Intergovernmental Conflict and Censorship:
Evidence from China's Anti-Corruption Campaign

Maiting Zhuang[*]

This version: September 5, 2019
Link to latest version

Abstract

Media censorship is prevalent in autocratic regimes, but little is known about how and why censorship might vary within a country. I study how Chinese newspapers report on officials caught during Xi Jinping's anti-corruption campaign, by collecting close to 40,000 articles in print and corresponding social media posts and comments. I find twice as many internet searches and comments on social media posts about corrupt officials from the same province compared to officials from other provinces. Yet, despite greater reader interest, local newspapers underreport corruption scandals involving high-level officials from their own province. Underreporting is greater when a newspaper does not rely on advertising revenue and a corrupt official is well connected. When newspapers do report high-level corruption at home, they deemphasise these stories, by making them shorter, less negative and featuring official sounding headlines without references to corruption or the anti-corruption campaign. City-level newspapers report less about corruption in their own city relative to other cities in the same province, but are more likely to report corruption within their provincial government than corresponding provincial newspapers. These results suggest that intergovernmental conflict within an autocracy can lead to diverging media censorship strategies by different levels of government.

Keywords: media censorship, newspapers, corruption, intergovernmental conflict.

JEL Codes: H77, P26, L82.

A large share of the world population lives in countries where the press is not free.[1] Media censorship is key for maintaining public support and the stability of many non-democratic regimes. But is censorship always uniform within an autocracy, or could its extent and direction vary? This paper uses a large dataset of local Chinese newspaper articles to show that a conflict of interest between different parts of an autocratic government can lead connected media outlets to publish different content.

An important part of Chinese president Xi Jinping's domestic policy is an anti-corruption campaign that has already punished over one million officials.[2] This campaign is run by a central party agency and is covered extensively in the *central* state media. Some media outlets are even "invited to investigate" officials who were named during the campaign (Repnikova, 2017). For the central government, the large number of (mostly local) officials caught signals its determination to eliminate corruption. In contrast, local governments may worry that these investigations could implicate other local officials still in power or otherwise damage their image and credibility. Thus, they may have different preferences over the media coverage of the campaign to the central government. In particular, local governments may have an incentive to selectively censor news about official corruption from their own administration. If they also have the ability to control local media coverage beyond central censorship guidelines, we should observe that the local media suppresses or downplays news about corrupt officials from the same locality.

In this paper, I study how different Chinese newspapers report about officials who are suspected of corruption. From the start of the anti-corruption campaign until the end of 2014, the central agency responsible posted announcements of investigations into 408 officials from 14 Chinese provinces on their website. I collect close to 40,000 articles about each of these officials, before and after their investigation has been announced, using an online newspaper archive (Wisenews), which contains 99 Chinese newspapers. As the corruption investigation announcements are independent of newspapers' reporting decisions, I can study *whether* and *how* newspapers publish articles about official corruption.

Using internet search and social media data, I show that people are more interested in corruption scandals that happen in their own province compared to those elsewhere. Searches on Baidu (China's most popular search engine) for officials under investigation are on average more than twice as high in the official's home province than the next highest province. I collect data on all posts about high-level corruption scandals that newspapers published on their official Sina Weibo accounts (Chinese equivalent to Twitter), along with close to 80,000 comments on these posts and information about the commenters. On any given social media post, there are almost twice as many comments coming from the home province of the corrupt official mentioned, relative to comments from other provinces. Comments about home-province officials are also more negative and tend to be posted by less active social media users.

I compare how the likelihood of an article being published about a suspected official changes after their corruption scandal breaks, depending on whether the official and the newspaper are from the same province. My estimation strategy uses only variation over time within an official-newspaper pair and controls for a full set of seasonal factors. I

---

[1] According to Freedom House, only 13 percent of the world population enjoys a free press. "Freedom of the Press 2017", `https://freedomhouse.org/report/freedom-press/freedom-press-2017`, last accessed: 20 April 2019.

[2] "Charting China's 'great purge' under Xi". *BBC*, 23 October 2017. `http://www.bbc.com/news/world-asia-china-41670162`.

find that local newspapers report less about central government investigations of high-level local officials from the same province, relative to officials from other provinces. After an investigation is announced, the daily probability of an article about a corrupt high-ranking official being published by a newspaper from the same province declines by around 6 percentage points (90 percent of the pre-announcement mean probability). These estimates in this paper may be a lower bound on the true extent of media bias due to local government control, as there is higher reader demand for this type of reporting. The extent of underreporting is more than twice as large for corrupt officials who served on the provincial party standing committee with the incumbent party secretary, suggesting that this censorship strategy is designed to avoid implicating other officials still in power. This effect is largest for official party newspapers, which are directly government owned and rely on government subsidies and subscriptions, and smaller for commercial newspapers that have to compete for readership and advertising revenue.

I also find that provincial governments have imperfect control over lower levels of government. Compared to provincial-level newspapers, newspapers from lower levels of government report *more* about corruption scandals involving officials from their own provincial government. Both provincial and sub-provincial newspapers compete for similar local audiences and they report similarly about corruption in other provinces. This finding suggests that the extent of underreporting is closely tied to the direct relationship between a newspaper and a local government and appears inconsistent with a localised censorship strategy by the central government or strategic overreporting by neighbouring provinces. These and other alternative hypotheses are discussed in greater detail in Section 4.4

Corruption scandals involving *low-level* officials are generally reported more frequently in newspapers from the same province, with some some exceptions. Municipal newspapers report less about corrupt low-level officials from the same city compared to officials from other cities in the same province. Using a case study, where a high-ranking official and his network of low-level associates were sequentially investigated, I show that corruption scandals involving well-connected low-level officials are underreported while their patron was in power.

Even when newspapers report on corruption in their home province, they are careful not to emphasise these stories. Compared to articles about corruption elsewhere, articles about high-level corrupt officials in a newspaper's home province are on average 30 percent shorter. In addition, their headlines are designed to appear uninteresting. Relative to an average article about the anti-corruption campaign, articles about local high-level corruption are twice as likely to include the exact text of the official announcement as their headline. Their headlines are also around one third less likely to include any mentions of corruption and references to Xi Jinping and the central anti-corruption campaign. Articles about local corruption scandals also tend to be more positive, with their headlines and full texts including more positive and fewer negative words.

This paper contributes to the literature on media bias. Media bias can be driven by supply-side factors, such as advertising revenue (e.g., Di Tella and Franceschelli, 2011; Beattie et al., 2017) and preferences of media owners (e.g., Durante and Knight, 2012; Enikolopov, Petrova and Zhuravskaya, 2011), as well as demand-side factors, such as reader's ideologies (e.g., Gentzkow and Shapiro, 2010; Puglisi and Snyder Jr, 2011). In this paper, I study the competition between different supply side factors in determining media bias, i.e., whether the content of local newspapers reflects the preferences of the central or local governments in an autocracy.

The focus on the use of local media by subnational governments in an autocracy distinguishes this paper from many other studies of media in non-democracies (see, e.g., Adena et al., 2015; Yanagizawa-Drott, 2014).[3] Until very recently, few large-scale quantitative studies of media censorship in China existed. Censorship on the Chinese internet, China's so-called "Great Firewall", has been the subject of a number of studies, such as King, Pan and Roberts (2013, 2014). Online censorship can be measured directly through the removal of social media posts. In contrast, newspaper articles that are prevented from being published, or even written in the first place, cannot be observed. Qin, Strömberg and Wu (2018) study the trade-off between Chinese newspapers' financial and political goals by constructing a measure of bias based on the published contents of 117 Chinese newspapers. Their measure is positively correlated with being an official party newspaper and negatively correlated with advertising revenue. In this paper, I measure media bias by using the announcements of official corruption investigations, which are exogenous to the reporting decisions of newspapers. This enables me to construct counterfactuals for articles that were not published, as well as compare reporting across published articles.

This paper also contributes to the wider literature on political decentralisation and the functioning of bureaucracies, see Mookherjee (2015) for a recent review of the literature. While China's relatively well-managed bureaucracy has been credited as a contributing factor to China's fast growth (see Li and Zhou, 2005), persistent agency problems are revealed in a number of areas, where local officials adopt policies counter to the central government's intentions (e.g., Fisman and Wang, 2017; Jia and Nie, 2017). In this paper, dissent within the political hierarchy manifests itself through different media censorship strategies. The fact that the media can be captured by local bureaucrats has wider implications. Papers, such as Egorov, Guriev and Sonin (2009) and Lorentzen (2014), have shown theoretically that an autocratic ruler can benefit from using the media to monitor local officials and citizens' grievances. Other than broadcast official propaganda, it is also the role of the Chinese media to act as a watchdog and report local malfeasance upwards (Shirk, 2011). The findings of this paper cast serious doubt on the ability of newspapers in China to fulfill this role, as local newspapers tend to carry the least news about local corruption, despite having an informational advantage and an interested readership. This echoes the results of Pan and Chen (2018) who show that lower level governments in China conceal online complaints from their superiors.

The rest of the paper is structured as follows. The next section provides institutional background on newspaper censorship and the current anti-corruption campaign in China. Section 2 describes the data. The empirical strategies and main results are discussed in Sections 3 and 4, respectively. Robustness checks are reported in Section 5. Section 6 concludes.

1 Background

1.1 Newspapers

China has one of the world's largest newspaper markets, both in terms of circulation and advertising revenue. The newspaper market grew quickly until 2013 and newspapers remain a relevant source of news due to their strong online presence (Sparks et al., 2016). All newspapers are regulated and licensed by the State Administration of Press, Publication, Radio, Film and Television (SAPPRFT). According to SAPPRFT, 1894

---

[3]See also Prat and Strömberg (2013) for a recent general survey on the political economy of media.

newspapers were published in 2016, printing a total of 39 billion copies.[4]

General-interest newspapers in China can be divided into two main types: official party newspapers and commercial newspapers. Official party newspapers, called Dailies, are the party's "throat and tongue". They are heavily subsidised and rely on subscriptions by government agencies and enterprises (see, e.g., Shirk, 2011). Qin, Strömberg and Wu (2018) find that Dailies carry the most biased media content. Entertainment-oriented commercial newspapers, called Evenings or Metros, rely on advertising revenues and compete for customers. According to Stockmann and Gallagher (2011), commercial newspapers are seen as more credible by the public.

Dailies are directly owned and operated by the government, but some commercial newspapers are owned by other (government-owned) newspapers.[5] In contrast to directly government-owned newspapers, top personnel decisions at subsidiary newspapers are made by their parent newspapers.

The hierarchy of Chinese newspapers mirrors that of the government bureaucracy. National newspapers, such as the People's Daily – the official newspaper of the central CCP leadership, are owned by the central government and are available in the entire country. Around 90 per cent of newspapers in China are local newspapers. These are owned by local governments of different levels, that is, provincial governments own provincial-level papers and prefecture governments own prefectural-level papers etc. The majority of local newspapers are only circulated in their local areas.[6]

1.2 Censorship of newspapers

Like all media in China, newspapers are subject to strict government control, both before and after publication. Ex ante, central and local propaganda bureaus issue reporting guidelines and hold meetings with chief editors. Published content in newspapers is monitored and failure to comply with directives can result in demotions, dismissals or even jail sentences for journalists and editors (Shirk, 2011).

Online leaks reveal further details about the operation of Chinese propaganda departments. The following are some examples of leaked directives from the Guangdong province propaganda department to local media:

> *Do not independently investigate, report, or comment on the series of corruption cases in Maoming, with the exception of those which are arranged unified manner.* (December 5, 2012)[7]

> *No media are to sensationalize the topic of government officials making public their personal assets or related issues. Do not place reports on the front page, and do not lure readers to coverage.* (January 24, 2013)[8]

---

[4]"National news and publishing situation 2016". *SAPPRFT*, 27 September 2017. `http://www.sapprft.gov.cn/sapprft/govpublic/6677/1633.shtml`.

[5]No news outlet in China is truly independent of the government, as this government decree illustrates *"[. . .] no matter who its investors are, a news provider is a publicly owned resource"* that has *"[. . .] just one shareholder: the Chinese Communist government"* (He, 2004).

[6]Only the most successful provincial papers, such as the *Southern Weekend*, are available outside of their province.

[7]Henochowicz, Anne. 2012. "Ministry of Truth: Dispatch from Guangdong". *China Digital Times*, 20 December. `https://chinadigitaltimes.net/2012/12/ministry-of-truth-dispatch-from-guangdong`.

[8]Henochowicz, Anne. 2013. "Ministry of Truth: Guangdong People's Congress". *China Digital Times*, 27 January. `https://chinadigitaltimes.net/2013/01/ministry-of-truth-guangdong-`

These guidelines illustrate how local media are forbidden from investigating and reporting certain stories. During especially politically sensitive times, such as major party meetings, the media are advised not to report any negative news. The tone and framing of articles on is also important and editors often copy official government announcements or articles by approved central news agencies to avoid sanctions.

Official corruption can be a sensitive topic and such stories are frequently censored. He (2004) cites anecdotal evidence that local propaganda bureaus are stricter than their central counterpart when it comes to reporting local corruption, consistent with the results of this paper.[9]

1.3 Anti-corruption campaign

When Chinese President Xi Jinping came to power at the end of 2012, he vowed to end corruption in the Chinese Communist Party (CCP) for fear it would otherwise "doom the party and the state". He introduced new laws and regulations to curb excessive spending by officials. The CCP's Central Commission for Discipline Inspection (CCDI) was endowed with new powers to investigate the over 80 million CCP members for corruption and other misdemeanors. Over 1.34 million officials have been punished in Xi's first five years in office.[10] Contrary to press speculations, Lorentzen and Lu (2018) and Francois, Trebbi and Xiao (2016) find little evidence that the campaign is driven by factional politics.

The majority of corruption investigations are conducted internally by the CCDI or its regional inspection teams. However, the CCDI publishes information about some cases directly on its website. These generally involve higher ranking officials or those involved in wider corruption cases. For example, on 17 October 2013, the CCDI posted:

> *The Deputy Secretary of the Municipal Party Committee and Mayor of Nanjing, Ji Jianye, is suspected of seriously violating discipline and engaging in illegal behaviour and is currently under investigation.*[11]

Even though corruption is not always mentioned in the announcements, the vast majority of subsequent prosecutions are for corruption. The investigation announcements tend to be unexpected and almost always result in criminal proceedings or, at a minimum, the end of the official's career.

The anti-corruption campaign is covered extensively in the central state media. For instance, the *People's Daily* website has a dedicated section on the campaign, where articles about new regulations, individual cases and progress reports are published. The large number of reported investigations is seen as a sign of the central leadership's determination and success in eliminating corruption. Online and print versions of newspapers are an important source of news for detailed accounts about the anti-corruption campaign. Television stations dedicate their more limited broadcasting time to special programmes about the campaign. For instance, the CCDI has partnered with China's national broadcaster CCTV to produce and air several documentaries about the campaign, featuring on-air confessions of corrupt officials. A fictionalised anti-corruption TV drama (*In the*

---

peoples-congress.

[9]According to Shirk (2011), the central government uses local media reports to monitor local officials.

[10]"Charting China's 'great purge' under Xi". *BBC*, 23 October 2017. http://www.bbc.com/news/world-asia-china-41670162.

[11]"南京市委副书记、市长季建业涉嫌严重违纪违法正接受调查.." *CCDI*, 17 October 2013. http://www.ccdi.gov.cn/scdc/zggb/zjsc/201607/t20160704_115803.html.

Name of the People), in part financed by the national agency responsible for prosecuting corruption cases, attracted a record number of viewers.[12] This is, reportedly, the first time since 2004 that national censors allowed the broadcast of a TV show about high-level corruption and the government has since commissioned more films and TV shows about the campaign.

2 Data and Descriptive Statistics

2.1 Newspapers

The sample of newspapers consists of all Chinese-language general interest newspapers available on WiseNews, an online archive based in Hong Kong, from the end of 2010 to the end of 2015, yielding a total of 99 newspapers. WiseNews selects influential publications from large cities and is representative of newspapers in large metropolitan areas in China (Qin, Strömberg and Wu, 2018).[13] I collect information on newspaper ownership from the SAPPRFT website and China Journalism Yearbooks.

Panel A of Appendix Table A1 presents summary statistics of the newspapers in the sample. 52 percent of the sample are subsidiary newspapers and 48 percent are directly government-owned newspapers, of which 21 percent are official party papers. The majority of newspapers are provincial newspapers (55 percent), central papers represent 18 percent of the sample, while sub-provincial papers represent 27 percent. Appendix Figure A1 shows the 14 out of 31 province-level administrative regions in mainland China, including all four province-level municipalities, that have at least one newspaper in the sample. This represents slightly more than half of China, both in terms of population (53 percent) and GDP (57 percent).[14]

2.2 Officials under investigation

The sample of officials under investigation contains 408 individuals. These are all the officials named on the CCDI website from the start of the anti-corruption campaign in 2012 until the end of 2014, who come from the 14 provinces (and central government) that are covered by the WiseNews sample.[15] Appendix Figure A3 shows the distribution of officials and newspapers across provinces. All CCDI posts include the time of announcement and the official's name, occupation and location. Knowing an officials' occupation is generally sufficient to determine their rank in the government hierarchy (see Appendix Figure A2 for the distribution of ranks).

Panel B of Appendix Table A1 shows summary statistics of the officials in the sample. 9 percent of officials (34 officials) in the sample are high-ranking.[16] In terms of their sector

---

[12]Ming, Cheang. 2017. "China's anti-graft campaign may be headed for a screen near you". *CNBC*, 30 March. https://www.cnbc.com/2017/03/30/in-the-name-of-the-people-chinas-anti-graft-campaign-makes-it-to-the-small-screen.html.

[13]Of these, 76 are available for the entire time period. WiseNews provides content for 105 newspapers, of which three are in English and another three only report about health and sports. I also consolidate local editions of the same newspaper. Newspapers with multiple editions on Wisenews are Nanfang Ribao, Nanfang Dushibao, Guangzhou Ribao and Nanguo Chengbao.

[14]Source: National Bureau of Statistics of China, based on figures from 2015.

[15]Across all provinces, the CCDI named a total of 804 officials over this time period. This is a small subset of all officials investigated during this time period.

[16]In China, an official of vice-provincial rank or higher is considered a high-ranking official (see, e.g., Li and Zhou, 2005).

of work, 78 percent of the sample (320 officials) are government officials, mostly from local governments (72 percent of the sample, 295 officials) and 22 percent are officials from state-owned institutions (hospitals, universities etc.) and enterprises.

2.3 Articles

I collect all articles about these 408 corrupt officials, by searching all 99 newspapers in the WiseNews database using the official's name and one other keyword, typically their occupation, organisation name or location, between two years prior and one year following the CCDI announcement. This results in a total of 39,271 unique articles.[17] By construction, each article can be matched to a newspaper and one or more officials.[18] I have the following article-level information: word count, whether it includes an image, page number and headline.[19]

Summary statistics are reported in Panel C of Appendix Table A1. The mean word count of articles is around 1300 words, though most of the articles are shorter with 55 percent of articles in the sample having less than 1000 characters. The average article is placed on page 8 of the newspaper and more than half of the articles are on pages one to five. Around 9 percent of articles include an image. Among articles published following a corruption scandal, 22 percent of article headlines include direct references to corruption and eight percent of headlines include keywords associated with the central anti-corruption campaign.[20] 1 percent of articles about the corruption scandals copy the official CCDI announcement. 27 percent of articles mention more than one official under investigation, as newspapers frequently summarise the progress of the anti-corruption campaign by listing several corrupt officials in one report.[21]

3 Empirical strategy

3.1 Are corruption scandals underreported?

To test the underreporting of corruption scandals, I take into account all articles that can *potentially* be published by constructing a daily panel dataset for the 40,392 official-newspaper pairs over the entire sample period, resulting in around 40 million official-newspaper-day observations (summary statistics are reported in Panel D of Appendix Table A1). Appendix Figure A4 shows the fraction of official-newspaper pairs with at least one published article by day relative to the investigation announcement at 0. The mean daily probability of publication is 0.1 percent.

---

[17] As newspapers from the same province will often omit the province's name when referring to an official, I exclude province names from the keywords. There are a total of 50,720 articles about the 804 officials from all provinces.

[18] There are a total of 53,412 article-official pairs for the 408 officials under investigation and 72,519 pairs for all 804 in total. I identify articles to be the same if they are published by the same newspaper on the same day with the same headline, word count, image and layout information.

[19] Page numbers are constructed by matching WiseNews information on the article layout to electronic and hard copies of the newspapers. The reported page numbers are absolute, not relative, as the total number of pages varies from issue to issue.

[20] This includes direct references to the anti-corruption campaign, the 18th National Congress, Xi Jinping and mentions of "tigers" and "flies", which refer to high- and low-level officials who were caught during the campaign.

[21] Appendix Table A2 shows the characteristics of officials with no articles and only articles before or after the investigation announcement.

The main hypothesis I test in this paper is that local newspapers *selectively* under-report the central anti-corruption campaign in response to local government pressure. I compare how local newspapers report about corrupt officials from their own province compared to officials from other provinces, before and after an investigation announcement. I also extend this analysis to the municipality level, by comparing how municipal newspapers report about corrupt officials from their own city, relative to officials from other cities of the same province and from cities in other provinces. I use a linear probability model to estimate the effect of a newspaper and corrupt official being from the same province on the probability of reporting the corruption scandal of the form

$$y_{o,n,t} = \beta_1 P_{o,t} + \beta_2 (P_{o,t} \times S_{o,n}) + \mathbf{X'_t}\delta + \theta_{n,t} + \gamma_{o,n} + \epsilon_{o,n,t} \tag{1}$$

with the indices denoting official (o), newspaper (n) and day (t), respectively. $y_{o,n,t}$ is a dummy variable for whether an article about official o was published in newspaper n on day t.[22] $P_{o,t}$ is a dummy variable that takes the value of 1 for all days t after an investigation into official o has been announced. $S_{o,n}$ is a dummy variable that takes the value of one if official o and newspaper n are from the same province. $\mathbf{X_t}$ is a set of controls for potential seasonality in article publication, and includes month and year fixed effects, day-of-the-week fixed effects and fixed effects for important Chinese holidays (such as Chinese New Year, Dragon Boat Festival, Mid-Autumn Festival) and major CCP events (such as the 18th Party Congress, different Plenary Sessions of the CCP Central Committee and sessions of the National People's Congress and the National Committee of the Chinese People's Political Consultative Conference) over the sample period. I also control for the provincial-level counterparts to the national CCP events ($\theta_{n,t}$). All regressions include official-newspaper-pair fixed effects ($\gamma_{o,n}$), which account for time-invariant characteristics of officials, newspapers and official-newspaper pairs.[23] These fixed effects account for the direct effect of $S_{o,n}$, but the coefficient of interest β_2 on the interaction term $P_{o,t} \times S_{o,n}$ varies over time within an official-newspaper pair. To account for possible correlation of the error term across both officials and newspapers, I use two-way clustered standard errors at the official and newspaper levels (see Cameron, Gelbach and Miller, 2011).

Official corruption is newsworthy, so we should expect β_1 to be positive, i.e., the likelihood of an article being published about an official should increase after they are placed under investigation. The expected sign of β_2 - the change in the daily probability of publication post-announcement when an official and newspaper are from the same province - is ex ante ambiguous. On the demand side, newspaper readers should be *more* interested in officials from the same province, as their activities could directly impact readers' welfare. There are competing supply-side factors that could bias reporting. As the central government broadcasts the results of its anti-corruption campaign throughout the country, there are few reasons for it to selectively censor local newspaper coverage. Provincial governments, however, could pressure local newspapers not to publish corruption stories about officials from the same province, leading to a negative effect on β_2. There are a number of reasons for this type of censorship. For instance, publicising the investigations could encourage members of the public to come forward and implicate officials still in

---

[22]In Section 5, I show that the results are robust to using a Logit or Poisson model with the article count as dependent variable.

[23]Time-invariant characteristics of official-newspaper pairs could include, for example, a newspaper readership's general level of interest in an official's activities, or the relative administrative rank of the official and the newspaper's editors.

8

power, supervisors and colleagues could be accused of condoning corruption and revealing large-scale corruption could damage the reputation of the local government.

3.1.1 Heterogeneity by official rank

The extent of local government censorship should depend on the expected benefits and costs. I expect the benefits of censorship to be larger for corruption scandals involving previously more powerful officials. Equation 2 allows for heterogeneous effects based on official rank:

$$
\begin{aligned}
y_{o,n,t} =& \alpha_1(P_{o,t} \times H_o) + \alpha_2(P_{o,t} \times (1 - H_o)) + \alpha_3(P_{o,t} \times S_{o,n} \times H_o) + \\
& \alpha_4(P_{o,t} \times S_{o,n} \times (1 - H_o)) + \mathbf{X_t'}\delta + \theta_{n,t} + \gamma_{o,n} + \epsilon_{o,n,t}
\end{aligned}
\tag{2}
$$

where H_o indicates a high-ranking official and the rest of the notation follows equation 1. Under local government censorship, there should be a larger bias in local media reporting about high-level corruption, i.e. in equation 2, $\alpha_3 < \alpha_4 < 0$ and $\alpha_1 > \alpha_2 > 0$ as high-level corruption is likely to attract more reader interest.

3.1.2 Heterogeneity by newspaper type

The cost of censorship could also differ by newspaper type. As outlined in Section 1, there are three different types of local newspapers: government-owned party papers, government-owned commercial papers and subsidiary commercial papers. It might be easier for governments to ensure compliance of directly government-owned newspaper, for example, because propaganda department officials are directly involved in editorial matters and personnel decisions. In this case, both party papers and government-owned commercial papers should censor more than subsidiaries. Alternatively, the relevant cost of censorship could be financial. Commercial newspapers regardless of ownership rely on advertising revenue and underreporting on corruption scandals could lower readership and advertising. In this case, we should primarily see censorship in party papers which are financed through subsidies. Equation 3 allows for heterogeneity based on newspaper type

$$
\begin{aligned}
y_{o,n,t} =& \mu_1(P_{o,t} \times O_n) + \mu_2(P_{o,t} \times G_n) + \mu_3(P_{o,t} \times I_n) + \mu_4(P_{o,t} \times S_{o,n} \times O_n) + \\
& \mu_5(P_{o,t} \times S_{o,n} \times G_n) + \mu_6(P_{o,t} \times S_{o,n} \times I_n) + \mathbf{X_t'}\delta + \theta_{n,t} + \gamma_{o,n} + \epsilon_{o,n,t}
\end{aligned}
\tag{3}
$$

where O_n, G_n and I_n indicate whether newspaper n is an official party paper, a government-owned commercial paper or subsidiary commercial paper, respectively. The rest of the notation is as above.

If directly government-owned newspapers can be censored more efficiently, then we should expect $\mu_4 = \mu_5 < \mu_6 < 0$, i.e. there is more underreporting in party papers and government-owned commercial papers. If, on the other hand, newspapers trade off political and financial costs, then we should expect (both government-owned and subsidiary) commercial newspapers to underreport less, i.e. $\mu_4 < \mu_5 = \mu_6 < 0$. I also estimate the interaction between newspaper types and official ranks. We should expect to see a more pronounced pattern for high-ranking officials.

3.1.3 Municipality-level censorship

I can also test whether similar patterns of underreporting occur at the municipality level. I replace the main regressor $S_{o,n}$ (whether official and newspaper are from the same province) in the above equations with two mutually exclusive dummy variables that take

the value of one if the newspaper and official are from the same city ($SC_{o,n}$) and if they are from different cities of the same province ($DC_{o,n}$). If the selective underreporting of local corruption scandals also extends to the city level, then the coefficients on $SC_{o,n}$ should be smaller than those on $DC_{o,n}$. All other controls and standard errors are as before.

3.2 How are corruption scandals reported?

Given the central media's coverage of the anti-corruption campaign, it will be difficult for local governments to suppress news of investigations altogether. However, they could still attempt to deemphasise unfavourable news (see Section 1). To determine whether a newspaper writes different articles about corrupt officials from their own province compared to those from other provinces, I estimate the following equation

$$y_{o,n,t} = \kappa_1 \tilde{H}_{o,n} + \kappa_2 \tilde{L}_{o,n} + \mathbf{X_t'}\delta + \theta_{n,t} + \gamma_o + \gamma_n + \epsilon_{o,n,t} \tag{4}$$

on the sample of articles published after the investigation announcement. I use the following quantifiable characteristics of the articles as dependent variables: (log) word count, page number and dummy variables for whether an article includes an image, mentions multiple officials, has a headline that is a copy of the official announcement message, includes references to corruption and the centrally-run anti-corruption campaign, as well as measures of the sentiment expressed in an article. As there can be multiple officials mentioned in one article, the regressors $\tilde{H}_{o,n}$ and $\tilde{L}_{o,n}$ denote the share of high- and low-level officials from the same province in the total number of officials in each article. The equation includes official and newspaper fixed effects, standard errors are clustered at the newspaper level and the rest of the notation is as before.[24]

4 Results

4.1 Underreporting of corruption scandals

Table 1 reports the estimation results for equation 1 in the first two columns. The estimated coefficient on the post-announcement dummy (β_1) is statistically significant and positive. The probability of newspapers publishing an article about an official from another province on any given day increases by 0.3 percentage points, after this official's corruption scandal breaks (column 2). This effect is three times as large as the mean probability of publication (0.1 percent). The estimated β_2 is negative, but not statistically significantly different from zero, suggesting that newspapers on average do not report less about corruption in their own province.

4.1.1 Low vs. high-level officials

There should be more incentives to censor corruption scandals involving high-level officials. Figure 1 shows the fraction of official-newspaper pairs with at least one article. Corruption scandals involving low-level officials are reported *more* frequently in newspapers from the same province compared to those from other provinces (Panel A), but high-level corruption scandals are reported *less* in their own province *in absolute terms.*

---

[24]The results are robust to using article-official pairs and clustering standard errors at the article, official and newspaper level.

Columns 3 and 4 of Table 1 report estimation results for equation 2. Local newspapers underreport corruption scandals involving high-level officials from their own province, α_4 is negative and statistically significant. The daily probability of an article about a high-ranking official being published in a newspaper from the same province falls by 5.7 percentage points after the CCDI announcement, which is 87 percent of the pre-announcement mean of 6.6 percent. A newspaper from the same province as a high-level official writes on average 17 fewer articles in the year following that official's corruption scandal, compared to writing 24 articles per year before the scandal.[25] The positively estimated α_3 shows that the probability of reporting in a same-province newspaper increases by 0.2 percentage points following low-level corruption scandals (83 percent of the pre-announcement average). In contrast, the likelihood of an article about corruption scandals being published in newspapers from other provinces is four times higher for high-ranking compared to low-ranking officials.

4.1.2 Reader demand for corruption news stories

The selective underreporting of scandals involving high-level officials from the same province is consistent with a supply-side explanation, i.e., newspapers are pressured by their provincial governments to hide negative news. The fact that newspapers carry more news about low-level local officials after they are investigated for corruption suggests that newspaper readers are more interested in learning about corruption in their own province. Such a positive reader demand effect implies that the here measured media bias is likely an *underestimate* of the true extent of censorship. I obtained data from Baidu (China's most popular search engine) on the number of searches for officials under investigation by province.[26] Figure 2 shows the relative number of Baidu searches from each official's own province compared to the next highest outside province in the year following the investigation announcement. On average, there were more than twice as many searches about corrupt officials in their own province.[27]

Newspapers' reach extends beyond their print subscribers and many newspapers are very active in creating news content on social media. I collected data on all posts about corrupt high-level officials on their official Sina Weibo accounts (the most popular Chinese micro-blogging website, similar to Twitter), along with the comments about these posts and information about the commenters. I estimate the following equation on a dataset of Weibo comments by province

$$y_{o,n,p} = \lambda_1 \tilde{H}_{o,p} + \lambda_2 N_{n,p} + \mathbf{X}'_{\mathbf{t}}\delta + \theta_{n,t} + \gamma_o + \gamma_n + \gamma_p + \epsilon_{w,p} \tag{5}$$

where p denotes the self-reported province of people who comment on a Sina Weibo post about official o from newspaper n. $\tilde{H}_{o,p}$ is the share of high-level officials mentioned in the post who are from province p and $N_{n,p}$ is a dummy variable that takes the value of one if the newspaper is from province p. γ_p are province fixed effects and the rest of the notation is as above. Standard errors are clustered at the newspaper level.

Column 1 of Table 2 shows that people write on average twice as many comments on posts by newspapers from the same province or posts about officials from the same province. Column 2 shows that they are also faster at commenting on posts by their

---

[25]The pre-announcement mean is calculated for a high-level official in a newspaper from the same province, before the investigation announcement.

[26]Data were only available for 214 officials, including almost all high-level officials.

[27]This difference is larger for high-level compared to low-level officials, but not statistically significant.

local newspapers. These findings are consistent with local newspapers attracting a local readership (online as well as offline) and readers being more interested in local corruption scandals. Columns 3 and 4 analyse the average sentiment expressed in comments. Comments on posts about officials from the same province contain one third fewer positive words and a similar number of negative words. Comments on posts by newspapers from the same province contain both fewer positive and negative words. Columns 5 to 8 show how the average characteristic of commenters changes depending on whether a post is about an official from the same province. Commenters on these posts are less likely to be verified (column 5) and have fewer posts, followers and friends (column 7-8). These results suggest that posts about corruption at home attracts users who are normally less active on social media.

The fact that some people search the internet for corrupt officials from their own province does not mean that the local government censorship is ineffective. By writing fewer articles, less attention is drawn to these unfavourable stories (see Section 4.3). And although local governments are unable to prevent citizens from accessing alternative sources of information, experimental evidence from China suggests that citizens' demand for politically sensitive information is low (Chen and Yang, forthcoming).

4.1.3 Party newspapers vs. commercial newspapers

Figure 3 shows how each newspaper type reports corruption scandals involving high-level officials from the same province, with the full set of coefficients of equation 3 shown in Appendix Table A3. All types of local newspapers underreport, but they do so to a varying degree. Official party newspapers censor the most, followed by government-owned commercial newspapers and subsidiary newspapers. The difference in estimates between party newspapers and other types of newspapers is statistically significant. While the point estimate for subsidiary newspapers is less negative than for government-owned newspapers, the difference is not significant. These findings suggest that competition for advertising revenue could be a mitigating factor that leads commercial newspapers to censor less than party newspapers. However, it is important to note that none of the newspaper types in the sample are entirely free of local government influence. All types of newspapers increase their reporting when a *low-level* local official is investigated for corruption, with no statistically significant differences in magnitudes. There is also no difference between newspaper types when reporting about either high- or low-ranking officials from other provinces.

4.1.4 Within-province censorship

Media bias differs not only across province borders, but also within provinces. In Figure 4 and Appendix Table A4, I compare reporting in provincial and sub-provincial newspapers about corrupt officials from only the provincial governments. Panel A shows that, compared to a sub-provincial paper from the same province, a provincial newspaper underreports twice as much about corruption scandals involving high-level officials from the provincial government. This difference is statistically significant at the 1 percent level. This finding suggests that even within a province, incentives between different levels of government are not always aligned. In this case, the provincial government has

imperfect control over the reporting of newspapers from lower-level governments.[28] In contrast, Panel B shows that provincial newspapers are significantly more likely to report corruption scandals involving *low-ranking* provincial officials than sub-provincial newspapers. In neither case are there significant differences in the reporting of out-of-province newspapers by their administrative rank.

Within a province, corruption scandals involving city-level officials are only underreported by newspapers from the same city, while newspapers from other cities of the same province increase their reporting post-investigation, see Figure 5 and Appendix Table A5.[29] The difference in the coefficients is significant at the 5 percent level. The results suggest that local government censorship extends down to the municipal level.

4.2 Mechanism: Reasons for censorship

What are the reasons for this kind of censorship? Officials named by the CCDI are immediately removed from power following the investigation announcement and will almost invariably be found guilty. However, their former colleagues and superiors still have an incentive to censor negative coverage to limit the damage to their own reputation.[30] Two pieces of indirect evidence support this explanation.

4.2.1 Proximity to the incumbent party secretary

Proximity to powerful officials determines the extent of censorship. Using the People's Daily website and Baidu Baike, I collect CV information for all high-ranking officials in the sample and all provincial party secretaries over the sample period.[31] I determine whether each investigated official has served on the province's party standing committee at the same time as the incumbent party secretary. This is a good measure of proximity to the head of the province, as without their approval, an official is unlikely to become a standing committee member. In addition, as committees meet regularly and normally have fewer than 15 members, it is hard for a provincial party secretary to distance himself from a fellow committee member. Figure 6 shows that there is more than twice as much underreporting of corruption scandals involving officials who are close to the party secretary compared to officials who are not by their own province's newspapers. This difference is significant at the 1 percent level, while there is no significant difference in reporting by the out-of-province media (see Appendix Table A6).[32]

---

[28] Qin, Strömberg and Wu (2018) also find that newspapers from lower levels of governments are less biased, thus eroding the political goals of higher level governments.

[29] Almost all municipal officials are low ranking.

[30] Theoretically, the incentives of remaining officials could also work in the opposite direction, if they intend to distance themselves from the investigated officials. The empirical findings regarding high-level corruption scandals suggest that this mechanism is less important. As high-ranking officials occupy, by definition, one of the top two ranks of officials within a province, they are seen to represent the provincial government.

[31] Baidu Baike is a popular web encyclopedia, similar to Wikipedia. Provincial party secretaries are the highest ranked officials in their province.

[32] A similar pattern holds when comparing officials who are in active government with those in nominally high-ranking advisory positions, that are commonly regarded as semi-retirement postings with little real power. Results not shown.

4.2.2 Case study: Wan Qingliang

The case of Wan Qingliang, former party secretary of Guangzhou city and Guangdong province standing committee member, also illustrates the importance of being connected to powerful individuals. Wan was one of the most powerful officials in the province and was widely considered a rising star in the party, before he was unexpectedly investigated for corruption. Wan was reportedly implicated by previous CCDI investigations of his former aides. In the sample of officials there are five lower level officials from Guangzhou who were connected to Wan (hereafter Wan-group officials), according to ChinaFile's summary of the anti-corruption campaign.[33]

I construct a sample of Wan-group officials and other corrupt officials with the same administrative ranks from Guangdong province and newspapers from Guangdong province. Rather than comparing reporting before and after each low-level official's investigation was announced, I split the post-announcement period into two: before and after Wan Qingling's investigation. Consistent with previous results, newspapers from Guangdong increase their reporting about the low-level officials similarly following their corruption investigation (Figure 7 and Appendix Table A7). While Wan was in power, newspapers from Guangzhou (where he was party secretary) report significantly less about officials connected to Wan relative to other newspapers from Guangdong province (Panel A). Guangzhou newspapers also reported less about Wan-group officials compared to the other officials, though this difference is not statistically significant (p-value of 0.12).[34] After Wan was removed from power, Guangzhou newspapers started to report about Wan-group officials as much as other newspapers and as much as about the other officials (Panel B). Taken together, these results suggest that corruption scandals in local newspapers are censored in order to avoid implicating other officials who are still in power.

4.3 Deemphasising corruption scandals

The results of equation 4 (Table 3) suggest that newspapers deemphasise stories about corruption scandals involving high-level officials from the same province. Articles about corrupt high-level officials from the same province have on average 30 percent fewer words than articles about officials from other provinces (column 1). The effect of increasing the share of own-province low-level officials on article length is negative, but insignificant. Column 2 shows that the probability of an article containing an image decreases with the share of officials from the same province. Articles mentioning officials from the same province are placed significantly further forward in a newspaper, by around three pages compared to an average page number of 10 (column 3).

In one third of articles, newspapers name more than one official in order to chart the progress of the anti-corruption campaign. Column 4 shows that as the share of own-province high-level officials in an article goes from zero to one, the probability of an article referring to multiple officials is 13 percentage points lower. For low-level officials, this probability declines by 6 percentage points. One explanation is that newspapers selectively omit the names of corrupt officials from their own provinces when summarising

---

[33] "Catching Tigers and Flies". *ChinaFile*. `https://anticorruption.chinafile.com`, last accessed: 20 April 2019. ChinaFile names government organs and major media outlets as sources for its information. Although I focus on small city-level newspapers, I cannot rule out that this connection measure might be endogenous.

[34] The point estimate on the interaction between Wan officials and Guangzhou newspapers before Wan's investigation is negative, but imprecisely estimated.

the campaign.

Analysing the text of article headlines provides further details into the framing of corruption stories. In order to avoid controversy when covering sensitive topics, newspapers often copy official announcements or articles by established party mouthpieces. In column 5, the dependent variable is a dummy that takes the value of 1 when an article headline copies the CCDI announcement. Relative to a mean probability of copying of 0.9 percent, an article with only high-level officials from the same province is 1.6 percentage points more likely to be copied from the CCDI. Headlines about corrupt high-level officials from the same province are also 6 percentage points less likely to include references to "corruption", which is close to 30 percent of the average probability (column 6). Ten percent of articles in the sample explicitly refer to the anti-corruption campaign in their headlines, consistent with the central government's efforts to publicise the campaign results. Yet, when newspapers write about high-level officials from their own province, headlines are 3.5 percentage points less likely to reference the campaign (column 7). Headlines about low-level officials are also more likely to copy the official announcement and less likely to refer to the anti-corruption campaign.

Table 4 shows that articles about corruption scandal in a newspaper's home province also differ in tone. I measure the sentiment of each article's headline and body using the National Taiwan University Semantic Dictionary.[35] The text of articles about high-level officials from the same province contain on average 3.3 more positive words (relative to a mean of 23, column 3), and 3.5 fewer negative words (relative to a mean of 19, column 4). A similar pattern holds for low-level officials and the articles' headlines (columns 1 and 2). Taken together, the results of this section suggest that even when newspapers write about corruption scandals in their own province, they are more cautious in their writing and do not try to draw attention to these types of stories.

4.4 Interpretation of results

I interpret these results as evidence that local governments in China selectively censor unfavourable news rather than adhering to a central propaganda strategy. When a high-level provincial official is investigated for corruption by the CCDI, the provincial government attempts to limit how much attention is paid to this news story. This strategy involves pressuring local newspapers to publish fewer stories about the scandal and to make published stories less attractive to readers, for example by featuring official sounding headlines that do not contain any mentions of corruption. The extent of censorship depends on the relationship between a newspaper and their local government. Party newspapers that are government-owned and financed by public subsidies exhibit the largest bias in reporting compared to newspapers that rely on advertising revenue and have to compete for readership. Sub-provincial level newspapers censor less than their provincial-level counterparts when it comes to high-level corruption in the provincial government. Just as the censorship strategy of provincial governments runs counter to the objectives of the central government, sub-provincial governments in turn can only be imperfectly controlled by provincial governments. And while corruption scandals of low ranking officials are generally not censored in their own province, municipal newspapers underreport about corrupt officials from the same city. A rationale for censorship appears to be protecting powerful incumbent officials, as underreporting is more severe when a corrupt official is well-connected.

---

[35]The texts are tokenised and pre-processed using jieba.

In this section, I will outline some alternative hypotheses and show that they cannot be convincingly reconciled with all of the results. One alternative hypothesis is that the observed pattern is driven by the relative demand of newspaper readers for articles about corruption in their own compared to other provinces, possibly because readers are already informed about corruption at home. Two pieces of evidence reported in Section 4.1.1 suggest that this alternative hypothesis is unlikely. First, there were more than twice as many internet searches about corrupt officials in their home province, which suggests greater interest in corruption at home. Newspapers also report more about low-level corruption in their home province. Second, newspapers wrote more stories about high- relative to low-level officials from other provinces, suggesting that readers tend to be more interested in high-level corruption. These results suggest that readers demand more coverage of high-level corruption in their home province, implying that the results are a lower-bound estimate of the true extent of local government censorship.

An alternative demand-side explanation is that readers' demand varies by article type. Prior to an investigation, most articles about local officials are about their official activities (e.g. speeches and policies), whereas afterwards, articles focus on the investigation. If readers are less interested in corruption scandals than official activities, this could explain a drop in the publication probability of articles about local high-level officials after the investigation date, but not the results for low-level officials and officials from other provinces. This alternative hypothesis also cannot explain why high-level corruption scandals are reported less in *absolute* terms in their own province compared to other provinces (Figure 1). Or why newspapers deemphasise corruption scandals involving high-level officials from their own relative to other provinces (Table 3). In Section 5.2, I find no evidence that local newspapers are more likely to report about corruption involving officials from their own province *before* these officials are investigated.

On the supply side, these results could be driven by provincial governments *overreporting* corruption in other provinces. This story is difficult to reconcile with the findings in Appendix Tables A3 and A4. Newspapers who are more closely linked to the provincial government (official party or provincial newspapers) do not report more about corruption in other provinces, but do significantly underreport high-level corruption at home relative to other newspapers. Anecdotally, provincial governments collude when it comes to censorship.[36] If out-of-province newspapers also underreport corruption, then this would bias the estimated effect towards zero.

Another alternative interpretation is that these results reflect the central government's optimal strategy. In order to avoid social unrest in response to local corruption, the centre may choose to publicise the campaign using examples from other areas. A number of findings are difficult to rationalise with this hypothesis. First, although sub-provincial newspapers circulate in areas directly affected by corrupt high-level provincial officials, they underreport less than provincial papers (Appendix Table A4). Second, when newspapers report corruption cases in their home province, they are less likely to reference the centre and its anti-corruption campaign (Table 3). Third, underreporting depends on corrupt officials' proximity to powerful local leaders. It appears unlikely that the local population would be this well-informed about local networks (see Section 4.2).

In this paper I use the term censorship loosely to refer to any changes in media

---

[36]Local newspapers do focus their investigative reporting on corruption in other provinces (Repnikova, 2017). However, local governments have been trying to stop this practice by putting pressure on a newspaper's government of origin, lobbying the central government to pass laws forbidding cross-border reporting or even resorting to intimidation tactics and violence (He, 2004).

content due to government pressure. I am unable to distinguish between cases where local governments actively pressure newspapers into reporting less or differently and cases where editors and journalists self-censor, i.e., choose to change their reporting for fear of repercussions.

5 Robustness and extensions

5.1 Time-profile of estimates

I examine the time profile of the estimated coefficients over the sample period by estimating the following equation

$$y_{o,n,t} = \sum_{s=-24}^{12} \mathbf{D'_{o,n,t+s}} \alpha_\mathbf{s} + \mathbf{X'_t} \delta + \gamma_{o,n} + \epsilon_{o,n,t} \tag{6}$$

where $\mathbf{D'_{o,n,t+s}}$ includes all regressors from equation 2 for each month of the sample and the rest of the notation is as before.[37]

Figure 8 (and Appendix Table A8) shows all estimated coefficients of interest. Prior to the announcement, the estimated effects are not statistically different from zero suggesting that the investigation announcements were surprises. The month of the investigation has the largest impact on publication probability. For out-of-province newspapers, the effects fade gradually over the year following the announcement (Panels B and D). For newspapers from the same province, publication probability is only significantly higher in the month of the announcement for low-level officials (Panel C), while underreporting on high-level corruption persists for the year following the announcement (Panel A).

5.2 Articles about corruption pre-announcement

Newspapers could disproportionately publish articles that accuse officials from their own province of corruption, which then results in CCDI investigations. In this case, newspapers might report less about these officials after the investigation announcement, as they will have already reported these stories. This type of reverse causality appears inconsistent with the data. Figure 9 shows the share of newspaper articles that contain references to corruption in their headline. Articles that are published after the investigation announcement are seven times more likely to mention corruption. Before the investigation announcement, newspapers are more likely to mention corruption in connection with both high- and low-level officials from other provinces and only 0.6 percent of article headlines about high-level officials from the same province contain words related to corruption.[38]

I re-estimate equations 1 and 2 excluding 9 high- and 33 low-level officials for whom at least one article was published before the investigation announcement by a newspaper from the same province with the word "corruption" in the headline. Appendix Table A10 shows that in this restricted sample newspapers underreport high-level corruption scandals from the same province, although the estimated magnitude is smaller.

---

[37]Due to collinearity, I omit month fixed effects.

[38]The majority of the 59 pre-announcement articles about same-province high-level officials which include "corruption" referred to *anti-corruption* speeches and activities, while 9 headlines were ambiguous.

5.3 Flexible official- and newspaper-province specification

As a specification test, I replace the dummy variable for whether or not a newspaper is from the same province as an official ($S_{o,n}$) in equation 1 with a vector of indicators for the provinces of newspapers and high-level official. Table 5 shows how newspapers from a province (columns) reports about officials in another province (rows) after that official is investigated for corruption. The diagonal entries (in bold) show the equivalent to α_3 separately for each province with at least one corrupt high-level official. All but one (for Hunan) of the diagonal entries are estimated to be negative, while the off-diagonal entries are all estimated to be positive. This suggests that the dummy ($S_{o,n}$) specification is supported by the data.[39]

5.4 Functional form specification

I test the sensitivity of estimates to different functional form specifications by estimating equations 1 and 2 using a Poisson QML estimator with the number of articles as the dependent variable ($\ln(N_{o,n,t})$) and clustered bootstrapped standard errors at the newspaper and official level. Results are shown in the first two columns of Appendix Table A9.[40] Since different assumptions underlie these models, their estimates are not directly comparable. However, all models point to the underreporting of corruption scandals involving high-level officials, when they are from the same province as a newspaper. As a comparison, columns 3 and 4 of Appendix Table A9 show the OLS results for the Poisson sample which omits all official-newspaper pairs where no article was ever published. In the restricted sample, the estimates generally become larger in magnitude and gain in significance.

5.5 Selection of officials and newspapers

Both official and newspaper samples are non-randomly selected. While all regressions control for official-newspaper-pair (or official and newspaper) fixed effects, differential sample selection could still be problematic. The CCDI could investigate groups of high-level officials from the same province at the same time, but only post an announcement about one official. If out-of-province newspapers report about the CCDI announcement and local newspapers write about the entire group, then using the sample of CCDI announcements could generate the result that newspapers underreport high-level corruption in their own province. To rule out this channel, I compare the list of high-level corrupt officials with a number of different news reports and data on the anti-corruption campaign by ChinaFile, which also uses media sources to identify corrupt officials. I find no mention of any additional corrupt high-level official during the sample period.[41] ChinaFile uses a

---

[39]Given the small number of provinces, it is difficult to draw any firm conclusions about heterogeneity across provinces. Two provinces with the largest absolute coefficients (Sichuan and Jiangxi) are commonly identified as two of the three provinces (the other province - Shanxi - is not in the sample) that experienced the most intensive anti-corruption crackdown during this period (see, e.g., Lorentzen and Lu, 2018). One of the province-level municipalities with the highest coefficient (Chongqing) is the province of Bo Xilai, who was widely seen as Xi Jinping's rival.

[40]The results are also robust to using a fixed-effects logit estimator with $\ln(\frac{y_{o,n,t}}{1-y_{o,n,t}})$ as the dependent variable. Results not shown.

[41]Some media sources also mention high-ranking military officers as part of the campaign. Military officers are not comparable to government and party officials, as it has its own CCDI-equivalent agency and media organisations.

wider definition of high-ranking officials or "Tigers". They identify 41 high-ranking officials, which includes the original 34 used in this paper and an additional seven officials that I classed as low ranking.[42] Using this alternative definition does not change the main results significantly, see Appendix Table A11. Due to differences in data sources, there are also discrepancies among low-level officials: 16 officials are in the CCDI, but not in the ChinaFile sample and 46 officials vice versa.

Selection of newspapers could affect the results about differences in censorship across newspaper type, if types of newspapers are concentrated in specific provinces. As all major newspapers in Guangdong province are in the WiseNews database, the results reported in Section 4.2.2 suggest that the results are not purely driven by selection. I also match each subsidiary newspapers to their parent newspaper (an official party newspaper) when both are in the sample and make a direct paired comparison about their reporting. Figure 10 shows the difference in the mean probability of publication before and after an official is investigated, i.e., a negative number means that newspapers report less about an official after their scandal relative to before. Each point represents a subsidiary and parent newspaper pair, with values for the subsidiary shown on the y-axis and the parent newspaper on the x-axis. Panel A shows that for high-level officials from the same province all points lie above the 45 degree line, subsidiary newspapers underreport less about high-level corruption in their own province relative to their parent newspaper, confirming the results of Section 4.1.3.

6   Conclusion

Media censorship is a common tool for many autocratic governments to maintain control over their population. In this paper, I show how conflicts of interest within an autocratic government can lead to diverging media censorship strategies in the context of local newspaper reporting about a centralised anti-corruption campaign in China. Local newspapers report differently on corrupt officials, depending on three factors: an official's importance, a newspaper's owners and the relationship between newspaper and official. Despite greater interest from newspaper readers, newspapers write fewer articles about corruption scandals involving high-level officials from their own province relative to scandals involving officials from other provinces. These articles are also shorter and feature less interesting headlines without references to corruption or the anti-corruption campaign. Corruption scandals of well-connected officials are censored more. Commercial newspapers, which compete for readers and advertising revenue, tend to censor less. Within a province, different levels of government may also pursue conflicting censorship strategies. Articles about high-level corruption in the provincial government are more often suppressed in provincial newspapers relative to city newspapers. These results illustrate how different media content under censorship can reveal agency problems in an autocracy.

References

Adena, Maja, Ruben Enikolopov, Maria Petrova, Veronica Santarosa, and Ekaterina Zhuravskaya. 2015. "Radio and the Rise of The Nazis in Prewar Germany." *Quarterly Journal of Economics*, 130(4): 1885–1939.

[42]The difference mainly stems from officials working at state-owned enterprises.

Beattie, Graham, Ruben Durante, Brian Knight, and Ananya Sen. 2017. "Advertising Spending and Media Bias: Evidence from News Coverage of Car Safety Recalls." *National Bureau of Economic Research Working Paper*, 23940.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2011. "Robust Inference with Multiway Clustering." *Journal of Business & Economic Statistics*, 29(2): 238–249.

Chen, Yuyu, and David Y. Yang. forthcoming. "The Impact of Media Censorship: 1984 or Brave New World?" *American Economic Review*.

Di Tella, Rafael, and Ignacio Franceschelli. 2011. "Government Advertising and Media Coverage of Corruption Scandals." *American Economic Journal: Applied Economics*, 3(4): 119–151.

Durante, Ruben, and Brian Knight. 2012. "Partisan Control, Media Bias, and Viewer Responses: Evidence from Berlusconi's Italy." *Journal of the European Economic Association*, 10(3): 451–481.

Egorov, Georgy, Sergei Guriev, and Konstantin Sonin. 2009. "Why Resource-Poor Dictators Allow Freer Media: A Theory and Evidence from Panel Data." *American Political Science Review*, 103(04): 645–668.

Enikolopov, Ruben, Maria Petrova, and Ekaterina Zhuravskaya. 2011. "Media and Political Persuasion: Evidence from Russia." *American Economic Review*, 101(7): 3253–3285.

Fisman, Raymond, and Yongxiang Wang. 2017. "The Distortionary Effects of Incentives in Government: Evidence from China's." *American Economic Journal: Applied Economics*, 9(2): 202–218.

Francois, Patrick, Francesco Trebbi, and Kairong Xiao. 2016. "Factions in Nondemocracies: Theory and Evidence from the Chinese Communist Party." *National Bureau of Economic Research Working Paper*, 22775.

Gentzkow, Matthew, and Jesse M. Shapiro. 2010. "What Drives Media Slant? Evidence from US Daily Newspapers." *Econometrica*, 78(1): 35–71.

He, Qinglian. 2004. "Media Control in China." *China Rights Forum*, 1: 11–28.

Jia, Ruixue, and Huihua Nie. 2017. "Decentralization, Collusion, and Coal Mine Deaths." *Review of Economics and Statistics*, 99(1): 105–118.

King, Gary, Jennifer Pan, and Margaret E. Roberts. 2013. "How Censorship in China Allows Government Criticism but Silences Collective Expression." *American Political Science Review*, 107(02): 326–343.

King, Gary, Jennifer Pan, and Margaret E. Roberts. 2014. "Reverse-Engineering Censorship in China: Randomized Experimentation and Participant Observation." *Science*, 345(6199): 1251722.

Li, Hongbin, and Li-An Zhou. 2005. "Political Turnover and Economic Performance: The Incentive Role of Personnel Control in China." *Journal of Public Economics*, 89(9): 1743–1762.

Lorentzen, Peter. 2014. "China's Strategic Censorship." *American Journal of Political Science*, 58(2): 402–414.

Lorentzen, Peter L., and Xi Lu. 2018. "Personal Ties, Meritocracy, and China's Anti-Corruption Campaign." *SSRN Electronic Journal*.

Mookherjee, Dilip. 2015. "Political Decentralization." *Annual Review of Economics*, 7(1): 231–249.

Pan, Jennifer, and Kaiping Chen. 2018. "Concealing Corruption: How Chinese Officials Distort Upward Reporting of Online Grievances." *American Political Science*

Review, 1–19.

Prat, Andrea, and David Strömberg. 2013. "The Political Economy of Mass Media." In *Advances in Economics and Econometrics.* , ed. Daron Acemoglu, Manuel Arellano and Eddie Dekel, 135–187. Cambridge University Press.

Puglisi, Riccardo, and James M. Snyder Jr. 2011. "Newspaper Coverage of Political Scandals." *Journal of Politics,* 73(3): 931–950.

Qin, Bei, David Strömberg, and Yanhui Wu. 2018. "Media Bias in China." *American Economic Review,* 108(9): 2442–2476.

Repnikova, Maria. 2017. *Media Politics in China: Improvising Power under Authoritarianism.* Cambridge University Press.

Shirk, Susan L. 2011. *Changing Media, Changing China.* Oxford University Press.

Sparks, Colin, Haiyan Wang, Yu Huang, Yanhua Zhao, Nan Lü, and Dan Wang. 2016. "The Impact of Digital Media on Newspapers: Comparing Responses in China and the United States." *Global Media and China,* 1(3): 186–207.

Stockmann, Daniela, and Mary E. Gallagher. 2011. "Remote Control: How the Media Sustain Authoritarian Rule in China." *Comparative Political Studies,* 44(4): 436–467.

Yanagizawa-Drott, David. 2014. "Propaganda and Conflict: Evidence from the Rwandan Genocide." *Quarterly Journal of Economics,* 129(4): 1947–1994.

Figures

Figure 1: Fraction of official-newspaper pairs with article per given day by official rank
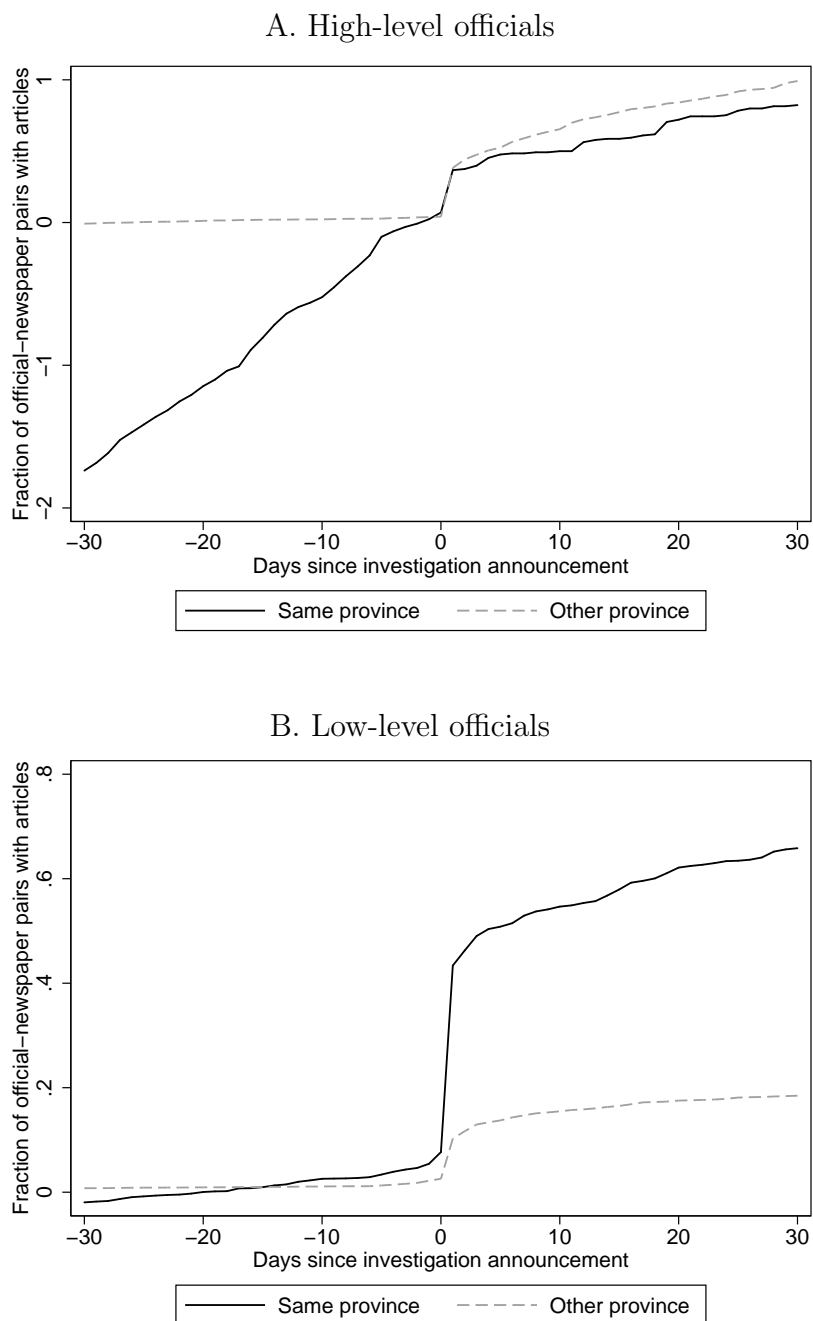
A. High-level officials

B. Low-level officials

These charts show the fraction of official-newspaper pairs with at least one article per day on the y-axis, cumulated from the day of the announcement.

Figure 2: Search intensity for corrupt officials on Baidu
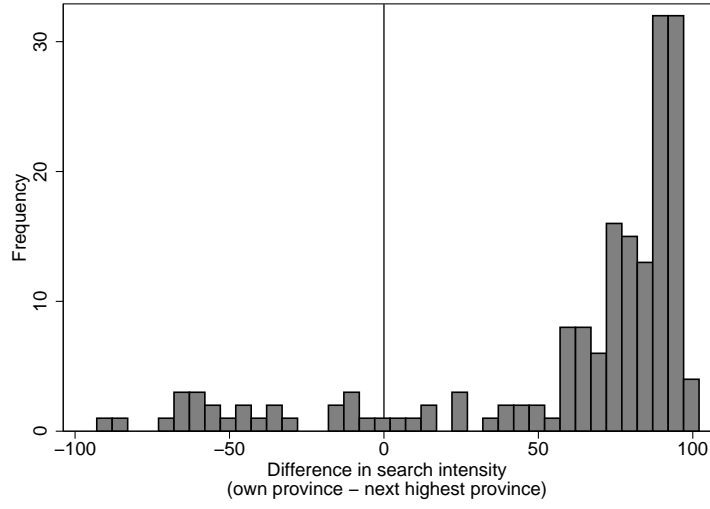
This chart shows the distribution of the difference in Baidu search intensity for each official in their own province relative to the next highest other province in the year following the investigation announcement (x-axis). A number greater than zero means that there were more searches in an official's own province. The maximum number of searches is normalised to 100. Data are only available for 214 officials, of which data for 21 officials are based on partial information about the year post-announcement (since September 2013).

Figure 3: Heterogeneity by newspaper type
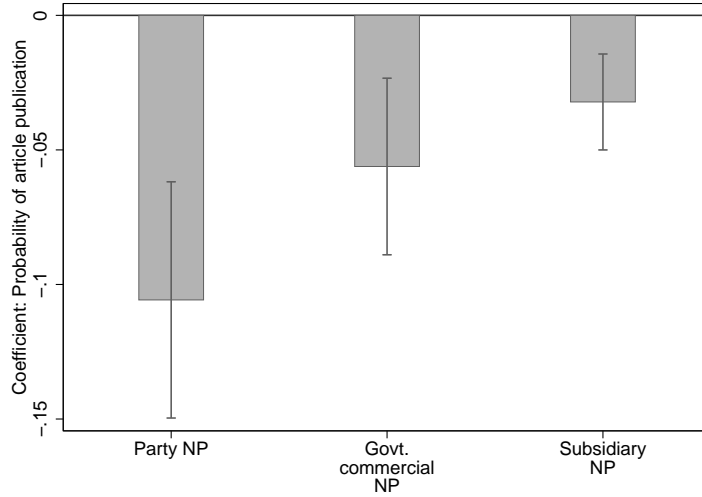
This chart shows coefficients from a regression of a dummy for whether an article is published on the interaction terms between dummies for three different newspaper types (party newspapers, government-owned commercial newspapers and subsidiary newspapers) as in equation 3 and an indicator for high-level officials from the same province as a newspaper in the post-scandal period ($H_o \times S_{o,n} \times P_{o,t}$). The regression includes official-newspaper-pair fixed effects and all seasonal controls from column 2 in Table 1. Standard errors are clustered by newspaper and official. The point estimate is shown in the grey bars and 90% confidence intervals are marked by the black lines. The full set of coefficients is shown in Appendix Table A3.

23

Figure 4: Heterogeneity by newspaper administrative level
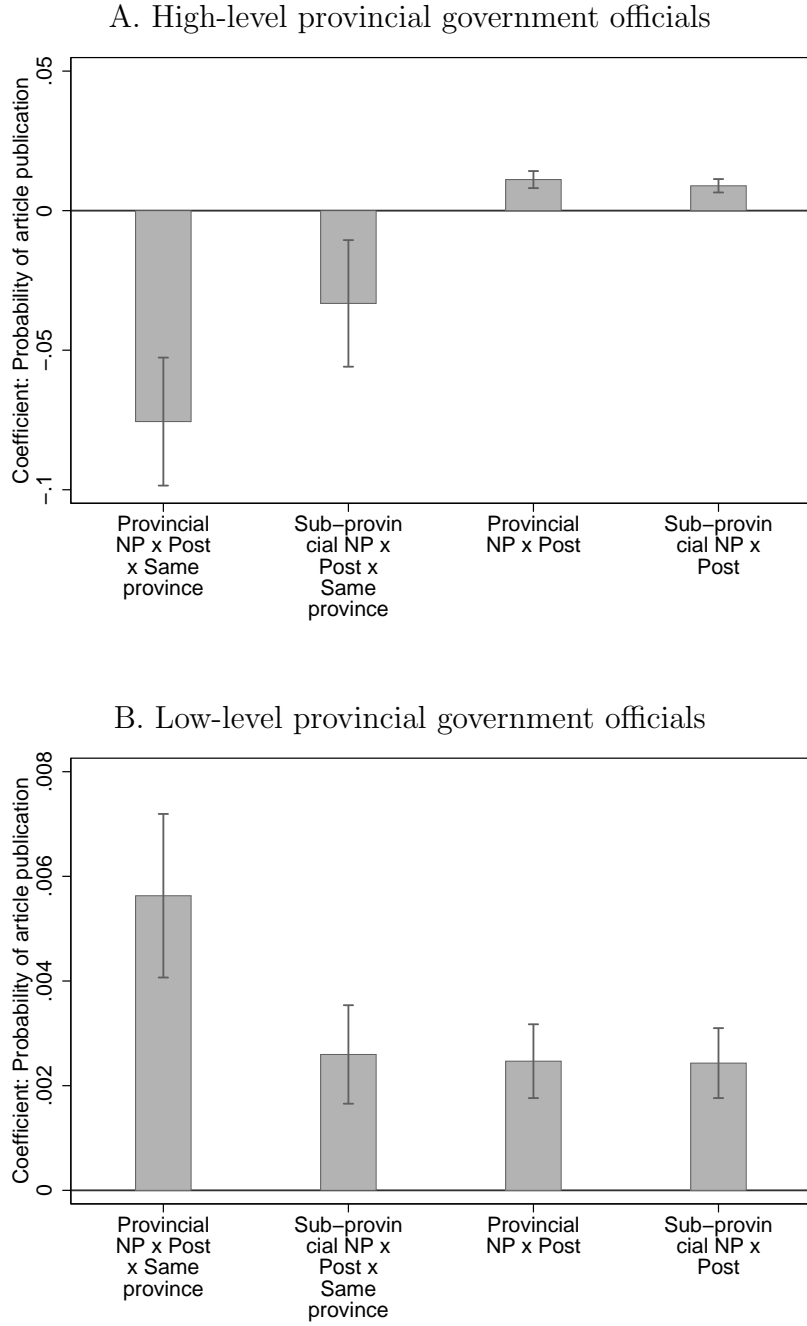
A. High-level provincial government officials

B. Low-level provincial government officials

These charts show coefficients from a regression of a dummy for whether an article is published on the interaction terms between two dummy variables for provincial-level and sub-provincial level newspapers and an indicator for the post-scandal period and newspapers from the same province ($P_{o,t} \times S_{o,n}$) as high-level (Panel A) and low-level provincial government officials (Panel B). The sample is restricted to provincial government officials and local newspapers. Coefficients on the same interaction terms for out-of-province newspapers in the post-scandal period ($P_{o,t}$) are shown as a comparison. The regression includes official-newspaper-pair fixed effects and all seasonal controls from column 2 in Table 1. Standard errors are clustered by newspaper and official. The point estimate is shown in the grey bars and 90% confidence intervals are marked by the black lines. The full set of coefficients is shown in Appendix Table A4.

Figure 5: Municipality-level results

This chart shows coefficients from a regression of a dummy for whether an article is published on an indicator for the post-scandal period ($P_{o,t}$) and its interactions with two dummy variables that takes the value of 1 if an official and a newspaper are from the same city ($P_{o,t} \times SC_{o,n}$) and if an official and a newspaper are from different cities of the same province ($P_{o,t} \times DC_{o,n}$). The sample is restricted to sub-provincial newspapers and officials. The regression includes official-newspaper-pair fixed effects and all seasonal controls from column 2 in Table 1. Standard errors are clustered by newspaper and official. The point estimate is shown in the grey bars and 90% confidence intervals are marked by black lines. The coefficients are reported in Appendix Table A5.
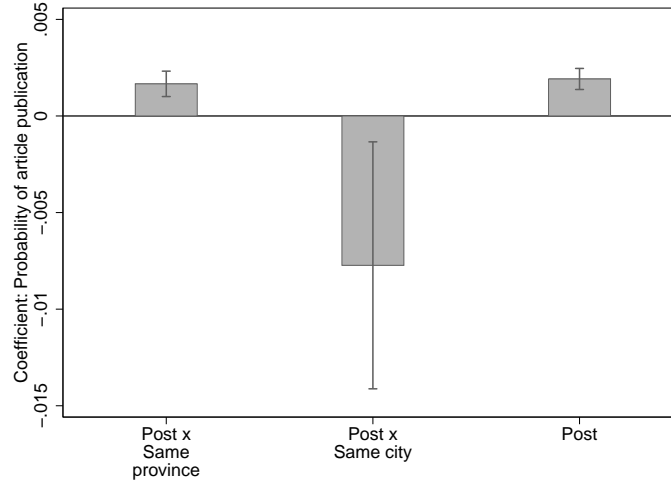
Figure 6: Proximity to the incumbent party secretary
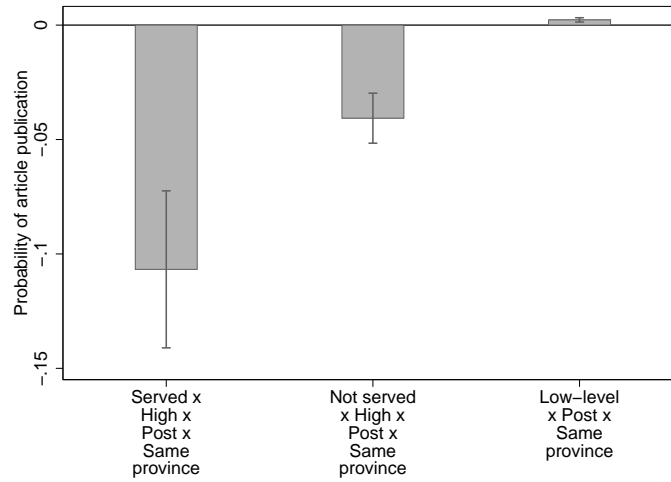

This chart shows coefficients from a regression of a dummy for whether an article is published on the interaction between a dummy variable for whether a high-level official o has served on the provincial party standing committee with the incumbent party secretary of the province (served vs. not served) and a post-scandal dummy interacted with an indicator for whether the newspaper and official are from the same province ($P_{o,t} \times S_{o,n}$). The coefficient on low-level officials from the same province post-scandal is shown as a comparison. The regression includes official-newspaper-pair fixed effects and all seasonal controls from column 2 in Table 1. Standard errors are clustered by newspaper and official. The point estimate is shown in the grey bars and 90% confidence intervals are marked by black lines. The full set of coefficients is shown in Appendix Table A6.

Figure 7: Officials connected to Wan Qingliang

A. Before Wan Qingliang's investigation announcement
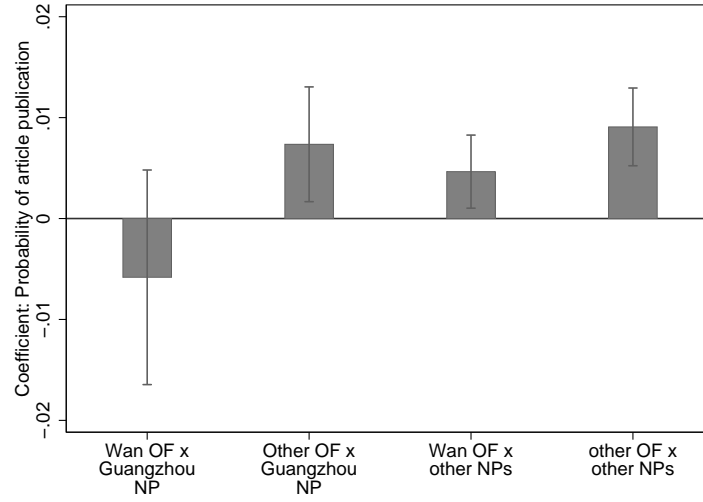
B. After Wan Qingliang's investigation announcement
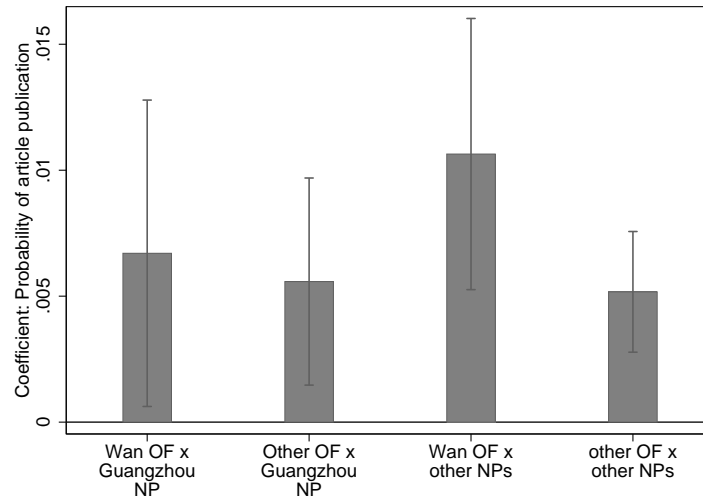
These charts show coefficients from a regression of a dummy for whether an article is published on an indicator for the period after each individual official's corruption scandal ($P_{o,t}$) and before Wan Qingliang's investigation announcement (Panel A) or after his investigation announcement (Panel B), interacted with two sets of mutually exclusive dummies: 1) for whether a newspaper is from Guangzhou city (Guangzhou NP) or from Guangdong province (other NPs) and 2) for whether an official is connected to Wan Qingliang (Wan OF) or other Guangdong officials of the same ranks (other OF). The regression includes official-newspaper-pair fixed effects and all seasonal controls from column 2 in Table 1. Standard errors are clustered by newspaper and official. The point estimate is shown in the grey bars and 90% confidence intervals are marked by the black lines. The full set of coefficients is shown in Appendix Table A7.

Figure 8: Time profile of estimates
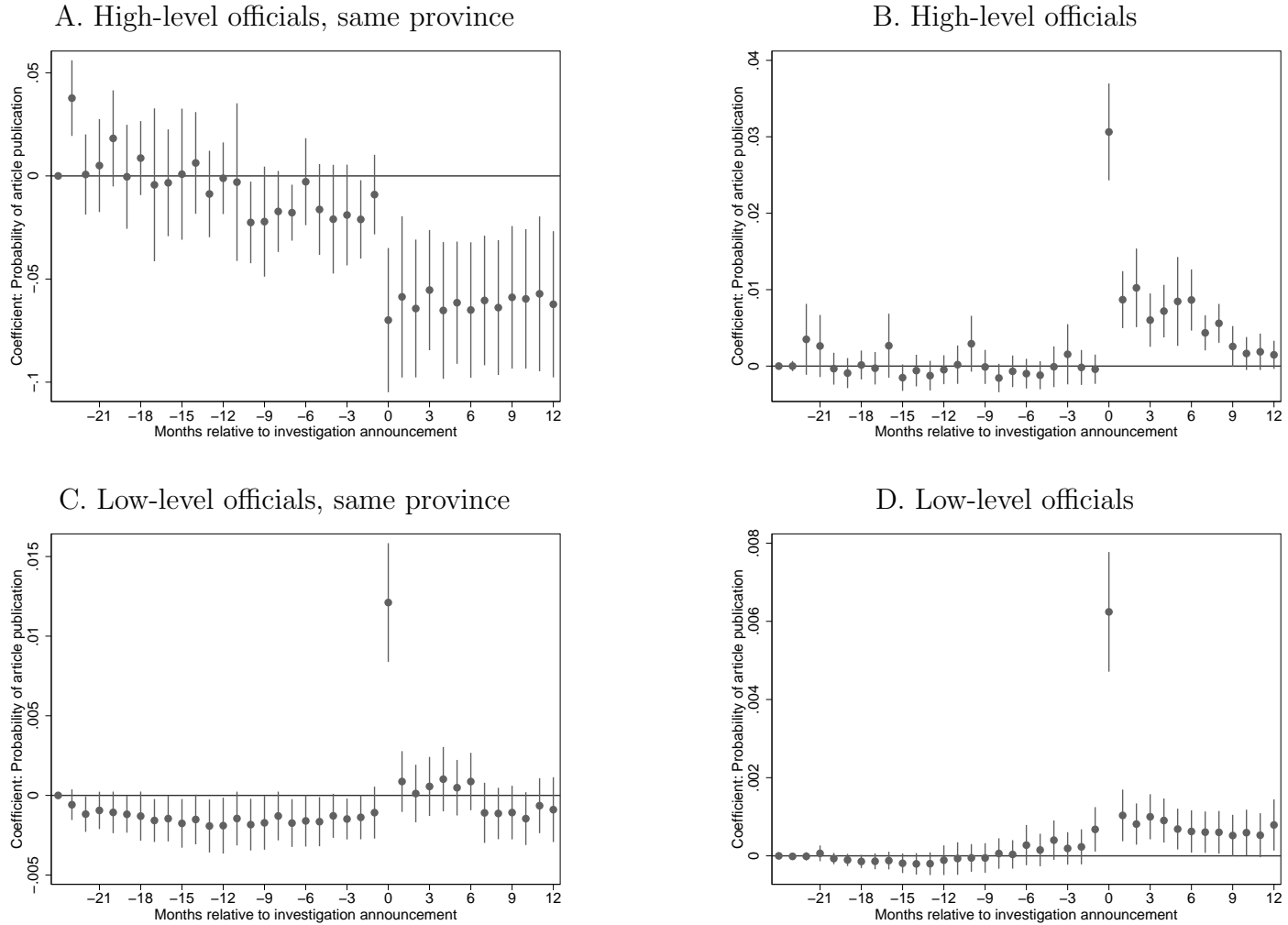
A. High-level officials, same province

B. High-level officials

C. Low-level officials, same province

D. Low-level officials

These charts show coefficients from a regression of a dummy for whether an article is published on a vector of month-since-announcement dummies interacted with dummies for whether an official is high-ranking and whether a newspaper and official are from the same province, see equation 6. The point estimates are shown in dots with 90% confidence intervals. On the x-axis are months since the investigation announcement. Month 24 before the announcement is the omitted category. The regression includes official-newspaper-pair fixed effects and all seasonal controls from column 2 in Table 1 except for month fixed effects. Standard errors are clustered by newspaper and official. Coefficient estimates are reported in Appendix Table A8.

Figure 9: Share of articles with "corruption" in the headline
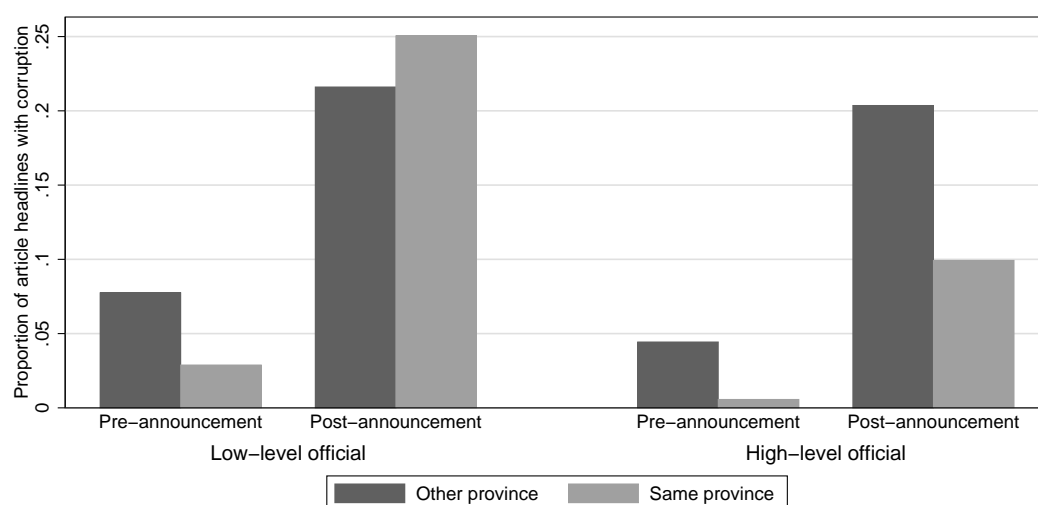
This chart shows the fraction of articles that have a headline which includes references to corruption, by whether or not the article appeared before or after the investigation announcement, whether the official and newspaper are from the same or different provinces and whether the official is low or high ranking. The chart is based on a sample of 32,906 articles that only mention one official.

Figure 10: Subsidiary and their parent newspapers

A. High-level officials, same province

B. High-level officials

C. Low-level officials, same province

D. Low-level officials
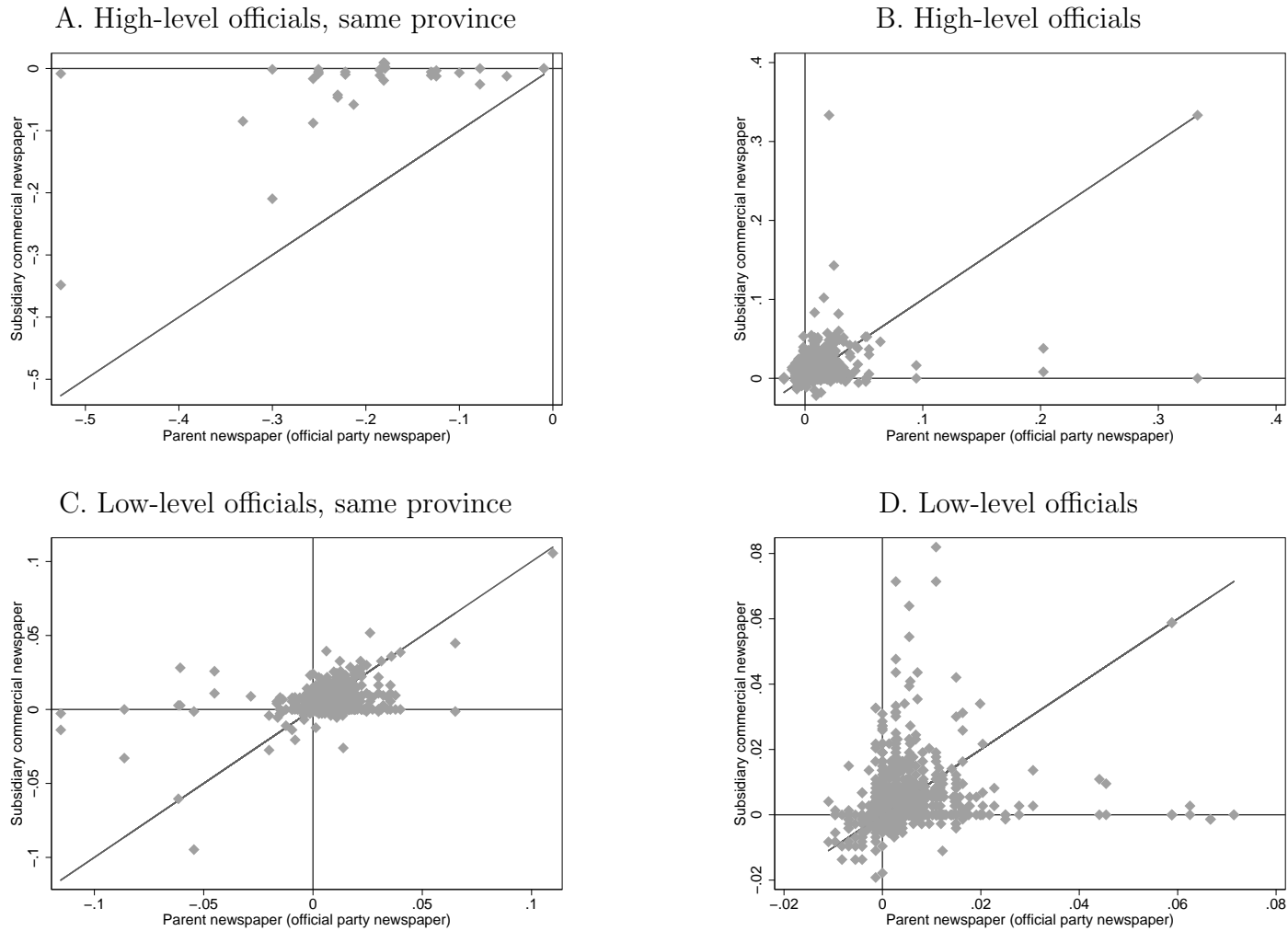
These charts show the difference in the mean daily probability of reporting about a given official before and after their corruption investigation announcement for a subsidiary newspaper (y-axis) relative to their parent newspaper (x-axis). Each point is a subsidiary-parent newspaper pair. A negative number means that reporting is lower after the announcement than before. The 45 degree line is shown for comparison.

Tables

Table 1: Underreporting of corruption scandals

| | Dummy for whether an article is published | | | |
	(1)	(2)	(3)	(4)
Post-scandal x Same province	-0.00080	-0.00086		
	(0.0010)	(0.0011)		
Post-scandal	0.0015***	0.0030***		
	(0.00021)	(0.00041)		
High-level official x				
Post-scandal x Same province			-0.057***	-0.057***
			(0.013)	(0.013)
High-level official x Post-scandal			0.0077***	0.0089***
			(0.0013)	(0.0014)
Low-level official x				
Post-scandal x Same province			0.0023***	0.0023***
			(0.00052)	(0.00060)
Low-level official x Post-scandal			0.00084***	0.0023***
			(0.00012)	(0.00030)
day-of-week FE		Yes		Yes
month FE		Yes		Yes
year FE		Yes		Yes
holiday FE		Yes		Yes
CCP meeting FE		Yes		Yes
No. of obs.	39,929,390	39,929,390	39,929,390	39,929,390
R-squared	0.043	0.043	0.045	0.045
Mean dep. var.	0.0012	0.0012	0.0012	0.0012

This table shows results from regressions of a dummy for whether an article is published on interactions between dummies for the post-scandal period, whether a newspaper and official are from the same province, whether an official is high-ranking and different controls, see equations 1 and 2. All regressions include official-newspaper-pair fixed effects. Standard errors in parentheses are clustered by newspaper and official. Significance at the 1%, 5% and 10% is denoted by ***, ** and *, respectively.

Table 2: Comments on Sina Weibo posts about high-level corrupt officials

	(1) Number of comments	(2) Time to first comment	(3) Positive words	(4) Negative words	(5) Verified	(6) Characteristics of commenters Posts	(7) Followers	(8) Friends
Share of high-level officials from same province	5.52***	36.6	-0.15*	-0.060	-0.017	-1815.0**	-44347.1	-27.3
	(1.23)	(3412.0)	(0.083)	(0.078)	(0.017)	(747.2)	(51790.6)	(51.1)
Newspaper from same province	6.62***	-3719.1*	-0.10*	-0.12**	0.028	707.7	28734.5	-21.3
	(2.33)	(2193.6)	(0.052)	(0.057)	(0.019)	(1033.0)	(156762.4)	(40.9)
No. of obs.	3,638	3,638	3,638	3,638	3,638	3,638	3,638	3,638
R-squared	0.20	0.25	0.070	0.055	0.072	0.12	0.034	0.089
Mean dep. var.	6.27	2498.9	0.52	0.74	0.11	6495.7	140058.1	611.3

This table shows results from regressions of different Sina Weibo comments characteristics by province on the share of high-level officials from the same province mentioned in each post and a dummy variable for whether the newspaper is from the same province, see equation 5. The dependent variables are the following: 1) number of comments, 2) time to first comment (in minutes), 3) average number of positive words, 4) average number of negative words, 5) percentage of commenter accounts that are verified, 6) number of posts by commenters, 7) number of followers of commenters and 8) number of friends of commenters. Positive and negative words are classified using the National Taiwan University Semantic Dictionary, following pre-processing using jieba. All regressions include official, newspaper and province fixed effects, as well as seasonal controls from column 2 in Table 1. Standard errors in parentheses are clustered by newspaper. Significance at the 1%, 5% and 10% is denoted by ***, ** and *, respectively.

Table 3: Framing of corruption news stories

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
						Article headlines contains	
	Log words	Includes image	Page number	Multiple officials	Official announcement	Corruption	Anti-corruption campaign
Share of high-level officials from same province	-0.30***	-0.0038	-2.94***	-0.13***	0.016*	-0.058***	-0.035*
	(0.096)	(0.013)	(0.70)	(0.021)	(0.0094)	(0.018)	(0.018)
Share of low-level officials from same province	-0.031	-0.0090	-4.15***	-0.059***	0.010**	0.0083	-0.020**
	(0.057)	(0.0066)	(0.61)	(0.012)	(0.0046)	(0.011)	(0.0084)
No. of obs.	18,458	18,458	17,186	18,458	18,458	18,458	18,458
R-squared	0.36	0.37	0.43	0.48	0.10	0.15	0.18
Mean dep. var.	6.46	0.091	10.2	0.31	0.0086	0.22	0.10

This table shows results from regressions of different characteristics of articles published post-scandal on the share of high- and low-level officials from the same province in the total number of officials mentioned in each article, see equation 4. The dependent variables are the following: 1) log words in article, 2) dummy for whether article includes an image, 3) page number of article, 4) dummy for whether article refers to more than one official, 5) dummy for whether article headline copies official CCDI investigation announcement, 6) dummy for whether article headline contains references to corruption and 7) dummy for whether article headline includes keywords related to the anti-corruption campaign. All regressions include official and newspaper fixed effects, as well as seasonal controls from column 2 in Table 1. Standard errors in parentheses are clustered by newspaper. Significance at the 1%, 5% and 10% is denoted by ***, ** and *, respectively.

Table 4: Sentiment of corruption news stories

	Headline		Body	
	Positive (1)	Negative (2)	Positive (3)	Negative (4)
Share of high-level officials from same province	0.17***	0.051	3.27*	-3.48***
	(0.042)	(0.048)	(1.84)	(0.98)
Share of low-level officials from same province	0.044***	-0.087**	-0.20	-2.10***
	(0.015)	(0.033)	(1.10)	(0.64)
No. of obs.	18,458	18,458	18,458	18,458
R-squared	0.15	0.12	0.25	0.26
Mean dep. var.	0.24	0.49	22.8	19.3

This table shows results from regressions of the number of positive and negative words included in a post-scandal's article's headline (columns 1 and 2) and body (columns 3 and 4) on the share of high- and low-level officials from the same province in the total number of officials mentioned in each article, see equation 4. Positive and negative words are classified using the National Taiwan University Semantic Dictionary, following pre-processing using jieba. All regressions include official and newspaper fixed effects, as well as seasonal controls from column 2 in Table 1. Standard errors in parentheses are clustered by newspaper. Significance at the 1%, 5% and 10% is denoted by ***, ** and *, respectively.

Table 5: Coefficients for all newspaper and official provinces (high-level officials only)

Official province	Newspaper province									
	Centre	Sichuan	Tianjin	Shandong	Guangdong	Guangxi	Jiangxi	Hubei	Hunan	Chongqing
Centre	0.008***	0.009***	0.011***	0.014***	0.014***	0.015***	0.017***	0.014***	0.014**	0.014***
	(0.003)	(0.002)	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)	(0.003)	(0.006)	(0.003)
Sichuan	0.008***	**-0.071***	0.008	0.016***	0.014***	0.013***	0.026***	0.013***	0.013*	0.015***
	(0.003)	(0.041)	(0.007)	(0.004)	(0.002)	(0.004)	(0.009)	(0.003)	(0.007)	(0.003)
Tianjin	0.011***	0.008***	**-0.07***	0.018***	0.021***	0.018***	0.018***	0.019***	0.019**	0.017***
	(0.002)	(0.002)	(0.014)	(0.005)	(0.002)	(0.003)	(0.004)	(0.003)	(0.007)	(0.002)
Shandong	0.008**	0.005**	0.01	**-0.003**	0.01	0.011**	0.015*	0.011**	0.013*	0.014**
	(0.004)	(0.002)	(0.006)	(0.007)	(0.007)	(0.005)	(0.008)	(0.005)	(0.007)	(0.007)
Guangdong	0.009***	0.008***	0.012**	0.013*	**-0.04***	0.013***	0.018**	0.014***	0.017**	0.012***
	(0.003)	(0.003)	(0.005)	(0.007)	(0.018)	(0.005)	(0.007)	(0.005)	(0.008)	(0.002)
Guangxi	0.007***	0.008***	0.008*	0.014***	0.014***	**-0.029***	0.018***	0.013***	0.014**	0.018***
	(0.002)	(0.002)	(0.005)	(0.004)	(0.002)	(0.007)	(0.003)	(0.003)	(0.006)	(0.001)
Jiangxi	0.009***	0.009***	0.011***	0.014***	0.014***	0.015***	**-0.071***	0.012***	0.011***	0.012***
	(0.002)	(0.002)	(0.003)	(0.004)	(0.003)	(0.005)	(0.04)	(0.003)	(0.004)	(0.002)
Hubei	0.009***	0.011***	0.009	0.019***	0.018***	0.018***	0.022***	**-0.03**	0.014**	0.012***
	(0.003)	(0.003)	(0.006)	(0.002)	(0.003)	(0.005)	(0.004)	(0.018)	(0.006)	(0.002)
Hunan	0.011***	0.01***	0.014	0.018***	0.019***	0.019***	0.022***	0.016***	**0.01***	0.012***
	(0.003)	(0.003)	(0.009)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.002)
Chongqing	0.007***	0.008***	0.01***	0.009**	0.009***	0.01***	0.013***	0.011***	0.01**	**-0.086***
	(0.002)	(0.003)	(0.002)	(0.004)	(0.003)	(0.002)	(0.003)	(0.002)	(0.004)	(0.006)

This table shows results from the regression of a dummy for whether an article is published on the interaction between a post-scandal-period dummy and indicators for each newspaper's (in columns) and official's province (in rows) for high-level officials. (Only provinces with at least one high-level official caught during the campaign are reported.) In bold are the coefficients when an official and a newspaper are from the same province. The regression includes official-newspaper-pair fixed effects, as well as seasonal controls from column 2 in Table 1. There are 3,458,778 observations in this regression, the R-squared is 0.076 and the mean dependent variable is 0.006. Standard errors in parentheses are clustered by newspaper and official. Significance at the 1%, 5% and 10% is denoted by ***, ** and *, respectively.

A Appendices (for online publication)

A.1 Appendix I: Detailed sample descriptions

Figure A1: Map of provinces in sample

This map shows provinces and province-level municipalities in the newspaper sample in dark grey. Source: GADM version 3.6. See `www.gadm.org`. WGS84 projection.

Figure A2: Geographic distribution of officials and newspapers
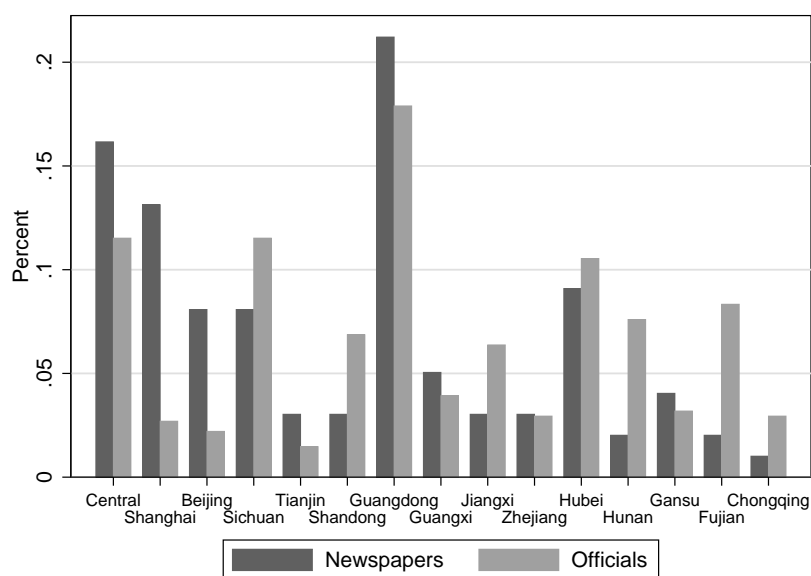
This chart shows the distribution of officials and newspapers in the sample across provinces.

Figure A3: Distribution of officials' ranks
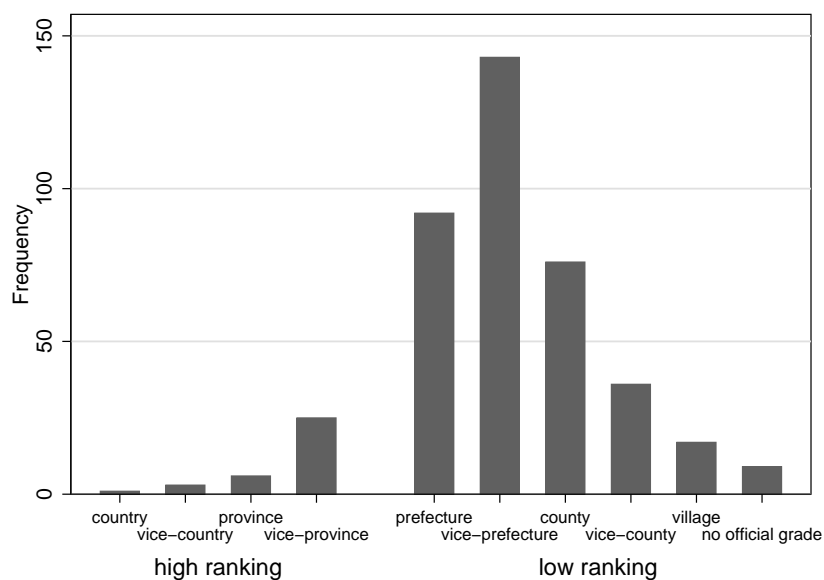
This chart shows the distribution of official ranks in the sample.

Figure A4: Fraction of official-newspaper pairs with article per given day

A. Full sample

B. Two-weeks before and after investigation announcement
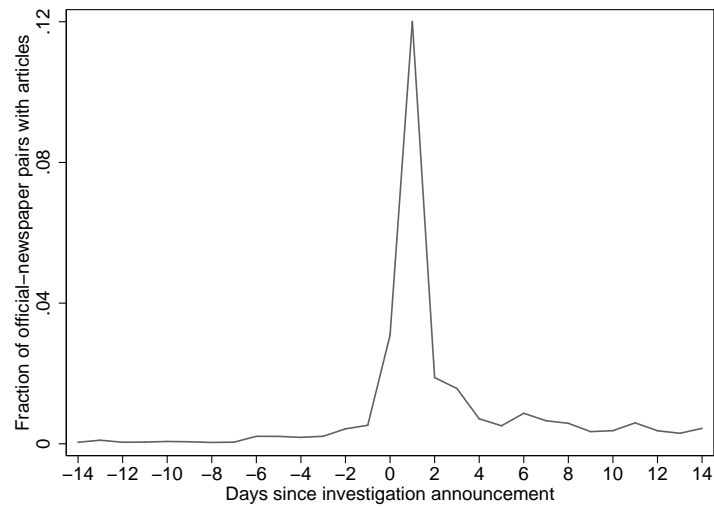
These charts show the fraction of official-newspaper pairs with at least one article per day on the y-axis.

Table A1: Summary statistics

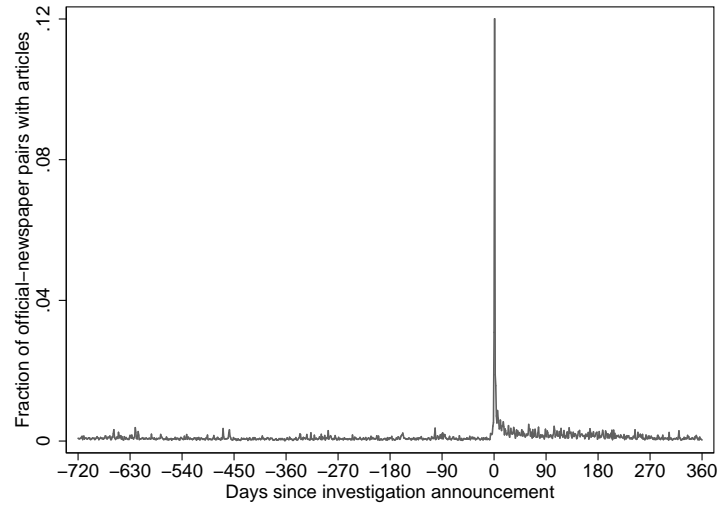|	mean	sd	min	max	obs
A. Newspapers					
Party NP	0.21	0.41	0	1	99
Govt-owned commercial NP	0.27	0.45	0	1	99
Subsidiary NP	0.52	0.50	0	1	99
Central NP	0.18	0.39	0	1	99
Provincial NP	0.55	0.50	0	1	99
Sub-provincial NP	0.27	0.45	0	1	99
Issues per week	5.97	2.02	1	7	99
B. Officials					
High-ranking	0.09	0.28	0	1	408
Government	0.78	0.41	0	1	408
Local govt.	0.72	0.45	0	1	408
Provincial govt.	0.19	0.39	0	1	408
Sub-provincial govt.	0.53	0.50	0	1	408
SOE	0.13	0.34	0	1	408
Other institution	0.08	0.27	0	1	408
C. Articles					
Number of words	1332	1733	12	32767	39,205
Log (words)	6.67	1.07	2	10	39,205
Incl. image	0.09	0.29	0	1	39,172
Page number	8.32	8.48	1	134	35,172
Front page	0.15	0.36	0	1	35,172
Multiple officials	0.27	0.44	0	1	16,473
Copy of official announcement	0.01	0.10	0	1	16,473
Corruption (in headline)	0.22	0.41	0	1	16,473
Campaign (in headline)	0.08	0.28	0	1	16,473
D. Official-Newspaper-Day Panel					
Article	0.001	0.034	0	1	39,929,391
Number of articles	0.001	0.042	0	11	39,929,391
Same province	0.076	0.266	0	1	39,929,391

This table reports summary statistics for newspapers (Panels A), officials under investigation (Panel B) and newspaper articles (Panel C), where the last four rows refer only to articles published following a corruption scandal. Panel D shows summary statistics for the main regressor and dependent variables in the constructed official-newspaper-day panel. The number of observations varies due to data limitations.

Table A2: Officials with no articles and only articles pre- or post-scandal

Dummy for officials with	no articles	only articles post-scandal	only articles pre-scandal
Govt. official	0.015	0.0070	0.017
	(0.019)	(0.059)	(0.023)
High-level official	-0.020*	-0.17***	-0.025*
	(0.012)	(0.046)	(0.014)
Shanghai	-0.012	0.36**	-0.014
	(0.016)	(0.16)	(0.019)
Beijing	-0.013	0.44**	-0.015
	(0.017)	(0.18)	(0.020)
Sichuan	-0.011	0.091	-0.012
	(0.015)	(0.065)	(0.017)
Tianjin	-0.0042	0.53***	-0.0045
	(0.0088)	(0.20)	(0.010)
Shandong	0.062	0.54***	0.061
	(0.054)	(0.10)	(0.055)
Guangdong	-0.012	0.25***	-0.014
	(0.016)	(0.073)	(0.019)
Guangxi	-0.012	0.32**	0.049
	(0.016)	(0.13)	(0.063)
Jiangxi	0.026	0.13	0.025
	(0.041)	(0.082)	(0.043)
Zhejiang	-0.012	0.16	0.069
	(0.016)	(0.12)	(0.081)
Hubei	0.0099	0.28***	0.0081
	(0.029)	(0.089)	(0.031)
Hunan	0.055	0.62***	0.086*
	(0.043)	(0.099)	(0.051)
Gansu	-0.010	0.30**	-0.012
	(0.014)	(0.14)	(0.016)
Fujian	0.17**	0.49***	0.22***
	(0.066)	(0.098)	(0.076)
Chongqing	-0.010	0.36**	-0.012
	(0.016)	(0.14)	(0.019)
No. of obs.	408	408	408
R-squared	0.12	0.38	0.15
Mean dep. var.	0.029	0.30	0.042

There are 12 officials with no articles in the sample, 124 officials with only articles following the corruption scandal and 17 officials with only articles before the corruption scandal. This table shows the probability an official belonging to each of these three cases regressed on official characteristics. Robust standard errors in parentheses. Significance at the 1%, 5% and 10% is denoted by ***, ** and *, respectively.

A.2   Appendix II: Regression tables for figures in main text

Table A3: Heterogeneity by newspaper (NP) type

	Dummy for whether an article is published
Party NP x High-level official x Post-scandal x Same province	-0.11***
	(0.026)
Govt. commercial NP x High-level official x Post-scandal x Same province	-0.056***
	(0.020)
Subsidiary NP x High-level official x Post-scandal x Same province	-0.032***
	(0.011)
Party NP x Low-level official x Post-scandal x Same province	0.0018*
	(0.00094)
Govt. commercial NP x Low-level official x Post-scandal x Same province	0.0016
	(0.0013)
Subsidiary NP x Low-level official x Post-scandal x Same province	0.0028***
	(0.00065)
Party NP x High-level official x Post-scandal	0.0088***
	(0.0014)
Govt. commercial NP x High-level official x Post-scandal	0.0074***
	(0.0021)
Subsidiary NP x High-level official x Post-scandal	0.0099***
	(0.0016)
Party NP x Low-level official x Post-scandal	0.0023***
	(0.00029)
Govt. commercial NP x Low-level official x Post-scandal	0.0021***
	(0.00033)
Subsidiary NP x Low-level official x Post-scandal	0.0024***
	(0.00033)
p(Party NP = Govt. commercial NP) - same province	0.099
p(Party NP = Subsidiary NP) - same province	0.005
p(Govt.commercial NP = Subsidiary NP) - same province	0.205
p(Party NP = Govt. commercial NP) - not same province	0.495
p(Party NP = Subsidiary NP) - not same province	0.465
p(Govt.commercial NP = Subsidiary NP) - not same province	0.264
No. of obs.	39,929,390
R-squared	0.046
Mean dep. var.	0.0012

This table shows results from a regression of a dummy for whether an article is published on interactions between dummies for three different newspaper types (party newspapers, government-owned commercial newspapers and subsidiary newspapers, see equation 3) and dummies for the post-scandal period, whether a newspaper and official are from the same province and whether an official is high-ranking, which are partly shown in Figure 3. P-values for a Wald-test of equality between coefficients on newspaper type for high-level officials are reported at the bottom. The regression includes official-newspaper-pair fixed effects and all controls from column 2 in Table 1. Standard errors in parentheses are clustered by newspaper and official. Significance at the 1%, 5% and 10% is denoted by ***, ** and *, respectively.

Table A4: Heterogeneity by newspaper (NP) administrative level

	Dummy for whether an article is published
Provincial NP x High-level official x Post-scandal x Same province	-0.076*** (0.014)
Sub-provincial NP x High-level official x Post-scandal x Same province	-0.033** (0.014)
Provincial NP x Low-level official x Post-scandal x Same province	0.0056*** (0.00094)
Sub-provincial NP x Low-level official x Post-scandal x Same province	0.0026*** (0.00057)
Provincial NP x High-level official x Post-scandal	0.011*** (0.0018)
Sub-provincial NP x High-level official x Post-scandal	0.0089*** (0.0014)
Provincial NP x Low-level official x Post-scandal	0.0025*** (0.00042)
Sub-provincial NP x Low-level official x Post-scandal	0.0024*** (0.00040)
p(Provincial NP = Sub-provincial NP) - High-level officials, same province	0.006
p(Provincial NP = Sub-provincial NP) - High-level officials, not same province	0.177
p(Provincial NP = Sub-provincial NP) - Low-level officials, same province	0.003
p(Provincial NP = Sub-provincial NP) - Low-level officials, not same province	0.794
No. of obs.	10,813,270
R-squared	0.045
Mean dep. var.	0.0014

This table shows results from a regression of a dummy for whether an article is published on interactions between dummies for either provincial-level or sub-provincial level newspapers and dummies for the post-scandal period, whether a newspaper and official are from the same province and whether an official is high-ranking, which are partly shown in Figure 4. The sample is restricted to provincial government officials and local newspapers. P-values for a Wald-test of equality between the coefficients on provincial and sub-provincial newspapers are reported at the bottom of the table. The regression includes official-newspaper-pair fixed effects and all controls from column 2 in Table 1. Standard errors in parentheses are clustered by newspaper and official. Significance at the 1%, 5% and 10% is denoted by ***, ** and *, respectively.

Table A5: Municipality-level results

	Dummy for whether an article is published
Post-scandal x Same province	0.0017***
	(0.00039)
Post-scandal x Same city	-0.0077**
	(0.0038)
Post-scandal	0.0019***
	(0.00032)
p(Same province = Same city)	0.023
No. of obs.	6,647,982
R-squared	0.074
Mean dep. var.	0.0011

This table shows results from a regression of a dummy for whether an article is published on an indicator for the post-scandal period ($P_{o,t}$) and its interactions with two dummy variables that takes the value of 1 if an official and a newspaper are from the same city ($P_{o,t} \times SC_{o,n}$) and if an official and a newspaper are from different cities of the same province ($P_{o,t} \times DC_{o,n}$), which are partly shown in Figure 5. The sample is restricted to city-level government officials and newspapers. The p-value of a Wald-test for equality between the coefficients on same city and same province is shown at the bottom of the table. The regression includes official-newspaper-pair fixed effects and all controls from column 2 in Table 1. Standard errors in parentheses are clustered by newspaper and official. Significance at the 1%, 5% and 10% is denoted by ***, ** and *, respectively.

Table A6: Proximity to the incumbent party secretary

	Dummy for whether an article is published
Served x High-level official x Post-scandal x Same province	-0.11***
	(0.021)
Not served x High-level official x Post-scandal x Same province	-0.041***
	(0.0066)
Served x High-level official x Post-scandal	0.010***
	(0.0023)
Not served x High-level official x Post-scandal	0.0088***
	(0.0014)
Low-level official x Post-scandal x Same province	0.0023***
	(0.00057)
Low-level official x Post-scandal	0.0023***
	(0.00030)
p(Not served = Served) - Same province	0.000
p(Not served = Served) - Not same province	0.504
No. of obs.	39,929,390
R-squared	0.046
Mean dep. var.	0.0012

This table shows results from a regression of a dummy for whether an article is published on the interaction between a dummy variable for whether a high-level official o has served on the provincial party standing committee with the incumbent party secretary of the province (served vs. not served) and dummies for the post-scandal period, whether a newspaper and official are from the same province and whether an official is high-ranking, which are partly shown in Figure 6. P-values at the bottom of the table are from Wald-tests for the equality of coefficients on high-level officials that served with the party secretary and those who did not. The regression includes official-newspaper-pair fixed effects and all controls from column 2 in Table 1. Standard errors in parentheses are clustered by newspaper and official. Significance at the 1%, 5% and 10% is denoted by ***, ** and *, respectively.

Table A7: Officials connected to Wan Qingliang

	Dummy for whether an article is published
Wan OF x Guangzhou NP x pre Wan-investigation x Post-scandal	-0.0058
	(0.0062)
Wan OF x Guangdong NP x pre Wan-investigation x Post-scandal	0.0046**
	(0.0021)
other OF x Guangzhou NP x pre Wan-investigation x Post-scandal	0.0074**
	(0.0033)
other OF x Guangdong NP x pre Wan-investigation x Post-scandal	0.0091***
	(0.0022)
Wan OF x Guangzhou NP x post Wan-investigation x Post-scandal	0.0067*
	(0.0035)
Wan OF x Guangdong NP x post Wan-investigation x Post-scandal	0.011***
	(0.0031)
other OF x Guangzhou NP x post Wan-investigation x Post-scandal	0.0056**
	(0.0024)
other OF x Guangdong NP x post Wan-investigation x Post-scandal	0.0052***
	(0.0014)
p(Wan OF x Guangzhou NP = Wan OF x other NPs) - pre Wan-scandal	0.068
p(Wan OF x Guangzhou NP = other OF x Guangzhou NP) - pre Wan-scandal	0.122
p(Wan OF x Guangzhou NP = Wan OF x other NPs) - post Wan-scandal	0.404
p(Wan OF x Guangzhou NP = other OF x Guangzhou NP) - post Wan-scandal	0.752
No. of obs.	1,191,804
R-squared	0.034
Mean dep. var.	0.0043

This table shows how newspapers from Guangzhou and other Guangdong province newspapers report about officials connected to Wan Qingliang and other Guangdong officials of the same ranks, both before and after Wan Qingliang's investigation announcement, as shown in Figure 7. P-values at the bottom of the table are from Wald-tests for equality of different set of coefficients. All regressions include official-newspaper-pair fixed effects and all controls from column 2 in Table 1. Standard errors in parentheses are clustered by newspaper and official. Significance at the 1%, 5% and 10% is denoted by ***, ** and *, respectively

Table A8: Coefficient estimates of dynamic model

Month	High-level official x Same province	Low-level official x Same province	High-level official	Low-level official
-23	0.038***	-0.00058	0.0000011	-0.000014
	(0.011)	(0.00058)	(0.00040)	(0.000043)
-22	0.00069	-0.0012*	0.0035	-0.000012
	(0.012)	(0.00067)	(0.0028)	(0.000062)
-21	0.0050	-0.00094	0.0026	0.000064
	(0.014)	(0.00070)	(0.0024)	(0.00012)
-20	0.018	-0.0011	-0.00034	-0.000069
	(0.014)	(0.00079)	(0.0013)	(0.000087)
-19	-0.00041	-0.0012*	-0.00091	-0.00010
	(0.015)	(0.00070)	(0.0012)	(0.000093)
-18	0.0087	-0.0013	0.00014	-0.00014
	(0.011)	(0.00093)	(0.0011)	(0.00010)
-17	-0.0043	-0.0016*	-0.00028	-0.00014
	(0.022)	(0.00081)	(0.0013)	(0.00012)
-16	-0.0033	-0.0015*	0.0027	-0.00012
	(0.016)	(0.00086)	(0.0025)	(0.00014)
-15	0.00084	-0.0018*	-0.0015	-0.00019
	(0.019)	(0.00092)	(0.0010)	(0.00015)
-14	0.0063	-0.0015	-0.00059	-0.00020
	(0.015)	(0.00094)	(0.0012)	(0.00016)
-13	-0.0087	-0.0019*	-0.0012	-0.00020
	(0.013)	(0.0010)	(0.0012)	(0.00017)
-12	-0.0011	-0.0019*	-0.00048	-0.00011
	(0.010)	(0.0011)	(0.0011)	(0.00023)
-11	-0.0030	-0.0015	0.00019	-0.000069
	(0.023)	(0.0010)	(0.0015)	(0.00025)
-10	-0.023*	-0.0018*	0.0029	-0.000053
	(0.012)	(0.00098)	(0.0022)	(0.00021)
-9	-0.022	-0.0017*	-0.00011	-0.000054
	(0.016)	(0.0010)	(0.0013)	(0.00023)
-8	-0.017	-0.0013	-0.0016	0.000061
	(0.012)	(0.00092)	(0.0011)	(0.00023)
-7	-0.018**	-0.0017*	-0.00068	0.000037
	(0.0082)	(0.00090)	(0.0012)	(0.00022)
-6	-0.0028	-0.0016	-0.00099	0.00028
	(0.013)	(0.00097)	(0.0012)	(0.00031)
-5	-0.016	-0.0016*	-0.0012	0.00015
	(0.013)	(0.00093)	(0.0011)	(0.00025)
-4	-0.021	-0.0013	-0.000086	0.00040
	(0.016)	(0.00083)	(0.0016)	(0.00030)
-3	-0.019	-0.0015*	0.0015	0.00019
	(0.015)	(0.00077)	(0.0024)	(0.00025)
-2	-0.021*	-0.0014	-0.00017	0.00023
	(0.011)	(0.00083)	(0.0014)	(0.00027)
-1	-0.0090	-0.0011	-0.00041	0.00068*
	(0.012)	(0.00098)	(0.0012)	(0.00034)
0	-0.070***	0.012***	0.031***	0.0062***
	(0.021)	(0.0022)	(0.0038)	(0.00092)
1	-0.059**	0.00087	0.0087***	0.0010**
	(0.024)	(0.0011)	(0.0022)	(0.00040)
2	-0.064***	0.00012	0.010***	0.00081**
	(0.020)	(0.0011)	(0.0031)	(0.00032)
3	-0.055***	0.00056	0.0060***	0.00100***
	(0.018)	(0.0011)	(0.0021)	(0.00035)
4	-0.065***	0.0010	0.0072***	0.00090***
	(0.020)	(0.0012)	(0.0021)	(0.00034)
5	-0.061***	0.00048	0.0085**	0.00069**
	(0.018)	(0.0011)	(0.0035)	(0.00031)
6	-0.065***	0.00087	0.0087***	0.00062*
	(0.020)	(0.0011)	(0.0024)	(0.00033)

Table A8: Coefficient estimates of dynamic model

Month	High-level official x Same province	Low-level official x Same province	High-level official	Low-level official
7	-0.060***	-0.0011	0.0043***	0.00061*
	(0.019)	(0.0011)	(0.0014)	(0.00032)
8	-0.064***	-0.0011	0.0056***	0.00060*
	(0.020)	(0.00097)	(0.0015)	(0.00033)
9	-0.059***	-0.0011	0.0026	0.00052
	(0.021)	(0.0010)	(0.0016)	(0.00032)
10	-0.060***	-0.0015	0.0017	0.00059*
	(0.020)	(0.00100)	(0.0013)	(0.00035)
11	-0.057**	-0.00065	0.0019	0.00053
	(0.023)	(0.0010)	(0.0014)	(0.00034)
12	-0.062***	-0.00089	0.0015	0.00079**
	(0.021)	(0.0012)	(0.0011)	(0.00039)

This table shows results from a regression of a dummy for whether an article is published on a vector of month-since-announcement dummies interacted with dummies for whether an official is high-ranking and whether a newspaper and official are from the same province, see equation 6 and Figure 8. Month 24 before the announcement is the omitted category. The regression includes official-newspaper-pair fixed effects and all seasonal controls from column 2 in Table 1 except for month fixed effects. Standard errors in parentheses are clustered by newspaper and official. Significance at the 1%, 5% and 10% is denoted by ***, ** and *, respectively. The number of observations is 39,546,638, the R-squared of the regression is 0.047 and the mean dependent variable is 0.0012.

A.3 Appendix III: Additional robustness checks

Table A9: Poisson and OLS comparison

	Poisson: Number of articles published		OLS: Dummy for whether an article is published	
	(1)	(2)	(3)	(4)
Post-scandal x Same province	-1.54***		-0.0041**	
	(0.11)		(0.0016)	
Post-scandal	2.58***		0.0097***	
	(0.065)		(0.00090)	
High-level official x				
Post-scandal x Same province		-3.47***		-0.064***
		(0.18)		(0.013)
High-level official x Post-scandal		2.48***		0.014***
		(0.068)		(0.0017)
Low-level official x				
Post-scandal x Same province		-0.82***		0.0012
		(0.078)		(0.00075)
Low-level official x Post-scandal		2.75***		0.0082***
		(0.070)		(0.00068)
No. of obs.	11,820,620	11,820,620	11,820,620	11,820,620
Log likelihood				
or R-squared	-253,973	-251,377	0.042	0.044
Mean dep. var.	0.0045	0.0045	0.0040	0.0040

The first two columns of this table show results from regressions of number of articles published on interactions between dummies for the post-scandal period, whether a newspaper and official are from the same province, whether an official is high-ranking and different controls, see equations 1 and 2 and Section 5, using the Poisson QML estimator. Standard errors are cluster bootstrapped by newspaper and official. The last two columns show the OLS results (using a dummy for whether an article was published as dependent variable) for the same restricted sample. All regressions include official-newspaper-pair fixed effects and all seasonal controls as in Table 1. Standard errors in parentheses are clustered by newspaper and official. Significance at the 1%, 5% and 10% is denoted by ***, ** and *, respectively.

Table A10: Underreporting on corruption scandals - Restricted sample of officials with no pre-announcement "corruption" articles from same-province newspapers

	Dummy for whether an article is published	
	(1)	(2)
Post-scandal x Same province	0.0018***	
	(0.00066)	
Post-scandal	0.0028***	
	(0.00040)	
High-level official x Post-scandal x Same province		-0.039***
		(0.011)
High-level official x Post-scandal		0.0086***
		(0.0015)
Low-level official x Post-scandal x Same province		0.0031***
		(0.00055)
Low-level official x Post-scandal		0.0023***
		(0.00029)
No. of obs.	35,779,913	35,779,913
R-squared	0.017	0.018
Mean dep. var.	0.00089	0.00089

This table shows the estimation results for equations 1 and 2 for a restricted sample, excluding 42 officials with at least one article published in a same-province newspaper before the CCDI investigation, where the headline included references to corruption. All regressions include official-newspaper-pair fixed effects and all seasonal controls as in Table 1. Standard errors in parentheses are clustered by newspaper and official. Significance at the 1%, 5% and 10% is denoted by ***, ** and *, respectively.

Table A11: Underreporting on corruption scandals - ChinaFile alternative definition of high- vs. low-ranking officials

	Dummy for whether an article is published
	(1)
High-level official x Post-scandal x Same province	-0.045***
	(0.012)
High-level official x Post-scandal	0.0085***
	(0.0012)
Low-level official x Post-scandal x Same province	0.0023***
	(0.00060)
Low-level official x Post-scandal	0.0022***
	(0.00029)
No. of obs.	39,929,390
R-squared	0.045
Mean dep. var.	0.0012

This table shows the estimation result for equation 2 using the alternative ChinaFile definition of high- vs. low-ranking officials. All regressions include official-newspaper-pair fixed effects and all seasonal controls as in Table 1. Standard errors in parentheses are clustered by newspaper and official. Significance at the 1%, 5% and 10% is denoted by ***, ** and *, respectively.