

# Does Online Word-of-Mouth Increase Demand? (and How?)

## Evidence from a Natural Experiment\*

**Stephan Seiler**

Stanford University

**Song Yao**

Northwestern University

**Wenbo Wang**

Hong Kong University of  
Science and Technology

This draft: March 23, 2017

We leverage a temporary block of the Chinese microblogging platform Sina Weibo due to political events, to estimate the causal effect of online word-of-mouth content on product demand in the context of TV show viewership. Based on this source of exogenous variation, we estimate an elasticity of TV show ratings (market share in terms of viewership) with respect to the number of relevant comments (comments were disabled during the block) of 0.016. We find more post-show microblogging activity increases demand, whereas comments posted prior to the show airing do not affect viewership. These patterns are inconsistent with informative or persuasive advertising effects, and suggest complementarity between TV consumption and anticipated post-show microblogging activity.

**Keywords:** Microblogging, Advertising, Social Media, Word of Mouth, Natural Experiment

\*We thank Tomomichi Amano, Eric Anderson, Yuxin Chen, Pradeep Chintagunta, Jörg Claussen, JP Dube, Jun Kim, Pedro Gardete, Brett Gordon, Wes Hartmann, Ayelet Israeli, Anja Lamrecht, Carl Mela, Blake McShane, Harikesh Nair, Sridhar Narayanan, David Reiley, Navdeep Sahni, Brad Shapiro, Ken Wilbur, Florian Zettelmeyer, Juanjuan Zhang, and seminar participants at London Business School, INSEAD, Duke University, Temple University, University of British Columbia, University of California at San Diego, University of Colorado, University of Florida, University of Minnesota, and Stanford GSB, as well as conference participants at the Workshop on Economics of Advertising and Marketing (Vilnius), Greater-China Conference on Mobile Big Data Marketing (Hong Kong), Big Data Conference (NYU), Searle Conference on Internet Commerce and Innovation (Northwestern Law), Marketing Science (Baltimore), Quantitative Marketing and Economics Conference (Northwestern), 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, HKUST IEMS Research Grant #16BM07, and the McManus Research Grant at the Kellogg School of Management, Northwestern University. We thank Ipetram Inc., Zhanbo Zhao, and Rio Luk for their help with data collection. 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, the exchange of information between consumers has become increasingly public, and word-of-mouth (WOM), which used to be confined to a small circle of contacts in the offline world, has tremendously increased in visibility in the online context. The proliferation of WOM via social media platforms such as Facebook or Twitter begs the question of how strongly brand-related online WOM can affect product sales. One obstacle to answering this question is the fact that causal inference is particularly difficult in the realm of online WOM due to the fact that firms are not directly in control of the amount of WOM. Therefore, randomized field experiments, which have increased in prominence in studies of advertising effectiveness (Lewis and Reiley 2014, Blake et al. 2015, Sahni 2015, Gordon et al. 2016), are difficult to implement in the case of WOM.

In this paper, we leverage a natural experiment, the temporary shutdown of the most popular Chinese microblogging outlet Sina Weibo due to political events, to estimate the causal effect of online WOM on product demand in the context of TV show viewership. Based on this source of exogenous variation, the paper makes the following contributions. First, we quantify the *causal* impact of WOM on product sales. Second, we analyze and identify the behavioral mechanism underlying the effect with a particular focus on complementarities between microblogging and TV consumption, a mechanism that has received little attention in previous work.<sup>1</sup>

Two key findings emerge from our analysis. First, we find an elasticity of TV viewership with respect to microblogging activity of 0.016, which is substantially lower than most estimates in the previous literature. For instance, Sonnier et al. (2011), Dhar and Chang (2015), and Liu (2015) estimate WOM elasticities between 0.59 and 1.04 across a variety of markets.<sup>2</sup> We suspect that part of this discrepancy can be attributed to the fact that previous papers were limited in their ability to deal with the endogeneity issues, due to the absence of (quasi-) experimental variation. Our estimated elasticity is slightly smaller than elasticities typically estimated for TV advertising of around 0.03 (Gordon and Hartmann (2013), Shapiro (2016), Tuchman (2016)), rather than substantially larger as other estimates would suggest. This finding is timely due to the recent increase in marketing spending on social media and a general belief among marketing practitioners that WOM can be a highly effective way to reach customers. For example, according to a study by the American Marketing Association, 64% of marketing executives believe WOM is the most effective form of marketing.<sup>3</sup> Our findings caution against overestimating the effectiveness of WOM relative to other marketing channels.<sup>4</sup>

---

<sup>1</sup>One notable exception is Lovett and Staelin (2016), who estimate a structural model that allows for complementarity between WOM and TV viewing.

<sup>2</sup>The meta-analysis of You et al. (2015) finds an average elasticity of 0.2. For the subset of elasticities that share characteristics with our study (e.g., private consumption), the average elasticity is roughly equal to 0.6. Lovett and Staelin (2016) estimate a more moderate elasticity of 0.04.

<sup>3</sup>“Marketers Say ‘Word of Mouth Marketing’ Is More Effective than Traditional Marketing; Forecast Big Increase in Social Media Spending,” November 19, 2013, *PR Newswire*.

<sup>4</sup>Our comparison is based on the effectiveness of “per unit of WOM” and does not take the cost of providing WOM into account. Despite a lower elasticity, WOM could compare favorably in terms of return-on-investment to TV advertising, due to cost differences.

The second key contribution of the paper is to identify the behavioral mechanism by which microblogging affects consumers' TV viewing decisions. We do so by exploiting detailed information on the timing of microblogging activity as well as an extensive content analysis. In a first step, we distinguish two possible behavioral channels that could drive the effect of WOM on product demand. WOM could serve as a complementary activity to product consumption, and thus its presence increases demand because it enhances the utility derived from watching the show. Alternatively, WOM could affect demand in a similar fashion as traditional advertising by informing or persuading people to watch a specific show. Exploiting differences in the timing of microblogging activity, we find that microblogging activity *after a show* has aired is the primary driver of viewership, whereas the amount of activity prior to the show does not have an impact on viewership. This pattern suggests the complementarity between TV viewing and the consumption of (post-show) microblogging content increases TV ratings (market shares in terms of viewership). Instead, if informative or persuasive effects were important, we would expect pre-show microblogging activity to affect ratings. Next, we investigate which type of post-show activity most strongly affects ratings, and find that microblogging activity expressing sentiment has the strongest effect. Interestingly, both positive and negative sentiment affects ratings positively, suggesting an engaging (and potentially controversial) post-show debate is the key driver of TV show ratings.

Our analysis proceeds in the following steps. First, we provide details on the reasons behind the shutdown of the microblogging platform Sina Weibo, and show the political events that triggered the shutdown are unlikely to have had any direct effect on TV show viewership. The reason behind the temporary block was the defection of a prominent government official in early 2012, which led to a series of related events over a period of three months. Roughly midway through this time period, the Chinese government limited the functionality of Sina Weibo for three days. The censorship did not block Sina Weibo entirely, but only disabled the commenting function. In other words, users were still able to post tweets but were unable to comment on those tweets. Using Google trends and Baidu search-volume data as well as counts of news stories pertaining to the scandal, we show people's interest in the scandal increased at three distinct points in time that are associated with major events of the scandal. None of these events, however, coincides with the Weibo censorship and the spikes in interest do not lead to a change in TV viewership. These patterns provide our first piece of evidence that a direct effect of the block is unlikely.

In a second step, we use a series of difference-in-differences regressions to further rule out a direct effect of the block or other contemporaneous events on TV ratings. We show TV show ratings dropped in mainland China, but not in Hong Kong, where Twitter rather than Sina Weibo is used. Furthermore, within mainland China, ratings decreased more in cities with higher Weibo usage. Finally, we find the block only affected shows with a high level of microblogging activity, whereas shows with little to no activity did not experience any change in ratings during the censorship. These differential effects across geographies and shows provide a second piece of evidence that the block affected TV ratings via the channel of reduced Weibo comments.

Having established that the block only affected ratings via its impact on online WOM, we

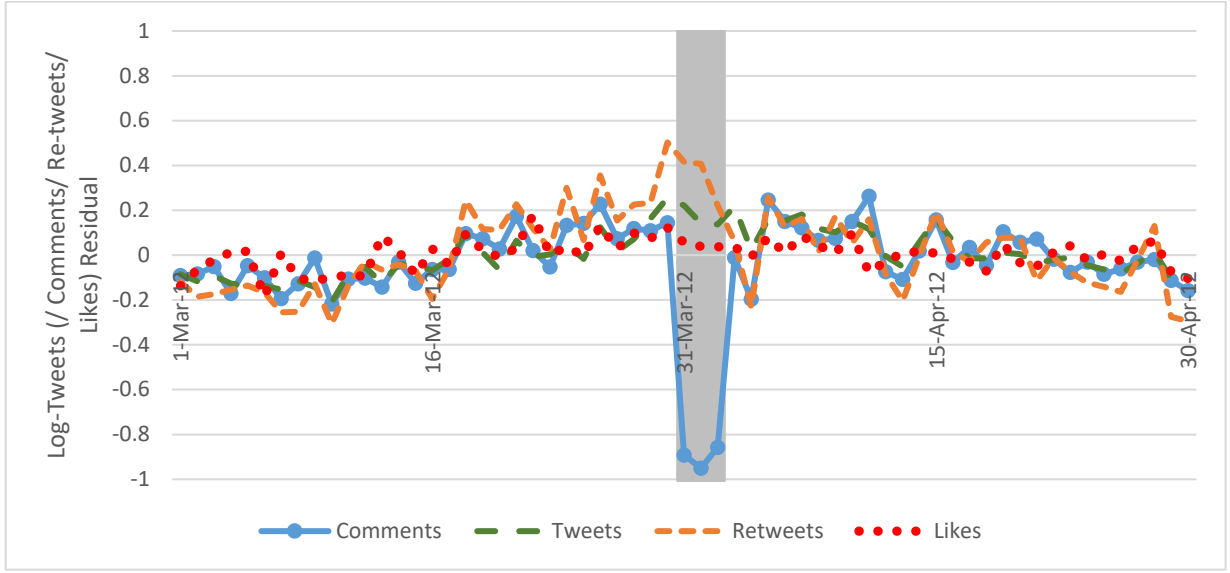


Figure 1: **The Evolution of Sina Weibo Activity over Time.** The graph plots the average daily value of the residuals from a set of regressions of  $\log(\text{Comments})$ ,  $\log(\text{Tweets})$ ,  $\log(\text{Re-Tweets})$  and  $\log(\text{Likes})$  on show and weekday fixed effects. The gray band indicates the three days of the block.

proceed in a third step to estimate the elasticity of ratings with respect to online WOM. To obtain this elasticity estimate, we run an instrumental variable (IV) regression in which we regress episode-level ratings on the number of relevant comments pertaining to the episode as well as show and weekday fixed effects. 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. The key variation this IV regression exploits is illustrated by the two graphs provided in Figures 1 and 2, which display the time series of the movement in Weibo comments<sup>5</sup> and show ratings, respectively. Figure 1 plots the average daily residuals from a regression at the episode level of (log) comments onto show (and weekday) fixed effects, thus isolating the time series variation in comments. The graph highlights the decrease in comments during the block, which occurred March 31 to April 2 and is indicated with a gray band. Figure 2 presents a similar graph for TV show ratings, which decreased during the time period of the Sina Weibo block relative to the typical ratings variation over time. Our IV-regression relates the drop in comments to the drop in ratings during the block. As a further piece of evidence for the ratings decrease during the block, we also compute the change in ratings for the top 10 shows (based on their average ratings) at the individual show level. We find all 10 shows experienced lower ratings during the censorship, and the decrease is statistically significant for most shows. Detailed results are reported in Table A1 in the appendix.

<sup>5</sup>We define the number of comments pertaining to a specific episode of a show as the total number of relevant comments posted on the day the episode aired.

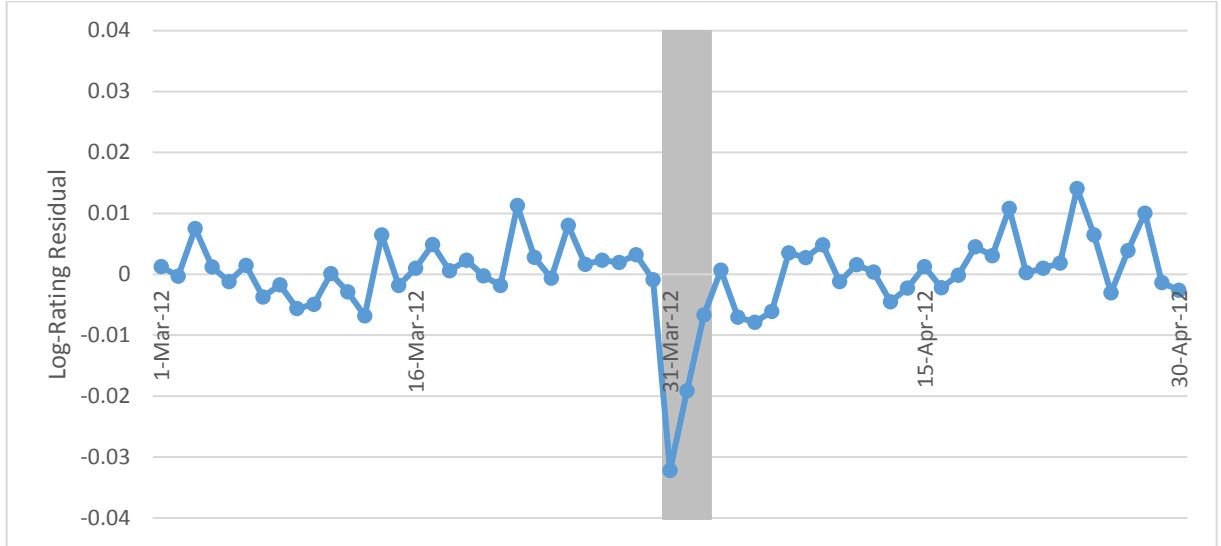


Figure 2: **The Evolution of Ratings over Time.** The graph plots the average daily value of the residuals from a regression of  $\log(\text{Rating})$  on show/weekday-pair fixed effects and time-varying controls. The gray band indicates the three days of the block. Time-varying controls include holidays and a control for the duration of the table tennis world tournament.

We note the decrease in ratings during the block is modest in magnitude. In Figure 2, we therefore control more rigorously for factors that lead to volatility in ratings over time than in most of our regressions. These controls allow us to isolate the ratings reduction during the censorship more clearly.<sup>6</sup> As a comparison, we present a similar graph that only controls for show and weekday fixed effects (as we do in all of our regressions) in Figure A1 in the appendix. Volatility of ratings outside of the block is higher, and hence the decrease during the block is less visually salient (but still statistically significant).<sup>7</sup> This lack of salience might not be surprising due to the fact that, similar to the impact of advertising (see Lewis and Rao (2015)), we would not expect WOM variation to have dramatic effects on ratings. Although none of our regressions suffer from a lack of sufficient statistical power, we nevertheless emphasize that in a setting with a small effect magnitude like ours, it is reassuring that we are able to consistently detect the effect across a series of difference-in-differences specifications. We also probe the robustness of our results to different ways of constructing standard errors, and find the statistical power of our results to be unaffected.

In a fourth and final step, we explore the mechanism underlying the impact of WOM on TV show ratings. As outlined above, we investigate heterogeneity in the effect of comments on ratings as a function of their timing (before or after the show) as well as their content, and find post-show comments expressing (positive or negative) sentiment are the primary driver of viewership. We interpret these data patterns as evidence that the complementarity of TV and microblogging

<sup>6</sup>Specifically, we control for show/weekday-pair fixed effects, holidays, and the duration of the table tennis world tournament (sport events are not included in the set of shows in our estimation sample).

<sup>7</sup>We also note our panel of shows is unbalanced due to the fact that most shows do not air every day. The time series of average ratings is therefore affected by compositional changes in shows over time.

consumption is the main channel through which WOM affects demand in our setting.

Our paper contributes to various streams of literature. First, in terms of the substantive questions addressed, our paper relates to the growing literature on social media as a marketing channel. This general area can be further divided into studies of firms' own advertising as well as user-generated content pertaining to specific brands or products.<sup>8</sup> This distinction is sometimes also referred to as paid/owned versus earned media. Within the former category, Tucker (2014) shows advertising based on personal information can increase clicks, Aral and Walker (2014) analyze how firms can generate viral marketing strategies on social networks, and Lambrecht et al. (2015) study whether Twitter promotions are more effective on early trend adopters. Petrova et al. (2016) study the impact of Twitter activity on donations to political campaigns. Gong et al. (2017) conduct a field experiment to measure microblogging's effect on demand, by varying the level of tweets generated by the firm.

Within the realm of social media marketing that is not under the firm's control, one can further distinguish studies of customer reviews as well as the analysis of online conversations about brands, which our setting falls into. This distinction is helpful due to the fact that the mechanisms underlying both types of WOM are distinct. Reviews are typically written post-consumption and typically do not involve a conversation between users. Therefore, the mechanism pertaining to consumption complementarity that we investigate is specific to online conversations and less likely to apply to reviews. A series of papers uses quasi-experimental methods to understand the impact of the valence and volume of reviews on sales. Chevalier and Mayzlin (2006) and Zhu and Zhang (2010) use difference-in-differences approaches across platforms, Luca (2016) and Anderson and Magruder (2012) employ a regression discontinuity design. Chintagunta et al. (2010) base their analysis on the sequential roll-out of movies across different markets.

With regards to papers analyzing the impact of online conversations, we are not aware of any that use experimental or quasi-experimental methods. To the best of our knowledge, our paper is the first to leverage a natural experiment to estimate the effect of WOM in this context. The bulk of the literature on (non-review) WOM uses some form of dynamic panel model to relate sales and WOM in different time periods to each other (Godes and Mayzlin (2004), Trusov et al. (2009), Villanueva et al. (2008)). Although these approaches allow us to better understand the temporal correlation patterns of WOM and sales, the nature of the variation is unlikely to recover a causal effect. These methodological differences manifest themselves in the estimated elasticity, which we find to be significantly lower than estimates found in the prior literature.

In terms of the identification strategy, our approach is related to a small set of papers that measure advertising effectiveness by exploiting quasi-experimental variation. Hartmann and Klapper (2017) exploit across-market variation in TV advertising exposure during the Super Bowl.

---

<sup>8</sup>In addition, a significant literature studies various aspects of interactions of users within social media platforms, but does not analyze the impact on consumers' purchase decisions. Trusov et al. (2010) propose a model to identify influential users on social media websites. Zhang and Zhu (2011), Aaltonen and Seiler (2016), Shriver et al. (2013), Toubia and Stephen (2013), and Ahn et al. (2016) investigated why users contribute content to social network platforms such as Wikipedia and Twitter. Ameri et al. (2016) estimate the impact of observational learning and WOM on users' adoption decisions on an anime network.

Shapiro (2016) and Tuchman (2016) use ad-exposure variation in the vicinity of borders between neighboring TV markets to study anti-depressant and e-cigarette advertising. Sinkinson and Starc (2015) use the crowding out of regular ads due to political campaigns’ ads as a source of exogenous variation. Our study similarly exploits quasi-experimental variation, in our case provided by the censorship event on Sina Weibo.

Finally, in terms of the behavioral mechanism underlying our findings, our paper relates to Becker and Murphy (1993), who propose a model of complementarity between product consumption and advertising, and Tuchman et al. (2016), who empirically test this model. In our setting, consumption complementarity arising from consumers’ social interactions rather than advertising plays a key role. We are aware of two studies that study a similar phenomena. Lovett and Staelin (2016) structurally estimate the impact of “enhanced enjoyment” of live TV consumption which arises from viewers interacting after the show. Gilchrist and Sands (2016) show that social interactions increase movie viewership due to consumers wanting to enjoy a “shared experience” with others.

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 a set of difference-in-differences regressions. In section 5, we implement an IV regression to measure the elasticity of viewership with respect to the number of comments on Sina Weibo tweets. In sections 6 and 7, we investigate the behavioral mechanism driving our results and dynamic effects of microblogging on demand. Finally, we provide some concluding remarks.

## 2 Data and Descriptive Statistics

We rely on two separate sources of data in this paper. First, we use detailed episode-level data on TV viewership (i.e., ratings) across a large set of shows, as well as geographic locations in China. Second, we assemble a unique data set of microblogging activity by scraping Sina Weibo content 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, and each show may comprise multiple episodes if the show is a series. 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 major cities in mainland China and Hong Kong (see Figure A2 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.<sup>9</sup> In Hong Kong, we obtain data for all six local channels. For each city and channel combination, we collect ratings 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.<sup>10</sup> Our time window therefore covers roughly a month before and after the block.

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. Because some shows were aired (usually on different dates and times) on multiple channels, we have 193 show-channel pairs. For Hong Kong, we have 112 shows and 132 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 episode across all cities in mainland China. When computing the average market share, we weigh 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 episode is the episode’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.<sup>11</sup> The shows in our sample come from a variety of genres ranging from TV serials and reality shows to reporting of current events and shows geared toward children. In Table A2 in the appendix, we report the rating distribution across shows for each genre separately.

We also decompose the variation in ratings into the across-show as well as the time-series variation within shows, by computing the residuals from a regression of ratings onto show fixed effects. The standard deviation of the residuals is reported in the second to last column of Table 1. We provide the same descriptive statistics for shows in Hong Kong in Table A3 in the appendix. Finally, we note that many shows in China are broadcasted at a high frequency, 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

<sup>9</sup>We exclude local channels from this calculation.

<sup>10</sup>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 and comments occur after 8 a.m., we treat the three calendar days March 31 to April 2 as the time period of the block.

<sup>11</sup>The market share for each episode is defined as the share of households watching the particular episode at the time of its broadcasting among all households that own a TV. Because many consumers do not watch any TV at a given point in time, the rating numbers are generally quite low. The ratings only measure live TV viewership and do not include delayed viewing.



	<i>Mean</i>	<i>S.D.</i>	<i>90th Perc.</i>	<i>95th Perc.</i>	<i>Max</i>	<i>S.D. (Time Series Only)</i>	<i>Obs</i>
<b><u>Ratings</u></b>							
Rating	0.434	0.483	0.941	1.225	4.166	0.158	7,899
Log Rating	0.322	0.260	0.663	0.800	1.642	0.090	7,899
<b><u>Microblogging</u></b>							
Comments	11,311	76,757	345	6,670	2,506,875	48,962	7,899
Log Comments	2.18	3.03	5.84	8.81	14.73	1.25	7,899
Tweets	2,294	11,047	125	4,083	109,923	5,659	7,899
Re-Tweets	68,192	960,259	622	21,813	30,142,880	921,605	7,899
Likes	386	9,640	1	27	487,849	9,290	7,899
<b><u>Show-Level Comments</u></b>	<i>Mean</i>	<i>S.D.</i>	<i>Min</i>	<i>10th Perc</i>	<i>90th Perc.</i>	<i>Max</i>	<i>Obs</i>
All Shows	14,406	83,159	0	0	416	786,398	193
TV Series	24,505	73,876	1	7	37,534	257,320	23
Reality Shows	48,632	169,142	0	0	111,139	786,398	40
Children Shows	26	69	0	0	45	279	16
News	5,359	21,261	0	0	260	88,665	50
Other Shows	49	152	0	0	110	1,052	64
Established Shows	16,158	93,862	0	0	281	786,399	137
New Shows	10,120	48,269	0	0	1,333	257,320	56
Re-Run	59	107	0	0	176	416	22
Current Show	16,252	88,206	0	0	571	786,399	171
Daily Frequency	7,060	35,994	0	0	279	257,320	118
Less than Daily Frequency	25,965	125,192	0	0	1,302	786,399	75

Table 1: **Descriptive Statistics: TV Ratings and Sina Weibo Activity.** The unit of observation is an episode in the top panel and a show in the lower panel.

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.<sup>12</sup> On Sina Weibo, users can engage in four different types of activities:

<sup>12</sup>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.



Figure 3: **Example of a Tweet with Comments.** The displayed tweet is about a popular Chinese dance competition show, similar to the US show “Dancing with the Stars.”

tweeting, retweeting, “like”-ing, and commenting. Commenting is the type of activity primarily affected by the block. It allows a user to respond to an existing tweet with some content of her own (a “comment”). All comments on a given tweet are listed below that tweet, forming a thread of interactive discussion among users who are interested in the subject of that tweet. These users can read and reply to each other’s content in that thread, even though they do not necessarily follow each other. The conversational nature of comments distinguishes it from other types of activity (e.g., tweets, re-tweets, and likes) and is likely to play a role in the behavioral mechanism we identify later. Figure 3 displays an example of a tweet together with a series of comments pertaining to the tweet.

We obtained the microblogging data by scraping the Sina Weibo website. Specifically, for all shows contained in our data, we scraped every tweet mentioning the show during March and April of 2012. We further collected the number of comments on each tweet,<sup>13</sup> as well as the number of re-tweets and likes.<sup>14</sup> 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.<sup>15</sup> We note this definition includes microblogging activity both before and after the show aired. Based on a cursory check of the content of tweets and comments, we found a calendar day to be a good approximation for the time window that delineates content pertaining to a specific episode. Pre-show tweets typically contain content regarding consumers’ anticipation of the upcoming episode, whereas post-show tweets contain discussion of the episode that aired earlier in the day. We return in more detail to the timing of activity when investigating the mechanism by which microblogging affects demand in section 6.

We report descriptive statistics for the different types of microblogging activity in the lower portion of the first panel in 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 is about five times larger than the number of tweets. In terms of other types of user activity, we observe a large number of re-tweets related to the shows in our sample. 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 comments per episode in our data is equal

---

<sup>13</sup>One can also comment on a re-tweet, but we only consider the comments on the “original” tweet.

<sup>14</sup>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. 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.

<sup>15</sup>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. Other types of activity typically occur within a short time window after the original tweet. Specifically, we find that more than 50% of comments were posted within 41 minutes of the original tweet (to establish this pattern we analyze a sample of 12,000 comments for which we collected timestamps).



Figure 4: **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.

to 11,311, 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.

To explore which show characteristics correlate with the amount of activity on Sina Weibo, we report the across-show distribution of comments for specific subsets of shows in the bottom panel of Table 1. The unit of observation is a show (rather than an episode), and for each show, we compute the average number of comments per episode, excluding any episodes that aired during the block. We first decompose the comment distribution by show genre. In our sample, we identify four major categories of TV shows, namely, TV series, reality shows, news / reporting on current events, and children’s shows. The remaining shows belong to a variety of genres such as weather reports, finance-related shows, historical documentaries, and so on, and we lump them together in a residual category. TV series and reality shows garner substantially more comments per episode than children’s shows and shows in the “other” category. News reporting also receives relatively high activity, largely because consumers debate the news coverage and related political events on Sina Weibo. We provide a similar decomposition for old versus new shows, where the latter is defined as a show that started airing during our sample period, as well as whether the show is a re-run or a show that is airing for the first time. Finally, we also analyze differences in activity across shows that air daily versus at a lower frequency. We find significant heterogeneity exists in activity within each genre, and the subset of shows in the new/old, re-run/first-time as well as daily/less frequent categories. Therefore, while observable show characteristics do predict the extent of microblogging activity somewhat, there are strong show-specific factors that drive the amount of Sina Weibo activity.

### 3 The Block

In this section, we provide details on the censorship of Sina Weibo and the events leading up to it. Other than providing background information on the source of exogenous variation we exploit, the timeline of the scandal also helps us assess whether the block and the political events driving it might have affected TV viewership through any channel other than the reduction in microblogging activity. The timing of the censorship event thus provides a first piece of evidence in support of our exclusion restriction: The block only affected TV ratings via its impact on microblogging.

In February 2012, a political scandal erupted in China after Wang Lijun, a top government official in the city of Chongqing, defected to the US Consulate. His superior, Bo Xilai, was later removed from his post and arrested. By March 2012, many rumors related to the political scandal appeared on the Internet, especially on social media websites. To remove such rumors, 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.<sup>16</sup> Specifically, the commenting function was disabled during that time period.<sup>17</sup> Figure 4 displays the announcement of the 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 origins of the block were political events, the events that triggered the block might have had 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 is unrelated to the reduction in the number of comments. 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 and involved several major events, none of which coincided with the block. Second, we find that during those major events, TV viewership remained unchanged, while

---

<sup>16</sup>The government gave no official statement linking the Weibo block explicitly to political events, but news sources clearly connected them (see, e.g., “Coup Rumors Spur China to Hem in Social Networking Sites,” *New York Times*, March 31, 2012, accessed November 2015).

<sup>17</sup>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 held nearly 90%. The authorities also closed down 16 minor websites for spreading political rumors.

it dropped during the Sina Weibo block. Therefore, interest in the political events does not seem to affect viewership.

To establish the first point, we obtain data from the Chinese search engine Baidu regarding the number of search queries for the names of the two Chinese officials involved in the scandal: Wang Lijun and Bo Xilai. In Figure 5, we plot the time series of the daily number of searches for the two names from January to June 2012, where the three days of the Sina Weibo block are indicated in gray. The graphs display the time window of the entire series of events from early February to mid-April and feature three pronounced 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 rather than the Baidu search data, and found the time-series patterns to be very similar.

Second, we find no 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 episode-level (log) ratings onto show and weekday fixed effects and a dummy for the time period of the block. On top of these variables, we then also include the (log) search volume of both names in the regression. We find the coefficients on both variables 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 results from this regression in Table A4 in the appendix. We also show robustness to using the search-volume variables 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-volume and key-event dummy variables on TV show ratings.

Furthermore, to provide a more exhaustive assessment of interest in the scandal, we provide three additional pieces of evidence. First, we collect the Baidu search volume for a wider set of queries related to the scandal, namely, “Gu Kailai” (Bo Xilai’s wife), “Bo Guagua” (Bo Xilai’s son), “Chongqing” (the city where the scandal started), and “US consulate” (Wang’s visit to the US consulate triggered the scandal). Second, we collect data on news coverage pertaining to the political events from Baidu news as well as Google news and compute the number of news articles published on either website that contain the name Bo Xilai or Wang Lijun. Third, Internet users in China were conceivably seeking out information about the scandal from sources of Western news or social media platforms that were blocked in mainland China (e.g., Twitter, Facebook, etc.). We can capture some of this activity by collecting the Baidu search volume for the query “Fan Qiang,”<sup>18</sup> a keyword Chinese people use to find VPN software to circumvent the restrictive Internet control. Although the search for VPN software is not specific to the political scandal, we expect to see an

---

<sup>18</sup>This Chinese word translates to “jumping over the wall.”

increase in attempts to access censored sources during these politically sensitive events. Across all three sets of time series, (1) the larger set of Baidu queries, (2) news coverage, and (3) searches for the VPN software, we find no evidence for an increase in interest in the scandal around the time of the censorship, and these additional graphs show similar patterns to the graphs presented in Figure 5, with spikes around the three main events of the scandal. We report more details on how the graphs are computed in section B of the appendix.

## 4 Difference-in-Differences Analysis

In this section, we analyze the magnitude of the ratings change during the censorship event and provide further evidence that the block affected TV viewership via its effect on the amount of TV-show-related microblogging activity on Sina Weibo. The latter is achieved via a set of difference-in-differences regressions, and complements the analysis in the previous section in terms of ruling out a direct effect of the block.

As a starting point to this analysis, we implement a “simple-difference” regression that analyzes the drop in ratings around the time of the block. Specifically, we run a regression of episode-level log-ratings on a dummy for the three days of the block, show fixed effects, as well as day-of-the-week dummies:

$$\text{LogRating}_{jt} = \alpha * \text{Block}_t + \delta_j + \text{Weekday}_t' \gamma + \varepsilon_{jt}. \quad (1)$$

$\text{LogRating}_{jt}$  denotes the logarithm of the rating of show  $j$  on day  $t$ .  $\text{Block}_t$  is a dummy variable equal to 1 for the three days of the block.  $\delta_j$  is show  $j$ ’s fixed effect.  $\text{Weekday}_t$  is a vector of day-of-the-week dummies.  $\varepsilon_{jt}$  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 is potentially problematic if it happened during the same time window and also affected viewership. 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 implement 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.

### 4.1 Geography-based Analysis

We start with an analysis along the geographical dimension, which leverages the fact that Sina Weibo is the predominant microblogging platform in mainland China but not in Hong Kong, where

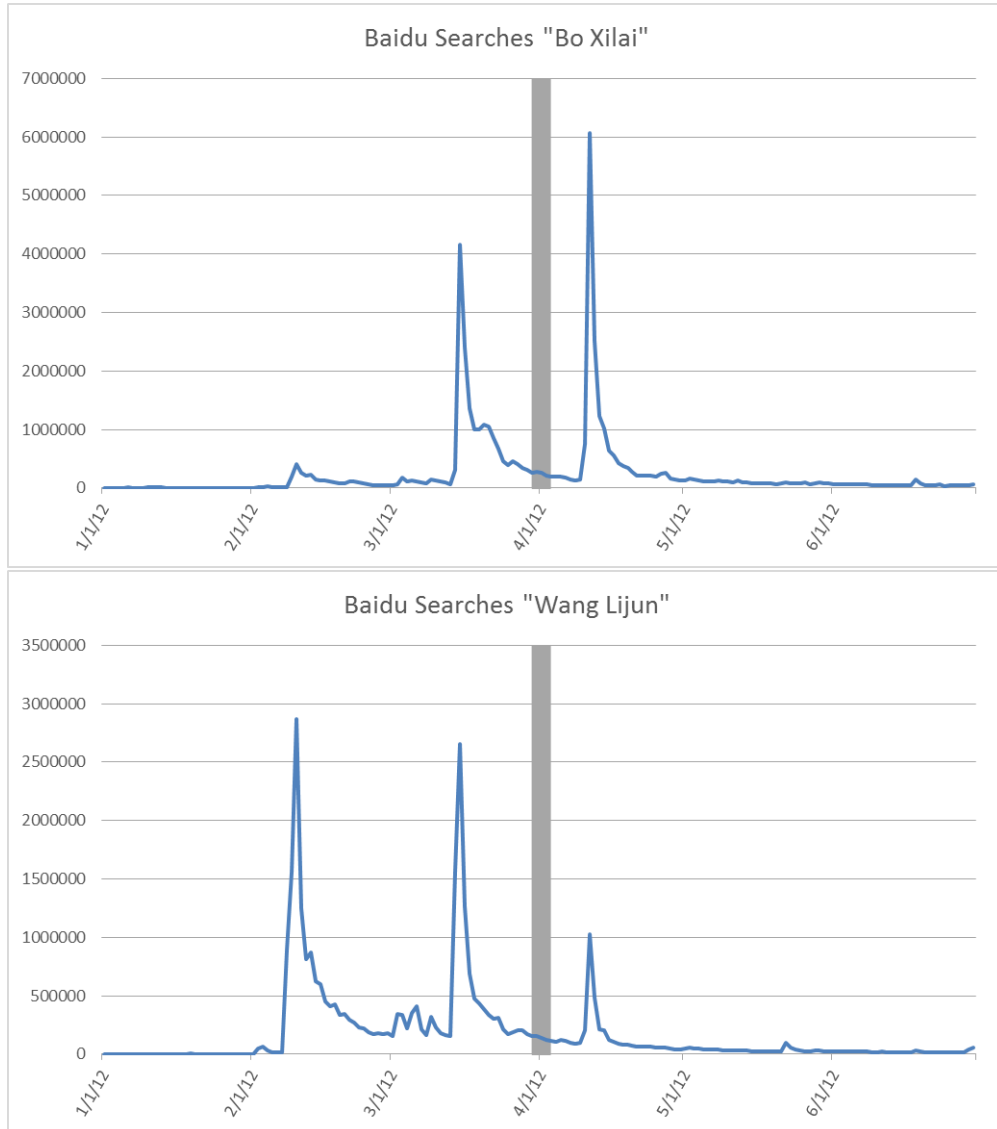


Figure 5: **Time Series of the Baidu Search Volume for the Names of the Two Politicians Involved in the Scandal from January to June 2012.** The three days of the Sina Weibo block are indicated in gray.



Twitter is also available. 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. 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. We re-iterate that the shows in Hong Kong and mainland China are different because Hong Kong consumers primarily watch local channels, which do not overlap with the major channels in mainland China. We are hence comparing different sets of shows between the two locations. Later, we run a further set of regressions based on a sample of overlapping shows. Similar to the previous regression, we include show and day-of-the-week dummies and cluster standard errors at the show level:

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

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, and 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 evidence that Hong Kong constitutes a valid control group, we compute the correlations of the Baidu search volume regarding the scandal between mainland China and Hong Kong. The Baidu search-volume levels 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 A6 in the appendix, which shows the volumes are highly correlated across the two locations with identical peaks in activity.<sup>19</sup> 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.

The results from this regression are reported in column (2) of Table 2. The primary coefficient of

---

<sup>19</sup>We also attempted to collect data on the set of other sources that indicate interest in scandal, which we discussed earlier in section 3 (see in particular the last paragraph in that section) by geography. However, for most variables, the data are quite sparse when cutting it by geography. For example, “Gu Kailai” and “US Consulate” had a low volume of Baidu searches and barely appeared in Google news articles in Hong Kong. We therefore only include graphs for the two additional keywords for which we are able to obtain sufficient data, “Bo Guagua” and “Chongqing.” The correlations between Hong Kong and mainland China for those two search queries are equal to 0.86 and 0.82, respectively. The graphs of these time series are presented in Figure A6 in the appendix. Generally, we would like to compare other sources as well, such as online discussions of the scandal in the two geographies. However, because such discussions were censored in mainland China, we cannot implement such a comparison.

	(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 (respective shows)	HK and Shenzhen (mainland shows)	24 Cities in Mainl. China	24 Cities in Mainl. China
Censor Dummy	-0.017*** (0.005)	0.005 (0.010)	0.002 (0.010)	-0.008*** (0.002)	-0.010 (0.006)	-0.008 (0.006)
Mainland × Censor Dummy		-0.026** (0.012)				
Shenzhen × Censor Dummy			-0.035** (0.014)	-0.017* (0.010)		
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	15,798	189,576	189,576
Shows	193	325	325	193	193	193
R <sup>2</sup>	0.881	0.964	0.951	0.774	0.479	0.479

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.

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. Furthermore, 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), are similar to the previous regression comparing Hong Kong and the entire mainland.

We then proceed to exploit one slightly different dimension of the data. The regressions in column (2) and (3) both compare different sets of shows in Hong Kong and mainland China /

Shenzhen. However, Hong Kong residents can receive mainland channels and we are therefore able to compare the ratings for the *same set of shows* in Hong Kong and Shenzhen. To leverage the overlap in shows, we run the same specification as in column (3), but base the regression on the sample of mainland shows and their ratings in both Shenzhen and Hong Kong. Therefore, our sample for this regression includes every episode twice, once for Shenzhen and once for Hong Kong. Instead of show fixed effects, we include a separate fixed effect for each show/city-pair. The results from this regression are reported in column (4) and, as expected, the drop in ratings is significantly larger in Shenzhen relative to Hong Kong. We also note the (un-interacted) censor dummy is negative and significant, most likely because mainland shows are tweeted about on Sina Weibo and hence consumers in Hong Kong can in principle access this information. Furthermore, the market share of mainland shows in Hong Kong shows is small, and therefore the selected group of Hong Kong-based consumers that watch mainland shows might also read the relevant comments on Sina Weibo.<sup>20</sup> Therefore, we would not necessarily expect to see no effect in Hong Kong. However, the stronger impact of the block in mainland China supports our assertion that the censorship affected mainland consumers more strongly because they rely relatively more on Sina Weibo.<sup>21</sup>

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. In the same vein as the regression presented in the previous paragraph, these regressions compare the same set of shows across different geographical locations. 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 heterogeneity across cities, we run our baseline regression with an additional interaction term of the censor dummy and the local Sina Weibo usage variable<sup>22</sup> (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. To interpret the magnitude of the coefficient, note that we scale 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 sees ratings drop by 3.7%. We also implement a specification in which we interact the censor dummy with an indicator for whether

---

<sup>20</sup>We report the rating distribution for mainland shows in both Shenzhen and Hong Kong in the lower panel of Table A3 in the appendix.

<sup>21</sup>In a similar spirit, we could in principle also analyze shows of Hong Kong origin that air in Hong Kong as well as in Shenzhen. However, only two of the six Hong Kong channels can be received in Shenzhen, and the market share of those shows is small in Shenzhen. Furthermore, Hong Kong shows are primarily targeted at local consumers, and hence they are not talked about much on Sina Weibo. Therefore, the small subset of consumers in Shenzhen that watch the show are unlikely to use Sina Weibo to inform themselves and talk about these shows.

<sup>22</sup>We define Weibo usage as the number of times the three keywords were searched in a specific city divided by the population size of the respective city. Because the scale of this variable is hard to interpret, we normalize the variable to lie between zero and one.

the city has an above-median usage of Sina Weibo and again find a significant negative effect.<sup>23</sup>

## 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 mainland China. Specifically, we group shows based on Sina Weibo activity outside of the block. If the block only affected TV viewing via 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 per episode of each show in the month before and after the block.<sup>24</sup> In other words, we categorize shows by the “normal” amount of commenting activity 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 three and 45 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:

$$\begin{aligned} \text{LogRating}_{jt} = & \alpha * \text{Block}_t + \alpha_M * \text{Block}_t * \text{Medium}_j + \alpha_H * \text{Block}_t * \text{High}_j \\ & + \delta_j + \text{Weekday}'_t \gamma + \varepsilon_{jt}. \end{aligned} \quad (3)$$

*Medium<sub>j</sub>* and *High<sub>j</sub>* 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 small in magnitude and insignificant. The medium-level shows experience a stronger drop of 1.3% in their ratings (i.e., 0.5% + 0.8%). However, the decrease for medium-activity shows is not significantly different from the effect for low-activity shows. For shows with a strong presence on Sina Weibo, we find an effect of larger magnitude. For this category of shows, ratings dropped by 3.1% during the block (i.e., 0.5% + 2.6%), and the estimated coefficient is significantly different from the effect for low-activity shows.<sup>25</sup>

<sup>23</sup>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.

<sup>24</sup>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. We report descriptive statistics for the average number of comment per episode (outside the block) in the lower panel of Table 1.

<sup>25</sup>Consumers could conceivably substitute away from high- toward lower-activity shows. This type of substitution between shows does not seem to occur, because none of the groups of shows captured in the regression experiences an increase in ratings. In section E in the appendix, we further explore whether consumers substitute between shows or to the outside option of not watching any TV. We find that comments primarily lead consumers to substitute away from the outside option rather than other shows.

Dependent Variable	(1)	(2)	(3)	(4)	(5)
	Log Rating	Log Rating	Log Rating	Log Rating	Log Rating
Censor Dummy	-0.005 (0.005)	-0.012** (0.005)			
Medium Weibo Activity × Censor Dummy	-0.008 (0.011)			-0.007 (0.009)	
High Weibo Activity × Censor Dummy	-0.026** (0.012)			-0.024** (0.011)	
Weibo Activity (Unit: 100,000 Com.) × Censor Dummy		-0.029** (0.012)			-0.030*** (0.011)
TV Series × Censor Dummy			-0.025 (0.020)	-0.005 (0.021)	-0.015 (0.016)
Kids Show × Censor Dummy			-0.003 (0.007)	0.002 (0.007)	-0.003 (0.007)
Current Events Show × Censor Dummy			-0.024*** (0.006)	-0.015** (0.006)	-0.021*** (0.006)
Reality Show × Censor Dummy			-0.020 (0.015)	-0.005 (0.014)	-0.002 (0.015)
Other Shows × Censor Dummy			-0.008 (0.007)	-0.001 (0.009)	-0.008 (0.007)
Show FEs	Yes	Yes	Yes	Yes	Yes
Weekday Dummies	Yes	Yes	Yes	Yes	Yes
Observations	7,899	7,899	7,899	7,899	7,899
Shows	193	193	193	193	193
R <sup>2</sup>	0.881	0.882	0.881	0.881	0.882

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

To 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. We find a significant effect for the interaction term, which we report in column (2) of Table 3. 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.

Apart from differences in activity levels on Sina Weibo, we can also characterize shows along other dimensions that might influence how strongly they are affected by the censorship event. We thus proceed to analyze the differential impact of the block on different show genres, new versus old shows, and so on. Our analysis in this regard is based on the different show attributes that were presented earlier in the lower panel of Table 1. We first analyze differences in the impact of the block for shows belonging to different genres. Specifically, we can categorize shows in our

sample into four major categories, namely, TV drama series (23 shows), reality shows (40), news / reporting on current events (50), and children’s shows (16). The remainder of shows (64) belong to a variety of genres and are treated as a residual category. Results from a regression that interacts a dummy variable for each of the five categories with the censor dummy is reported in column (3) of Table 3.<sup>26</sup> In terms of magnitude, we observe a large negative coefficient for TV series, reality shows, as well as news shows, but only a small coefficient for children’s shows and the residual category. Although the five coefficients are not significantly different from each other, we do find that TV series, reality shows, and news shows as a group experience a significantly larger decrease than the other two genres.<sup>27</sup>

Furthermore, the genres with larger effect sizes are tightly correlated with the type of genres that experience higher Sina Weibo activity. The data on activity levels across genre were presented earlier in Table 1, and we find an average number of comments per episode of 24,504 / 48,631 / 5,358 for TV series / reality shows / news shows and much fewer comments, 26 and 48 per episode, for children shows and other shows, respectively. We therefore conjecture that the differential impact we observe across different show genres proxies for the key driver of effect heterogeneity, which is the average activity level on Sina Weibo of each show. To further explore this assertion, we estimate a regression that includes interactions of the censor dummy with both show genres as well as with activity levels (see column (4)). When doing so, the effect sizes of the genre interaction terms decrease substantially, whereas the effect magnitudes on the interaction terms with activity levels are very similar to the specification in column (1), which includes only those terms. The one exception is the effect on news shows, which is significantly different from zero; however, the five genre interaction terms are not significantly different from each other. Results are similar when including genre dummy interaction together with a linear interaction term with Weibo activity in column (5). We hence conclude that genre has no major role in predicting the ratings drop during the block, after controlling for differences in the shows’ average activity level.

Finally, we also explore heterogeneity between shows that started airing recently versus longer-running shows, shows that air daily versus at a lower frequency, as well as re-runs versus new shows. When exploring effect heterogeneity along each of these three dimensions, we find no significant differences across shows aired at a different frequency and for new versus established shows.<sup>28</sup> The absence of a significant effect is most likely due to the fact that the activity level in terms of the average number of comments does not differ much across those two dimensions. The only characteristic that is predictive of the impact of the block is whether the show is a re-run or a

---

<sup>26</sup>We include interactions with an exhaustive set of genre dummies in this regression and thus do not include the censor dummy without interaction.

<sup>27</sup>We also note that although we find a significant effect only for news shows, our results are robust to dropping news shows from the sample entirely. When estimating equation 1 without news shows, we obtain a coefficient (standard error) on the censor dummy of -0.016 (0.007). In comparison, the coefficient (standard error) based on all shows is equal to -0.017 (0.005).

<sup>28</sup>In each case, the show characteristic is defined as a binary distinction. We hence run our baseline regression with the censor dummy as well as an interaction of the dummy with a dummy for one of the binary values in each case. The coefficients (standard errors) in each of the cases are as follows. Interaction of the censor dummy with the new-show dummy: 0.001 (0.013); interaction with daily-frequency dummy: -0.010 (0.013).

current show. Re-runs are significantly less affected by the censorship. Similar to the heterogeneity patterns across genres, this finding is most likely driven by the fact that re-runs do receive less attention on Sina Weibo, and the average number of comments is much smaller (59 per episode) for re-runs relative to current shows (16,000 comments per episode).<sup>29</sup>

### 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. 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; hence, strong time trends in TV viewership are unlikely. Second, our setting is slightly richer than other difference-in-differences settings, because the treatment is temporary. 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 coefficients of interest are robust to including them. We implement this robustness check for the main regressions presented in this section. Specifically, we re-run four difference-in-differences regressions across geographies and across types of show (one regression of HK vs. mainland China, two regressions across-cities within mainland China, and one regression across-shows with different 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 A5 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 vein, 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 out of 10 relevant coefficients across all regressions, we find

---

<sup>29</sup>The coefficient (standard error) on the interaction of the re-run dummy with the block dummy is equal to 0.029 (0.010).

only two to be significant (one at the 1% level, the other at the 10% level).<sup>30</sup> In both cases, the sign of the estimated coefficients has the opposite sign relative to the “true” treatment coefficient. Furthermore, we run the identical set of falsification tests for a one-week shift of the block. This approach allows us to hold fixed the days of the week during which the (placebo) block happens. We implement the analysis for the week before and after the block across all four specifications and find one coefficient out of 10 is significant at the 10% level. All other coefficients are statistically insignificant. The results from these placebo tests provide further evidence that differential time trends are unlikely to contaminate our results.

Finally, we run a robustness check with regards to the computation of standard errors, which are clustered at the show level in all the regressions reported above. As an alternative, we follow Bertrand et al. (2004) and compute the average rating of each show separately for all episodes during and outside of the block, thus reducing the data to exactly two observations per show (one for during the block and one for outside the block).<sup>31</sup> For each pair of treatment and control groups across three specifications,<sup>32</sup> we then compute the difference-in-differences estimate and the corresponding standard error. We find our results remain qualitatively similar. In the case of the comparison of shows in mainland China and Hong Kong, as well as shows with different microblogging activity levels, precision increases when collapsing the data. For the specification based on cities with different Sina Weibo penetration rates, precision decreases and the estimated effect is significant only at the 10% level.<sup>33</sup>

#### 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 by adjusting other types of marketing activity. 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 only one type of marketing activity (WOM in this case). 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.

---

<sup>30</sup>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.

<sup>31</sup>To mirror our regression specification, we first regress ratings on show and weekday fixed effects and then compute the average rating residual for each show and time period. We also weigh shows by the number of episodes aired during the relevant time period when computing average ratings.

<sup>32</sup>We do not apply this approach to the regression based on an interaction with the (continuous) Weibo penetration rate variable with the censor dummy, because no binary distinction exists between treatment and control group in that specification.

<sup>33</sup>The p-values for the three main specifications (mainland China / Hong Kong, low- / high-activity shows, low / high Sina Weibo penetration cities) are equal to 0.012, 0.012, and 0.077 when collapsing the data, relative to 0.028, 0.026, and 0.006 when clustering standard errors at the show level.



## 5 Elasticity Estimate: Instrumental Variable Regression

In this section, we proceed to measure the impact of comments onto ratings, which is the key estimate of interest to evaluate the impact of online WOM in our context. So far, we analyzed the impact of the block on ratings. We now add data on microblogging activity at the episode-level and estimate the impact of comments on ratings by isolating the variation in comments caused by the censorship event. Specifically, we run an IV 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 during the block. The difference-in-differences regressions in the previous section as well as the timeline of the scandal presented in section 3, both provide evidence that the block affected TV show ratings via its effect on Sina Weibo activity and did not directly affect ratings. The absence of any direct effect of the block on ratings is the key identifying assumption and allows us to exclude the block from the second stage of the IV regression.

Similar to the regression framework used in the previous section, we control for show fixed effects and day of the week, and cluster standard errors at the show level. Formally, we run the following regression:

$$\text{LogRating}_{jt} = \alpha * \text{LogComments}_{jt} + \delta_j + \text{Weekday}_t' \gamma + \varepsilon_{jt}, \quad (4)$$

where  $\text{LogComments}_{jt}$  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.

### 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 comments 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 number of comments onto show fixed effects and weekday dummies, as well as a dummy for the three days of the block.<sup>34</sup> Columns (1) and (2) in Table 4 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.<sup>35</sup>

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 A6 in the appendix. With the exception of a significant positive effect (only in the log-specification) for re-tweets, we find no evidence that the block affected any type of activity on Sina Weibo other than comments. The fact that other types of activity are unchanged during the duration of the

---

<sup>34</sup>We also include a dummy for whether an episode was actually aired on the specific day.

<sup>35</sup>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 Episode-Level Comments	Log Number of Episode-Level 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 Episode Level Comments (Unit: 10,000)					0.024*** (0.009)	
Log Number of Episode-Level Comments						0.016*** (0.005)
Censor Dummy	-10,351** (5,172)	-1.003*** (0.086)	-14,751** (6,504)	-1.066*** (0.104)		
F-Stat on Censor Dummy	4.00	136.38	5.14	105.14		
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
R <sup>2</sup>	0.483	0.789	0.595	0.839	0.845	0.877

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.

censorship is important because we are attributing the change in ratings to the change in comments during the block. If other types of activity also changed, the exclusion restriction is violated because the block will affect ratings not only via comments but also through its indirect impact on other types of microblogging activity.<sup>36</sup> We also note that ideally we would like to study the impact of all four activities onto ratings, but such an analysis would require four separate instruments, whereas we have access to only one.

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 period with each episode for which we have ratings data. We define the number of relevant comments pertaining to a specific episode as the number of comments that mentioned the show on the same calendar day on which the particular episode aired.<sup>37</sup> The results from these “episode-

<sup>36</sup>Due to the (weak) uptake of re-tweets during the block, if anything, the effect of ratings per unit of comment might be larger than our estimate.

<sup>37</sup>This measure includes comments that are posted after the episode airs. These comments are included because consumers might value reading/writing them after watching the show, and they are able to anticipate the number of post-show comments based on comments pertaining to previous episodes of the same show. As we show in more

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 first stage estimate is highly significant as the F-statistic on the censor dummy (the excluded instrument) shows, which mitigates weak-instrument concerns (Rossi, 2014).

## 5.2 Second Stage and Effect Magnitude

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 0.016, and hence doubling the number of comments leads to a 1.6% increase in ratings.<sup>38</sup> To gauge the magnitude of this effect, we first note that doubling the number of comments leads to a movement of roughly 20% of a standard deviation of the typical within-show fluctuation in ratings over time (see Table 1) and thus constitutes an economically important effect.<sup>39</sup>

At the same time, our estimate is substantially lower than most estimates in the previous literature. For instance, Sonnier et al. (2011) estimate a short-run elasticity of (positive) WOM of 0.64, Dhar and Chang (2015) find an elasticity of 1.04 for music sales, and Liu (2015) estimates an elasticity of 0.59 for box office revenue in the release week of a movie. These estimates (as well as those from several other papers in the WOM literature) are more than an order of magnitude larger than our elasticity estimate of 0.016. Although the different papers do study WOM in different product markets and on different WOM platforms, we suspect that part of the discrepancy is due to the fact that previous papers were limited in their ability to deal with issues of endogeneity due to the nature of their data. We note that one other more moderate elasticity estimate is from Lovett and Staelin (2016), who find an elasticity of 0.04.

Importantly for the optimal allocation of marketing budgets across channels, we obtain an estimate for the impact of WOM that is smaller than magnitude levels found for TV advertising of around 0.03 (Gordon and Hartmann (2013), Tuchman (2016), Shapiro (2016)). Therefore, the seemingly higher impact of WOM relative to traditional advertising when comparing correlational studies might be an artifact of endogeneity bias rather than actual differences in effectiveness. Of course, we estimate the impact of WOM only for one particular product market and one specific type of WOM. Hence, whether the more modest effect size constitutes a finding that holds across a variety of markets and forms of WOM is an issue that is beyond of the scope of this study and is left for future research.

---

detail in section 6 below, post-show comments turn out to play a crucial role in our setting.

<sup>38</sup>This result is closely related to the regressions of log ratings on the block dummy presented earlier. In particular, column (1) of Table 2 represents the reduced form of the IV regression presented above. Hence, the multiplication of the first-stage coefficient of the block dummy and second-stage coefficient on log comments is equal to the coefficient on the block dummy in column (1) of Table 2.

<sup>39</sup>An alternative way to assess the economic relevance of the estimated effect is to consider the impact of microblogging on revenue from TV advertising (which is tied to ratings). We implement such a calculation in the appendix (section C) and find that a one-standard-deviation shift in the number of comments (76,000 comments), entails a change in advertising revenue per episode of US\$190,000.

Finally, it is instructive to assess, using a back-of-the-envelope calculation, how many households changed their viewership behavior relative to the number of active and passive Sina Weibo users. The average show in our sample has a rating of 0.434% (i.e., 0.434% of households were watching the show). In 2012, China had 456 million households, and the block moves ratings by 1.6%. Therefore,  $450,000,000 * 0.00434 * 0.016 = 32,000$  fewer households watched the average episode. With respect to the amount of microblogging, the average episode received 11,311 comments. Under the assumption that these comments originate from different users, and assuming a ratio of readers to contributors of 100:1<sup>40</sup>, the block makes 3% of Sina Weibo readers ( $32,000 / (11,300 * 100)$ ) not tune into the respective show. We note this calculation depends crucially on the ratio of readers to contributors, which we do not observe in the data. Hence, we use an external number as an approximation.

### 5.3 The Nature of the Experimental Variation and Its Impact on Effect Magnitude

As with any natural experiment, our analysis is predicated on the specific variation induced by the natural experiment. Several aspects of this variation are useful to keep in mind when interpreting the magnitude of the effect and its external validity.

First, our natural experiment provides a substantial shift in microblogging activity by fully eliminating it during the block. This type of large deviation is likely to yield a different effect on ratings than a change in microblogging activity at the margin. Given the nature of our shock, we are unable to trace out the response curve in our setting. However, we posit that the most likely scenario is one with decreasing returns from additional microblogging activity, and hence the marginal effect will be smaller than the one estimated from a larger deviation from the status-quo level.

Second, the censorship event eliminated microblogging activity for all shows rather than for any one individual show. Therefore, the effect on ratings from a uni-lateral change, holding microblogging for all competitor shows constant, could conceivably lead to a larger reaction in terms of ratings, because consumers might substitute to other shows that are still receiving comments. To assess whether such competitive effects are likely to matter in our context, we investigate whether consumers substitute primarily between shows or toward the outside option during the block. We find the latter to be the dominant channel, and the primary source of the loss in ratings is due to consumers watching less TV overall. Therefore, the effect of a unilateral decrease in microblogging might not be very different from the effect due to a decrease across all shows. We describe the analysis of substitution patterns in detail in section E of the appendix.

Finally, our estimate isolates the elasticity of comments while holding all other types of microblogging activity constant. One would expect a larger elasticity when increasing all activity by a certain percentage, rather than just increasing comments.

---

<sup>40</sup>See [https://en.wikipedia.org/wiki/1%25\\_rule\\_\(Internet\\_culture\)](https://en.wikipedia.org/wiki/1%25_rule_(Internet_culture)), retrieved on June 6th, 2016.

## 6 Mechanism: Advertising or Consumption Complementarity?

We now turn to analyzing the behavioral mechanism underlying the estimated effect of microblogging on demand. To structure our analysis of the mechanism, we base it on the three main conceptual frameworks with regard to the impact of advertising outlined in Bagwell (2007): informative and persuasive effects, and complementarity between advertising and product consumption. Although we study WOM rather than traditional advertising, the distinction between the three possible channels applies to any kind of information provision, and hence also applies to WOM. Informative effects can arise if microblogging reminds the consumer of a show’s existence or provides additional information about features of the show that the consumer values. Persuasive effects increase the consumers’ appreciation of the show without delivering information about the show or its content. Finally, complementarity between WOM and TV show viewership arises if consumers derive a higher utility from viewing the TV show when they are also able to engage in microblogging (either actively or passively).

Among the three mechanisms, only the complementarity channel allows for an active role of the consumer in her decision to obtain information. Informative and persuasive effects are entirely passive; that is, the consumer gets exposed to certain information by reading content on Sina Weibo, and this exposure might affect her choice regarding which TV show to watch. Instead, complementarity implies the consumer actively decides how much information to consume. In the context of advertising, the consumer might decide to watch certain ads as a function of her past consumption of the product (Becker and Murphy (1993)). In the case of microblogging, the case for an active role is even stronger because reading tweets and comments is arguably a more conscious choice than watching an ad. Furthermore, the consumer might also decide to contribute information by commenting herself, a possibility that is absent with regards to firms’ advertising. We therefore believe studying complementarity is particularly interesting in this context because it seems a priori to be a more important mechanism in the context of WOM relative to traditional advertising.

Below, we first separate complementarity from informative/persuasive effects based on the timing of comments. Second, we distinguish between informative and persuasive effects based on the actual content of comments. The latter part requires additional data on microblogging content. We gathered such data for a random sample of around 12,000 comments<sup>41</sup> across all shows in our sample and assign each comment to the following categories: informative, expressing sentiment, or neither. In the case of sentiment, we further categorized those comments into positive or negative sentiment. Across all shows in our data, we find that 25% of comments express sentiment, of which 24% are positive and 1% are negative. Only 1% of comments contain informative content (e.g., the time at which the episode airs). In Table A7 in the appendix, we provide more detailed descriptive statistics on the distribution of content attributes across shows.

---

<sup>41</sup>We provide more detail on the sampling process in section D in the appendix.

## 6.1 Consumption Complementarity

We first turn to disentangling the complementary function of microblogging from the informative and persuasive channels. For comments to affect consumers via the latter two channels, consumers need to be exposed to comments on Sina Weibo before the show actually airs. Such pre-show comments can inform consumers about the show or persuade them to watch the show due to the enthusiastic comments of other users. Therefore, when pre-show comments disappear during the block, some subset of consumers will not be informed about or persuaded to watch the show, which might explain the decrease in ratings.

Alternatively, consumers might gain utility from engaging in microblogging (either reading or contributing) *after* the show to discuss their opinions about the episode. In this scenario, when the censorship event removes the ability to comment after a show aired, the consumer’s utility from watching the show decreases, and hence fewer consumers tune in. In principle, complementarity could also apply to pre-show comments. For instance, consumers might have a discussion about the plot of the upcoming episode, and being able to engage in such a discussion increases the utility from watching the show. Therefore, any impact of post-show comments is only consistent with the complementarity channel, whereas an impact of pre-show comments is consistent with all three channels and will require further investigation. We also note that for post-show comments to influence viewership, we require consumers to be able to anticipate the (lack of) consumption of post-show microblogging activity when deciding which show to watch. Due to the salient announcement of the block (see Figure 4), this condition is likely to be met in our context.

Based on this reasoning, the ideal variation to distinguish between complementarity and the other channels would be random variation in whether pre- or post-show comments were disabled across shows. In our case, the censorship event eliminated both pre and post comments for all shows in our sample, and we hence do not have a source of exogenous variation that differentially affects both types of comments. In the absence of such variation, we turn to exploring effect heterogeneity across different types of shows in the following way: for each show in our data, we compute the average number of comments per episode that were posted before and after the show aired.<sup>42</sup> We then analyze whether the block differently affected shows that typically had more activity before versus after the show. The idea behind this analysis is that shows with predominantly pre-show activity suffered mostly in terms of losing that pre-show activity during the block and vice versa for shows with mostly post-show comments. Hence, if post-show activity is the main driver of ratings, we should see shows with mostly post-show activity experience a larger ratings drop during the block.

We report results from the relevant regressions in Table 5. We start by replicating column (1) of

---

<sup>42</sup>In section 4.2, we used the total number of comments throughout the day on which a specific show aired. We now categorize comments with regards to whether they were posted before/during/after the episode aired. We primarily focus on pre- and post-show comments, but also investigate the impact of comments posted during the show. When including during-show comments in addition to pre- and post-show comments, we find during-show comments have no significant impact on ratings, and the coefficients on pre and post comments are similar to the specification without during-show comments presented in Table 5.

Table 3 in the first column, which splits shows by the average amount of activity (regardless of the timing of comments) and interacts dummies for high- and medium-activity shows with a dummy for the block. We then run the same regression, but separately split the number of pre- and post-show comments into three equally sized bins each and include interaction terms of the block dummy with those variables in the regression reported in column (2). Doing so, we find that shows with high post-show microblogging activity experienced the strongest decline in ratings during the censorship event. Instead, shows with high pre-show activity did not experience a drop in ratings during the block of Sina Weibo. Based on the reasoning presented above, we take these patterns as evidence of a complementarity effect rather than an informative or persuasive advertising effect.

We then delve deeper into the nature of post-show comments that are predictive of the impact of the block. We first include the number of post-show comments expressing sentiment (again using a three-bin split) on top of the total number of post-show comments. Doing so, we find the coefficient on high total post-show comments becomes insignificant, whereas the coefficient on high post-show comments expressing sentiment is negative and significant (see column (3) of Table 5). These results show that enthusiastic post-show discussions are the primary driver of complementarity with show viewership. The post-show comments that do not express sentiment do not seem to affect viewership of the show. When we further split comments into those containing positive and negative sentiments in column (4) of Table 3, we find both types of comments lead to a larger drop in ratings during the block. This finding is particularly interesting because most prior papers that study valence find negative WOM leads to a decrease in product demand (see, e.g., Sonnier et al. (2011)). However, most of these papers have a mechanism in mind whereby WOM affects the consumer prior to making a purchase decision, and hence negative content will lead to a decrease in purchases. Instead, in the case of complementarity, negative comments are not necessarily a bad thing. Instead, they could simply be the sign of a heated debate among viewers of the show. We conclude that a stronger emotional engagement on Sina Weibo after an episode airs is the type of activity that most strongly affects viewership.

We also note the webpage layout for comments on Sina Weibo is one that allows for a discussion between users, because it displays comments in a common thread underneath the original tweet (see Figure 3). The users who contribute comments do not need to follow each other and can use the comment section to participate in a conversation with others who are interested in the same subject. The interactive element of comments is therefore conducive to generating complementarity between TV show viewership and post-show consumption of microblogging content. We furthermore conjecture this mechanism might be unique to comments and applies less to tweets and re-tweets, because they do not allow for a dialogue between users in the same way comments do. Therefore, our results with regards to consumption complementarity might not extend to other forms of microblogging activity. As an additional piece of evidence that user engagement is high for post-show comments, we also note the fraction of comments expressing sentiment is higher for post-show relative to pre-show comments (see Table A7 in the appendix).

Finally, we investigate whether the level of commenting activity either before or after a show

Dependent Variable	(1) Log Rating	(2) Log Rating	(3) Log Rating	(4) Log Rating
Censor Dummy	-0.005 (0.005)	-0.001 (0.007)	-0.002 (0.007)	-0.002 (0.007)
Medium <b>Daily</b> Activity × Censor Dummy	-0.008 (0.011)			
High <b>Daily</b> Activity × Censor Dummy	-0.026** (0.012)			
Medium <b>Pre-Show</b> Activity × Censor Dummy		-0.007 (0.010)	-0.007 (0.012)	-0.007 (0.012)
High <b>Pre-Show</b> Activity × Censor Dummy		0.011 (0.020)	0.024 (0.019)	0.028 (0.019)
Medium <b>Post-Show</b> Activity × Censor Dummy		-0.007 (0.009)	-0.007 (0.009)	-0.008 (0.009)
High <b>Post-Show</b> Activity × Censor Dummy		-0.041** (0.020)	0.001 (0.018)	0.005 (0.019)
Medium <b>Post-Show (Any) Sentiment</b> Comments × Censor Dummy			0.007 (0.014)	
High <b>Post-Show (Any) Sentiment</b> Comments × Censor Dummy			-0.060*** (0.016)	
Medium <b>Post-Show Positive Sentiment</b> Comments × Censor Dummy				0.017 (0.014)
High <b>Post-Show Positive Sentiment</b> Comments × Censor Dummy				-0.039** (0.017)
Medium <b>Post-Show Negative Sentiment</b> Comments × Censor Dummy				-0.017 (0.014)
High <b>Post-Show Negative Sentiment</b> Comments × Censor Dummy				-0.041** (0.018)
Show FEs	Yes	Yes	Yes	Yes
Day of the Week Dummies	Yes	Yes	Yes	Yes
Observations	7,899	7,899	7,899	7,899
Shows	193	193	193	193
R <sup>2</sup>	0.881	0.881	0.881	0.881

Table 5: **Timing and Content: The Differential Impact of Weibo Activity.** The unit of observation is an episode. Standard errors are clustered at the show level.

is systematically correlated with show characteristics such as genre, whether the show is a re-run, and so on. We find that shows with high pre-show commenting activity and shows with high post-show commenting activity have similar characteristics (i.e., they tend to be of a similar genre,



are less likely to be a re-run, etc.).<sup>43</sup> This finding reassures us a correlation of high post-show (relative to pre-show) activity with other show characteristics is not the reason for a differential decrease in ratings for shows with a high level of post-show comments. Table A8 in the appendix provides detailed descriptive statistics on the distribution of show characteristics as a function of commenting activity before and after a particular show.

## 6.2 Informative and Persuasive Effects

Due to the support we find for complementarity over an informative or persuasive effect, the distinction between the latter two is more secondary in our context. Nevertheless, the insignificant impact of pre-show comments could be masking the fact that some types of pre-show comments do in fact influence ratings. We therefore run a set of regressions (not reported in the table) in which we include the number of pre-show comments that are informative or express sentiment. The idea behind this regression is that an informative effect would predict that comments that are informative in nature have a larger impact, whereas a persuasive effect would suggest comments expressing sentiment have a greater impact. When including both types of comments both individually and separately in the regression, we find they are consistently insignificant. We therefore conclude our analysis does not support the notion that WOM affects ratings via an informative or persuasive channel.

## 6.3 Other Patterns in Support of Consumption Complementarity

A few other patterns of effect heterogeneity across different types of shows also provide some additional evidence for the consumption complementarity channel identified above. First, we find no difference in the ratings drop during the block for new versus established shows (see footnote 28 in section 4.2). If informative effects were important, we would expect to see a larger impact for new shows of which consumers are less aware. The absence of a differential effect for new shows is therefore consistent with the absence of informative effects.

Second, we present evidence in the next section that WOM does not have any longer-term effects on ratings. In other words, after controlling for comments pertaining to the current episode, comments that were posted about prior episodes do not have any impact on current ratings. If the anticipation of the microblogging activity after the show is the main driver of ratings, we would not expect past microblogging activity to affect current ratings. Instead, other types of mechanisms, most importantly the persuasive channel, are more likely to imply dynamic effects in the sense that consumers who are exposed to WOM in multiple time periods are more likely to be persuaded to watch a specific show. We therefore see the absence of dynamic effects as an additional piece of evidence in support of the complementarity channel.

---

<sup>43</sup>Similarly, the difference in characteristics between shows with low/medium/high pre-show comments is also small, albeit slightly more pronounced. The same is true when comparing characteristics across shows with low/medium/high levels of post-show comments.

	(1)	(2)	(3)	(4)
Dependent Variable Type of Regression	Log Ratings IV 2nd Stage	Log Ratings IV 2nd Stage	Log Ratings IV 2nd Stage	Log Ratings IV 2nd Stage
Log Number of Comments	0.016*** (0.005)	0.015*** (0.005)	0.015*** (0.004)	0.017*** (0.005)
Lagged Log Comments (n-1)		0.00005 (0.00496)	-0.00437 (0.00516)	-0.00398 (0.00514)
Lagged Log Comments (n-2)			0.00530 (0.00512)	0.00396 (0.00568)
Lagged Log Comments (n-3)				0.00048 (0.00487)
First-Stage F-Stat(s)	105.14	59.54, 64.31	40.54, 47.50 43.75	30.72, 36.13 38.07, 32.13
Show FEs	Yes	Yes	Yes	Yes
Day of the Week Dummies	Yes	Yes	Yes	Yes
Observations	7,899	7,733	7,567	7,399
Shows	193	193	193	193
R <sup>2</sup>	0.877	0.880	0.880	0.882

Table 6: **Dynamic Effects.** The unit of observation is an episode. Standard errors are clustered at the show level. Lagged comments are instruments with equivalently lagged censor dummies.

## 7 Dynamics and Long-Term Effects

In this section, we explore whether changes in microblogging activity have any longer-term impact beyond influencing ratings of the current episode. To analyze dynamic effects of microblogging onto TV show ratings, we re-run our baseline regression and include lagged comments (with different lag structures) as additional regressors. In keeping with our identification strategy for the contemporaneous effect of comments on ratings, we instrument each lagged-comments variable with an equivalently lagged censor dummy variable. In this way, we are only exploiting variation in past commenting activity that is caused by the censorship event. Column (1) in Table 6 replicates our baseline regression without lagged effects. In column (2), we add comments pertaining to the previous episode of the same show as an additional regressor. As before, the relevant set of comments is defined as all comments posted on the day the specific episode aired. Lagged comments are defined in the same way, but in relation to the day on which the previous episode aired. When including the “one-episode-lagged” number of comments, we find the coefficient on the lagged term to be small and statistically insignificant. The coefficient on contemporaneous comments remains almost unchanged and statistically significant. We further explore regressions with up to three periods of lags, and consistently find small and insignificant effects on all lagged terms. We also note the first-stage regressions have strong predictive power. The table reports the F-stat for all first-stage

regressions for each specification.<sup>44</sup>

In summary, the results reported above show no evidence of dynamic WOM effects in our setting, and only contemporaneous microblogging activity has an impact on TV show viewership. This finding is in contrast to several papers in the previous WOM literature that have emphasized the importance of dynamic considerations (Villanueva et al. (2008), Sonnier et al. (2011)). At the same time, as discussed in the previous section, the absence of a dynamic effect is consistent with Sina Weibo comments affecting ratings due to consumption complementarity between TV viewing and microblogging.

## 8 Conclusion

Promoting products through social media websites such as Twitter, and allowing users to discuss and voice their opinions about a product, has become a common practice for many firms. But whether this new marketing channel can effectively enhance demand remains unclear. Similar to evaluating the effectiveness of other types of marketing activity, measuring the effect of microblogging using field data is challenging, and the correlation between microblogging activity of a product and its demand does not necessarily imply causality. Furthermore, obtaining credible evidence on the impact of WOM is particularly challenging because the firm does not directly control WOM, and hence field experiments are difficult to implement. In this paper, we leverage a natural experiment to identify the causal effect of WOM on product demand, and investigate the mechanism through which WOM affects demand.

Several novel findings emerge from our analysis. First, the magnitude of the estimated WOM elasticity is significantly lower than the magnitudes obtained in previous studies. Our findings therefore caution against overstating the impact of WOM, and highlight the importance of using exogenous variation to obtain a causal estimate. The difference in magnitude is particularly striking in our setting: our estimate is more than an order of magnitude lower than many other estimates in the WOM literature. Furthermore, although correlational estimates suggest WOM is more effective than TV advertising (0.12 for TV advertising (Sethuraman et al. (2011)) vs. 0.2 for WOM (You et al. (2015))), the opposite is true for causal estimates in both realms (0.016 for WOM obtained in this study vs. 0.03 for TV advertising). We note the set of causal estimates for both WOM and TV advertising is much smaller, and therefore further work is required to establish whether the relative performance of both channels generalizes beyond the existing studies.

Second, we analyze the mechanism by which WOM influences demand in our setting, and find that complementarity between TV and microblogging consumption is the main driver. Particularly, we find the anticipation of more post-show microblogging activity leads to an increase in TV ratings. Several intriguing managerial implications emerge from this analysis. First, rather than attempting to increase WOM prior to the show airing in order to remind or persuade con-

---

<sup>44</sup>When including lagged terms, the regressions contain multiple endogenous regressors; therefore, a separate first stage (and a separate F-stat) exists for each of the endogenous variable. We report the F-stats for all excluded instruments for each of the first-stage regressions in each case.

sumers, fostering an active discussion after the show is more important. Second, contrary to other mechanisms, consumption complementarity does not entail any dynamic effects of microblogging, which is consistent with what we find empirically. Finally, our findings show both positive and negative post-show comments have a positive impact on ratings, and therefore a discussion that engages consumers appears to be the key driver of ratings. Our findings with regards to negative comments are particularly interesting because one would expect them to decrease ratings if the effect operated through informative or persuasive channels. Instead, the case is less clear for our proposed channel of complementarity with post-show microblogging, and the conventional wisdom of avoiding negative content might not necessarily apply here.

In summary, both with regards to the effect magnitude as well as the proposed mechanism and its implications, our paper provides a set of novel findings that further the understanding of the impact of WOM on demand.

## References

- AALTONEN, A. AND S. SEILER (2016): “Cumulative Growth in User Generated Content Production: Evidence from Wikipedia,” *Management Science*, 62, 2054–2069.
- AHN, D.-Y., J. A. DUAN, AND C. F. MELA (2016): “Managing User-Generated Content: A Dynamic Rational Expectations Equilibrium Approach,” *Marketing Science*, 35, 284–303.
- AMERI, M., E. HONKA, AND Y. XIE (2016): “Word-of-Mouth, Observational Learning, and Product Adoption: Evidence from an Anime Platform,” *Working Paper*.
- ANDERSON, M. AND J. MAGRUDER (2012): “Learning from the Crowd: Regression Discontinuity Estimates of the Effects of an Online Review Database,” *Economic Journal*, 122, 957–989.
- ARAL, S. AND D. WALKER (2014): “Tie Strength, Embeddedness, and Social Influence: A Large-Scale Networked Experiment,” *Management Science*, 60, 1352–1370.
- BAGWELL, K. (2007): “The Economic Analysis of Advertising,” in *Handbook of Industrial Organization*, ed. by M. Armstrong and R. Porter, Elsevier Science, vol. 3, 1701–1844.
- BECKER, G. S. AND K. M. MURPHY (1993): “A Simple Theory of Advertising as a Good or Bad,” *The Quarterly Journal of Economics*, 108, pp. 941–964.
- BERTRAND, M., E. DUFLO, AND S. MULLAINATHAN (2004): “How Much Should We Trust Differences-In-Differences Estimates?” *The Quarterly Journal of Economics*, 249–275.
- BLAKE, T., C. NOSKO, AND S. TADELIS (2015): “Consumer Heterogeneity and Paid Search Effectiveness: A Large Scale Field Experiment,” *Econometrica*, 83, 155–174.
- 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.
- CHINTAGUNTA, P. K., S. GOPINATH, AND S. VENKATARAMAN (2010): “The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation Across Local Markets,” *Marketing Science*, 29, 944–957.
- DHAR, V. AND E. A. CHANG (2015): “Does Chatter Matter? The Impact of User-Generated Content on Music Sales,” *Journal of Interactive Marketing*, 23, pp. 300–307.
- GILCHRIST, D. S. AND E. G. SANDS (2016): “Something to Talk About: Social Spillovers in Movie Consumption,” *Journal of Political Economy*, 24, 1339–1382, press: Marginal Revolution, September 1, 2016 Boston Globe, September 9, 2016.
- 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 (2017): “Tweeting as a Marketing Tool - Field Experiment in the TV Industry,” *The Journal of Marketing Research*, forthcoming.
- GORDON, B., F. ZETTELMEYER, N. BHARGAVA, AND D. CHAPSKY (2016): “A Comparison of Approaches to Advertising Measurement: Evidence from Big Field Experiments at Facebook,” *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 (2017): “Super Bowl Ads,” *Marketing Science*, forthcoming.
- 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 (2015): “The Unfavorable Economics of Measuring the Returns to Advertising,” *The Quarterly Journal of Economics*, 130, 1941–1973.
- LEWIS, R. A. AND D. H. REILEY (2014): “Online ads and offline sales: measuring the effect of retail advertising via a controlled experiment on Yahoo!” *Quantitative Marketing and Economics*, 12, 235–266.
- LIU, Y. (2015): “Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue,” *Journal of Marketing*, 70, pp. 74–89.
- LOVETT, M. J. AND R. STAELIN (2016): “The Role of Paid, Earned, and Owned Media in Building Entertainment Brands: Reminding, Informing, and Enhancing Enjoyment,” *Marketing Science*, 35, pp. 142–157.
- LUCA, M. (2016): “Reviews, Reputation, and Revenue: The Case of Yelp.com,” *Working Paper*.
- PETROVA, M., A. SEN, AND P. YILDIRIM (2016): “Social Media and Political Donations: Evidence from Twitter,” *Working Paper*.
- ROSSI, P. E. (2014): “Invited Paper – Even the Rich Can Make Themselves Poor: A Critical Examination of IV Methods in Marketing Applications,” *Marketing Science*, 33, 655–672.
- SAHNI, N. (2015): “Effect of Temporal Spacing between Advertising Exposures: Evidence from an Online Field Experiment,” *Quantitative Marketing and Economics*, 13, pp. 203–247.
- 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. (2016): “Positive Spillovers and Free Riding in Advertising of Prescription Pharmaceuticals: The Case of Antidepressants,” *The Journal of Political Economy*, forthcoming.
- 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.
- SINKINSON, M. AND A. STARC (2015): “Ask Your Doctor? Direct-to-Consumer Advertising of Pharmaceuticals,” *Working Paper*.
- 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.
- 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.

- TRUSOV, M., R. E. BUCKLIN, AND K. PAUWELS (2009): “Effects of Word-of-Mouth Versus Traditional Marketing: Findings from an Internet Social Networking Site,” *Journal of Marketing*, 73, 90–102.
- TUCHMAN, A. (2016): “Advertising and Demand for Addictive Goods: The Effects of E-Cigarette Advertising,” *Working Paper*.
- TUCHMAN, A., H. S. NAIR, AND P. GARDETE (2016): “An Empirical Analysis of Complementarities between the Consumption of Goods and Advertisements,” *Working Paper*.
- TUCKER, C. E. (2014): “Social Networks, Personalized Advertising, and Privacy Controls,” *Journal of Marketing Research*, 51, 546–562.
- VILLANUEVA, J., S. YOO, AND D. M. HANSSENS (2008): “The Impact of Marketing-Induced Versus Word-of-Mouth Customer Acquisition on Customer Equity Growth,” *Journal of Marketing Research*, 45, 48–59.
- YOU, Y., G. G. VADAKKEPATT, AND A. M. JOSHI (2015): “A Meta-Analysis of Electronic Word-of-Mouth Elasticity,” *Journal of Marketing*, 79, 19–39.
- 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.
- ZHU, F. AND X. M. ZHANG (2010): “Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics,” *Journal of Marketing*, 74, 133–148.

## A Appendix: Data Collection

During our data collection, Sina Weibo changed the way tweets are displayed when querying the webpage for a specific keyword. 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. In the paragraphs below, we outline in detail how we deal with adjusting the pieces of data we obtained after June 2015, in order to make these data comparable to the data collected earlier, 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, we obtained the daily number of tweets 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 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 is unlikely to occur within a day.

We employ a similar type of rescaling to obtain the number of comments, re-tweets, and likes. 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, re-tweets, or likes.

## B Interest in the Scandal: Additional Evidence

In this section, we provide more details on the broader assessment of how much attention people paid to the scandal (summarized in the final paragraph of section 3 in the main text).

First, we widened the set of keywords for which we track the Baidu search volume. To implement such an analysis, we analyzed news articles pertaining to the scandal in order to identify a set of keywords that are relevant to the scandal and that users might be searching for on Baidu. Based on this preliminary analysis, we then computed the Baidu search volume from January to June 2012 for the following search queries: “Gu Kailai” (Bo Xilai’s wife), “Bo Guagua” (Bo Xilai’s son), “Chongqing” (the city where the scandal started), and “US consulate” (Wang’s visit to the US consulate triggered the scandal). Those four graphs are presented in Figure A3. For an easier comparison, we also include the two graphs on “Bo Xilai” and “Wang Lijun” already shown in the paper. Although the intensity of activity varies, across all graphs, we can clearly distinguish the three salient events of the scandal. The individual keywords load relatively more on some of the events in an intuitive way. For instance, the first event is Wang Lijun’s visit to the US consulate, and hence “US consulate” search queries spike during that time (but less so during the other two major events). Most importantly for our analysis, not one of the graphs shows a spike during the censorship event in late March / early April.

Second, we obtained additional data on news coverage pertaining to the political events from Baidu news as well as Google news. We use these data to construct a set of graphs in the following way: for every day during the January to June 2012 window, we count the number of news articles published on either website that contain the name “Bo Xilai” or “Wang Lijun”.<sup>45</sup> Figure A4 displays the time series of news coverage, for Baidu news and Google news, respectively. A similar picture (relative to the Baidu search graphs) emerges from these graphs. We observe spikes in the number of relevant articles around the three major events of the scandal, but no uptake in news coverage around the time of the censorship. The Google news graph shows slightly less pronounced movements around the main events of the scandal. However, the level of coverage on Google is generally quite low, and more importantly, no increase occurs in the level of coverage around the censorship event.

Finally, we provide a third piece of evidence that indirectly tracks interest in the scandal along yet another dimension. Specifically, Internet users in China might have been seeking out information about the scandal from sources of Western news or social media platforms that were blocked in mainland China (e.g., Twitter, Facebook, etc.). To do so, Chinese Internet users needed to use a VPN software product. We can capture some of this activity by computing the Baidu search volume

---

<sup>45</sup>We count any article that contains one of the names anywhere in the body of text. The name does not necessarily have to appear in the title of the article.

for the query “Fan Qiang,” a keyword Chinese people use to find VPN software to circumvent the restrictive Internet control. Although the search for VPN software is not specific to the political scandal, we still expect it to correlate with access to censored sources during these politically sensitive events. The time series of Baidu searches for this keyword is reported in Figure A5. Similar to the other patterns reported above, we again find spikes in activity around the time of the main events of the scandal, but no activity increase at the time of the censorship event.

Taken together, the different graphs present a fairly exhaustive picture regarding the information users consumed. Importantly for our context, all of those pieces of evidence show no spike in interest in the scandal around the time of the censorship.

## C Impact on TV Advertising Revenue

As an alternative way to 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  $(ad\ price * ad\ duration * rating)$ .<sup>46</sup> Because of the multiplicative structure of this pricing formula, the estimated 1.6% increase in ratings when doubling the number of comments will translate into a 1.6% 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 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.<sup>47</sup>

Remember that in the level specification, we estimated an increase of 0.024 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\$25,000 ( $ad\ price * ad\ duration * \Delta rating = 3,700 * 277 * 0.024 = 25,000$ ). A one-standard-deviation shift implies a change in the number of comments of about 77,000 (see Table 1), which entails a change in advertising revenue per episode of over US\$190,000 ( $=25,000 * 77,000 / 10,000$ ). Hence, the impact of microblogging on advertising revenue is substantial.

---

<sup>46</sup> 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.

<sup>47</sup> 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.

## D Content Analysis: Data Collection

To analyze microblogging content, we randomly sample comments across all shows in our data. The sampling procedure was implemented as follows: we sample up to 200 comments per show. However, not all shows have as many comments. If a show has fewer than 200 comments, we sample all comments of the specific show. In the case of a larger number of comments, we randomly sample 200. This sampling procedure yields a total of 12,266 comments.

Next, we assign comments into three categories. We define a comment as “informative” if it contains information about the show, such as time to be aired, actors/actresses to be featured, plot details, and so on. A comment is considered to represent “sentiment” if it expresses an opinion, attitude, or feeling toward the show using either language or emojis (e.g., “awesome show, cannot wait!,” “worst show ever!,” or a smiley face). A comment can fall into both categories; that is, the two types are not mutually exclusive. Depending on the valence of sentiment, a sentimental comment was further coded into positive or negative sentiment. Finally, some comments fall into a residual category of comments that contain neither information nor sentiment. These comments typically discuss a show’s content in an emotionless tone (e.g., “Haven’t seen XYZ (the name of a talk show host) on TV screen for a while.”). Eight research assistants who are native Chinese speakers and blind to the research purpose coded the content of comments and resolved any disagreements by discussion.

Based on this procedure, for each show, we then compute the fraction of its comments that contain (1) information, (2.1) positive sentiment, (2.2) negative sentiment, or (3) neither. For our analysis involving content in section 6.1, we compute for each show the total number of informative, sentimental, and so on comments per episode. These variables are calculated by multiplying the total number of comments per episode by the fraction of informative, sentimental, etc. comments obtained from the content sample. The lower panel of Table A7 reports summary statistics of the content data.

## E 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 analysis that exploits 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 although 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 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.), we do not observe ratings of local TV channels. 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 with only 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<sup>48</sup> 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).<sup>49</sup> Interestingly, the 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 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.

---

<sup>48</sup>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 are 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.

<sup>49</sup>This regression also includes a linear time trend because total viewership experienced a 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).

## F Appendix: Additional Tables and Figures

(1) # Episodes (# Episodes During Block in Parenthe- ses)	(2) Average Rating Pre- Censor	(3) Average Rating During- Censor	(4) Average Rating Post- Censor	(5) Decrease During Censor	(6) P-value for Test of Difference in Means (During vs. Outside Censor)	(7) Decrease During Censorship (3 Days Before/After Censor Only)
61 (3)	3.646	3.251	3.365	-0.259	0.043	-0.205
61 (3)	2.310	2.047	2.141	-0.181	0.052	-0.200
9 (1)	2.466	1.556	1.952	-0.653	n/a	n/a
34 (4)	1.994	1.659	1.919	-0.295	0.312	n/a
72 (6)	1.048	1.150	1.678	-0.432	0.009	0.0002
39 (6)	1.276	1.257	1.519	-0.210	0.028	-0.011
72 (6)	0.800	0.982	1.485	-0.399	0.004	-0.027
52 (3)	1.059	1.015	1.383	-0.230	0.068	-0.065
30 (6)	1.194	1.060	1.224	-0.152	0.029	-0.128
26 (2)	0.913	0.892	1.265	-0.314	0.048	-0.078

Table A1: **Rating Changes during the Censor: 10 Highest-Rated Shows.** Each row denotes an individual show. Shows are displayed in order of their average rating per episode. No difference-in-means test is available for show (3) because only one episode aired during the block. Shows (3) and (4) did not air any episodes 3 days before/after the block. Ratings are in levels, not log-transformed.

	<i>Mean</i>	<i>S.D.</i>	<i>Min</i>	<i>10th Perc</i>	<i>90th Perc.</i>	<i>Max</i>	<i>Obs</i>
All Shows	0.441	0.438	0.038	0.074	0.935	3.509	193
TV Series	0.744	0.414	0.154	0.207	1.381	1.582	23
Reality Shows	0.511	0.443	0.052	0.154	0.955	2.208	40
Children Shows	0.448	0.185	0.123	0.138	0.744	0.758	16
News	0.325	0.380	0.041	0.052	0.648	2.228	50
Other Shows	0.378	0.481	0.038	0.071	0.688	3.509	64
Established Shows	0.401	0.457	0.038	0.068	0.767	3.509	137
New Shows	0.540	0.374	0.045	0.125	1.062	1.582	56
Re-Run	0.481	0.229	0.123	0.207	0.815	1.001	22
Current Show	0.436	0.458	0.038	0.071	0.935	3.509	171
Daily Frequency	0.451	0.478	0.038	0.065	1.062	3.509	118
Less than Daily Frequency	0.426	0.369	0.043	0.103	0.825	2.209	75

Table A2: **Descriptive Statistics: Ratings Variation across Types of Shows.** The unit of observation is a show.

		<i>Mean</i>	<i>S.D.</i>	<i>90th Perc.</i>	<i>95th Perc.</i>	<i>Max</i>	<i>S.D. (Time Series Only)</i>	<i>Obs</i>
<b><u>Hong Kong Shows</u></b>	Rating	2.85	6.29	12.57	19.78	36.17	0.67	3,528
	Log-rating	0.74	0.90	2.61	3.03	3.61	0.15	3,528
<b><u>Overlapping Shows</u></b>								
Shows of Mainland	Ratings in Shenzhen	0.247	0.392	0.700	1.100	3.120	0.128	7,899
Origin (193 Shows)	Ratings in Hong Kong	0.024	0.032	0.057	0.077	0.455	0.022	7,899

Table A3: **Descriptive Statistics: Show Ratings in Hong Kong & “Overlapping Show” for the Hong Kong and Shenzhen Market.** The top panel shows the rating distribution across episodes for all shows in Hong Kong (not including shows of mainland origin). The lower panel reports ratings in Hong Kong and Shenzhen for all shows of mainland origin.

	(1)	(2)	(3)	(4)	(5)
Dependent Variable	Log-Rating	Log-Rating	Log-Rating	Log-Rating	Log-Rating
Censor Dummy	-0.0168*** (0.0053)	-0.0168*** (0.0053)	-0.0168*** (0.0053)	-0.0168*** (0.0053)	-0.0163*** (0.0052)
Log Search Volume “Bo Xilai”		-0.0003 (0.0013)		0.0002 (0.0024)	
Log Search Volume “Wang Lijun”			-0.0010 (0.0021)	-0.0011 (0.0033)	
Dummy March 14-18 (Bo Xilai Removed from Office)					0.0038 (0.0036)
Dummy April 10-17 (Bo Xilai Arrested)					0.0025 (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 A4: **Robustness Check: Controls for Saliency of the Political Scandal.** The unit of observation is an episode. Standard errors are clustered at the show level.



Type of Analysis	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variable	Mainland & HK		24 Mainland Cities			Across Shows		
	Log-Rating	Log-Rating	Log-Rating	Log-Rating	Log-Rating	Log-Rating	Log-Rating	Log-Rating
Censor Dummy	0.0052 (0.0104)	0.0055 (0.0104)	-0.0098 (0.0065)	-0.0096 (0.0064)	-0.0079 (0.0062)	-0.0077 (0.0061)	-0.0050 (0.0053)	-0.0054 (0.0053)
Medium Activity							-0.0082 (0.0105)	-0.0078 (0.0105)
× Censor Dummy							-0.0259** (0.0115)	-0.0252** (0.0112)
High Activity								
× Censor Dummy								
Mainland Dummy	-0.0260** (0.0118)	-0.0262** (0.0117)						
× Censor Dummy								
SW Penetration			-0.0265* (0.0142)	-0.0266* (0.0142)				
× Censor Dummy								
>Median SW Penet.					-0.0163*** (0.0059)	-0.0163*** (0.0059)		
× Censor Dummy								
Time Trend		-0.00036 (0.00025)		0.00006 (0.00019)		0.00008 (0.00017)		-0.00026 (0.00022)
Mainland Dummy		0.00046 (0.00029)						
× Time Trend								
Sina Weibo Penet.				0.00013 (0.00035)				
× Time Trend								
>Median SW Penet.						0.00001 (0.00012)		
× Time Trend								
Medium Activity								0.00032 (0.00032)
× Time Trend								0.00094** (0.00041)
High Activity								
× Time Trend								
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 A5: **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 (177.480)	12,080 (33,141)	-10,351** (5,173)	-180.841 (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	1.304	1.571	0.716
Censor Dummy	0.072 (0.048)	0.256*** (0.090)	-1.003*** (0.086)	0.014 (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 A6: **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. Each row/column combination displays the result from a separate regression.

	Percentage				Activity per hour (Unit: 10,000)			
	<i>Tweets</i>	<i>Comments</i>	<i>Re-tweets</i>	<i>Likes</i>	<i>Tweets</i>	<i>Comments</i>	<i>Re-tweets</i>	<i>Likes</i>
<b><u>Timing</u></b>								
Pre-show	54.18	62.36	51.54	66.26	1.32	6.99	30.44	0.23
During-show	12.27	12.30	8.72	6.75	13.79	63.64	237.80	1.10
Post-show	33.56	25.35	39.74	26.99	4.63	16.11	133.11	0.54
<b><u>Content (of Comments)</u></b>							<i>Mean</i>	<i>Mean</i>
							<i>(Pre-Show</i>	<i>(Post-Show</i>
Fraction of Comments			<i>Mean</i>	<i>S.D.</i>	<i>10th</i>	<i>90th</i>	<i>Comments)</i>	<i>Comments)</i>
Pertaining to Each	Informative		0.018	0.037	0	0.059	0.019	0.011
Category (Distribution	Sentiment		0.329	0.249	0	0.735	0.303	0.347
Across Shows Is	Positive Sentiment		0.313	0.246	0	0.725	0.293	0.309
Displayed)	Negative Sentiment		0.016	0.026	0	0.043	0.011	0.038

Table A7: **Descriptive Statistics: The Timing and Content of Microblogging Activity.**

	(1)	(2)	(3)	(4)	(5)	(6)
	Pre-Show Comments			Post-Show Comments		
Level of Commenting Activity	Low	Medium	High	Low	Medium	High
<b>Genre</b>						
TV Series	0.030	0.079	0.250	0.036	0.068	0.250
Reality Shows	0.121	0.079	0.047	0.107	0.082	0.063
Children Shows	0.288	0.317	0.172	0.107	0.397	0.234
News	0.394	0.397	0.203	0.518	0.301	0.203
Other Shows	0.167	0.127	0.328	0.232	0.151	0.250
<b>Other Characteristics</b>						
Established Shows	0.667	0.778	0.688	0.589	0.836	0.672
Re-Run	0.136	0.095	0.109	0.143	0.096	0.109
Daily Frequency	0.591	0.683	0.563	0.500	0.726	0.578

Table A8: **Show Characteristics by Level of Pre- and Post-Show Commenting.** Each cell shows the fraction of shows with the row characteristics among shows with a specific level of commenting (as indicated by the column heading).

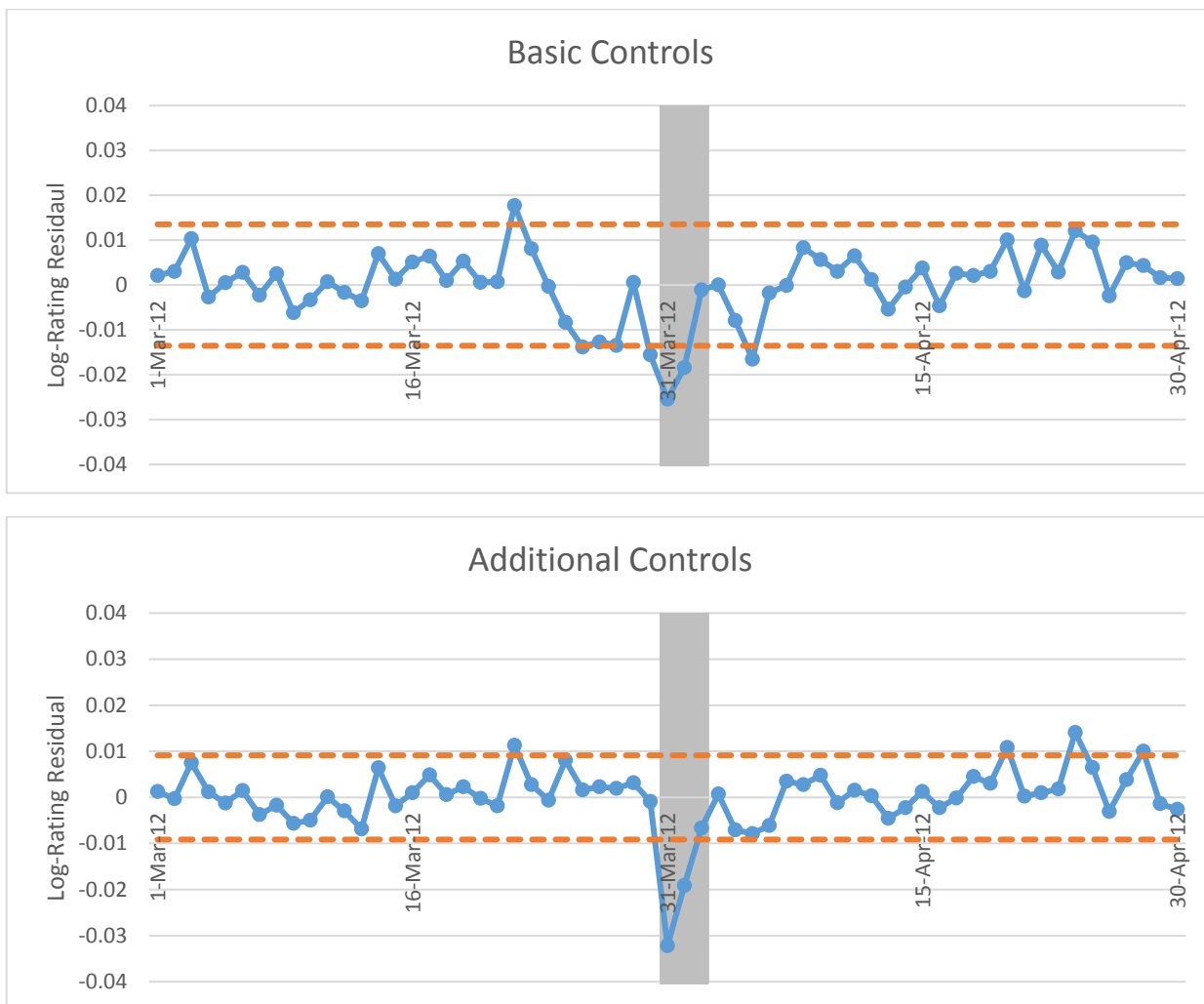


Figure A1: **The Evolution of Ratings over Time with Basic and More Extensive Controls.** The graphs plot the average daily value of the residuals from a regression of  $\log(\text{Rating})$  onto control variables. Basic controls (top panel) include show fixed effects and weekday dummies. Additional controls include show/weekday-pair fixed effects as well as controls for holidays and the duration of the table tennis world tournament. Both graphs are based on all shows that aired at least one episode in March or April 2012 (our regression excludes shows that did not air an episode during the block, because those shows do not provide variation to identify the censorship effect). The gray band indicates the three days of the block. Dashed lines indicate  $2 \times \text{SD}$  above and below the average daily residual in each graph.



Figure A2: Geographical Distribution of the Cities in Our Data: 24 Mainland Cities and Hong Kong

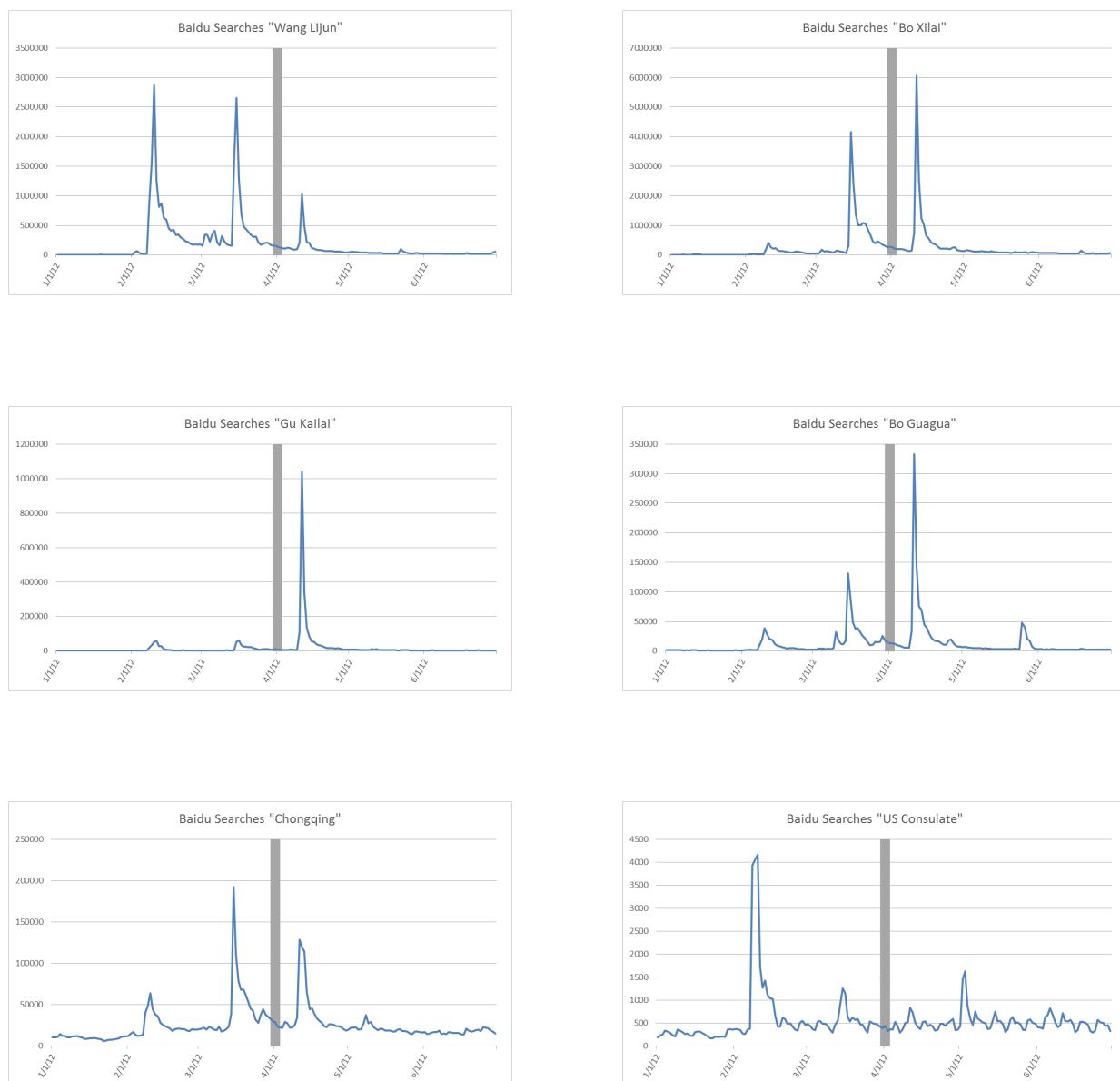


Figure A3: **Baidu Search Volume for Keywords Related to the Political Scandal.** The graphs display the daily search volume on Baidu from January to June 2012 for the respective keyword. The gray bands indicate the three days of the block.

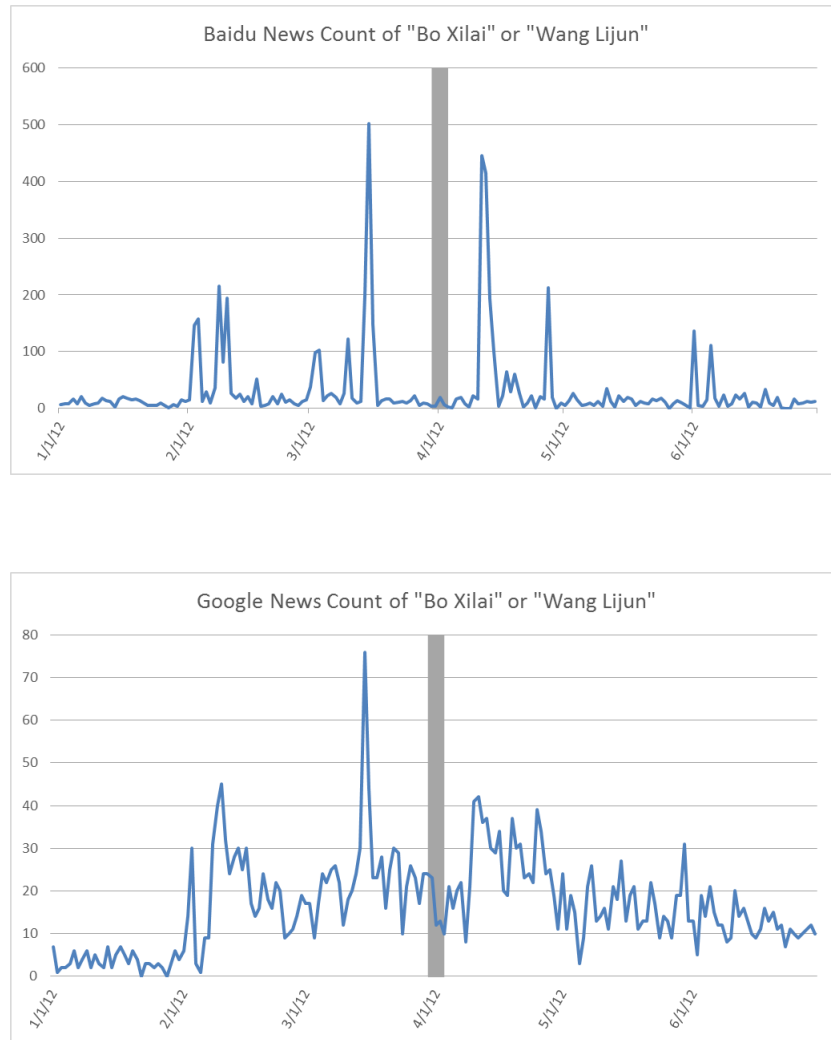


Figure A4: **Baidu/Google News Coverage Related to the Political Scandal.** The graphs display the daily count of news stories on Baidu and Google News containing the respective name from January to June 2012. The gray bands indicate the three days of the block.



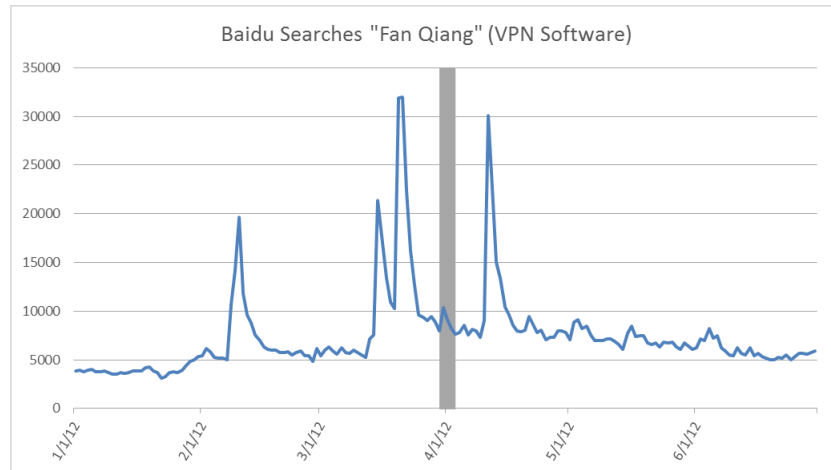


Figure A5: **Baidu Search Volume for the Keyword “Fan Qiang” (VPN Software).** The graph displays the daily search volume on Baidu from January to June 2012 for the keyword “Fan Qiang.” The gray band indicates the three days of the block.

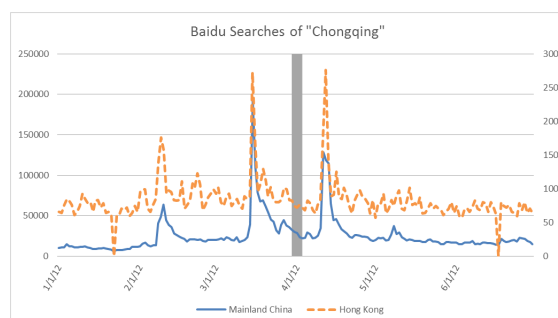
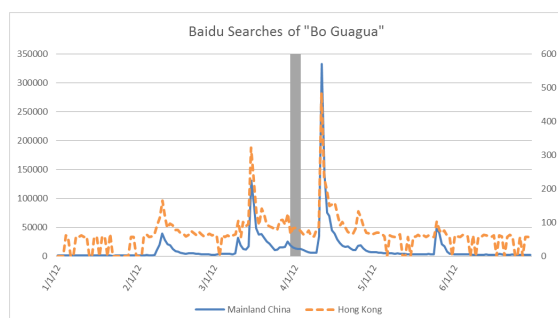
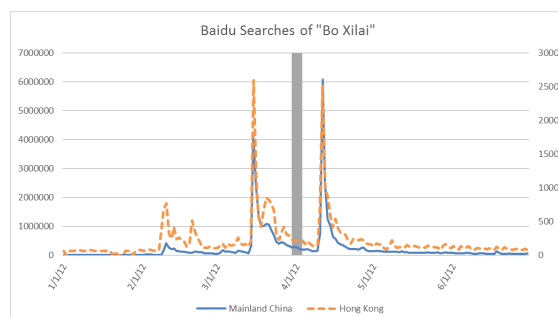
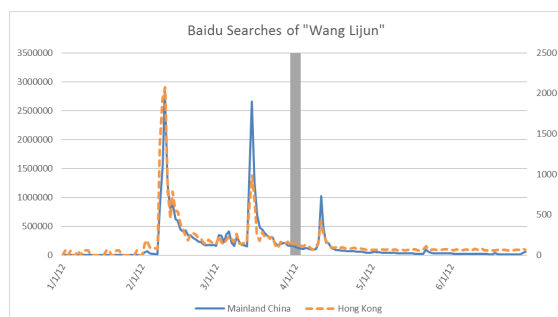


Figure A6: **Comparison of the Baidu Search Volume between Mainland China and Hong Kong.** The graphs display the daily search volume on Baidu from January to June 2012 for the respective keyword in mainland China and Hong Kong. The gray bands indicate the three days of the block.