# Appendix A

This appendix provides information on data construction. The data used in the analysis is built using the following three steps. Figure A1 presents the flow diagram of data processing.

Figure A1: Flow Diagram of Data Processing



## A.1: Enrolment and Raw Data

We used a Weibo message crawler[[1]](#footnote-1) (PMSinaMsgCrawler) to randomly crawl Weibo’s overlay and got 659,692 users’ posts (about 26,608,054 posts) on Weibo from January 1, 2011 to June 30, 2012. All valid user accounts on Sina Weibo have a 10-digit identity code. The crawler generated randomly possible account codes between 1,000,000,000 and the maximum (in this case, 2,909,292,288) and tested their validity using the computer program on the Sina Weibo website. If the account was active, which means the user wrote at least one post per month, we crawled that user’s posts and the related information of each post (number of times a certain post was reposted, number of comments it received, etc.). Consistent with findings of Fu and Chau (2013), we find that the active accounts on average contribute 80% of the total posts. Our sample, and the corresponding number of times posts were reposted, account for 59.97% of the total posts during the study period.

Figure A2 shows the time distribution of the number of posts in our sample. It illustrates a steady increase in the number of posts during this period. The sudden increase from April to May 2012 may be due to the commercialization of Sina Weibo. During this period, many e-commerce shops opened Weibo accounts, and the new e-commerce platform and interactive advertising increased the overall number of posts.[[2]](#footnote-2)

Figure A2: Time Distribution of Sample Posts



In order to test the representativeness of our data, in Figure A3 we compare our sample with the general pattern of registered Weibo users, which is a typical indicator to measure Weibo development before 2013.[[3]](#footnote-3) It shows that both patterns are quite similar.

Figure A3: Time Distribution of Sample Posts and the Number of Weibo Users



## A.2: Experts’ Annotation and Fact Finding

We define scandals as negative information that violates the CCP’s discipline and state laws. According to the CCP’s Criminal Law and Regulations on Disciplinary Sanctions, we divide the scandals into four categories—economic crimes (i.e., bribery, embezzlement and misappropriation), dereliction of duty, sex scandals, and inappropriate speech and behavior. Next, we randomly selected 100,000 posts and assigned three research assistants to independently annotate all the posts and build the scandal-related corpus.[[4]](#footnote-4) We only kept posts that all three RAs classified as scandals to extract the keywords. Table A1 lists all the annotation keywords across the four types of scandals.

Table A1: Annotation Keywords

|  |
| --- |
| **Economic Issues** |
| account | accusation | ambitions | assets | backing |
| bank | banknote | behind the scenes | blackmail | bribe |
| bureau director | business | cash | collude | collusion |
| Commission for Discipline Inspection | consumption | corrupt officials | corruption | cost |
| dark | data | deal with | death penalty | deprivation |
| desire | detect | director-general | discipline | dismiss |
| double designation | end | escape | essential | estate |
| execution by shooting  | expose | expose | extravagant | fame and wealth |
| family clan | fire | flee | foreign | frugal |
| fugitive | grease | greedy | guilt | honest |
| hook up | hotbed | in private | in secret | income |
| injustice | institution | interpersonal relationship | investigation | investment |
| jail | large amount of | large amount of money | law | long-term |
| lost | luxury car | material | material comforts | million |
| money | moneybags | network resources | official | one hundred Yuan |
| overseas | partnership | party member | personnel | position |
| power | profit | project | property | provincial party committee |
| public fund | public vehicles | punish | purpose | report |
| resignation | resources | returns | sanction | satisfy |
| scale | secretary | special task force | stock | surrender |
| temptation | ten million | track | tradition | transaction |
| transfer | treat | up to | use | value |
| villa | wanted | wealth | will | wipe out |
|  |  |  |  |  |
|  |  |  |  |  |
| **Dereliction of Duty** |
| accident | approve | avoid | awe | bidding |
| block | carry out | case | citizens | collusion |
| communication | compensate | conflict | construction | contribution |
| cooperate | counterfeit | countryside | court | cover up |
| cry | demolitions | department | development | dialogue |
| dilemma | dishonest | dislocated | dissatisfaction | distribution |
| duties | effectiveness | embarrass | enforcement | enterprise |
| examine | fiscal | forced confession | framed | function |
| funds | gathering | governance | group | hukou |
| illegal | illicit | image | in principle | industry |
| injured | insensitive | inspect | interviews | intimidate |
| investigate | land | leaders | livelihood | lost |
| misconduct | malfeasance | mediocre officials | mistake | mob |
| on duty | on site | party member | people | petition |
| policy | political performance | pollution | problem | profit |
| project | property right | protest | public | public credibility |
| public employment | public opinion | purpose | real estate | rectify |
| relief | rent seeking | report | reputation | residence |
| resources | response | responsibility | review | risk |
| reimburse | rule of law | service | shady deal | silence |
| strike | subsidy | supervise | suspend | tough |
| trial | truth | typical | unemployment | village head |
| violence | vow | warning | weiquan | work |
|  |  |  |  |  |
|  |  |  |  |  |
| **Sex Scandals** |
| actress | adore | adultery | affairs | affectionate |
| age | ambiguous | amour | appearance | avoid a conflict of interest |
| beauty | buzz | character | child | clarify |
| close | complex | concubine | connection | consumption |
| contact | date | demand | derailed | dinner party |
| divorce | drunk | emptiness | entertainment | enthusiasm |
| exposure | fake news | family | female | figure |
| get a room | get together | gift | give | good-looking |
| gossip | grapevine | group photo | guesthouse | harassment |
| hook | hotels | illegitimate child | image | image |
| infectious | intimate | intimidate | introduce | jewelry |
| journalist | just | lady | lascivious | late at night |
| live together | lonely | lover | made up | maintain |
| maintain | married | media | midnight | mobile phone |
| money worship | monthly | moral | obscene | office secretary |
| on a business trip | peach blossom luck | photograph | place | plump |
| pornographic | position | pregnancy | pretty | private |
| private life | prostitute | raise | rape | regret |
| relationship | room | rumor | sauna | scale |
| scandal | second wife | seduce | service | sexy |
| shopping | SMS | social interaction | somewhere | style |
| temptation | third wheel | trading power for sex | transactional sex | trick |
| unspoken rules | video | whoring | wife | young |
|  |  |  |  |  |
|  |  |  |  |  |
| **Inappropriate Speech and Behavior** |
| absurd | addicted | apologize | arrogant | atmosphere |
| attack | attention | attitude | awareness | background |
| bad | bar | brand | bureaucratic jargon | capricious |
| carat | cause trouble | children of officials | circle | connive |
| constraint | consumption | cost | cover | criticism |
| criticize | dalliance | defamation | delete | deliberately |
| dinner | discourse | dispute | drug | drunk |
| enjoy | entertainment | exaggeration | expect | experience |
| extravagance | fact | famous brand | fashion | feast |
| fight | foreign | forward | framed | free |
| furious | headlines | high profile | human flesh search | hype |
| impoliteness | indulgence | influences | Internet | interview |
| laughing | lesson | limit | limited edition | waste |
| low profile | luxury | make up | momentum | mood |
| negative | netizens | objective | official shield one another | onlookers |
| personality | pleasure | pressure | price | privilege |
| propaganda | protection | public | public area | public opinion |
| public relations | raving | reason | reception | recognition |
| red tape | refute | regret | remarks | report |
| response | scene | select | self-blame | sensation |
| show off | society | spa | speech | squander |
| superstition | swagger | tedious | threaten | tone |
| tourism | upper class | urban management officers | waste |  |

Then we employed a keywords matching method on 26.6 million posts. The keywords are based on the scandal-related corpus we built. This is to identify the scandal-related posts, and we got around 271,000 scandal-related posts. We again employed the human annotation to identify specific cases of scandals. We only kept cases that all three coders considered to be scandals. We obtained 441 cases including officials’ names and the details of the scandal, which included economic crimes (i.e., bribery, embezzlement and misappropriation) (n = 207), dereliction of duty (n = 112), sex scandals (n = 43), and inappropriate speech and behavior (n = 188).

## **A.3: Recalling and Calculating Reposted Times**

To investigate the online discussions on each specific case, we performed a content analysis of the raw data as in A.1 based on the officials’ names. We chose to implement approximate matching based on the similarity between two strings rather than simple string matching, where only identical names are matched. [[5]](#footnote-5) We first subset the text that contains the name of the involved officials. Then, we used the officials’ names and their positions as the substring to compare with the subset text to calculate the distance. In this method, we can exclude those cases sharing same first two characters with the involved two-character names (21% of the cases are two-character names). In particular, we used a python library named Fuzzywuzzy, which used Levenshtein Distance to calculate the differences between sequences and patterns that was developed by SeatGeek (https://github.com/seatgeek/fuzzywuzzy). This algorithm has efficient performance in calculating the similarity between short texts. The partial ratio function in Fuzzywuzzy allows us to perform substring matching. It takes the shortest string and compares it with all the substrings of the same length. We set the similarity threshold as 60% after comparing the different outcomes of the threshold. Then we calculated the total number of times each related post was reposted for each scandal case. The first day when there were several reposts about a scandal was regarded as the “scandal broken” date. We double-checked this date when we searched the Internet to identify the speed of the government response based on the news statement and made adjustments where necessary.

# Appendix B

We define scandals as negative information that violates the CCP’s discipline and state laws. Based on the regulation and laws of the party and the state, we divide the scandals into four categories—economic crimes, dereliction of duty, sex scandals, and inappropriate speech and behaviour.

According to the Criminal Law (Chapter 8) and Regulations on Disciplinary Sanctions of the Communist Party of China, an act is defined as corrupt and violating discipline in an economic aspect if it involves (1) accepting or offering bribes; (2) embezzlement; (3) misappropriation; (4) failure to explain the sources of a large amount of properties; (5) concealment of deposits abroad; (6) dividing state-owned assets; and (7) dividing confiscated property. Officials involved in corruption or violating discipline in an economic aspect will receive penalties from one-year fixed-term imprisonment to death sentence based on the circumstances. The following is a screenshot of an example of economic issues.



According to the Criminal Law and Regulations on Disciplinary Sanctions of the Communist Party of China, an act is defined as a dereliction of duty if it involves (1) abuses of power; (2) neglecting one’s duty; (3) engaging in malpractice for personal gain; (4) violating the provisions of the Law on Guarding State Secrets. Note that there are two types of dereliction of duty in China: the crime of dereliction of duty violating Chapter 9 of the Criminal Law and the dereliction of duty violating the Regulations on Disciplinary Sanctions of the CCP. For the former, the involved officials will receive penalties with no more than ten years of fixed-term imprisonment. For the latter, the involved officials will get less punishment with demotion or removal, etc. The website of the National Supervisory Commission (NSC) has provided detailed explanations on the two types of dereliction of duty.[[6]](#footnote-6) According to their explanation, the NSC acknowledged the difficulties of committing the crimes of dereliction of duty in practice, and most cases of dereliction of duty, in reality, belong to the latter. In terms of our dataset, all of cases belong to the latter. The following is a screenshot of an example of the dereliction of duty.



According to Article 127 of the Regulations on Disciplinary Sanctions of the Communist Party of China, “where improper sexual relations occur with others, causing a negative impact, a sanction shall be given…Where professional power, mentoring relationships, staff relationships or other similar relationships are exploited to have sexual relations with others, a heavier sanction shall be given…”. We define an act as a sex scandal if the official violates Article 127. The involved officials with sex scandals will receive a sanction of expulsion or below. The following is a screenshot of an example of sex scandal.



According to the Regulations on Disciplinary Sanctions of the Communist Party of China, the CCP will sanction behaviours such as “luxurious lifestyles, hedonism, or the pursuit of vulgar interests (Article 126),” “inappropriate acts in a public place that violate the public order and good morals (Article 128)” and “serious violation of social morality (Article 129).” The involved officials with inappropriate behaviours will receive a sanction of expulsion or below. The following is a screenshot of an example of inappropriate speech and behaviour.

****

In the baseline model, we also use the full dependent variable instead of four category simplification to run the baseline model. The results of speed of government response are robust, while the coefficient on officials’ punishment is insignificant. This may be because punishment itself is highly affected by the scandal types. Generally, penalties for different types of scandals are different. For the dereliction of duty, sex scandals, and inappropriate behaviours and talks, they only violate the party’s own intraparty regulations—Regulations on Disciplinary Sanctions of the CCP. While the economic crimes further violate the Criminal Law (Chapter 8) and the involved officials will receive penalties from a fixed-year imprisonment to a death sentence which is more severe than the other three types. With regard to sex scandals, the involved officials will receive more severe punishment since the CCP always sees morality as a key source of its legitimacy (Tong, 2011; Chu, 2013; Yan, 2021).

# Appendix C

Table C1: Definitions of Scandal Types

|  |  |  |
| --- | --- | --- |
| Scandal Types | Definition | # of Scandals |
| Economic crimes | (1) accepting or offering bribes; (2) embezzlement; (3) misappropriation; (4) failure to explain the sources of a large number of properties; (5) concealing deposits abroad; (6) dividing state-owned assets; and (7) dividing confiscated property | 207 |
| Dereliction of duty | (1) abuses of power; (2) neglecting one’s duty; (3) engaging in malpractice for personal gain; (4) violating the provisions of the Law on Guarding State Secrets | 112 |
| Sex scandal | (1) improper sexual relations with others; (2) exploiting professional power, mentoring relationships, staff relationships or other similar relationships to have sexual relations with others | 43 |
| Inappropriate speech and behaviour | (1) luxurious lifestyles, hedonism or the pursuit of vulgar interests; (2) inappropriate acts in a public place that violate the public order and good morals; (3) serious violation of social morality | 118 |

Table C2: Descriptive Statistics

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | Obs | Mean | SD | Min | Max |
| Internet Reposts (*ln*)  | 441 | 8.045 | 3.361 | 0 | 16.259 |
| Whether the government responds within one month  | 441 | 0.361 | 0.481 | 0 | 1 |
| Severity of punishment  | 436 | 4.88 | 3.14 | -1 | 10 |
| Rainfall (8:00-20:00) | 163,161 | 28.962 | 95.474 | 0 | 3,610 |
| Scandals | 441 | 2.089 | 1.226 | 1 | 4 |
|  Economic crimes | 207 |  |  |  |  |
| Dereliction of duty | 112 |  |  |  |  |
| Sex scandal | 43 |  |  |  |  |
| Inappropriate speech and behaviour | 118 |  |  |  |  |
| Officials’ administrative level | 439 | 3.392 | 2.009 | 0 | 8 |
| Officials’ position | 436 | 4.436 | 1.669 | 1 | 6 |

Table C3: Online Discussion and Severity of Punishment: Concerning Sampling Bias

|  |  |  |
| --- | --- | --- |
|   | (1) | (2) |
|  | Disciplining | Disciplining |
| VARIABLES | Tobit Model | Restricted Sample |
|   |  |  |
| Repost (ln) | 0.115\*\* | 0.043\*\* |
|  | (0.045) | (0.019) |
| Dereliction of duty | -4.091\*\*\* | -2.356\*\*\* |
|  | (0.352) | (0.206) |
| Sex scandals | -2.525\*\*\* | -1.567\*\*\* |
|  | (0.566) | (0.351) |
| Inappropriate speech | -5.353\*\*\* | -1.945\*\*\* |
|  | (0.447) | (0.218) |
| Officials’ position | Yes | Yes |
| Officials’ administrative level | Yes | Yes |
| Constant | 3.988\*\*\* |  |
|  | (1.297) |  |
| Observations | 430 | 350 |

Note: The type of economic crime is the omitted reference category. Standard errors clustered at the official level in parentheses. \*\*\* p < 0.01, \*\* p < 0.05, \* p < 0.1

Table C4: External Validity: IV-2SLS Estimation

|  |  |  |  |
| --- | --- | --- | --- |
|   | (1) | (2) | (3) |
|  | First-stage | Second-stage | Second-stage |
| VARIABLES | Baidu Index (ln) | Respond within one month | Punishment Severity |
|   |   |   |   |
| Rainfall (ln) | 0.162\*\* |  |  |
|  | (0.076) |  |  |
| Repost (ln) |  | 1.867\*\*\* | 1.059\* |
|  |  | (0.626) | (0.559) |
| Dereliction of duty | -1.092\* | 3.480\*\*\* | -1.286 |
|  | (0.615) | (1.174) | (1.024) |
| Sex scandals | -0.124 | 2.909\*\*\* | -1.373\* |
|  | (0.569) | (0.977) | (0.738) |
| Inappropriate speech | -0.963\*\* | 3.112\*\*\* | -3.182\*\*\* |
|  | (0.487) | (1.003) | (0.801) |
| Year fixed effects | Yes | Yes | Yes |
| Provincial fixed effects | Yes | Yes | Yes |
| Constant | 8.886\*\*\* | -21.185\*\*\* | -5.706 |
|  | (1.147) | (6.773) | (5.925) |
|  |  |  |  |
| Observations | 187 | 107 | 172 |

Note: Scandals related to economic crimes serve as the omitted reference category. Standard errors in parentheses. \*\*\* p < 0.01, \*\* p < 0.05, \* p < 0.1

1. The PMSinaMsgCrawler is one of the most widely used open-source trawlers to collect Weibo posts. It was developed by Pameng (<http://cnpameng.com/>) to blend crowdsourcing with big data analytics. We crawled the data at the end of 2012. [↑](#footnote-ref-1)
2. <https://www.huxiu.com/article/16983/1.html>. [↑](#footnote-ref-2)
3. The number of registered users is the only way to measure Weibo’s development in Sina’s financial reports during the study period. After 2013, Sina cited the number of active monthly users instead in its financial reports. [↑](#footnote-ref-3)
4. Each assistant read the related articles of Criminal Law and Regulations on Disciplinary Sanctions of the CCP before they started to code. This method follows the spirit of inter-annotation agreement in natural language processing (Pustejovsky and Stubbs 2012) and can avoid the bias proposed by King *et al.* (2017a). In addition, since there is no text corpus for political scandal tagging, we cannot directly apply a supervised learning method to identify the scandals. We tried an unsupervised method of identification, but the results were not as expected. Thus, the expert annotation we used serves as a valuable resource to develop the supervised machine learning method. [↑](#footnote-ref-4)
5. There are two scenarios for the duplicate names. First, officials’ names were the same as famous people’s names. If this is the case, those posts discussing famous people will be mistakenly considered as posts discussing officials. And posts discussing famous people will attract a lot of reposts, leading to the overestimation of our dependent variables. Here we use Sogou dictionaries of entertainment and sports stars to disambiguate official names. Fortunately, none of the officials’ names are the same as entertainment and sports stars. Second, officials’ names were the same as ordinary people’s names. Based on our estimation, the proportion of wrong posts due to the second scenario of duplicate names is just 1.3%, and they only have a minimal number of reposts that can be ignored. [↑](#footnote-ref-5)
6. <https://www.ccdi.gov.cn/hdjln/nwwd/202108/t20210812_142290.html> . [↑](#footnote-ref-6)