The Impact of Earned Media on Demand: Evidence from a Natural Experiment*

Stephan Seiler
Stanford University

Song Yao
Northwestern University

Wenbo Wang
Hong Kong University of Science and Technology

This draft: November 18, 2015

We leverage a temporary block of the Chinese microblogging platform Sina Weibo due to political events to estimate the causal effect of user-generated microblogging content on product demand in the context of TV show viewership. Using a set of difference-in-differences regressions, we show viewership decreased more strongly in geographical areas with a higher Sina Weibo penetration, and only for shows with a high activity level on Sina Weibo. We quantify the effect on viewership in units of comments on tweets (comments were disabled during the block) by instrumenting the number of relevant comments with a dummy for the time period of the block, and find an elasticity of 0.02. In terms of the behavioral mechanism, we find more pre-show microblogging activity increases demand, whereas the ability to engage in microblogging during showtime as a complementary activity to TV consumption does not affect product demand.

Keywords: Microblogging, Advertising, Social Media

*We thank Tomomichi Amano, Eric Anderson, Yuxin Chen, Pradeep Chintagunta, Jun Kim, Ken Wilbur, Florian Zettelmeyer, Juanjuan Zhang, and seminar participants at Temple University, University of British Columbia, and Stanford GSB, as well as conference participants at the Big Data Conference (NYU), Marketing Science (Baltimore), and Yale China India Insights for their valuable comments. This research is partially funded by the Hong Kong Research Grants Council Early Career Scheme #699213 and the McManus Research Grant at the Kellogg School of Management, Northwestern University. We also thank Audrey Siyu Zhao for being an excellent Research Assistant. All errors are our own.

1 Introduction

With the advent of web 2.0, microblogging via platforms such as Twitter has become increasingly popular among Internet users. Since 2013, more than 500 million tweets have been posted on Twitter each day[1] and the user base of Twitter has grown exponentially in the past several years, especially among consumers with higher education and income levels. By 2014, nearly 20% of the US population were Twitter users.[2] Consequently, microblogging websites have developed into a crucial channel through which firms advertise and promote a variety of products. While firms use microblogs to post their own tweets as a form of advertising (paid media), earned media, where individual consumers tweet and comment about the firm's products, also plays a crucial, if not more important, role. Many marketers believe the latter can be an effective way to reach customers who might be more likely to be influenced by their friends than by a firm's advertising message. A recent report by Gallup based on a large-scale consumer survey argues that "consumers are highly adept at tuning out brand-related Facebook and Twitter content. [... Companies] best bet is to engage their existing customers and inspire them to advocate on their behalf."[3]

Despite such claims, empirical evidence on the causal impact of earned media on product demand is scant and the effectiveness of earned media is less well understood than more traditional marketing tools. We posit the lack of empirical evidence is in large part due to the fact that firms are not directly in control of the amount of earned media provided; therefore, implementing a randomized field experiment is difficult. Furthermore, similar to the analysis of other marketing activity such as TV advertising, using observational data to estimate the causal effect of microblogging on demand is problematic because products that are tweeted about more frequently are likely to differ in other respects. Such endogeneity concerns, combined with the inability to experimentally induce variation, constitute a major obstacle to obtaining a credible estimate of earned-media effectiveness. In this paper, we use a natural experiment to circumvent both issues and estimate the causal effect of user-generated microblogging content (i.e., earned media) on product demand in the context of TV show viewership.[4] This approach allows us to answer the question of whether earned media can enhance product demand and to quantify the magnitude of the effect.

We leverage plausibly exogenous variation in microblogging activity that originated from the temporary shut-down of Sina Weibo, the most popular microblogging outlet in China. In early 2012, the defection of a prominent government official led to a political scandal in China. During this period, the Chinese government limited the functionality of Sina Weibo for three days. The government's intervention was aimed at suppressing microblogging related to the political scandal, but it effectively reduced activity across the platform as a whole, even in domains that were entirely unrelated to politics. We focus on one such domain, the microblogging activity regarding TV

[1] Twitter Official Blog, "New Tweets per second record, and how!" 2013.

[2] Pew Research Center, "Social Media Update 2014," 2015.

[3] "The myth of social media," June 11, 2014, *Wall Street Journal*.

[4] We note the term earned media is broader than user-generated content on microblogs and includes brand mentions in traditional news outlets such as newspaper articles, as well as consumer word-of-mouth in other types of social media.

shows, and demonstrate that the effect of the block of Sina Weibo on microblogging activity led to a significant reduction in viewership. Importantly, the nature of the censorship was such that it primarily affected user-generated content, namely, earned media. Specifically, the censorship did not block Sina Weibo entirely, but only disabled the commenting function. In other words, users (and TV companies) were still able to post tweets but were unable to comment on those tweets. We also note that whereas firms provide a fraction of content on Sina Weibo through their official accounts, this fraction is relatively small, making up only 0.18% of the total tweets in our setting.

A further advantage of our setting is the magnitude of the variation induced by our natural experiment. Commenting, which the censorship blocked entirely and hence reduced to zero, is a commonly used feature on microblogging websites, and the number of comments pertaining to the TV shows in our data is about five times larger than the number of original tweets. Therefore, the block removes a substantial amount of content.[5] Specifically, the average episode in our data lost over 6,000 comments.[6] A field experiment would unlikely be able to replicate such a large shock to the amount of microblogging activity. This feature of the data is particularly relevant here because studies on the impact of advertising have noted that statistical power to measure economically relevant effects is often small (see Lewis and Rao, forthcoming). Although we are considering user-generated content rather than advertising, we believe that issues of statistical power are likely to also apply here.

Our analysis proceeds in multiple steps. We first use a series of difference-in-differences regressions to show that during the block, TV viewership dropped in a pattern that is consistent with the block affecting viewership via decreasing the amount of activity on Sina Weibo, rather than through a direct effect of the block or the political scandal that triggered it. Our setting is a particularly rich one in which to conduct this type of analysis and allows us to investigate viewership changes from various vantage points. First, we show TV viewership dropped across all shows in our sample in mainland China. In comparison, for Hong Kong, we do not find a decrease in viewership for a comparable set of shows. Hong Kong provides a natural control group in our setting because Sina Weibo has a close to zero penetration rate in Hong Kong, because Twitter is available to Hong Kong users (it is blocked in mainland China) and constitutes the primary microblogging platform in Hong Kong. Second, among Chinese mainland cities, we find the block affected viewership more strongly in cities where Sina Weibo had high penetration rates, but had little effect in cities where the penetration rates were low. Third, we find that within the set of shows we observe in our sample, only those shows with a strong presence on Sina Weibo experienced a drop in viewership during the block. Instead, for shows with a low level of activity, we do not observe any change in viewership. Finally, we provide evidence that the Sina Weibo block occurred mid-way through a prolonged period of about three months during which the political scandal played out, and did

[5] As we show in more detail later, the block did not affect other types of activity on Sina Weibo. Specifically, the number of tweets, re-tweets, and likes (i.e., the other options available to influence content) did not change during the block.

[6] We define the number of comments pertaining to a specific episode of a show as the number of comments posted on the day of the episode and before the episode started airing.

not coincide with any major event of the scandal. These patterns provide evidence that any direct effect of the block (or other contemporaneous events) is unlikely.

In a final step, we quantify the impact of microblogging on product demand as a function of the number of comments. To implement such an analysis, we run an instrumental variable regression in which we regress viewership on the level of activity on Sina Weibo as measured by the number of relevant comments on the day a specific show aired. We use a dummy for the time period of the block as an instrument for the number of comments in order to isolate the part of the variation driven by the natural experiment. Doing so, we find that a 100% increase in the number of comments regarding a specific show leads to a 2% increase in viewership. We further evaluate potential mechanisms through which earned media on Sina Weibo affects TV viewership. We find that earned media works similarly to traditional advertising such that more user-generated content *before a show* leads to more TV consumption. By contrast, earned media content created *during the show* does not increase its viewership. In other words, the consumption of TV shows and the creation of related microblogging content do not appear to be complementary to each other.

Our paper relates to the literature on measuring advertising effectiveness by exploiting a specific source of exogenous variation. This approach is most closely related to a series of papers that also aim to exploit non-experimental variation to measure the effect of advertising. Hartmann and Klapper (2015) explored the variation in TV advertising exposure among consumers, using show ratings across markets during the Super Bowl Finals, and found a significant boost in product demand due to Super Bowl ads. To measure the effect and spillover of TV pharmaceutical advertising, Shapiro (2014) used ad-exposure variation created naturally by borders between neighboring TV markets. Other researchers have used field experiments to investigate the effectiveness of banner ads (e.g., Sahni, 2015a; Sahni, 2015b), sponsored search advertising (e.g., Blake et al., 2014), and direct mail advertising (Bertrand et al., 2010). A particularly related paper is Gong et al. (2015), who conducted a field experiment to measure microblogging's effect on demand by varying the level of tweets and retweets generated by one specific firm. In comparison, we capture variation in user-generated content rather than the firms' own content provision. Furthermore, we are able to exploit a shock that was broader in scope and affected many different products (in our case, TV shows) in the market. Our paper also relates more broadly to the growing literature on social media and user-generated content. To name a few, Godes and Mayzlin (2004) use an online newsgroup's discussion to study the effect word-of-mouth on TV ratings, Chevalier and Mayzlin (2006) investigate how online reviews affect book sales, Trusov et al. (2010) proposed a model to identify influential users on social media websites, and Lambrecht et al. (2015) studied whether Twitter promotions are more effective on early trend adopters or influential users. Zhang and Zhu (2011), Aaltonen and Seiler (2015), Shriver et al. (2013), and Toubia and Stephen (2013) investigated why users contribute content to social network platforms such as Wikipedia and Twitter. Sonnier et al. (2011) consider how online communication affects sales. Stephen and Galak (2012) and Lovett and Staelin (forthcoming) investigate the interaction between paid media (e.g., firm paid advertising), earned media, and own media (e.g., content on a firm's own website). We contribute to

		Mean	S.D.	90th Perc.	95th Perc.	Max	S.D. (Time Series Only)
Ratings	Rating	0.434	0.483	0.941	1.225	4.166	0.158
	Log Rating	0.322	0.260	0.663	0.800	1.642	0.090
Microblogging	Comments	5,803	43,656	192	1,941	1,159,749	33,679
(Pre-show Activity on the Same Day)	Log Comments	1.78	2.79	5.26	7.57	13.96	1.46
	Tweets	1,006	5,508	67	967	83,299	3,698
	Re-Tweets	21,472	250,858	365	5,791	12,613,689	235,508
	Likes	189	7,762	0	4	486,478	7,497

Table 1: **Descriptive Statistics: TV Ratings and Sina Weibo Activity.** The unit of observation is an episode. Our sample contains 7,899 episodes.

the literature by establishing and measuring the effect of microblogging on demand with a clean identification strategy.[7]

The remainder of the paper is structured as follows. The next two sections outline the data and provide details on the nature of the Sina Weibo block. In section 4, we present results from difference-in-differences regressions. In section 5, we implement an instrumental variable regression to measure the elasticity of viewership with respect to the number of comments on Sina Weibo tweets. In Section 6, we investigate the behavioral mechanism driving our results. Finally, we provide some concluding remarks.

2 Data

We rely on two separate sources of data in this paper. First, we use detailed show-level data on TV viewership (i.e., TV show ratings) across a large set of channels, as well as geographic locations in China. Second, we assembled a unique and highly detailed data set of microblogging activity by scraping Sina Weibo for the set of shows that appear in the TV data. Below, we outline in detail the two data sets and provide descriptive statistics.

2.1 TV Ratings Data

We obtain data on viewership for a large set of TV shows in mainland China and Hong Kong from CSM Media Research, the leading TV-rating data provider in China. The data are reported at the episode-city level. Each episode belongs to a specific show that comprises multiple episodes if

[7] Zhang and Zhu (2011) investigate the impact of a censorship event on Wikipedia that blocked activity from a subset of users. They analyze how the blockage of some users affected the contributions of users whose accounts were not blocked, and find that contributions from those users decreased because of the loss of part of their audience. Although not identical to the censorship of Sina Weibo, the variation exploited in Zhang and Zhu (2011) shares some similarities with the Sina Weibo block in our data.

the show is a series.[8] Because of constraints imposed by the data provider, we do not observe the full universe of shows. Instead, we work with data from an extensive subset of shows and cities. Specifically, we select a set of 24 demographically and economically diverse major cities in mainland China and Hong Kong (see Figure B1 for the cities' locations). For each city (except Hong Kong), the set of shows is identical. Next, we select the top 20 national channels based on their market share in mainland China, which covers roughly 80% of the mainland market.[9] In Hong Kong, we obtain data for all six local channels. Finally, we also collect data on viewership of the six Hong Kong channels in Shenzhen, which is located in mainland China but is adjacent to Hong Kong. Due to geographical proximity, Shenzhen residents can receive Hong Kong channels. For each city and channel combination, we collect market shares in terms of viewership for all TV shows that ran between 6 p.m. and 12 a.m. every day from March 1 to April 30, 2012. We choose these two months based on the timing of the Sina Weibo block, which took place from March 31 to April 2, 2012.[10] Our time window therefore covers roughly a month before and after the block.

For mainland China, our data comprise a total of 24 cities, 20 channels, and 907 shows. For Hong Kong, we have six channels and 508 shows. In a final step, we further narrow the set of shows down to the ones that provide relevant variation for our analysis. Specifically, all of our later regressions will analyze the change in viewership for a given show during the block relative to episodes of the same show that aired before or after the block. We therefore focus on shows for which we observe multiple episodes and that aired at least one episode during the block and one episode before or after the block. This selection leaves us with a total of 166 shows in mainland China.[11] Because some shows were aired (usually on different dates and times) on multiple channels, we have 193 unique show-channel pairs. For Hong Kong, we have 112 shows and 132 unique show-channel pairs. In all of our analysis, a show/channel combination constitutes the cross-sectional unit; going forward, we simply refer to such a combination as a "show."

Because the set of mainland shows is identical across cities, we base most of our analysis on the average market share of each show-episode across all cities in mainland China. When computing the average market share, we weight the observations from the different cities by their population size. Table 1 presents descriptive statistics for the aggregated data on the 193 mainland shows for which we observe a total of 7,899 individual episodes. Our main outcome measure for each show is the show's rating, that is, its market share in terms of viewership (measured from 0 to 100 rating points). In the first two rows of Table 1, we report ratings as well as log-ratings across all 7,899 episodes. The average episode has a rating of 0.434, but the distribution is relatively skewed with a large standard deviation and a maximum rating of 4.166 in our sample.[12] We therefore prefer to

[8]Some shows we observe only once in our data, as is the case, for instance, for a special event or movie. Our later analysis is going to focus on shows for which we have repeated observations.

[9]We exclude local channels from this calculation.

[10]Strictly speaking, the block was in place from 8 a.m. on March 31 to 8 a.m. on April 3. Because we analyze the effect of microblogging on prime time TV shows that air after 6 p.m., and because most relevant tweets occur after 8 a.m., we treat the three calendar days March 31 to April 2 as the time period of the block.

[11]One of the 20 channels does not air any serial shows, and hence we drop all shows from this channel due to our selection rule for shows.

[12]The market share for each show is defined as the share of households watching the particular show among all

use log-rating as the dependent variable in most of our analysis. We also decompose the variation in ratings into the across-show as well as the time-series variation within shows, by computing the residual from a regression of ratings onto show fixed effects. The standard deviation of the residuals is reported in the last column of the table. The amount of variation at the across- and within-show level provides us with a benchmark to later assess the magnitude of the estimated effect of Weibo activity on TV viewership (section 5). We provide the same descriptive statistics for shows in Hong Kong in Table B1 in the appendix. Finally, we note that many shows in China are broadcasted at a high frequency (often at a daily interval) and the average interval between two consecutive episodes is 1.48 days. We therefore have a substantial panel dimension in our data. For the two months of data in our sample, we observe 7,899 episodes for 193 shows and hence have about 40 observations per show.

2.2 Microblogging Data

Our second data set measures the amount of activity related to each TV show on Sina Weibo, which is the primary microblogging website in China, with 61 million daily active users and around 100 million daily tweets.[13] On Sina Weibo, users can engage in four different types of activities: tweeting, retweeting, "like"-ing, and commenting. Commenting is the type of activity primarily affected by the block. It allows a user to re-tweet an existing tweet with some content of her own (a "comment"). Figure 1 displays an example of a tweet together with a series of comments pertaining to the tweet.[14] The numbers of comments, re-tweets, and likes are displayed below the original tweet followed by a list of comments with the most recent comments listed first.

We obtained the microblogging data by scraping the Sina Weibo website for information relevant to each show in our data. Specifically, for all shows contained in our data, we scraped every tweet mentioning the show during March and April of 2012.[15] We further collected the number of comments on each tweet as well as the number of re-tweets and likes from the counts displayed below each tweet (see Figure 1).[16] To relate the amount of Sina Weibo activity to a particular episode, we calculate the number of relevant tweets, re-tweets, comments, and likes that users posted on the day that a particular episode aired and before the time at which the episode aired.[17]

households who own a TV. Because many consumers do not watch any TV at any given point in time, the rating numbers are generally quite low.

[13] Retrieved from Weibo's SEC filings (http://ir.weibo.com/phoenix.zhtml?c=253076&p=irol-sec, accessed November, 2015) in April 2014. Daily number of users and tweets are calculated based on data from December 2013.

[14] The tweet is about a popular Chinese dance competition show, similar to the US show "Dancing with the Stars."

[15] Sina Weibo allows users to search existing tweets by "keyword," and returns the full set of tweets that include the specific "keyword" (the user can also define the relevant time period over which to search). We define as the relevant tweets for each show the set of tweets that mention the show's name, and hence we can scrape them by simply entering the show's name as a "keyword" into the Weibo search algorithm.

[16] During the time period of our data collection, Sina Weibo altered the algorithm that displays historical tweets and now only displays a subset of the full set of relevant tweets for a specific keyword. Luckily, most of our data collection was already completed when this change happened. However, we collected some pieces of data after the change and we therefore need to rescale those data to reflect the fact that Sina Weibo now displays only a subset of all tweets. Section A of the appendix provides more details on how we handled this issue.

[17] Strictly speaking, we use the timestamp of each tweet to define whether the tweet falls into the relevant time window. For all other types of activity, we use the timestamp associated with the tweet to which they belong.

Figure 1: **Example of a Tweet with Comments.**

We report descriptive statistics for these episode-level variables in the lower portion of Table 1.

We first note that comments, which are the type of activity primarily affected by the block, are frequently used, and for the average episode, the number of comments posted prior to the episode airing is more than five times larger than the number of tweets. Therefore, although tweets are likely to be more salient, comments make up a large fraction of user-generated content in our data. In terms of other types of user activity, we also find a large number of re-tweets related to the shows in our sample. This finding is perhaps unsurprising, because re-tweets are a lower-involvement tool and do not generate any new content, but help disseminate existing content. Finally, the "like" feature is used relatively sparsely. These four types of user activity represent the exhaustive set of options for participating on Sina Weibo as a user. Although our primary focus will be on comments and their decrease during the block, we also track other types of activity and later assess whether the block indirectly affected them.

In terms of the distribution of comments across episodes, we find a highly skewed distribution. The average number of pre-show comments per episode in our data is equal to 5,800, but some episodes in the right tail of the distribution are mentioned in a substantially larger number of comments. Similar patterns also hold for tweets, re-tweets, and likes. Furthermore, all four types of activity are highly correlated with each other. For example, log-comments and log-tweets have a correlation of 0.93. We also calculate the within-show variation in activity by regressing comments, as well as each of the other variables, onto show fixed effects. We report the standard deviation of the residuals from these regressions in the last column of Table 1. Finally, we also note that most tweets originate from private users rather than the firms themselves. We obtained data on the activity on each show's official Sina Weibo account (if the show has an account) and found the firm's own tweets make up only 0.18 percent for the average show in our sample. Hence, the vast majority of tweets (and other activity) on Sina Weibo is made up of user-generated content rather than content provided by the shows' producers or the TV channels airing the shows.

3 The Block

In February 2012, a political scandal erupted in China after a top government official defected. By March 2012, many rumors related to the political scandal appeared on the Internet, especially on social media websites. To stop the rumors from spreading further, and to remove rumors that had already been posted, Sina Weibo announced early on March 31, 2012, that from 8 a.m. on that day until 8 a.m. on April 3, the microblogging platform would be partially blocked.[18] Specifically, the commenting function that allows a user to re-tweet an existing tweet with some content of her own (a "comment") was disabled during that time period.[19] Figure 2 displays the announcement of the

[18] The government gave no official statement linking the Weibo block explicitly to political events, but news sources clearly connected them (see for example "Coup Rumors Spur China to Hem in Social Networking Sites", *New York Times*, March 31, 2012, accessed November 2015).

[19] Another Chinese microblogging website, Tencent Weibo, also blocked the comment function during the same period. We focus on Sina Weibo because Tencent Weibo only had about 9% of the market share, whereas Sina Weibo

Figure 2: **Announcement of the Sina Weibo Block.** See main text for the English translation. The message was displayed on the Sina Weibo homepage (weibo.com) for the duration of the block.

block that appeared on the Sina Weibo homepage and was visible to anybody using the platform. The statement reads in English as follows:

To all Weibo users:

Recently, there have been many harmful and illegal rumors that appeared in the comment section of Weibo. To effectively remove these rumors, from March 31 8a.m. to April 3 8a.m., we will suspend the comment function on Weibo. After removing the rumors, we will reopen the comment function. Such an action is necessary. It is for the purpose of creating a better environment for users' communications. We ask for your understanding and forgiveness. Thank you for your support.

Sina Weibo

March 31, 2012

Because the origin of the block were political events, one might wonder whether the events that triggered it might have a direct effect on TV viewership. For instance, one could imagine consumers paid more attention to the political events during those three days and thus were less likely to watch TV. Such behavior would lead to a correlation between the time period of the block and TV viewership that was not triggered by the reduced number of tweets and in fact was not at all related to the block itself. Two pieces of evidence, however, speak against the presence of such a direct effect. First, the political scandal unfolded over the course of about three months, from February to April 2012. Several important events were related to the scandal during this period, and the block does not coincide with any of them. Second, we find that during the most salient moments of the scandal (which are quite pronounced, as we show below), TV viewership remained

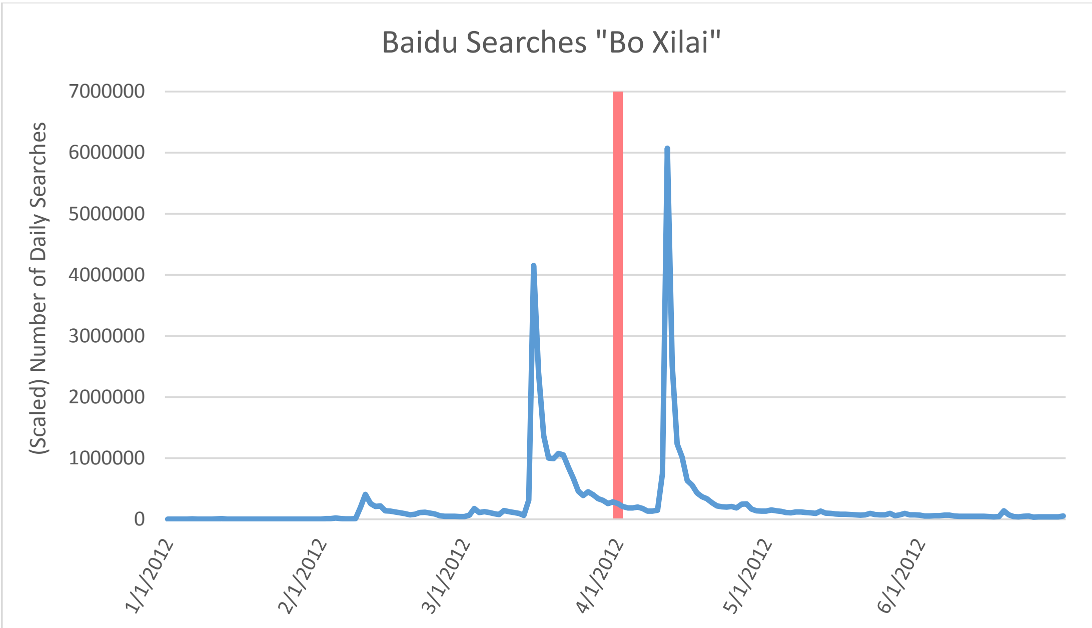
held nearly 90%. The authorities also closed down 16 minor websites for spreading political rumors.

unchanged, while it dropped during the Sina Weibo block. Therefore, interest in the political events did not seem to have an effect on viewership.

To substantiate the first point above, we obtain search index data from the major Chinese search engine Baidu for search queries regarding the two Chinese officials involved in the scandal: Wang Lijun and Bo Xilai.[20] We obtained the search index of a keyword using Baidu's patented algorithm that is based on the search volume of the focal keyword for a particular period. Accordingly, it is reasonable to believe the search indices of the two names are highly correlated with the attention people paid to the political scandal. In Figure 3, we present the time series of search indices for the two names from January to May 2012, where the three days of the Sina Weibo block are indicated in red. The graphs display the time window of the entire series of events from early February to mid-April and, very saliently, feature three prominent peaks across the two time series: in early February, mid-March, and mid-April. These peaks correspond respectively to (1) Wang Lijun traveling to the US consulate, (2) Bo Xilai's dismissal from his municipal post, and (3) Bo's suspension from the party's Central Committee. The censorship of Sina Weibo occurred between the latter two of the three peaks and during a time period in which searches for the two names were at a fairly low level. In other words, the block happened at a time when people paid little attention to the political scandal. We also analyzed similar graphs based on Google trends data rather than the Baidu search index, and found the time series patterns to be very similar.

The second and maybe even more important point is that we do not find any evidence that TV viewership changed as a function of the saliency of events related to the political scandal. We establish this pattern by using our later regression framework (see section 4, equation 1) in which we regress (log) show ratings onto show fixed effects and a dummy for the time period of the block. On top of these variables, we then also include the (log) search indices of both names in the regression. We find the coefficients on both search indices are small and insignificant, whereas the coefficient on the block dummy remains significant and its magnitude is not affected by the inclusion of the additional variables. We report detailed results from this regression in Table B2 in the appendix. In that table, we show robustness to using the search indices individually and jointly as controls as well as a specification with a series of dummy variables for the peaks of the scandal. Across all specifications, we find a clear and precise null effect of the search indices and key event dummies on TV show ratings. In summary, the arguments outlined above show that interest in the political scandal was unlikely to affect TV ratings directly. Furthermore, even if it did affect ratings, such a finding is unlikely to contaminate our analysis, because the block did not coincide with any major event related to the scandal.

[20] Wang Lijun is the name of the official who defected and Bo Xilai was Wang's supervisor.



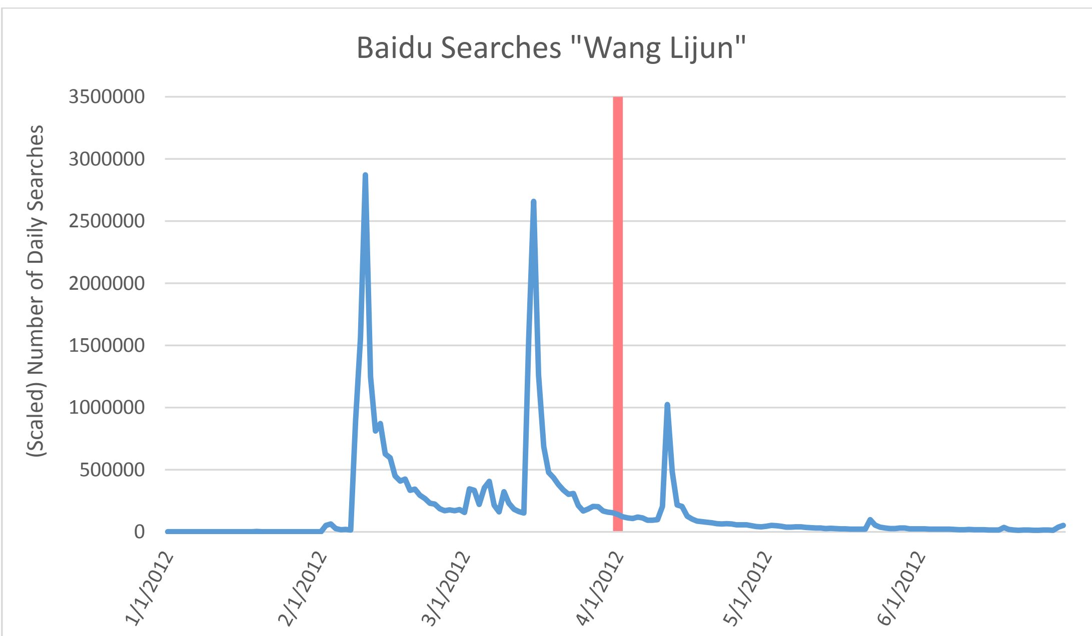


Figure 3: **Time Series of Search Indices for the Names of the Two Politicians Involved in the Scandal from January to June 2012.** The three days of the Sina Weibo block are indicated in red.

4 Difference-in-Difference Estimation

In this section, we establish that the block affected TV viewership via its effect on the amount of TV-related tweeting activity on Sina Weibo. The simplest approach to this end would be an analysis of the time series of show-specific viewership around the time of the censor. Although this approach has potential issues, we provide it as a benchmark case for the later set of regressions. Specifically, we run a regression of the show-specific log-rating on a dummy for the three days of the block, show fixed effects, as well as day-of-the-week dummies:

$$LogRating_{jt} = \alpha * Block_t + \delta_j + Weekday_t'\gamma + \varepsilon_{jt}. \tag{1}$$

$LogRating_{jt}$ denotes the logarithm of the rating of show j on day t. $Block_t$ is a dummy variable equal to 1 for the three days of the block. δ_j is show j's fixed effect. $Weekday_t$ is a vector of day-of-the-week dummies. ε_{jt} is the regression error term. Standard errors are clustered at the show level. The results from this regression are reported in column (1) of Table 2 and show a significant drop in viewership during the Sina Weibo block.

This simple first regression might cause worry over two potential issues. First, the block (and the political events related to it) might have had a direct impact on viewership. Second, any unrelated event that happened during the same time window and that might also have affected viewership is potentially problematic. Based on the reasoning provided in the previous section, we believe a direct effect of the block is an unlikely scenario in our setting. Nevertheless, as an alternative way to deal with both issues just outlined, we implemented an analysis that relies on the fact that different geographical areas, as well as different types of shows, should be differentially affected if the block affects viewership because of the change in microblogging activity but not through any other channel. Our setting is particularly rich in this regard because we have clear predictions where and for which shows the effect of the block should be stronger. In particular, Hong Kong residents were not subject to the block, because Twitter rather than Weibo is the primary microblogging platform. Accordingly, TV viewership should not have been affected in Hong Kong during the block. Second, across the 24 mainland cities in our data, the penetration rates of Sina Weibo show large variation. Consequently, the block should only have affected in cities where Sina Weibo was more widely adopted. Furthermore, we also know the extent to which individual shows have a presence on Sina Weibo, by computing the number of show-related comments outside the block period. We can hence analyze whether viewership changes differentially across shows with different amounts of activity on Sina Weibo. Next, we explain these analyses in detail.

4.1 Geography-based Analysis

We start with an analysis along the geographical dimension that leverages the fact that Sina Weibo is the predominant microblogging platform in mainland China but not in Hong Kong, where Twitter is available and the primary microblogging platform. Due to this alternative, the usage of Sina

Weibo in Hong Kong is very low. Specifically, for the set of shows in our sample, Hong Kong users generate only 0.5% of the amount of tweets relative to the neighboring city Shenzhen in mainland China, which is comparable in population size.[21] We take this finding as evidence that the usage of Sina Weibo is negligibly small in Hong Kong.

We run a difference-in-differences regression using data on all shows in mainland China as well as Hong Kong in which we interact the block-dummy with a dummy for mainland China. Similar to the previous regression, we include show and day-of-the-week dummies and cluster standard errors at the show level:

$$LogRating_{jt} = \alpha * Block_t + \beta * Block_t * Mainland_j + \delta_j + Weekday_t'\gamma + \varepsilon_{jt}. \tag{2}$$

If the block affects TV viewership via lowering the amount of TV-show-related activity on Sina Weibo, we would expect the block to only affect viewership in mainland China, but not in Hong Kong. Instead, if the block has a direct effect in the sense that the political events surrounding it led to lower TV viewing in general, we would expect to see the reduction in viewership in both Hong Kong and mainland China. Similarly, any unrelated event that happened during the period of the block is likely to affect both geographies comparably due to their cultural and economic similarity. As a further piece of evidence that Hong Kong constitutes a valid control group, we compute the correlations of Baidu search indices regarding the scandal between mainland China and Hong Kong. The Baidu search indices in Hong Kong and mainland China for the two involved politicians show a strong correlation of 0.89 and 0.94, respectively. We also provide a plot of the time series of searches in Figure B2 in the appendix, which shows the indices are highly correlated across the two locations with identical peaks in activity. Hence, any direct effect of the political scandal that triggered the block is likely to have affected Hong Kong in a similar fashion as the mainland. The identifying assumption is hence that absent the effect of the block on viewership through its impact on Sina Weibo, the time series of TV viewership in Hong Kong and mainland China would have been identical.[22]

The results from this regression are reported in column (2) of Table 2. The primary coefficient of interest is the interaction of the block with the mainland dummy, which represents the difference-in-differences estimate. We find a negative and significant effect of -0.026. This effect is slightly larger, but not significantly different from, the “simple-difference” regression reported in column (1) of the table. We also note the coefficient on the censor dummy is small in magnitude and not significantly different from zero. In other words, for Hong Kong, we observe no difference in viewing behavior during the time period of the Sina Weibo block.

[21]To compute the penetration rate of Sina Weibo in Hong Kong relative to Shenzhen, we collected additional data on the number of tweets mentioning a specific show by city of origin. Note that ideally, we would like to observe the number of users reading posts on Sina Weibo by city. Here, we use the number of posts at the city level as a proxy for readership and general usage of the platform.

[22]We note that we do not need to make the assumption that Twitter (which is used in Hong Kong) and Sina Weibo (which is used in mainland China) are similar. Instead, our identifying assumption is that the block of Sina Weibo affected TV shows in mainland China, whereas Hong Kong did not experience a similar shock at the same point in time.

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	Log-Rating	Log-Rating	Log-Rating	Log-Rating	Log-Rating	Log-Rating
Sample	Mainland China	HK and Mainland China	HK and Shenzhen	Shenzhen (inc. HK shows)	24 Cities in Mainl. China	24 Cities in Mainl. China
Censor Dummy	-0.017***	0.005	0.002	-0.005	-0.010	-0.008
	(0.005)	(0.010)	(0.010)	(0.009)	(0.006)	(0.006)
Mainland × Censor Dummy		-0.026**				
		(0.012)				
Shenzhen × Censor Dummy			-0.035**			
			(0.014)			
Mainland Show × Censor Dummy				-0.023*		
				(0.013)		
Sina Weibo Penetration × Censor Dummy					-0.027*	
					(0.014)	
Above Median Penet. × Censor Dummy						-0.016***
						(0.006)
Show FEs	Yes	Yes	Yes	Yes	Yes	Yes
Weekday Dummies	Yes	Yes	Yes	Yes	Yes	Yes
City FEs	n/a	n/a	n/a	n/a	Yes	Yes
Observations	7,899	11,427	11,427	9,533	189,576	189,576
Shows	193	325	325	255	193	193

Table 2: **Difference-in-Differences Regressions: Geographical Differences.** The unit of observation is an episode in columns (1) to (4) and an episode/city combination in columns (5) and (6). Standard errors are clustered at the show level.

We run two further robustness tests to confirm the finding from the regression above. First, in an effort to make the treatment and control groups even more similar to each other, we run the same regression as before but substitute the aggregated mainland data with city-level data from Shenzhen. Shenzhen neighbors Hong Kong, and the two cities are only separated by a river. The results from this regression, which are reported in column (3),[23] are similar to the previous regression comparing Hong Kong and the entire mainland. We find a negative and significant effect of the interaction of the block with the Shenzhen dummy variable. Finally, we exploit one slightly different dimension of the data. In Shenzhen, the local Hong Kong shows can also be received due to the geographical proximity. We therefore collected ratings data in Shenzhen for shows of Hong Kong origin. Contrary to mainland shows, the Hong Kong shows have little or no presence on Sina Weibo, because they are primarily targeted at the Hong Kong population, which does not use Sina

[23]We reiterate that the set of shows in Shenzhen is identical to the aggregated mainland data. The mainland data are obtained by aggregating the city-level data across 24 cities with the identical set of shows.

Weibo.[24] Hence, Hong Kong shows broadcasted in Shenzhen provide another natural control group for the mainland shows. This test also has the further advantage that it compares different groups of shows for the same geographical entity, Shenzhen. Regression results for this specification are reported in column (4) of Table 2 and again show similar patterns, although statistical power is slightly weaker and the interaction effect is only significant at the 10% level.

In a final set of regressions based on geographical differences in the strength of the effect, we explore heterogeneity across different cities in mainland China. Specifically, for the 24 mainland cities in our sample, we compute a proxy for the local usage of Sina Weibo. Similar to the comparison with Hong Kong, where Sina Weibo usage is close to zero, we would expect TV ratings in cities with higher penetration rates of Sina Weibo to be more affected by the block. To construct a proxy for the city-level penetration rates of Sina Weibo, we collect the frequency of Baidu searches for the terms "Sina Weibo Registration," "Sina Weibo Logon," and "Sina Weibo" at the local level for six months before February 2012, when the political scandal started. To explore the heterogeneity across cities, we run our baseline regression with an additional interaction term of the censor dummy and the local Sina Weibo usage variable (as well as city fixed effects). Results from this regression are reported in column (5) of Table 2 and we find a negative effect on the interaction term that is significant at the 10% level. In terms of magnitude, we scaled the penetration-rate variable in such a way that it takes on values between 0 and 1. Therefore, the lowest penetration city experiences a rating decrease of 1%, whereas the highest penetration city saw ratings drop by 3.7%. We also implement a specification in which we interact the censor dummy with an indicator for whether the city has an above-median usage of Sina Weibo and again find a negative effect. The statistical precision is higher in this specification and the interaction term is significant at the 1% level.[25]

4.2 Across-Show Analysis

In a second and complementary set of regressions, we analyze differences in the effect of the block across different shows in our sample. To simplify the analysis, we concentrate solely on the ratings data for mainland China in terms of geography.[26] As described earlier, substantial variation exists in the amount of activity both within and across shows on Sina Weibo (see Table 1). Here, we use the across-show differences in Sina Weibo activity outside the block to group shows and evaluate whether the block affected them differently. Specifically, if the block only affected TV viewing via

[24] Although the Hong Kong shows do not have official accounts on Sina Weibo, users in mainland China (and in particular in Shenzhen) might still use Sina Weibo to talk about these shows. We checked whether this is the case and found that Sina Weibo activity regarding Hong Kong shows that originates from Shenzhen does exists, but the level of activity is fairly low. Shenzhen users' tweets about Hong Kong shows constitute 17% of the volume of tweets regarding mainland shows. Hence, we would expect the block to matter less for Hong Kong shows relative to mainland shows.

[25] We also ran the same regression using Internet-penetration rates as a proxy for Sina Weibo usage at the local level and find similar results.

[26] That is, we do not use data for Hong Kong or the city-level data for mainland China in this part of the analysis.

its effect on Sina Weibo, we would expect to see a stronger decrease in ratings for shows with a more active presence on Sina Weibo and no effect for shows that have little to no Sina Weibo activity associated with them.

We test this hypothesis using data on the average number of comments pertaining to an episode of each show in the month before and after the block.[27] In other words, we are categorizing shows by the "normal" amount of commenting activity regarding the show for an episode that did not air during the block. In a first regression, we categorize shows into three equally sized bins according to their level of activity on Sina Weibo. We note the lowest activity category contains shows with less than one comment per episode; hence, these shows have almost no activity. Medium- and high-activity shows are characterized by at least one and 22 comments per episode, respectively. Column (1) of Table 3 shows results from a regression in which we interact the block dummy with dummies for the different activity levels. As before, we control for show fixed effects and day-of-the-week dummies and cluster standard errors at the show level:

$$\begin{aligned} LogRating_{jt} = & \alpha * Block_t + \alpha_M * Block_t * Medium_j + \alpha_H * Block_t * High_j \\ & + \delta_j + Weekday_t'\gamma + \varepsilon_{jt}. \end{aligned} \tag{3}$$

$Medium_j$ and $High_j$ denote dummies for two of the activity groups. Low-activity shows constitute the omitted category, and the block dummy captures the effect on those shows. We find the effect for the low-activity shows is very small in magnitude. The medium-level shows experience a stronger drop of 1.7% in their ratings (i.e., 0.4% + 1.3%). Although the decrease for medium-activity shows is not significantly different from the effect for low-activity shows, we note the decrease is statistically significantly different from zero (at the 10% level). For shows with a strong presence on Sina Weibo, we find an effect of larger magnitude. For this category of shows, ratings dropped by 2.9 percent during the block (i.e., 0.4% + 2.5%) and the estimated coefficient is significantly different from the effect for low-activity shows. Furthermore, the effect is large in magnitude and corresponds to about 30% of the typical show-level fluctuation (one standard deviation) in log-ratings over time of 0.09, which we report in Table 1.

To further probe the robustness of the results, we also interact the average number of comments per episode (in units of 100,000 comments) linearly with a dummy for the time period of the block. In this regression, the interaction captures whether shows with a stronger presence on Sina Weibo experienced a sharper decrease in ratings. We find a significant effect for the interaction term, which we report in column (2) of Table 3. To get a sense of the magnitude of the effect, note the show with the largest amount of microblogging activity received an average of 280,000 comments per episode. The predicted decrease for the top show therefore equals about 21% according to this

[27] Results are similar when grouping shows according to only their pre-block level of activity. However, a small set of shows (14) are not observed before the block (they started airing during the block). Therefore, using post-block data allows us to assign an activity level also to that set of shows. As we show in section 6.3 below, the block had little post-block long-term impact on microblogging activity. Accordingly, the April activity can still be considered a good measure of "normal" Weibo activity.

Dependent Variable	(1) Log Rating	(2) Log Rating
Censor Dummy	-0.004	-0.011**
	(0.005)	(0.005)
Medium Weibo Activity	-0.013	
× Censor Dummy	(0.011)	
High Weibo Activity	-0.025**	
× Censor Dummy	(0.011)	
Weibo Activity (Unit: 100,000 Comments)		-0.071***
× Censor Dummy		(0.023)
Show FEs	Yes	Yes
Day-of-the-Week Dummies	Yes	Yes
Observations	7,899	7,899
Shows	193	193

Table 3: **Difference-in-Differences Regressions: Across-Show Differences.** The unit of observation is an episode. Standard errors are clustered at the show level.

regression specification. Because the distribution of show-related tweets is highly skewed, we prefer the first specification, which splits shows into three bins rather than the interaction with a linear term.

4.3 Robustness Checks

One concern with any difference-in-differences setting is the existence of pre-existing differential time trends between the treatment and control group. This concern typically applies to regressions that compare treatment and control group observations before and after a policy change. The identifying assumption in such a setting is that the treatment group, had it not been treated, would have followed the same trajectory over time as the control group. Therefore, the existence of differential time trends might cast doubt on the validity of this assumption. We note that in our context, this issue is less likely to be a concern. First, we study the evolution of TV show ratings over a relatively short time horizon of two months, and hence strong time trends in TV viewership are unlikely. Second, and more importantly, our setting is slightly richer than other typical difference-in-differences settings because the treatment is temporary in our case. We thus have observations for both before and after the treatment period. Therefore, the identification argument boils down to the assumptions that during the three days of the block, the treatment group would have evolved in the same way as the control group had it not been treated. Differential time trends between both groups are therefore less likely to lead to a spurious result.

Nevertheless, we also explicitly test for the presence of differential time trends and assess whether our key coefficient of interest is robust to including such time trends. We implement this robustness check for the entire set of regressions presented in this section. Specifically, we re-run the four difference-in-differences regressions across geographies and across type of show (one regression of HK versus mainland China, two regressions across-city within mainland China, and one regression across Weibo activity levels), but also include a linear time trend and a linear time trend interacted with a treatment-group dummy for the respective regression. We report results from this set of regressions in Table B3 in the appendix. The table replicates the original regressions for easier comparison and then reports a version of each regression that also includes time trends for treatment and control groups. As expected, we find little evidence of any time trends for either treatment or control group in any of the specifications. Furthermore, the coefficients of interest of the censor dummy interacted with the relevant treatment-group dummy are almost unchanged across all specifications.

As a further test in the same vain, we also implement a set of placebo regressions in which we move the treatment period from late March / early April to either mid-March or mid-April. To make things as comparable as possible, we implement two placebo treatment periods that are three days long, which is equal to the duration of the actual block. We run two placebo tests for each of the four specifications listed above and among a total of 10 relevant coefficients across all regressions, we find only two to be significant (one at the 1% level, the other at the 10% level).[28] In both cases, the sign of the estimated coefficients has the opposite sign relative to the "true" treatment coefficient. The results from these placebo tests provides further evidence that differential time trends are unlikely to contaminate our results.

4.4 Firms' Response and Other Marketing Activity

One further concern with any external shock to one particular kind of marketing activity is that firms might react to this shock by adjusting marketing activity through other channels. Due to the short duration of the block, we think such a scenario is unlikely in our case. Furthermore, the censorship was not announced beforehand, and firms were unlikely to have been able to anticipate the block. Therefore, the short duration and suddenness of the block constitutes a feature of our natural experiment that provides a clean setting to isolate the effect of changing relevant marketing activity (earned media in this case) in only one channel. Any longer-term shock is likely to trigger a response from firms, which would constitute a further obstacle to correctly identifying the effect of interest.

[28] Three difference-in-differences regressions across geographies each contains one relevant interaction term, whereas the across-show regression contains two relevant interactions (because we are considering three different levels of Weibo activity), giving us a total of five relevant coefficients across all four regressions. For each regression, we implement two placebo tests and hence obtain a total of 10 relevant coefficients.

4.5 Difference-in-Differences Summary

In summary, the two sets of difference-in-differences regressions show that across geographies and TV shows, we observe patterns of rating changes during the block that are consistent with an effect of the block on ratings via reducing show-related microblogging activity. Specifically, we find that shows in Hong Kong, where Twitter rather than Sina Weibo is primarily used, did not experience a reduction in viewership. Instead, for the average mainland show, we do find a significant drop in ratings during the block. For the 24 mainland cities, we also find that cities with high penetration rates of Sina Weibo experienced a significant drop in viewership. By contrast, the block had a smaller effect in cities with low Weibo usage. We then further decompose the movement in mainland show ratings into shows with different amounts of activity on Sina Weibo, and find the decrease in ratings is primarily driven by a set of high-activity shows. These shows experienced an almost 3% drop in ratings, whereas ratings of lower activity mainland shows were not affected.

5 Instrument Variable Regressions

The difference-in-differences regressions in the previous section provide evidence that the block affected TV show ratings via its effect on Sina Weibo activity. Although the movement in ratings induced by the block is large relative to the typical movement in ratings for a given show over time, one would ideally like to relate the amount of decrease in earned media on Sina Weibo induced by the block directly to the change in TV ratings. To provide a measurement of this kind, we run an instrumental variable regression in which we regress ratings on the total number of comments related to the specific episode and instrument the number of comments with a dummy for whether the episode aired on a day during the block. Similar to the regression framework used in the previous section, we control for show fixed effects, day of the week, and cluster standard errors at the show level. Formally, we run the following regression:

$$LogRating_{jt} = \alpha * LogComments_{jt} + \delta_j + Weekday_t'\gamma + \varepsilon_{jt}, \tag{4}$$

where $LogComments_{jt}$ denotes the (log) number of comments related to an episode of show j, which aired on day t. This variable is instrumented with a dummy for whether on day t Sina Weibo was blocked. Below, we first present details of the first-stage regression, which documents the decrease in comments during the block, and then present the second-stage results.

5.1 First Stage

Before proceeding to the actual first stage of our IV regression, we assess the effect of the block on activity on Sina Weibo more broadly. To this end, we analyze the time series of show-specific activity in the simplest possible way by calculating the number of tweets, re-tweets, comments, and

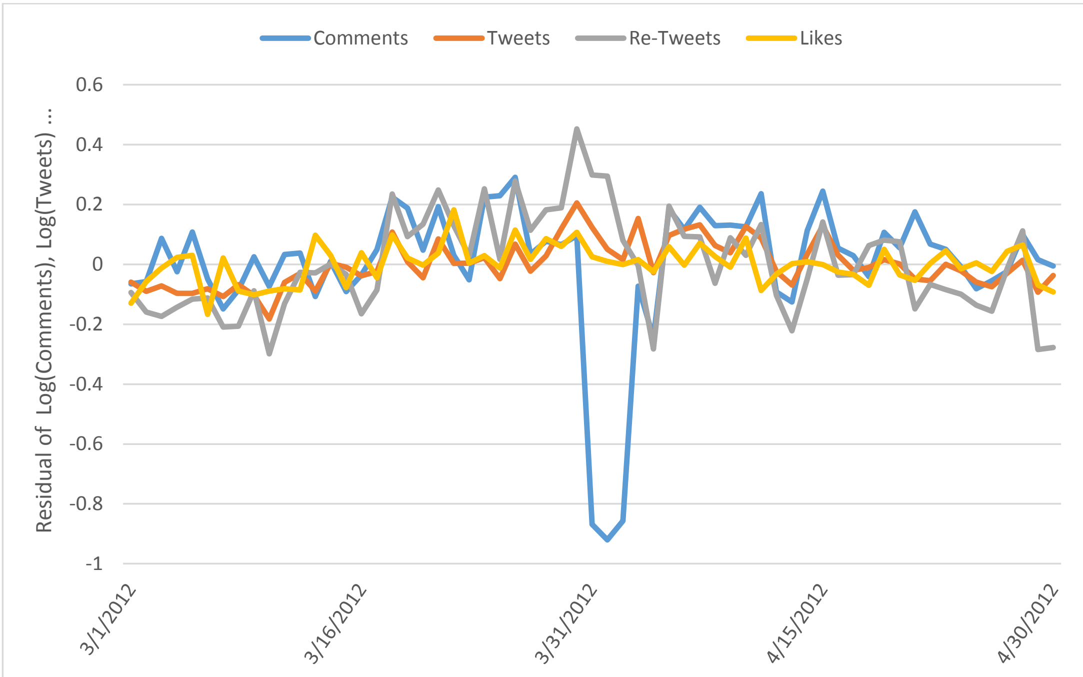


Figure 4: **The Evolution of Sina Weibo Activity over Time.** The graph plots the average value of the residuals from a set of regressions of log(Comments), log(Tweets), log(Re-Tweets) and log(Likes) on show fixed effects for each day of the sample period.

likes for each day in March and April 2012 for each show contained in our sample (regardless of whether an episode actually aired on the specific day). We then regress the different daily activity measures onto show fixed effects, weekday dummies as well as a dummy for the three days of the block.[29] Columns (1) and (2) in Table 4 report the results from two regressions that use the number of comments, the type of activity being primarily affected by the block, as the dependent variable. The two columns report results using the number of comments in either levels or logs. For both specifications, we unsurprisingly find the block caused a substantial drop in the number of comments. We also note the levels regression is much lower in precision, presumably due to a very skewed distribution of daily comments and therefore a much higher variance in the dependent variable in levels. In our preferred specification in logs, the impact of the block corresponds to a shift of 70% of a standard deviation in the time-series variation of log-comments.[30] We repeat the same type of regression also for the other three types of user involvement on Sina Weibo: tweets, re-tweets, and likes. We report the results from these regressions in Table B4 in the appendix. With the exception of a positive effect in the log-specification for re-tweets, we find no evidence that the

[29]We also include a dummy for whether an episode was actually aired on the specific day.

[30]In the top panel of the table, we report the standard deviation of the residuals when regressing the dependent variable in the respective column onto show fixed effects. This metric reflects the amount of variability over time after controlling for cross-sectional differences across shows.

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	Number of Daily Comments	Log Number of Daily Comments	Number of Pre-Show Comments	Log Number of Pre-Show Comments	Rating	Log Rating
Type of Regression	OLS	OLS	IV 1st Stage	IV 1st Stage	IV 2nd Stage	IV 2nd Stage
Unit of Observation	Show/Day	Show/Day	Episode	Episode	Episode	Episode
Standard Deviation of DV (Control. for Show FEs)	47,503	1.304	33,679	1.463	0.158	0.090
Number of Pre-Show Comments (Unit: 10,000)					0.055***	
					(0.019)	
Log Number of Pre-Show Comments						0.018***
						(0.006)
Censor Dummy	-10,351**	-1.003***	-6,441**	-0.922***		
	(5,172)	(0.086)	(2,905)	(0.097)		
F-Stat on Censor Dummy	4.00	136.38	4.92	89.79		
Show FEs	Yes	Yes	Yes	Yes	Yes	Yes
Day of the Week Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	10,126	10,126	7,899	7,899	7,899	7,899
Shows	166	166	193	193	193	193

Table 4: **Instrumental Variable Regressions.** The unit of observation is a show/day combination in columns (1) and (2) and an episode in the remaining columns. Standard errors are clustered at the show level in all regressions. The regressions in column (1) and (2) also include a dummy for whether an episode of the show was aired on the specific day.

block affected any type of activity on Sina Weibo other than comments. We note that therefore, if anything, the effect of ratings per unit of comment might be larger than our estimate, because we find weak evidence of an uptake of re-tweets during the block (which might have caused ratings to decrease less than they otherwise would have). Finally, in Figure (4), we provide some graphical evidence that clearly illustrates the impact of the block and also provides a sense of the magnitude of the change in Sina Weibo activity during the block. To generate the graph, we regress each (log) activity measure onto show fixed effects and weekday dummies but do not include a dummy for the block. We then compute the residuals from this regression and plot the average value of the residuals over time. The blue line shows the very pronounced drop in the number of comments during the block.[31] Also, relative to the fluctuation in comments over the remainder of the time window, the block-induced decrease is large.

We next present the actual first stage of our IV specification, which we run at the episode level. To implement such a regression, we need to associate Sina Weibo activity over a specific time

[31]The graph also displays the (more modest) increase in log re-tweets (represented by the grey line) during the block.

period with each episode for which we have ratings data. To compute such a measure, we rely on the timestamps of comments and the starting time of each individual episode. More specifically, we define the number of relevant comments pertaining to a specific episode as the number of comments that mentioned the show on the same day before the particular episode aired. The results from these "episode-centric" regressions are reported in columns (3) and (4) and are very similar to their counterparts using daily data in the first two columns. We also note that, especially for the log-specification, the magnitudes of the F-statistic on the censor dummy (the instrumental variable) are considerable, which mitigates weak-instruments concerns (Rossi, 2014).

5.2 Second Stage

We report results from the second stage in columns (5) and (6) of Table 4 for the level and log specification, respectively. In our preferred log-log specification, we find a statistically significant coefficient of 0.018, and hence doubling the number of comments leads to a 1.8% increase in ratings. To gauge the magnitude of this effect, note the standard deviation of log ratings over time is equal to 9% (see Table 1); and hence doubling the number of comments leads to a movement in log-rating of 20% of a standard deviation. We also note the results are closely related to the regressions of log ratings on the block dummy presented early. In particular, column (1) of Table 2 represents the reduced form of the IV regression presented above. In the IV regression, the coefficient on the block in the first stage is equal to 0.922 and hence corresponds to a decrease of almost 100% in comments, which is exactly what we would expect due to the nature of the block. Moving log-comments by 0.922 in turn reduces log-ratings by 0.017. This number is identical to the coefficient on the block dummy in column (1) of Table 2. As a further benchmark for the effect magnitude, note the standard deviation of the average rating across all episodes is equal to 0.260 (see column (2) Table 1). Therefore, doubling the number of comments will lead to movement of 6.5% of a standard deviation of the ratings distribution (0.017/0.260). Results are very similar in the level-level specification as well. The block reduces the number of comments by 6,400, which in turn shifts ratings by $0.055 * 0.64 = 0.035$. This magnitude again represents about 20% of the standard deviation of the time series of the ratings variable (0.035/0.158).

The identifying assumption in this regression is closely related to our previous analysis, and we hence reiterate it only briefly here. We are identifying the causal effect of earned media in the form of comments on ratings, if the block only affected ratings because of its impact on the number of comments, and is therefore excluded from the second stage of the regression. The discussion in section 3 as well as the difference-in-differences regressions in the previous section provide evidence to support this assumption. Specifically, our earlier regressions document that shows in Hong Kong as well as low-Weibo-activity shows in mainland China experienced no change in their ratings. Had the block had any direct effect on TV viewership, we would have expected it to affect viewership of those shows as well. The absence of any change in ratings in those control groups thus provides support for the exclusion restriction.

5.3 Effect Magnitude

To assess the magnitude of the estimated effect, comparing it to other estimates of the impact of other marketing tools on product demand is instructive. Although relatively little research has addressed earned media,[32] a large literature looks at paid-media effectiveness and in particular on advertising effects primarily in the realm of TV advertising. For instance, Gordon and Hartmann (2013) report an elasticity of 0.034 for political advertising campaigns. The meta-analysis conducted by Sethuraman et al. (2011) reports a median elasticity of 0.05 and finds that across a large set of studies, about 40% of estimated elasticities lie between 0 and 0.05. One would of course expect to see variance in the effectiveness depending on the product category as well as the specific marketing tool, and hence we would not expect our estimate to necessarily replicate the ones found in the previous literature. Our estimate of 0.018 nevertheless is roughly similar to other estimates, although it is on the lower side with regards to previous findings. Despite the lower elasticity, earned media is still likely to be an attractive marketing tool because microblogging activity can most likely be generated at a lower cost. Therefore, in terms of return-on-investment, microblogging might look similar (or even favorable) relative to more traditional forms of marketing activity. Because firms control user-generated content generation only indirectly, making an assessment of the costs associated with generating an active social media following is difficult. We therefore do not attempt to translate our estimate into the corresponding return-on-investment.

Finally, our estimate isolates the elasticity of comments alone while holding all other types of activity constant. This is due to the unique features of our data, but typically the different types of activity will tend to move together. Hence, one would expect a larger elasticity when increasing all activity by a certain percentage, rather than just increasing comments. Therefore, we conclude that microblogging is similar in its effectiveness to traditional advertising and can potentially be delivered at a lower cost. Our findings are therefore encouraging with respect to the ability of user-generated microblogging content to increase product demand.

To further assess the magnitude and economic relevance of the effect of Sina Weibo activity on ratings, we investigate the impact of the decrease in comments on advertising revenue. In China, TV advertising is priced per second and rating point; therefore, total ad revenue at the episode level is given by $(ad\ price * ad\ duration * rating)$.[33] Because of the multiplicative structure of this pricing formula, the estimated 2% increase in ratings when doubling the number of comments will translate into a 2% increase in revenue from TV advertising. To also compute the absolute magnitude in monetary terms, we obtained additional data on ad duration as well as ad prices. Specifically, we complement our TV show data with data on the full set of ads that aired during the episodes contained in our data. Furthermore, we obtain advertising prices at the channel/hour level for the 20 channels in our sample during prime time. We find an average ad duration of

[32] Recently, Lovett and Staelin (forthcoming) estimate the elasticities of earned media conditional on *exposure* is 0.44. However, our estimate is on the overall volume instead of on exposure. So the two estimates cannot be easily compared.

[33] Advertising is shown across the entire mainland China for the set of channels in our sample. Hence, an ad has one national price, rather than the more granular media market-level pricing in the United States.

277 seconds, that is, around 4.5 minutes per episode. The average ad price is equal to US$3,700 per second and rating point.[34] Remember that in the level specification, we estimated an increase of 0.055 in ratings for an increase of 10,000 comments (see column (5) of Table 4). Therefore, adding an additional 10,000 comments leads to an increase in ad revenue per episode of US$57,000 ($adprice * adduration * \Delta rating = 3,700 * 277 * 0.055 = 57,000$). The average episode in our sample has 5,800 comments; hence, in the absence of this amount of activity, ad revenue would drop by about US$33,000 dollars per episode ($= 57,000 * 5,800/10,000$). However, great heterogeneity exists in the amount of microblogging activity across shows and episodes. A one-standard-deviation shift implies a change in the number of comments of about 44,000 (see Table 1),[35] which entails a change in advertising revenue per episode of US$250,000 (=57,000*44,000/10,000). Hence, the impact of microblogging on advertising revenue can be very substantial. We also note that advertising prices in the United States are higher; hence, a shift in microblogging activity in the US market will entail an even larger impact on advertising revenue.[36]

Finally, as another way to put the magnitude of the effect into context, we also translate the ratings change to the total amount of TV consumption by Chinese households in section 6.4 below.

6 Mechanism and Additional Results

6.1 Advertising Effect or Consumption Complementarity?

Microblogging activity, and more specifically comments, can lead to an increase in product demand in various ways. First, they could provide information by reminding the consumer of the show or persuade consumers that already know about the show to watch it. Both channels are similar to the dichotomy of informative versus persuasive advertising, which is often used as a taxonomy for paid media, that is, marketing efforts such as TV advertising, which are directly under the firm's control. Second, being able to engage in microblogging about the show while watching it might enhance consumers' viewing experience. In other words, microblogging might be a complementary activity with respect to TV consumption.

In this section, we test whether the effect we find is due to complementarity versus an informative or persuasive effect due to microblogging activity prior to the show. We refer to the latter channel as an "advertising effect" because it is similar to the way more traditional forms of advertising influence consumers. The key difference that will allow us to disentangle the two effects is the

[34] Advertising prices per ratings point are not directly reported. However, we have data on advertising expenditure (per second of ad duration) at the hour/channel level as well as data on average ratings. We can hence divide expenditure by ratings to obtain the cost per rating point. For our calculation, we use the average advertising price per rating point (and per second) across all 20 channels during prime time.

[35] Note the standard deviation *within* shows over time is equal to 34,000 and hence is not much smaller than the standard deviation across episodes.

[36] We implemented a back-of-the-envelope calculation comparing US and Chinese advertising prices and found the price difference (per rating point) to be roughly equivalent to the ratio of GDP between the United States and China. During our sample period in 2012, the ratio was equal to 2:1 (US relative to China).

following: if complementarity is the main driver, microblogging activity *during* the show has an impact on ratings; instead, in the case of an advertising effect, *pre*-show microblogging will be the primary driver of ratings changes. Ideally, we would want to observe exogenous shocks that differentially affect both types of microblogging activity. Unfortunately, the block led to a reduction of comments regardless of when they were posted and hence does not provide us with separate sources of variation to disentangle the two channels.

We therefore analyze the issue in the following way: For each show in our data, we compute the average number of comments per episode that were posted before and during the show.[37] The average number of pre-show comments was previously used to group shows according to their activity level on Weibo in section 4.2 and is defined as the number of comments pertaining to a specific show on the same day and before the specific episode of the show aired. We now also add to the analysis the number of comments posted during the episode, which is defined as the number of comments posted at any point in time while the episode was being aired. We proceed to analyze whether the block differently affected shows that typically have more activity before versus during the show.

Similar to the analysis in Table 3, we split the number of comments variable into three equally sized bins for both pre- and during-show comments. We then regress ratings onto a censor dummy as well as interactions terms with the activity level of the show in terms of both types of comments. We report results in the first three columns of Table 5, first separately for pre- and during-show comments and then jointly. As reported before, we find that shows with higher activity in terms of pre-show comments experienced a stronger decrease in ratings during the block.[38] We find no such pattern for shows with higher levels of commenting during the show. The difference between the two types of comments is particularly clear when both sets of variables are included in column (3). In this regression, high pre-show activity significantly predicts a stronger decrease in ratings. By contrast, the coefficients on the two dummies for during-show comments have a positive sign, are small in magnitude, and not significantly different from zero. We replicate the same type of analysis using an interaction term with the number of pre- and during-show comments as a continuous variable in columns (4) to (6) of the same table, and find similar results. The number of pre-show comments are the main driver of rating changes, whereas the amount of comments posted during the show are not predictive. We also note the distributions of pre- and during-show comments are relatively similar, and hence the coefficients for both types of comments are comparable in terms of the variance in the underlying variables. Specifically, pre- and during-show comments respectively have a mean of 4,826 (5,448) and a standard deviation of 28,300 (40,644).

We hence conclude that microblogging does operate in a similar fashion to other traditional marketing tools. The novel channel through which microblogging could in theory influence product

[37] For each show, the average number of pre- and during-show comments are computed across all episodes that did not air during the block.

[38] As one would expect, the two commenting variables are positively correlated. However, some amount of independent variation exists in each variable that allows us to include both in the same regression. The correlation coefficient of the two variables is equal to 0.67.

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	Log Rating	Log Rating	Log Rating	Log Rating	Log Rating	Log Rating
Censor Dummy	-0.004	-0.011**	-0.005	-0.011**	-0.015***	-0.011**
	(0.005)	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)
Medium Pre-show Activity	-0.013		-0.014			
× Censor Dummy	(0.011)		(0.010)			
High Pre-show Activity	-0.025**		-0.026*			
× Censor Dummy	(0.011)		(0.015)			
Medium During-show Activity		-0.000	0.005			
× Censor Dummy		(0.013)	(0.012)			
High During-show Activity		-0.017	0.002			
× Censor Dummy		(0.012)	(0.016)			
Pre-show Activity				-0.071***		-0.084***
× Censor Dummy				(0.023)		(0.021)
During-show Activity					-0.041**	0.020
× Censor Dummy					(0.016)	(0.034)
Show FEs	Yes	Yes	Yes	Yes	Yes	Yes
Day of the Week Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7,899	7,899	7,899	7,899	7,899	7,899
Shows	193	193	193	193	193	193

Table 5: **The Differential Impact of Pre- versus During-show Weibo Activity.** Pre- and during-show activity (the last two variables reported in the table) are measured in units of 100,000 comments.

demand, which is complementarity to TV consumption, does not turn out to be empirically relevant in our setting. We note that our analysis only shows that during-show comments have no *immediate* effect on ratings. However, one could imagine the ability to engage in microblogging during the show deepens consumer engagement and might thus increase repeat viewing. For instance, the findings in Joo et al. (2014) regarding online search being triggered by TV ads suggests such longer-term interaction effects between different types of media might exist. Although we find no dynamic effects of the censorship block on ratings (see section 6.3 below) in our setting, we cannot fully rule out such a mechanism.

6.2 Content Analysis

Furthermore, with respect to informative versus persuasive effects of microblogging, we believe that due to the nature of comments, a persuasive effect is more likely. A show's official account would typically send a reminder tweet regarding a specific episode and such activity was not blocked due to the censorship. Therefore, comments are mostly used to express enthusiasm about a specific episode and can hence be understood as more persuasive in nature.

In order to provide systematic evidence to support this assertion, we complement our data with an analysis of the content of a random subsample of comments. Specifically, we randomly sample 1,000 comments from the full set of comments[39] in our data and categorized comments into either "informative" or "persuasive", whereas the latter category is defined as comments that express sentiment about the show (e.g. "awesome show, cannot wait!"). More detail on how the sample was collected and the coding of comments into different categories are provided in Section C of the appendix. Consistent with our expectations, we find that 48.3% of comments are persuasive and only 5% are informative. Examples for an informative and persuasive comment are respectively: "It is rebroadcasted every morning at 8am" (posted under the tweet "Don't miss [Daily Express] at 11pm tonight") and "The preview makes me hungry, lol. Strongly recommend the show to all @food_lovers and who @love2eat!". The remainder of comments which are neither informative nor persuasive contain information not directly related to the show (e.g. "Oh no, I am stuck in traffic and might not make it home in time to watch the show").[40]

Based on this analysis of comment content, we conclude that comments primarily affect viewing behavior via a persuasive effect. We also note that the majority, about 80% percent, of persuasive comments express positive sentiment and only a small fraction of comments express negative or mixed sentiment about a show.

6.3 Dynamics and Long-Term Effects

Next, we explore whether the block had any long-term effects beyond decreasing ratings during its duration. In our context, dynamic effects could arise from two sources. First, state dependence could be present in microblogging behavior. For instance, the lower activity level during the block could make engaging in microblogging after the block when Sina Weibo was fully functional again less attractive for users. Such an effect could arise, for example, if users care about the involvement of other users on the same topic and use past activity as a measure for potential future engagement and readership on the topic. Other research has shown such audience effects exists on user-generated content platforms (see, e.g., Zhang and Zhu 2011, Shriver et al. 2013, and Toubia and Stephen 2013). A second mechanism that could lead to a longer-term impact of the censorship is a dynamic response in ratings to the decrease in microblogging similar to a goodwill stock advertising model. In other words, the decrease in current microblogging will decrease the goodwill stock and thus lead to a prolonged effect of the block on ratings.

To assess both of these aspects, we regress show-level ratings and comments onto the censor dummy (and the usual control variables) and also include a further series of dummies for five three-day intervals immediately after the block. We note the block lasted for three days and hence

[39] We note that our full data set contains millions of comments and simply retrieving each tweet (without the content) with the corresponding timestamp and count of comments, re-tweets and likes took a professional company about 2 months to collect. We are therefore not able to scale the content analysis up. However, we believe that the random subsample provides a useful additional piece of evidence in understanding the way in which comments affect consumer viewing behavior with regards to TV shows.

[40] The split between informative and persuasive content is different for original tweets. Based on sampling 1,000 random tweets we find that 57.4% are informative while only 28.8% express sentiment.

	(1)	(2)
Dependent Variable	Log-Rating	Log-Comments
Censor Dummy	-0.017***	-0.914***
	(0.006)	(0.101)
Days 1-3 after the Block	-0.012**	-0.175**
	(0.006)	(0.077)
Days 4-6 after the Block	-0.005	0.205*
	(0.008)	(0.113)
Days 7-9 after the Block	0.006	0.157
	(0.005)	(0.101)
Days 10-12 after the Block	-0.004	-0.274**
	(0.006)	(0.110)
Days 13-15 after the Block	0.003	-0.083
	(0.005)	(0.097)
Show FEs	Yes	Yes
Day-of-the-Week Dummies	Yes	Yes
Observations	7,899	7,899
Shows	193	193

Table 6: **Dynamic Effects.** The unit of observation is an episode. Standard errors are clustered at the show level.

the post-block time intervals the dummies capture are of the same length as the block itself.[41] Results from both regressions, which are reported in Table 6, reveal some interesting results. In terms of the movement in ratings, we find ratings remain at a lower level in the first three days after the block and then return back to their "normal" level. The magnitude in the three days following the block is equal to about two thirds of the the decrease during the block. None of the four other time dummies for later time periods are statistically significant. For the regression using comments as the dependent variable, we also find a significant and negative effect in the first three days after the block. However, in the case of comments, the decrease after the block is much smaller in magnitude relative to the more substantial decrease during the block. Furthermore, when inspecting all coefficients in column (2) of Table 6, we find that several coefficients are statistically significant and have similar magnitude, some of them with a positive and others with a negative sign. The comments time series therefore seems to fluctuate more, and the negative coefficient during the three days immediately after the block constitutes a fairly typical movement in terms of its magnitude. The effect during the censorship event instead is substantially larger in magnitude, as Figure 4 presented earlier shows. We therefore conclude the evidence for a lasting decrease in comments after the block is weak.

[41] We also note results are very similar when we run the analysis at the day level rather than using three-day blocks. For ease of exposition, we report only results for the more aggregated version using dummies for consecutive three-day intervals.

Taken together the results above show that microblogging activity can lead to a longer-lasting impact on ratings. However, the longer-term effects are not caused by dynamics in microblogging activity, but are more likely due a dynamic effect of current commenting onto future ratings. These patterns also lend some further support to the dominance of the "advertising effect" over complementarity of microblogging and TV consumption discussed in the previous paragraph. If complementarity was the main driver, the immediate return of comments to its original level should also lead to ratings immediately returning to their regular levels. Instead, an advertising effect can lead to a prolonged decrease in ratings, even if comments do not remain at a lower level. An alternative explanation for the observed pattern is state dependence in show viewership. Consumers that missed an episode of a show during the block might be less likely to return to watch the show at a later point in time.

6.4 Substitution Patterns

One interesting question in our context is whether the decrease in viewing for shows due to the block leads to consumers substituting to other shows or whether it reduces overall TV viewership. Part of our previous analysis exploiting differences in the effect magnitude across shows with different amounts of Sina Weibo activity can shed some light on this issue. Specifically, the results presented in Table 3 provide evidence that while ratings for shows with high Sina Weibo activity decreased, other types of shows experienced a more modest or no decrease. However, we do not observe an increase in ratings for any type show. Hence, for the set of shows included in our data, no shows seem to benefit from the block by absorbing the decrease in viewership for the high-Sina-Weibo activity shows.

However, shows that our sample does not contain might have seen their ratings increase during the block because of viewers substituting across shows. Although we observe ratings for all shows on the top 20 *national* channels during prime time (6 p.m. to 12 a.m.), a non-negligible amount of local TV channels exist whose ratings we do not observe. Unfortunately, we do not have ratings data at the show level for the whole universe of local and national channels. However, we do observe the share of consumers watching *any* TV during a particular time window. Although this information provides us only with one time series of the share of total viewership over time, we can nevertheless use it to assess whether any discernible decrease occurs during the block.

When regressing daily (log) total viewership[42] during prime time on a dummy for the block as well as weekday dummies, we find a significant coefficient of -0.027 (standard error of 0.010).[43] Interestingly, this magnitude of this effect is similar to the estimated effect on ratings at the show level (see, e.g., column (2) of Table 2). In other words, the percentage decrease in total viewership

[42]The total viewership data are reported in a similar fashion as the show-level rating data. We observe the total share of consumers watching TV in 24 cities and aggregate the city-level data into the national share of viewership using population weights. The total viewership data is recorded at the minute level; here we use the average daily share for the prime-time window from 6 p.m. to 12 a.m.

[43]This regression also includes a liner time trend because total viewership experienced a seasonal decrease during our sample period. Without the linear time trend, the coefficient (standard error) on the block dummy is equal to -0.027 (0.020).

during the block is similar to the percentage decrease in ratings for the shows in our sample. This finding suggests the block resulted in consumers watching less TV rather than substituting to other shows. When using the viewership share in levels as the dependent variable, we also obtain a significant coefficient of -0.808 (standard error is equal to 0.295).

As another way to gauge the magnitude of our effect, we can also translate the overall drop in TV viewing (expressed in percentage terms above) into the total household-hours change in TV viewership during prime time. This metric is easy to calculate by scaling the 0.808-percentage-point drop in ratings up to the population-hour level. The average household experienced a reduction of 0.05 hours ($0.00808 * 6 = 0.05$), that is, about three minutes. At the country level, for the entire mainland China, which comprises 456 million households based on 2012 data, this number corresponds to a decrease of 23 million hours in total TV consumption ($0.05 * 456 = 23$). We also note that prime-time TV makes up the bulk of total TV viewing, and presumably most shows with high Sina Weibo activity air during that time. Therefore, when we estimate the effect of the block on viewership using data for the entire day, we find a smaller coefficient on the block of -0.269 (standard error is equal to 0.122).[44] This finding translates into a per-household reduction of 0.065 hours ($0.00269 * 24 = 0.065$), which is only marginally larger than the magnitude we find for prime time alone.

6.5 Effect Heterogeneity

We further explore two other potentially important dimensions of heterogeneity in the estimated effect, namely, across different genres of shows and new versus old shows. However, we find no evidence that the block led to a differential decrease in ratings along those dimensions. In other words, the block affected shows similarly across genres as well as vintage. We present the results of that analysis in the appendix.

7 Conclusion

Promoting products through social media websites such as Twitter has become a common practice for many firms, but whether this new media channel can effectively enhance demand remains ambiguous. Similar to evaluating the effectiveness of other types of marketing activity, measuring the effect of microblogging using field data where the correlation between microblogging activity of a product and its demand does not necessarily imply causality is challenging. Furthermore, obtaining credible evidence on the impact of microblogging activity by individual users, rather than the firms themselves, is particularly hard because the firm does not directly control earned media. In this paper, relying on a natural experiment, we provide a clean empirical identification of the effect of earned media on product demand. In the context of TV show ratings, we demonstrate that user-generated microblogging activity significantly increases product demand.

[44] This coefficient is for the level specification and hence should be compared with the equivalent specification for prime time only, which yielded a coefficient of -0.808.

The natural experiment we exploit for identification is the following: in March of 2012, Sina Weibo partially shut down its website for three days due to political events, blocking users' ability to comment on other users' tweets. Relying on this censoring event, we investigate how ratings of TV shows were affected when their microblogging activity exogenously decreased. Using a series of difference-in-differences estimators, we show the decrease in ratings only occurred for shows that had more intensive Weibo activity outside the block. Furthermore, TV shows in Hong Kong, where consumers use Twitter rather than Sina Weibo as their microblogging platform, were unaffected. Similarly, TV viewership was also less affected in mainland cities where Sina Weibo had low penetration rates. Using the block as an instrumental variable, we also quantify the magnitude of the effect of commenting activity on TV show ratings per unit of comment, and find an elasticity of 0.02 of TV show ratings with respect to Sina Weibo comments. In terms of the behavioral mechanism underlying our estimated effect, we find earned media works similarly to traditional advertising such that more user-generated content *before a show* leads to more consumption. By contrast, microblogging activity *during the show* does not increase its viewership. In other words, the concurrent consumption of TV shows and the creation of related microblogging content do not appear to be complementary to each other.

Several extensions beyond this research are possible. For example, a study of the interaction between paid media and earned media on microblogging websites, and how a firm's paid media on the platforms can influence earned media, would be interesting. Furthermore, earned media through microblogging may differ in their sentiment pertaining to a product. Further disentangling the effects of positive, negative, and neutral user-generated content would therefore be interesting.

References

Aaltonen, A. and S. Seiler (2015): "Cumulative Growth in User Generated Content Production: Evidence from Wikipedia," *Management Science*, forthcoming.

Bertrand, M., D. Karlan, S. Mullainathan, E. Shafir, and J. Zinman (2010): "What's Advertising Content Worth? Evidence from a Consumer Credit Marketing Field Experiment," *The Quarterly Journal of Economics*, 125, 263–306.

Blake, T., C. Nosko, and S. Tadelis (2014): "Consumer Heterogeneity and Paid Search Effectiveness: A Large Scale Field Experiment," *Econometrica*, forthcoming.

Chevalier, J. A. and D. Mayzlin (2006): "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 43, 345–354.

Godes, D. and D. Mayzlin (2004): "Using Online Conversations to Study Word-of-Mouth Communication," *Marketing Science*, 23, 545–560.

Gong, S., J. Zhang, P. Zhao, and X. Jiang (2015): "Tweeting Increases Product Demand," *Working Paper*.

Gordon, B. R. and W. R. Hartmann (2013): "Advertising Effects in Presidential Elections," *Marketing Science*, 32, pp. 19–35.

Hartmann, W. R. and D. Klapper (2015): "Super Bowl Ads," *Working Paper*.

Joo, M., K. C. Wilbur, B. Cowgill, and Y. Zhu (2014): "Television Advertising and Online Search," *Management Science*, 60, pp. 56–73.

Lambrecht, A., C. Tucker, and C. Wiertz (2015): "Should You Target Early Trend Propagators? Evidence from Twitter," *Working Paper*.

Lewis, R. A. and J. M. Rao (forthcoming): "The Unfavorable Economics of Measuring the Returns to Advertising," *Quarterly Journal of Economics*.

Lovett, M. J. and R. Staelin (forthcoming): "The Role of Paid, Earned, and Owned Media in Building Entertainment Brands: Reminding, Informing, and Enhancing Enjoyment," *Marketing Science*.

Rossi, P. E. (2014): "Invited PaperÑEven the Rich Can Make Themselves Poor: A Critical Examination of IV Methods in Marketing Applications," *Marketing Science*, 33, 655–672.

Sahni, N. (2015a): "Advertising Spillovers: Field Experimental Evidence and Implications for Returns from Advertising," *Working Paper*.

——— (2015b): "Effect of Temporal Spacing between Advertising Exposures: Evidence from an Online Field Experiment," *Working Paper*.

Sethuraman, R., G. J. Tellis, and R. A. Briesch (2011): "How Well Does Advertising Work? Generalizations from a Meta-analysis of Brand Advertising Elasticities," *Journal of Marketing Research*, 48, pp. 457–471.

Shapiro, B. T. (2014): "Positive Spillovers and Free Riding in Advertising of Prescription Pharmaceuticals: The Case of Antidepressants," *Working Paper*.

Shriver, S. K., H. S. Nair, and R. Hofstetter (2013): "Social Ties and User-Generated Content: Evidence from an Online Social Network," *Management Science*, 59, 1425–1443.

Sonnier, G. P., L. McAlister, and O. J. Rutz (2011): "A Dynamic Model of the Effect of Online Communications on Firm Sales," *Marketing Science*, 30, 702–716.

Stephen, A. T. and J. Galak (2012): "The Effects of Traditional and Social Earned Media on Sales: A Study of a Microlending Marketplace," *Journal of Marketing Research*, 49, 624–639.

Toubia, O. and A. T. Stephen (2013): "Intrinsic vs. Image-Related Utility in Social Media: Why Do People Contribute Content to Twitter?" *Marketing Science*, 32, 368–392.

Trusov, M., A. Bodapati, and R. E. Bucklin (2010): "Determining Influential Users in Internet Social Networks," *Journal of Marketing Research*, 47, 643–658.

Zhang, X. M. and F. Zhu (2011): "Group Size and Incentives to Contribute: A Natural Experiment at Chinese Wikipedia," *American Economic Review*, 101, 1601–15.

A Appendix: Data Collection

During our data collection, Sina Weibo changed the way in which tweets are displayed when querying the webpage for a specific keyword. Specifically, after June 2015, the number of tweets Sina Weibo displayed for a given keyword was smaller than the number of tweets we obtained when implementing the identical query prior to June 2015. We were not able to obtain any official statement regarding the motivation or the specific nature of the change. However, by June 2015, we had completed most of our data collection, and the change therefore has relatively little impact on us. In the paragraphs below, we outline in detail how we deal with adjusting the pieces of data that we obtained after June 2015, in order to make these data comparable to the older data and to "offset" the reduction in the number of tweets being displayed after June 2015.

Prior to June 2015, we collected the number of tweets (but not re-tweets, comments, and likes) at the daily level for all shows in our sample. We collected only the aggregated number of tweets, and did not obtain data on individual tweets with their corresponding timestamp. Obtaining the daily number of tweets for a keyword used to be easy because Sina Weibo displayed the total number of tweets when being queried for a specific keyword and time window. Collecting each tweet and then computing the total number manually was therefore unnecessary (this feature of the webpage has also disappeared since our original data collection). Later, because we wanted to compute the number of tweets prior to a specific episode airing, we went back to the data to scrape each tweet (with its timestamp) individually. By this time, Sina Weibo had changed the way results were displayed, and the total number of daily tweets was different relative to our previous data collection.

Fortunately, the fact that we scraped data before and after the change allows us to assess the magnitude of the change and to adjust the new data to make it comparable to the older data. Whenever possible, we use the data we initially collected directly because it contains the full set of tweets. For instance, the daily number of tweets was obtained before the change. However, to compute the pre-show tweets on a specific day, we need to use the timestamp on each tweet, which only the newer data contain. We hence rescale the new data by the ratio of daily tweets between the two data sets. For example, assume the new data contain 100 daily tweets for a specific show, 80 of which were posted before the show. The old data instead contain 200 daily tweets and hence the new data only display half of them. In this case, we assume the total number of pre-show tweets is 160. This calculation is based on the assumption that the subset of tweets Sina Weibo selects is random with respect to the time of the day at which the tweet is posted. We think this assumption is likely to hold. We explored, for instance, whether the daily ratio of tweets between the two data sets varied across the 60 days (two months) contained in our sample and found they did not (when regressing the daily ratio of tweets onto date fixed effects, the date fixed effects are jointly insignificant). We therefore think any systematic selection of the tweets being displayed are unlikely to occur within a day.

We employ a similar type of rescaling to obtain the number of pre-show comments and re-tweets. Similar to the logic for computing pre-show tweets, the rescaling is based on the assumption that the algorithm that selects which tweets to display does not select tweets based on the number of comments and/or re-tweets.

B Effect Heterogeneity

In this section, we further explore whether the censorship event affected certain types of shows more or less. We first investigate whether the content of the specific show affects the degree to which ratings react to microblogging activity. To this end, we generate a set of dummies for the specific genre to which a show belongs. From our entire set of 166 shows, we generate the following four categories: TV drama series (19 shows), children's shows (15), current events (36), and reality shows (40). All other shows belong to a diverse set of other genres such as music, lifestyle shows, historical documentaries, and so on, and we hence treat all remaining shows as a residual category. Using our new show-type definitions, we proceed to run a regression of ratings on four interaction terms of the censor dummy with a dummy for each of the four main categories described above. The censor dummy coefficient therefore captures the effect of the block on the shows belonging to the residual category.

We report results from this regression in column (1) of Table 7. We find little evidence of any heterogeneity in the effect across categories or possible substitution. Only the dummy on the "current events" category is marginally significant. When testing for the joint significance of all four category interaction terms, we are unable to reject the hypothesis that all four are equal to zero. Interestingly, when including the show-type interactions together with dummies for different levels of Sina Weibo activity, among the two dimensions of show characteristics, we find the activity level matters in terms of effect magnitude, but show type does not. We report the results from this regression in column (3) of Table 7. Column (2) replicates the results from a regression of ratings on activity-level interactions with the censor dummy (this regression is identical to column (1) of Table 3). The results in column (3) show the activity level continues to predict the effect magnitude of the block regardless of the genre of the show. The coefficients on the activity interaction terms are almost unchanged, whereas the estimated coefficients on the genre interactions are all insignificant. In other words, the block had a stronger impact on shows *within* each genre with higher activity levels, providing further evidence that the type of show does not matter for how effective microblogging activity is in increasing ratings.

We also explore whether the block had a stronger impact on new shows. One might expect such an effect, for instance, if informative effects are important and new shows might therefore be more reliant on microblogging activity as a means to inform consumers about the show. However, across a set of different regression specifications, we do not find any evidence of newer shows being more or less affected by the block. For instance, when including an interaction term of the block with a dummy for whether the show is within its first five episodes into a regression of (log) ratings on the

	(1)	(2)	(3)
Dependent Variable	Log-Rating	Log-Rating	Log-Rating
Censor Dummy	-0.008	-0.004	0.001
	(0.007)	(0.005)	(0.009)
Medium Weibo Activity		-0.013	-0.011
× Censor Dummy		(0.011)	(0.009)
High Weibo Activity		-0.025**	-0.023**
× Censor Dummy		(0.011)	(0.011)
TV Drama Series	-0.016		-0.006
× Censor Dummy	(0.021)		(0.022)
Children Show	0.005		0.003
× Censor Dummy	(0.010)		(0.009)
Current Events	-0.016*		-0.014
× Censor Dummy	(0.009)		(0.009)
Reality Show	-0.012		-0.005
× Censor Dummy	(0.017)		(0.017)
Show FEs	Yes	Yes	Yes
Day-of-the-Week Dummies	Yes	Yes	Yes
Observations	7,899	7,899	7,899
Shows	193	193	193

Table 7: **Effect Heterogeneity Across Genres.** The unit of observation is an episode. Standard errors are clustered at the show level.

block dummy, we find a coefficient (standard error) on the interaction term of 0.0025 (0.0138).[45]

C Content Analysis: Data Collection

To analyze microblogging content, we randomly sampled 1,000 tweets and 1,000 comments. The sampling procedure occurred as follows. First, we set the probability of a show being drawn equal to the ratio of the number of comments pertaining to that specific show relative to the number of comments pertaining to all shows in our sample. If a show is drawn, we then retrieve the set of all the tweets that mention the show's name and have at least one comment. From the set of tweets that qualify under this condition, we randomly sample one tweet. Finally, we randomly pick one comment from all the comments posted in response to the sampled tweet. This procedure yields a final sample of 1,000 comments, paired with 1,000 original tweets.

We define a comment or tweet as "informative" if it contains show information such as time to be aired, actors/actresses to be featured, story plot, story behind the scenes, and so on. A

[45]The coefficient (standard error) on the block dummy is equal to -0.0171 (0.0057). Other specifications, such as an interaction of the episode count (as a continuous variable) with the block dummy, yield similarly insignificant results.

comment or tweet is considered to represent "sentiment" if it expresses an opinion, attitude, or feeling toward the show (e.g., "awesome show, cannot wait!," "worst show ever!"). A comment or tweet can fall into both categories; that is, the two types are not mutually exclusive. Depending on the valance of sentiment, a comment was further coded into category: "positive," "negative," and "mixed" sentiment. Two research assistants who are native Chinese and blind to the research purpose coded comments and tweets, and resolved any disagreements by discussion.

We find the following results. First, only 5% of the comments are informative, whereas 48.3% express sentiments. The remainder contain information not directly related to the show (e.g., "Oh no, I am stuck in traffic and might not make it home in time to watch the show"). Out of all comments that fall into the sentiment category, 80.7% express positive sentiment, 15.1% express negative sentiment, and the remainder are mixed. For original tweets, the results are as follows: 57.4% of tweets are informative, whereas only 28.8% express sentiments. Out of all tweets in the latter category, 70.1% express positive sentiment, 24.3% express negative sentiment, and the remainder are mixed.

D Appendix: Additional Tables and Figures

	Mean	S.D.	90th Perc.	95th Perc.	Max	S.D. (Time Series Only)
Rating	2.85	6.29	12.57	19.78	36.17	0.67
Log-rating	0.74	0.90	2.61	3.03	3.61	0.15

Table B1: **Descriptive Statistics: Show Ratings in Hong Kong.**

	(1)	(2)	(3)	(4)	(5)
Dependent Variable	Log-Rating	Log-Rating	Log-Rating	Log-Rating	Log-Rating
Censor Dummy	-0.0168***	-0.0168***	-0.0168***	-0.0168***	-0.0163***
	(0.0053)	(0.0053)	(0.0053)	(0.0053)	(0.0052)
Log Search Index		-0.0003		0.0002	
"Bo Xilai"		(0.0013)		(0.0024)	
Log Search Index			-0.0010	-0.0011	
"Wang Lijun"			(0.0021)	(0.0033)	
Dummy March 14-18					0.0038
(Bo Xilai Removed from Office)					(0.0036)
Dummy April 10-17					0.0025
(Bo Xilai Arrested)					(0.0042)
Show FEs	Yes	Yes	Yes	Yes	Yes
Day-of-the-Week Dummies	Yes	Yes	Yes	Yes	Yes
Observations	7,899	7,899	7,899	7,899	7,899
Shows	193	193	193	193	193

Table B2: **Robustness Check: Controls for Saliency of the Political Scandal.** The unit of observation is an episode. Standard errors are clustered at the show level.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Type of Analysis	Mainland & HK		24 Mainland Cities				Across Shows	
Dependent Variable	Log-Rating	Log-Rating	Log-Rating	Log-Rating	Log-Rating	Log-Rating	Log-Rating	Log-Rating
Censor Dummy	0.0052	0.0055	-0.0098	-0.0096	-0.0079	-0.0077	-0.0039	-0.0042
	(0.0104)	(0.0104)	(0.0065)	(0.0064)	(0.0062)	(0.0061)	(0.0052)	(0.0052)
Medium Activity × Censor Dummy							-0.0129	-0.0125
							(0.0106)	(0.0106)
High Activity × Censor Dummy							-0.0246**	-0.0239**
							(0.0115)	(0.0112)
Mainland Dummy × Censor Dummy	-0.0260**	-0.0262**						
	(0.0118)	(0.0117)						
SW Penetration × Censor Dummy			-0.0265*	-0.0266*				
			(0.0142)	(0.0142)				
>Median SW Penet. × Censor Dummy					-0.0163***	-0.0163***		
					(0.0059)	(0.0059)		
Time Trend		-0.00036		0.00006		0.00008		-0.00023
		(0.00025)		(0.00019)		(0.00017)		(0.00023)
Mainland Dummy × Time Trend		0.00046						
		(0.00029)						
Sina Weibo Penet. × Time Trend				0.00013				
				(0.00035)				
>Median SW Penet. × Time Trend						0.00001		
						(0.00012)		
Medium Activity × Time Trend								0.00031
								(0.00033)
High Activity × Time Trend								0.00082**
								(0.00040)
Show FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Weekday Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	11,427	11,427	189,576	189,576	189,576	189,576	7,899	7,899
Shows	325	325	193	193	193	193	193	193

Table B3: **Robustness Check: Differential Time Trends.** The unit of observation is an episode in columns (1), (2), (7), and (8) and an episode/city combination in columns (3) to (6). Standard errors are clustered at the show level.

	(1)	(2)	(3)	(4)
Dependent Variable	Number of Tweets	Number of Re-tweets	Number of Comments	Number of Likes
Standard Deviation of DV (after Controlling for Show FEs)	4,182	2,896,767	47,503	10,202
Censor Dummy	3.027	12,080	-10,351**	-180.841
	(177.480)	(33,140.867)	(5,172.914)	(265.899)
Show FEs	Yes	Yes	Yes	Yes
Day-of-the-Week Dummies	Yes	Yes	Yes	Yes
Observations	10,126	10,126	10,126	10,126
Dependent Variable	Log Number of Tweets	Log Number of Re-Tweets	Log Number of Comments	Log Number of Likes
Standard Deviation of DV (after controlling for show FEs)	0.824	1.571	1.304	0.716
Censor Dummy	0.072	0.256***	-1.003***	0.014
	(0.048)	(0.090)	(0.086)	(0.050)
Show FEs	Yes	Yes	Yes	Yes
Day-of-the-Week Dummies	Yes	Yes	Yes	Yes
Observations	10,126	10,126	10,126	10,126

Table B4: **The Effect of the Block on Different Types of Activity on Sina Weibo.** The unit of observation is a show/day combination. Standard errors are clustered at the show level.

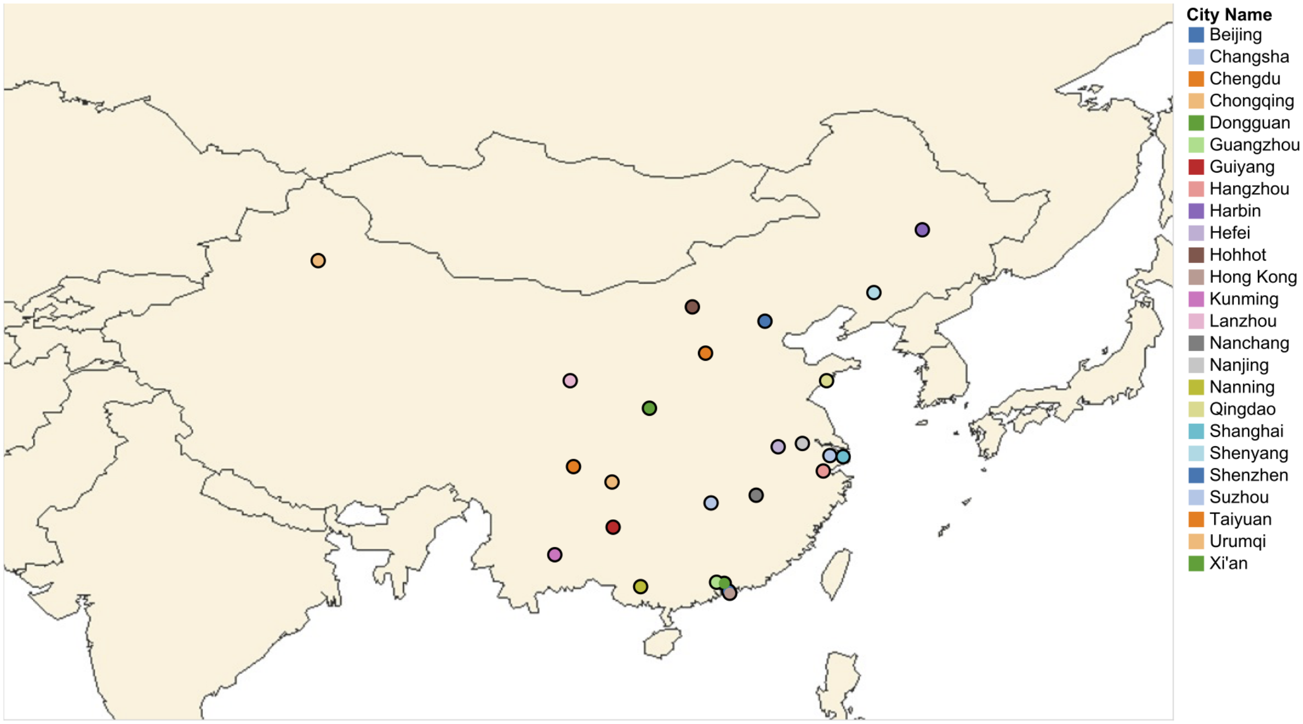
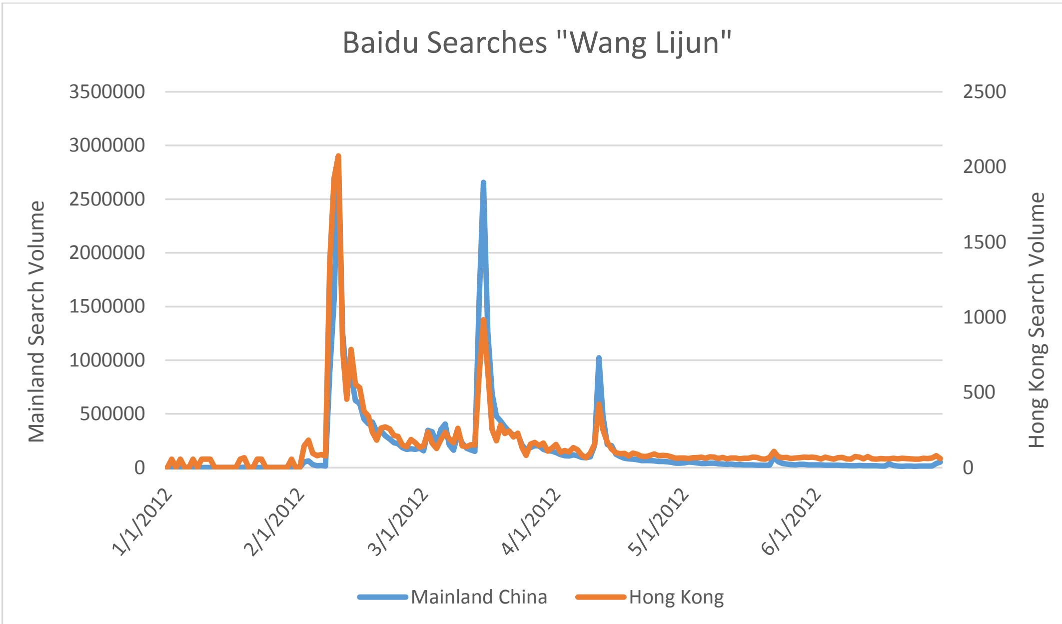


Figure B1: **Geographical Distribution of the Cities in Our Data: 24 Mainland Cities and Hong Kong**



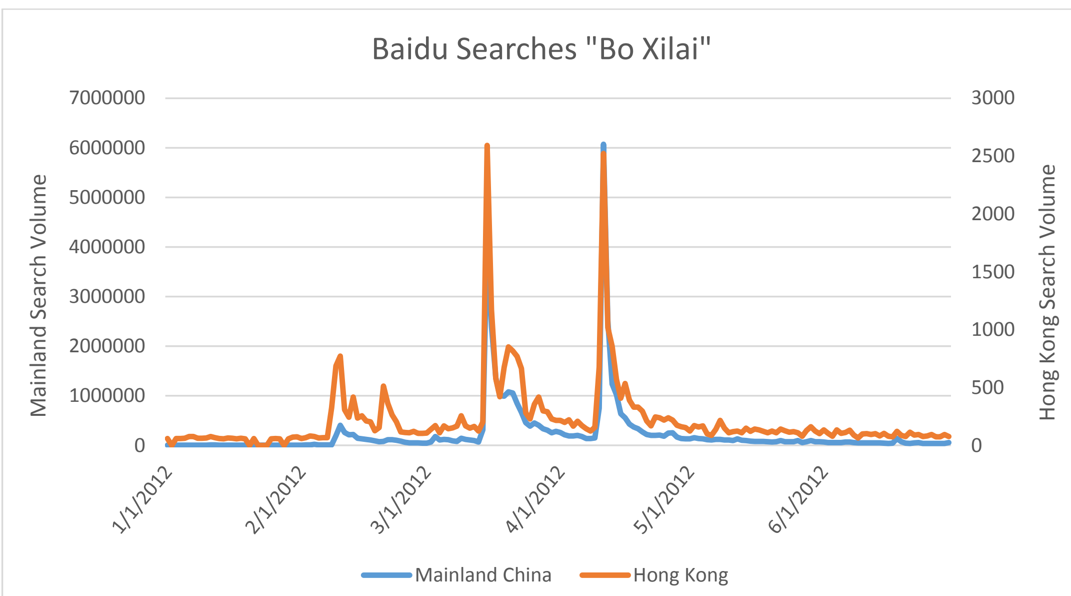


Figure B2: **Search Indices Comparison between Mainland China and Hong Kong.**