HOW ALLOWING A LITTLE BIT OF DISSENT HELPS CONTROL SOCIAL MEDIA: IMPACT OF MARKET STRUCTURE ON CENSORSHIP COMPLIANCE

Zhenqi (Jessie) Liu
Johns Hopkins University, United States

Abstract
This paper studies the role of market structure in regulatory compliance through a unique empirical example: censorship via content removal by three major live-streaming platforms in China. Based on 30 unexpected sensitive events, I first present reduced-form evidence that the largest platform censored a higher number of keywords and complied faster on average than the smaller platforms. I then develop and estimate a structural model where three platforms compete for users by choosing whether to comply with the government's censorship requests. By complying immediately, platforms may lose users who prefer to evade censorship by switching out. By delaying compliance, platforms incur a legal cost; however, the delay allows them to attract new users from their competitors who quickly comply with the government's censorship requests. My counterfactual analysis predicts that if the less compliant small platform were to be shut down or merged with the medium platform, the remaining platforms could comply less often in equilibrium due to increased strategic incentives. This suggests that decentralizing online market power might help an authoritarian government control social media. (JEL: P26, L82, D22)

The editor in charge of this paper was Pierre Dubois.

Acknowledgments: The author thanks Pinar Yildirim, Camilo Garcia-Jimeno, Juan Pablo Atal, Andrew Shephard, Masashi Crete-Nishihata, Jeffery Knockel, Ying Xie, Mathias Reynaert, Andrew Ching, Muxin Li, Ginger Zhe, David Yang, Karam Kang, Michael Keane, Harun Alp, and seminar participants at the University of Pennsylvania, University of Illinois at Chicago, University of Delaware, Johns Hopkins Carey Business School, University of Oxford, University of Hong Kong, City University of Hong Kong, and University of Western Ontario, as well as attendees at the Digital Economics Conference at TSE, the 2021 Virtual Meeting of the International Industrial Organization Conference, and the NBER Political Economy Summer Institute for helpful comments. The author is especially grateful to Masashi Crete-Nishihata and Jeffery Knockel of the Citizen Lab for sharing and facilitating access to the primary data used in this project. This research was partially funded by the Johns Hopkins Carey Business School. Special thanks to Whitney Birenbaum for her invaluable copy-editing assistance.

E-mail: jzliu@jhu.edu

Journal of the European Economic Association 2025 00(0):1–38
https://doi.org/10.1093/jeea/jvaf018

1. Introduction

On August 12, 2015, a series of massive explosions occurred in the city of Tianjin in northeast China. According to a BBC report,[1] over 100 people were killed and nearly 800 others were injured. Shortly thereafter, the public started to speculate over the cause of this event. Some rumors circulated that an overheated chemical warehouse, which had operated illegally for years because of its owner's strong political ties with the government, was the culprit. After the blasts, local residents seeking compensation for their homes protested in front of the daily press conference venue.[2] During the protest, the Chinese government disseminated urgent albeit vague guidelines to social media platforms, requesting them to censor online discussions on the Tianjin Explosion.[3] Some social media platforms complied immediately by censoring user messages that contained keywords such as "Tianjin massive explosion" by implementing a method called "automated keyword filtering." Some other platforms, however, delayed compliance despite the potential risk of penalization from the Chinese government. In fact, social media companies in China "appear to have a degree of flexibility in determining when and what specific keywords to block" (Knockel 2016) and the decision of whether to comply is not as simple as it may seem on the surface.

Beyond China, other countries have also been pushing for stricter regulations on social media platforms following data breach scandals and instances of terrorist groups using social media as a recruiting device (Arnold 2018). For example, in 2017, Germany's parliament passed a law forcing social networks to delete hate-speech postings and misinformation within 24 hours, or they would face fines of up to €50 million (Oltermann 2018). However, firms have not always complied with those regulations in a prompt manner: A German government-funded survey found that Twitter erased only one in every hundred user messages that violated the regulation, and none of the deletions took place within 24 hours (Lomas 2017). In the United States, a social media platform, Parler, has gained its popularity amongst users who are banned from popular sites such as Twitter and Facebook due to their content moderation policies following the 2020 presidential election.[4] These examples[5] suggest that, if the implementation of a regulation creates a significant competitive disadvantage for a firm, the firm may have an incentive to minimize the impact of this regulation by delaying its compliance or not complying at all. How, then, does market structure affect such competitive incentives of a firm, and consequently the market-level compliance?

1. https://www.bbc.com/news/world-asia-china-33844084.

2. https://www.theatlantic.com/international/archive/2015/08/chinas-response-tianjin-explosions/401426/.

3. https://www.wsj.com/articles/BL-CJB-27514.

4. https://www.nytimes.com/2020/11/11/technology/parler-rumble-newsmax.html.

5. It is important to note that content moderation in those settings is related to censorship in terms of the similar trade-off it induces on platforms, albeit with notable differences when it comes to user welfare to be discussed in Section 2.

This paper studies these questions through a unique empirical example: censorship compliance via content removal by online platforms in China. This type of censorship targets user-generated content related to politically sensitive events based on a central order from the authoritarian government. Using a novel dataset on three major live-streaming platforms, I measure each platform's compliance behavior by examining when and how many keywords the platform has added to its own blacklist following a sequence of salient events. This dataset encompasses the comprehensive history of blacklisted keywords used by each of the three platforms over a 2-year period (2015–2017). If a user's message includes any blacklisted keywords in a chat on the platform, the message will either go undelivered or be replaced with asterisks, typically paired with a warning sign indicating a delivery failure. I exploit the unexpected occurrence of 30 sensitive events during the data collection period, such as the 2015 Tianjin Explosion mentioned above. Each salient event triggered the Chinese government's censorship request and surveillance, as well as the need for platforms to comply. By comparing the timing and frequency of platforms' blacklists update using an event study approach, I find that platforms of different sizes exhibit different compliance behavior: The largest platform not only blacklisted a higher number of keywords on average, it also complied faster than the smaller platforms during most events. Moreover, conditional on a platform's own decision of not censoring, its user traffic is positively correlated with its competitors' censorship decisions.

Motivated by this empirical pattern, I develop a structural model of oligopolistic competition to investigate the relationship between platforms' size, political pressure, and their compliance with censorship regulations. In this model, a platform's profit depends on the number of users and the legal cost of failing to comply with the censorship regulations. For each unexpected sensitive event, online platforms receive requests for censoring a set of keywords related to these events and decide simultaneously whether to comply immediately. Users incur heterogeneous disutility from being censored and may evade censorship by switching to another platform. On the one hand, platforms are pressured to remove certain user-generated content immediately. If they fail to do so, they may be subject to a fine and risk being temporarily shut down by the government (King, Pan, and Roberts 2013). The potential shutdown can be viewed as a radical form of legal or reputation costs for the platform. On the other hand, by strategically delaying censorship compliance, a platform could attract users—who are trying to evade censorship—from other platforms.

The model predicts that while large platforms comply more often than their small competitors due to higher legal costs, consolidating market power via merging or shutting down small platforms does not necessarily cause more censorship compliance in the marketplace. If a market hosts *fewer* platforms, two factors are at play: first, each platform captures a larger market share and bears higher legal costs of non-compliance; second, platforms have increased strategic incentives to differentiate from other obedient competitors by not complying immediately, now that users have fewer options to switch to. Following this change in the market structure, whether a platform is more or less likely to censor depends on which of the two forces dominates. If even a slight increase in a platform's size alarms the government and significantly increases the platform's legal cost of non-compliance, the former political pressure would dominate and generate more censorship in the marketplace. If, on the other hand, limiting the number of alternatives significantly increases a platform's chance to capture more switching users, then the latter strategic incentive would dominate and cause platforms to censor less often in equilibrium. To quantify the relative magnitude of these two forces and derive meaningful counterfactual predictions, I estimate the model by exploiting variations in platforms' market shares across different events and the correlation between censorship decisions of a platform and its competitors. My counterfactual analysis shows that shutting down a small platform with the goal of censoring users could backfire and lead to an unintended consequence where the overall compliance turns out to be lower.

Past research (Edmond 2013) suggests that an authoritarian regime's chances of survival decline with the number of information sources unless there are strong economies of scale in information control. For this very reason, authoritarian regimes, such as China and Russia, have always been heavy-handed in regulating private media outlets to preserve political power. However, in this digital age, where nearly half of the global population owns a social media account (Newberry 2019), blindly penalizing emerging platforms for non-compliance carries a significant cost. My findings suggest that decentralizing online market power could help an authoritarian government maintain sufficiently high market-level censorship: tolerating—or even encouraging—a bit of dissent on small platforms allows big platforms to censor more effectively as it mitigates their strategic incentives. This might be one of the reasons why, unlike in the United States, where the market is dominated by a handful of mainstream social media platforms, Chinese social media remains "very fragmented and localized (Chiu, Ip, and Silverman 2012)."

The rest of the paper is organized as follows. Section 2 discusses the related literature. Section 3 describes the data and institutional background. Section 4 presents the event-study analysis and results. Section 5 proposes the model and discusses a series of model predictions. Section 6 describes the estimation strategy and presents the estimation results. Section 7 provides policy-relevant counterfactual predictions. Section 8 concludes.

2. Literature Review

This paper contributes to multiple strands of literature that lie at the crossroads of economics, political science, and legal studies.

2.1. Media Capture and Bias

Most studies of media capture in non-democracies have examined the effectiveness of government's media strategy at the regime level (King, Pan, and Roberts 2013; Edmond 2013; Lorentzen 2014; Gehlbach and Sonin 2014; Qin, Strömberg, and Wu 2017; Hobbs and Roberts 2018; Roberts 2018; Chen and Yang 2019). For example, Chen and Yang (2019) show in a field experiment that China's censorship apparatus remains robust to a large number of citizens despite their access to uncensored internet due to a combination of low demand for uncensored information and moderate social transmission. Other recent studies have documented local media's lack of censorship due to their misaligned political (Zhuang 2022) and economic (Qin, Strömberg, and Wu 2018) incentives from the central leadership. For instance, Zhuang (2022) provides evidence that local governments in an autocracy use local media to further their own political goals through underreporting corruption scandals. In this paper, I propose a new mechanism where censorship could remain effective by leveraging market asymmetry *despite* misaligned incentives and increasing competition. This paper also adds to research on documenting the source of media bias, albeit from a different angle. Existing work has studied supply-side factors, such as advertising revenue (Beattie et al. 2021) and the preferences of media owners (Durante and Knight 2012; Yildirim, Gal-Or, and Geylani 2013), as well as demand-side factors, such as consumer ideologies (Gentzkow, Shapiro, and Sinkinson 2014; Knight and Tribin 2019; Simonov and Rao 2022) that lead to media's strategic choice of information disclosure. Using a novel dataset on the blacklists of entertainment-based social media platforms following a sequence of nation-wide political events, I present direct measures of supply-side censorship compliance and focus on the *timing* of content moderation, instead of the content itself.

2.2. Regulatory Compliance and Content Moderation

This paper also contributes to the existing literature on regulatory compliance of firms in a unique empirical setting. It is well acknowledged in the literature that larger firms are subject to significantly higher reputation costs and more "equipped" to react to regulation (Griffith 2015). My event-study result that bigger platforms comply faster is consistent with the literature in that they do face higher "legal costs" due to larger user base, although all focal platforms in my empirical setting employ similarly sophisticated combination of surveillance technologies as documented in Section 3.3. However, most existing research in this domain fails to recognize the role of market structure in shaping the incentives of firms' regulatory compliance. This paper, on the other hand, highlighted that the overall compliance could go up due to intensified competition, even if, on average, the reputational costs of firms become smaller.

In this context, censorship and content moderation emerge as two key strategies for regulating information dissemination. Both share common trade-offs for social media platforms, potentially driving at least *some* users away. For instance, social media giants, such as Twitter, Facebook, and YouTube, have long been criticized by some conservatives and celebrities for their content regulation policies "censoring free speech" (Hooker 2019; Barrett 2022; Oremus 2022). On the other hand, non-compliance of regulation might involve legal penalties such as fines and reputational costs and, in the extreme case of censorship, the threat of a temporary or permanent shutdown. Despite these similarities, censorship and content moderation diverge in their motivations, methods, and particularly on user welfare. Censorship, often wielded by governmental bodies or authoritarian regimes, involves the restriction of information deemed inappropriate or contrary to established norms or laws. Content moderation, on the other hand, is typically carried out by private entities such as social media platforms or online forums to enforce community guidelines and terms of service. While censorship focuses on top–down control and is often driven by political motives, content moderation generally aims to foster a positive user experience and mitigate risks, such as harassment, hate speech, misinformation, or other forms of harmful content. In that regard, platforms might have an interest in moderating content to protect their users aside from the legal pressure.

2.3. Platform Competition in Two-Sided Markets

This paper is related to extensive literature on platform competition in two-sided markets. In the seminal papers by Rochet and Tirole (2003, 2006), a platform is defined as an intermediary that enables interactions between two or multiple sides of end-users by appropriately charging each side. Most social media platforms are "audience makers," according to Evans (2003) that match advertisers (on one side) to audiences (on the other side) who consume content for free. This type of two-sided markets are characterized by strong indirect network effects and often asymmetric pricing (Caillaud and Jullien 2003; Crémer, De Montjoye, and Schweitzer 2019). While the focal platforms in this paper allow users to access live-streamed content and interact through comments, they monetize primarily through viewer interactions, such as virtual gift sending and tipping, rather than through advertising revenue (see Section 3 for detailed documentation). Their unique business model sets the focal platforms apart from other social media platforms in that their strategic incentives mainly depend on users' utility but not so much on advertisers' profits. This is contrary to, for example, the case of YouTube where advertisers might refuse to be associated with videos including sensitive content and thus "push" for quick compliance with regulation. While a detailed analysis is beyond the scope of this paper, the presence of advertising profits could weaken the larger platform's strategic incentive to delay compliance, since it would risk losing ad revenue. In contrast, it could strengthen the smaller platforms' incentive to delay, as they forgo such profits and benefit from the larger platform's quicker compliance in equilibrium. The overall scope of censorship may rise or fall, depending on the relative weights of user profit versus advertising revenue.

2.4. Discrete Choice Models

Lastly, this paper draws from and contributes to empirical work on the discrete choice models of firms' strategic decisions. The identification and estimation of discrete games have been studied in several structural contexts (Sweeting 2004, 2009; Bajari et al. 2006; Bramoullé, Djebbari, and Fortin 2009; De Paula and Tang 2012). Methodologically, the paper most related to mine is Wan and Xu (2014), where they propose an inference procedure for a static binary decision game of incomplete

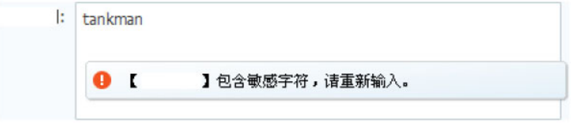


FIGURE 1. A screenshot of blacklisted keyword that triggers censorship on YY. The text following the warning sign is translated as "[The message you sent] contains sensitive words. Please try again."

information that allows for the correlation of private signals. Unlike Wan and Xu (2014), I develop a structural model that micro-founds the strategic interaction between firms by modeling the demand-side behavior explicitly. This paper shows how the strategic interaction term in the "reduced-form" profit function can be mapped from a set of structural parameters with important economic implications in my framework.

3. Data and Institutional Background

Section 3.1 describes each dataset at length. Section 3.2 introduces the institutional background of censorship in China. Section 3.3 provides an overview of the social media platforms studied in this paper.

3.1. Data

The primary source of data is provided by the Citizen Lab.[6] Researchers from the Citizen Lab decrypted the entire blacklists of keywords using several reverse engineering techniques (Knockel et al. 2015). By using tools for finding cryptographic constants inside a program's address space as it is running, they identified the files that stored blacklisted keywords and the URLs from which keyword updates were downloaded for the three most popular live-streaming platforms in China: YY, 9158, and Sinashow.[7] These built-in lists of keywords perform checks to determine whether any of these keywords are present in users' chat messages before the messages are sent. If a user's message contains a blacklisted keyword in a chat, his/her message will either be undelivered or replaced by asterisks, and generally accompanied with a warning sign such as "The message you sent contains restricted words. Please try again." Figure 1 provides an example of a censored message on YY that contains keyword "tankman," a reference to the Tiananmen Square Protest. At any time, a platform may update its blacklist as it deems necessary. Official reports show that automated keyword filtering

6. The Citizen Lab (https://citizenlab.ca) is an interdisciplinary laboratory based at the Munk School of Global Affairs, University of Toronto.

7. For example, Sinashow comes installed with a binary database of keywords in a file named "*Word_410.ucw*" and downloads updates for it from http://www.51uc.com/uc_interface/down_policy/Word_410.ucw. This file is a custom binary container storing sensitive GBK-encoded keywords that have been encrypted (Knockel et al. 2015).

TABLE 1. Content themes and related categories of blacklisted keywords.

Theme	Example categories	Example translations and original keywords
Event	Recurring events	“64 memorial” (悼念64)
	Unexpected events	“Tianjin nuclear explosion” (天津核爆炸)
Social	Gambling, illicit goods and services	“Crystal meth formula” (冰毒配方)
	Prurient interests	“Adult video” (成人视频)
Political	Communist Party of China	“Inner-party division” (党内分歧)
	Religious movements	“Dalai Lama” (达赖喇嘛)
	Ethnic groups	“East Turkistan Muslim” (东突穆斯林)
Technology	General technical terms	“Internet TV software” (网络电视软件)
	URLs	“app.box.com”, “freelibs.org”
	Applications and services	“VPN800”, “Encryption Router” (加密路由器)
People	Government officials	“Xi Jinping” (习近平)
		“Ruthless Xi”(包子心手辣)
	Dissidents	“Wuerkaixi” (吾尔开希)
Misc	Keywords with unclear contexts	“Too well” (太恩)

Notes: Based on the contextual information, three researchers from the Citizen Lab coded each blacklisted keyword into 1 of 80 content categories grouped under six general themes according to a codebook. This table listed some keywords (in parentheses) and their translations in some example categories from each of the six themes. In the Chinese language, words are concatenated together, whereas English words are separated by whitespace or punctuation. In other words, the original keywords on the blacklists convey information much the same way as English phrases do.

censored 90% of the “inappropriate messages” on interactive live-streaming platforms due to its efficiency and timeliness.[8]

To uncover the complete history of censorship within the 2-year study, the Citizen Lab performed an hourly download and decrypted all three platforms’ keyword blacklists between February 2015 and August 2017. The Citizen Lab used a combination of machine and human translation to translate the keywords to English and analyzed the context behind each one. Based on these translations and contextual information, three researchers coded each keyword into 1 of 80 content categories grouped under six general themes according to a codebook developed in Crandall et al. (2013). A total of 5,089 unique keywords[9] were identified and grouped from the blacklists of all three platforms. Table 1 provides some example categories of keywords grouped into each theme. The blacklisted keywords target a variety of content, including issues related to sensitive events, Chinese politics, circumvention tools, pornography, gambling, and illicit drugs. Notably, the Chinese government is most likely to surveil the platforms for compliance of censorship during critical times when information has the greatest impact, such as during elections, periods of civil unrest, and sensitive political anniversaries (Crandall et al. 2013). My study thus focuses only on the keywords grouped into the “event” category. While the keywords in other categories such as “social,” “political,” and “people” also convey contextual information on a platform’s censorship behavior, they are mostly generic terms that fail to be associated with any particular event and, hence, do not fit into the event-study framework. Keywords in the “event” category reference 49 unique events. I extracted

8. http://www.cac.gov.cn/2017-02/21.

9. including phone numbers and URLs.

the event dates from major news outlets and cross-checked them with Wikipedia entries, if available. The Online Appendix (Table I.1 and Table I.2) provides detailed descriptions and reference links for all the events included in this study. Based on the event dates, I separate these events into two distinct groups: unexpected and recurring events.[10]

(1) **Unexpected** events (e.g. South China Sea disputes): The event originally occurred after February 2015.

(2) **Recurring** events (e.g. Memorial Day of Tiananmen Protest): The event originally occurred before February 2015.

Among the 49 unique events, 30 of them are "unexpected" events and the remaining 19 are "recurring" events. In the empirical analysis, I will focus on the 30 "unexpected" events for two reasons: (1) as both users and platforms could anticipate the (anniversary) date of recurring events, chances are that users start engaging in event-related conversations even prior to the event date and thus platforms may need to add related keywords preemptively; (2) for some recurring events, related keywords remain in platforms' blacklists from previous calendar years. Even if these platforms do not add any new words during the anniversary, the pre-existing keywords would remain in effect for censoring. Therefore, the number of newly added keywords could be negatively correlated with the number of existing keywords during the event window for a recurring event and thus do not reflect accurately the extent of disruption users might experience.

If a platform added any keywords to its blacklist related to an event on a specific date, that particular event is defined to be "censored" by this platform on that calendar day. Specifically, I consider only the timing of a keyword first *added* to a platform's blacklist and disregard other subsequent operations such as deletion. This is for two reasons: first, in the data, there is only a small fraction ($<0.5\%$) of keywords that were added then removed *within* a month; second, adding keywords affects users' experience explicitly by warding off their messages, while the effect of deleting keywords is more subtle and most users are not even aware of this practice unless they try the same word *both* before and after it was removed. Table 2 presents the summary statistics of censored keywords associated with 30 unexpected events across three platforms.

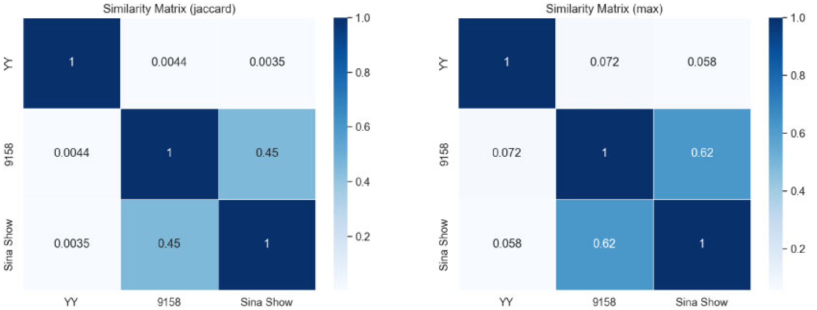
Figure 2 illustrates the overlap of blacklisted keywords across platforms in the event category. Following Knockel et al. (2015), in panel (a), I cluster the keywords by platform and compute their Jaccard similarity coefficient, that is, the size of the intersection of two sets divided by the size of their union ($J(x, y) = |x \cap y|/|x \cup y|$). In panel (b), I compute the keyword list x's similarity to list y as similarity(x, y) = max(% of x in y, % of y in x). Using either method, I find very little similarity between the keyword list of YY and that of the 9158 or Sina Show. On the other hand, I observe

10. Knockel et al. (2015) included a subset of the events studied in this paper. My classification rule for the events are slightly different from theirs. They separated the events into three categories: scheduled events, recurring events, and current events. The former two correspond to the "recurring events" and the last one corresponds to the "unexpected events" based on my definition.

TABLE 2. Summary statistics on platform censorship during unexpected events.

Platform	Total Number of Keywords Censored Per Event				
	Mean	Std. Dev	Min	Max	Obs.
YY	16.13	37.04	0	188	30
9158	1.80	3.48	0	15	30
Sina Show	2.20	3.82	0	15	30

Notes: The summary statistics have been calculated based on the number of keywords linked to each unexpected event that appeared on the blacklists of all three platforms. There are in total 30 unexpected events as defined above.



(a) Jaccard Similarity

(b) Max Similarity

FIGURE 2. Similarity heatmap of blacklisted keywords across platforms. Following Knockel et al. (2015), I present two heatmaps based on different similarity metrics to illustrate the overlap of blacklisted keywords across platforms in the "event" category. In panel (a), I compute the Jaccard similarity coefficient of two platforms' keyword lists, that is, the size of the intersection of two sets divided by the size of their union. In panel (b), I compute list x's similarity to list y as similarity(x, y) = max(% of x in y, % of y in x).

moderate overlap between the keyword lists of 9158 and Sina Show, especially based on the second similarity measure. These findings reveal that (1) there is heterogeneity in platforms' censored lists and (2) Sina Show (the smallest platform of all three) tends to censor a subset of (but not exact) keywords blocked by 9158 (the medium platform), possibly inherited from an internal list developed by their shared parent company.[11]

I developed a secondary dataset by scraping each platform's daily traffic from Siterankdata.com. This data source provides the history of websites' global traffic ranks reported by Alexa[12] over a 5-year window. Alexa.com is an American web traffic analysis company that provides web traffic data and global rankings "based on a combined measure of Unique Visitors and Pageviews" on 30 million websites.[13]

11. For a detailed discussion on the ownership information of these three firms, please refer to Section 3.3.

12. https://www.alexa.com.

13. https://support.alexa.com/hc/en-us/articles/200449744-How-are-Alexa-s-traffic-rankings-determined.

TABLE 3. Summary statistics on platform daily alexa ranks.

Platform	Daily Alexa Rank				Monthly Unique Visitors (est.)	
	Mean	Std. Dev	Min	Max	Mean	Std. Dev.
YY	7,937	3,117	3,838	16,324	11,692,000	1,315,395
9158	84,475	21,585	49,687	183,707	396,345	101,173
Sina Show	1,002,753	958,710	82,725	4,300,155	7,095	5,745

Notes: The summary statistics are calculated based on the daily Alexa ranking data between July 5, 2015 and August 15, 2017. The last two columns refer to the (estimated) monthly unique visitors of each platform from siterankdata.com based on their bi-monthly reported estimates. To minimize the measurement error, this paper only uses the daily Alexa rank as a proxy for the platform traffic or user size.

Specifically, the higher a platform's rank is (i.e. smaller rank number), the more popular it is on a global scale. In other words, the closer a site is to rank number 1, the more visitors it requires to improve its rank. In contrast, a small change in the number of visitors to a small site will result in a large change in its rank. Table 3 provides the summary statistics on daily Alexa ranks of the three platforms and their estimated daily unique visitors.

I use the log of the inverse of a platform's global ranking as a proxy for user traffic because studies have shown that the ranking of a website follows Zipf's law (Adamic 2000): The relationship between the number of visitors to a platform and its global rank by popularity is nearly linear on a log-log plot, with the slope being -1.

3.2. Institutional Background

The market structure of online platforms and the institutional environment of censorship in China offer a particularly suitable setting to study the question of interest in this paper for two main reasons. First, instead of passively executing orders of "Big Brother," social media companies in China use their discretion to determine when and what specific content to block, despite the legal and regulatory pressure from the government (Knockel et al. 2015). In fact, "domestic censorship in China is deeply fragmented and decentralized" (Bamman, O'Connor, and Smith 2012). The Chinese government directs companies to censor their own content according to a list of vague guidelines (Stern and Hassid 2012). As a result, private companies are held accountable for any content published on their own platforms: The fact that a story has already been published elsewhere, and therefore presumably approved by the authorities, provides no legal cover (Initiative 2005; Human Rights Watch 2006; King, Pan, and Roberts 2013). According to one leaked log of company decisions, Sina Weibo, one of China's most popular social media sites, instructs employees to "not be stricter than Tencent," its main competitor, and to hold off on implementing government directives until "urged to block content a second time" (Miller 2018). Second, online platforms in China are subject to a high frequency of censorship requests from the central government due to the frequent occurrence of unexpected politically sensitive events. This makes it empirically possible to study platforms' interactions and detect potential strategic responses.

3.3. Interactive Live-Streaming Market

The "interactive live-streaming" or "zhibodashang" (直播打赏) market refers to a digital entertainment industry characterized by platforms that enable users to broadcast live video content in real-time while also providing interactive features that allow viewers to engage with the content creators. These platforms enable viewers to reward content contributors with virtual gifts that can be purchased with real money (Xiang 2016). Typically, the virtual gifts take the form of visually appealing items that helps the audience stand out in a live chat and fosters their engagement and interaction with creators in real-time. Platforms, in turn, take a cut of each virtual item. This market initially emerged in China in the late 2000s and has evolved into a lucrative business model. The key features of this market set it apart from other chat and social media apps, as it focuses on live content creation and monetization through viewer interactions such as virtual gift sending and tipping (Medium 2016).

This paper studies three social media platforms that operate in this interactive live-streaming market: YY, 9158, and Sinashow. YY, developed by YY Inc and registered in Guangzhou, is the largest live-broadcasting platform in China in terms of its user population during the sample period. Initially known for creating an online gaming portal and a voice messaging service for gamers, YY made its initial foray into virtual gifts through live-streaming performances by amateur singers. For example, users on YY exchange "virtual roses" as a form of currency, with top users said to spend as much as $20,000 per month (Geron 2012). Virtual gifts pulled in about 1.5 billion Chinese yuan ($230 million USD) for YY in the fourth quarter of 2015 (Medium 2016). In contrast to the sales of virtual gifts, according to YY's annual financial statement, online advertizing and promotion accounted for only 1.1% and 0.5%, of YY's total net revenues in 2015 and 2016, respectively (YY Inc. 2016). Tian Ge Interactive Holdings Limited, registered in Hangzhou, owns two other platforms: 9158 and Sina Show. Tian Ge's 9158 is an early adopter of the virtual gift model, hosting amateur singers and pioneering this concept. As many as 6 billion virtual gifts (of over 910 types developed) were exchanged monthly on 9158.com, according to its 2015 corporate report (Tian Ge Interactive Holdings Limited 2015). Notably, Tian Ge generates revenues "primarily through sales of virtual currency to its third-party distributors on the live social video platforms," with advertising ranking as only its fifth-largest source of income, following revenue from gaming, e-commerce, and financial investment (Tian Ge Interactive Holdings Limited 2016). Not only did YY and Tian Ge list each other as close competitors on their corporate presentations and websites for investors, but they were also both market leaders: by revenue, YY owns a share of 40% and Tian Ge 28% in 2015, respectively (YY Inc. 2015). According to a 2016 industry report published by iResearch, a leading Chinese consulting company, YY and 9158 are the top two "leading platforms in the interactive livestreaming market" with "more than 10 million monthly active users (MAU)" (iResearch 2016, p20-24). While Sina Show had a smaller market share ($\approx 8.1\%$) compared to the other two platforms, it held the highest user rating (9.4) among the most popular live-broadcasting apps in 2016, followed by YY (9.2) and 9158 (8.8), based on another widely circulated industry report (Jiemian.com 2016). Because these platforms rely on active daily user engagement for revenue, censorship that disrupts user activity imposes significant costs. Figure A.1 in the Online Appendix shows a screenshot of the user interface on one of the platforms examined in this study.

Furthermore, all three platforms manage content using a similar mix of surveillance technologies. For example, YY employs a system that periodically "sweeps" the platform for offensive content, supported by a "24-hour surveillance" data security team. Tian Ge, in its public filings, describes a detection system that randomly captures and processes screenshots of video chat rooms every 1–3 minutes (Knockel et al. 2015). While both Sina Show and 9158 operate under the same parent company (Tian Ge), which developed the detection technology, they are managed by separate subsidiaries with distinct headquarters and administrative teams: 9158 was founded in 2008 by Jinhua Jiuyuewoba Network Technology Co., Ltd., whereas Sina Show spun off from Sina UC and registered under Xinxiu Dongli Culture Media Co., Ltd. in 2010.[14] As shown in Figure 2, the two platforms share only moderate overlap in their blacklisted keywords, suggesting that they do not rely on an identical censorship list. Online Appendix F explores an alternative scenario in which Sina Show and 9158 coordinate censorship decisions to maximize joint profits under common ownership. The main results remain qualitatively robust under this alternative.

To simplify the notation, I will refer to YY, 9158, and Sina Show as the big, medium, and small platforms, respectively, in all subsequent sections.

4. Event Study Analysis

In general, a platform's decision to censor content can be influenced by various internal and external factors, such as a surge in user activity (which increases the likelihood of encountering sensitive content), the preferences of the platform owner, or requests from local governments. It is thus difficult to identify the impact of censorship on platforms' traffic due to endogeneity issues: oftentimes a platform's censorship decision is triggered by its own abnormal traffic. However, an outbreak of nation-wide events, especially the "unexpected" ones, can be considered as exogenous shocks that trigger a platform's self-censorship. Therefore, I adopt an event-study approach to analyze how a platform's censorship intensity, measured by the number of blacklisted keywords associated with an event, dynamically changes as the event unfolds. Due to the reasons mentioned in Section 3.1, I will focus only on the "unexpected events."

The goal of this event study is to establish three facts. First, the big platform censored on average more keywords and also reacted faster than the medium and small

14. https://www.qcc.com.

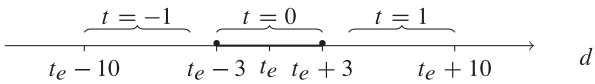


FIGURE 3. An illustration on how the event-time dummy variables are constructed.

platforms. Second, all three platforms have experienced a decline in the post-event traffic on average, but this decline is insignificant for the big platform and it happened to platforms of different sizes at varying times and magnitude. Third, conditional on a platform's own decision of not censoring, its user traffic is positively correlated with its competitors' censorship decisions. In particular, the last result is significant for small and medium platforms.

4.1. Platform Censorship Behavior

In order to understand the censorship behavior of platforms following a sequence of unexpected events, I consider the following econometric model:

$$R^i_{ed} = \sum_t \beta^i_t D^t_{ed} + \gamma_y + \varphi_q + \chi_e + \varepsilon^i_{ed}, \tag{1}$$

where R^i_{ed} denotes the number of keywords associated with event e censored by platform $i(= Big, Medium, Small)$ on calendar day d, γ_y is a year fixed effect, φ_q is a seasonal (quarter) fixed effect, and χ_e is an event fixed effect. The D^t_{ed} are a series of "event-time" dummies that equal 1 when the calendar day d is within t weeks of event e. Therefore, β^i_t represents the time trend of platform i's censorship intensity (i.e. number of keywords censored), conditional on seasonal and event fixed effects. Formally, the event-time dummy variables are given by

$$D^t_{ed} = \mathbb{1}\left[t = \lfloor \frac{d - t_e + 3}{7} \rfloor\right],$$

where $\mathbb{1}[\cdot]$ denotes an indicator function that equals 1 if the condition inside the brackets is true, and 0 otherwise; t_e is the date when an event $e \in \{E_1, E_2, \ldots, E_{30}\}$ occurred. Figure 3 provides a visual representation of how the event-time dummy variables are constructed relative to the calendar dates.

It is well-known that not all βs can be identified in this case because the dummy variables are perfectly collinear in the presence of event fixed effects. Hence, I normalize $\beta_{-1} = 0$. All post-event coefficients can be considered as treatment effects relative to 1 week before the event occurs. I also impose endpoint restrictions on β^i_t:

$$\beta^i_t = \begin{cases} \bar{\beta} & \text{if } t \geq 8 \\ \underline{\beta} & \text{if } t \leq -8. \end{cases}$$

This assumes that any dynamics wears off after 8 weeks. Because the sample is unbalanced in event time, these endpoint coefficients give unequal weight to events that happen very early or very late in the sample. For this reason, I restrict the sample to

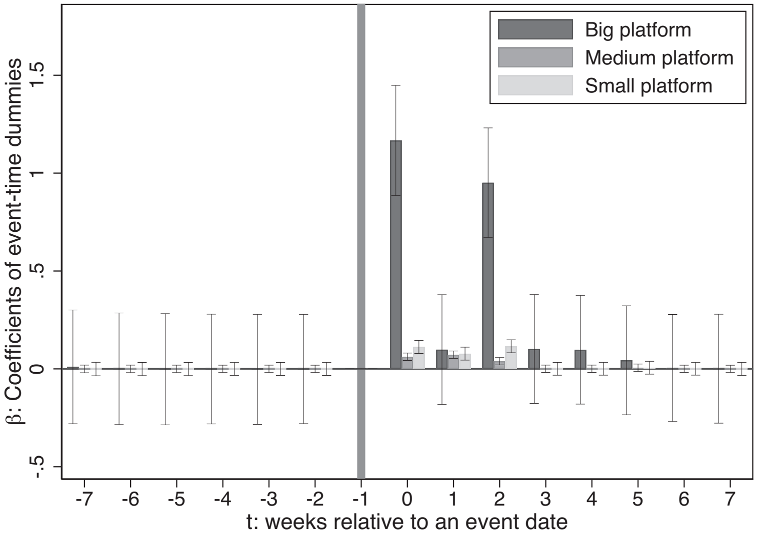


FIGURE 4. Coefficient plot of event-time dummies in regression (1). This figure plots the estimated coefficients (β_t^i, $-7 \leq t \leq 7$) of the event-time dummies from separate regressions for all three platforms ($i = Big, Medium, Small$).

a 7-week window prior to and after each event. Figure 4 plots the estimated censorship coefficients (β_t^i) from separate regressions for big, medium, and small platforms, respectively.

From Figure 4, we see that no event-related keywords were added prior to an event, corroborating the fact that those events were "unexpected." It also shows that the big platform censored on average more and most keywords within the event week, whereas the medium and small platforms tended to delay censoring some of their keywords by 1–2 weeks relative to the big platform.

4.2. Platform Traffic

In order to study the time trend of platforms' traffic in the aftermath of these events, I consider a similar econometric model as in the last section except that the dependent variable X_{ed}^i now denotes the log of user size (i.e. log (1/Alexa rank)) of platform i on calendar day d. Specifically, I consider the following econometric model:

$$X_{ed}^i = \sum_t b_t^i D_{ed}^t + \gamma_y + \varphi_q + \chi_e + \varepsilon_{ed}^i. \quad (2)$$

If events are indeed "unexpected" to a platform (and their users), the following null hypothesis should be true:

$$H_0 : b_t^i = 0, \forall t < -1, i = Big, Medium, Small.$$

TABLE 4. Joint test statistics of pre-event coefficients.

Platform i	Null Hypothesis	F statistics	P-value
Big	$b^B_{-2} = \ldots = b^B_{-7} = 0$	0.938	0.466
Medium	$b^M_{-2} = \ldots = b^M_{-7} = 0$	1.442	0.194
Small	$b^S_{-2} = \ldots = b^S_{-7} = 0$	0.503	0.806

In other words, the above condition requires that there are no platform-specific traffic trends preceding these events. I test this hypothesis by constructing F-statistics with robust standard errors clustered at the event level.

Table 4 reports the test statistics for the joint significance of pre-event coefficients for each platform. All three F-statistics fail to reject the null hypothesis, suggesting that there is no pre-event trend in traffic. This result shows that the event classification scheme described in Section 3.1 produces relatively accurate dates and supports the nature of unexpected events. Figure 5 plot the estimated traffic coefficients (b^i_t) from regression (2) for the (a) big, (b) medium, and (c) small platforms, respectively. Specifically, the left three panels plot the traffic trend using regression (2) across all 30 unexpected events (i.e. $e \in \{1, \ldots, 30\}$); The right three panels plot the traffic trend using regression (2) conditional on only events that were not censored by the platform but censored by at least one of its competitors within the event week (i.e. $e \in \{e' | R^i_{e'd} = 0, \forall d \in [t_e - 3, t_e + 3]$ and $\exists d \in [t_e - 3, t_e + 3] \quad s.t. \sum_{j \neq i} R^j_{e'd} > 0\}$).[15]

The left three panels (a, c, and e) in Figure 5 show that all three platforms have experienced a significant decline in the post-event traffic on average, but this decline happened to platforms of different sizes at varying times and magnitude. In particular, the medium platform's traffic declined immediately following the event ($b^M_0 = -0.044$, $p < 0.05$), followed by the small platform whose traffic began to show a decline on average 2 weeks after the event ($b^S_2 = -0.168$, $p < 0.05$). In contrast, the big platform's traffic declined on week 4, but the decline was significant compared to that of its competitors ($b^B_4 = -0.036$, $p = 0.20$). On the other hand, the big and medium platforms lost no more than 10% of their users during the post-event window, whereas the small platforms lost more than 20% of its users 2 weeks after an event occurred. This drastic change in the traffic of small platforms is likely caused by the fact that even a small change in the number of visitors to a small site usually leads to a large change in its Alexa rank. The right three panels (b, d, and f) in Figure 5 reveal a reverse traffic trend compared to the left three panels: There are significant upward post-event traffic trends on the small[16] and medium platforms and, in general,

15. Among the 30 unexpected events, there are, respectively, 3, 10, and 10 such events for the big, medium, and small platforms.

16. There is, however, a delayed growth in the small platform's traffic following the big platform's censorship. To explore the reasons behind this delay, I looked into another data source, App Annie, that recorded the daily app download rank of the small platform in 2016. Detailed analysis is provided in Online Appendix G on page 23. Results from the analysis suggest that it may be caused by users' behavioral intertia in the first 2 weeks. The initial response to censorship might be attempts to circumvent it rather

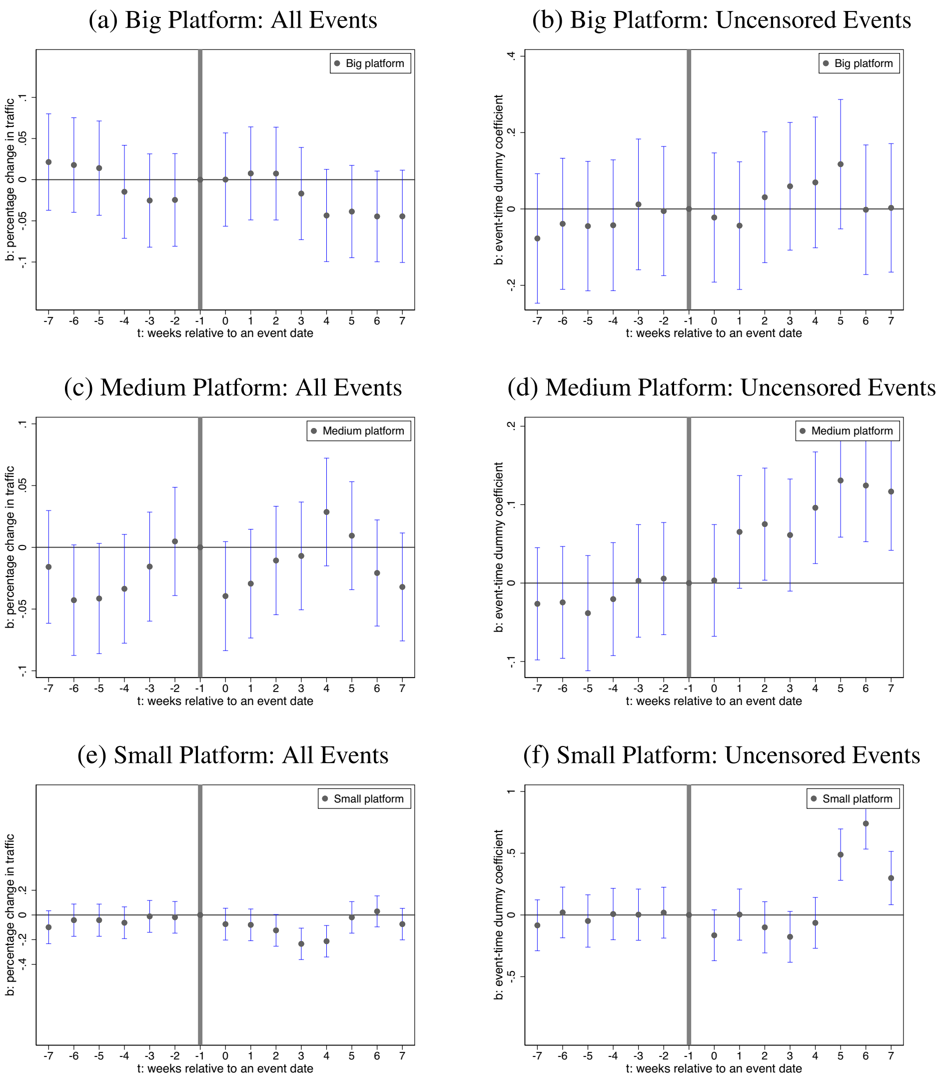


FIGURE 5. Coefficient plots of event-time dummies in regression (2). The left three panels (a, c, and e) in this figure plot the estimated coefficients (b_t^i) of the event-time dummies in regression (2) for each platform ($i = Big, Medium, Small$), respectively across all 30 unexpected events. The right three panels (b, d, and f) plot the estimated coefficients of event-time dummies in regression (2) across events that were not censored by the platform but were censored by at least one of its competitors within the event week. Since the dependent variable is a proxy of the log of user size, the coefficient of each event-time dummy represents the percentage change in platform traffic in that specific week relative to 1 week before the event date.

less decline in traffic on all three platforms. This last result suggests that conditional on a platform's own decision of not censoring, its traffic is positively correlated with its competitors' censorship decisions.

5. A Structural Model of Strategic Censorship

The event-study analysis shows that platforms of different sizes censor at different times and intensities and their traffic is responsive to the censorship activity. In this section, I explicitly model how users on each platform may stay or switch to another platform conditional on their preferences and allow platforms to strategically exploit users' switching patterns, anticipating their competitors' censorship decisions.

5.1. Platform Payoffs

Consider a market with $N \geq 2$ platforms. Each platform is indexed by $i(= 1, 2, .., N)$ and let $-i$ denote the set of i's opponents. Every time an unexpected event occurs, I consider platforms playing an independent[17] game, where each platform simultaneously makes a decision about whether or not to add keywords in week 0, that is, within the week of event occurrence. The simultaneous-move assumption is intended to capture the extent of each platform's compliance with the government's censorship request within the event week (i.e., ± 3 days of event date). This structure reflects a distortion in the decisions when platforms delay compliance, where they could have acted immediately. This assumption is consistent as well with the fact that each platform's blacklist is carefully encrypted and considered a corporate secret and hence *not* observed by other platforms (Knockel et al. 2015). Although it is not entirely impossible for platforms to hack each other's keyword lists occasionally, the assumption of incomplete information is reasonably consistent with reality for at least two reasons. First, Chinese neitizens often use euphemisms to bypass censorship, making the exact blocked keywords less obvious, even if they pertain to the same event. Figure 2 shows that the Jaccard similarity between large and medium/small platforms is less than 0.01, indicating very limited overlap between their keyword blacklists, including those keywords added with a delay. Second, it is technically challenging to decode the encoded keyword blacklist, as documented in Knockel et al. (2015). Additionally, private firms frequently change the URLs for download updates to ensure security, making even the encoded versions difficult to trace.

than seeking alternatives. Only after realizing the difficulty or ineffectiveness of circumvention efforts might users start exploring other options, contributing to the delayed growth in the small platform's traffic.

17. In Online Appendix H, I have employed a non-parametric runs test (Edgington 1961) to explore whether the platforms' censorship decisions are serially correlated over time. While the big platform's decisions exhibit weak evidence of positive serial correlation ($P = 0.04$), the results for the medium and small platforms are not statistically significant, suggesting no strong evidence against the null hypothesis of serial independence.

Hence, platform i has a decision variable $a_i \in \{0, 1\}$ denoting its choice to censor (1) or not (0).[18] Platforms gain utility from the mass of users they would retain after an event, where the dollar value of each user is normalized to 1. Platforms are assumed to be horizontally differentiated *outside* the event window due to their differentiated user-generated content with some of the most popular live performers signing exclusive contracts with only one platform (Medium 2016). The horizontal features of each platform are reflected in the utility function of captive users who obtain a higher match value from using their "favorite" platform compared to using other platforms prior to an event. I assume that the cost of switching back is negligible so switching users will always return to their favorite platforms before the next event occurs. They are thus referred to as the "captive users." In other words, switching is temporary because users typically switch out either to seek out further information or to engage in time-sensitive discussions about an event. As an event becomes less news-worthy over time, captive users tend to return to their favorite platforms for entertainment activities.[19]

Let platform i's payoff (π_i) from censoring and not censoring be:

$$\pi_i = \begin{cases} D_i(a, x) & a_i = 1 \\ D_i(a, x) - (c_0 + c_1 x_i + \varepsilon_i) & a_i = 0, \end{cases}$$

where $D_i(a, x)$ is the remaining mass of users on platform i conditional on the censorship decisions of each platform and their competitors. The terms in the parentheses $(c_0 + c_1 x_i + \varepsilon_i)$ reflect the legal cost[20] of not censoring an event, which is assumed to be a linear function of x_i, the mass of "captive users" of platform i. c_0 can be viewed as a fixed penalty from non-compliance, while c_1 is the expected loss per user from a temporary shut down by the government due to non-compliance. ε_i is the private signal observed only by platform i. However, private signals are allowed to be correlated across platforms and the joint distribution of these private signals are assumed to be known by all platforms. The correlation between private signals captures the fact that some unobserved event-specific variables may affect all platforms' censorship decisions. This assumption is also motivated by the anecdote that social media companies in China receive private information of censorship guidelines

18. Unlike in the event study analysis in Section 4, I focus on the extensive, instead of intensive, margin of platforms' censorship decisions in the structural model and thus reduce their choice variable to a binary one for two important reasons. First, this approach significantly simplifies the strategy space and eases the complexity of the solution concept. Second, censoring or not has a straightforward impact on platforms' users, whereas the impact of censoring X keywords as opposed to $X + 1$ keywords on users is much less so. This is because some keywords may be used more frequently than others and thus blacklisting a small number of frequently used keywords could affect even more users than blacklisting a large number of less common words. In other words, varying the number of censored keywords could affect the platform's traffic in a non-monotonic fashion.

19. This behavioral assumption also ensures the static nature of the game, which is essential for estimation given the limited number of events.

20. Note that this "legal cost" function reflects both the legal pressure toward platforms of various sizes and the potential technology difference of compliance. However, the latter is less likely the main driver for size-dependent differences given that the companies of interest use a similar mix of surveillance technologies, as documented in Section 3.3.

from different “resources distributed across several bureaucracies” (Miller 2018). These guidelines, to the platforms’ awareness, possibly originates from the same central directive.[21] As a result, platforms exploit both their private information *and* the correlation between private information to form a rational expectation of their rivals’ choices.

5.2. *User Switching Behavior*

There is a mass of 1 users in the market. Each platform i is the favorite of a share of x_i users such that $\sum_{i=1}^{N} x_i \leq 1$.[22] Before events occur, users are assumed to single-home[23] and be active only on their favorite platform. After an event occurs, users obtain some disutility from being censored calibrated by an individual taste parameter θ. On all platforms, assume that θ is drawn from a Pareto distribution[24] with a shape parameter denoted by $\alpha(> 0)$ and its location parameter normalized to 1. Let $F(\cdot)$ denote the cumulative distribution function of variable θ:

$$F(\theta > \underline{\theta}) = \begin{cases} \left(\frac{1}{\underline{\theta}}\right)^{\alpha} & \underline{\theta} \geq 1 \\ 1 & \underline{\theta} < 1. \end{cases}$$

Upon observing the censorship actions of their own platform, users decide to switch ($s = 1$) or not to switch ($s = 0$) to other platforms. If they choose to stay when the platform censors, depending on their distaste for censorship, they will have to endure some degree of disutility. Alternatively, if they choose to switch, an expected fixed cost of switching $\gamma(\geq 1)$ is incurred. Prior to switching, users do not know whether other

21. The potential influence of local governments is assumed away in this paper for two important reasons. First and foremost, the sensitive events studied in this paper are high profile on a national scale, and, hence, social media companies are more likely to be pressured directly or indirectly by the central government whose directives they ought to prioritize. Second, the event study results documented in Section 4 suggest that interferences from the local government, if there are any, seem to be secondary and outweighed by influences from outside the local jurisdictions where the companies are headquartered. Specifically, YY, the largest platform, is registered in Guangzhou, while 9158 and Sina Show, the medium and small platforms, are headquartered in Zhejiang and Beijing, respectively. Guangzhou, a coastal city close to Hong Kong, is one of the most liberal regions (second only to Shanghai) in mainland China (Wong 2015; Pan and Xu 2018). However, the largest platform, headquartered in Guangzhou, is the one that censors the most among the three platforms, implying under more influence from outside its jurisdiction.

22. Note that the total mass of platform users is weakly less than 1 but does not need to sum to 1, allowing for greater flexibility in the identification of parameters detailed in Section 6.1.

23. According to Qianfan Analysis, one of the leading analytics platform in China, there were 15,540,200 monthly active users on YY (big platform) and 563,000 on 9158 (medium platform) in 2016, respectively . The multi-homing users on both platforms were estimated at only 51,100, which is less than 10% of the active users on the medium platform. In addition, even for users with multiple accounts, they mainly used one app at a time from each product category (e.g. live-streaming, messaging, etc.). Hence, changes in traffic on a particular platform could also be interpreted as its multi-homing users consuming content on another affiliated platform, not necessarily suggestive of a change in the total number of users overall.

24. This distributional assumption on θ is innocuous. Any distribution that satisfies $F(\theta > 0) = 1$ and $F(\theta > \gamma) > 0$ (e.g. log-normal) will deliver a demand function that is linear in the binary actions (i.e., $a \in \{0, 1\}$) of a firm. For model tractability, I assume that θ is drawn from a Pareto distribution as it is scale-free and associated with a simple CDF (Cumulative Distribution Function).

platforms are implementing censorship until they switch and start using them. In other words, users do *not* observe the censoring decisions of other platforms at the time of decision-making. This is because they would have to enter the exact same keywords on another platform in order to know whether they are censored or not (see Section 3 for detailed discussions). Thus, I assume that users who have switched from platform i will end up using one of its competitor platforms $j(\neq i)$ with equal probability. [25] If switching users find out that the new platform also censors their messages, they will leave the new platform immediately for their outside options, such as watching TV, listening to music, and so on.[26] Formally, a user with taste parameter θ from platform i chooses $s \in \{0, 1\}$ to maximize his/her utility conditional on platforms' censorship actions:

$$\max_{s\in\{0,1\}} u_i(s; \theta) = v_0 - (1 - s)\theta a_i - s\gamma, \tag{3}$$

where $v_0(> \gamma)$ denotes the reserved utility of captive users using the platform and the utility of their outside options is normalized to 0. A user θ on platform i prefers to switch if and only if $u_i(s = 1; \theta) \geq u_i(s = 0; \theta)$:

$$\theta a_i \geq \gamma. \tag{4}$$

Let $\underline{\theta_i}(a)$ denote the threshold user on platform i who is indifferent between switching and not switching conditional on platforms' actions. Any user with θ above this threshold $\underline{\theta_i}(a)$ will switch or otherwise will stay. Note that if their favorite platform does not censor, no user will switch out as switching is costly (i.e., $v_0 - \gamma > 0 \Rightarrow \underline{\theta_i}(a_i = 0) \to +\infty$). If the platform does censor, however, only a fraction of users would switch out while the remaining users will stay despite being censored (i.e., $\underline{\theta_i}(a_i = 1) = \gamma \in [1, +\infty)$). Therefore, for any platform i, the share of "switchers" is given by

$$F(\theta \geq \underline{\theta_i}(a)) = \left(\frac{1}{\underline{\theta_i}(a)}\right)^{\alpha} = \begin{cases} \left(\frac{1}{\gamma}\right)^{\alpha} & a_i = 1 \\ 0 & a_i = 0. \end{cases} \tag{5}$$

Taking log of both sides when $a_i = 1$ in equation (5), we can rewrite it as follows:

$$log[F(\theta \geq \underline{\theta_i}(a_i = 1))] = -\alpha log(\gamma). \tag{6}$$

25. This behavioral assumption greatly simplifies the equilibrium solution and estimation of this structural model. It is also consistent with the institutional background where a platform owner has significant information advantage over the platform users in the sense that the owner observes its own event-specific private signal (ε) and knows the distribution of competitor platforms' signals whereas the users do not. Therefore, without knowledge of the private signals—such as the urgency of the event requiring censorship per government request—users cannot update their beliefs about other platforms' censorship decisions based on observing their own platform's actions. Discussions on how the main predictions are affected by incorporating consumer expectation into the model are provided in the Conclusion section.

26. I refer to these "outside options" as any form of leisure activities that users may adopt after they give up seeking information or discussing about the sensitive events online.

Based on the above equation, α can be viewed as the "switching elasticity" that characterizes how responsive users are to changes in the switching cost. All else equal, the larger is the switching cost γ, the larger is the fraction of users that will always stay regardless of the censorship activity. In sum, the mass of remaining users when platform i chooses a_i and its opponents choose a_{-i} is given by

$$D_i(a,x) = x_i - \underbrace{x_i F(\theta \geq \underline{\theta_i}(a))}_{\text{outgoing users}} + \underbrace{(1-a_i)\sum_{j\neq i}\frac{x_j}{N-1}F(\theta \geq \underline{\theta_j}(a))}_{\text{incoming users}}, \tag{7}$$

which depends on the relative size of outgoing and incoming users, affected by both i and its competitors' actions. Plugging equation (5) into equation (7), we obtain the following demand function:

$$D_i(a,x) = \begin{cases} x_i - x_i\left(\frac{1}{\gamma}\right)^{\alpha} & a_i = 1 \\ x_i + \sum_{j\neq i}\frac{x_j}{N-1}\left(\frac{1}{\gamma}\right)^{\alpha} a_j & a_i = 0. \end{cases}$$

Specifically, given the private signal ε_i and its rivals' choices $\{a_j\}_{j\neq i}$, platform i's (realized) payoff when censoring relative to not censoring can be expressed as follows:

$$\pi_i(a_i = 1, a_{-i}, x) - \pi_i(a_i = 0, a_{-i}, x) = \beta_0 + x_i\beta_1 + \sum_{j\neq i}(\delta x_j)a_j + \varepsilon_i, \tag{8}$$

where

$$\beta_0 = c_0, \tag{9}$$

$$\beta_1 = c_1 - \gamma^{-\alpha}, \tag{10}$$

$$\delta = -\frac{1}{(N-1)(\gamma)^{\alpha}}. \tag{11}$$

Since platform i will choose $a_i = 1$ if and only if equation (8) is positive, we can rewrite the profit function of platform i into the following relative form:

$$v_i(x) = \begin{cases} \beta_0 + \beta_1 x_i + \sum_{j\neq i}(\delta x_j)a_j + \varepsilon_i & \text{if } a_i = 1 \\ 0 & \text{if } a_i = 0. \end{cases} \tag{12}$$

Note that equation (12) corresponds to the reduced-form profit function widely adopted in the literature on social interactions (e.g. Bajari et al. 2006; De Paula and Tang 2012; Aradillas-Lopez and Gandhi (2013)), and (δx_j) is referred to as the "strategic interaction" term. In my framework, this term maps to a function of platforms' user size scaled by some structural parameters. Since $\delta x_j < 0$ for any given state variable $x \in \Omega_X$, a_j is clearly a strategic substitute for a_i. Furthermore, conditional on the size distribution of platforms, the magnitude of strategic interaction is larger when users' switching cost (γ) is smaller or the "switching elasticity" (α) is larger. This is because larger values of these two parameters imply a higher likelihood that a greater share of users will switch to a non-censoring competitor, making a platform's censorship

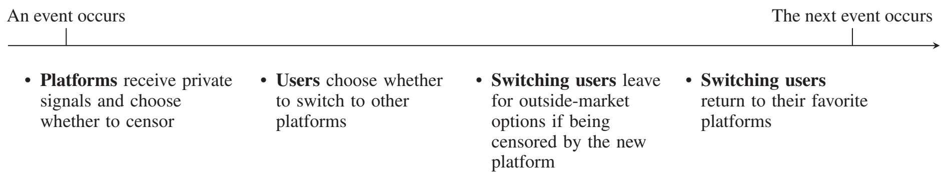


FIGURE 6. Timing of the game.

decision more sensitive to its rivals' actions. Figure 6 summarizes the timing of the game.

5.3. Bayesian Nash Equilibrium Strategies

Without loss of generality, this paper assumes that the outcome observed is the result of a pure strategy[27] Bayesian Nash Equilibrium (BNE). In this game, each platform ($i = 1, 2, \ldots, N$) simultaneously chooses $a_i \in \{0, 1\}$. A state of the game is described by (X, ε), where $X = (X_1, X_2, \ldots, X_N)$ and $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N)$. $X_i \in (0, 1]$ is platform i's user share prior to an event that is publicly observed. $\varepsilon_i \in \mathbb{R}$ is the private signal observed only by platform i. Let $F_{\varepsilon|X}$ be the conditional distribution function of ε given X. Let $\Omega_X \equiv \{(X_1, X_2, \ldots, X_N)| \sum_{i=1}^{N} X_i \leq 1\}$ denote the support of X. In equilibrium, platform i chooses action 1 if and only if its expected payoff is greater than if it chooses action 0, that is,

$$a_i = \mathbb{1}\left[\beta_0 + x_i\beta_1 + \sum_{j \neq i}(\delta x_j) \cdot \mathbb{P}(a_j = 1|X, \varepsilon_i) + \varepsilon_i \geq 0\right], \tag{13}$$

where $\mathbb{1}[\cdot]$ is the indicator function. The term $\mathbb{P}(a_j = 1|X, \varepsilon_i)$ is platform i's expectation on its rival j's action, based on player i's private information and the publicly observed state variable. Equation (13) defines a set of simultaneous equations. Following Athey (2001) and Wan and Xu (2014), this paper adopts a particular class of BNEs referred to as "monotone pure strategies" (MPSEs). With an MPSE[28] in this framework, there exists a sequence of cutoffs $\varepsilon^* = (\varepsilon_1^*, \varepsilon_2^*, \ldots, \varepsilon_N^*) : \Omega_X \to \mathbb{R}^N$ such that for each player i,

$$a_i = \mathbb{1}\left[\varepsilon_i \geq \varepsilon_i^*(X)\right]. \tag{14}$$

That is, if the private shock (i.e. the private cost of not censoring a particular event) is sufficiently large, platform i will choose to censor, that is, $a_i = 1$. Athey (2001) proves

27. It is well known that Harsanyi's purification theorem has endorsed the empirical appeal of pure strategy equilibria. Harsanyi (1973) has shown that the existence of private information in payoffs will induce players using pure strategies approximately with the prescribed probabilities associated with mixed-strategy equilibria of a complete information game.

28. Note that if private signals (ε_i) are assumed to be independent, all BNE solutions to this game are MPSEs.

that an MPSE exists whenever a Bayesian game obeys a Spence-Mirrlees single-crossing condition. Hence, in this model, I consider the following assumptions made to ensure the single-crossing condition.

ASSUMPTION 1 (*Bounded positive regression dependence*). The conditional pdf $f_{\varepsilon|X}$ exists and is assumed to be common knowledge. For any $i \in \{1, 2, \ldots, N\}$ and $(t, x) \in \mathbb{R} \times \Omega_X$,

$$0 \leq \frac{\partial \mathbb{P}(\varepsilon_{-i} \geq t | X = x, \varepsilon_i)}{\partial \varepsilon_i} \leq (N-1)(\gamma)^{\alpha} - \left| \frac{\partial \mathbb{P}(\varepsilon_{-i} \geq t | X = x, \varepsilon_i)}{\partial t} \right| .$$

Assumption 1 implies that platform i's best response is non-decreasing in its private signal when its rivals also adopt a monotone strategy. In a parametric case, where ε conforms to a joint normal distribution with mean zero and unit variance, it suffices that the correlation coefficient between ε_i and ε_j is bounded above, vis-a-vis the magnitude of $(N-1)(\gamma)^{\alpha}$.

THEOREM 1. *If Assumption* 1 *holds, then for any public state variable $x \in \Omega_X$, there exists a unique MPSE where each player's equilibrium strategy is non-decreasing.*

Proof. See Appendix. □

Given that the equilibrium is monotone conditional on $X = x$, platform i receives zero expected payoff when the value of its private shock equals $\varepsilon_i^*(x)$, that is,

$$\varepsilon_i^*(x) = - \underbrace{\left[c_0 + (c_1 - (\gamma)^{-\alpha})x_i\right]}_{\text{legal cost (net of economic gains)}} + \underbrace{\sum_{j \neq i} \frac{x_j}{(N-1)(\gamma)^{\alpha}} \mathbb{P}\left[\varepsilon_j \geq \varepsilon_j^*(x) | \varepsilon_i = \varepsilon_i^*(x)\right]}_{\text{strategic incentive}}. \tag{15}$$

Let $\mathbb{P}(\varepsilon_i \geq \varepsilon_i^*(x))$ denote the equilibrium probability of censorship for platform i. Note that the first two terms in bracket on the right hand side of equation (15) correspond respectively to the expected *legal cost* $(c_0 + c_1 x_i)$ and the expected *economic gains* $(x_i \gamma^{-\alpha})$ for platform i if it does not comply to censor (i.e. $a_i = 0$); the economic gains refer to the amount of users platform i would have managed to keep relative to the case where the platform complies to censor ($a_i = 1$). The third term of equation (15) stands for the expected *strategic incentives* between competitor platforms discussed before. The strategic incentive stems from the expected traffic brought about by potential switching users from other platforms.

From equation (15), we know that each platform's equilibrium strategy varies with respect to market covariates. However, upstream agents, such as an anti-trust regulator or an authoritarian government, are in general more concerned about the overall compliance level in the market and how it affects an average user as opposed to that of a particular platform. Then, how will the overall level of compliance (censorship) change as a result of a more concentrated or decentralized marketplace?

If a market becomes more concentrated by hosting fewer platforms, two factors will be at play: first, each platform captures a larger market share and bears higher legal costs of non-compliance; second, each platform is also subject to higher strategic incentives to differentiate from other obedient competitors by not censoring, now that users have fewer options to switch to. Following this change in the market structure, whether a platform is more or less likely to censor during a salient event depends on which of the above two forces dominates. If even a slight increase in the platforms' sizes alarms the government and significantly increases the risk of non-compliance, the former legal cost concern would dominate and lead to more censorship in the marketplace. If, on the other hand, limiting the number of alternatives significantly increases a platform's chance to capture switching users, then the latter strategic incentive would dominate and cause platforms to censor less often in equilibrium. Online Appendix B provides a formal characterization of how the equilibrium outcome varies with changes in key market covariates, including the legal costs per user (c_1), the number of competitors (N), and the correlation of private signals (ρ).

To quantify the relative magnitude of these two forces in my empirical context and draw policy-relevant counterfactual predictions, I proceed in the next section by estimating the model with the data described in Section 3. Specifically, I define an index to measure the market-level compliance, which I refer to as the "scope of censorship" (SC). In a market of N platforms, the index is defined as below

$$SC \equiv \mathbb{E}\left[\sum_{i=1}^{N} x_i a_i(x)\right]. \tag{16}$$

In other words, SC equals the expected share of platform users in the market affected by censorship during a salient event.[29]

6. Model Estimation

This section discusses the empirical strategies of identifying and estimating the model. I also illustrate the estimators' finite sample performance through a Monte Carlo experiment in the Appendix. In particular, I consider each event in the data as an independent game and assume that the structural parameters in my model remain the same across different games (events). I assume that factors that may have differentiated the platforms *outside* the event window other than their traffic are orthogonal to platforms' censorship decisions. I measure platforms' respective traffic by estimating the number of users using their average daily Alexa ranking 1 week prior to each event. A platform is considered to have censored an event if and only if it added

29. Note that in the context of this paper, censorship imposed by the government is exogenously given, and, hence, the "scope of censorship" refers to the market-level compliance by the platforms. In other words, this paper focuses on firms' compliance behavior that affects downstream users in a given political environment.

any keywords related to that event during the event week (i.e., week 0). I identify the legal cost parameters by exploiting variation in a platform's own traffic across events, and the strategic incentive by exploiting variation in its competitors' traffic. The latter identification strategy is conditional on the existence of exclusion restriction in this model: Changes in the traffic of a competitor platform only affect the platform's choice probabilities through the strategic interaction term. I show that the model parameters $(c_0, c_1, \gamma^{-\alpha})$ are identified up to scale without imposing further parametric assumptions on the error structure. Finally, I present the estimation results and discuss their implications.

6.1. Identification

I now formally describe the identification strategy in this subsection. The identification strategy follows Wan and Xu (2014) and takes two steps: First, I show that estimable bounds for the equilibrium beliefs can be derived under Assumption 1 and they can be arbitrarily close to each other as long as there exists one regressor that has sufficiently large independent variations (Lemma 1). Second, I show that model parameters $(c_0, c_1, (\gamma)^{-\alpha})$ are point identified up to scale by following Manski and Tamer (2002)'s interval–observed regressor approach. Third, in Figure 7, I present evidence from raw data that shows considerable variation in platforms' traffic across different events.

THEOREM 2. *If Assumption* 1 *holds, this structural model can be represented as a semi-parametric binary regression model where*

$$a_i = \mathbb{1}[\varepsilon_i \geq -\beta_0 - \beta_1 x_i - \sum_{j \neq i} \delta x_j \sigma_{ij}(x)], \forall i = 1, 2, .., N \tag{17}$$

and

$$\mathbb{P}(\sigma_{ij}^0(x) \leq \sigma_{ij}(x) \leq \sigma_{ij}^1(x)) = 1, \tag{18}$$

where

$$\sigma_{ij}(x) \equiv \mathbb{P}(\varepsilon_j \geq \varepsilon_j^*(x) | X = x, \varepsilon_i = \varepsilon_i^*(x)),$$
$$\sigma_{ij}^0(x) \equiv \mathbb{P}(\varepsilon_j \geq \varepsilon_j^*(x) | X = x, \varepsilon_i < \varepsilon_i^*(x)),$$
$$\sigma_{ij}^1(x) \equiv \mathbb{P}(\varepsilon_j \geq \varepsilon_j^*(x) | X = x, \varepsilon_i \geq \varepsilon_i^*(x)).$$

Proof. This proof extends Theorem 1 in Wan and Xu (2014) to N players. See Appendix. □

Note that the bounds $\sigma_{ij}^0 = \mathbb{P}(a_j = 1 | X = x, a_i = 0)$ and $\sigma_{ij}^1 = \mathbb{P}(a_j = 1 | X = x, a_i = 1)$ are non-parametrically estimable.[30]

30. Moreover, they collapse to be the same, that is, $\sigma_{ij}^0(X) = \sigma_{ij}^1(X)$ if and only if ε_i and ε_j are independent conditional on X for any $i \neq j$.

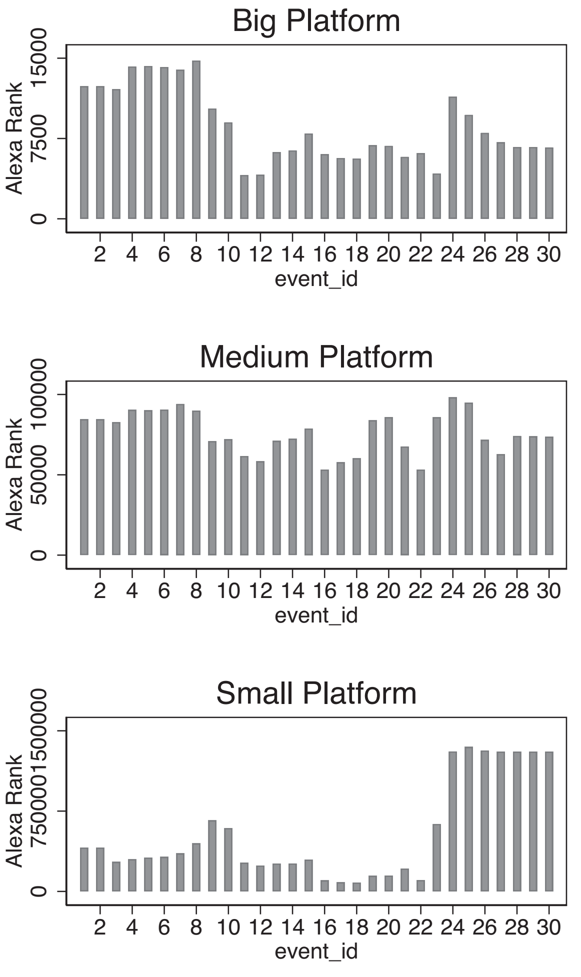


FIGURE 7. Pre-event traffic across different events. The three panels (top, middle, and bottom) plot the weekly average Alexa Rank of big, small, and medium platforms, respectively, 1 week prior to the event date. Note that an increase in one's Alexa Rank implies a decline in its traffic because traffic is inversely correlated with the Alexa Rank.

LEMMA 1. *For any* $\varepsilon > 0$ *and* $x_i \in \Omega_{X_i}$,

$$\lim_{t \to 1} \mathbb{P}\left(\sigma_{ij}^{1}(X) - \sigma_{ij}^{0}(X) \geq \varepsilon | X_i = x_i, X_j = t\right) = 0.$$

Lemma 1 shows that the non-parametrically estimable upper and lower bounds for the equilibrium beliefs can be arbitrarily close to each other when one of the platforms captures a dominant market share.

Proof. See Appendix. □

TABLE 5. Coefficient variation of pre-event traffic across events.

Platform i	Mean	Std. Dev	Std. Dev/Mean	Min	Max	Obs
Big	8,664	3,435	40%	4,091	14,778	30
Medium	76,344	13,037	17%	53,203	98,252	30
Small	537,541	458,901	85%	82,725	1,349,259	30

ASSUMPTION 2. $Median(\varepsilon_j|X = x) = 0$ for all $x \in \Omega_X$.

Assumption 2 allows for heteroskedasticity of unknown form.

ASSUMPTION 3. $\beta_1 = c_1 - \gamma^{-\alpha} \neq 0$. There exists $i, j (i \neq j)$ such that the distribution of X_i conditional on X_j has everywhere positive density with respect to the Lebesgue measure.

Assumption 3 implies an exclusion restriction. According to Theorem 2 from Wan and Xu (2014), $(\beta_0, \beta_1, \delta)$ is pointed identified up to scale if Assumptions 1–3 hold. Recall that in my framework,

$$\beta_0 = c_0, \tag{19}$$

$$\beta_1 = c_1 - \gamma^{-\alpha}, \tag{20}$$

$$\delta = -\frac{1}{(N-1)(\gamma)^{\alpha}}. \tag{21}$$

Therefore, (c_0, c_1) is point identified when we normalize $(\gamma)^{-\alpha}$ to 1.[31]

Variation in Market Structure. Figure 7 illustrates each platform's weekly average Alexa Rank 1 week prior to the event dates across different unexpected events, and Table 5 presents the summary statistics of the data used to plot Figure 7. Table 5 shows that there is considerable variation in the pre-event traffic of both big and small platforms, with their coefficents of variation (SD/Mean) being 40% and 85%, respectively. Note that, however, the changes in a platform's *pre-event* traffic is unlikely to be correlated with its subsequent censorship because of the exogenous nature of the unexpected events. There are a few non-event-related factors that might have led to the variations in the platform traffic over time: first, the live-streaming industry has experienced a sizable growth over the course of 2016 while the rate of growth is heterogeneous across different platforms (Jiemian.com 2016); second,

31. While it is reasonable to assume that the cost coefficients of not censoring is positive, I remain agnostic ex-ante about the sign of the cost coefficients c_0 or c_1 but only require that the expected total cost $c_0 + c_1 x_i > 0$. This approach accommodates factors beyond legal pressure, which may affect the sign, and provides greater flexibility in estimation. Meanwhile, platforms' censorship decisions are strategic substitutes given the users' switching behavior as modeled in my framework, meaning the sign of the term $(\gamma)^{-\alpha}$ is ex-ante known to be positive. Therefore, I choose to normalize this term, whose sign is definite.

market expansion has incentivized the platforms to offer competitive contracts to top live performers in the market, resulting in their high turnover rates and subsequently their loyal followers switching platforms. As discussed above, high variations in platforms' traffic are conductive to the model identification.

6.2. Estimation Methods

In this section, I present the estimation procedures of applying the Nested Fixed Point Algorithm (NFPX) proposed by Rust (1987) to my empirical setting.[32] Assume that private signals follow a joint normal distribution each with a mean of zero, and variance–covariance matrix Σ, where the variance of private signals is σ^2 on the leading diagonal of the matrix and the variance is multiplied by a correlation coefficient $\rho \geq 0$ as off-diagonal elements. For each iteration, I solve the equilibrium strategies for each platform in the inner loop and implement Maximum Likelihood Estimation in the outer loop. Given the state variables $X = (x_1, x_2, x_3)$ and a set of parameter guesses[33] for c_0, c_1, ρ, and σ, I calculate each platform's equilibrium cut-off values (ε_i^*) by iterating on the following system of equations:

$$\varepsilon_i^* = -c_0 - \left(c_1 - (\gamma)^{-\alpha}\right) x_i + \sum_{j \neq i} \frac{x_j (\gamma)^{-\alpha}}{N-1} \left[1 - \Phi\left(\frac{\varepsilon_j^* - \rho\varepsilon_i^*}{\sigma\sqrt{1-\rho^2}}\right)\right], \forall i = 1, 2, 3, \tag{22}$$

where $\gamma^{-\alpha}$ is normalized to 1. The inner loop iterates until the cutoff values converge to a fixed point. Using the cutoff values, I can construct platforms' choice probabilities and form the joint likelihood function. The joint distribution of platforms' choice probabilities conditional on X, which for notational convenience I abbreviate to $P^{a_1 a_2 a_3}$, has in total eight elements:

$$P^{111} = \mathbb{P}\left(\varepsilon_1 \geq \varepsilon_1^*(X), \varepsilon_2 \geq \varepsilon_2^*(X), \varepsilon_3 \geq \varepsilon_3^*(X)\right), \tag{23}$$

32. In the Online Appendix, I also present the results from adopting the two-step MMSE method proposed by Wan and Xu (2014). The two-step MMSE does not require any parametric assumption on the distribution of private signals. However, it generates less efficient estimates with wider confidence intervals. On the other hand, the NFXP imposes joint normality on the private signals, but the likelihood objective function in the Nested Fixed Point algorithm is differentiable and converges much faster than the score-type objective function in the two-step method. Given the relatively small sample size of 30 events, the parametric setting of NXFP method is more suitable in this context. As a robustness check, I present results from both methods in Online Appendix D and illustrate the finite sample performance of these two methods through a Monte Carlo experiment in Online Appendix E.

33. Note that it is necessary to restrict the parameter space to ensure that the game admits a unique monotonic pure strategy BNE, otherwise the sub-sample with a unique equilibrium for any parameter value would essentially have a size of zero. By Assumption 1, a sufficient condition that ensures the uniqueness of equilibrium is to impose an upper bound on the correlation coefficient and a lower bound on the payoff variance such that $\rho \leq \bar{\rho}$ and $\sigma^2 \geq \frac{1}{8\pi}\left(\frac{1+\rho}{1-\rho}\right)$. Note that as the upper bound of correlation coefficient $\bar{\rho}$ approaches 1, the lower bound on σ would approach infinity. Hence, I restrict the search of ρ to a compact parameter space $[0, \bar{\rho}]$, where $\bar{\rho} = 0.8$. This bound is further justified ex-post, as the estimated values of ρ lie well below this upper bound.

TABLE 6. Estimation results on censorship data.

	$\hat{c}_0$	$\hat{c}_1$	$(\gamma)^{-\alpha}$	$\hat{\rho}$	$\hat{\sigma}$
NFXP	−0.1183	1.1827	1	0.1187	0.2247
	[−0.1560, −0.0660]	[0.8164, 1.5249]	[·]	[0.0000, 0.6747]	[0.1995, 0.4244]

Notes: Table 6 reports the parameter estimates of interests from the Nested Fixed Point Estimation. Recall that the two-step Modified Maximum Score Estimation (MMSE) method does not specify the correlation structure of private signals and it requires a scale normalization. Thus, to generate comparable results between the two methods, I normalize $(\gamma)^{-\alpha}$ to 1. The bootstrapped 95% confidence intervals are in brackets. The number of bootstrap replications is 500.

$$P^{110} = \mathbb{P}\left(\varepsilon_1 \geq \varepsilon_1^*(X), \varepsilon_2 \geq \varepsilon_2^*(X), \varepsilon_3 < \varepsilon_3^*(X)\right), \tag{24}$$

$$P^{100} = \mathbb{P}\left(\varepsilon_1 \geq \varepsilon_1^*(X), \varepsilon_2 < \varepsilon_2^*(X), \varepsilon_3 < \varepsilon_3^*(X)\right), \tag{25}$$

$$P^{101} = \mathbb{P}\left(\varepsilon_1 \geq \varepsilon_1^*(X), \varepsilon_2 < \varepsilon_2^*(X), \varepsilon_3 \geq \varepsilon_3^*(X)\right), \tag{26}$$

$$P^{011} = \mathbb{P}\left(\varepsilon_1 < \varepsilon_1^*(X), \varepsilon_2 \geq \varepsilon_2^*(X), \varepsilon_3 \geq \varepsilon_3^*(X)\right), \tag{27}$$

$$P^{010} = \mathbb{P}\left(\varepsilon_1 < \varepsilon_1^*(X), \varepsilon_2 \geq \varepsilon_2^*(X), \varepsilon_3 < \varepsilon_3^*(X)\right), \tag{28}$$

$$P^{000} = \mathbb{P}\left(\varepsilon_1 < \varepsilon_1^*(X), \varepsilon_2 < \varepsilon_2^*(X), \varepsilon_3 < \varepsilon_3^*(X)\right), \tag{29}$$

$$P^{001} = \mathbb{P}\left(\varepsilon_1 < \varepsilon_1^*(X), \varepsilon_2 < \varepsilon_2^*(X), \varepsilon_3 \geq \varepsilon_3^*(X)\right). \tag{30}$$

The probabilities given in (23)–(30) are fully determined since $\varepsilon_1, \varepsilon_2, \varepsilon_3$ are assumed to be jointly normal. Given data consisting of G observations $(a_{1g}, a_{2g}, a_{3g}, X_g)$ for $g = 1, \ldots, G$, the log-likelihood function can then be calculated as

$$\mathcal{L}(\theta; X) = \sum_{g=1}^{G} \ln P^{a_{1g}a_{2g}a_{3g}}(\theta; X_g), \tag{31}$$

where $P^{a_{1g}a_{2g}a_{3g}}(\theta; X_g)$ denotes the probabilities defined in (23)–(30) evaluated at the point $(a_{1g}, a_{2g}, a_{3g}, X_g)$ for the parameter vector $\theta \equiv (c_0, c_1, \rho, \sigma)$. The outer loop updates the parameter guess and iterates until convergence using the simulated annealing algorithm.

6.3. Estimation Results

As introduced in Section 3, the dataset includes 30 events ($G = 30$), and the state variable X is measured by the market share of platforms 1 week prior to each event using its respective traffic ranking data. Table 6 reports the parameter estimates from both estimation methods. These estimates are accompanied by bootstrapped 95% confidence intervals. The number of bootstrap replications is 500.

TABLE 7. Censorship probability of platforms.

	Big platform	Medium platform	Small platform
Observed	0.4000	0.2000	0.1333
Predicted (NFXP)	0.4018	0.1948	0.1851

Note: The first and second row of the table report each platform's observed censorship frequency across the 30 unexpected events (described in Section 3) and the model-predicted censorship probability, respectively.

6.3.1. Parameter Interpretations. Estimation results confirm that first, the legal cost of non-compliance is positively ($\hat{c_1} > 0$) correlated with a platform's user size. Second, the correlation of platforms' private signals is moderately positive ($\hat{\rho} > 0$). However, for the small platform, the legal cost net of economic gains ($(c_1 - \gamma^{-\alpha})x_{Small}$) is dominated by its strategic incentives $(x_j/(N-1)\gamma^{\alpha})$, $j \in \{Big, Medium\}$. The last result suggests that, if this market becomes more concentrated, users are not necessarily more likely to experience censorship. I will further investigate the implications of this result through two counterfactual exercises in Section 7.

6.3.2. Model Fit. Table 7 presents the observed and predicted censorship probability of each platform across the 30 unexpected events during 2015–2017. The model fits the observed level well, suggesting that the model is able to recover the market covariates that influence the censorship decisions of platforms in the arrival of an average event in my data.

7. Counterfactual Predictions

Authoritarian governments are known for shutting down a firm at will if the firm does not comply with regulations. Why, then, does the government allow those non-compliers to stay in business, instead of permanently shutting them down? Moreover, one might expect digital platforms to benefit from mergers and acquisitions due to strong network effects, especially given that live-streaming platforms in China provide "highly homogeneous products and content" to consumers (Liu and Li 2016). Yet we observe a "very fragmented and localized" social media landscape in China (Chiu, Ip, and Silverman 2012). From a regulator's perspective, do platform mergers always lead to higher level of compliance?

In this section, I simulate two counterfactual exercises that shed some light on these questions.[34] First, I explore the scenario following a permanent shutdown of

34. Note that the counterfactual analysis on mergers presented in this section abstracts from several important elements and complex mechanisms discussed by regulators and researchers Crémer, De Montjoye, and Schweitzer (2019). The existence of network effects, extreme economies of scale, switching costs, vertical integration, and so on, can significantly affect the impact of such mergers. Hence, the results in this paper are by no means conclusive and should only be viewed as one of many possible mechanisms among that could potentially influence the outcome of a merger. The author thanks an anonymous referee for pointing this out.

TABLE 8. Counterfactual simulation results.

	Nested Fixed Point Estimation	
	Permanent shutdown	Merging two platforms
Δa_M	−0.0796	−0.0762
	[−0.1144, −0.0636]	[−0.1110, −0.0609]
Δa_B	0.0015	−0.0013
	[−0.0034, 0.0046]	[−0.0040, 0.0002]
ΔSC	−0.727%	−2.305%
	[−3.465%, 0.681%]	[−4.365%, −1.273%]
$\Delta \tilde{SC}$	0.815%	0.139%
(w/o strategic incentive)	[−0.623%, 1.561%]	[−0.126%, 0.267%]

Notes: The first two rows in Table 8 report the estimated change in the equilibrium censorship probability of the medium and big platforms, respectively, if the small platform is permanently shutdown (column 1), assuming that users of the small platform are split between the large and medium platforms, and if there is a merger between the medium and small platforms (column 2). The third row in Table 8 reports the estimated percentage change in the scope of censorship in the market. The fourth row in Table 8 reports the counterpart results in the above two cases if the strategic incentives are terminated under the assumption that users cannot switch between platforms. The bootstrapped 95% confidence intervals are in brackets. The number of bootstrap replications is 500.

the smallest platform. Then, I examine the case following a merger of the medium and small platforms[35] in the market. Finally, I compare the results from the above two cases with those in an alternative setup where there is no strategic interaction between platforms under the assumption that users cannot switch between platforms. Specifically, I define ΔSC as the percentage change in the the scope of censorship after such interventions, given by

$$\Delta SC \equiv [SC_{\text{after intervention}} - SC_{\text{before intervention}}]/SC_{\text{before intervention}}. \tag{32}$$

I assume that the correlation structure of private signals remains the same before and after the intervention.

Table 8 reports the simulation results. The point estimates are accompanied by bootstrapped 95% confidence intervals (in the square bracket). The number of bootstrap replications is 500. Both counterfactual exercises show that strategic incentives substantially mitigate the impact of increasing legal costs typically associated with market concentration. First, note that permanently shutting down the smallest platform could backfire and lead to an unintended outcome where the overall level of compliance is lower in the marketplace. This is because if the non-compliant smallest platform was absent, the remaining two platforms would share the

35. As mentioned in Section 3, while the medium and small platforms are listed under the same parent company Tian Ge, they are now operated by two different subsidiary companies and manage their own keyword blacklists separately. Prior to this "merger" exercise, each platform owner makes censorship decisions only based on the expected traffic of their own platform and users can freely switch between the two platforms. However, after the "merger" exercise, two platforms become one with the new platform owner making censorship decisions based on the joint market share. Meanwhile, users can only switch out to the big platform, but they can no longer "switch between" the medium and small platforms. Note that users can still switch to their outside-market options like watching TV or listening to music.

whole market, both of which would not only face higher legal costs but also have stronger strategic incentives to differentiate by not censoring. Their strategic incentives would increase as they would expect to attract more switching users from its then only competitor. As a result, while the big platform complies slightly more often, the medium platform complies significantly less often in equilibrium due to stronger strategic incentives. When the latter effect dominates, market concentration can reduce the overall scope of censorship.

On the other hand, market concentration via merging two platforms also leads to a decline in the scope of censorship. While the merged two platforms are under higher political pressure than when they were separated due to larger market share, the big platform now competes with a much larger market player and thus has stronger strategic incentives to differentiate by not censoring. In this case, the latter effect dominates the former and thus the overall level of compliance in the market is lower following the merger. Note that, however, if we neglect the strategic incentives in platforms' decision-making process by assuming that users cannot switch between platforms, then market concentration would have led to an increase in the scope of censorship.

8. Conclusion

This paper studies the relationship between the market structure of online platforms and their compliance with censorship regulations under political pressure. Using panel data on three major live-streaming platforms in China, this paper adopts an event study approach to explore how quickly and intensively platforms comply with censoring users' messages following a sequence of unexpected salient events. The event study analysis shows that the compliance behavior is different across platforms of different sizes: The largest platform not only censored a higher number of keywords on average, it also complied faster than the smaller platforms.

Motivated by the empirical patterns, I propose a structural model of oligopolistic competition. In the model, a platform's profit depends on its own compliance decision as well as that of its competitors, induced by the switching behavior of users with a diverse taste for censorship. The model predicts that if platforms are highly asymmetric, small platforms have strong incentives to differentiate from their big competitors by not censoring, while big platforms find it more costly to delay censorship. However, market concentration comes with two countervailing forces: first, each platform captures a larger market share and bears higher legal costs of non-compliance; second, platforms have more strategic incentives to differentiate from competitors via delaying their own compliance, now that users from their competitors have fewer options to switch to. Moreover, conditional on a platform's own decision of not complying, its user traffic is positively correlated with its competitors' compliance decision. As a result, whether a platform is more or less likely to comply with censorship following market concentration depends on which of the two forces dominates. If even a slight increase in a platform's size alarms the government and significantly raises its cost of non-compliance, the former political pressure would dominate and generate more censorship in the marketplace. If, on the other hand, limiting the number of alternatives encourages more users to switch between platforms due to sufficiently lower switching or search cost, then the latter strategic incentive would dominate and cause platforms to censor less often in equilibrium. This paper quantifies the relative magnitude of these two forces by exploiting the variation in platforms' market share across different events. Based on the model estimates, this paper simulates two policy-relevant counterfactual experiments. The counterfactual analyses suggest that merging or permanently shutting down small platforms both turn out to lower the equilibrium scope of censorship compliance in the marketplace.

My findings suggest that decentralizing online market power could help an authoritarian government maintain sufficiently high market-level censorship with minimal enforcement. Although utilizing market structure might seem like an indirect and subtle strategy to achieve desired information control, rather than imposing higher political penalties directly on all platforms, it could serve as an effective alternative when the marginal enforcement cost is already high. After all, there is a limit to which the government can increase its political penalty without excessively disrupting the economy or inciting more protests. Coincidentally, unlike the US market, which is dominated by a handful of mainstream social media platforms, Chinese social media remains "very fragmented and localized (Chiu, Ip, and Silverman 2012)."

Beyond China, the framework proposed in this paper also offers useful insights on regulating misinformation in Western democracies. While most people dislike misinformation and prefer its removal, fake news often takes time to verify and can become an accepted "alternative truth" for some before it is exposed as false. When two segments of users co-exist, one is quick to identify "misinformation" and the other takes it as the alternative truth; removing the same piece of content pleases the former at the expense of upsetting the latter. If large platforms are expected to be more responsible for removing misinformation or to take actions faster, the latter group may disproportionately switch to small platforms that receive less legal attention every time a piece of misinformation turns viral. Subsequently, social media mergers and acquisitions not only affect the parties involved, but they could also significantly distort other small incumbents' incentives to comply with the regulation—a distortion that may exacerbate the spread of misinformation and create more "echo chambers" (Bénabou and Tirole 2016) in the long run. Hence, policymakers should consider this spillover effect when forming expectations of a social media platform's compliance with regulations.

Several limitations remain in this study, some of which, I seek to explore in ongoing research. First, the number of keywords on a platform's blacklist may not be a perfect indicator for its censorship intensity because I do not directly observe the amount of messages censored by a platform. For example, some keywords may be used more frequently than others and thus blacklisting a small number of frequently used keywords could affect more users than blacklisting a large number of less common words. Second, this paper is limited to studying only the events that were observed in the 2-year dataset of the three platforms; there may exist other events that were censored by some other platforms or purposefully ignored by all three platforms. Third, this paper abstracts away from modeling the user belief and assumes that users switch to different platforms with equal probability. One may naturally consider an alternative assumption where users switch to a platform with a probability that is inversely correlated with the platform's equilibrium censorship probability. Under this alternative assumption, a small platform becomes even less likely to censor, as users are more likely to switch there than to larger platforms. As a result, the presence of a small platform further reduces the strategic incentives for a large platform to censor. In that regard, my counterfactual results likely underestimate the effect predicted by the alternative model. Fourth, this paper assumes that the legal cost from non-compliance is linear in platforms' user size, which minimizes the number of parameters required for estimation given the available data. However, it is an intriguing question as to whether or how the implications may change under a non-linear cost function. Finally, this paper assumes that all users dislike censorship to some extent, but ignores the possibility of users who may be in favor of censorship. In reality, platforms may host both segments of users. Future research can extend the model and explore these issues.

References

Adamic, Lada A. (2000). "Zipf, Power-Laws, and Pareto-a Ranking Tutorial." Xerox Palo Alto Research Center, Palo Alto, CA, http://ginger.hpl.hp.com/shl/papers/ranking/ranking.html.

Aradillas-Lopez, Andres and Amit Gandhi (2013). "Robust Inference of Strategic Interactions in Static Games." Working Paper, University of Wisconsin.

Arnold, Andrew (2018). "Do We Really Need To Start Regulating Social Media?" The Forbes.

Athey, Susan (2001). "Single Crossing Properties and the Existence of Pure Strategy Equilibria in Games of Incomplete Information." *Econometrica*, 69, 861–889.

Bajari, Patrick, Han Hong, John Krainer, and Denis Nekipelov (2006). "Estimating Static Models of Strategic Interaction." NBER Working Paper.

Bamman, David, Brendan O'Connor, and Noah Smith (2012). "Censorship and Deletion Practices in Chinese Social Media." First Monday, 17. https://firstmonday.org/ojs/index.php/fm/article/download/3943/3169. Accessed on March 17, 2017.

Barrett, Paul (2022). "*Musk says Twitter is Biased Against Conservatives–Facts Say Otherwise*." The Hill.

Beattie, Graham, Ruben Durante, Brian Knight, and Ananya Sen (2021). "Advertising Spending and Media Bias: Evidence from News Coverage of Car Safety Recalls." *Management Science*, 67, 698–719.

Bénabou, Roland and Jean Tirole (2016). "Mindful Economics: The Production, Consumption, and Value of Beliefs." *Journal of Economic Perspectives*, 30, 141–64.

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin (2009). "Identification of Peer Effects Through Social Networks." *Journal of Econometrics*, 150, 41–55.

Caillaud, Bernard and Bruno Jullien (2003). "Chicken & Egg: Competition Among Intermediation Service Providers." *RAND Journal of Economics*, 34(2), 309–328.

Chen, Yuyu and David Y. Yang (2019). "The Impact of Media Censorship: 1984 or Brave New World?" *American Economic Review*, 109(6), 2294–2332.

Chiu, Cindy, Chris Ip, and Ari Silverman (2012). "*Understanding Social Media in China*." McKinsey Quarterly.

Crandall, Jedidiah R, Masashi Crete-Nishihata, Jeffrey Knockel, Sarah McKune, Adam Senft, Diana Tseng, and Greg Wiseman (2013). "Chat Program Censorship and Surveillance in China:

Tracking TOM-Skype and Sina UC." First Monday, 18. https://firstmonday.org/ojs/index.php/fm/article/download/4628/3727. Accessed on March 20, 2017.

Crémer, Jacques, Yves-Alexandre De Montjoye, and Heike Schweitzer (2019). "Competition Policy for the Digital Era." Report for the European Commission, 1. https://op.europa.eu/en/publication-detail/-/publication/21dc175c-7b76-11e9-9f05-01aa75ed71a1/language-en. Accessed November 2024.

De Paula, Aureo and Xun Tang (2012). "Inference of Signs of Interaction Effects in Simultaneous Games with Incomplete Information." *Econometrica*, 80, 143–172.

Durante, Ruben and Brian Knight (2012). "Partisan Control, Media Bias, and Viewer Responses: Evidence from Berlusconi's Italy." *Journal of the European Economic Association*, 10, 451–481.

Edgington, Eugene S. (1961). "Probability Table for Number of Runs of Signs of First Differences in Ordered Series." *Journal of the American Statistical Association*, 56, 156–159.

Edmond, C. (2013). "Information Manipulation, Coordination, and Regime Change." *The Review of Economic Studies*, 80, 1422–1458.

Evans, David S. (2003). "Some Empirical Aspects of Multi-Sided Platform Industries." *Review of Network Economics*, 2, 191–209

Gehlbach, Scott and Konstantin Sonin (2014). "Government Control of the Media." *Journal of Public Economy*, 118, 163–171.

Gentzkow, Matthew, Jesse M. Shapiro, and Michael Sinkinson (2014). "Competition and Ideological Diversity: Historical Evidence from us Newspapers." *The American Economic Review*, 104(10), 3073–3114.

Geron, Tomio (2012). "*YY.com: China's Unique Real-Time Voice and Video Service with a Virtual Goods Twist*." The Forbes.

Griffith, Sean J. (2015). "Corporate Governance in an Era of Compliance." *William & Mary Law Review*, 57, 2075.

Harsanyi, John C. (1973). "Games with Randomly Disturbed Payoffs: A New Rationale for Mixed-Strategy Equilibrium Points." *International Journal of Game Theory*, 2, 1–23.

Hobbs, William R. and Margaret Roberts (2018). "How Sudden Censorship can Increase Access to Information." *American Political Science Review*.

Hooker, Matthew P. (2019). "Censorship, Free Speech & Facebook: Applying the First Amendment to Social Media Platforms via the Public Function Exception." *Washington Journal of Law, Technology & Arts*, 15, 36.

Human Rights Watch (2006). "Corporate Complicity in Chinese internet Censorship." 18, 8(C)https://www.hrw.org/report/2006/08/10/race-bottom/corporate-complicity-chinese-internet-censorship, retrieved March 2017.

Initiative, OpenNet (2005). "Internet Filtering in China 2004–2005: A Country Study." http://www.opennetinitiative.net/studies/china, retrieved 20 March 2007.

iResearch (2016). "2016 Report on the Chinese Live Video Streaming Market." iResearch Consulting Group. https://www.iresearch.cn/. retrieved 20 March 2017.

Jiemian.com (2016). "2016 Q1 Report on the Chinese Live-streaming Market." Jiemian News. https://www.jiemian.com/, retrieved 20 March 2017.

King, Gary, Jennifer Pan, and Margaret Roberts (2013). "How Censorship in China Allows Government Criticism but Silences Collective Expression." *American Political Science Review*, 107, 326–343.

Knight, Brian and Ana Tribin (2019). "The Limits of Propaganda: Evidence from Chavez's Venezuela." *Journal of the European Economic Association*, 17, 567–605.

Knockel, J (2016). "*One App, Two Systems: How WeChat Uses One Censorship Policy in China and another Internationally*." Citizen Lab, Munk School, University of Toronto.

Knockel, Jeffrey, Masashi Crete-Nishihata, Jason Q. Ng, Adam Senft, and Jedidiah R. Crandall (2015). "Every Rose Has Its Thorn: Censorship and Surveillance on Social Video Platforms in China."In *5th USENIX Workshop on Free and Open Communications on the Internet (FOCI 15)*, USENIX Association.

Liu, R. and D. Li (2016). "The Pulse of Live Streaming in China: Understanding the New Chinese Lifestyle and the Business Behind it." China Tech Insights. https://mat1.gtimg.com/chinatechinsights/file/20160926/The_pulse_of_live_streaming_in_China.pdf, retrieved 20 March 2017.

Lomas, Natasha (2017). “Facebook, Twitter Still Failing on Hate Speech in Germany as New Law Proposed.” Tech Crunch. https://techcrunch.com/2017/03/14/facebook-twitter-still-failing-on-hate-speech-in-germany/, retrieved 14 March 2017.
Lorentzen, P. (2014). “China’s Strategic Censorship.” *American Journal of Political Science*, 58, 402–414.
Manski, Charles F. and Elie Tamer (2002). “Inference on Regressions with Interval Data on a Regressor or Outcome.” *Econometrica*, 70, 519–546.
Medium (2016). “Virtual Gifts Are Still the Top Earner in China’s Live Video Streaming Market.” Medium. https://medium.com/@traceyxiang/virtual-gifts-are-still-the-top-earner-in-chinas-live-video-streaming-market-7b41c328a0f4, retrieved 6 May 2016.
Miller, Blake (2018). “The Limits of Commercialized Censorship in China.” Working Paper, Department of Government, London School of Economics and Political Science.
Newberry, Christina (2019). “130+ Social Media Statistics that Matter to Marketers in 2019.”https://blog.hootsuite.com/social-media-statistics-for-social-media-managers/, retrieved 5 March 2019.
Oltermann, Philip (2018). “Tough New German Law Puts Tech Firms and free Speech in Spotlight.” The Guardian. https://www.theguardian.com/world/2018/jan/05/tough-new-german-law-puts-tech-firms-and-free-speech-in-spotlight, retrieved 5 January 2018.
Oremus, Will (2022). “How Social Media ’censorship’ Became a Front Line in the Culture War.” *The Washington Post*. https://www.washingtonpost.com/technology/2022/10/09/social-media-content-moderation/, retrieved 9 October 2022.
Pan, Jennifer and Yiqing Xu (2018). “China’s Ideological Spectrum.” *The Journal of Politics*, 80, 254–273.
Qin, Bei, David Strömberg, and Yanhui Wu (2017). “Why does China Allow Freer Social Media? Protests Versus Surveillance and Propaganda.” *Journal of Economic Perspectives*, 31, 117–40.
Qin, Bei, David Strömberg, and Yanhui Wu (2018). “Media Bias in China.” *American Economic Review*, 108(9), 2442–76.
Roberts, Margaret E. (2018). “Censored.” In *Censored*. Princeton University Press.
Rochet, Jean-Charles and Jean Tirole (2003). “Platform Competition in Two-Sided Markets.” *Journal of the European Economic Association*, 1, 990–1029.
Rochet, Jean-Charles and Jean Tirole (2006). “Two-Sided Markets: A Progress Report.” *The RAND Journal of Economics*, 37, 645–667.
Simonov, Andrey and Justin Rao (2022). “Demand for Online News under Government Control: Evidence from Russia.” *Journal of Political Economy*, 130, 000–00.
Stern, Rachel E. and Jonathan Hassid (2012). “Amplifying Silence: Uncertainty and Control Parables in Contemporary China.” *Comparative Political Studies*, 45, 1230–1254.
Sweeting, Andrew (2004). “Coordination Games, Multiple Equilibria and the Timing of Radio Commercials.” *Journal of Economics and Management Strategy*, 15(4), 909–942.
Sweeting, Andrew (2009). “The Strategic Timing Incentives of Commercial Radio Stations: An Empirical Analysis Using Multiple Equilibria.” *The RAND Journal of Economics*, 40, 710–742.
Tian Ge Interactive Holdings Limited (2015). “Corporate Presentation, April 2015.”https://www.tiange.com/Upload/TiangeCorporatePPTLongVersion_en.pdf, retrieved 20 March 2017.
Tian Ge Interactive Holdings Limited (2016). “2016 Annual Report.”https://www.tiange.com, retrieved 20 March 2017.
Wan, Yuanyuan and Haiqing Xu (2014). “Semiparametric Identification of Binary Decision Games of Incomplete Information with Correlated Private Signals.” *Journal of Econometrics*, 182, 235–246.
Wong, Chun Han (2015). “Where China’s Liberals Live, and Why.” The Wall Street Journal. https://blogs.wsj.com/chinarealtime/2015/04/15/where-chinas-liberals-live-and-why/, retrieved 15 April 2017.
Xiang, Tracey (2016). “The World Of Chinese Interactive Online Video Shows.” TechNode. https://technode.com/2016/02/19/the-world-of-chinese-interactive-online-video-shows/, retrieved 19 February 2016.

Yildirim, Pinar, Esther Gal-Or, and Tansev Geylani (2013). "User-Generated Content and Bias in News Media." *Management Science*, 59, 2655–2666.
YY Inc. (2015). "Corporate Presentation of YY Inc., July 2015."https://investors.yy.com. Accessed March 20, 2017.
YY Inc. (2016). "2016 Annual Report on Form 20-F of YY Inc."https://ir.joyy.com/financial-information/annual-reports. Accessed March 20, 2017.
Zhuang, Maiting (2022). "Intergovernmental Conflict and Censorship: Evidence from China's Anti-Corruption Campaign." *Journal of the European Economic Association*, 20, 2540–2585.

Supplementary Data

Supplementary data are available at *JEEA* online.

Journal of the European Economic Association 2025 00(0):1–38
https://doi.org/10.1093/jeea/jvaf018