Online Appendix: 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

A 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".[45] 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 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 $(ad\ price * ad\ duration * rating)$.[46] 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

[45] 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.

[46] 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.

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.[47]

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.

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 valance 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

[47] 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.

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[48] 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).[49] 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.

[48]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.

[49]This regression also includes a liner 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 Parentheses)	(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.

	Mean	*S.D.*	*Min*	*10th Perc*	*90th Perc.*	*Max*	*Obs*
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.

		Mean	*S.D.*	*90th Perc.*	*95th Perc.*	*Max*	*S.D. (Time Series Only)*	*Obs*
Hong Kong Shows	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
Overlapping Shows								
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***	-0.0168***	-0.0168***	-0.0168***	-0.0163***
	(0.0053)	(0.0053)	(0.0053)	(0.0053)	(0.0052)
Log Search Volume		-0.0003		0.0002	
"Bo Xilai"		(0.0013)		(0.0024)	
Log Search Volume			-0.0010	-0.0011	
"Wang Lijun"			(0.0021)	(0.0033)	
Dummy March 14-18					0.0038
(Bo Xilai Removed from Office)					(0.0036)
Dummy April 10-17					0.0025
(Bo Xilai Arrested)					(0.0042)
Show FEs	Yes	Yes	Yes	Yes	Yes
Day-of-the-Week Dummies	Yes	Yes	Yes	Yes	Yes
Observations	7,899	7,899	7,899	7,899	7,899
Shows	193	193	193	193	193

Table 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.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Type of Analysis	Mainland & HK		24 Mainland Cities				Across Shows	
Dependent Variable	Log-Rating	Log-Rating	Log-Rating	Log-Rating	Log-Rating	Log-Rating	Log-Rating	Log-Rating
Censor Dummy	0.0052	0.0055	-0.0098	-0.0096	-0.0079	-0.0077	-0.0050	-0.0054
	(0.0104)	(0.0104)	(0.0065)	(0.0064)	(0.0062)	(0.0061)	(0.0053)	(0.0053)
Medium Activity							-0.0082	-0.0078
× Censor Dummy							(0.0105)	(0.0105)
High Activity							-0.0259**	-0.0252**
× Censor Dummy							(0.0115)	(0.0112)
Mainland Dummy	-0.0260**	-0.0262**						
× Censor Dummy	(0.0118)	(0.0117)						
SW Penetration			-0.0265*	-0.0266*				
× Censor Dummy			(0.0142)	(0.0142)				
>Median SW Penet.					-0.0163***	-0.0163***		
× Censor Dummy					(0.0059)	(0.0059)		
Time Trend		-0.00036		0.00006		0.00008		-0.00026
		(0.00025)		(0.00019)		(0.00017)		(0.00022)
Mainland Dummy		0.00046						
× Time Trend		(0.00029)						
Sina Weibo Penet.				0.00013				
× Time Trend				(0.00035)				
>Median SW Penet.						0.00001		
× Time Trend						(0.00012)		
Medium Activity								0.00032
× Time Trend								(0.00032)
High Activity								0.00094**
× Time Trend								(0.00041)
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	12,080	-10,351**	-180.841
	(177.480)	(33,141)	(5,173)	(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	0.256***	-1.003***	0.014
	(0.048)	(0.090)	(0.086)	(0.050)
Show FEs	Yes	Yes	Yes	Yes
Day-of-the-Week Dummies	Yes	Yes	Yes	Yes
Observations	10,126	10,126	10,126	10,126

Table 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)*			
		Tweets	*Comments*	*Re-tweets*	*Likes*	*Tweets*	*Comments*	*Re-tweets*	*Likes*
Timing									
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
Content (of Comments)								*Mean (Pre-Show*	*Mean (Post-Show*
Fraction of Comments				*Mean*	*S.D.*	*10th*	*90th*	*Comments)*	*Comments)*
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
Genre						
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
Other Characteristics						
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).

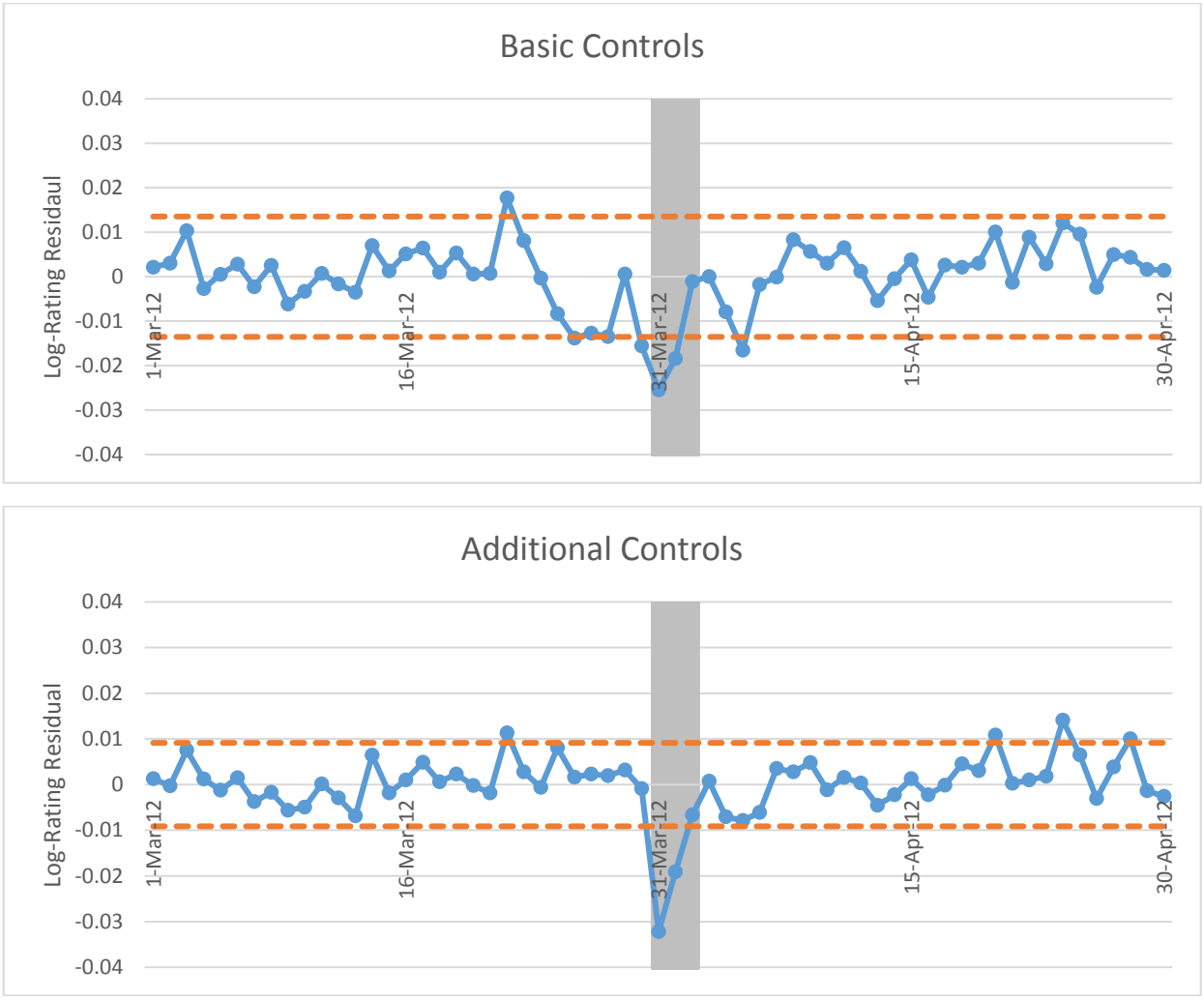

Figure A1: **The Evolution of Ratings over Time with Basic and More Extensive Controls.** The graphs plot the average daily value of the residuals from a regression of log(Rating) onto control variables. Basic controls (top panel) include show fixed effects and weekday dummies. Additional controls include 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*SD above and below the average daily residual in each graph.

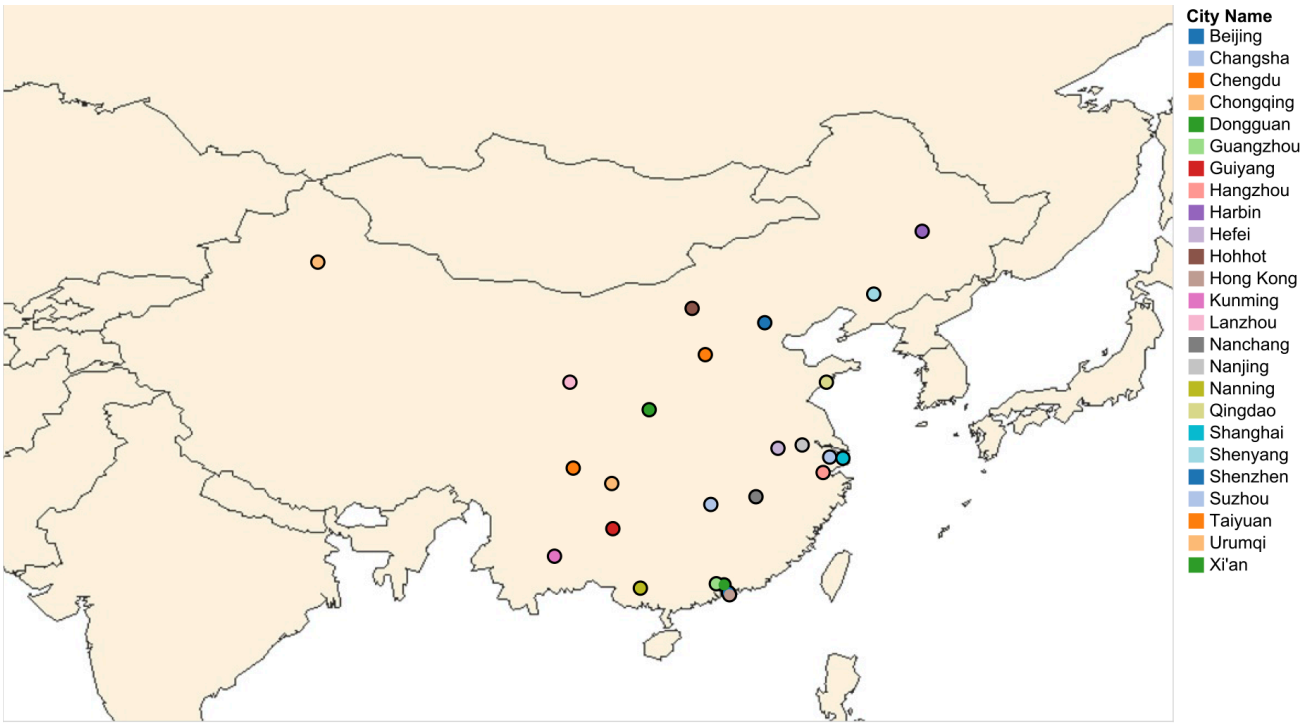

Figure A2: **Geographical Distribution of the Cities in Our Data: 24 Mainland Cities and Hong Kong**

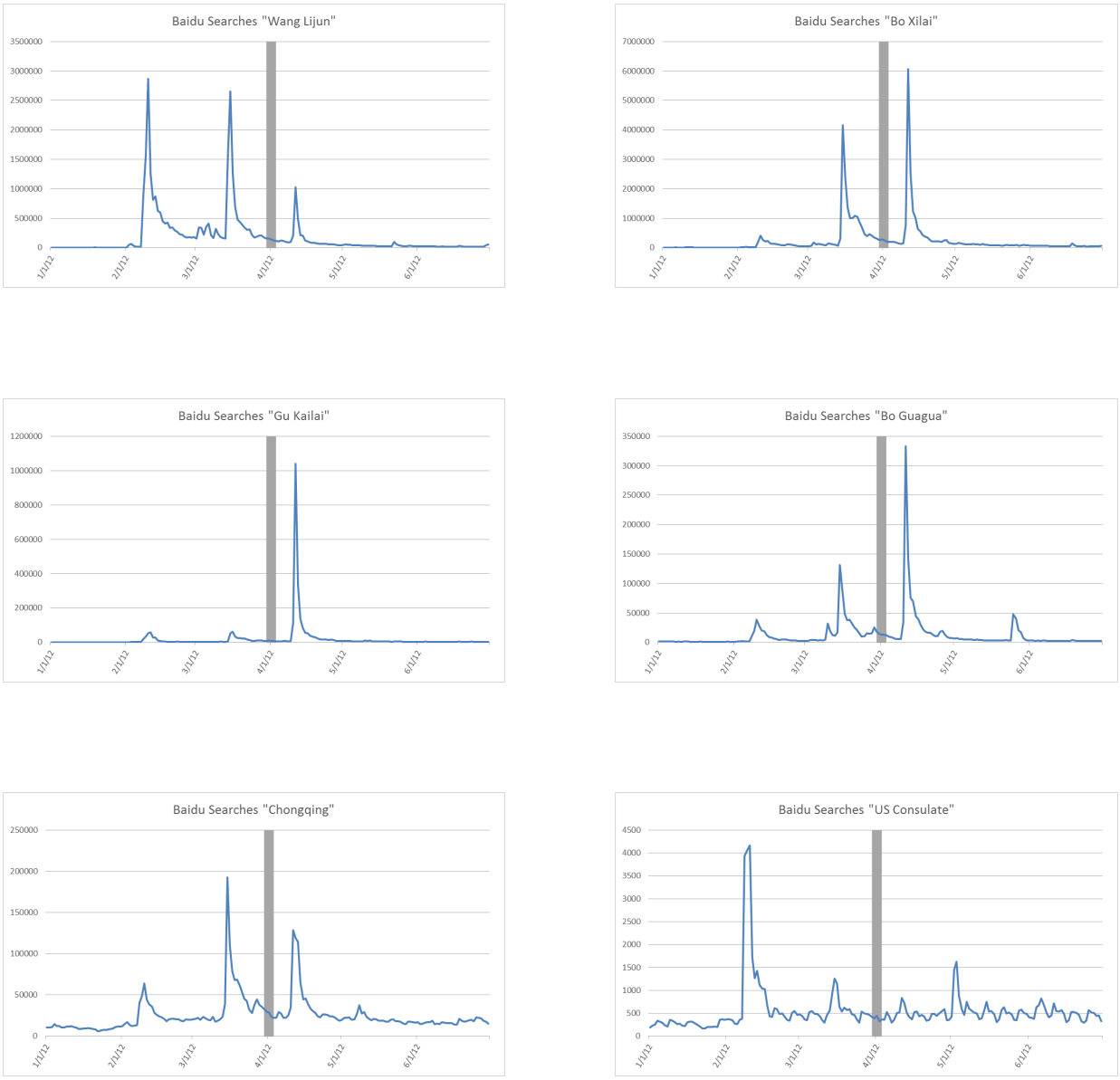

Figure A3: **Baidu Search Volume for Keywords Related to the Political Scandal.** The graphs display the daily search volume on Baidu from January to June 2012 for the respective keyword. The gray bands indicate the three days of the block.

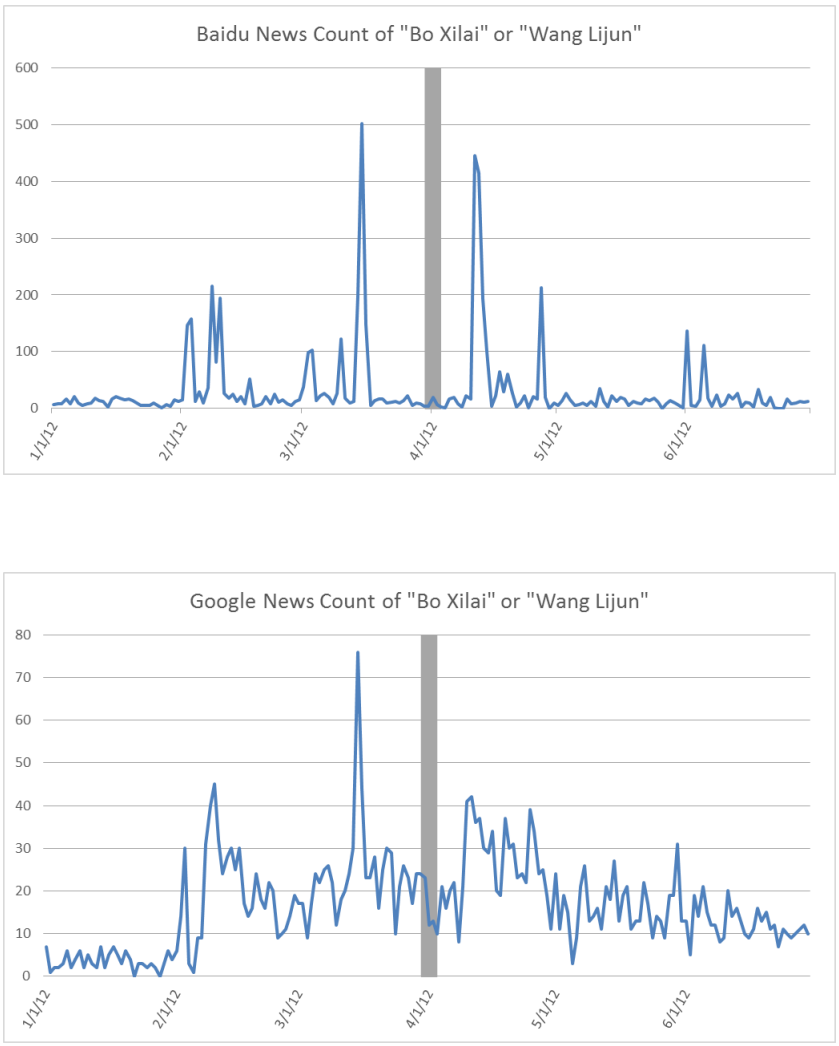

Figure A4: **Baidu/Google News Coverage Related to the Political Scandal.** The graphs display the daily count of news stories on Baidu and Google News containing the respective name from January to June 2012. The gray bands indicate the three days of the block.

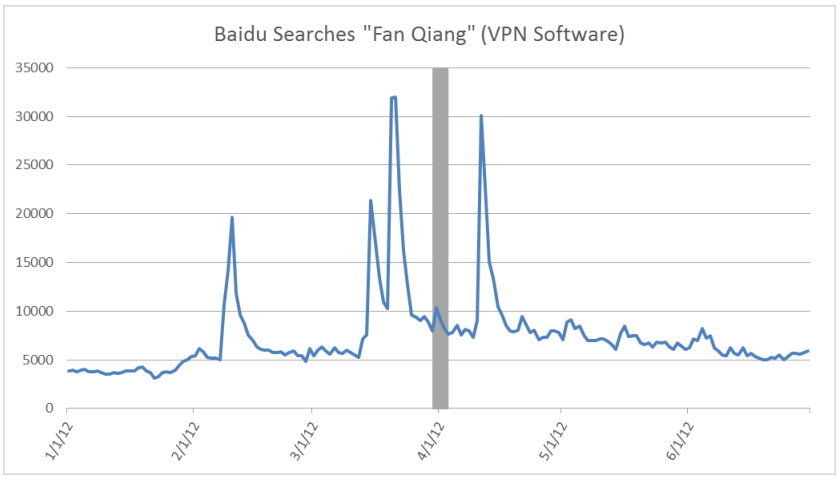

Figure A5: **Baidu Search Volume for the Keyword "Fan Qiang" (VPN Software).** The graph displays the daily search volume on Baidu from January to June 2012 for the keyword "Fan Qiang." The gray band indicates the three days of the block.

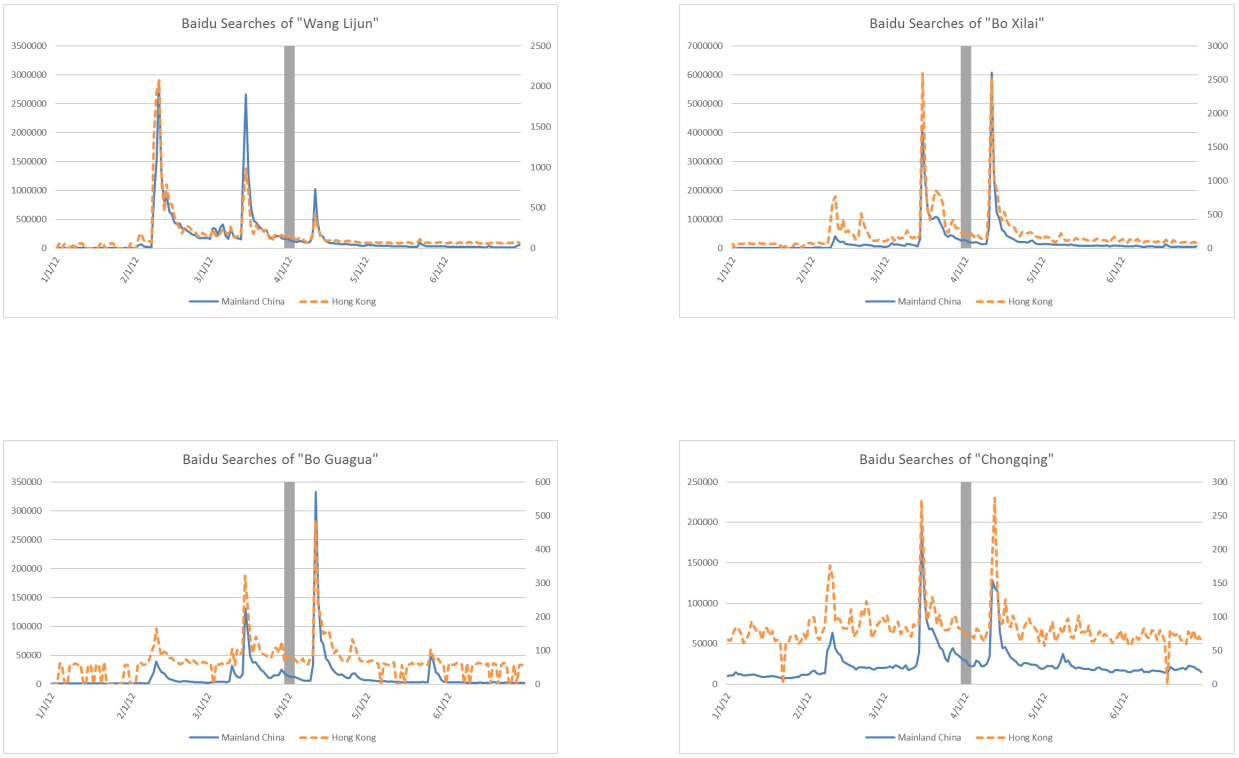

Figure A6: **Comparison of the Baidu Search Volume between Mainland China and Hong Kong.** The graphs display the daily search volume on Baidu from January to June 2012 for the respective keyword in mainland China and Hong Kong. The gray bands indicate the three days of the block.