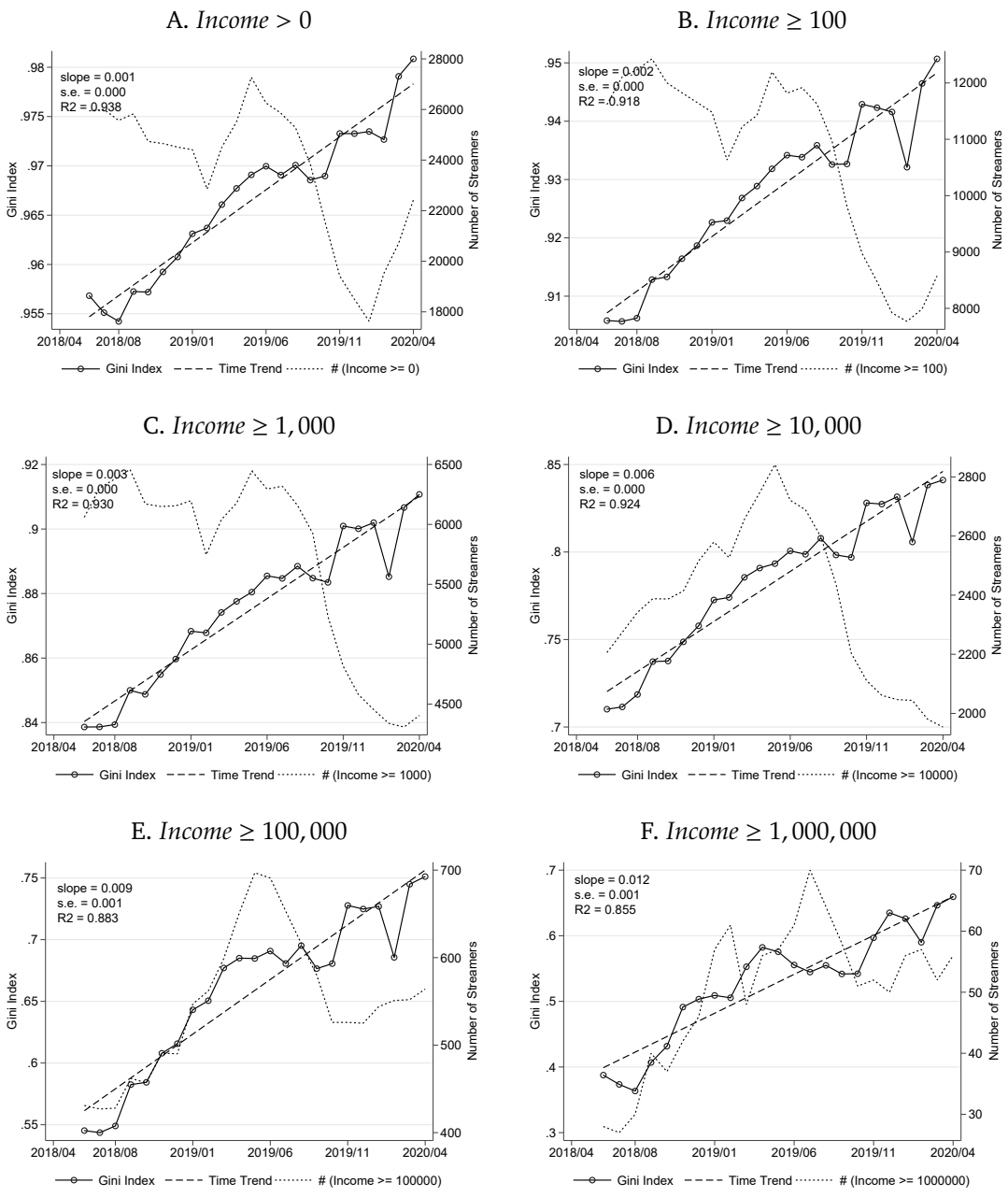


Figure 2. **Seasonality-adjusted Gini index.** This figure shows the seasonality-adjusted inequality trend as measured by the Gini index. We aggregate streamers' tipping income over the past three months and calculate the Gini index using the resulting income distribution. In Panel A, the dotted line 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.

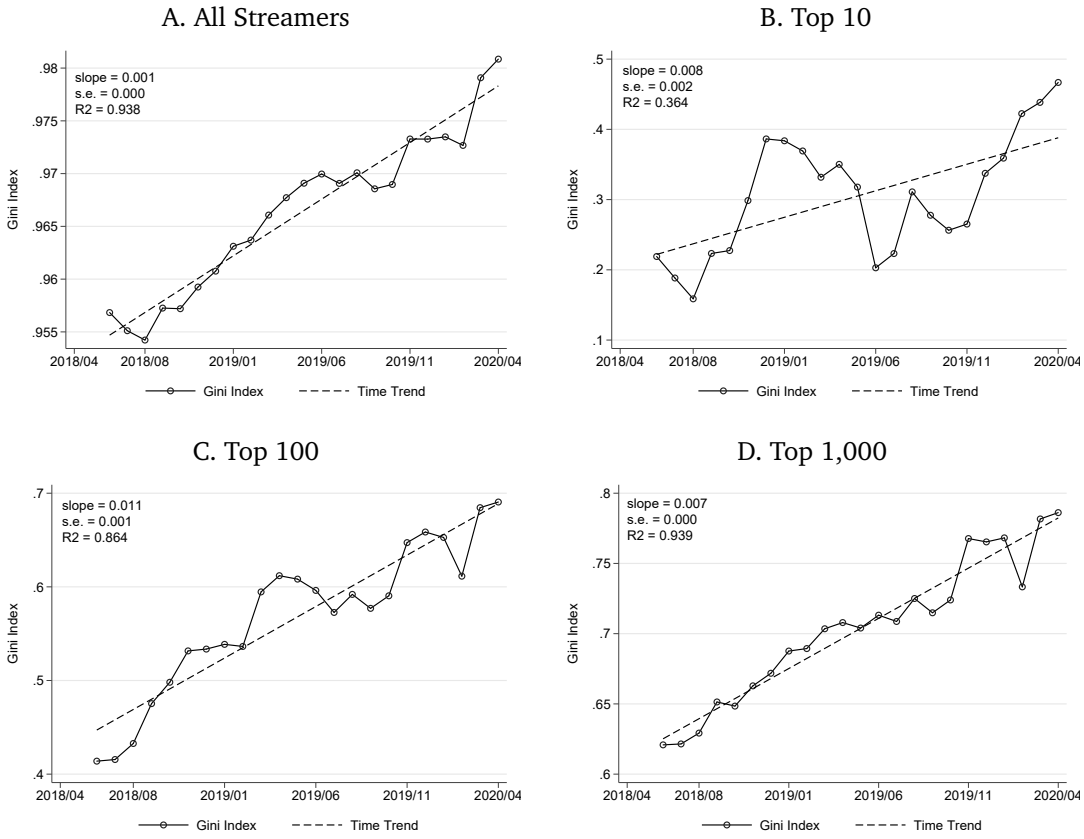


Figure 3. **Seasonality-adjusted Gini index among top streamers.** This figure shows the seasonality-adjusted inequality trend as measured by the Gini index within top streamers. We aggregate streamers' tipping income over the past three months, rank the streamers according to their aggregate income, 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.

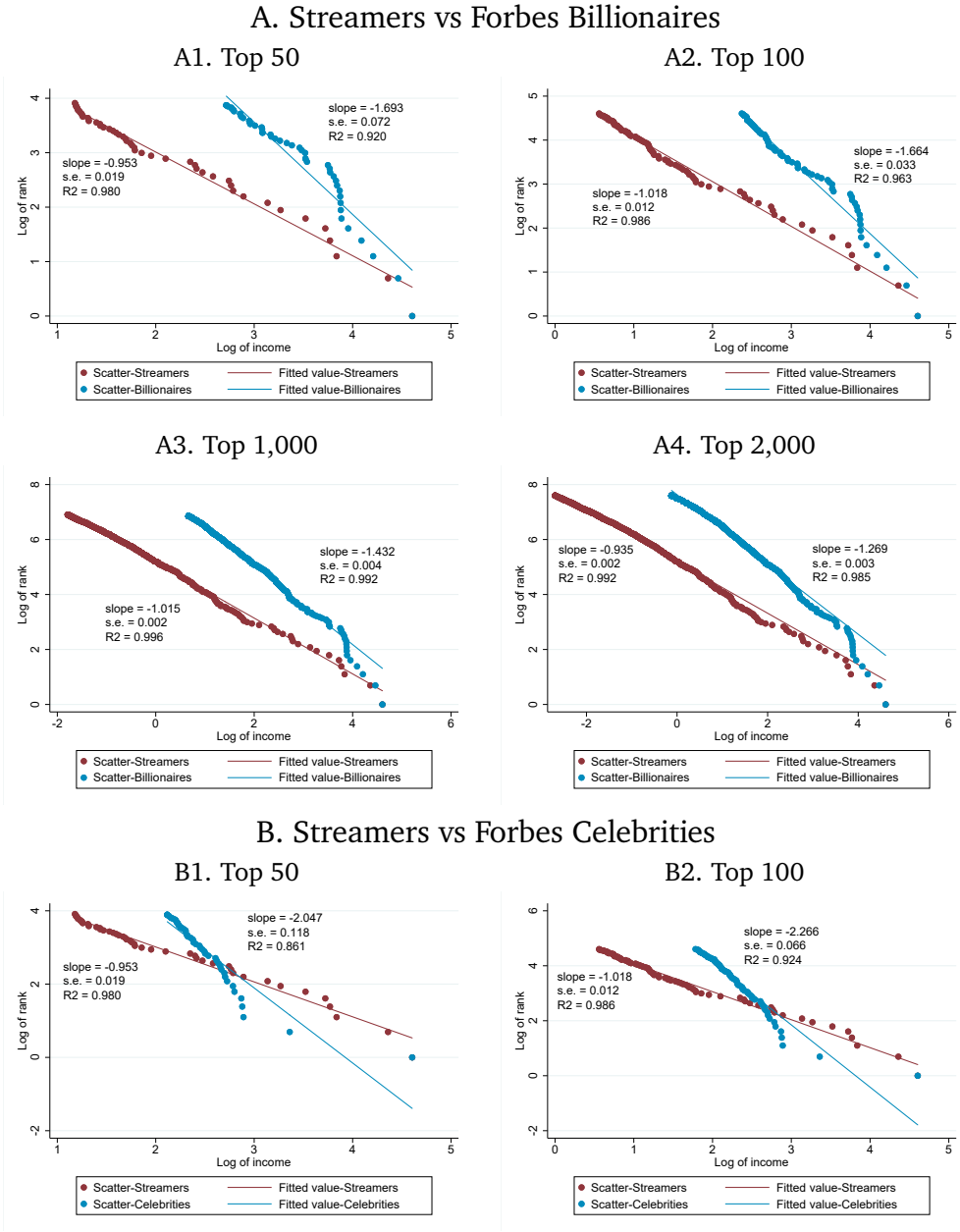


Figure 4. **PL exponent of top streamers vs PL exponent of Forbes.** Panel A compares PL exponent calculated using the streamers sample with that calculated using the 2020 Forbes' ranking of the world's richest billionaires. Panel B compares inequality among top streamers with inequality among 2020 Forbes' 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/`.

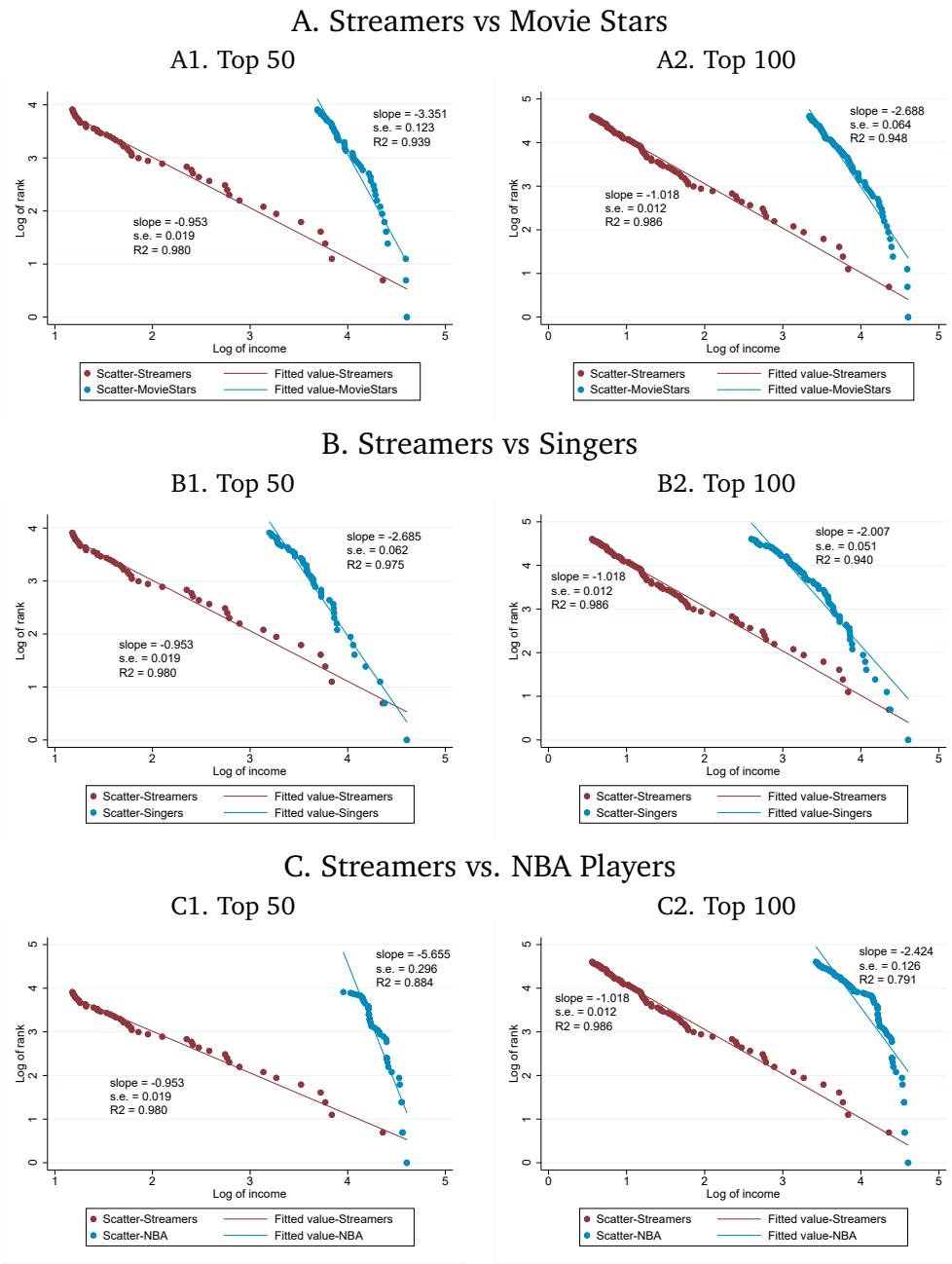


Figure 5. **PL exponent of top streamers vs. PL exponent of other superstars.** This figure compares the PL exponent calculated using the streamer's sample with that calculated using other superstars in the field of arts and sports. Streamers are ordered according to the total value of their virtual 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 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.[56]

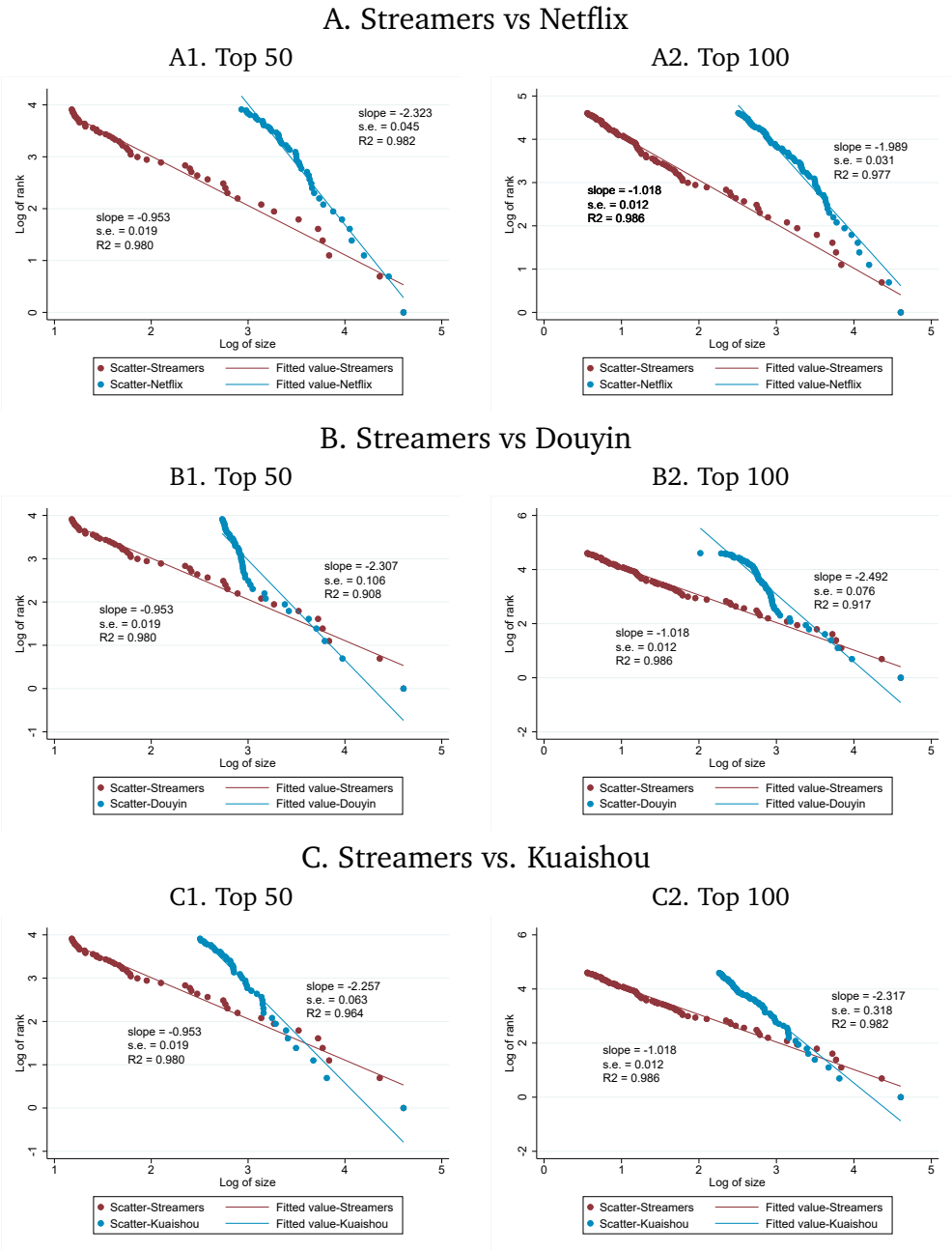


Figure 6. **PL exponent of top streamers vs. PL exponent of other digital platforms.** This figure compares the PL exponent calculated using the streamer's sample with that calculated using other superstars in the field of digital platforms. Streamers are ordered according to the total value of their virtual gifts during the sample period. In Panel A, films and television series are ordered based on their total hours viewed. In Panel B, short videos are ordered by their count of likes. In Panel C, content creators are ordered by their follower count. The log of rank is placed on the y-axis, and the log of streamers' total tipping income or size metrics of superstars in other digital platforms, is placed on the x-axis. The size variables are normalized by their max value so that the red distribution line 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.

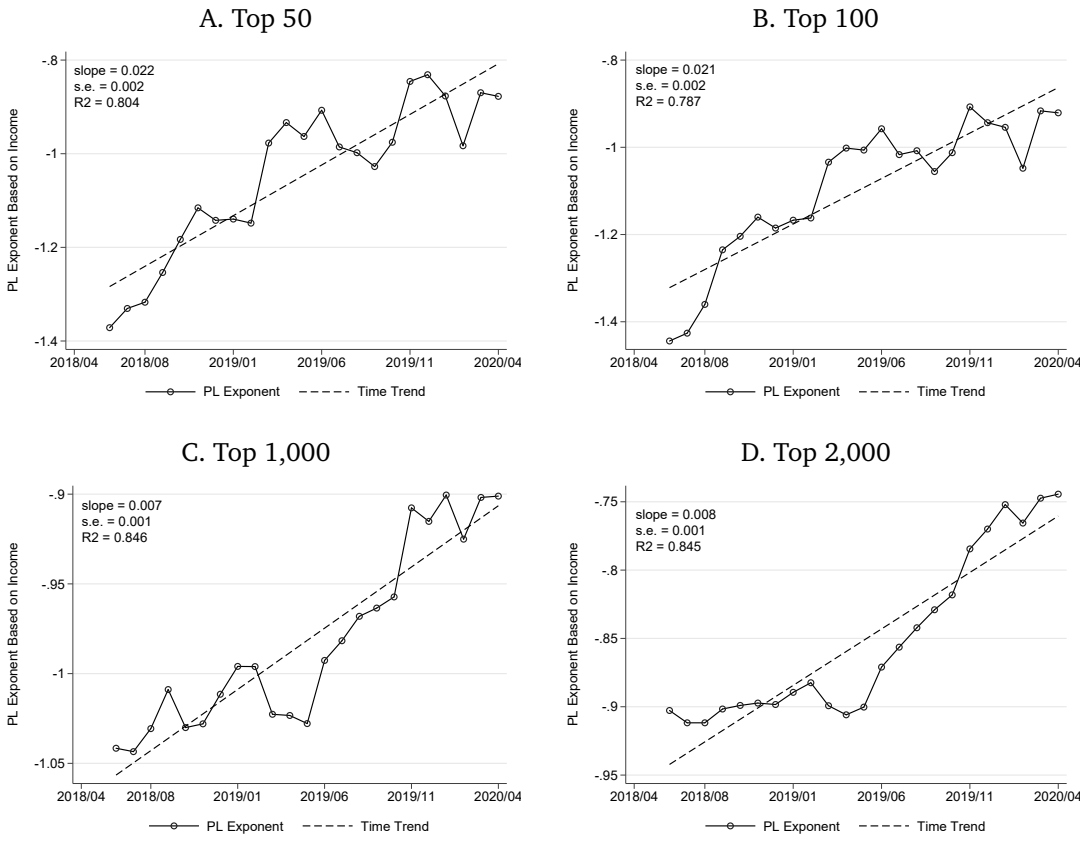
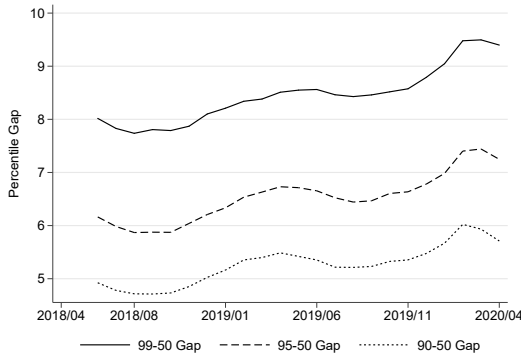
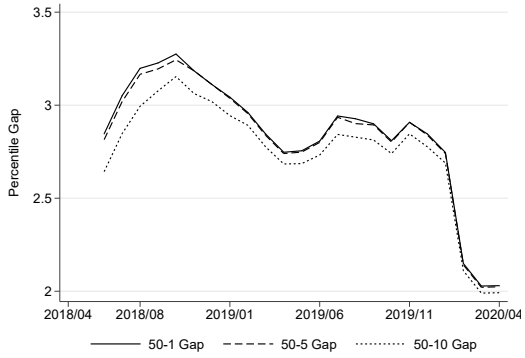


Figure 7. **Seasonality-adjusted PL exponent.** This figure shows the dynamic of inequality as measured by the seasonality-adjusted PL exponent. We aggregate streamers' tipping income over the past three months, rank them according to their aggregate income, and calculate the PL exponent using the resulting 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.

A. Percentile gaps: 99-50, 95-50, 90-50



B. Percentile gaps: 50-1, 50-5, 50-10



C. Percentile values: P95, P75, P50, P25, P5

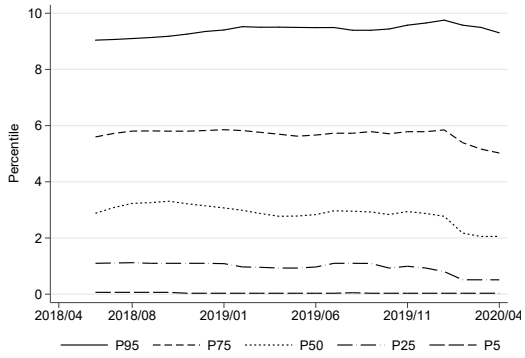


Figure 8. **Percentile gaps.** This figure shows the time trend of different percentile gaps of the log income variable. We aggregate streamers' tipping income over the past three months and calculate different percentile values using the distribution of the log income variable. The exact percentile 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 ($P5$).

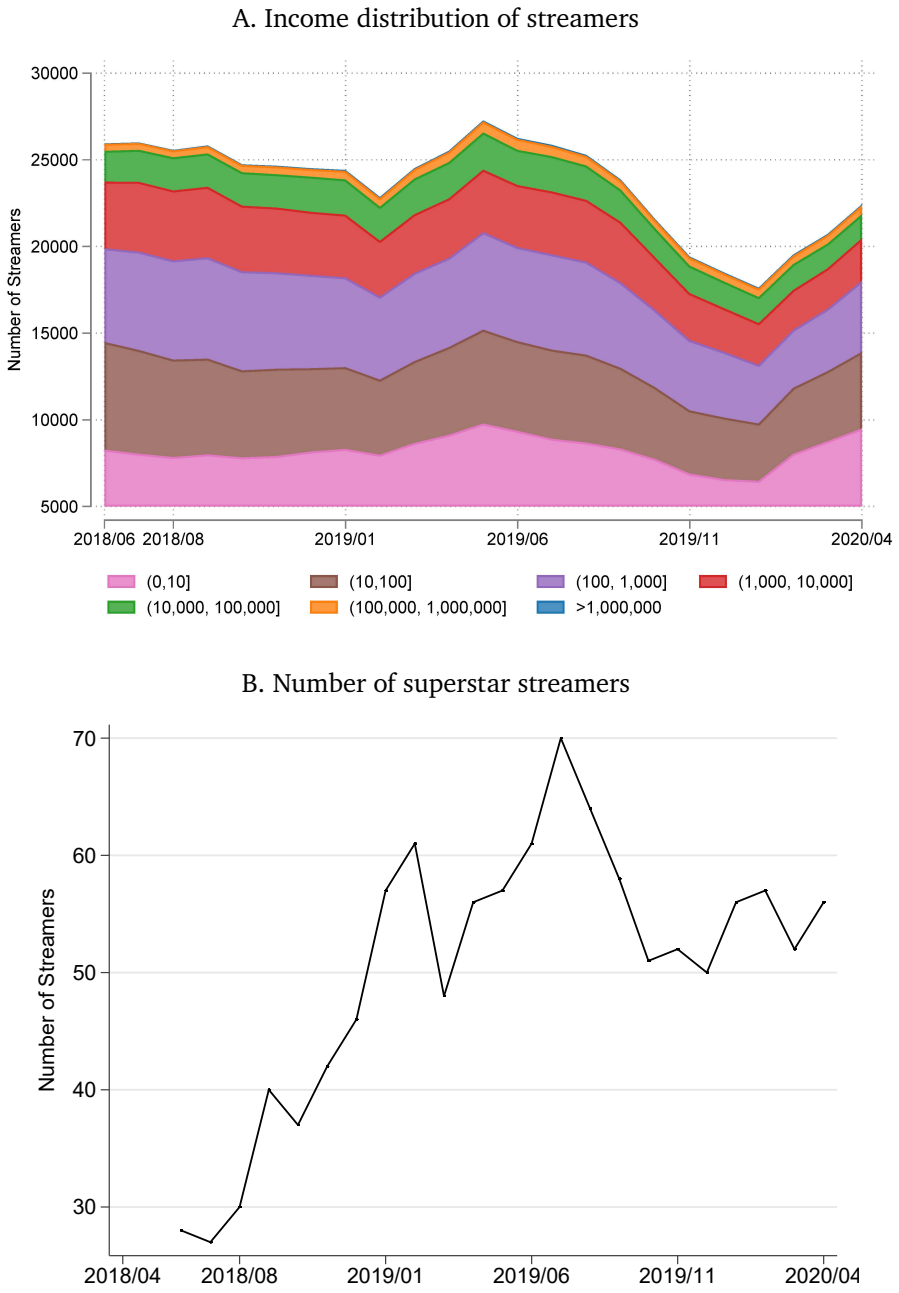


Figure 9. **Income distribution of streamers.** We aggregate streamers' tipping income over the past three months, and streamers are then divided into different groups according to their income. In Panel A, this stacked area chart not only tracks the total number of streamers who earned any positive 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.

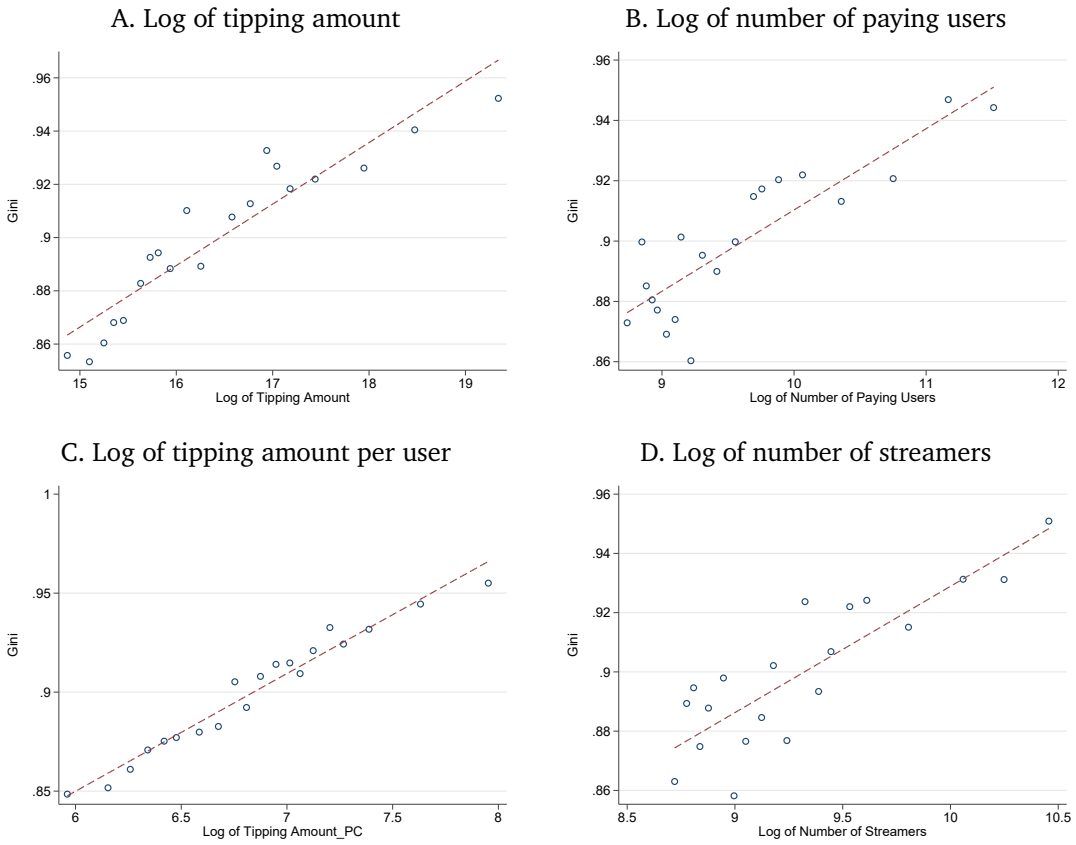


Figure 10. **Cross city relationships between market size and inequality.** This figure provides intuitive cross-city evidence that a broader market positively correlates with greater inequality. In Panels A to D, we use the total tipping amount, the total number of paying users, the tipping amount per 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.

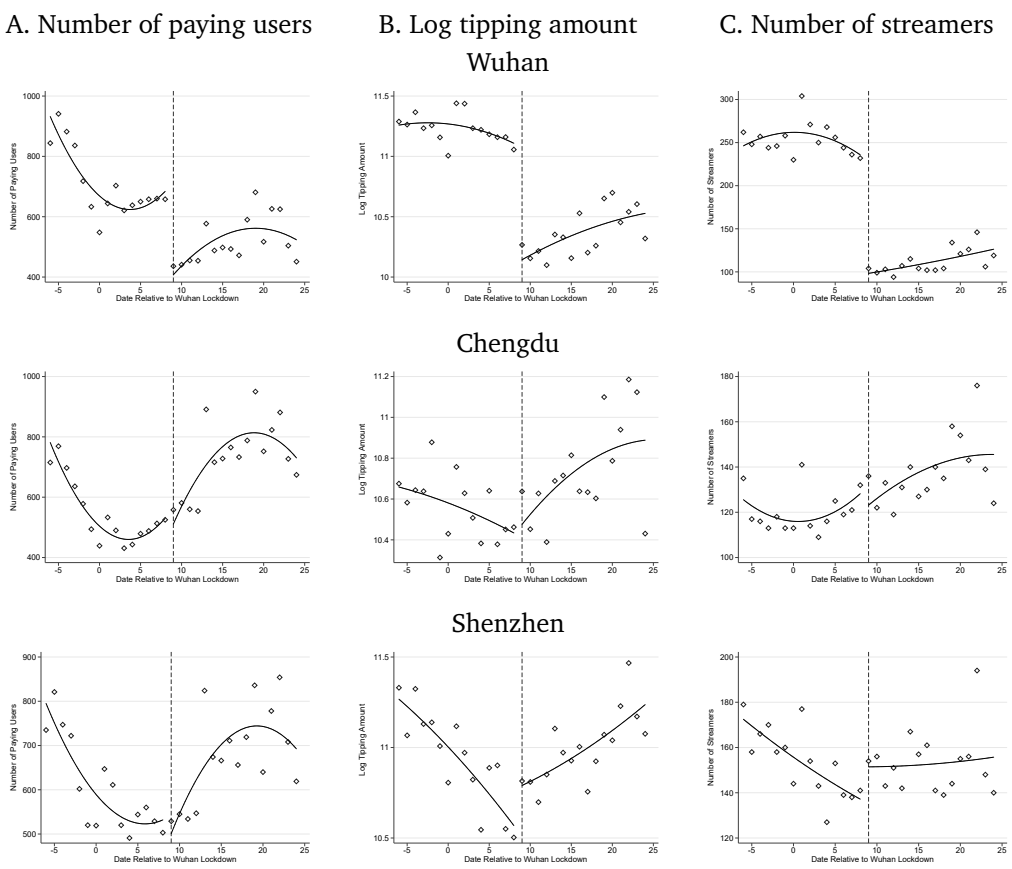


Figure 11. **Entertainment demand discontinuity in Wuhan.** The first row in this figure shows the discontinuity in entertainment demand in Wuhan around Day 9 (February 1, 2020). From that day on, the epidemic became serious (at least to an extent), the citizens of Wuhan became more concerned about it, 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.

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. [57]

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. **Summary statistics of main variables of Section 4.** This table shows the summary statistics of variables used in Section 4. Panel A summarizes the month-level variables used in Section **??**. $Gini50_t$ 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; $ln\,(Tipping\ Amount_t)$ is the log value of the total tipping income of all streamers in month t; $QuarterEnd_t$ is a dummy variable that equals one if the month is March, June, September, or December, and zero otherwise. *Fulltime* means that we further refine our sample by imposing an additional constraint: only those streamers who have logged a monthly streaming duration of at least 40 hours are included in our Gini coefficient computations. Panel B summarizes the city-month level variables used in Section 4.1.2 and Section 4.2. $Gini50_{j,t}$ 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; $ln\,(Tipping\ Amount_{j,t})$ is the log value of total tipping income of all streamers in city j in month t. $\Delta\, ln(Tipping\ Amount)$ 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. Panel C summarizes the city-level variables used in Section 4.3. *Population* is the city's residential population and *GDP* is local GDP. *Aggregate Index* is digital financial inclusion index of different cities, while *Coverage Breadth*, *Usage Depth*, and *Digitization Level* are index measuring different dimensions of digital finance development.

Variables	N	Mean	SD	Min	Median	Max
Panel A. Month level summary stats.						
Gini50	25	0.901	0.024	0.860	0.901	0.949
Gini500	25	0.851	0.036	0.785	0.856	0.920
Gini50_FullTime	25	0.828	0.040	0.755	0.830	0.898
Gini500_FullTime	25	0.810	0.044	0.728	0.814	0.888
ln(Tipping Amount)	25	18.692	0.347	18.119	18.760	19.383
Quarterend	25	0.320	0.476	0.000	0.000	1.000
Panel B. City-month level summary stats.						
Gini50	450	0.861	0.047	0.729	0.861	0.977
Gini500	450	0.762	0.080	0.537	0.766	0.960
Gini50_FullTime	450	0.858	0.047	0.722	0.859	0.977
Gini500_FullTime	450	0.766	0.079	0.548	0.768	0.961
ln(Tipping Amount)	450	14.908	0.864	13.134	14.849	17.234
Quarterend	450	0.320	0.467	0.000	0.000	1.000
Gini50	604	0.022	0.063	-0.173	0.024	0.246
Gini50_FullTime	604	0.020	0.066	-0.184	0.022	0.248
ln (Tipping Amount)	604	0.154	0.551	-2.083	0.159	1.916
ln(Tipping Amount_NewUsers)	604	0.029	1.430	-5.143	0.047	4.497
Panel C. City level summary stats.						
Gini50	92	0.900	0.032	0.823	0.903	0.966
Gini50_FullTime	92	0.899	0.030	0.826	0.900	0.958
ln(Tipping Amount)	92	16.438	1.155	14.770	16.151	19.658
ln(Population)	92	8.844	0.501	7.804	8.798	10.376
ln(GDP)	92	6.306	0.689	5.129	6.171	8.243
Aggregate Index	92	253.502	19.923	206.529	253.852	302.983
Coverage Breadth	92	241.669	24.945	181.948	246.198	290.318
Usage Depth	92	254.736	21.376	206.300	250.889	316.651
Digitization Level	92	290.346	12.562	260.608	288.739	328.998
Panel D. City-date level summary stats.						
Gini50	205	0.849	0.037	0.772	0.850	0.922
Gini50_FullTime	205	0.847	0.039	0.766	0.849	0.920
ln(Tipping Amount)	205	11.163	0.531	10.099	11.209	12.336

Table 3. **Time series estimation of the effect of market size on inequality using *Quarter End* as IV.** This table reports the time series level OLS and IV results of the effect of market size on inequality measure *Gini* using *Quarter End* to instrument the market size. We run the following regressions: $Gini_t = \alpha + \beta_{OLS} MarketSize_t + \eta_t + \epsilon_t$, and $Gini_t = \alpha + \beta_{IV} \hat{MarketSize_t} + \eta_t + \epsilon_t$, 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 instrument variable $Quarter\ End_t$ is a dummy variable that equals one if the month is March, June, September, or December and zeroes otherwise. Panel A presents the estimated coefficients of the Gini coefficients response to market size, while Panel B displays the outcomes following Z-score normalization. The first two columns in each panel report the OLS results while the last two columns in each panel report the 2SLS results. *Gini*50 is calculated using the income distribution of the streamers who receive more than 50 RMB in tipping income in month t. $Gini500_{jt}$ is calculated based on the income distribution of the streamers who receive more than 500 RMB in tipping income in month t. Robust standard errors are reported in parentheses. The Kleibergen and Paap F-statistics are reported in the last row. The sample includes the 25 months from April 2018 to April 2020. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A: Non-Standardization				Panel B: Standardization			
	OLS		IV: Quarterend		OLS		IV: Quarterend	
	Gini50	Gini500	Gini50	Gini500	Gini50	Gini500	Gini50	Gini500
ln (Tipping Amount)	0.072***	0.107***	0.077***	0.117***	1.051***	1.028***	1.123***	1.115***
	(0.005)	(0.008)	(0.006)	(0.010)	(0.070)	(0.075)	(0.087)	(0.096)
Year-Quarter FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	25	25	25	25	25	25	25	25
Adjusted R2	0.940	0.937	0.937	0.932	0.940	0.937	0.937	0.932
K. P. F-statistic			21.970	21.970			21.970	21.970

Table 4. **Panel estimation of the effect of market size on inequality using** *Quarter End* **as IV.** 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 $Gini_{jt} = \alpha + \beta_{IV} \ln(\hat{TippingAmount}_{jt}) + \theta_j + \eta_t + \epsilon_{jt}$, 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 instrument variable $Quarter\,End_t$ is a dummy variable that equals one when the month is March, June, September, and December and zeroes otherwise. In Columns (1), (3), (5), and (7), $Gini_{jt}$ is the Gini coefficient of the subsample of streamers whose incomes are more than 50 RMB from city j in month t. In Columns (2), (4), (6), and (8), $Gini_{jt}$ is the Gini coefficient of the subsample of streamers whose incomes are more than 500 RMB. Panel A presents the estimated coefficients of the Gini coefficients response to market size, while Panel B displays the outcomes following Z-score normalization. The first two columns in each panel report the OLS results β_{OLS}. The last two columns in each panel 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.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A: Non-Standardization				Panel B: Standardization			
	OLS		IV: Quarterend		OLS		IV: Quarterend	
	Gini50	Gini500	Gini50	Gini500	Gini50	Gini500	Gini50	Gini500
ln (Tipping Amount)	0.081***	0.135***	0.079***	0.127***	1.132***	1.142***	1.107***	1.075***
	(0.006)	(0.010)	(0.008)	(0.011)	(0.040)	(0.039)	(0.063)	(0.064)
City FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	450	450	450	450	450	450	450	450
Adjusted R2	0.892	0.889	0.849	0.836	0.892	0.889	0.849	0.836
K. P. F-statistic			101.869	101.869			101.869	101.869

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 4.2. To exclude the interference of the Covid-19 pandemic, we 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 $\Delta\ Gini_{j,stage} = \alpha+\beta\Delta\ \ln(\hat{Market\ Size}_{j,stage})+\epsilon_j$, 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. $Gini50$ 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 $\Delta\ \ln(Tipping\ Amount)$. 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*. Panel A presents the estimated coefficients of the Gini coefficients response to market size, while Panel B displays the outcomes following Z-score normalization. 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)
	$\Delta\ Gini50$					
	Stage 3- Stage2			Stage 3- Stage2		
	Panel A: Non-Standardization					
$\Delta\ \ln(Tipping\ Amount)$	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: Standardization					
$\Delta\ \ln(Tipping\ Amount)$	0.822***	0.496***	0.676***	0.570***	0.683***	0.834***
	(0.110)	(0.127)	(0.087)	(0.100)	(0.098)	(0.111)
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

Table 6. **Cross city estimation of the effect of market size on inequality using city features as IV.** This table shows the cross-city-level OLS and IV regression results. In the OLS regression, we run $Gini_j = \alpha + \beta_{OLS} MarketSize_j + \epsilon_j$. 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. Panel A presents the estimated coefficients of the Gini coefficients response to market size, while Panel B displays the outcomes following Z-score normalization. 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. 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). Robust standard errors are robust and reported in parentheses. K. P. F-statistics are reported in the last rows of each panel. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS	IV					
		ln(Population)	ln(GDP)	AggregateIndex	CoverageBreadth	UsageDepth	DigitizationLevel
Panel A: Non-Standardization							
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
Panel B: Standardization							
ln(Tipping Amount)	0.828***	0.738***	0.694***	0.690***	0.682***	0.699***	0.722***
	(0.054)	(0.069)	(0.058)	(0.067)	(0.071)	(0.103)	(0.107)
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		93.157	313.610	119.429	108.527	54.963	47.868

Table 7. **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. We use the log value of the tipping amount to proxy market size and use *Gini*50 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. Panel A presents the estimated coefficients of the Gini coefficients response to market size, while Panel B displays the outcomes following Z-score normalization. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

	20 Days before Day 9	20 Days after Day 9	After-Before
Panel A: Non-Standardization			
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: Standardization			
Gini50_Wuhan-Gini50_Chengdu	-0.660	-1.291	-0.632***
ln(Tipping Amount)_Wuhan-ln(Tipping Amount)_Chengdu	1.277	-0.647	-1.924***
Ratio			***0.328***
Gini50_Wuhan-Gini50_Shenzhen	-0.560	-1.278	-0.718**
ln(Tipping Amount)_Wuhan-ln(Tipping Amount)_Shenzhen	0.490	-1.196	-1.686***
Ratio			***0.426***
Gini50_Wuhan-Gini50_Beijing	-0.524	-1.452	-0.816**
ln(Tipping Amount)_Wuhan-ln(Tipping Amount)_Beijing	-0.355	-1.983	-1.397***
Ratio			***0.570***
Gini50_Wuhan-Gini50_Shanghai	-0.921	-1.553	-0.632***
ln(Tipping Amount)_Wuhan-ln(Tipping Amount)_Shanghai	-1.404	-2.723	-1.319***
Ratio			***0.480***

Table 8. **Robustness checks using *Fulltime* sample of streamers.** This table provides the robustness checks of identification part using *Fulltime* sample of streamers. We further refine our sample by imposing an additional constraint: only those streamers who have logged a monthly streaming duration of at least 40 hours are included in our Gini coefficient computations. We then replicate the results across four identification strategies using this revised Gini measure as the dependent variable. The results displays the outcomes following Z-score normalization. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)
Panel A: Market Size Seasonality & Time-series Estimation				
	OLS		IV: *Quarter End*	
	Gini50	Gini500	Gini50	Gini500
ln(*Tipping Amount*)	1.045***	1.048***	1.131***	1.123***
	(0.089)	(0.084)	(0.112)	(0.106)
Year-Quarter FEs	Yes	Yes	Yes	Yes
N	25	25	25	25
Adjusted R2	0.903	0.913	0.898	0.909
K. P. F-statistic			21.970	21.970
Panel B: Market Size Seasonality & Panel Estimation				
	OLS		IV: *Quarter End*	
	Gini50	Gini500	Gini50	Gini500
ln(*Tipping Amount*)	1.081***	1.062***	1.092***	1.064***
	(0.041)	(0.039)	(0.069)	(0.067)
City FEs	Yes	Yes	Yes	Yes
Year-Quarter FEs	Yes	Yes	Yes	Yes
N	450	450	450	450
Adjusted R2	0.859	0.867	0.809	0.808
K. P. F-statistic			101.869	101.869
Panel C: User Surge before a Capital Raising Event				
			Δ *Gini*50	
	Stage 3- Stage2		Stage 3- Stage2	
Δ ln(*Tipping Amount*)	0.761***	0.406***	0.510***	0.650***
	(0.129)	(0.141)	(0.120)	(0.109)
N	18	92	18	92
Adjusted R2	0.754	0.481	0.709	0.729
K. P. F-statistic	19.884	11.136	23.304	13.767
Panel D: City Features				
	OLS	ln(Population)	ln(GDP)	Aggregate Index
ln(*Tipping Amount*)	0.753***	0.642***	0.593***	0.582***
	(0.064)	(0.083)	(0.069)	(0.081)
N	92	92	92	92
Adjusted R2	0.563	0.550	0.536	0.533
K.P.F-statistic		93.167	314.423	119.429
Panel E. Event-based Gini Coefficient Response				
City		Δ(*Gini*)	Δ(*Market Size*)	Ratio
Chengdu		-0.458**	-1.924***	0.238
Shenzhen		-0.631**	-1.686***	0.374
Beijing		-0.726*	-1.628***	0.446
Shanghai		-0.208	-1.319***	0.158

Table 9. **The correlation of** *Gini_Income* **and** *Gini_Wage***.** For each month t, we construct the Gini coefficient $Gini_Income_t$ using the income distribution of streamers who receive virtual gifts worth more than 0, 50, and 500 RMB from users. Specifically, we have obtained data pertaining to the duration of the streamers' streaming time, which enables us to calculate the income they earn per unit of time. Subsequently, we leverage this calculated hourly wage along with the same sample to re-compute the Gini coefficient $Gini_Wage_t$. We run the regression: $Gini_Incomet = \alpha + \beta Gini_Wage_t + \epsilon_t$. Standard errors are robust and reported in parentheses. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

	Gini_Income		
	Income>0	Income>50	Income>500
Gini_Wage	0.372***	0.379***	0.421***
	(0.104)	(0.117)	(0.137)
N	25	25	25
Adjusted R-squared	0.327	0.281	0.260

Table 10. **Response coefficients of** *Gini_Income* **and** *Gini_Wage* **on market size.** Similar to Table 3, this table reports the time series level 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 $Gini_t = \alpha + \beta_{IV} \hat{MarketSize}_t + \eta_t + \epsilon_t$, where $Gini_t$ is the Gini coefficient calculated using the income distribution of streamers in month t; $MarketSize_t$ is log value of total tipping amount 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 results using *Gini_Income* as the dependent variables, while panel B reports the results using *Gini_Wage* as the dependent variables. In Columns (1), the dependent variable is calculated using the income distribution of the streamers who earn any positive income in month t. In Columns (2), $Gini_t$ is calculated using the income distribution of the streamers who receive more than 50 RMB in tipping income in month t. In Columns (3), $Gini_t$ 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)	(2)	(3)
	Panel A: Gini_Income		
log (Tipping Amount)	0.041***	0.077***	0.117***
	(0.003)	(0.006)	(0.010)
N	25	25	25
Adjusted R-squared	0.924	0.937	0.932
	Panel B: Gini_Wage		
log (Tipping Amount)	0.027**	0.041**	0.051*
	(0.013)	(0.020)	(0.027)
N	25	25	25
Adjusted R-squared	0.545	0.561	0.548
K.P. F-statistic		21.091	

Table 11. **Summary statistics of main variables of Section 5.3.** This table shows the summary statistics of variables used in Section 5.3. Panel A summarizes the streamer-month level variables used in Table 12. $ln(Streaming\ Time)_{jt}$ is the streaming time of streamer j in month t; $ln(Tipping\ Income)_{jt}$ is log value of total tipping income of streamer j in month t. $Quarter\ End_t$ is a dummy variable that equals one if the month is March, June, September, or December and zeroes otherwise. Panel B summarizes the streamer-date level variables used in Table . $ln(Tipping\ Income_{j,t})$, represents the income of streamer j over the preceding 7 days, while $ln(Streaming\ Time_{j,t})$ measures the daily broadcasting duration of the broadcaster. Our sample is restricted to *full-time* streamers based on their broadcasting duration three months prior to the lockdown in Wuhan (with an average monthly streaming time exceeding 40 hours). $Wuhan\ Dependence_j = \frac{Tipping\ Amount\ from\ Wuhan\ Users}{Total\ Tipping\ Amount}$, where the income calculations are based on data from the three months preceding the lockdown. And then *Wuhan Dependence* is operationalized as a categorical variable, taking values from 1 to 5. $Postt = 1$ indicates days after Wuhan Lockdown.

Variables	N	Mean	SD	Min	Median	Max
Panel A. Streamer-month level summary stats.						
ln(Streaming Time)	66,696	4.573	0.538	3.714	4.527	6.165
ln(Tipping Income)	66,696	8.459	2.540	0.002	8.855	17.780
Quarterend	66,696	0.323	0.468	0.000	0.000	1.000
Panel B. Streamer-date level summary stats.						
ln(Streaming Time)	10,776	6.474	1.362	1.616	6.845	8.669
ln(Tipping Income)	10,776	7.205	2.883	0.095	7.747	12.241
Post	10,776	0.425	0.494	0.000	0.000	1.000
Wuhan Dependence	10,776	2.948	1.387	1.000	3.000	5.000

Table 12. **The change of labor supply to market size using *Quarter End* as IV.** This table reports the time series level OLS and IV results of the effect of market size on streamers' labor supply time using *Quarter End* to instrument the market size. We run the following regressions: $ln(Streaming\ Time)_{jt} = \alpha + \beta_{OLS} MarketSize_{jt} + \theta_j + \eta_t + \epsilon_{jt}$, and $ln(Streaming\ Time)_{jt} = \alpha + \beta_{IV} \hat{MarketSize}_{jt} + \theta_j + \eta_t + \epsilon_{jt}$, where $ln(Streaming\ Time)_{jt}$ is the streaming time of streamer j in month t; $MarketSize_{jt}$ is log value of total tipping income of streamer j in month t, θ_j represents streamer-level fixed effects and η_t represents year-quarter fixed effects. The instrument variable $Quarter\ End_t$ is a dummy variable that equals one if the month is March, June, September, or December and zeroes otherwise. Panel A reports the elasticity of Gini coefficients response to market size, while Panel B reports results after z-score normalization. The first two columns in each panel report the OLS results, while the last two columns report the IV results. We examine samples of streamers whose monthly tipping income exceeds 50 and 500 RMB, respectively. Standard errors are clustered at the streamer level and reported in parentheses. The Kleibergen and Paap F-statistics are reported in the last row. The sample includes the 25 months from April 2018 to April 2020. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A: Non-Standardization				Panel B: Standardization			
	OLS		IV: QuarterEnd		OLS		IV: QuarterEnd	
	Income>0	Income>50	Income>0	Income>50	Income>0	Income>50	Income>0	Income>50
ln (Tipping Income)	0.199***	0.246***	0.130**	0.103**	0.941***	1.161***	0.614**	0.486**
	(0.004)	(0.003)	(0.063)	(0.051)	(0.017)	(0.015)	(0.298)	(0.242)
Streamer FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	54192	51020	54192	51020	54192	51020	54192	51020
Adjusted R2	0.600	0.623	0.214	0.198	0.600	0.623	0.214	0.198
K.P. F-statistic			29.523	55.728			29.523	55.728

Table 13. **The change of labor supply to market size using Covid-19 shock.** For streamer j and date t, we run the following 2SLS regression: $\log(Tipping\ Income_{j,t}) = \alpha + \beta Post_t \times Wuhan\ Dependence_j + \epsilon_{j,t}$, and $log(Streaming\ Time_{j,t}) = \alpha + \beta \log(\widehat{Tipping\ Income}_{j,t}) + \epsilon_{j,t}$, where the independent variable $Tipping\ Income_{j,t}$, represents the income of streamer j over the preceding 7 days, while the dependent variable, $Streaming\ Time_{j,t}$ measures the daily broadcasting duration of the broadcaster. Our sample is restricted to *full-time* streamers based on their broadcasting duration three months prior to the lockdown in Wuhan (with an average monthly streaming time exceeding 40 hours). The construction formula for the instrumental variable $Wuhan\ Dependence_j$ is: $Wuhan\ Dependence_j = \frac{Tipping\ Amount\ from\ Wuhan\ Users}{Total\ Tipping\ Amount}$, where the income calculations are based on data from the three months preceding the lockdown. And in our regression analysis, the variable *Wuhan Dependence* is operationalized as a categorical variable, taking values from 1 to 5. This categorization is derived from the quintile ranking of the continuous measure. Our analysis incorporates all streamers that are resident in Wuhan, spanning the period from 2020/1/4 to 2020/2/29 (4 weeks around Day 9). Robust standard errors are reported in parentheses. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

	Panel A: Non-Standardization		Panel B: Standardization	
	OLS	IV	OLS	IV
ln (Tipping Income)	0.318***	0.547***	0.674***	1.158***
	(0.003)	(0.063)	(0.006)	(0.133)
N	10776	10776	10776	10776
Adjusted R2	0.455	0.220	0.455	0.220
K.P. F-statistic		144.631		144.631

For Online Publication

June 1, 2026

A Appendix

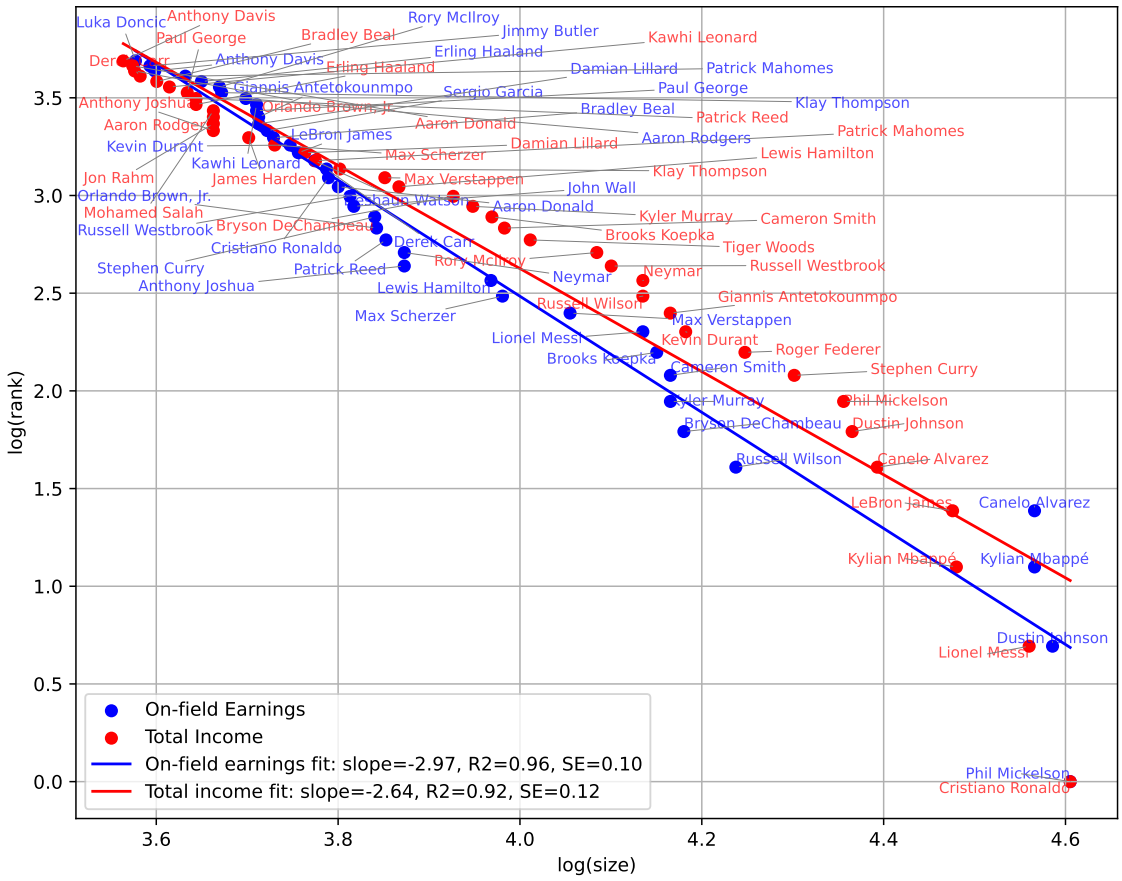


Figure A1. **PL distribution of on-field income and total income for top athletes.** This figure compares the PL exponent calculated using the top athletes' on-field income with that calculated using their total income. The income data for athletes is sourced from the Forbes World's Highest-paid Athletes 2023. The on-field earnings encompass prize money, salaries, and bonuses, while off-field earnings are estimated to include sponsorship deals, appearance fees, memorabilia and licensing income, along with cash returns from businesses in which the athlete has a significant stake. The total income is the sum of on-field and off-field earnings. The log of rank is placed on the y-axis, and the log of athletes' income, which we call their size, is placed on the x-axis. Each data point represents an athlete's income and their ranking. The blue line depicts the linear fit for their on-field earnings, while the red line illustrates the linear fit for their total income. The size variables are normalized by their max value so that the red line and the blue line can get started from the same position on the x-axis in the graph. The slope, R^2, and standard error of fit lines are depicted in the legend located at the bottom-left corner.

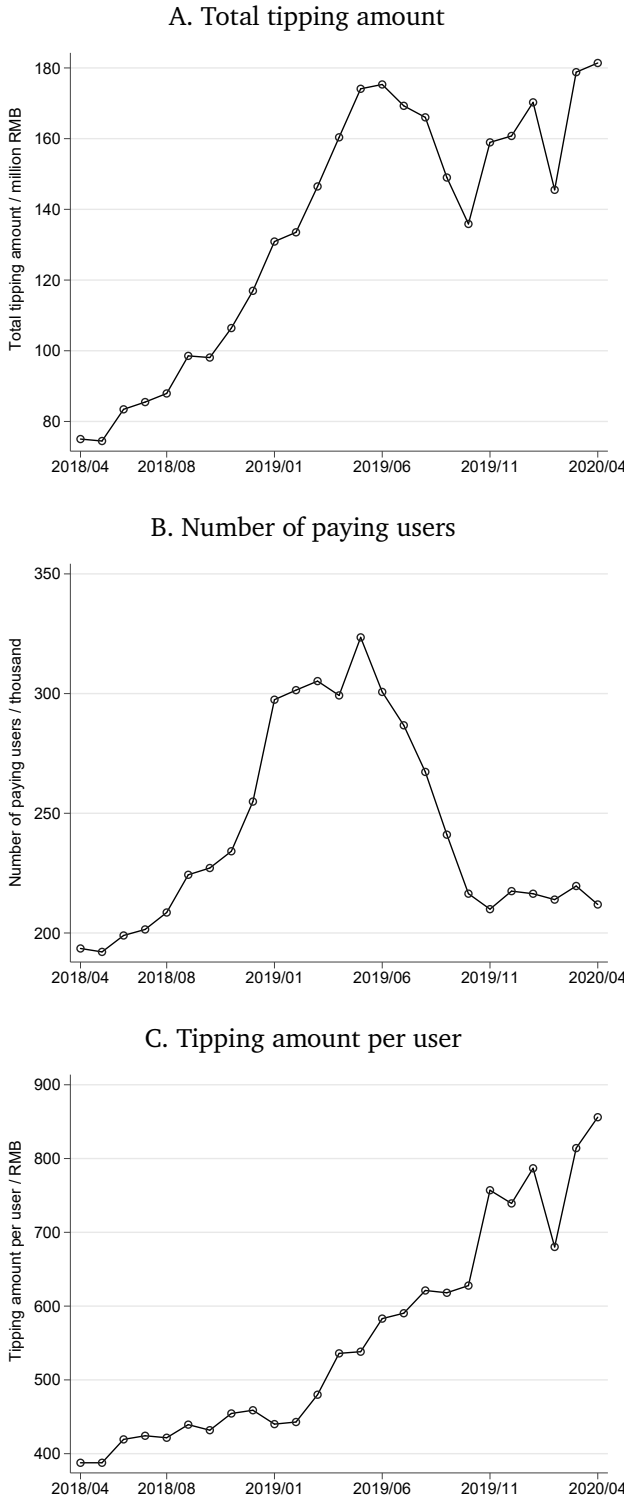


Figure A2. **Seasonality-adjusted tipping amount.** This figure shows the dynamics of seasonality-adjusted market size during the sample period. We calculate the average of the past three months to adjust for seasonality. Panel A plots the total amount of the virtual gifts received by the streamers in 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.

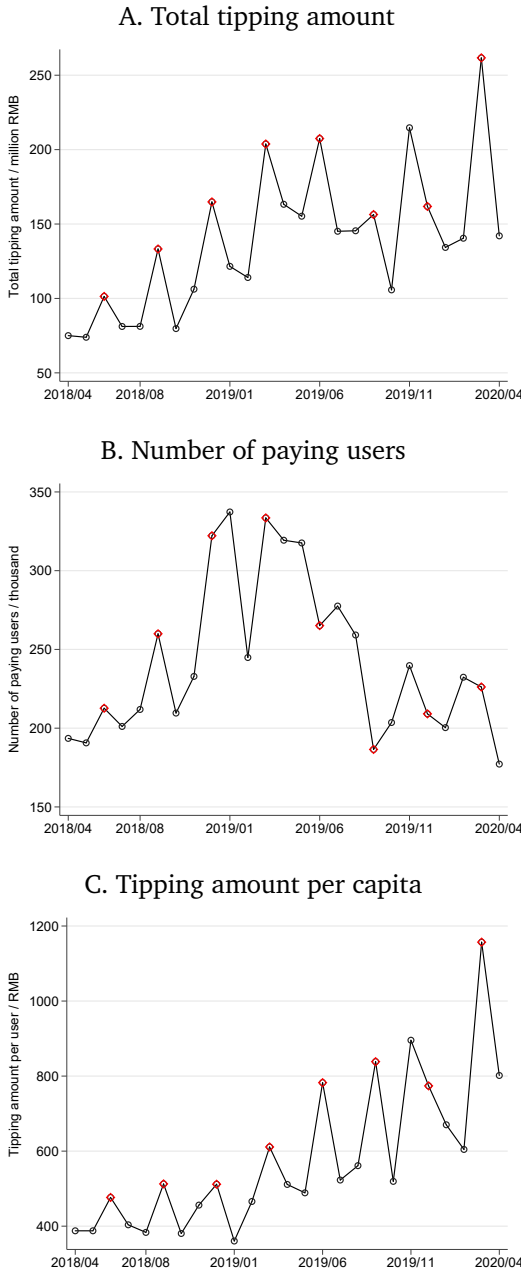


Figure A3. **Tipping amount.** This figure shows the dynamics of raw market size during the sample period. Panel A plots the total value of the virtual gifts received by the streamers in the sample (in million RMB). We further decompose this tipping amount into the number of paying users (in thousand) 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).

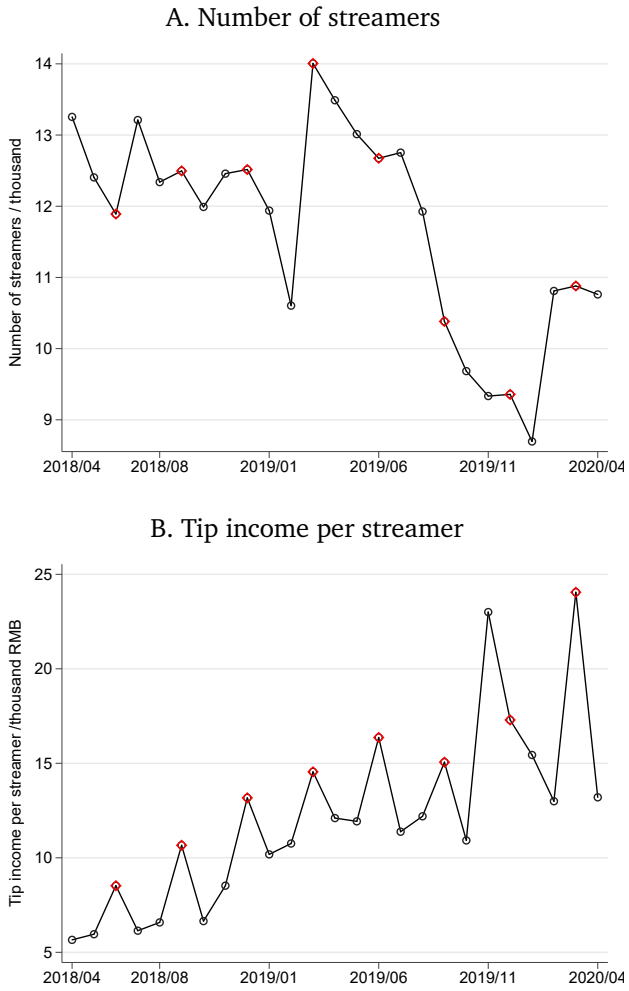


Figure A4. **Decompose raw tipping amount from the streamer side.** This figure decomposes the raw tipping amount from the streamer side. Panel A plots the number of streamers (in thousand) who can earn positive tipping income each month. Panel B shows the average tipping income (in thousand RMB) received by these streamers. The red dots mark the months at the end of each quarter (March, June, September, and December).

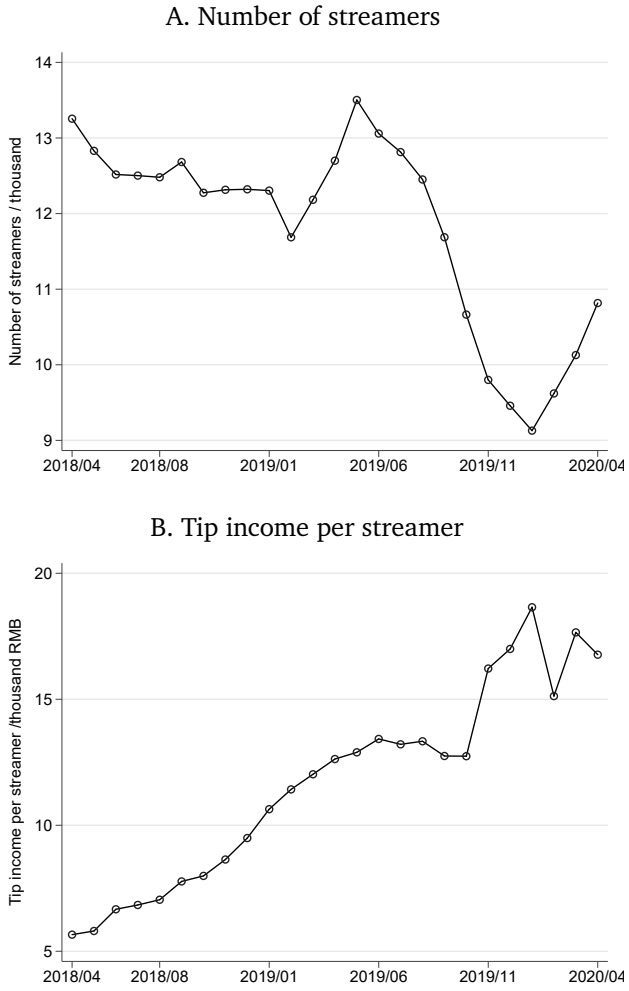


Figure A5. **Decompose seasonality-adjusted tipping amount from the streamer side.** This figure decomposes the seasonality-adjusted tipping amount from the streamer side. Panel A plots the number of streamers (in thousand) who can earn positive tipping income over the past three months. Panel B plots the average tipping income (in thousand RMB) received by these streamers.

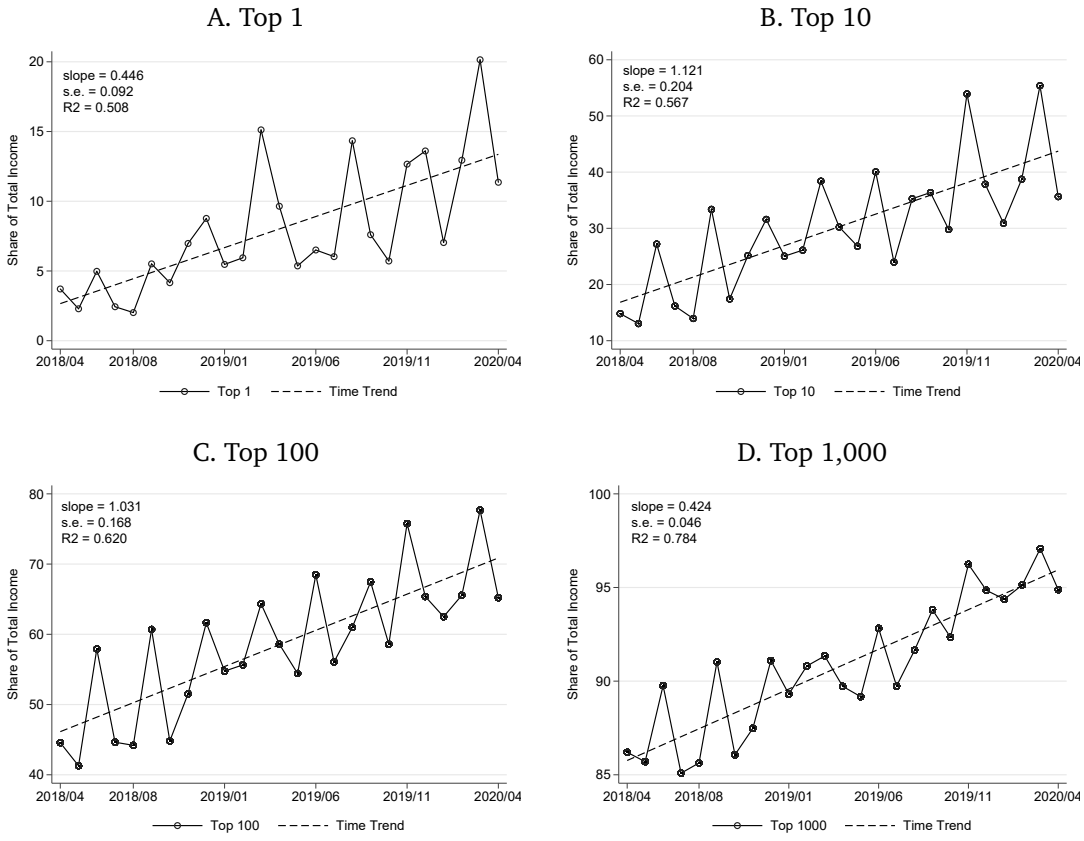


Figure A6. **Income share of top streamers.** This figure ranks the streamers according to their monthly income and calculates the top groups' share of total tipping income. The upper left corner of each sub-figure is marked with the slope, standard error, and R^2 of the linear fit line.

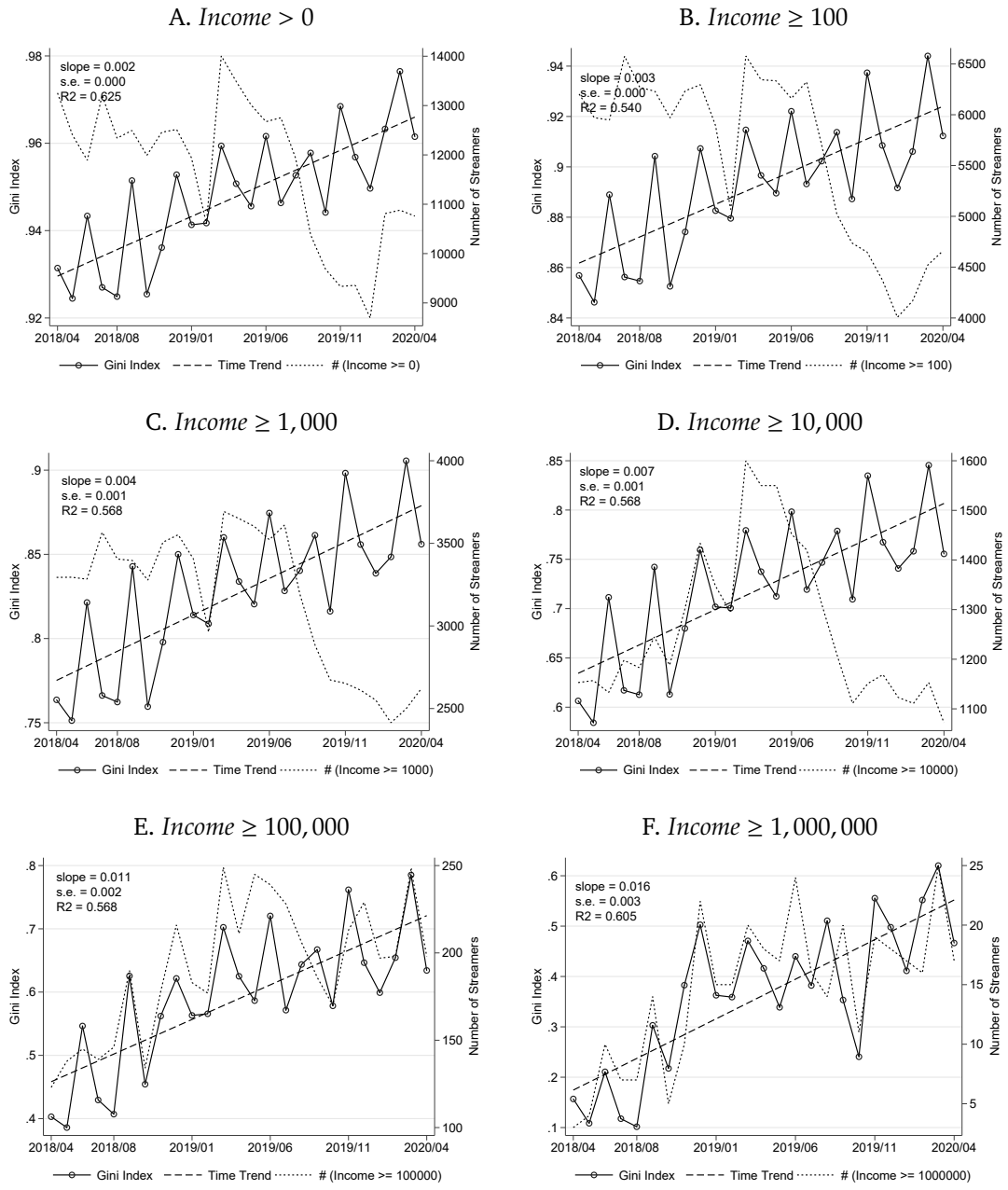


Figure A7. **Gini index.** This figure shows the inequality trend as measured by the Gini index, which is calculated using the distribution of streamers' monthly tipping income. In Panel A, the dotted line shows the number of streamers with positive monthly tipping income, the solid line shows the Gini 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.

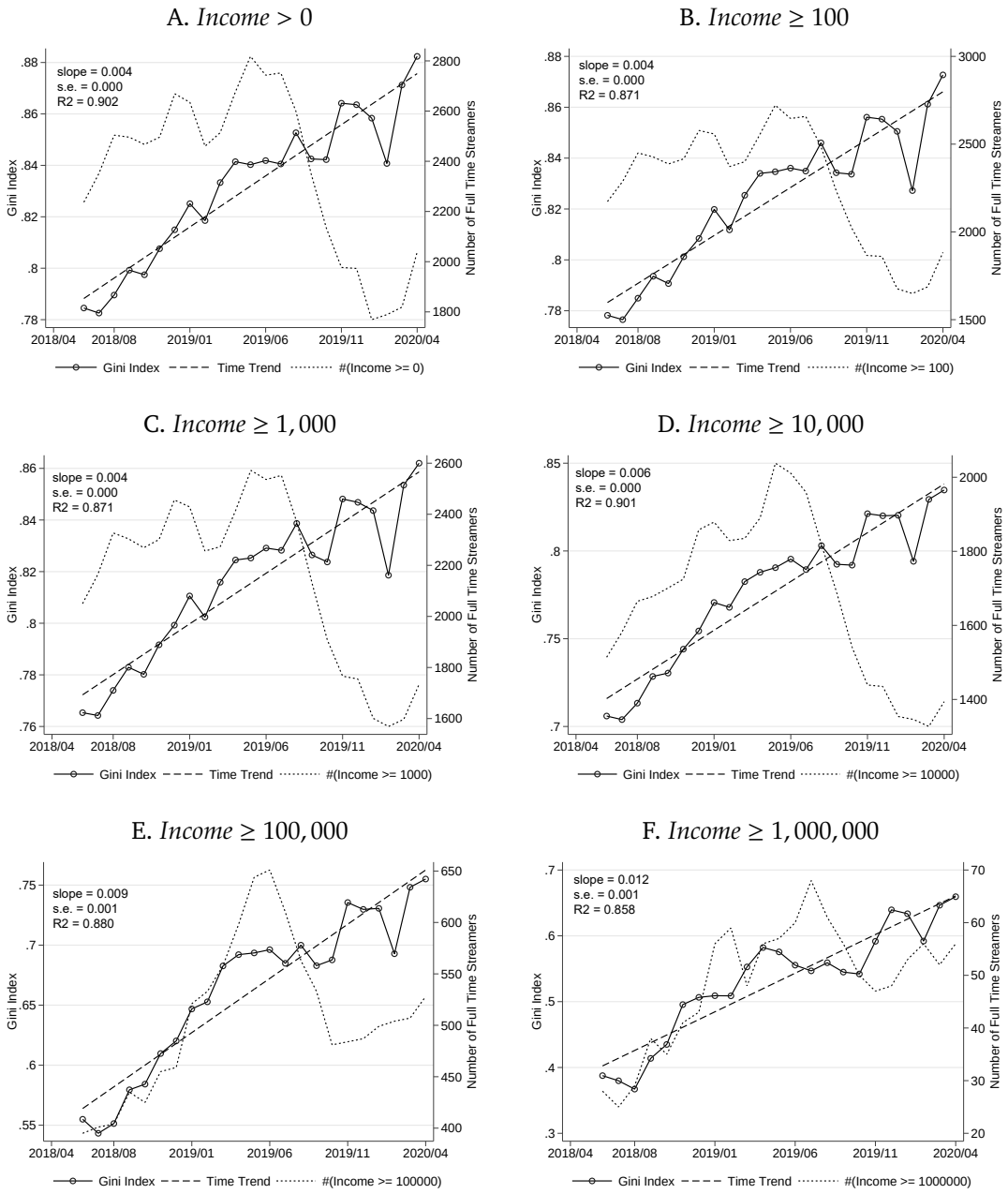


Figure A8. **Seasonality-adjusted Gini index within full-time streamers.** This figure shows the seasonality-adjusted inequality trend as measured by the Gini index among full time streamers. We refine the sample to include only those streamers with a monthly broadcasting duration exceeding 40 hours. We further 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 full time 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 full time 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.

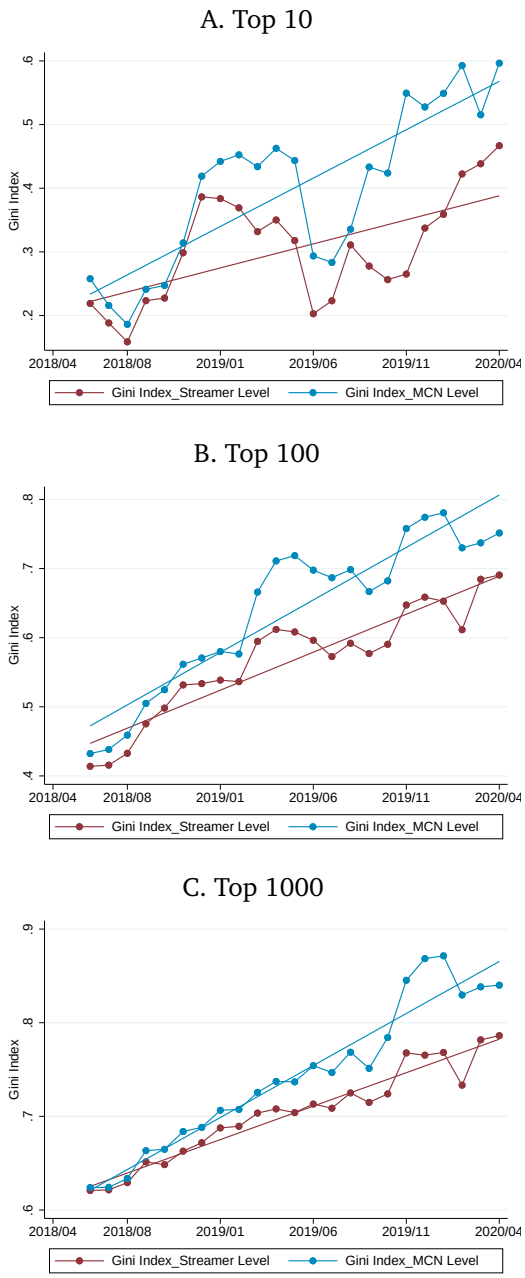


Figure A9. **Gini at MCN level vs. Gini at streamer level** This figure illustrates the dynamic changes in the Gini coefficient at both the MCN level (blue line) and the streamer level (red line). When calculating the MCN-level Gini coefficient, we initially aggregate the incomes of streamers within the same MCN to obtain the monthly MCN-level tipping income; streamers not part of any MCN are treated as individual MCNs. We then rank by MCN-level income to calculate the Gini coefficient between top entities. Additionally, a smoothing process has been applied using the income from the past three months.

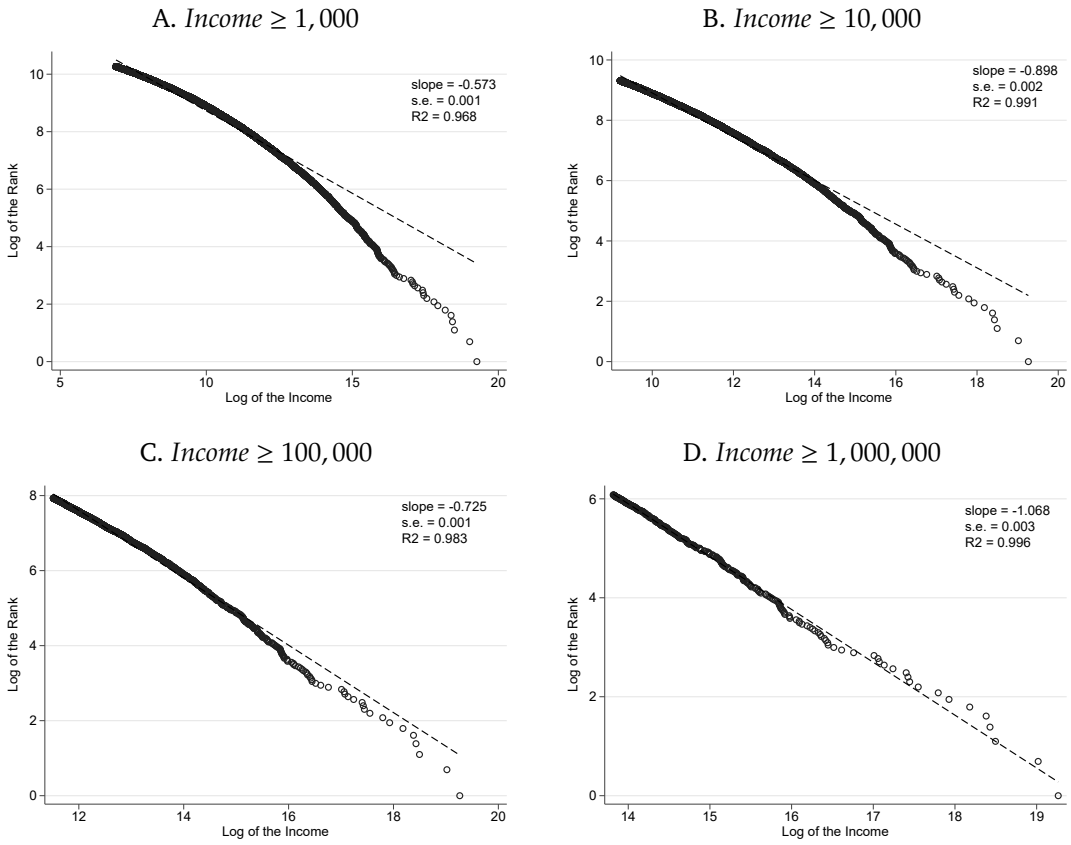


Figure A10. **Log tipping income versus log rank of the streamers.** Streamers are ordered according to the total value of their received virtual gifts during the whole sample period. The log value of rank is placed on the y-axis, and the log of streamers' total tipping income is placed on the x-axis. 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.

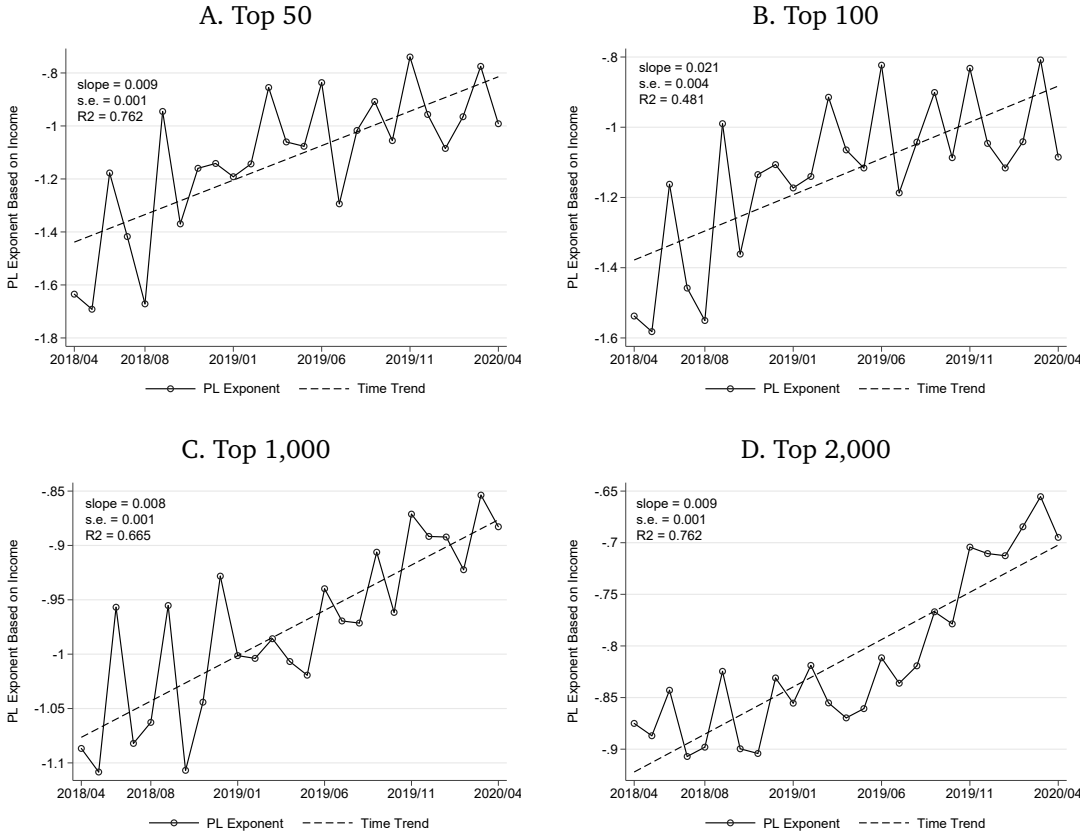


Figure A11. **PL exponent.** This figure shows the dynamic of inequality as measured by the PL exponent. We rank the streamers according to their monthly income and calculate the PL exponent using the resulting income distribution of those top streamers. Specifically, we run the linear regressions $\ln rank_{j,t} = \alpha_t + \beta_t \ln size_{j,t} + \varepsilon_{j,t}, \forall t$, 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.

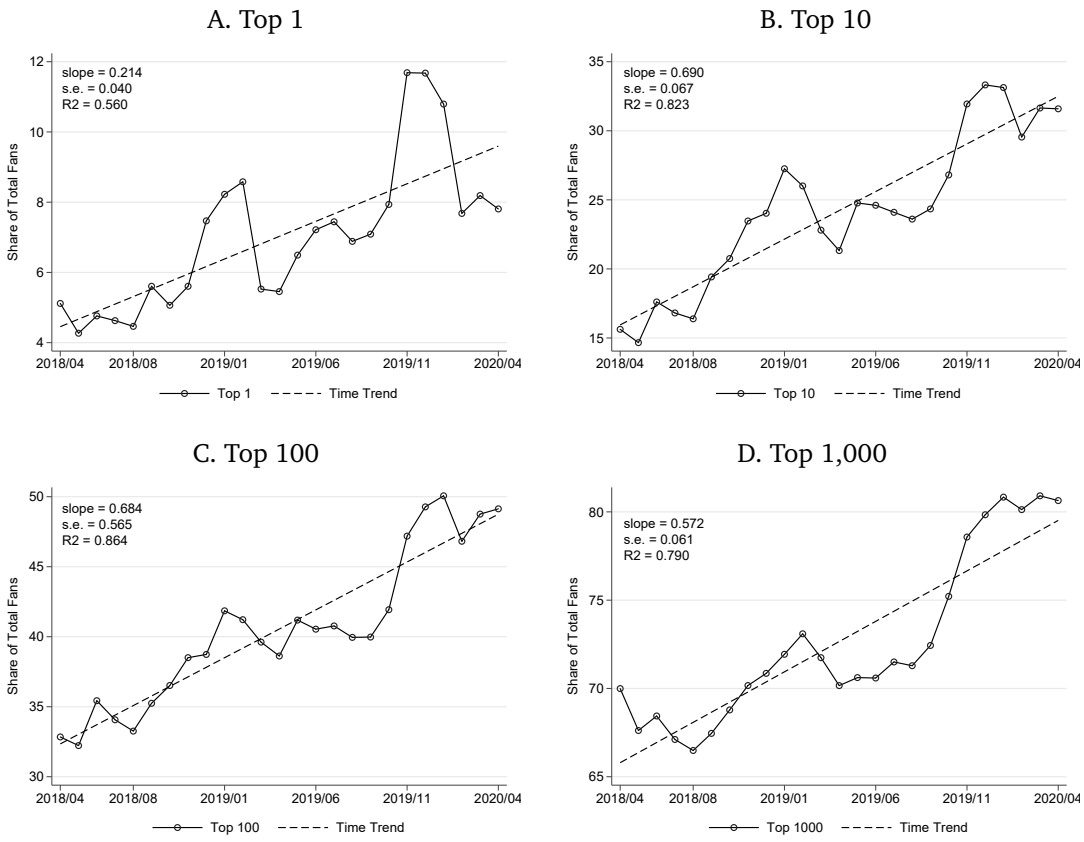


Figure A12. **Seasonality-adjusted fans share of top streamers.** In this figure, we rank the streamers by their aggregate number of loyal fans in the past three months, adjusted for seasonality, and we calculate the ratio of the number of loyal fans of the top streamer group to the total number of fans. 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.

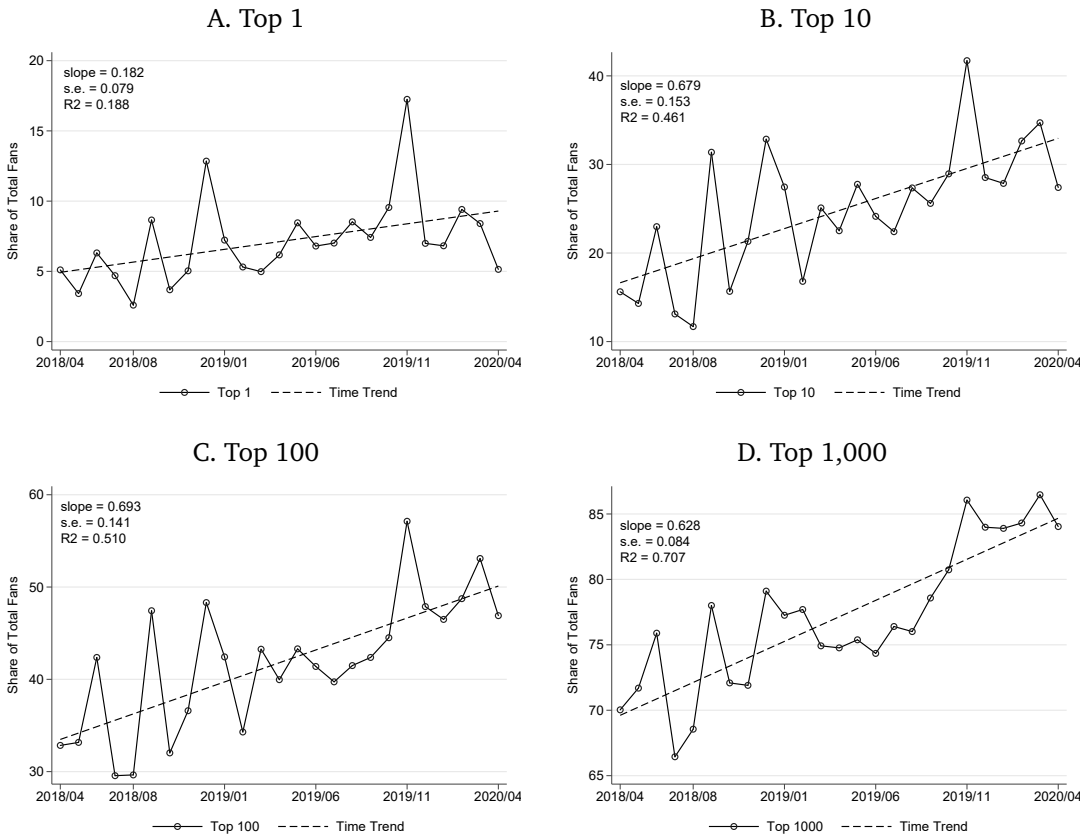


Figure A13. **Fans share of top streamers.** In this figure, we rank the streamers by their monthly number of loyal fans, and we calculate the ratio of the number of loyal fans of the top streamer group to the total number of fans. Suppose there are N streamers. The method to determine whether $user_i$ 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.

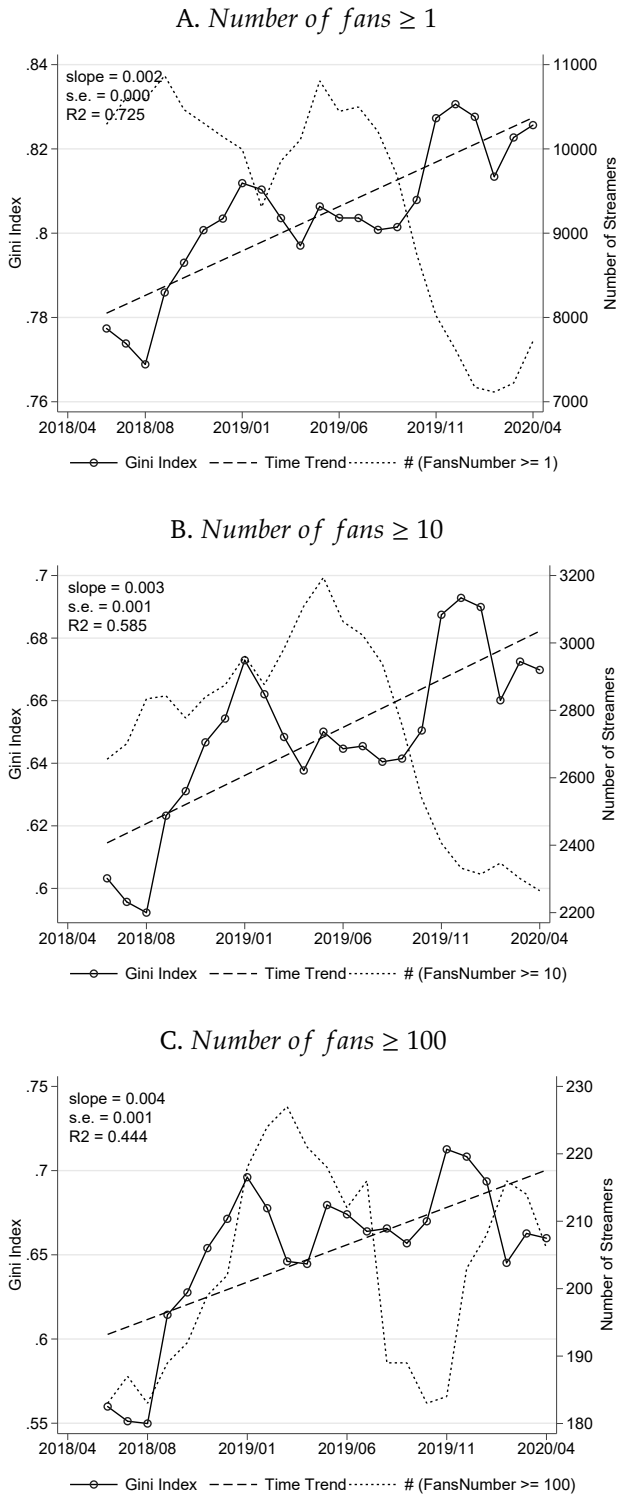
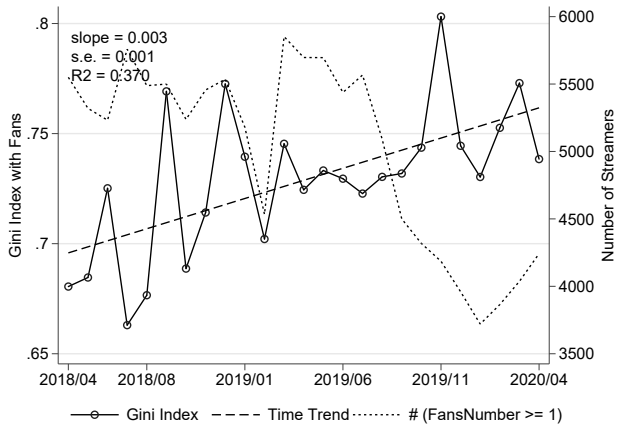
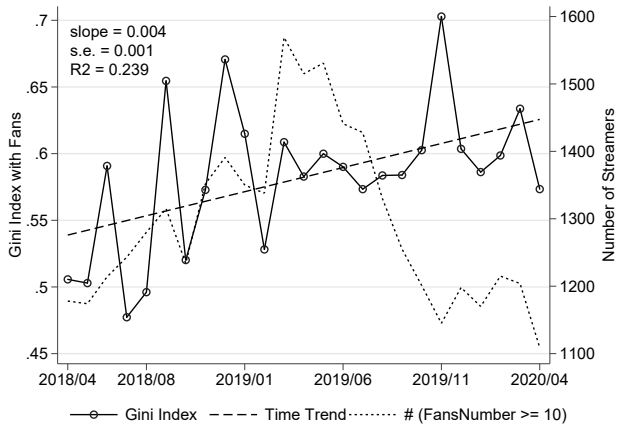


Figure A14. **Seasonality-adjusted Gini index with fans.** This figure shows the seasonality-adjusted inequality trend as measured by the Gini index. We aggregate streamers' loyal fans over the past three months and calculate the Gini index using the resulting fans distribution. In Panel A, the dotted 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* ≥ 1



B. *Number of fans* ≥ 10



C. *Number of fans* ≥ 100

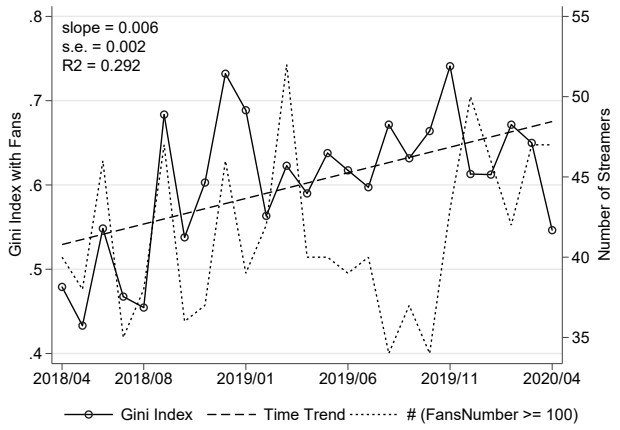


Figure A15. **Gini index with fans.** This figure shows the inequality trend as measured by the Gini index. We calculate the Gini index using streamers' monthly fans distribution. In Panel A, the dotted line shows the number of streamers with any loyal fans, the solid line shows the Gini coefficient 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.

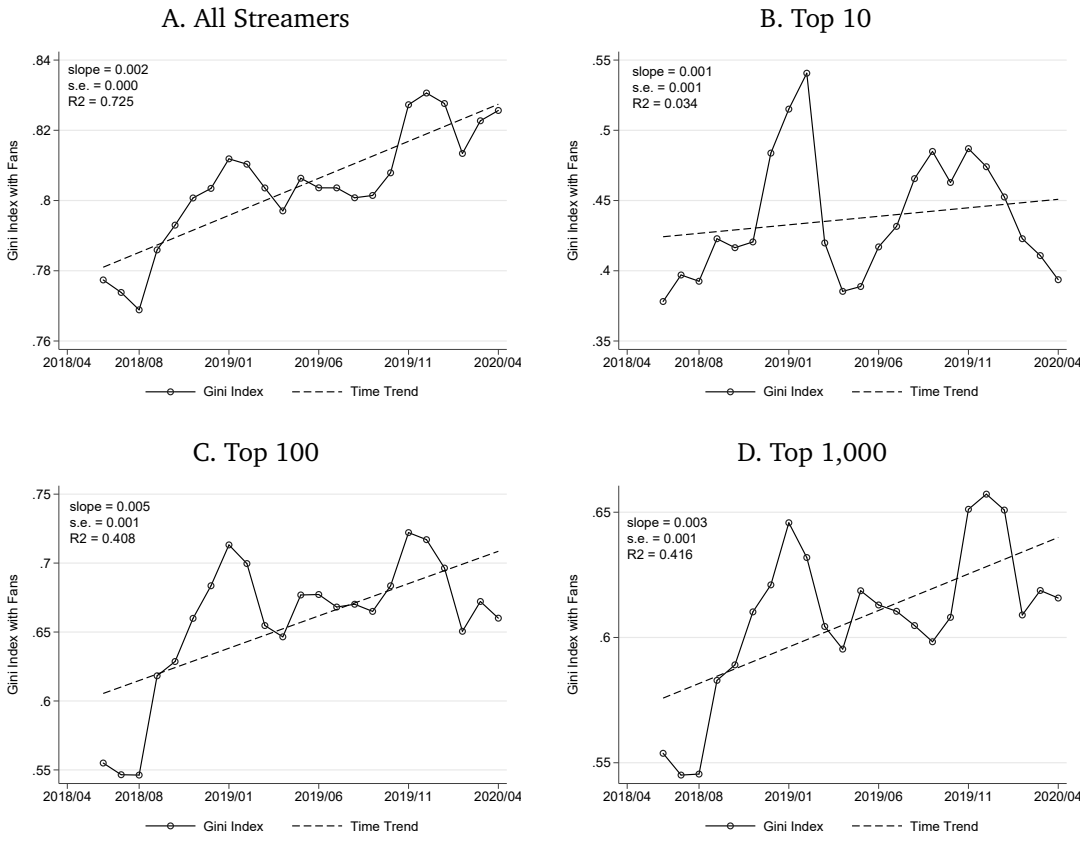


Figure A16. **Seasonality-adjusted Gini index with fans among top streamers.** This figure shows the seasonality-adjusted inequality trend as measured by the Gini index within top streamers. We aggregate streamers' loyal fans over the past three months, rank them according to their aggregate number of 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.

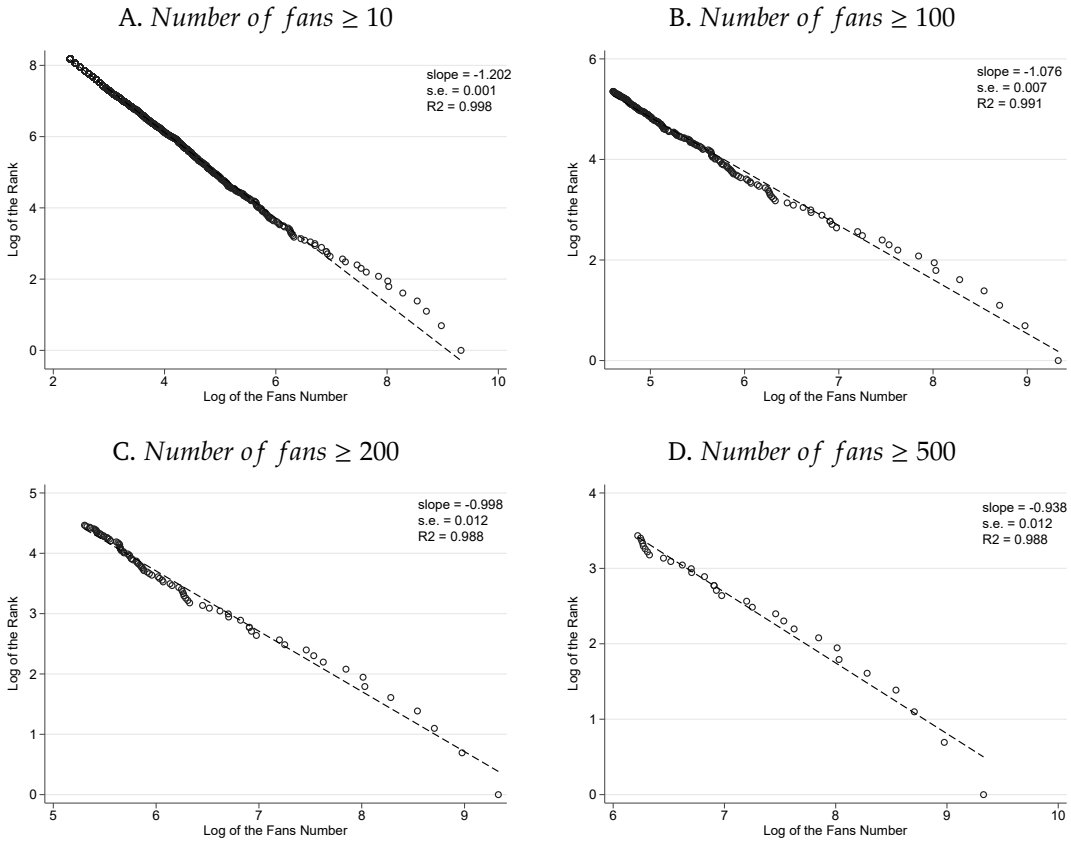


Figure A17. **Log number of fans versus log rank of the streamers.** Streamers are ordered according to the total number of their loyal fans during the whole sample period. The log value of rank is placed on the y-axis, and the log of streamers' total number of loyal fans is placed on the x-axis. From 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.

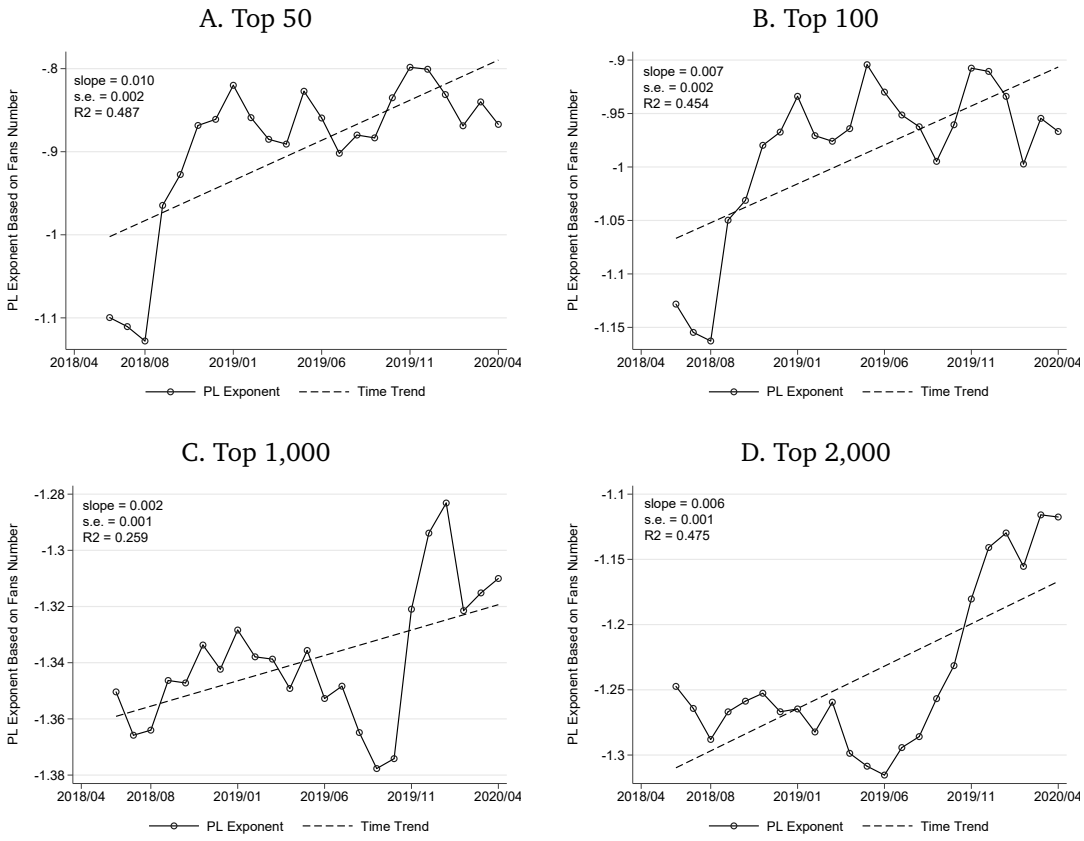


Figure A18. **Seasonally adjusted PL exponent with fans.** This figure shows the dynamic of inequality as measured by seasonality-adjusted PL exponent with fans. We aggregate streamers' number of loyal fans over the past three months, rank them according to their aggregate number of loyal fans, and calculate 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.

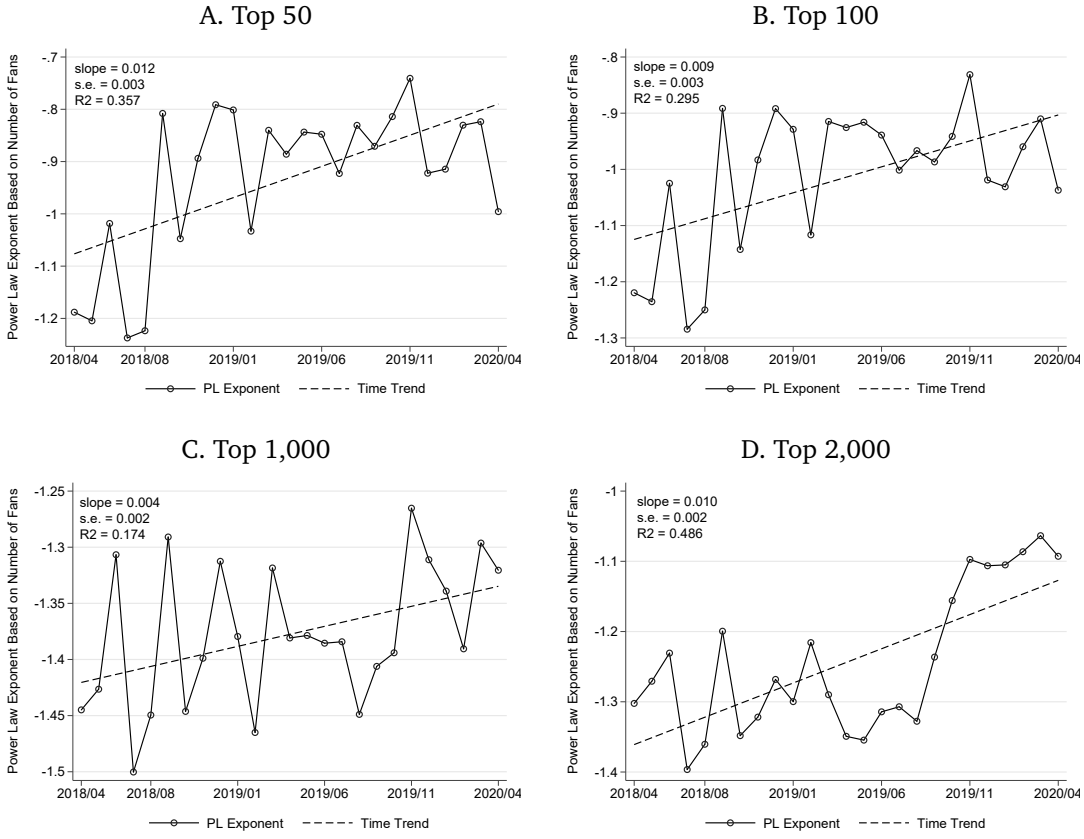


Figure A19. **PL exponent with fans.** This figure shows the dynamic of inequality as measured by the PL exponent with fans. We rank the streamers according to their monthly number of loyal fans and calculate the PL exponent using the resulting fans distribution of those top streamers. Specifically, 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.

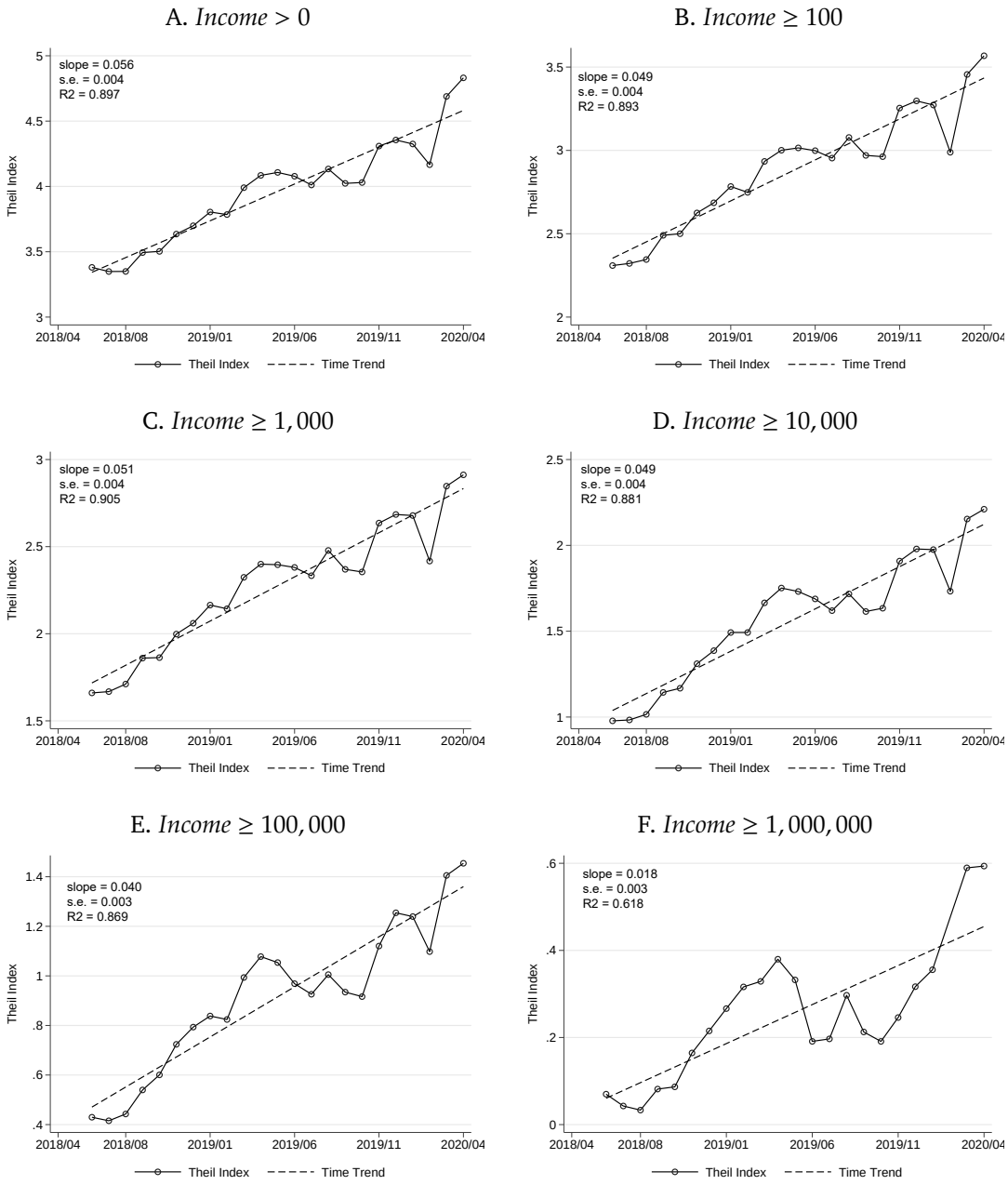


Figure A20. **Seasonality-adjusted Theil index.** This figure shows the seasonality-adjusted inequality trend as measured by the Theil index. We aggregate streamers' tipping income over the past three months and calculate the Theil index using the resulting income distribution. In Panel A, the solid line shows the Theil coefficient calculated based on this sample, and the dashed line is the linear fit time trend of the Theil 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.

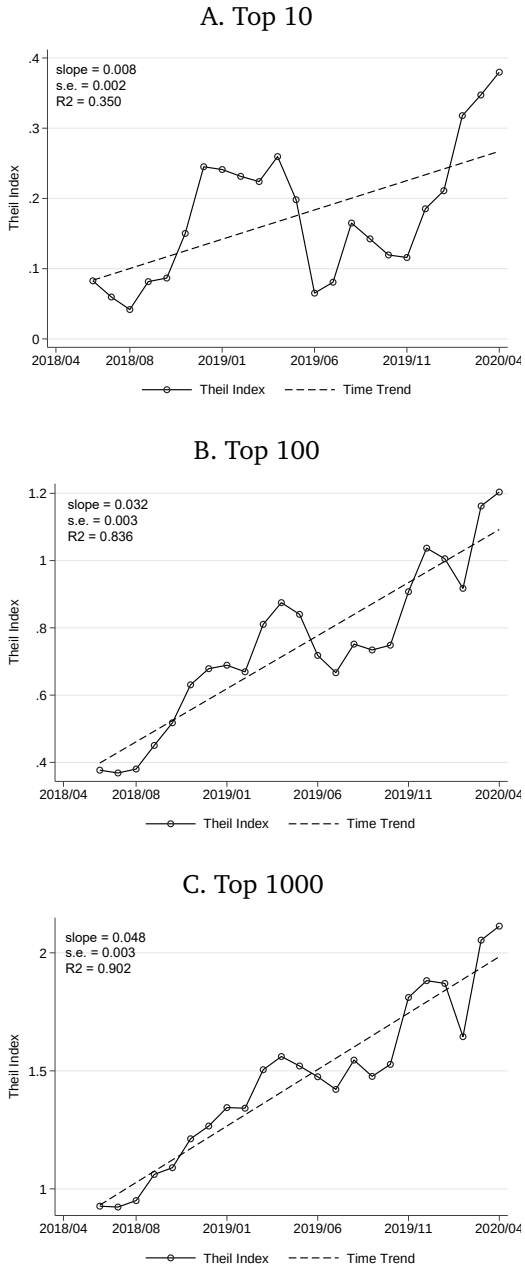


Figure A21. **Seasonality-adjusted Theil index among top streamers.** This figure shows the seasonality-adjusted inequality trend as measured by the Theil index within top streamers. We aggregate streamers' tipping income over the past three months, rank the streamers according to their aggregate income, and calculate the Theil index using the resulting income distribution of those top streamers. From Panel A to Panel C, the Theil 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.

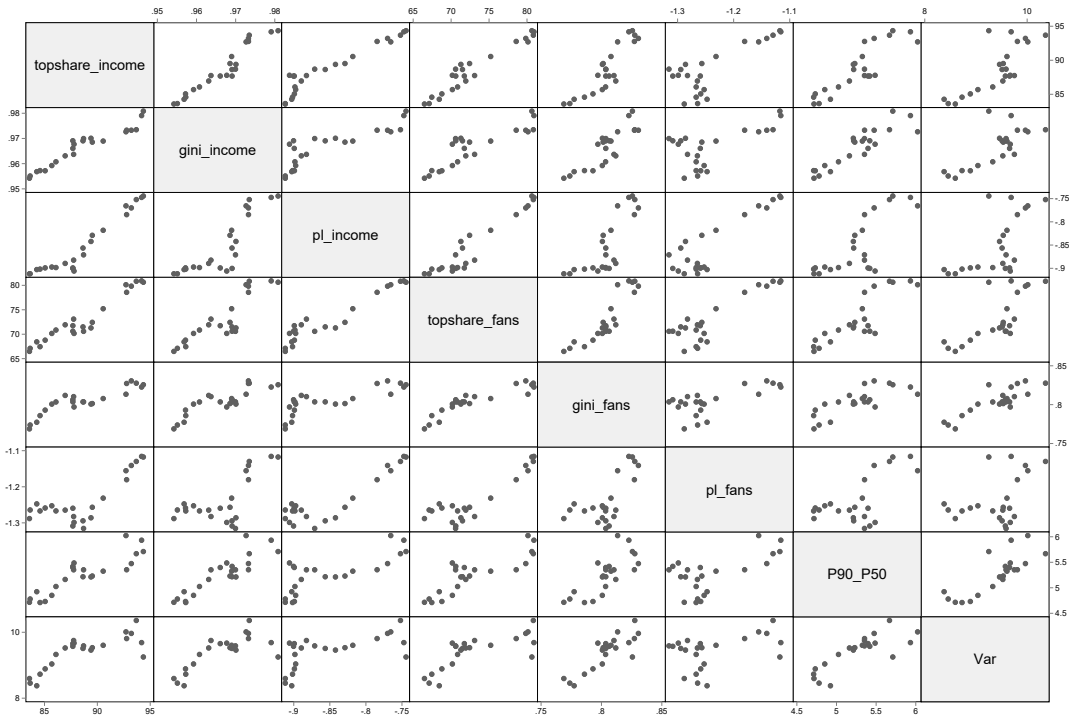


Figure A22. **Correlation matrix.** This figure reports the correlations between the various measures of inequality calculated in Section 3. The specific definitions of the variables shown in the figure are as follows: *topshare_income* refers to the income share of the top 1,000 streamers ordered by 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.

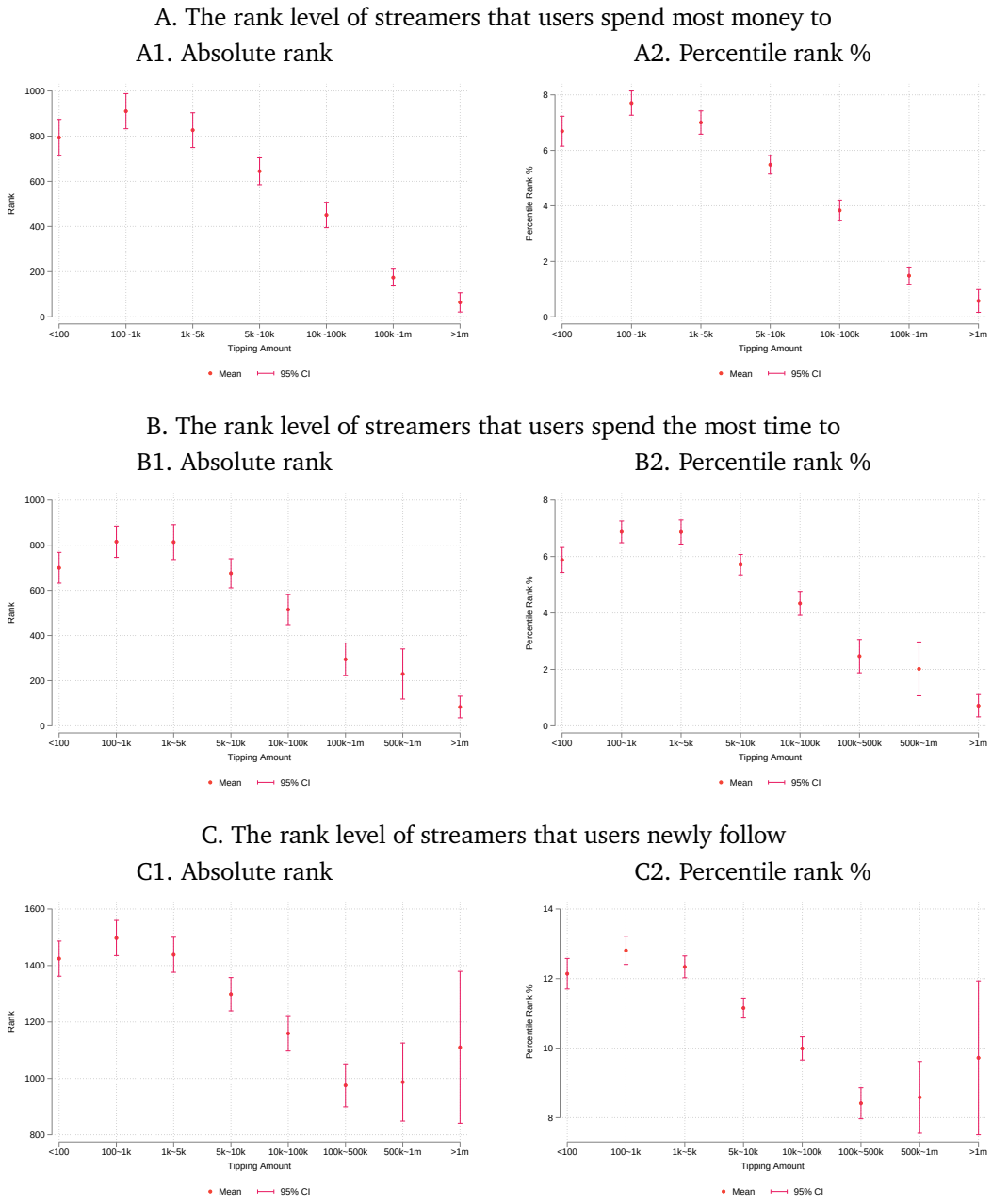


Figure A23. **The rank level of streamers that users appreciate with different consumption levels.** This figure provides evidence of whether active users who spend a lot would disproportionately appreciate the top streamers who already have cumulative advantages. For each month t, we divide users into 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.

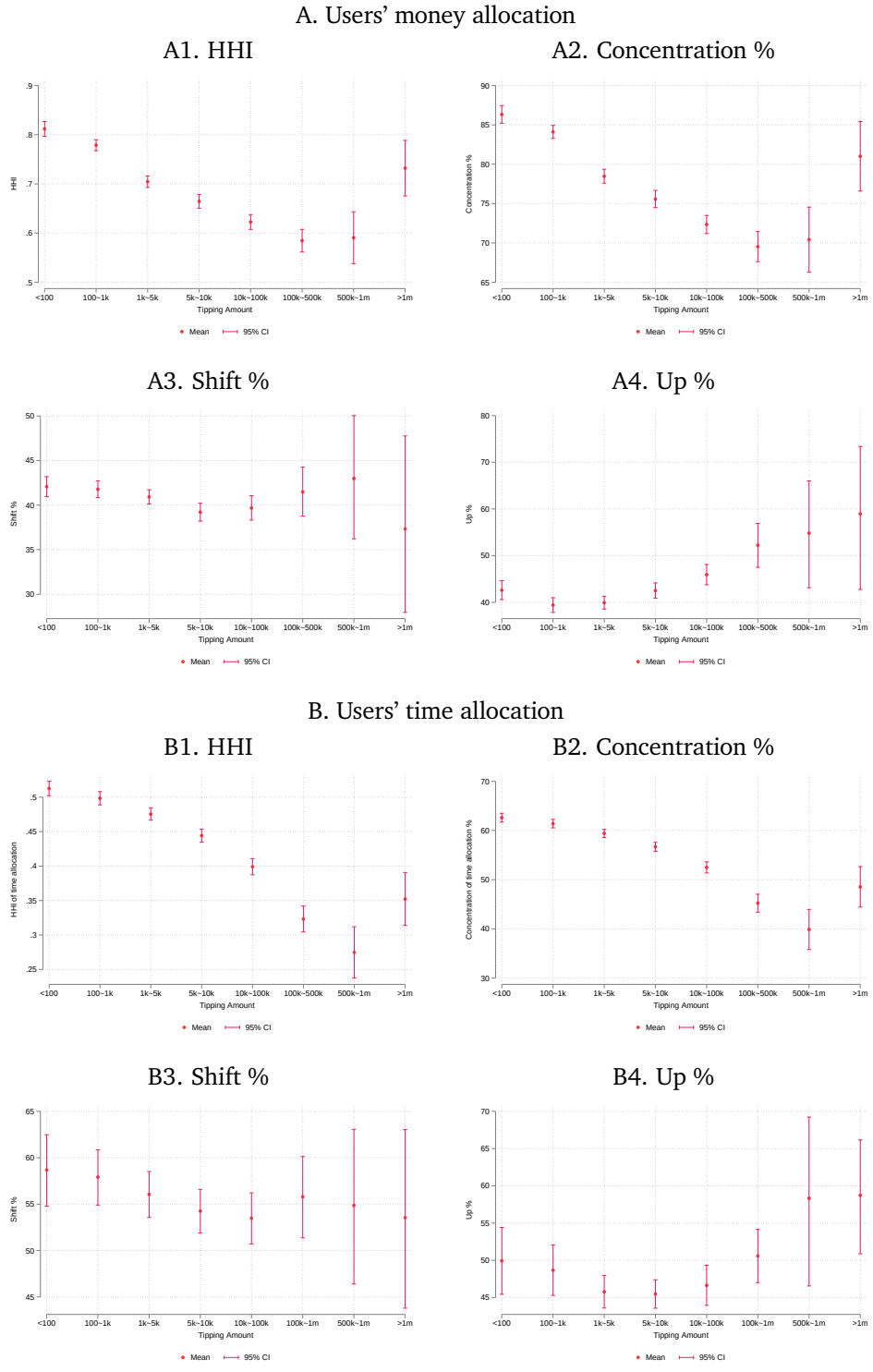


Figure A24. **Loyalty of users with different consumption levels.** This figure provides evidence whether users with high consumption levels would have exceptionally high loyalty. When measuring loyalty, we focus on users' money allocation in Panel A, while in Panel B, we investigate users' viewing time 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.

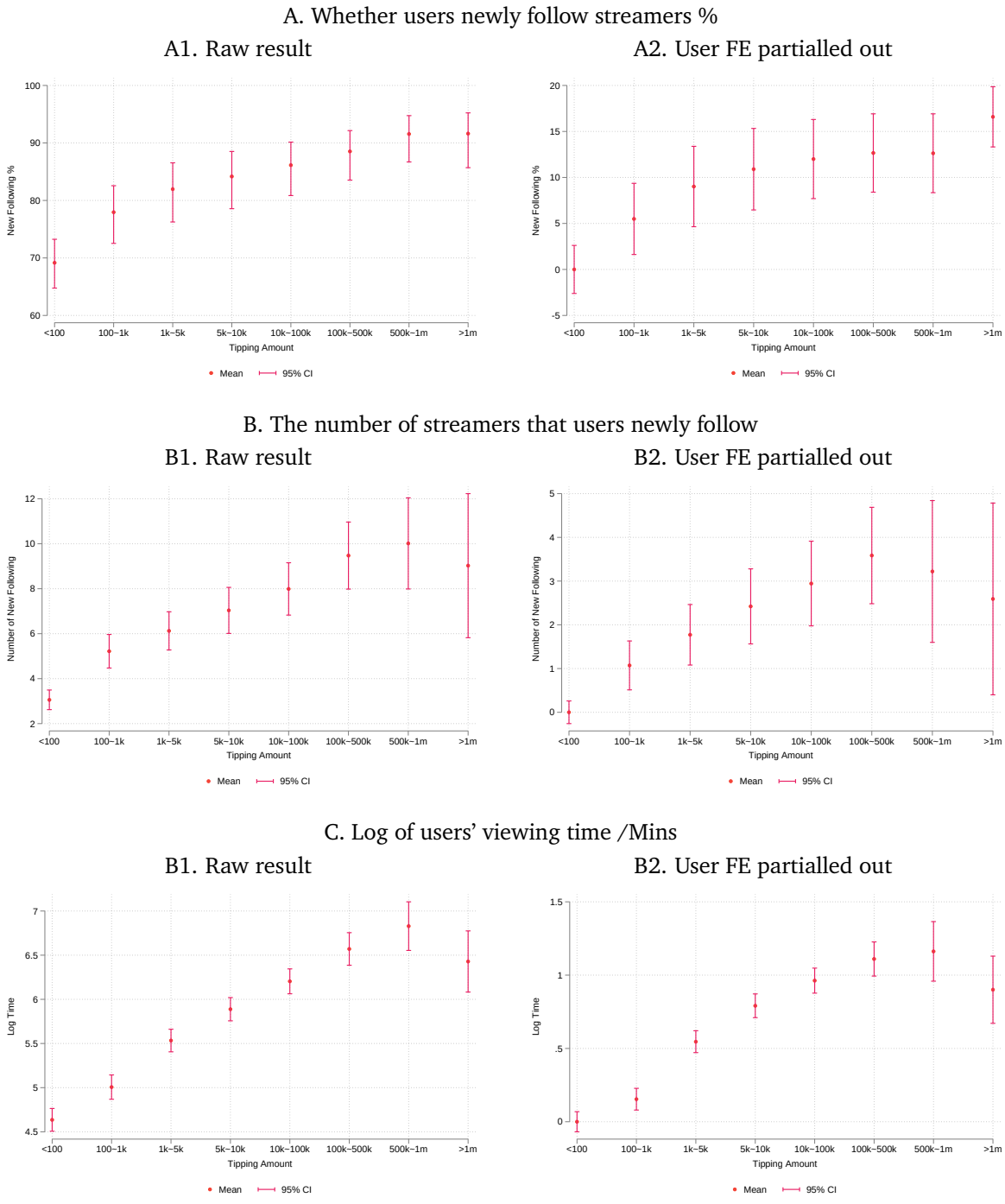


Figure A25. **Engagement of users with different consumption levels.** This figure illustrates whether users with deep pockets would also have high engagement. For each month t, we divide users into eight groups based on their monthly tipping amount. Panel A and Panel B show how active users are in exploring 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.

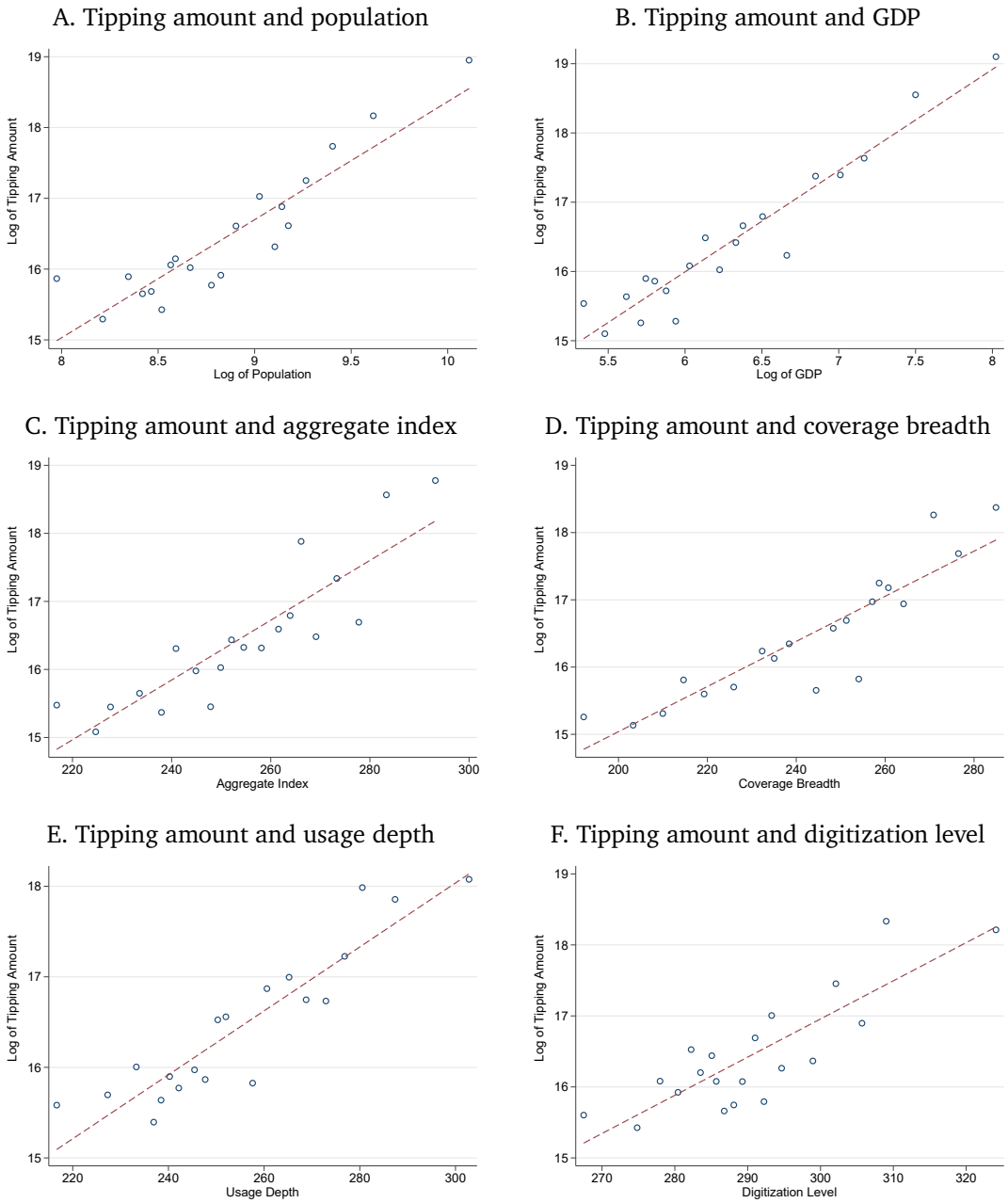
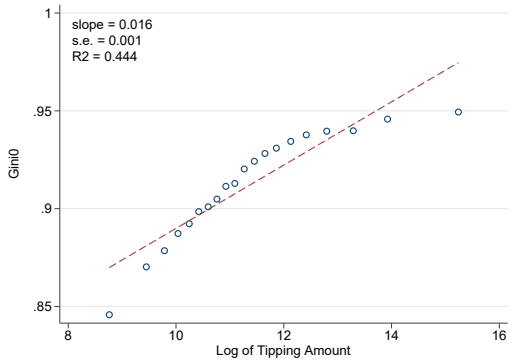
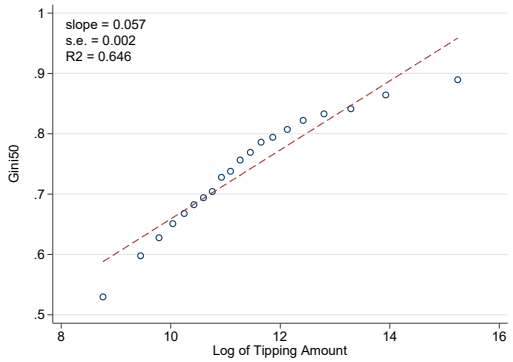


Figure A26. **Correlation between market size and city size.** In Section 4.3, we use one city's population, GDP, and digital financial inclusion index as instrumental variables for the platform's actual market size. This figure presents the correlation between the city size and local market size. We 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

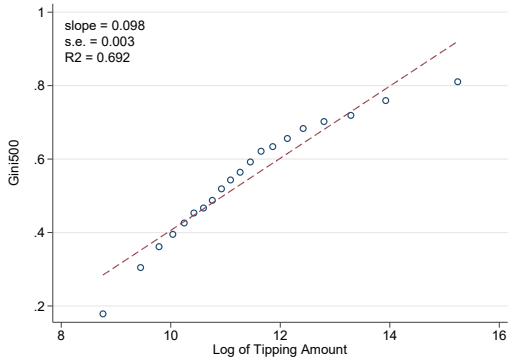


Figure A27. **City-month level relation between market size and inequality.** This figure provides intuitive city-month-level panel evidence that a broader market positively correlates with greater inequality. In this figure, we use total tipping amount as an indicator of local market size. From Panel 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.

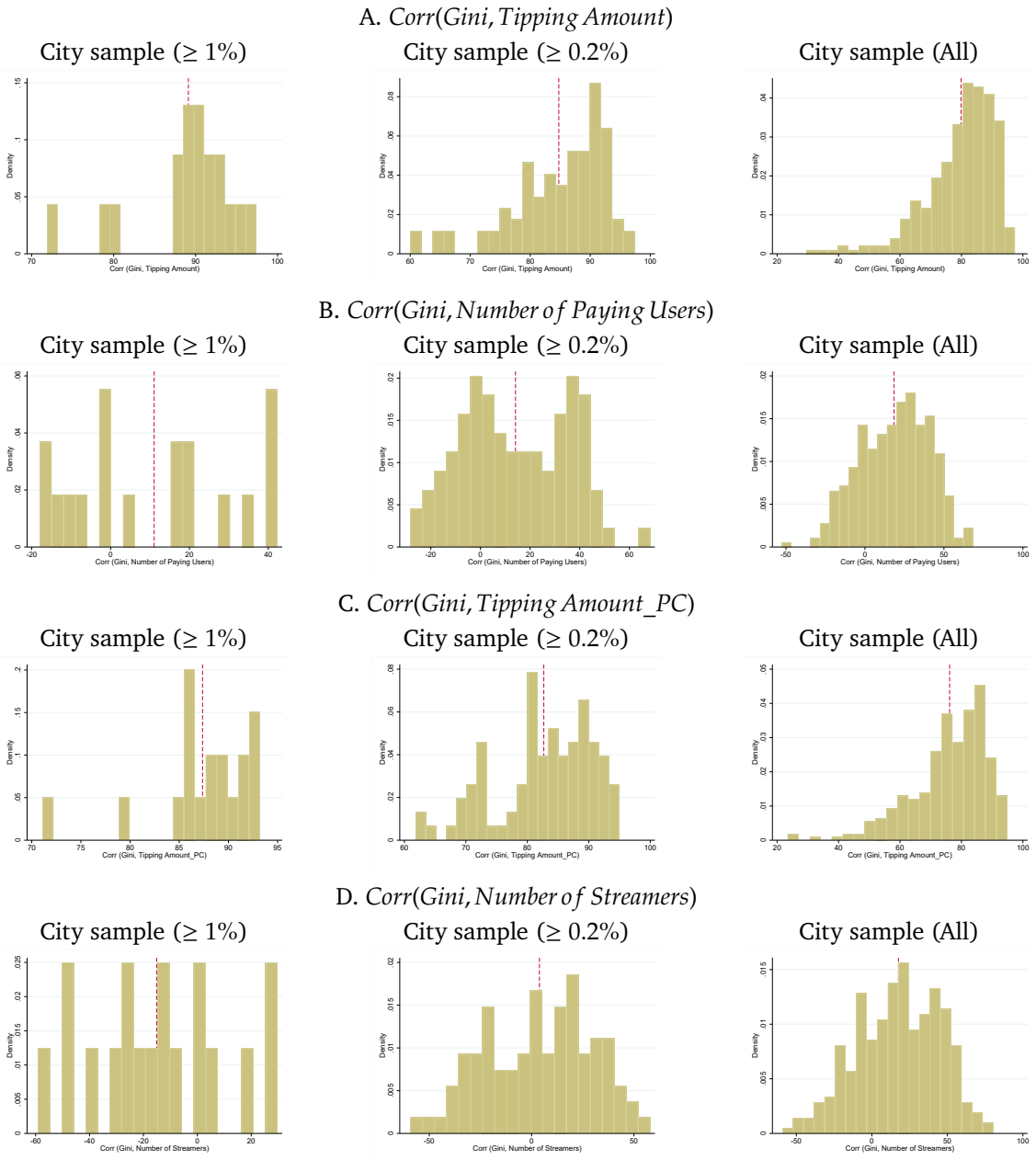


Figure A28. **Distribution of correlation between market size and Gini coefficients.** Using city-month level panel observations, we plot the density distribution of correlation between market size variables and Gini coefficients, and the red dashed line marks their mean value. The correlation coefficients 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.

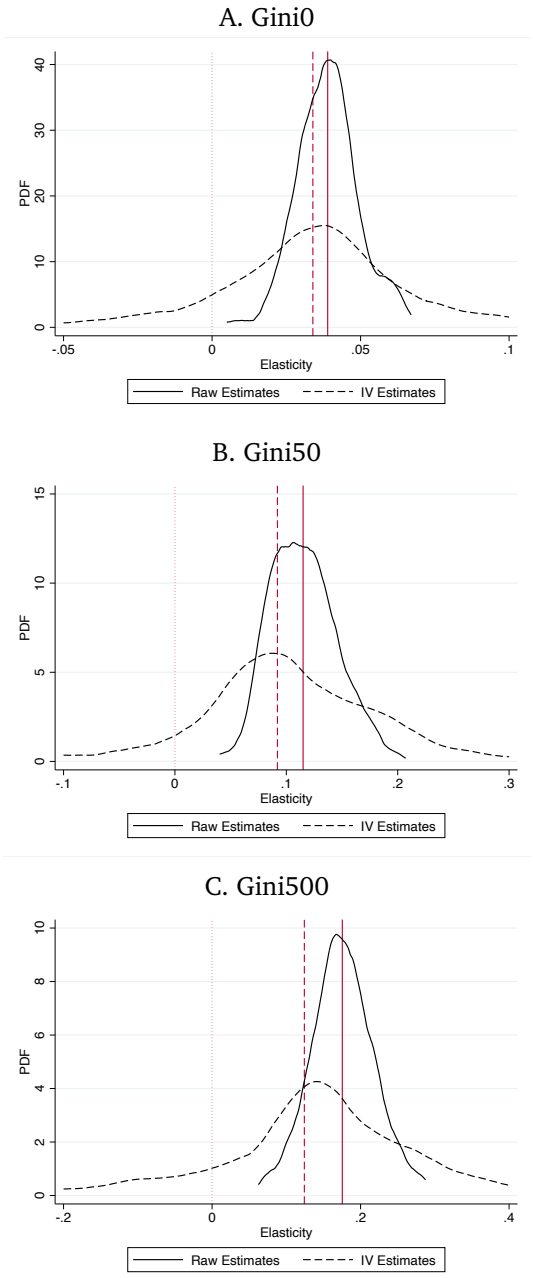


Figure A29. **OLS estimation versus IV estimation.** For each city j, we run the following OLS and IV regressions and then plot the probability distribution function of β_j: $Gini_{jt} = \alpha_j + \beta_{OLS,j} \ln(TippingAmount_{jt}) + \eta_t + \epsilon_{jt}, \forall j$ and $Gini_{jt} = \alpha_j + \beta_{IV,j} \ln(\hat{TippingAmount}_{jt}) + \eta_t + \epsilon_{jt}, \forall j$. In Panels A to C, we measure the inequality using $Gini0$, $Gini50$, and $Gini500$, 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.

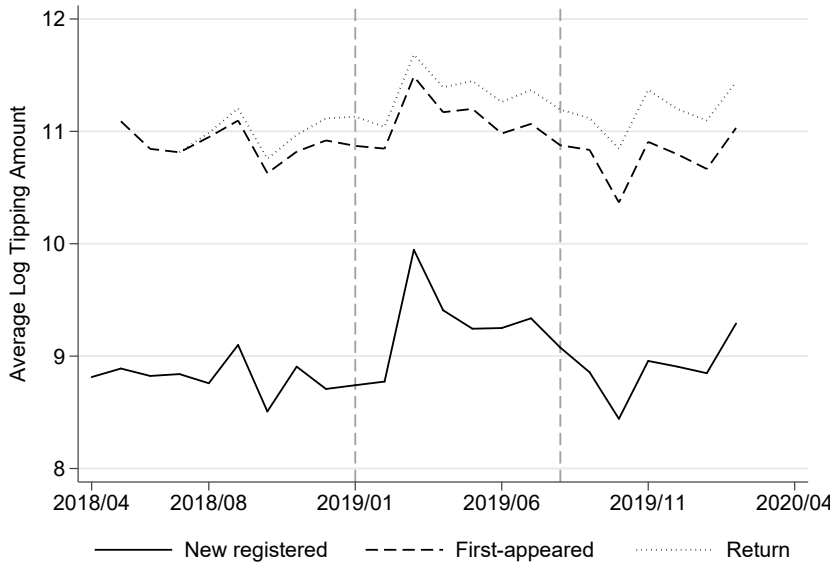


Figure A30. **The user growth rises and down around the capital raising event.** This figure plots the average log value of the tipping amount contributed by new users each month. We calculate the average value for each month based on the 302 cities. To exclude the impact of the Covid-19 pandemic, we 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.

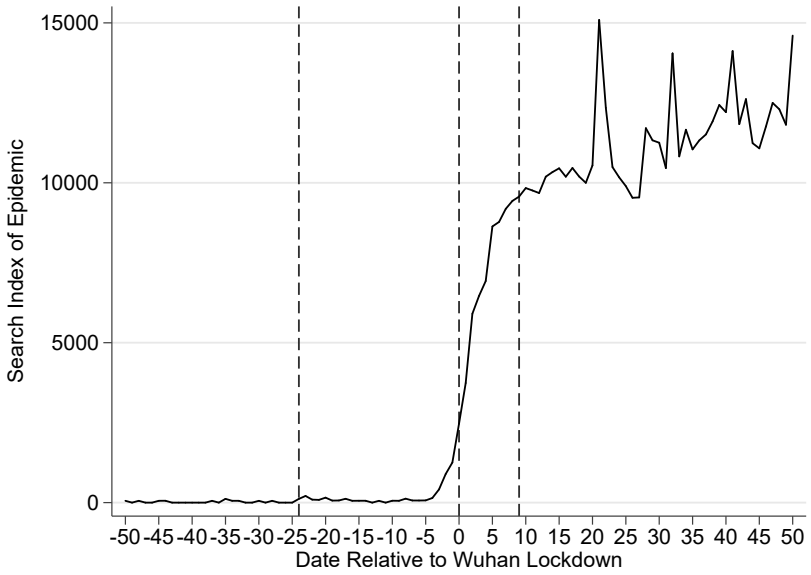


Figure A31. **Citizens' attention to the Covid-19 epidemic in Wuhan.** This figure shows Wuhan citizens' attention to the Covid-19 epidemic, as measured by the Baidu search index of the Chinese equivalent of the keyword "Epidemic". The search volume data were obtained from the website of "Baidu Index" (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.

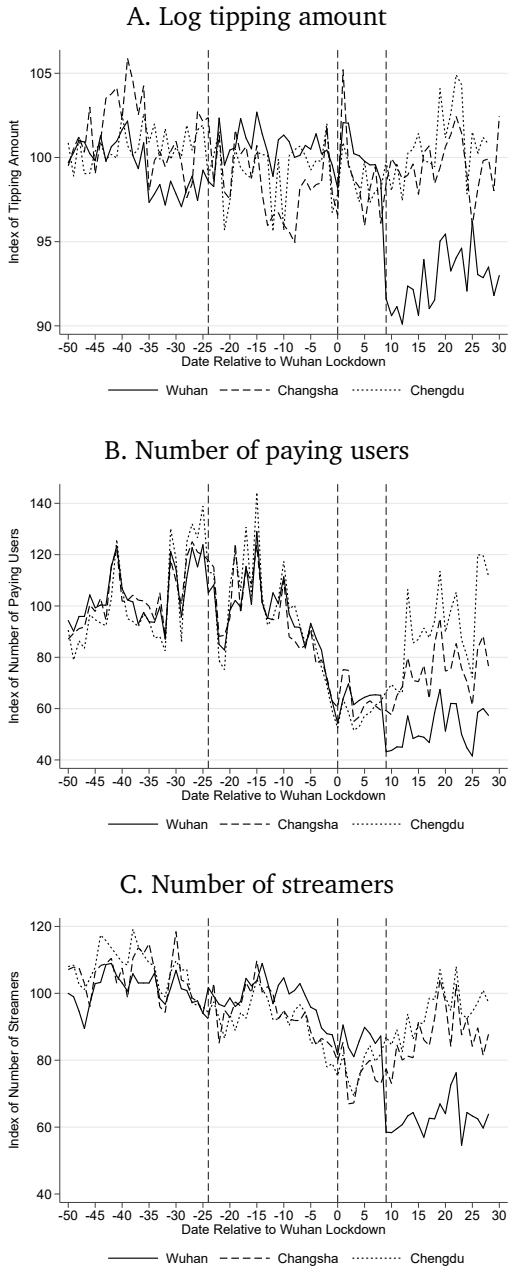


Figure A32. **Entertainment demand shock.** This figure plots the entertainment demand response to the Wuhan lockdown shock in three representative cities, the epicenter city Wuhan, its nearby city Changsha and the distant city Chengdu. The x-axis indicates the days relative to Wuhan lockdown on January 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.

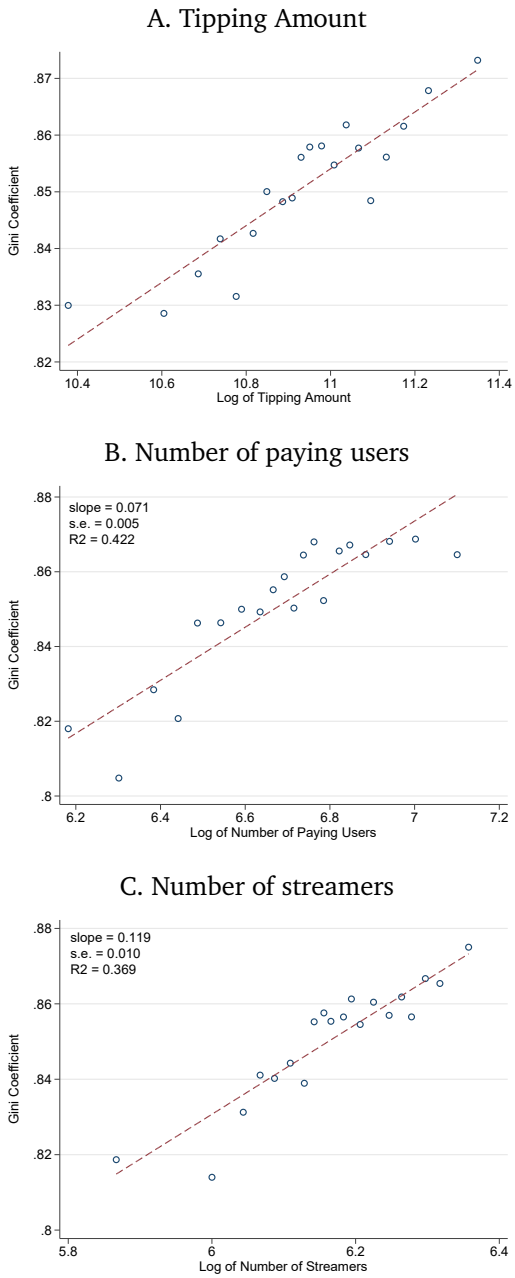


Figure A33. **City-daily level relation between market size and inequality.** This figure provides intuitive city-day-level panel evidence that a broader market positively correlates with greater inequality. We use $Gini50_{j,t}$ as a measure of inequality, which is calculated using the income distribution 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.

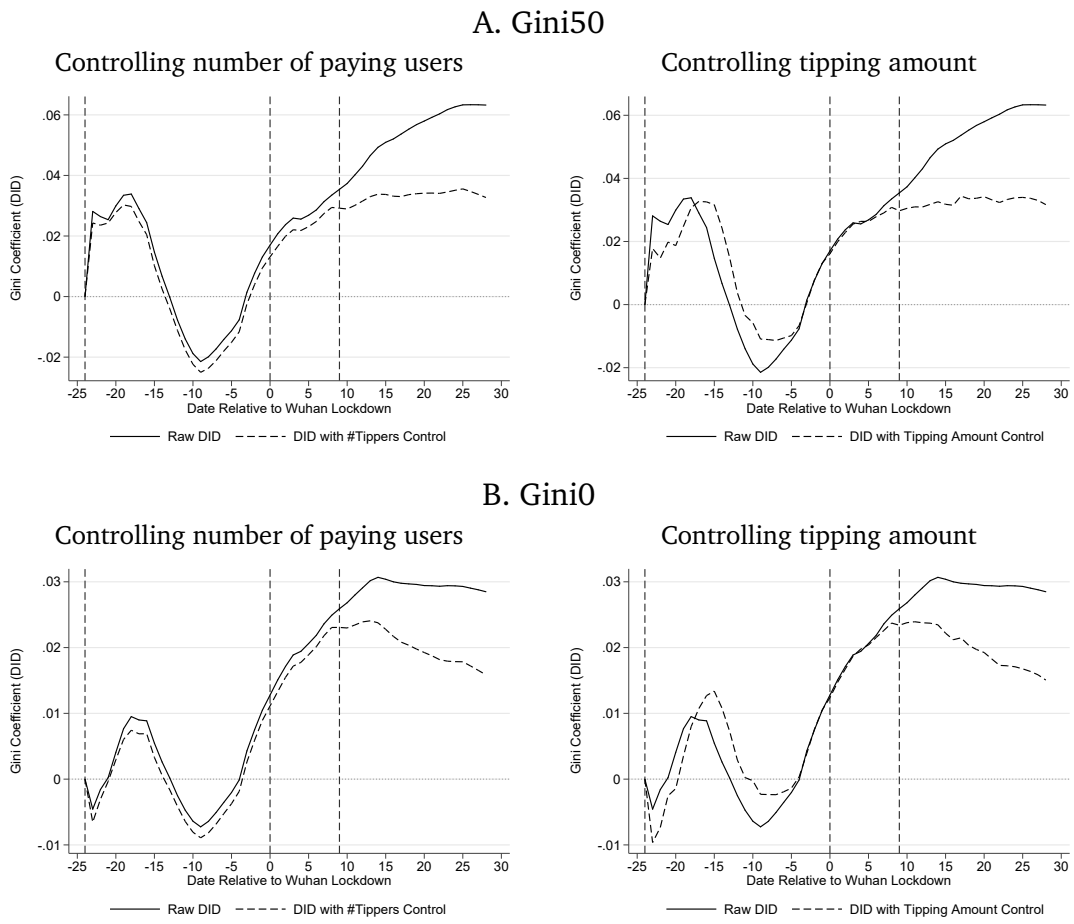
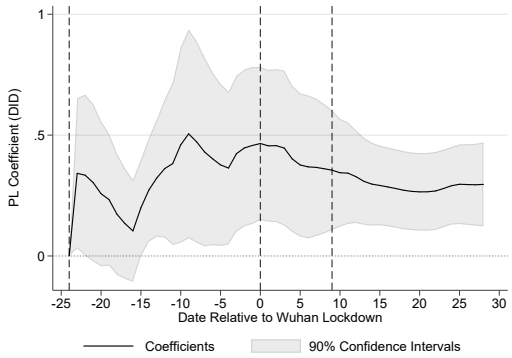
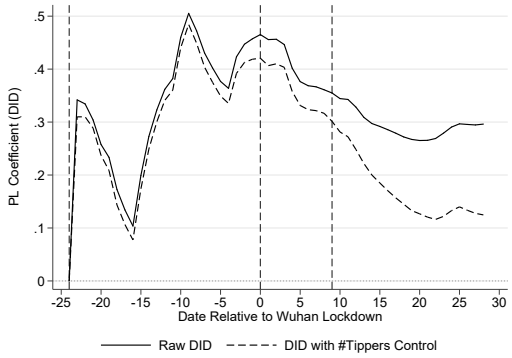


Figure A34. **The dynamic treatment effects with market size variables controlled.** In Figure **??**, we plot estimated $\beta_{raw,T}$ in the following regressions $Gini_{j,t} = \alpha_{raw,T} + \beta_{raw,T} ln(Distance_j) \times Post_{1t} + \gamma_{raw,T} Post_{1t} + \theta_j + \epsilon_{j,t}$, for $t \in [-50, T]$ 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

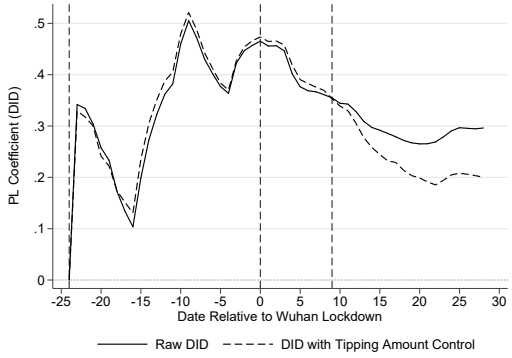


Figure A35. **The dynamic treatment effects using PL exponent as the dependent variable.** We estimate $\beta_{raw,T}$ in the DID regressions $PL_{j,t} = \alpha_{raw,T} + \beta_{raw,T} ln(Distance_j) \times Post_{1t} + \gamma_{raw,T} Post_{1t} + \theta_j + \epsilon_{j,t}$, for $t \in [-50, T]$ and $\forall T > -24$. 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.

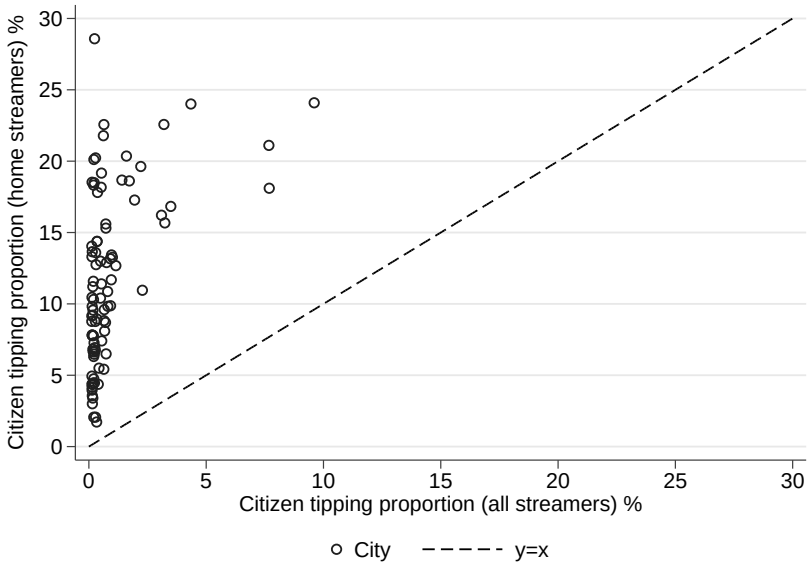


Figure A36. **Home bias effect in tipping behavior.** The horizontal axis denotes the proportion of income that users contribute to all streamers, while the vertical axis indicates the proportion of income that users contribute to streamers from the same city. The dotted line represents a 45° line. If the citizens of city j do not exhibit any preference for local streamers, city j would be positioned on this line.

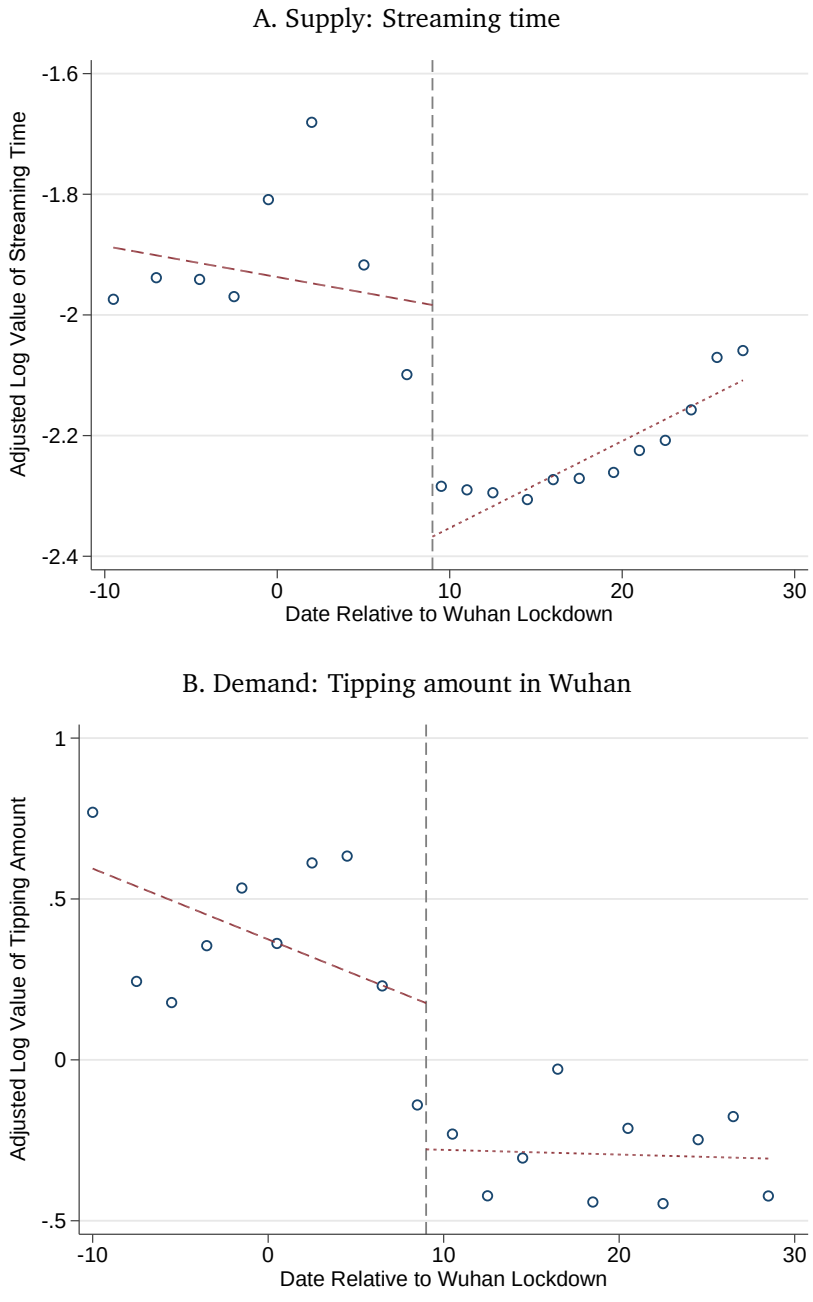


Figure A37. **Supply response to Covid-19 in Wuhan.** Panel A in this figure shows the discontinuity in streaming time for streamers in Wuhan around Day 9 (February 1, 2020). From that day on, the epidemic became serious (at least to an extent), the citizens of Wuhan became more concerned about it, and the government implemented a series of escalating protective measures. Panel B shows the discontinuity in demand side (log value of tipping amount). The red dashed line in each sub-figure is the linear fit line before and after the breakpoint, Day 9.

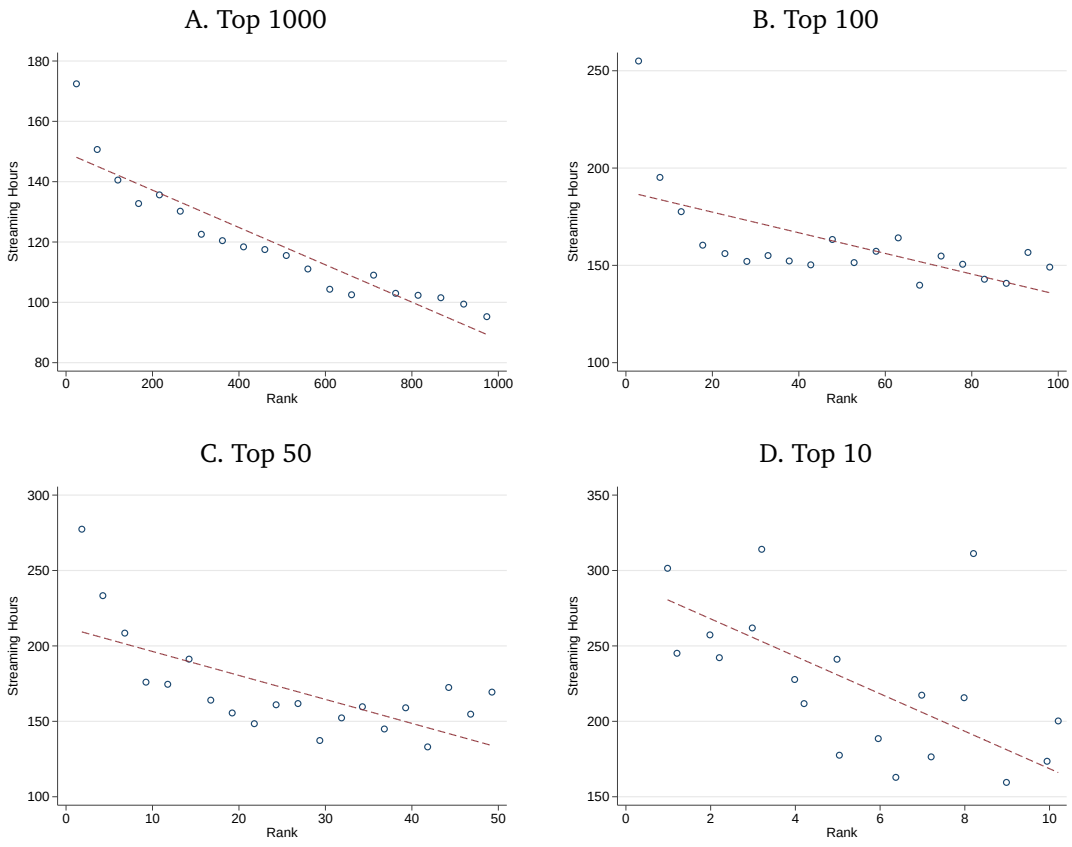


Figure A38. **The relationship between income rank and streaming hours of streamers per month.** This figure illustrates the correlation between streamers' working hours and their respective income ranks in the previous month. The x-axis represents the streamers' tipping income rank in month $t-1$, while the y-axis indicates their live streaming duration in month t. Specifically, we have utilized the streamers' broadcast records to determine their monthly working hours, thus providing a measure of the effort each streamer expends. In various subplots, we have limited the sample distribution to the top 1000, 100, 50 and 10 streamers, respectively. And in all subfigures, we control for year-month fixed effects.

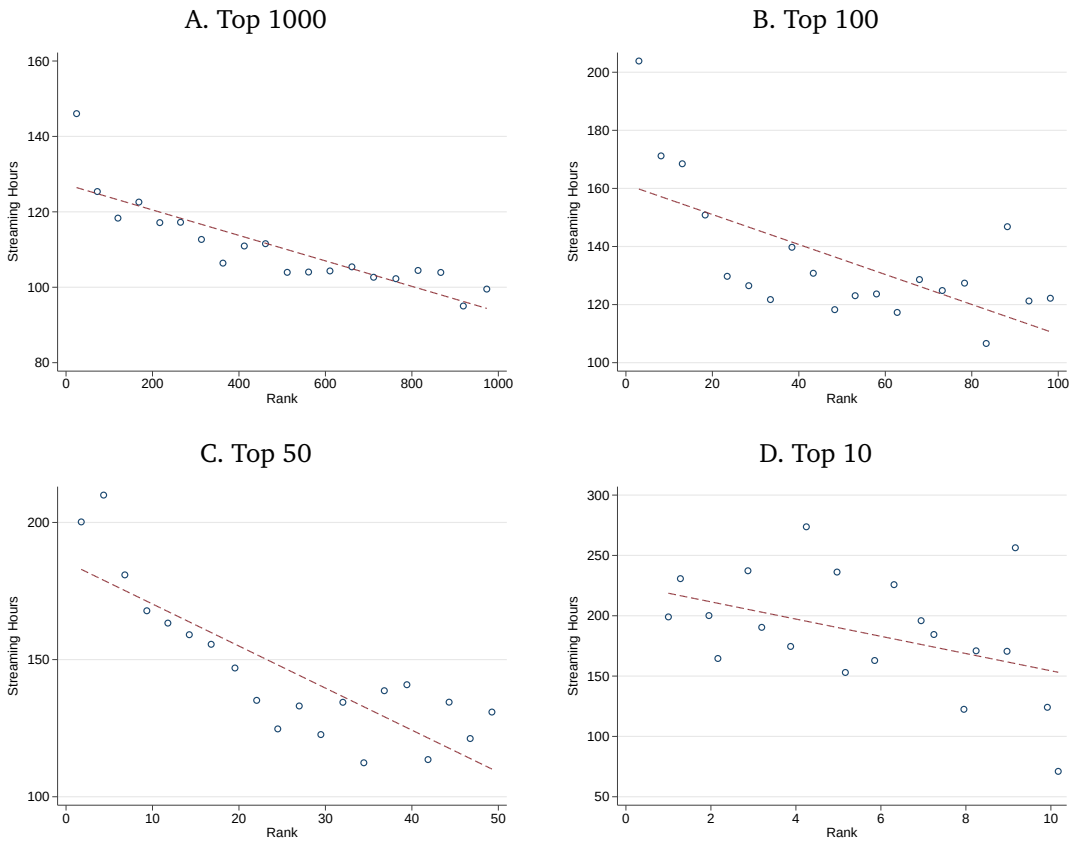


Figure A39. **The relationship between wage rank and streaming hours of streamers per month.** This figure illustrates the correlation between streamers' working hours and their respective wage (income per hour) ranks in the previous month. The x-axis represents the streamers' tipping income per hour rank in month $t-1$, while the y-axis indicates their live streaming duration in month t. Specifically, we have utilized the streamers' broadcast records to determine their monthly working hours, thus providing a measure of the effort each streamer expends. In various subplots, we have limited the sample distribution to the top 1000, 100, 50 and 10 streamers, respectively. And in all subfigures, we control for year-month fixed effects.

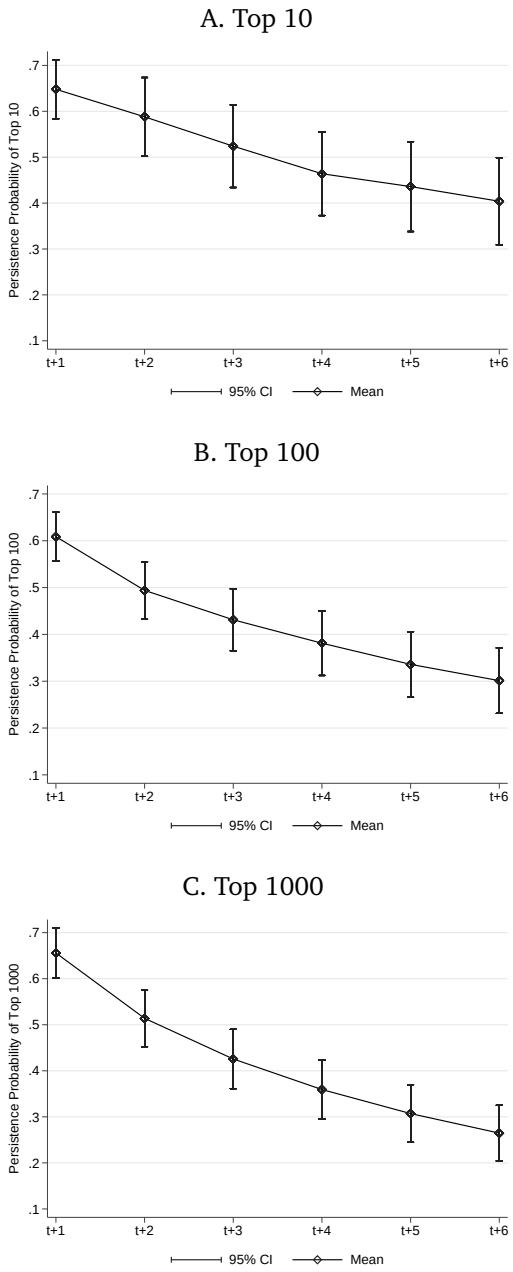


Figure A40. **Persistence probability of top streamers.** This figure plots the persistence probability of top streamers. Specifically, for month t, streamers are ranked by their tipping income, and the probability of remaining in the top tier at $t + 1$ to $t + 6$ months is calculated for top-tier streamers in month t. Then we depict the mean along with the 95% confidence interval. From Panel A to Panel C, we graph the dynamic persistence probabilities for the top 10, 100, and 1000 streamers, respectively.

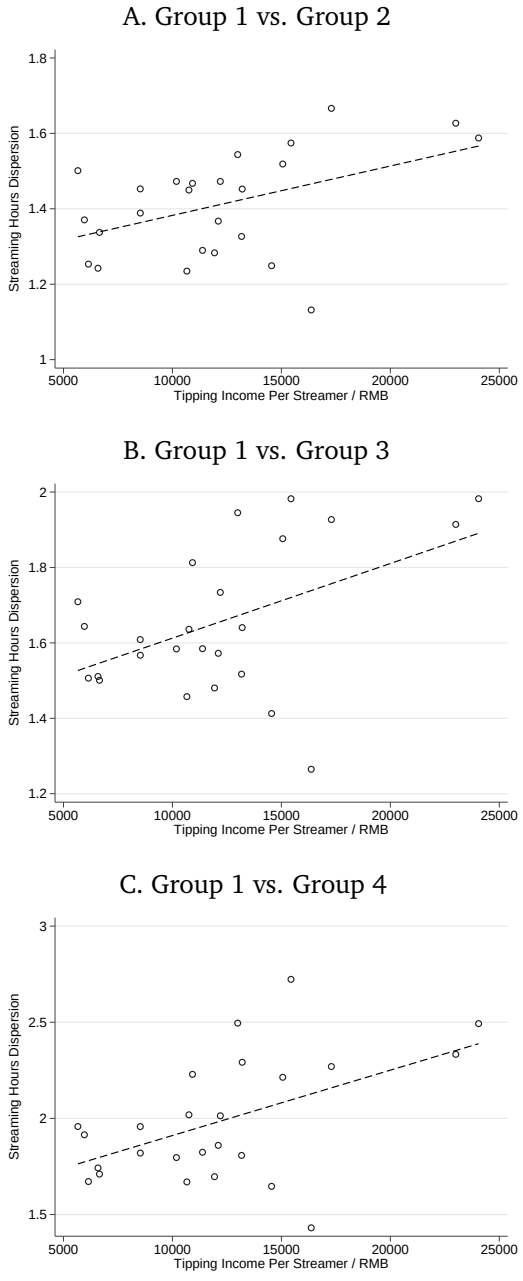


Figure A41. **Monthly relationship between market size and streaming hours dispersion.** Streamers' monthly tipping incomes are ranked and categorized into four distinct groups. Group 1 comprises streamers ranked 1-10, Group 2 encompasses those ranked 10-100, Group 3 includes ranks 100-500, and Group 4 consists of streamers ranked 500-1000. Streaming hours are averaged for each group on a monthly basis. From Panel A to Panel C, the dispersion in streaming hours is represented by the ratio of average streaming hours in Group 1 to those in Groups 2, 3, and 4, respectively. Market size is quantified as the average tipping income per streamer for each corresponding month.

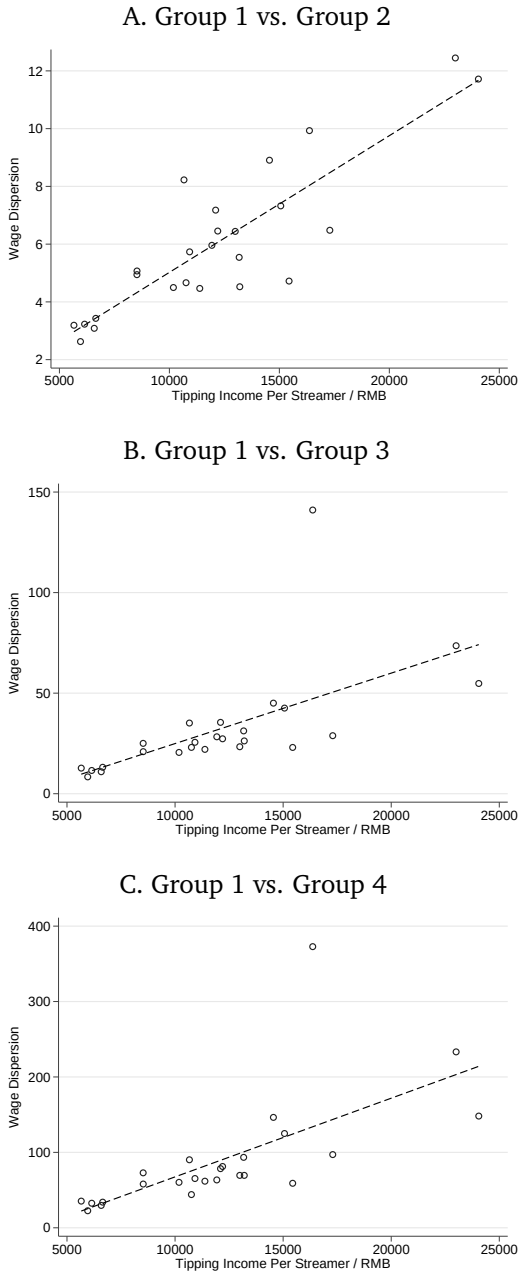


Figure A42. **Monthly relationship between market size and per-hour income dispersion.** Streamers' monthly tipping incomes are ranked and categorized into four distinct groups. Group 1 comprises streamers ranked 1-10, Group 2 encompasses those ranked 10-100, Group 3 includes ranks 100-500, and Group 4 consists of streamers ranked 500-1000. Tipping income per hour are averaged for each group on a monthly basis. From Panel A to Panel C, the dispersion in per-hour income is represented by the ratio of average per-hour income in Group 1 to those in Groups 2, 3, and 4, respectively. Market size is quantified as the average tipping income per streamer for each corresponding month.

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; *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.* 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 **??**. $Gini0_t$ 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 4.1.2. $Gini0_{j,t}$ 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 4.2. 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 Δ *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 $\Delta\, Gini_{j,stage} = \alpha + \beta\Delta \ln(\hat{Market\ Size}_{j,stage}) + \epsilon_j$, where Δ *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 Δ *Gini*, based on the income distribution of streamers whose incomes are positive from city j. In Panel B, *Gini*500 is used to calculate Δ *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: Δ *Gini*0						
$\Delta \ln$(*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: Δ *Gini*500						
$\Delta \ln$(*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 $\Delta Gini_{j,stage} = \alpha + \beta\Delta \ln(\hat{Market\ Size}_{j,stage}) + \epsilon_j$, 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 4.4. $Gini0_{j,t}$ 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

Table A20. **Robustness checks using Gini**$_{fans}$.*Thistableprovidestherobustnesschecksofidentificationpartusingfansnumbertocalc*
***, **, *indicatesignificanceatthe*1%, 5%, *and*10%*levels, respectively*.

	(1)	(2)	(3)	(4)
Panel A: Market Size Seasonality				
	Gini_fans	Gini_fans	Gini_fans	Gini_fans
ln (Num of Paying Users)	0.091***	0.082***	0.057***	0.142***
	(0.019)	(0.020)	(0.014)	(0.041)
Year-Quarter FEs	Yes	Yes	Yes	Yes
City FEs	No	No	Yes	Yes
N	25	25	450	450
Adjusted R2	0.677	0.671	0.701	0.537
K. P. F-statistic		21.091		23.542
Panel B: User Surge before a Capital Raising Event				
			Δ *Gini_fans*	
	Stage 2- Stage1		Stage 3- Stage2	
Δ ln(*Num of Paying Users*)	0.255**	0.125***	0.079*	0.179*
	(0.107)	(0.038)	(0.042)	(0.098)
N	18	92	18	92
Adjusted R2	-0.150	0.338	0.201	0.128
K. P. F-statistic	13.926	31.068	59.804	17.134
Panel C: City Features				
	OLS	ln(Population)	ln(GDP)	Aggregate Index
ln(*NumofPayingUsers*)	0.071***	0.060***	0.066***	0.082***
	(0.003)	(0.006)	(0.004)	(0.005)
N	92	92	92	92
Adjusted R2	0.718	0.700	0.714	0.704
K.P.F-statistic		164.788	591.956	93.492
Panel D: Event-based Gini Coefficient Response				
City		Δ(*Gini*)	Δ(*MarketSize*)	Ratio
Chengdu		-0.458**	-1.924***	0.238
Shenzhen		-0.631**	-1.686***	0.374
Beijing		-0.726*	-1.628***	0.446
Shanghai		-0.208	-1.319***	0.158

Table A21. **The relationship between streaming hours and rank of streamers.** In this table, we investigate the relationship between the streaming duration hours and ranking of live streamers. We restrict our sample to full-time streamers, defined as those whose average streaming duration has been at least 40 hours over the past three months. The dependent variable is the logarithm of the number of streaming hours in month t. In Columns (1)-(3), the independent variable is the streamer's tipping income ranking in month $t-1$. In Columns (4)-(6), the independent variable is the streamer's per-hour tipping income ranking in month $t-1$. In Columns (2) and (4), we control for year-month fixed effects, while in Columns (3) and (6), we account for both year-month and individual streamer fixed effects. Standard errors are clustered at the year-month and streamer level and reported in parentheses. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
	ln(Streaming Hours)					
ln(Rank)	-0.279***	-0.278***	-0.273***			
	(0.012)	(0.011)	(0.015)			
ln(Rank_wage)				-0.163***	-0.160***	-0.052***
				(0.011)	(0.011)	(0.014)
Year-Month FEs	No	Yes	Yes	No	Yes	Yes
Streamer FEs	No	No	Yes	No	No	Yes
N	43273	43273	39442	43273	43273	39442
Adj. R2	0.061	0.069	0.343	0.020	0.028	0.328