HAL
open science

Essays on Media and Government in China

Maiting Zhuang

To cite this version:

Maiting Zhuang. Essays on Media and Government in China. Economics and Finance. École des hautes études en sciences sociales (EHESS), 2020. English. NNT : 2020EHES0136 . tel-03168323

HAL Id: tel-03168323

https://theses.hal.science/tel-03168323

Submitted on 12 Mar 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

École des Hautes Études en Sciences Sociales

École Doctorale N°465 - Économie Panthéon-Sorbonne

Paris School of Economics

Thèse de Doctorat

Discipline : Analyse et politique économiques

MAITING ZHUANG

Essays on Media and Government in China

Thèse dirigée par: Ekaterina Zhuravskaya

Date de soutenance : le 30 novembre 2020

Rapporteurs 1 Matthew Gentzkow, Stanford University
2 David Strömberg, Stockholm University

Jury 1 Matthew Gentzkow, Stanford University
2 David Strömberg, Stockholm University
3 Emeric Henry, Sciences Po
4 Ruixue Jia, University of California San Diego
5 Hillel Rapoport, Paris School of Economics / Université Paris 1 Panthéon-Sorbonne
6 Ekaterina Zhuravskaya, Paris School of Economics / EHESS

In loving memory of my grandfather

Zhang Xiaolin

Acknowledgements

This thesis would not have been possible without the help and support of many people. First and foremost, I want to thank my PhD advisor Ekaterina Zhuravskaya whose knowledge and passion for research continues to be an inspiration to me. Her input has been invaluable from the inception of each project and despite her busy schedule, she has always treated my projects as seriously as her own. Thank you, Katia, for believing in me, pushing me to go further and simply refusing to take no for an answer!

To my wonderful friend and co-author Paul Dutronc-Postel - it has been a pleasure to discover the inner workings of the Chinese (and French) bureaucracy with you!

I want to thank all the members of the jury: Matthew Gentzkow for hosting me at Stanford University, reading and re-reading this thesis and always identifying the key areas of improvement for each paper; David Strömberg for carefully poring over this thesis, giving me detailed feedback and welcoming me to Stockholm; Hillel Rapoport and Emeric Henry for helping and supporting me all these years in my thesis committee until the very end of this journey; Ruixue Jia for lending her time and expertise in

agreeing to read this manuscript.

There are countless researchers whose comments and feedback have shaped this thesis. I want to thank the faculty of the Paris School of Economics' development group, in particular Sylvie Lambert, Karen Macours, François Libois, Oliver Vanden Eynde and Liam Wren-Lewis; the organisers of the many fun and stimulating India-China conferences over the years: Maëlys de la Rupelle, Thomas Vendryes, Clément Imbert and Guilhem Cassan and the organisers of the PEPES seminar series at Sciences Po: Quoc-Anh Do, Ruben Durante, Benjamin Marx. I am grateful to David Magolis for his support on the job market and Véronique Guillotin for her patience and kindness throughout all the years at the PSE. I gratefully acknowledge the funding provided by the PSE's development group, Labex OSE, G-MonD and a French government subsidy managed by the Agence Nationale de la Recherche under the framework of the Investissements d'avenir programme reference ANR-17-EURE-001.

This PhD has allowed me to meet an amazing group of people whom I wish to thank for their friendship: Annali Casanueva Artis for being a wonderful friend whether in Sheffield, Berkeley or Paris; Emanuela Migliaccio and Giulio Iacobelli for the sheer joy which they bring to everyday things; the occupants of office R6-58 à travers les âges: Irène Hu, Alexandra Jarotschkin, Alexia Lochmann, Etienne Madinier; everyone from the PSE's development group: Juliette Crespin-Boucaud, Sarah Deschênes, Victor Pouliquen, Juni Singh, Mattea Stein, Alessandro Tondini.

Finally, I would like to thank my family: my mother Fan Zhang for

listening patiently to all my research ideas and reading too many articles about corrupt officials; my father Rugang Zhuang whose PhD defence is one of my favourite childhood memories; my aunts, uncles and cousins in China for (unwittingly) providing much anecdotal evidence for my research; my grandmother Pan Zhaoxiang for sharing her life's stories with me and my husband Jonathan Lehne who is the first to hear any new idea and often the last to proofread any draft and whose patience, love and silliness pulled me through the hardest parts of this PhD.

Abstract

This thesis consists of three empirical research papers on the political economy of China. The first chapter studies how conflict within an autocratic elite affects media content, while the second chapter shows how media content can in turn influence public opinion. The third chapter analyses the motivation and behaviour of individuals as they rise up the autocratic hierarchy.

Chapter 1 offers an explanation for why media censorship varies within an autocratic country. I study how Chinese newspapers report about officials caught during Xi Jinping's anti-corruption campaign, by collecting close to 40,000 articles in print and the corresponding social media posts and comments. I show that individuals are significantly more likely to search for and comment on news about corrupt officials from their own province. 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 on high-level corruption at home, they deemphasise these stories, by

making them shorter, less negative and less likely to explicitly mention corruption. Similarly, 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 illustrate how intergovernmental conflict within an autocracy can lead to diverging media censorship strategies by different levels of government. I present suggestive evidence that this type of localised censorship can reduce the accountability of local governments.

Chapter 2 investigates whether stereotypes in entertainment media promote negative sentiment against foreigners. Despite close economic ties, anti-Japanese sentiment is high in China. I assemble detailed information on Chinese TV broadcasts during 2012 and document that around 20 percent of all TV shows aired during prime time were historical TV dramas set during the Japanese occupation of China during World War II. To identify the causal effect of media on sentiment, I exploit high-frequency data and exogenous variation in the likelihood of viewing Sino-Japanese war dramas due to channel positions and substitution between similar programmes. I show that exposure to these TV shows lead to a significant increase in anti-Japanese protests and anti-Japanese hate speech on social media across China. These effects are driven by privately rather than state-produced TV shows.

Chapter 3, co-authored with Paul Dutronc-Postel, illustrates how career incentives can affect bureaucrats' policy choices. We collect data on the career histories of the top bureaucrats of all Chinese prefectures between

1996 and 2014 and identify the causal effect of career incentives by exploiting variation in the ex ante competitiveness of promotions. Bureaucrats with a smaller starting cohort have a greater likelihood of promotion. This incentivises them to adopt a strategy that relies on real estate investment and rural land expropriation, resulting in faster growth in construction and GDP. We present suggestive evidence that the same incentives result in lower investment in education, public transport and health. We corroborate our findings using survey and remote sensing data, and show that land expropriations are associated with adverse outcomes for expropriated individuals, with subsequent arrests of local officials, and with the emergence of "ghost cities".

Keywords: Media, Censorship, Newspapers, Corruption, Intergovernmental Conflict, Ethnic Prejudice, Social Media, Protests, Bureaucracy, Personnel Management, Land Expropriation, China

Résumé

Cette thèse se compose de trois articles de recherche empirique sur l'économie politique de la Chine. Le premier chapitre étudie comment le conflit au sein d'une élite autocratique affecte le contenu des médias, tandis que le deuxième chapitre montre comment le contenu des médias peut à son tour influencer l'opinion publique. Le troisième chapitre analyse la motivation et le comportement des individus lorsqu'ils montent dans la hiérarchie autocratique.

Le premier chapitre explique pourquoi la censure des médias varie au sein d'un pays autocratique. J'étudie la façon dont les journaux chinois rendent compte des fonctionnaires arrêtés lors de la campagne anti-corruption de Xi Jinping, en rassemblant près de 40 000 articles imprimés et les posts et commentaires correspondants dans les médias sociaux. Je montre que des individus sont plus enclins à rechercher et à commenter sur des fonctionnaires corrompus de leur propre province. Pourtant, malgré un plus grand intérêt des lecteurs, les journaux locaux sous-rapportent les scandales de corruption impliquant des hauts fonctionnaires de leur propre province. Cette sous-rapportage est plus importante lorsqu'un journal ne dépend pas

des revenus publicitaires et qu'un fonctionnaire corrompu est bien connecté. Lorsque les journaux rapportent sur la corruption dans leur propre province, ils minimisent ces scandales, en les rendant plus courtes, moins négatives et moins susceptibles de mentionner explicitement la corruption. De même, les journaux municipaux rapportent moins sur la corruption dans leur propre ville que dans d'autres villes de la même province, mais sont plus susceptibles de signaler la corruption au sein de leur gouvernement provincial que les journaux provinciaux correspondants. Ces résultats illustrent comment les conflits intergouvernementaux au sein d'une autocratie peuvent conduire à des stratégies de censure des médias divergentes par différents niveaux de gouvernement. Je présente des preuves suggestives que ce type de censure localisée peut réduire la responsabilité et imputabilité des gouvernements locaux.

Le deuxième chapitre examine si les stéréotypes dans les médias de divertissement provoquent un sentiment négatif à l'égard des étrangers. Malgré des liens économiques étroits, le sentiment anti-japonais est élevé en Chine. Je rassemble des informations détaillées sur les émissions de télévision chinoises en 2012 et je documente qu'environ 20 pour cent de toutes les émissions de télévision diffusées aux heures de grande écoute étaient des dramatiques historiques qui se sont déroulées pendant l'occupation japonaise de la Chine au cours de la Seconde Guerre mondiale. Pour identifier l'effet causal des médias sur le sentiment, j'exploite les données à haute fréquence et la variation exogène de la probabilité de regarder des dramatiques de guerre sino-japonaises en raison des positions des chaînes et de la substitution entre des programmes similaires. Je montre que l'exposition à ces émissions de télévision conduit à une augmentation significative

des manifestations antijaponaises et des discours de haine anti-japonais sur les médias sociaux à travers la Chine. Ces effets sont attribuables à des émissions télévisées produites par des enterprises privées plutôt qu'à des émissions produites par l'État.

Le troisième chapitre, co-écrit avec Paul Dutronc-Postel, illustre comment les incitations peuvent affecter les choix politiques des bureaucrates. Nous collectons les historiques de carrière des hauts fonctionnaires de toutes les préfectures chinoises entre 1996 et 2014 et nous identifions l'effet causal des incitations en exploitant la variation ex ante du nombre de concurrents. Les cadres avec une cohorte initiale plus petite ont une plus grande probabilité de promotion. Cela les pousse à adopter une stratégie qui repose sur l'investissement immobilier et l'expropriation des terres rurales, et ce qui se traduit par une croissance plus rapide de la construction et du PIB. Nous présentons des preuves suggestives que les mêmes incitations entraînent une baisse des investissements dans l'éducation, les transports publics et la santé. Nous corroborons nos résultats en utilisant des données administratives, des données satellites, et des données d'enquête. Nous montrons que les expropriations de terres sont associées à des résultats négatifs pour les personnes expropriées, à des arrestations ultérieures de fonctionnaires locaux et à l'émergence de "villes fantômes".

Mots clés : Médias, Censure, Journaux, Corruption, Conflit Intergouvernemental, Préjugés Ethniques, Médias Sociaux, Manifestations, Bureaucratie, Gestion du Personnel, Expropriation de Terres, Chine

Introduction

Recent events around the world seem to contradict Francis Fukuyama's optimistic prediction for the "the universalisation of Western liberal democracy".[1] While rising populism and nationalism threaten democracies from within, non-democracies have become increasingly powerful and influential on the world stage. As freedom of the press is being eroded, more and more extreme views are being broadcast on social media.

Our understanding of the political processes underpinning many non-democracies is often limited. China, the world's largest non-democracy, operates a sophisticated system of censorship and propaganda in order to influence the opinions of billions of people. Moreover, it has been exporting its technology and know-how in these areas to other countries. These and other policies are implemented by China's vast bureaucracy which has been credited with generating rapid economic growth at potentially high social cost. This thesis relies on large-scale original data on various aspects of traditional and social media in China, as well as detailed information on

[1] In his 1989 essay, Francis Fukuyama stated that "we may be witnessing ... the end of history as such: that is, the end point of mankind's ideological evolution and the universalisation of Western liberal democracy as the final form of human government" (see Fukuyama, 1989).

Chinese officials' background and policies to shed light on different aspects of the political economy of China that have important lessons for other countries.

The majority of the world population lives in countries with some form of media censorship. The first chapter of this PhD re-examines government influence on the media and shows that censorship within an autocracy need not be uniform, but may be the outcome of strategic interactions between different branches of government. Since coming to power, Chinese President Xi Jinping has conducted a large-scale anti-corruption campaign which has punished millions of officials to date. While the campaign is covered extensively in the central state media, local officials might want to selectively censor news about corruption within their own region.

I collect around 40,000 articles from local Chinese newspapers to study how they report about individual officials who are caught during the campaign. Local newspapers remain an important source of local information in China, in part due to their strong online presence. Even though internet users have easy access to news sources from other parts of China, they still exhibit a "home bias" in preferring to engage with their local newspapers online.

Using internet-search and social-media data, I first document that people are significantly more interested in corruption scandals in their home province. The search volume on Baidu (China's most popular search engine) for an official under investigation is on average more than six times higher in that official's home province compared to other provinces. I also

collect local newspapers' posts on Sina Weibo (China's most popular microblogging site, similar to Twitter) about the same officials alongside close to 30,000 user comments on these posts. Social media users are also much more likely to comment on a post about a corrupt official from their home province.

Despite greater interest in local corruption stories, newspapers are less likely to report on corruption at home. Distinguishing between the incentives of three different levels of government (central, provincial and municipal), I find evidence that newspapers report in line with the interests of the government at their level, against the wishes of higher-level governments and their local readership. While local newspapers write more articles about *low-level* officials from their own province, they write fewer articles about investigations into *high-level* officials from their own province compared to officials from other provinces. I use text analysis to show that newspapers not only underreport corruption in their home province, but they also de-emphasise local corruption stories. Compared to articles about corruption in other provinces, articles about high-level corrupt officials from a newspaper's home province are shorter, less negative in tone and make fewer explicit mentions of corruption.

Newspapers that face more competition, rely more on advertising revenue and are not owned by local governments underreport less. Using information on individual officials' CVs and a case study, I show that this type of selective underreporting appears designed to protect other officials still in power.

The results of this chapter potentially have wider implications. Corruption among politicians and government officials is a widespread problem in many countries. In a non-democracy, without the threat of elections, local officials can often only be held accountable by higher levels of government. The Chinese anti-corruption campaign relies on tip-offs and complaints by the population to identify suspected corruption cases. Using internet search data, I show that by underreporting and deemphasising local corruption scandals, citizens could be discouraged from complaining about their local government officials.

The internet and social media were initially hailed as technologies that could promote freedom, but they can also be used as tools to promote hate crime. In the second chapter of this thesis, I investigate how stereotypes in entertainment media can fuel animosity against foreigners both online and offline. While there are many examples in history of state-sponsored propaganda leading to violence against specific groups, harmful consequences of biases in entertainment media can have even broader implications. The first reason is entertainment media's ubiquity and the second is a potential for vicious cycles: stereotypes in entertainment media often reflect underlying biases in the population, but seeing those reflected back on television or in cinemas could strengthen existing views.

I show that exposure to historical television dramas set during the Sino-Japanese war increases anti-Japanese protests and hate speech on social media in China today. Anti-Japanese sentiment in China has its roots in the Japanese occupation of China during World War II and is widespread even among younger generations. While China and Japan have strong

economic ties today, this resentment continues to adversely affect economic exchange. In 2012, territorial disputes over the Senkaku/Diaoyu islands in the East China Sea led to large-scale anti-Japanese protests in China.

I collect detailed data on the programming schedule of all major Chinese TV channels in 2012 and data on TV soap content and producers from official approval forms for TV show production. Around one fifth of all TV soaps aired in 2012 were dramas set during the Sino-Japanese war with highly negative depictions of Japanese soldiers. The share of distribution licenses awarded to this type of show (as a fraction of all domestically produced TV shows) increased by approximately eight fold from 2004 to 2011 and this increase mirrors trends in anti-Japanese sentiment in China according to survey data.

These historical television dramas are primarily intended to be a source of entertainment rather than an instrument for propaganda.. The majority of dramas are produced by private companies rather than the government. I show that there is no relationship between the propensity of a provincial television station to broadcast Sino-Japanese war dramas with that province's experience during the war or ties to Japan. I rule out strategic scheduling of historical TV dramas in response to unexpected increases in anti-Japanese sentiment, by using high-frequency data and accounting for Sino-Japanese war anniversaries and the intensity of the Senkaku/Diaoyu island conflict. The default channel position and substitution between TV programmes provide variation in the likelihood of historical TV drama viewership that is orthogonal to prior anti-Japanese sentiment.

Greater exposure to TV dramas set during the Sino-Japanese war increases the likelihood of anti-Japanese protests in China. Areas that were occupied and suffered more civilian casualties during the Sino-Japanese war respond more when being exposed to these TV dramas. I show that higher predicted viewership of Sino-Japanese war TV dramas also significantly increases anti-Japanese hate speech on social media. This increase is not purely driven by users directly discussing these TV shows. Exposure to these TV shows leads to an increase in nationalist sentiment expressed on social media, more discussion of the Senkaku/Diaoyu island conflict and calls for boycott of Japanese goods. The effect of TV shows on anti-Japanese and nationalist sentiment is driven by users writing new posts rather than reposting or forwarding content by other users. For both protests and social media posts, the effect is driven by privately rather than state produced TV shows.

While some aspects of these results are particular to the media environment and the history between China and Japan, one can draw a number of parallels to other contexts. Anti-Japanese protests in China are not isolated incidents. Displays of nationalism and anti-foreigner sentiment offline and online have become frequent around the world and underscore the importance of understanding the causes and propagation mechanisms of nationalism and prejudice. Activists, politicians and industry leaders have intuitively understood the power of entertainment media and the potential harmful consequences of racist and culturally inappropriate content. Responses have, for instance, included calls for more diversity in the industry, removing content and applying warning labels.

Power in China is concentrated in the hands of the Chinese Communist Party whose members are placed in all important sectors of society. For a better understanding of how policies are decided and carried out, it is crucial to understand who becomes a member of the ruling elite and what motivates them. The third chapter of this thesis, co-authored with Paul Dutronc-Postel, takes a closer look at the incentives facing prefecture party secretaries in China. These officials have great power over local developments and are at a critical stage in their career progression.

We compile a dataset of the complete career history for the heads of all of China's 334 prefectures from 1996 to 2014 using administrative and internet data. Identifying the causal effect of promotion incentives on policy choices is subject to many potential endogeneity concerns. A bureaucrat's unobserved personal characteristics may jointly determine his performance and his ability to advance in the hierarchy. Here we use variation in an individual bureaucrat's competitive environment as exogenous shocks to promotion incentives.

We find that the size of a prefecture party secretary's starting cohort (that is, the number of other prefecture party secretaries who start their term at the same time in the same province) leads to variation in the competitive pressure that an official faces, but is not correlated with the characteristics of the bureaucrat or his assigned prefecture. The effect of having more competitors on performance is ex ante ambiguous. While increased competition could incentivise bureaucrats to exert more effort, it could also have the opposite effect of discouraging effort. We develop a simple theoretical model to show that the Chinese promotion system

generates incentives similar to a contest between varying number of players for a fixed prize. In this scenario, a smaller starting cohort increases the likelihood of promotion and so encourages bureaucrats to exert more effort.

Our results show that career incentives push bureaucrats to expropriate more rural land and encourage construction and real estate investment, resulting in higher GDP growth. As these outcomes are potentially political sensitive and subject to manipulation, we corroborate our findings using survey, administrative and satellite data. While performance-based incentives may encourage faster growth, this appears to come at a cost of lower public goods provision which are less visible in the evaluation of performance. We further document the cost of land expropriations: individuals who were expropriated have worse outcomes later in life; cities where expropriations took place are more likely to be classed as so-called "ghost cities" and officials who undertook more expropriations are more likely to be arrested during the subsequent anti-corruption campaign.

These results could be seen as a cautionary tale for potential civil service reforms. While China's bureaucratic promotion system does result in higher economic growth rates, these might not be to the benefit of the local population that is unable to hold the bureaucrat accountable.

Introduction en Français

Les événements récents dans le monde semblent contredire la prédiction optimiste de Francis Fukuyama pour la "l'universalisation de la démocratie libérale occidentale". [2] Pendant que la montée du populisme et du nationalisme menace les démocraties de l'intérieur, les non-démocraties sont devenues de plus en plus puissantes et influentes sur la scène mondiale. Alors que la liberté de la presse s'érode, de plus en plus d'opinions extrêmes sont diffusées sur les médias sociaux.

Notre compréhension des processus politiques qui sous-tendent de nombreuses non-démocraties est souvent limitée. La Chine, la plus grande non-démocratie du monde, applique un système sophistiqué de censure et de propagande afin d'influencer les opinions de milliards de personnes. De plus, elle exporte sa technologie et son savoir-faire dans ces domaines vers d'autres pays. Ces politiques et d'autres encore sont mises en œuvre par une vaste bureaucratie chinoise qui a été reconnue pour avoir générer une croissance économique rapide à un coût social potentiellement élevé. Cette

[2] Dans son essai de 1989, Francis Fukuyama a déclaré que "nous assistons peut-être à ... la fin de l'histoire en tant que telle : c'est-à-dire le point final de l'évolution idéologique de l'humanité et l'universalisation de la démocratie libérale occidentale come forme finale de gouvernement humain" (voir Fukuyama, 1989).

thèse s'appuie sur des données originales à grande échelle sur divers aspects des médias traditionnels et sociaux en Chine, ainsi que sur des informations détaillées sur les antécédents et les politiques des fonctionnaires chinois afin de mettre en lumière différents aspects de l'économie politique de la Chine qui ont des leçons importants pour d'autres pays.

La majorité de la population mondiale vit dans des pays qui pratiquent une forme de censure médiatique. Le premier chapitre de ce doctorat réexamine l'influence du gouvernement sur les médias et montre que la censure au sein d'une autocratie n'est pas nécessairement uniforme, mais peut être le résultat d'interactions stratégiques entre différentes branches du gouvernement. Depuis son arrivée au pouvoir, le président chinois Xi Jinping a mené une vaste campagne de lutte contre la corruption qui a puni des millions de fonctionnaires à ce jour. Bien que la campagne soit largement couverte dans les médias de l'État central, les fonctionnaires locaux pourraient vouloir censurer de manière sélective les informations sur la corruption dans leur propre région.

Je collecte environ 40,000 articles de journaux chinois locaux pour étudier la façon dont ils rendent compte des fonctionnaires qui sont arrêtés pendant la campagne. Les journaux locaux restent une source importante d'informations locales en Chine, en partie en raison de leur forte présence en ligne. Même si les internautes ont facilement accès à des sources d'information provenant d'autres régions de Chine, ils préfèrent toujours s'adresser à leurs journaux locaux en ligne.

À l'aide de données sur des cherches sur Internet et de données sur les

réseaux sociaux, je montre d'abord que les gens s'intéressent beaucoup plus aux scandales de corruption dans leur province. Le volume de recherche sur Baidu (le moteur de recherche le plus populaire de Chine) pour un fonctionnaire sous enquête est en moyenne plus de six fois plus élevé dans la province d'origine de ce fonctionnaire que dans les autres provinces. Je recueille également des publications de journaux locaux sur Sina Weibo (le site de micro-blogging le plus populaire de Chine, similaire à Twitter) sur les mêmes fonctionnaires, ainsi que près de 30,000 commentaires d'utilisateurs sur ces publications. Les utilisateurs de médias sociaux sont également beaucoup plus susceptibles de commenter un publication concernant un fonctionnaire corrompu de leur propre province.

Malgré un plus grand intérêt pour les histoires de corruption locale, les journaux sont moins enclins à faire des reportages sur la corruption chez eux. Distinguant les motivations des trois différents niveaux de gouvernement (central, provincial et municipal), je trouve des preuves que les journaux rapportent en accord avec les intérêts du gouvernement à leur niveau, contre la volonté des gouvernements de niveau supérieur et de leur lectorat local. Alors que les journaux locaux écrivent plus d'articles sur les fonctionnaires *de bas niveau* de leur propre province, ils écrivent moins d'articles sur les enquêtes concernant les fonctionnaires *de haut niveau* de leur propre province par rapport aux fonctionnaires des autres provinces. J'analyse les textes des articles pour montrer que les journaux non seulement sous-rapport la corruption dans leur propre province, mais ils minimisent également les histoires de corruption locale. Par rapport aux articles sur la corruption dans d'autres provinces, les articles sur les hauts fonctionnaires corrompus de la province d'origine d'un journal sont plus courts, moins

négatif et font moins de mentions explicites de la corruption.

Les journaux qui font face à une plus grande concurrence, qui dépendent davantage des revenus publicitaires et qui ne sont pas la propriété des gouvernements locaux sous-rapport moins. En utilisant des informations sur les curriculum vitae des fonctionnaires individuels et une étude de cas, je montre que ce type de sous-rapportage sélective semble conçu pour protéger les autres fonctionnaires encore au pouvoir.

Les résultats de ce chapitre ont potentiellement des implications plus larges. La corruption parmi les politiciens et des fonctionnaires est un problème très répandu dans de nombreux pays. Dans une société non démocratique, sans menace d'élections, les fonctionnaires locaux ne peuvent souvent être tenus responsables que par des niveaux supérieurs du gouvernement. La campagne chinoise de lutte contre la corruption s'appuie sur les dénonciations et les plaintes de la population pour identifier les cas de corruption présumés. En utilisant des données sur des recherche sur Internet, je montre qu'en sous-rapportant et en minimisant les scandales de corruption locaux, les citoyens pourraient être découragés de se plaindre de leurs fonctionnaires locaux.

Internet et les médias sociaux ont été initialement salués comme des technologies susceptibles de promouvoir la liberté, mais ils peuvent également être utilisés comme outils pour promouvoir les crimes de haine. Dans le deuxième chapitre de cette thèse, j'étudie comment les stéréotypes dans les médias de divertissement peuvent alimenter l'animosité contre les étrangers à la fois en ligne et hors ligne. Il existe de nombreux exemples dans l'his-

toire de propagande soutenue par l'État menant à la violence contre des groupes spécifiques, mais les conséquences néfastes des préjugés dans les médias de divertissement peuvent avoir des implications encore plus larges. La première raison est l'omniprésence des médias de divertissement et la seconde est un potentiel de cycles vicieux : les stéréotypes dans les médias de divertissement reflètent souvent les préjugés sous-jacents de la population, mais le fait de voir ces préjugés reflétés à la télévision ou dans les cinémas pourrait renforcer les opinions existantes.

Je montre que l'exposition aux drames télévisés historiques se déroulant pendant la guerre sino-japonaise augmente les manifestations anti-japonaises et les discours de haine contre des japonais sur les médias sociaux en Chine aujourd'hui. Le sentiment anti-japonais en Chine a ses racines dans l'occupation japonaise de la Chine pendant la Seconde Guerre mondiale et est répandu même parmi les jeunes générations. Bien que la Chine et le Japon ont aujourd'hui des liens économiques solides, ce ressentiment continue d'affecter négativement les échanges économiques. En 2012, les conflits territoriaux concernant les îles Senkaku/Diaoyu dans la mer de Chine orientale ont conduit à des manifestations anti-japonaises à grande échelle en Chine.

Je recueille des données détaillées sur la grille de programmation de toutes les grandes chaînes de télévision chinoises en 2012 et des données sur le contenu et les producteurs de séries télévisées à partir des formulaires d'approbation officiels pour la production d'émissions de télévision. Environ un cinquième de tous les feuilletons télévisés diffusés en 2012 étaient des drames se déroulant pendant la guerre sino-japonaise, avec des représentations très négatives de soldats japonais. La part des licences de

distribution accordées à ce type de séries (en tant que fraction de toutes les séries télévisées produites au niveau national) a été multipliée par environ huit entre 2004 et 2011, et cette augmentation reflète les tendances du sentiment anti-japonais en Chine selon les données d'enquête.

Ces drames télévisés historiques sont principalement destinées à être une source de divertissement plutôt qu'un instrument de propagande. La majorité des drames sont produites par des sociétés privées plutôt que par le gouvernement. Je montre qu'il n'y a aucun rapport entre la propension d'une station de télévision provinciale à diffuser des drames de guerre sino-japonais avec l'expérience de cette province pendant la guerre ou les liens avec le Japon. J'écarte la programmation stratégique de drames télévisés historiques en réponse à une augmentation inattendue du sentiment anti-japonais, en utilisant des données à haute fréquence et en tenant compte des anniversaires de guerre sino-japonais et de l'intensité du conflit des îles Senkaku/Diaoyu. La position par défaut de la chaîne et la substitution entre les programmes télévisés offrent une variation de la probabilité d'audience des drames historique qui est orthogonale au sentiment anti-japonais antérieur.

Une plus grande exposition aux drames télévisés se déroulant pendant la guerre sino-japonaise augmente la probabilité de manifestations anti-japonaises en Chine. Les zones qui ont été occupées et ont subi plus de pertes civiles pendant la guerre sino-japonaise réagissent davantage lorsqu'elles sont exposées à ces series télévisés. Je montre que l'augmentation du nombre de téléspectateurs prévus pour les drames télévisés sur la guerre sino-japonaise augmente également considérablement les discours de haine

antijaponaise sur les médias sociaux. Cette augmentation n'est pas purement motivée par les utilisateurs qui discutent directement ces émissions de télévision. L'exposition à ces émissions de télévision entraîne une augmentation du sentiment nationaliste exprimé sur les médias sociaux, une plus grande discussion sur le conflit des îles Senkaku/Diaoyu et des appels au boycott des produits japonais. L'effet des émissions télévisées sur le sentiment antijaponais et nationaliste est dû au fait que les utilisateurs écrivent de nouveaux messages plutôt que de renvoyer ou de transmettre des publications d'autres utilisateurs. Pour les manifestations et les publications sur les réseaux sociaux, l'effet est provoqué par des émissions de télévision produites par des entreprises privées que par celles produites par l'État.

Même si certains aspects de ces résultats sont propres à l'environnement médiatique et à l'histoire entre la Chine et le Japon, on peut établir un nombre de parallèles avec d'autres contextes. Les manifestations antijaponaises en Chine ne sont pas des incidents isolés. Les manifestations de nationalisme et de sentiment anti-étranger en ligne et hors ligne sont devenues fréquentes dans le monde entier et soulignent l'importance de comprendre les causes et les mécanismes de propagation du nationalisme et des préjugés. Les activistes, les politiciens et les dirigeants d'entreprises ont intuitivement compris le pouvoir des médias de divertissement et les conséquences néfastes potentielles des contenus racistes et culturellement inappropriés. Les réponses ont, par exemple, inclus des appels à plus de diversité dans l'industrie, le retrait de contenus et l'application d'étiquettes d'avertissement.

Le pouvoir en Chine est concentré dans les mains du Parti Communiste

Chinois (PCC) dont les membres sont placés dans tous les secteurs importants de la société. Pour mieux comprendre comment les politiques sont décidées et mises en œuvre, il est essentiel de savoir qui devient membre de l'élite dirigeante et ce qui les motive. Le troisième chapitre de cette thèse, co-écrit avec Paul Dutronc-Postel, examine de plus près les motivations des secrétaires préfectoraux du PCC. Ces fonctionnaires ont un grand pouvoir sur les développements locaux et se trouvent à un stade critique de leur progression de carrière.

Nous compilons un ensemble de données sur l'historique complet des carrières des chefs de toutes les 334 préfectures de Chine de 1996 à 2014 en utilisant des données administratives et d'Internet. L'identification de l'effet causal des incitations à la promotion sur les choix politiques est sujette à de nombreuses préoccupations potentielles d'endogénéité. Les caractéristiques personnelles non observées d'un bureaucrate peuvent déterminer conjointement sa performance et sa capacité à progresser dans la hiérarchie. Ici, nous utilisons la variation dans l'environnement concurrentiel d'un bureaucrate individuel comme un chocs exogène pour identifies l'effet des incitations à la promotion.

Nous constatons que la taille de la cohorte initiale d'un secrétaire du parti préfectoral du PCC (c'est-à-dire le nombre d'autres secrétaires du PCC qui commencent leur mandat au même moment en autre préfectures dans la même province) entraîne une variation de la pression concurrentielle à laquelle un fonctionnaire est confronté, mais n'est pas corrélée avec les caractéristiques du bureaucrate ou de sa préfecture assignée. L'effet d'un plus grand nombre de concurrents sur les performances est ambigu ex ante.

Si une concurrence accrue pourrait inciter les bureaucrates à faire plus d'efforts, elle pourrait également avoir l'effet inverse de décourager les efforts. Nous développons un modèle théorique simple pour montrer que le système de promotion du PCC génère des incitations similaires à un concours entre un nombre variable de joueurs pour un prix fixe. Dans ce scénario, une cohorte initiale plus petite augmente la probabilité de promotion et encourage donc les bureaucrates à faire plus d'efforts.

Nos résultats montrent que les incitations à la carrière poussent les bureaucrates à exproprier davantage de terres rurales et encouragent la construction et l'investissement immobilier, ce qui entraîne une croissance plus élevée du PIB. Étant donné que ces résultats sont potentiellement politiquement sensibles et sujets à manipulation, nous corroborons nos résultats en utilisant des données d'enquête, administratives et satellitaires. Si les incitations fondées sur la performance peuvent favoriser une croissance plus rapide, cela semble se faire au prix d'une moindre fourniture de biens publics qui sont moins visibles dans l'évaluation des performances. Nous documentons en outre le coût des expropriations de terres : les personnes expropriées ont de pires résultats plus tard dans leur vie ; les villes où des expropriations ont eu lieu sont plus susceptibles d'être classées comme des "villes fantômes" et les fonctionnaires qui ont entrepris davantage d'expropriations sont plus susceptibles d'être arrêtés au cours de la campagne de lutte contre la corruption suivante.

Ces résultats pourraient être considérés comme un avertissement pour d'éventuelles réformes de la fonction publique. Bien que le système de promotion bureaucratique de la Chine entraîne effectivement des taux de crois-

sance économique plus élevés, ceux-ci pourraient ne pas profiter à la population locale qui n'est pas en mesure de tenir le bureaucrate responsable.

Contents

Acknowledgements **iii**

Abstract **vi**

Résumé **ix**

Introduction **xii**

Introduction en Français **xx**

1 Intergovernmental Conflict and Censorship: Evidence from China's Anti-Corruption Campaign **1**

1.1 Introduction 3

1.2 Background 11

1.2.1 Types of Newspapers 11

1.2.2 Censorship of Newspapers 13

1.2.3 Anti-Corruption Campaign 14

1.3 Data and Descriptive Statistics 17

1.3.1 Newspapers 17

1.3.2 Officials under Investigation 18

1.3.3 Articles 18

1.4 Empirical Strategy 20

1.4.1 Are Local Corruption Scandals Selectively Underreported? 21
1.4.2 How are Local Corruption Scandals Reported? . . . 24
1.5 Results 25
1.5.1 Underreporting of Corruption Scandals: "Tigers" versus "Flies" 25
1.5.2 Deemphasising Local Corruption Scandals 28
1.6 Mechanisms 30
1.6.1 Readers' Interest in Local Corruption 31
1.6.2 Effect of Underreporting on Local Government Accountability 33
1.6.3 Conditions for Newspaper Capture 35
1.6.4 Importance of Local Political Connections 39
1.6.5 Interpretation of Results 41
1.7 Robustness 43
1.7.1 Time-Profile of Estimated Coefficients 43
1.7.2 Flexible official- and newspaper-province specification 43
1.7.3 Functional form specification 44
1.7.4 Investigative Reporting 45
1.7.5 Selection of officials and newspapers 45
1.8 Conclusion 47
1.9 Figures 48
1.10 Tables 56

Appendices **62**
1.A Detailed Sample Description 62
1.B Descriptive Text Analysis of Newspaper Articles . . 68

1.B.1 Sentiment Analysis 68
1.B.2 Content Analysis 69
1.C Sina Weibo 70
1.D Additional Figures and Tables 72
1.E Additional Robustness Check Tables 80
1.F Regression Tables for Figures in Main Text 83

2 Anti-Japanese Protests, Social Media Hate Speech and Television Shows in China 90

2.1 Introduction 92
2.2 Background 97
2.2.1 Relationship between China and Japan 97
2.2.2 Media environment in China 102
2.3 Data 105
2.3.1 TV data 105
2.3.2 Social media data 106
2.3.3 Anti-Japanese protests 108
2.3.4 Other data 108
2.4 Empirical Strategy 109
2.4.1 Endogenous scheduling 110
2.4.2 Selection of TV audiences 111
2.5 Results 115
2.5.1 Anti-Japanese Protests 115
2.5.2 Anti-Japanese Hate Speech on Social Media 117
2.6 Robustness 120
2.6.1 Leads and Lags of Historical TV Drama Exposure . 120
2.6.2 Excluding TV Shows By Own-Province TV Channel 122

2.7 Conclusion 122
2.8 Figures 124
2.9 Tables 130

Appendices **139**
2.A Social Media 139
2.B Additional Descriptive Statistics 143
2.C Additional Results and Robustness Tables 146

3 Economic performance, land expropriation and bureaucrat promotion in China **153**
3.1 Introduction 155
3.2 Background 160
3.2.1 Chinese bureaucratic system 160
3.2.2 Land markets in China 162
3.3 Theoretical Framework 163
3.4 Data and descriptive statistics 164
3.4.1 Prefecture CCP secretaries 164
3.4.2 Land expropriations 166
3.4.3 Macroeconomic data 167
3.4.4 Remote sensing data 168
3.5 Empirical strategy 169
3.5.1 Effect of competition on promotion 169
3.5.2 Effect of competition on policy choices 171
3.6 Results 171
3.6.1 Promotions and the number of competitors 171
3.6.2 Economic growth and number of competitors 173

3.6.3 Construction-led growth 174
3.6.4 Consequences of land expropriations 175
3.6.5 Public goods provision 177
3.7 Robustness checks 178
3.7.1 Explaining variation in starting cohort size 178
3.7.2 Identification checks 179
3.7.3 Alternative specifications 180
3.7.4 Data quality concerns 182
3.8 Conclusion 184
3.9 Figures 186
3.10 Tables 189

Appendices **197**
3.A Robustness Tables and Figures 197
3.B Model 213
3.C Detailed sample description 216
3.C.1 Bureaucrats 216
3.C.2 Policy outcomes 219
3.D Measurement of GDP growth using night light data 221

Bibliography **223**

List of Figures **232**

List of Tables **234**

Chapter 1

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

I am deeply grateful to Ekaterina Zhuravskaya and Matthew Gentzkow for their advice and guidance. I also thank Ruben Durante, Emeric Henry, Philipp Ketz, Jonathan Lehne, Jennifer Pan, Hillel Rapoport, Jesse Shapiro, David Strömberg, Oliver Vanden Eynde, Noam Yuchtman and participants at the Paris School of Economics, CEPREMAP India-China Conference, Ronald Coase Institute Workshop on Institutional Analysis, 3rd Economics of Media Bias Workshop, 2018 MIEDC, 2018 GlaD Conference, 2018 EUDN PhD Workshop, 13th NEWEPS, 2019 OxDev, 2019 NEUDC, University of Zurich, UCLA Anderson School of Management, Stockholm Institute of Transition Economics, Uppsala University, University of Amsterdam, University of Warwick, University of Nottingham, ENS Lyon and Johns Hopkins SAIS Europe for valuable comments.

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 the corresponding social media posts and comments. I show that individuals are significantly more likely to search for and comment on news about corrupt officials from their own province. 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 on high-level corruption at home, they deemphasise these stories, by making them shorter, less negative and less likely to explicitly mention corruption. Similarly, 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 illustrate how intergovernmental conflict within an autocracy can lead to diverging media censorship strategies by different levels of government. I present suggestive evidence that this type of localised censorship can reduce the accountability of local governments.

Keywords: Media Censorship, Newspapers, Corruption, Intergovernmental Conflict

JEL Codes: D72, L82, H77, P26

1.1 Introduction

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 unique large-scale dataset of local Chinese newspaper articles, internet searches and social media comments to show that a conflict of interest between different parts of an autocratic government can lead connected media outlets to publish different content.

Corruption among politicians and government officials is a widespread problem in many countries. Without the threat of elections, local government officials in an autocracy can often only be held accountable by higher levels of government. In China, President Xi Jinping has conducted a large-scale anti-corruption campaign which has already caught and punished over one million officials since 2012, including hundreds of high-ranking officials.[2] This campaign relies heavily on tip-offs and complaints by the local population to identify suspected corruption cases and is extensively covered in the *central* state media.[3] Local officials, however, may fear that this campaign could lead to further scrutiny of their own performance and possibly implicate them in corruption scandals. Thus, they have an incentive to contravene central propaganda guidelines and selectively censor news

[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`.

[3] Selected media outlets are reportedly "invited to investigate" targeted officials (Repnikova, 2017).

about corruption within their own region.

In this paper, I study how Chinese newspapers associated with different levels of government report about officials who were investigated during the anti-corruption campaign. Local newspapers remain an important source for local information in China, in part due to their strong online presence. I collect all official announcements of corruption investigations by the central anti-corruption agency from the start of the campaign until the end of 2014 and identify the names of all officials under investigation. Using an online newspaper archive (Wisenews), which contains 99 major Chinese newspapers, I download close to 40,000 unique articles about these corrupt officials, before and after they were investigated. I study whether newspapers publish articles about official corruption and what these articles contain.

Using internet-search and social-media data, I show that people are more interested in corruption scandals in their own province compared to scandals in other provinces. The volume of searches on Baidu (China's most popular search engine) for an official under investigation is on average more than six times higher in that official's home province compared to other provinces. Aside from publishing articles in print, newspapers also post about high-level corruption scandals on their official Sina Weibo accounts (the Chinese equivalent of Twitter). I collect data on these posts, along with close to 30,000 comments on these posts and information about the commenters. Social media users are on average three times more likely to comment on a post about a corrupt official from their home province, relative to an official from another province. The social media data also

highlight the importance of local newspapers in the dissemination of local information both offline and online. Even though social media users have easy access to news sources from other parts of China, their posting behaviour shows a clear preference for news outlets from their own province.

Do local newspapers respond to this higher demand for information about corruption in their own province? In absolute terms, newspapers write more about corruption of *low-level* officials from their own province, but they write around 50 percent fewer articles about investigations into *high-level* officials from their own province compared to officials from other provinces. Taken into account that readers are more interested in corruption in their home province implies an even greater extent of underreporting of local high-level corruption scandals by local newspapers.

My main estimation strategy is a modified difference-in-differences estimator to compare how the likelihood of an article being published about a corrupt official changes after their corruption scandal breaks, depending on whether the official and the newspaper are from the same province. I identify the main effects only using variation over time within an official-newspaper pair and controls for a full set of seasonal factors. I find that local newspapers report less on central government investigations of high-level officials from the same province, relative to similarly ranked officials from other provinces. Following an announcement of a corruption investigation into a high-ranking local official, the probability that a newspaper publishes an article about this official declines by more than 70 percent of its pre-announcement mean.

I use text analysis to show that newspapers not only underreport corruption in their home province, but they also de-emphasise local corruption stories. Compared to articles about corruption in other provinces, articles about high-level corrupt officials from a newspaper's home province are on average 30 percent shorter and 40 percent less likely to mention the role of citizens' complaints in triggering these investigations. These articles are more positive in tone, with twice as many positive relative to negative words than the average article. Summaries of the anti-corruption campaign also appear more likely to omit names of officials from the same province. In addition, the headlines of articles about officials from the same province are designed to appear uninteresting. Relative to an average article about the anti-corruption campaign, articles about local high-level corruption are around one third less likely to include any mentions of corruption and references to Xi Jinping and the central anti-corruption campaign.

I compile a province-day panel of internet searches for citizens' complaint procedures and find that the amount and type of anti-corruption campaign coverage by local newspapers significantly affects these searches in their home province. This suggests that local newspapers' selective underreporting and deemphasising of corruption scandals in their own province could reduce the cost of corruption to local officials by reducing the probability that they are caught. There are two potential channels through which the media can affect accountability of local governments in an autocracy: information (local population learns about the effectiveness of complaints to higher levels of government from media coverage) and signaling (local population views newspaper censorship as a signal about the strength of their local government).

To better understand the relationship between newspapers and local governments, I explore how the extent of selective underreporting varies with newspapers' dependence on local governments and the importance of investigated officials in local networks. Newspapers that face more competition, rely more on advertising revenue and are not owned by local governments underreport less. These findings suggest some constraints to (local) media censorship in an autocracy and are in line with the theoretical predictions of Besley and Prat (2006) regarding conditions that facilitate media capture in a democracy.[4]

Within the same city, city- and provincial-level newspapers compete for the same readership, but city-level newspapers are associated with the municipal rather than the provincial government. Compared to their provincial-level counterparts, city-level newspapers report more about corruption scandals involving high-ranking officials from the provincial government. City-level newspapers report less about corrupt officials from their own municipal government, but more on corrupt officials from other municipalities within the same province. These results suggest that provincial governments can control news coverage of provincial newspapers, contrary to the central government's guidelines, while in turn, sub-provincial governments may act against the provincial government's interest.

Local political connections also affect measured reporting bias. News-

[4]Besley and Prat (2006) develop a model of endogenous media capture in a democratic setting and predict that media plurality, independent ownership and commercialisation of the media reduce media capture. In the context of this paper, media affects accountability not through elections but through higher levels of government. Blanket censorship could also reduce the population's trust in the local media and reduce the effectiveness of future propaganda (see, e.g. Kamenica and Gentzkow, 2011; Gehlbach and Sonin, 2014).

papers are less likely to publish articles about corrupt high-level officials with close ties to the head of their home province, compared to less well connected officials of the same ranks. This difference does not apply when reporting about officials from other provinces. A case study further illustrates how local newspapers underreport instances of corruption to avoid implicating officials still in power. A powerful high-level official and his network of low-level associates were investigated sequentially during the anti-corruption campaign. I compare articles about these connected low-level corrupt officials with articles about comparable but unconnected corrupt officials. Local newspapers underreport corruption scandals involving connected (relative to unconnected) officials before their high-level patron himself was removed from power, but not thereafter.

The heterogeneity of corruption reporting within an official's home province is inconsistent with two alternative hypotheses. In the first one, the central government's media strategy is to selectively censor its own anti-corruption campaign's progress in directly affected areas to maintain the population's trust in their local government. In the second, local governments strategically overreport corruption in neighbouring regions to divert attention from their own problems. Neither of these hypotheses would, for instance, explain the difference in underreporting between provincial and municipal newspapers in a corrupt official's home province, but not in other provinces.

This paper provides evidence that subnational governments in an autocracy use local media to further their own political aims rather than those of the regime's central leadership. Studies of propaganda and censorship in

non-democracies have so far mostly assumed a unified media strategy by the government (see, e.g., Adena et al. (2015); Yanagizawa-Drott (2014) and Chen and Yang (2019); King, Pan and Roberts (2017, 2013, 2014) in the context of Chinese online censorship.)[5] Qin, Strömberg and Wu (2018) construct a measure of bias based on the published contents of 117 Chinese newspapers. This measure is negatively correlated with advertising revenue, highlighting a trade-off between the political and economic goals of a newspaper's owner. In this paper, I use more fine-grained data to test whether the political goals of different parts of the government hierarchy are always aligned. Methodologically, I distinguish between two types of media bias: at the extensive margin, *whether* a newspaper reports local corruption and at the intensive margin, *how* a newspaper reports local corruption.

This paper also contributes to a broader literature on the functioning of bureaucracies and conflict within political hierarchies, see Mookherjee (2015) for a recent review of the literature. While China's relatively well-managed bureaucracy has been credited with contributing to China's fast growth (see Li and Zhou, 2005), local officials have adopted a number of policies counter to the central government's aims (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 local 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

[5]See also Anderson, Waldfogel and Strömberg (2016); Prat and Strömberg (2013) for recent surveys on the political economy of media.

can mitigate principal-agent problems by allowing (partially) free media to report local grievances upwards. Shirk (2011) suggests that the Chinese central leadership uses media reports to monitor the actions of local officials (see also Qin, Strömberg and Wu, 2017). The findings of this paper cast serious doubt on the ability of local newspapers 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.[6] This echoes the results of Pan and Chen (2018) who show that lower level governments in China conceal online complaints from their superiors. I also present suggestive evidence that this type of selective underreporting reduces local governments' accountability towards the central government and the population.

More broadly, this paper adds to our understanding of the factors underlying media bias. As media bias is an equilibrium outcome, it is generally difficult to distinguish between supply-side factors, such as advertising revenue (e.g., Di Tella and Franceschelli, 2011; Beattie et al., 2017) and the preferences of media owners (e.g., Durante and Knight, 2012; Enikolopov, Petrova and Zhuravskaya, 2011) versus demand-side factors, such as readers' ideologies (e.g., Gentzkow and Shapiro, 2010; Puglisi and Snyder Jr, 2011; Gentzkow, Shapiro and Sinkinson, 2014). In this paper, I use data on internet searches and social media comments to present direct estimates of reader demand. Taken into account reader demand, my findings suggest that the measured bias in the content of local Chinese newspapers

[6]Snyder and Strömberg (2010) show that when local media markets overlapped with political constituencies in the US, citizens were better informed about their local politicians and more able to hold them accountable. In China the perfect congruence between media market and political division enabled local governments to capture local media.

mostly reflects a particular supply-side factor, that is the preferences of local governments.

The rest of the paper is structured as follows. The next section provides institutional background on newspaper censorship and the anti-corruption campaign in China. Section 1.3 describes the data. The empirical strategy and main results are presented in Sections 1.4 and 1.5, respectively. Mechanisms explaining these results are discussed in Section 1.6 and robustness checks are reported in Section 1.7. Section 1.8 concludes.

1.2 Background

1.2.1 Types of 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).[7] All newspapers are regulated and licensed by the State Administration of Press, Publication, Radio, Film and Television (SAPPRFT).

General-interest newspapers in China can be divided into three different types, depending on their ownership structure and revenue sources: official party newspapers, government-owned commercial newspapers and

[7] Many newspapers in China publish online editions and actively disseminate news on social media platforms. Online news sites also frequently reprint and quote newspaper articles.

subsidiary commercial newspapers. Official party newspapers are owned and operated by the government. 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 party newspapers tend to be the most biased.

In contrast, commercial newspapers rely on advertising revenues. They compete for customers using more entertainment-oriented content. While some commercial newspapers are also government-owned, others are owned by other (government-owned) newspapers.[8] Top personnel decisions at subsidiary newspapers are made by their parent newspapers.

The geographic distribution and 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 Chinese newspapers are circulated locally.[9] These local newspapers are owned by local governments of different administrative levels, that is, provincial governments own provincial-level papers and prefecture governments own prefectural-level papers etc.

[8]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).

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

1.2.2 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 these guidelines can result in demotions, dismissals and jail sentences for journalists and editors (Shirk, 2011).

Apart from distributing propaganda and making profits, Chinese newspapers are also expected to report the performance of local officials to the central government (see, e.g. Shirk, 2011) and Qin, Strömberg and Wu (2018) find that central and party newspapers cover more news about official corruption in general. This role of local newspapers brings them into conflict with their local government officials. Local propaganda bureaus, which oversees newspapers, answer in principle to both the central propaganda bureau and the party leadership of their locality. In this paper, I find that local newspapers consistently underreport on important corruption cases in their home area rather than acting as a watchdog. This is consistent with anecdotal evidence (see He, 2004) and online leaks from the Guangdong province propaganda department:

> *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)[10]

[10] 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`.

> *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)[11]

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

1.2.3 Anti-Corruption Campaign

When Chinese President Xi Jinping came to power at the end of 2012, he vowed to end corruption in the CCP for fear it would otherwise "doom the party and the state". The Central Commission for Discipline Inspection (CCDI) was tasked with investigating the over 80 million CCP members and punished 1.34 million of them for corruption in Xi's first five years in office.[12] Unlike previous anti-corruption drive, Xi promised to investigate both "Tigers" (high-level officials) and "Flies" (low-level officials).[13] Con-

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

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

[13] In China, an official of vice-provincial rank or higher is considered a high-ranking official (see, e.g., Li and Zhou, 2005). Personnel decisions for low- and high-ranking officials are undertaken by different organisations. When reporting about the campaign, the Chinese media uses the same rank distinction.

trary to press speculations, Lorentzen and Lu (2018) and Francois, Trebbi and Xiao (2016) find little evidence that the campaign is driven by factional politics.

According to the CCDI, citizen complaints are the most important source of information for the anti-corruption campaign. In 2012, 41.8 percent of all investigations originated in citizen complaints.[14] Since then, the CCDI has streamlined complaint procedures, increased resources to process complaints and publicised the importance of complaints.[15] The number of complaints have steadily increased to reach 34 million in 2018.

While the majority of corruption investigations are conducted internally, the CCDI announces investigations into some officials directly on its website. These announcements include all "Tigers", as well as a selected set of "flies". The investigation announcements tend to be unexpected and almost always result in criminal charges or, at a minimum, the end of the official's career. A typical example is the following CCDI post from 17 October 2013:

[14]20.9 percent of cases were referred by auditors and other prosecutors, while 7.1 percent of cases were connected to earlier investigations. See "中央纪委：实名举报优先办理". *People's Daily*, 10 January 2013. `http://fanfu.people.com.cn/n/2013/0110/c64371-20151909.html`

Similarly, around 60 percent of investigated cases in Guangdong province in 2015 and 30 percent of cases in Beijing from 2013 to 2017 were sparked by citizen complaints. See "广东:今年查处省管干部中信访举报提供线索近6成". *CCDI*, 23 December 2015. `http://www.ccdi.gov.cn/yaowen/201512/t20151211_139494.html` and "纪委接受信访举报 5年来超15万件次". *Beijing Evening News*, 8 June 2018. Reprinted in *People's Daily* as "北京市纪委接受信访举报5年来超15万件次". `http://bj.people.com.cn/n2/2017/0608/c82840-30300319.html`.

[15]The CCDI conducted numerous press conferences and interviews national media outlets, emphasising the importance of citizen complaints for their investigations (often by citing specific cases), promising to process all complaints and reply to all non-anonymous complaints, as well as assuring citizens that there will be no retaliations against complainants.

> *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.*[16]

The anti-corruption campaign is covered extensively in the central state media. For instance, the *People's Daily* has a dedicated section on its website, where articles about new regulations, individual cases and the campaign progress are published. Online and print versions of other newspapers are also an important source of news for detailed accounts about the anti-corruption campaign. Television stations spend their limited broadcasting time on programmes about the campaign. For example, the CCDI and China's national broadcaster CCTV produced and aired several documentaries about the campaign, featuring on-air confessions of corrupt officials. A fictionalised anti-corruption TV drama (*In the Name of the People*), in part financed by the national agency responsible for prosecuting corruption cases, became China's most popular TV drama in recent years.[17] 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 on similar topics.

[16] "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`.

[17] 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`.

1.3 Data and Descriptive Statistics

1.3.1 Newspapers

The sample of newspapers consists of all Chinese-language general interest newspapers available on WiseNews, an online news 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).[18] I collect information on newspaper ownership from the SAPPRFT website and China Journalism Yearbooks.

Panel A of Appendix Table 1.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 1.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 more than half of China, in terms of population (53 percent) and GDP (57 percent).[19]

[18] 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, such as Nanfang Ribao, Nanfang Dushibao, Guangzhou Ribao and Nanguo Chengbao.

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

1.3.2 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, from the 14 provinces (and central government) that are covered by the WiseNews sample.[20] Appendix Figure 1.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. I use an officials' occupation to determine their rank in the government hierarchy (see Appendix Figure 1.A2 for the distribution of ranks).

Panel B of Appendix Table 1.A1 shows summary statistics of the officials in the sample. 9 percent of officials (34 officials) in the sample are high ranking. This represents the universe of all "Tigers" investigated over this time period from these provinces, see also 1.7.5. 78 percent of the sample (320 officials) are government officials, mainly from local governments (72 percent of the sample, 295 officials), and 22 percent are officials from state-owned institutions (hospitals, universities etc.) and enterprises.

1.3.3 Articles

I collect all articles about these 408 officials published by the 99 newspapers in the WiseNews database from two years prior to one year following

[20] Across all provinces, the CCDI named a total of 804 officials over this time period.

the CCDI announcement, resulting in 39,271 unique articles.[21] By construction, each article can be matched to a newspaper and one or more officials.[22]

Summary statistics of articles are reported in Panel C of Appendix Table 1.A1 and a more detailed text analysis of the content and sentiment of the articles is shown in Appendix 1.B. The average articles contains around 1300 words, though most of the articles are shorter. More than half of the articles are on pages one to five, with a mean page number of eight.[23] Among articles published following an investigation announcement, 22 percent of article headlines include direct references to corruption and eight percent of headlines include keywords associated with the central anti-corruption campaign.[24] One percent of article headlines 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 cases.[25]

In order to take into account all articles that can *potentially* be published, I construct 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

[21] I use the WiseNews search function and use the official's name and one other keyword, such as their occupation, organisation name or location.

[22] 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.

[23] I reconstruct absolute page numbers by matching WiseNews information on the article layout to electronic and hard copies of the newspapers.

[24] This includes direct references to the campaign, CCDI, the 18th National Congress, Xi Jinping, "Tigers" and "Flies".

[25] Appendix Table 1.A2 shows the characteristics of officials with no articles and only articles before or after the investigation announcement.

of Appendix Table 1.A1). Appendix Figure 1.A4 shows the fraction of official-newspaper pairs with at least one published article by day relative to the investigation announcement at day zero. The mean probability that a newspaper publishes an article about an official on a given day 0.1 percent.

1.4 Empirical Strategy

I compare whether and how local newspapers report about officials investigated during the anti-corruption campaign from their own province relative to officials from other provinces. There are three potential patterns of reporting. If the central government dictates a uniform media strategy surrounding the anti-corruption campaign, newspapers might report the same stories about each official, regardless of whether they are from the newspaper's home province. Newspapers could report more about investigated officials from their own province, either because they have more information about them or because their readers are more interested in their home-province officials (see Section 1.6.1 for evidence on reader demand). In contrast, if newspapers report less about corrupt officials from their own province, this could be in response to pressure from their local government - a scenario I refer to as local censorship.[26] There are a number of reasons for local censorship. Publicising the investigations could encourage members of the public to come forward and implicate officials still in

[26] I can not 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.

power. Supervisors and colleagues could be accused of condoning corruption. Revealing large-scale local corruption could damage the reputation of the local government. See Section 1.6.2 for evidence on citizen complaints and Section 1.6 for evidence that local censorship is stronger when it comes to well-connected officials.

1.4.1 Are Local Corruption Scandals Selectively Underreported?

I use a modified difference-in-differences approach to show how the extensive margin of reporting on an official changes when a corruption investigation is announced, depending on whether or not the official is from the same province as the newspaper. I estimate a linear probability model of the following form:

$$Article_{o,n,t} = \beta_1 Post_{o,t} + \beta_2(Post_{o,t} \times SameProvince_{o,n}) + \mathbf{X'_t}\delta + \theta_{n,t} + \gamma_{o,n} + \epsilon_{o,n,t} \tag{1.1}$$

with the indices denoting official (o), newspaper (n) and day (t), respectively. $Article_{o,n,t}$ is a dummy variable for whether an article about official o is published in newspaper n on day t. $Post_{o,t}$ is a dummy variable that takes the value of one for all days t after an investigation into official o has been announced. $SameProvince_{o,n}$ is a dummy variable for whether official o and newspaper n are from the same province.[27] All regressions include official-newspaper-pair fixed effects ($\gamma_{o,n}$), which account for time-invariant

[27] In Section 1.7, I show that the results are robust to different functional form assumptions and allowing for continuous $Post_{o,t}$ and flexible $SameProvince_{o,n}$ effects.

characteristics of officials, newspapers and official-newspaper pairs. These fixed effects absorb the direct effect of $SamePrivince_{o,n}$, but identify the coefficient of interest β_2 on the interaction term $Post_{o,t} \times SamePrivince_{o,n}$ off the time variation within an official-newspaper pair. I control for a comprehensive set of potential seasonal effects. $\mathbf{X_t}$ includes day-of-the-week, month and year 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). I also control for the provincial-level counterparts to the national CCP events ($\theta_{n,t}$). 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).

As corruption is newsworthy, we should expect β_1 to be positive, i.e., the likelihood of an article being published about an official increases when 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.

Heterogeneity by Official Rank The extent of local censorship should depend on the expected costs and benefits. The benefits are likely larger for censoring corruption scandals involving more powerful officials, while the costs could also likely be larger if readers are more interested in these

types of corruption cases as well. To test for this type of heterogeneity, I interact $Post_{o,t}$ and $SameProvince_{o,n}$ from equation 1.1 with dummy variables that indicate whether official o is a low- or high-ranking official.

Heterogeneity by Newspaper Type The cost of censorship could also differ by newspaper type. As outlined in Section 1.2, there are three types of local newspapers: government-owned party papers, government-owned commercial papers and subsidiary commercial papers. It might be easier for governments to censor directly government-owned newspaper, for example, because propaganda department officials are more 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 corruption scandals could lower readership and advertising. In this case, we should see more censorship in party papers which are financed through government subsidies. I test for this source of heterogeneity by interacting $Post_{o,t}$ and $SameProvince_{o,n}$ from equation 1.1 with indicators for whether newspaper n is an official party paper, a government-owned commercial paper or subsidiary commercial paper.

Within-Province Variation in Reporting The analysis in equation 1.1 can be extended to the municipality level, by replacing $SameProvince_{o,n}$ with two mutually exclusive dummy variables that take the value of one if the newspaper and official are from the same city ($SameCity_{o,n}$) and if they are from different cities of the same province ($DifferentCity_{o,n}$). If news-

papers are less likely to report corruption scandals from their own compared to other cities in the same province, then the coefficient on $SameCity_{o,n}$ should be smaller than that on $DifferentCity_{o,n}$.

1.4.2 How are Local Corruption Scandals Reported?

Newspapers can choose whether to report news, as well as how to frame it. I estimate the following equation to determine whether a newspaper writes different articles about officials who were revealed to be corrupt from their own province compared to those from other provinces

$$\begin{aligned} ArticleCharacteristic_{o,n,t} =& \kappa_1(High_o \times SameProvince_{o,n}) + \kappa_2(Low_o \times Same \\ & Province_{o,n}) + \mathbf{X_t'}\delta + \theta_{n,t} + \gamma_o + \gamma_n + \epsilon_{o,n,t} \end{aligned} \tag{1.2}$$

on the sample of articles published after the investigation announcements. I use the following quantifiable characteristics of articles as dependent variables: log word count, the balance of positive relative to negative words (see Appendix 1.B for more details about sentiment analysis), page number, indicators for whether an article mentions multiple officials and citizen complaints against officials. I also characterise the headlines of articles with dummy variables for whether the headline is a copy of the official announcement, includes explicit references to corruption and keywords associated with the centrally-run anti-corruption campaign. As there can be multiple officials mentioned in one article, the regressors $\tilde{High}_{o,n}$ and $\tilde{Low}_{o,n}$ 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 news-

paper fixed effects, standard errors are clustered at the newspaper level and the rest of the notation is as before.[28]

1.5 Results

1.5.1 Underreporting of Corruption Scandals: "Tigers" versus "Flies"

When reporting about the anti-corruption campaign, the media tends to distinguish between "Tigers" and "Flies". Figure 1.1 shows the fraction of newspapers that report at least one story about each official under investigation, compared to the number of internet searches on Baidu for each official. Newspapers write fewer articles about investigations of high-level officials from their own province than officials from other provinces (Panel A), but *more* articles about corrupt low-level officials from their own province (Panel C). Table 1.1 shows that in the ten days after a high-level official is investigated, own-province newspapers are 50 to 100 percent less likely to publish an article compared to out-of-province newspapers.

This absolute comparison does not account for larger reader interest in local news (see also Section 1.6.1). Panels B and D in Figure 1.1 show that there are on average 15,000 more searches on Baidu (China's most popular search engine) for a "Tiger" on the day of his investigation an-

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

nouncement (and 1,700 more searches for a "Fly") in that official's home province compared to other provinces. This disparity in internet searches and the fact that local newspapers do report more about low-level officials under investigation suggest that there is demand for local newspapers to report more about high-level corruption scandals in their own area. While it is difficult to ascertain the precise extent of reader interest, I present different scenarios in Table 1.1 using newspaper reporting about low-level officials' scandals, searches on Baidu and comments on Sina Weibo posts. The common assumption is that the ratio of own- versus other-province reporting of high-level officials should be the same as in the benchmarks absent local government intervention. Adjusting for reader interest implies a greater extent of underreporting.

Table 1.2 presents the estimation results for equation 1.1 that for the pooled set of all officials, the estimated coefficient on the post-announcement dummy (β_1) is statistically significant and positive (columns 1 and 2). The daily probability that a newspaper publishes an article about an official from another province increases by 0.3 percentage points, when it is announced that this official is investigated for corruption (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. On average, newspapers do not report differently about the anti-corruption campaign's investigations of officials from their own province.[29]

[29] One possible explanation for this finding is that local newspapers have to satisfy central government guidelines in publicising the campaign and writing about the majority of cases involving low-level officials from their home province helps mask the selective underreporting of high-level corruption. Another explanation is that suppressing all news about local corruption cases could alert readers to the bias in reporting and re-

Columns 3 and 4 reports the interaction of the same-province dummy with indicators for high- and low-level officials. Local newspapers underreport corruption cases of "Tigers" from their own province, but not those of "Flies". The daily probability of an article about a high-ranking official being published by a newspaper from the same province falls by 4.8 percentage points after the CCDI announcement, which is 73 percent of the pre-announcement mean of 6.6 percent. This means that a newspaper which used to write on average 24 articles per year about an official before he was investigated, writes six articles about him in the year following the investigation. The reverse pattern holds for low-level officials. The probability of reporting in a same-province newspaper increases by 0.5 percentage points (or 167 percent of the pre-announcement average) for a low-level official following an investigation announcement. The likelihood of reporting on both types of officials from other provinces increases following an investigation announcement, however, this increase is four times larger for high-ranking compared to low-ranking officials.

While investigations of low-level officials are generally not underreported by newspapers from the same province, there in one exception. City-level newspapers underreport corruption scandals involving officials from the same city, but not those involving officials from other cities of the same province, see Figure 1.3 and Appendix Table 1.F1.[30] The difference in the coefficients is statistically significant at the 5 percent level.

duce their trust in local newspapers and the local government (see, e.g., Kamenica and Gentzkow, 2011; Gehlbach and Sonin, 2014).

[30] Almost all city-level officials are low ranking, while high-ranking officials most occupy positions within the provincial government or party organisation.

1.5.2 Deemphasising Local Corruption Scandals

What do newspapers write in their published articles? I use an unsupervised machine learning algorithm to analyse the full text of all articles and find broadly two types of articles: the first type contains news about officials' work and policies ("Development" topic) and the second type references officials' corruption investigation ("Corruption" topic), see Appendix 1.B for more details. The probability that articles are about corruption increases sharply when an official was investigated for corruption (see Figure 1.2).[31] For 91 percent of post-announcement article-official pairs, corruption is identified as the dominant topic.[32]

I test whether newspapers report differently about officials from their own province in the year following the officials' investigation, by estimating 1.2. The resutls reported in Table 1.3 suggest that newspapers deemphasise stories about corruption scandals involving high-level officials from their own 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). There is a smaller, negative, but not significant effect for low-level officials. Column 2 shows that articles about officials under investigation from a newspaper's home province differ in tone. Compared to the average article, an article which only mentions high-level corrupt officials from the same province contains almost twice as

[31] Results are similar when using a topic model with nine topics, which has the highest coherence score out of a set of models with between two and twenty topics.

[32] I manually read a random sample of 218 articles that were published after an investigation announcement, stratified by whether the newspaper and official are from the same province and by the official's rank. All of these sample articles mention the official in connection with corruption.

many positive relative to negative words, while an article about low-level local officials contains around 50 percent more positive than negative words (column 2, see also Appendix 1.B).

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 the campaign.

Consistent with the anti-corruption campaign's efforts to encourage citizens' complaints against their local officials, 13 percent of post-investigation announcement articles discuss the role that accusations by members of the public played in official investigations. The probability of an article about a local "Tiger" containing references to citizen complaints is 5.5 percentage points lower, while there is no significant effect for articles about low-level officials from the same province (Column 5).

Analysing the text of article headlines provides further details into the framing of these news 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 one when an article headline copies the CCDI announcement. Relative to a mean probability of 0.9 percent, an ar-

ticle with only high-level officials from the same province is 1.6 percentage points more likely to be copied from the CCDI. Headlines about high-level officials from the same province are also 6 percentage points less likely to include keywords, such as "corruption", compared to 22 percent of articles in the sample (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 35 percent 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. Taken together, the results of this section suggest that even when newspapers write about corruption in their own province, they are more cautious in their writing and do not try to draw their readers' attention to these types of stories.[33]

1.6 Mechanisms

To explain what underlies this selective underreporting of local corruption, I first show that this pattern of reporting does not reflect demand from newspapers' readers and reduces the local population's propensity to com-

[33] Appendix Figure 1.D1 shows the average page number of articles. Newspapers place articles about high-level officials from their own province on average three pages further back in the paper once they are investigated for corruption, relative to the position of a typical pre-investigation article. While articles about high-level officials from a newspaper's own province are placed significantly before articles about officials from other provinces before an official is under investigation, this gap closes once officials are investigated.

plain about their officials. Underreporting is more severe when newspapers are more dependent on the local government and when officials are better connected in local power networks. I interpret these results as evidence that local governments strategically censor newspapers in their own locality even counter the interests of higher levels of government and this censorship reduces accountability.

1.6.1 Readers' Interest in Local Corruption

Internet Searches As shown in Figure 1.1, internet searches for officials are higher in that official's home province than in other provinces. In the year following an official's investigation announcement, there are on average six times more searches about corrupt officials in their own province than in the nine highest outside provinces (see Appendix Figure 1.D2 for a histogram of relative search volumes).[34] Appendix Table 1.D1 confirms that Baidu searches for a corrupt official are significantly higher in their home province than in the nine highest outside provinces, controlling for official and province fixed effects. The effect size is large and ranges from 80 index points for low-level officials to 84 index points for high-level officials on an index scale from 0 to 100 (compared to a mean of 23 index points). The difference in effect size for high- versus low-level officials is not statistically significant.

[34]Data are only available for a subset of 214 officials, including all but three "Tigers". For seven officials, the official's province was not among the ten highest provinces in terms of searches and I conservatively assume zero searches in these cases.

Social Media Comments Newspapers' reach extends beyond their print subscribers and many newspapers are very active in creating news content on social media. I collect data on all posts about high-level officials following their investigation on newspapers' official Sina Weibo accounts (the most popular Chinese micro-blogging website, which is similar to Twitter), along with comments about these posts and information about the commenters. (See Appendix 1.C for more details about the data collection process and for summary statistics of the Sina Weibo data.) I find that people are three times more likely to comment on a post about an official from their own province compared to an official from another province.

More formally, I construct a dataset of Weibo comments by commenter's province and estimate an analogous model to equation 1.2. Table 1.5 shows that people write three times more comments on posts by newspapers from their own province and twice as many comments on posts about officials from their own province, compared to the average number of comments by province (Columns 1 and 2). Columns 3 and 4 show that social media users are also faster to comment on posts by their local newspapers and about their local officials, although the estimates are imprecise.

Together with the Baidu results, these findings show that people are more interested in corrupt officials their own province. This greater interest in local affairs is not reflected in the publishing behaviour of local newspapers, who underreport news about local "Tigers". These social media findings also have important implications for the role of local newspapers in the spread of information in China. Not only is the print circulation of local newspapers restricted to their local area, but their online reach is also

greater among local users. This implies that even social media users who have easy access to alternative outside news sources could still be affected by bias in their traditional local news media.

1.6.2 Effect of Underreporting on Local Government Accountability

Complaints to higher levels of government are often the only way the population can hold their local officials accountable in an autocratic state like China.[35] Encouraging citizens to denounce their local officials for suspected corruption has been a key message during the anti-corruption campaign. I collect the daily search Baidu search volume for "non-anonymous complaint" by province, which is the most common (and encouraged) method for citizens to complain about officials.[36]

In Table 1.4 I regress the number of Baidu searches for complaints on investigations about officials and summary measures for the publication behaviour of provincial newspapers from the same province, while controlling for day and province fixed effects. Column 1 shows that on the day of a CCDI investigation announcement into a high-level official, the number of internet searches for complaints in that official's province increases by around one third of the mean. This effect is three times larger for high- compared to low-level officials.

[35]The practice of petitioning higher official goes back to imperial times and is nowadays managed by a set of Bureaus for Letters and Visits at the local and national level.

[36]The Chinese phrase is 实名举报. Results using other keywords, including the name of the CCDI's complaint website are similar.

Columns 2 to 5 of the same table shows that the anti-corruption campaign coverage by local newspapers affect internet searches for non-anonymous complaints.[37] Searches for complaints in a province increase by more than 50 percent of the mean when provincial newspapers report about corrupt officials from their own province (column 2). The effect of local newspaper reporting high- and low-level officials from their own province is 15 and 9 times larger, respectively, than the effects of the CCDI announcement. Reporting about high- but not low-level corruption scandals from other provinces also increases searches for complaints, but to a much smaller extent. How newspapers report about corruption also matters. Column 3 shows that internet searches for complaints increase significantly when local newspapers publish articles that discuss citizen complaints. Similarly, searches increase when article headlines contain words related to "corruption" and the anti-corruption campaign (results are not significant).

In Section 1.5, I show that local newspapers not only significantly underreport corruption scandals involving high-level officials from their own province, they also deemphasise these stories by, for instance, including fewer references to citizen complaints and less informative headlines. To the extent that searching online for complaint procedures correlates with the intention to lodge an actual complaint against corrupt officials, this underreporting and deemphasising of corruption scandals in their own area by local Chinese newspapers reduces the ability of the central government to monitor and punish local officials. In other words, local governments' control over newspapers reduces accountability of local officials to the pop-

[37] These results are calculated based on the sample of provincial newspapers in the Wisenews archive and assuming that they are representative for each province.

ulation and higher levels of government and, thus, reduces the cost of corruption.

Local newspaper reporting could affect accountability through their agenda-setting power, as well as by representing the local power structure. While the CCDI announcements are public, local newspapers still play an important role in disseminating this information.[38] When newspapers write fewer and less explicit articles about corruption investigations into local officials, local corruption might appear less salient or the local population might be less aware of the importance of citizen complaints for the anti-corruption campaign. The local population could also interpret newspapers' reluctance to report about local corruption as a sign that the local power structures remain intact or that the central government is unable to control the local government. Both of these factors could reduce citizen complaints either for fear of retaliation or a sense of inutility of complaining.[39]

1.6.3 Conditions for Newspaper Capture

Different newspapers underreport investigations of local "Tigers" to a varying extent. In general, the more a newspaper depends on the local government, the more it will underreport corruption.

[38]Many government and state-owned enterprise employees still receive party newspapers at their work place and for many older people newspapers remain one of the most important sources of local news. Local newspapers are also active on social media and many internet news websites copy articles from local newspapers.

[39]See, for instance, Edmond (2013) for a theoretical model where a regime uses propaganda to prevent being overthrown.

Provincial Newspapers versus Sub-Provincial Newspapers High-level officials mostly occupy positions in the provincial government or party organisation. The provincial government oversees provincial newspapers and municipal governments, which in turn oversee their own municipal newspapers. Within a city, provincial and sub-provincial newspapers compete for similar local audiences.[40] In Figure 1.4 and Appendix Table 1.F3, I compare articles in provincial and sub-provincial newspapers about investigations into officials from the provincial government. Panel A shows that, compared to a sub-provincial newspaper from the same province, a provincial newspaper underreports twice as much about 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. While the provincial government pressures provincial newspapers to underreport corruption within its ranks contrary to the wishes of the central government, it also only has imperfect control over newspapers from lower level governments.[41]

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.

[40]Provincial newspapers are circulated throughout the province, while sub-provincial newspapers have more local circulation.

[41]Qin, Strömberg and Wu (2018) find that newspapers from lower levels of governments are in general less biased.

Government Ownership, Advertising Revenue and Competition

Local general interest newspapers in China differ in terms of their ownership and commercialisation. Figure 1.5 shows how different types of newspapers report corruption scandals involving high-level officials from the same province (the full set of coefficients is reported in Appendix Table 1.F2). All local newspapers underreport, but party newspapers underreport the most, followed by government-owned commercial newspapers and subsidiary commercial newspapers. The difference in estimates between party newspapers and commercial newspapers is statistically significant, while the difference across commercial newspapers is not. All newspapers increase reporting when a *low-level* local official is investigated for corruption, by similar amounts. There is no difference between newspaper types when reporting about either high- or low-ranking officials from other provinces.

To test the impact of government ownership directly, I match each subsidiary newspaper to their parent newspaper (an official government-owned party newspaper) when both are in the sample. Appendix Figure 1.D3 shows the difference in the mean probability of publication before and after an official is investigated, that is, a negative number means that newspapers report less about an official after they are investigated 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. This pattern does not hold for low-level officials or out-of-

province newspapers. Using advertising revenue data from Qin, Strömberg and Wu (2018), Appendix Figure 1.D4 shows that newspapers with higher initial advertising revenue are more likely to publish articles about corruption scandals involving high-level officials from their own province relative to high-level officials from other provinces. This difference is statistically significant at the ten percent level. For low-level officials, there is no statistically significant relationship between advertising revenue and relative reporting.[42]

At the province level, Appendix Figure 1.D5 shows that there is less estimated underreporting of high-level corruption scandals in own-province newspapers when a province has a higher number of newspapers and a higher level of GDP proxying for a larger potential advertising market (see also Section 1.7.2). This relationship is statistically significant at the ten and five percent level, respectively, despite the small number of observations. These results illustrate that local governments in China face some constraints in censoring the local media. In line with the theoretical predictions of Besley and Prat (2006), media capture becomes more difficult when there is more competition between media outlets, media outlets are more commercialised and media outlets are not government owned.

[42]Qin, Strömberg and Wu (2018) find that reporting about corruption is negatively associated with advertising revenue prior to the anti-corruption campaign, while I find a significant and positive relationship for both high- and low-level officials, as well as for officials from the same province as a newspaper and for officials from other provinces.

1.6.4 Importance of Local Political Connections

While the investigated officials themselves are immediately removed from power, there could exist local corruption networks whose remaining members use their influence to reduce media coverage. I show in this section that investigations of better connected officials are more likely to be underreported.[43]

Proximity to the Head of Province With one exception, the highest ranking official of each province (the provincial CCP secretary) was not investigated in the earlier days of the anti-corruption campaign. The party secretary presides over the provincial standing committee of around 15 members, which is the top decision making body in each province. Using the People's Daily website and Baidu Baike, I collect information on the career path of all high-ranking officials and determine whether each official has served on the provincial standing committee with the incumbent party secretary.[44] This is a good proxy for connection to the head of province, as an official is unlikely to become a member of the standing committee without the party secretary's approval. Figure 1.6 shows that newspapers are more than twice as likely to underreport investigations involving officials who are connected to their province's party secretary, compared to similarly ranked but unconnected officials. 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 1.F4).

[43]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 suggest that this mechanism is less important.

[44]Baidu Baike is a popular Chinese web encyclopedia, similar to Wikipedia.

Case study: Wan Qingliang The case of Wan Qingliang, the 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, before he was unexpectedly investigated for corruption. He was reportedly implicated by previous CCDI investigations of his former aides. There are five low-level officials from Guangzhou in the sample who were connected to Wan (hereafter, Wan-group officials) according to ChinaFile.[45]

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 official was investigated, I split the post-investigation period into two: before and after Wan himself was investigated. While Wan was in power, newspapers from Guangzhou city (where he was party secretary) write significantly less about officials connected to Wan compared to other newspapers from Guangdong province ((Figure 1.7 Panel A and Appendix Table 1.F5). Guangzhou newspapers also reported less about Wan-group officials relative to other comparable officials, although this difference is statistically significant (p-value of 0.12).[46] After Wan was removed from power, there is no longer any difference in how much Guangzhou newspapers report about Wan-group officials compared to other officials. Reporting about Wan-group officials becomes similar in Guangzhou newspapers

[45] "Catching Tigers and Flies". *ChinaFile*. `https://anticorruption.chinafile.com`, last accessed: 20 April 2019. ChinaFile uses government organs and major media outlets as sources. Although I focus on local newspapers, I cannot rule out the endogeneity of this connection measure.

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

and other newspapers from Guangdong (Panel B). Taken together, these results suggest that corruption scandals in local newspapers are underreported in order to avoid implicating other (influential) officials who are still in power.

1.6.5 Interpretation of Results

I interpret these results as evidence that local governments in China selectively censor news that is unfavourable to them, 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. The extent of censorship depends on the relationship between a newspaper and their local government. Government-owned newspapers and newspapers that face less competition or rely less on advertising revenue are more biased. Sub-provincial level newspapers are less biased than their provincial-level counterparts when it comes to high-level corruption in the provincial government. Just as the central government cannot perfectly control provincial governments, sub-provincial governments in turn only partially answer to provincial governments. Censorship is more severe when a corrupt official is well-connected. Underreporting and deemphasising local corruption appears to reduce the local population's propensity to complain to higher levels of government, which is a key mechanism for holding local officials accountable in an autocracy.

The results are inconsistent with a number of alternative explanations. The internet search and social media results suggest that underreporting investigations of officials from the same province does not reflect a lack of reader interest. The heterogeneous effects by newspaper and official type are difficult to reconcile with supply-side stories that do not focus on the incentives and actions of the local government. One alternative hypothesis could be that provincial government overreport corruption in other provinces to distract from their own problems. Another alternative interpretation is that the central government chooses to publicise the campaign in provinces that were not directly affected in order to avoid social unrest. Section 1.6.3 shows that more dependent newspapers do not report more on corruption cases in other provinces, while they do underreport high-profile investigations in their home province. When high-level provincial officials are placed under investigation, these stories are underreported more in provincial relative to sub-provincial newspapers which are circulated in some of the same areas. Section 1.5.2 also suggests that home-province investigations are reported in a way to give less credit to the central anti-corruption campaign. Anecdotal evidence from China also casts doubt on these alternative interpretations. Local governments reportedly collude to censor cross-border reporting of corruption (He, 2004). People in China generally trust the central government more than their local governments and institutional channels exist for citizens to complain about their local officials to higher levels of government.

1.7 Robustness

1.7.1 Time-Profile of Estimated Coefficients

Instead of using a post-investigation period dummy variable, I estimate separate coefficients by month (relative to the announcement date) in Figure 1.8 (and Appendix Table 1.F6). Prior to the announcement, the estimated effects are not statistically different from zero suggesting that the investigation announcements were not anticipated. The month of the investigation has the largest impact on publication probability. For articles about officials from other province, the effects fade gradually over the year following the announcement (Panels B and D). The probability of publishing an article about a low-level officials from the same province is only significantly higher in the month of the announcement (Panel C), while underreporting on high-level officials from the same province persists for the year following the announcement (Panel A).

1.7.2 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 ($SameProvince_{o,n}$) with a vector of indicators for all the provinces of newspapers and high-level officials. Table 1.6 shows how newspapers from a province (in columns) reports on an official in another province (rows) after the official is investigated for

corruption. The diagonal entries (in bold) show the the effect of newspapers being from the same province as an official separately by province. All but one (for Hunan province) of the diagonal entries are estimated to be negative, while the off-diagonal entries are all estimated to be positive, lending support to the main specification.[47]

1.7.3 Functional form specification

I show the sensitivity of estimates to using a Poisson QML estimator with the number of articles as the dependent variable ($\ln(NumberArticles_{o,n,t})$) and clustered bootstrapped standard errors at the newspaper and official level.[48] Results are shown in the first two columns of Appendix Table 1.E1 and confirm that investigations of high-level officials from the same province are underrepored. For comparison, columns 3 and 4 of Appendix Table 1.E1 show the OLS results for the Poisson sample which omits all official-newspaper pairs where no article was published. In the restricted sample, the estimates tend to increase in magnitude and significance.

[47] 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). A large estimated effect is also seen for Chongqing, the province of Bo Xilai who was widely seen as Xi Jinping's rival. See also Section 1.6.3.

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

1.7.4 Investigative Reporting

It could be a potential concern for the empirical strategy if newspapers' investigative reporting of corruption in their own province triggers the anti-corruption investigations. This type of reverse causality appears inconsistent with the data. Appendix Figure 1.D6 shows the share of newspaper articles that refer to corruption in their headline. Before an investigation announcement, newspapers are more likely to mention corruption in connection with both 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.[49]

I re-estimate the main results excluding nine 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 1.E2 shows that in this restricted sample newspapers underreport high-level corruption scandals from the same province, although the estimated magnitude is smaller.

1.7.5 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.

[49]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.

For instance, if the CCDI investigates groups of high-level officials from the same province at the same time, but only names one official in its announcement and newspapers from the same province write about the entire group rather than only the named official. To rule out this channel, I compare the list of high-level corrupt officials from the CCDI announcements with a number of different news reports and data on the anti-corruption campaign by ChinaFile. I find no mention of any additional corrupt high-level official during the sample period.[50] ChinaFile uses a 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.[51] Using this alternative definition does not change the main results significantly, see Appendix Table 1.E3. 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 particular 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 1.6.4 suggest that the results are not purely driven by selection. The results at the individual newspaper level for parent versus subsidiary newspapers and newspaper advertising revenue in Section 1.6.3 are also inconsistent with pure selection.

[50]Some media sources also mention high-ranking military officers as part of the campaign. Military officers are not comparable to government and party officials.

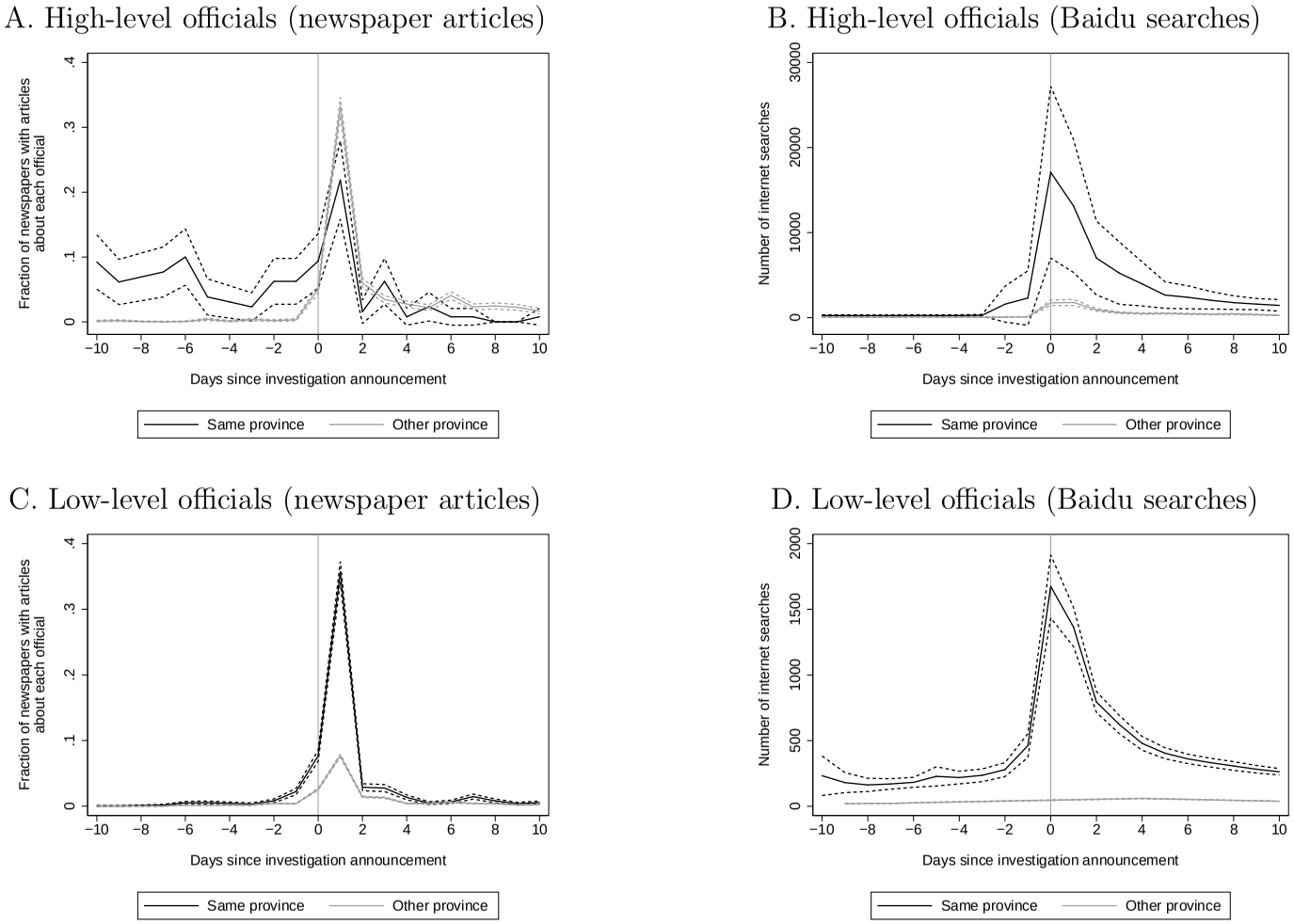
[51]The difference mainly stems from officials working at state-owned enterprises.

1.8 Conclusion

Media censorship is a common tool for autocratic governments to maintain control over their population. In this paper, I show how conflicts of interest within an autocracy can lead to diverging media censorship strategies in the context of a centralised anti-corruption campaign in China. I use an original dataset of Chinese newspaper articles to compare how local newspapers report investigations of officials from their own area relative to other areas. Despite greater interest from their readers, newspapers write fewer articles about corruption scandals involving high-level officials from their own province relative to officials from other provinces. The articles are also shorter, more positive, feature less interesting headlines and are less likely to mention the role of the central leadership. Investigations of well-connected local officials are underreported to a greater extent. Economic competition for advertising revenue and readership mitigates some of the reporting bias. These findings suggest that local governments pressure newspapers from their own areas to suppress stories that are unfavourable to them. I show suggestive evidence that this type of local censorship affects the ability of higher levels of government to monitor and punish local officials.

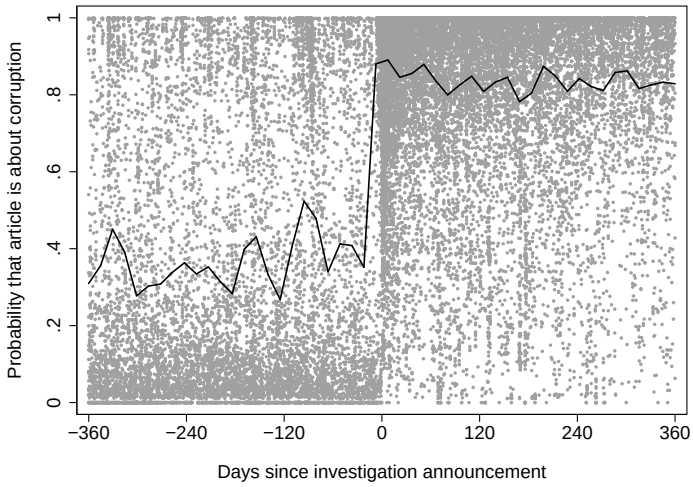
1.9 Figures

Figure 1.1: Newspaper Articles versus Internet Searches



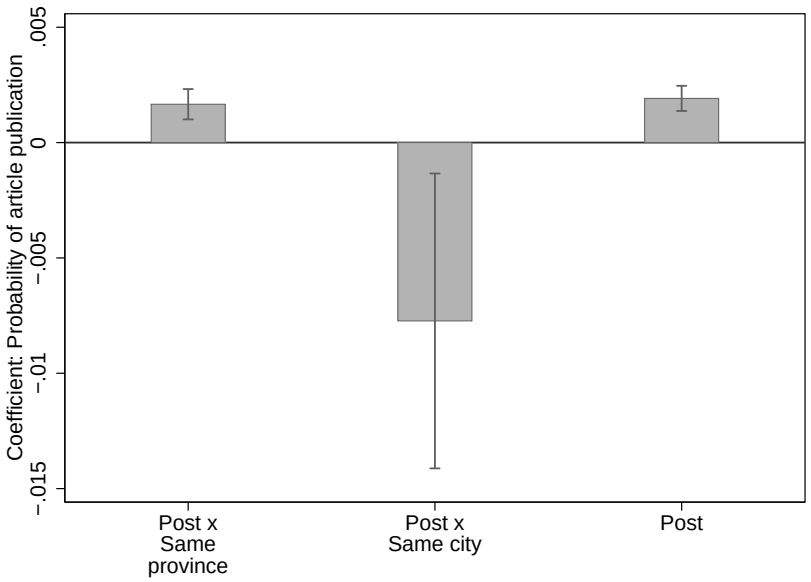
Notes: These charts show the average fraction of newspapers that publish at least one article about each investigated official (Panels A and C) and the average number of Baidu searches for each official (Panels B and D) in the solid lines with the 90 percent confidence intervals in the dotted lines. Newspaper articles about and internet searches for officials from the same province are shown in black (officials from other provinces are shown in grey). The x-axis shows the days before and after the announcement of an investigation into an official.
Newspapers publish more articles about officials when they are investigated for corruption. Newspapers write more articles about high-level officials under investigation from other provinces compared to officials from their own province in absolute terms. The reverse pattern holds for low-level officials. Baidu searches are always higher for officials from the same province suggesting more interest from the population.

Figure 1.2: Evolution of "Corruption" Articles



Notes: This chart shows a scatter plot for all article-official pairs. On the y-axis is the estimated probability of an article belonging to the "Corruption" topic using a two-topic topic model (see Appendix 1.B for more details) and on the x-axis is the day that the article was published relative to the day that the official mentioned in the article was investigated. The black line shows the fitted values of a linear polynomial smoother using an Epanechnikov kernel.
After an official's investigation is announced, articles about that official are more likely to be classified as "Corruption".

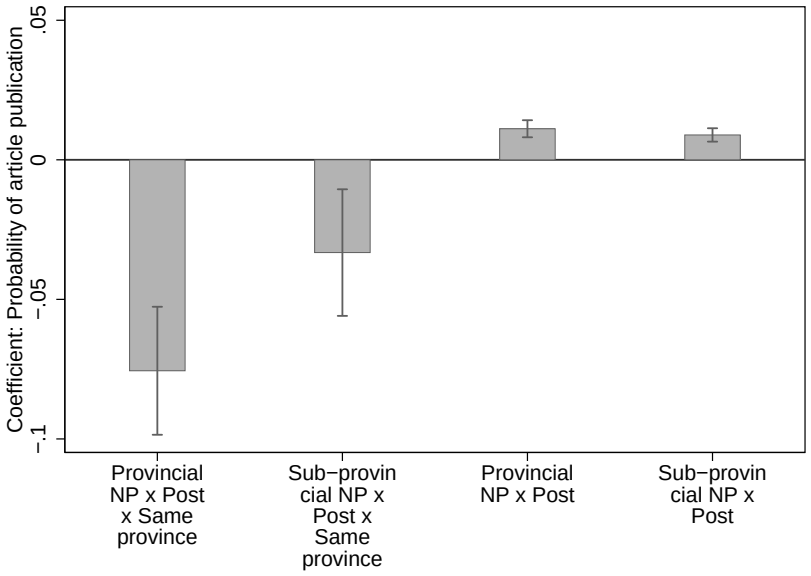
Figure 1.3: Municipality-Level Underreporting



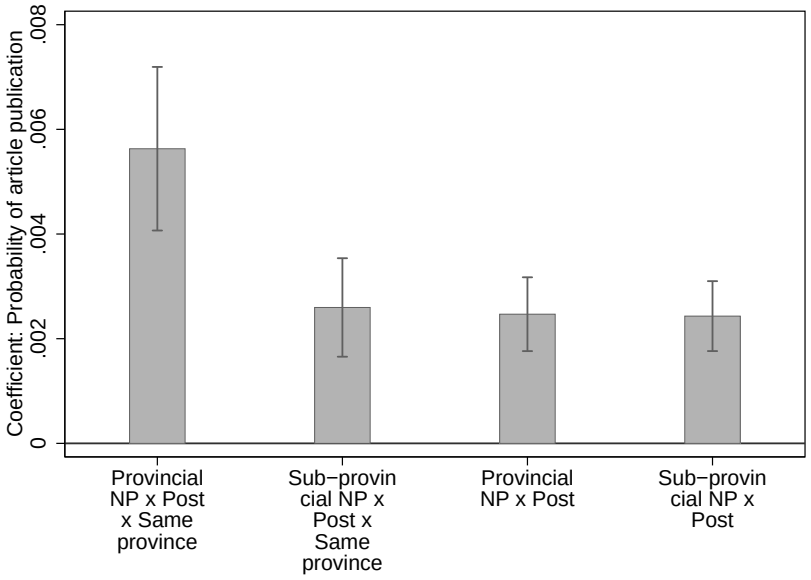
Notes: This chart shows coefficients from a regression of a dummy for whether an article is published on an indicator for the post-scandal period ($Post_{o,t}$) and its interactions with two dummy variables that takes the value of 1 if an official and a newspaper are from the same city ($Post_{o,t} \times SameCity_{o,n}$) and if an official and a newspaper are from different cities of the same province ($Post_{o,t} \times DifferentCity_{o,n}$). 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 1.2. 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 1.F1.
City-level newspapers underreport corruption scandals involving officials from the same city, but not those involving officials from other cities in the same province.

Figure 1.4: Provincial Newspapers versus Sub-Provincial Newspapers

A. High-level Provincial Officials

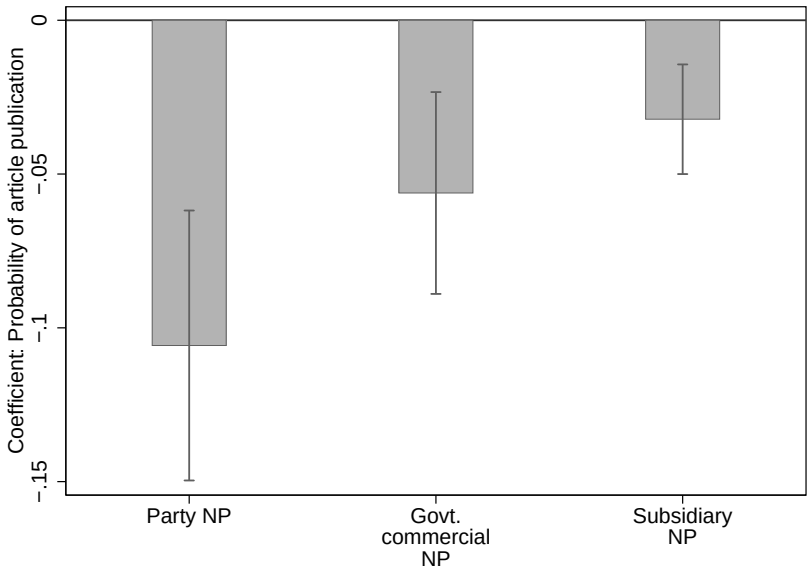


B. Low-level Provincial Officials



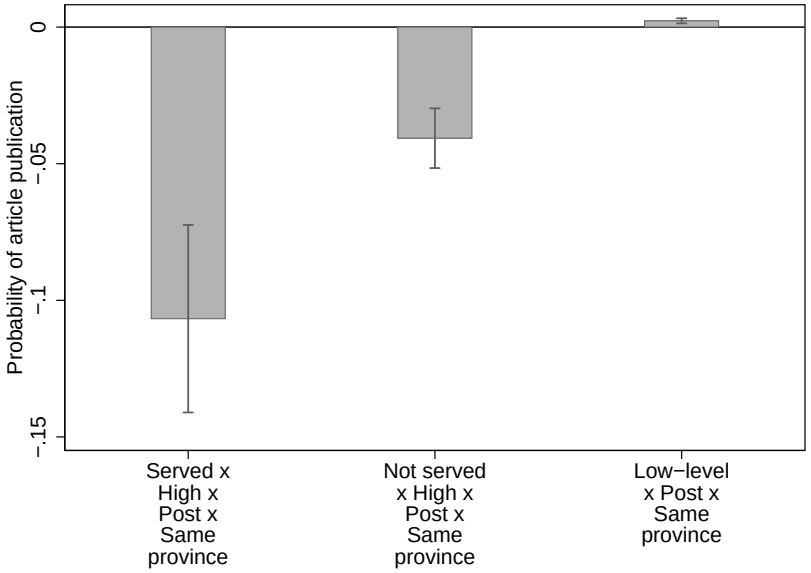
Notes: These charts show coefficients from a regression of a dummy for whether an article is published on the interaction terms between two dummy variables for provincial-level and sub-provincial level newspapers and an indicator for the post-scandal period and newspapers from the same province ($Post_{o,t} \times SameProvince_{o,n}$) 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 ($Post_{o,t}$) are shown as a comparison. The regression includes official-newspaper-pair fixed effects and all seasonal controls from column 2 in Table 1.2. 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 reported in Appendix Table 1.F3.
Provincial newspapers are underreport investigations into high-level officials from the provincial government to a greater extent than sub-provincial newspapers from the same province. Provincial newspapers publish more articles about low-level provincial officials than sub-provincial newspapers.

Figure 1.5: Party Newspapers versus Commercial Newspapers



Notes: 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) and an indicator for high-level officials from the same province as a newspaper in the post-investigation period ($HighLevel_o \times SameProvince_{o,n} \times Post_{o,t}$). The regression includes official-newspaper-pair fixed effects and all seasonal controls from column 2 in Table 1.2. 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 reported in Appendix Table 1.F2.
All newspapers underreport, but party newspapers underreport significantly more than both types of commercial newspapers. Party newspapers are heavily subsidised by local governments, while commercial newspapers depend on advertising revenues.

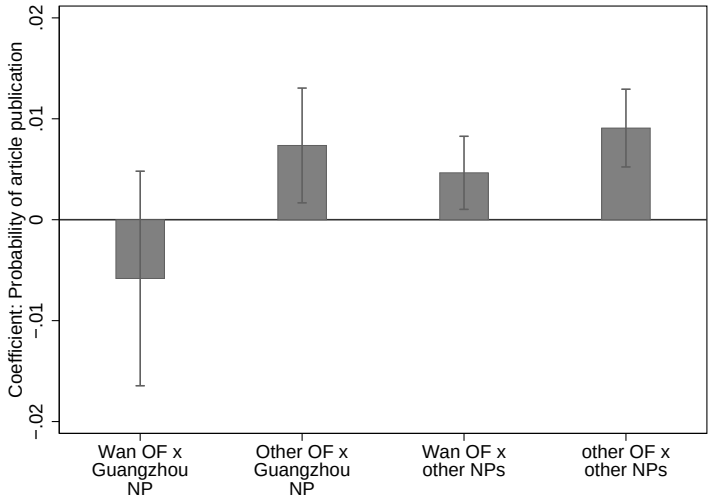
Figure 1.6: Proximity to the Head of the Province



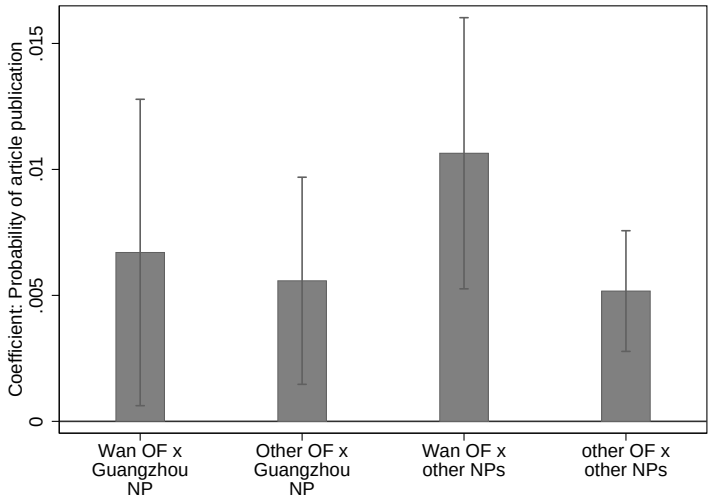
Notes: This chart shows coefficients from a regression of a dummy for whether an article is published on the interaction between a dummy variable for whether a high-level official has served on the party standing committee with the incumbent party secretary of the province (served versus not served) and a post-investigation dummy interacted with an indicator for whether the newspaper and official are from the same province ($Post_{o,t} \times SameProvince_{o,n}$). The coefficient on low-level officials is shown for comparison. The regression includes official-newspaper-pair fixed effects and all seasonal controls from column 2 in Table 1.2. 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 1.F4.
Newspapers underreport corruption investigations involving high-level officials who are connected to the top official of their province by twice as much as officials who are of the same administrative rank, but less well connected.

Figure 1.7: Case Study: Wan Qingliang's Network

A. Before Wan Qingliang's investigation announcement



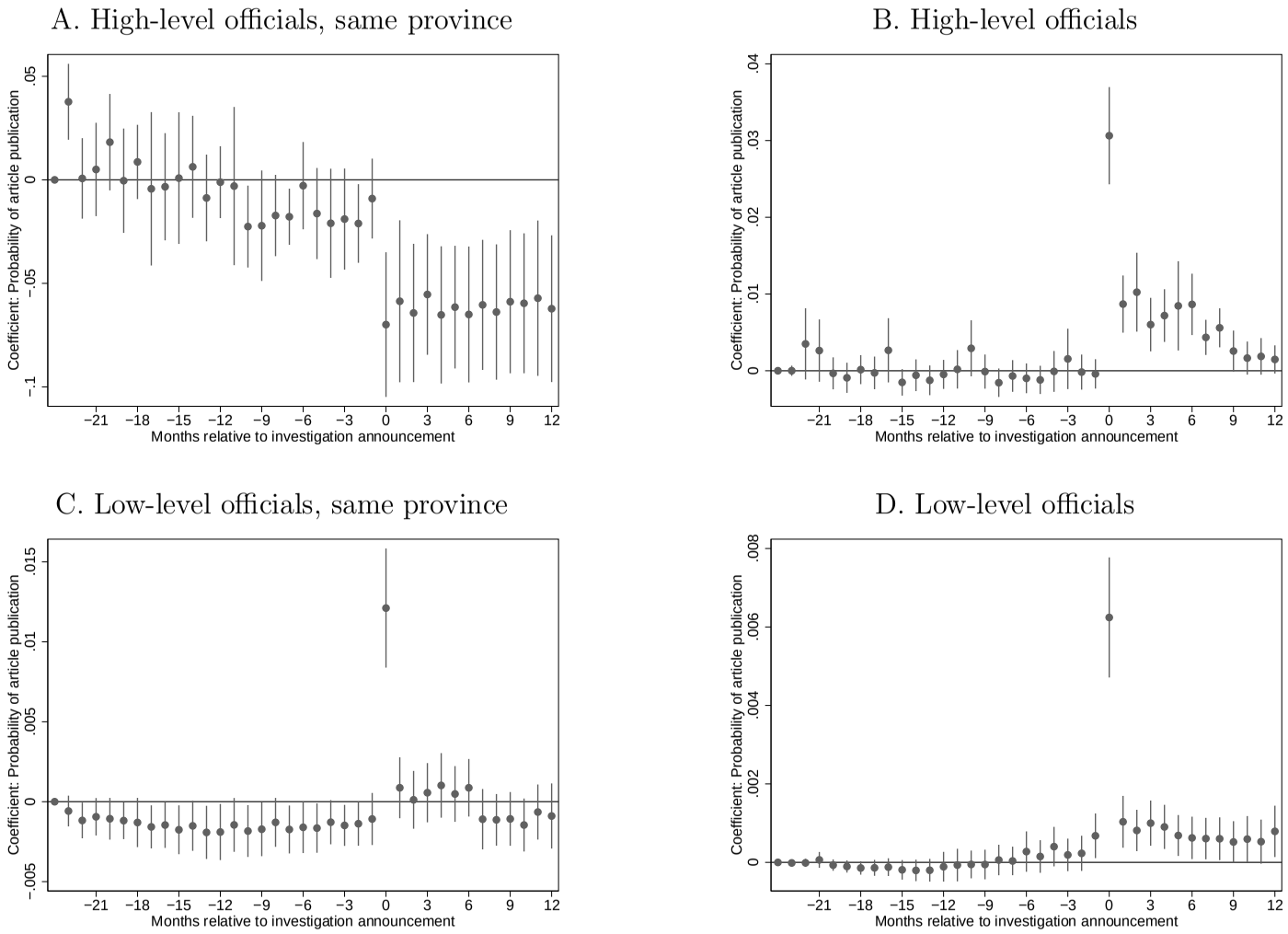
B. After Wan Qingliang's investigation announcement



Notes: These charts show coefficients from a regression of a dummy for whether an article is published on an indicator for the period after each official's corruption investigation ($Post_{o,t}$) and before Wan Qingliang's investigation announcement (Panel A) or after his investigation announcement (Panel B), interacted with two sets of dummy variables: 1) whether a newspaper is from Guangzhou city (Guangzhou NP) or from Guangdong province (other NPs) and 2) 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 1.2. 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 1.F5.

Before Wan Qingliang was investigated, newspapers from his city underreported investigations of his associates relative to other comparable officials and relative to other newspapers in the province. Following Wan Qingliang's own investigation, there were no longer any differences in how his city's newspapers report investigations of connected and unconnected officials and how they versus other newspapers report about connected officials.

Figure 1.8: Time Profile of Estimated Coefficients



Notes: 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- or low-ranking and whether a newspaper and official are from the same province. 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 1.2 except for month fixed effects due to collinearity. Standard errors are clustered by newspaper and official. Coefficient estimates are reported in Appendix Table 1.F6.
Investigation announcements do not appear to be anticipated by newspapers. The difference in reporting about officials from the same versus other provinces is only significant in the month of the announcement for low-level officials, but persists for the entire year post-announcement period for high-level officials.

1.10 Tables

Table 1.1: Extent of Underreporting of High-level Corruption Scandals by Newspapers from the Same Province Relative to Newspapers from Other Provinces

	Day 1 (1)	Days 2 to 10 (2)	Days 11 to 365 (3)
Fraction of own-province newspapers reporting	0.22	0.015	0.0081
Fraction of other-province newspapers reporting	0.33	0.030	0.0083
Underreporting by own-province relative to other-province newspapers (percentage of actual reporting by own-province newspapers)			
Equal reporting	52%	104%	3%
Proportional reporting benchmarks			
Low-level corruption scandals	606%	341%	615%
Low-level corruption scandals (adjusted for pre-scandal differences)	1224%	728%	1242%
Baidu searches	918%	899%	232%
Comments on Sina Weibo	138%	220%	62%

Notes: This table calculates the extent of underreporting of high-level corruption scandals by newspapers from the same province using a number of different benchmarks for three time periods following an investigation announcement: day 1 (column 1), days 2 to 10 (column 2) and days 11 to 365 (column 3) after the announcement. The first two rows show the fraction of newspapers that report about a high-level official from the same and other provinces, respectively. The bottom panel shows the percentage shortfall in reporting by own-province newspapers relative to different counterfactuals. 1) Equal reporting: own-province newspapers report the same (absolute) amount as out-of-province newspapers. In the following rows, the ratio of own-province to out-of-province reporting is assumed to be equal to various benchmarks. 2) Low-level corruption scandals: ratio of own-province to out-of-province reporting following investigation announcements into low-level officials. 3) Low-level corruption scandals (adjusted for pre-scandal differences): the above ratio multiplied by the pre-announcement ratio of own- vs. other-province for high- vs. low-level officials. 4) Baidu searches: ratio of the number of Baidu searches for high-level officials in their own relative to other provinces (time periods are calculated from day $t-1$). 5) Comments on Sina Weibo: ratio of the number of comments from users from the same province as an official compared to users from other provinces on newspapers' Sina Weibo posts about investigations of high-level officials (non time-variant, excludes user comments on newspapers from their own province). This counterfactual exercise provides simple extrapolations. As the dependent variable is the fraction of newspapers is reporting, the maximum value cannot be larger than 1.

Newspapers are less likely to report about high-level corruption scandals in their own province than in other provinces, in absolute terms. Implied underreporting increases when using different benchmarks to account for greater interest by newspaper readers in corruption in their local area.

Table 1.2: Underreporting of Local 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)
Seasonal controls		✓		✓
Official-newspaper-pair FE		✓		✓
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

Notes: 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- or low-ranking and different controls, see equation 1.1. All regressions include official-newspaper-pair fixed effects. Observations are at the official-newspaper-day level. Standard errors in parentheses are clustered by newspaper and official. Significance at the 1%, 5% and 10% is denoted by ***, ** and *, respectively.

When an official is investigated for corruption, more articles about them are published by newspapers from other provinces. Newspapers from the same province as a low-level official under investigation also publish more articles about them in the period after an investigation announcement, however, newspapers publish fewer articles about high-level officials from their home province post-announcement.

Table 1.3: Framing of Articles Post-Investigation Announcement

	(1) Log words	(2) Positive relative to negative words	(3) Multiple officials	(4) Citizen complaints	(5) Article headlines contains: Official announcement	(6) Article headlines contains: Corruption	(7) Article headlines contains: Anti-corruption campaign
Share of high-level officials from same province	-0.30***	6.75***	-0.13***	-0.055***	0.016	-0.058***	-0.035*
	(0.096)	(1.61)	(0.021)	(0.013)	(0.0094)	(0.018)	(0.018)
Share of low-level officials from same province	-0.031	1.90***	-0.059***	0.0046	0.010**	0.0083	-0.020**
	(0.057)	(0.69)	(0.012)	(0.011)	(0.0046)	(0.011)	(0.0084)
No. of obs.	18,458	18,458	18,458	18,458	18,458	18,458	18,458
R-squared	0.36	0.19	0.48	0.23	0.099	0.15	0.18
Mean dep. var.	6.46	3.49	0.31	0.13	0.0086	0.22	0.10

Notes: 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 1.2. The dependent variables are the following: 1) log word count article, 2) number of positive minus negative words contained in the article text, 3) page number of article, 4) dummy for whether article refers to more than one official, 5) dummy for whether an article discusses citizens' complaints and accusations of official corruption, 6) dummy for whether article headline copies official CCDI investigation announcement, 7) dummy for whether article headline contains references to corruption and 8) dummy for whether article headline includes keywords related to the anti-corruption campaign. Positive and negative words are classified using the National Taiwan University Semantic Dictionary, following pre-processing using jieba, see Appendix 1.B. All regressions include official and newspaper fixed effects, as well as seasonal controls from column 2 in Table 1.2. Observations are at the article level. Standard errors in parentheses are clustered by newspaper. Significance at the 1%, 5% and 10% is denoted by ***, ** and *, respectively.
These results suggest that newspapers deemphasise news about officials under investigation who are from their own province.

Table 1.4: Internet Searches for Complaints against Officials

	Baidu searches for non-anonymous complaints				
	(1)	(2)	(3)	(4)	(5)
Number of high-level officials from same province investigated	8.45*	1.01	1.14	1.29	1.32
	(4.69)	(7.88)	(7.91)	(7.97)	(7.94)
Number of low-level officials from same province investigated	2.45***	2.01*	2.58**	2.63**	2.66**
	(0.89)	(1.22)	(1.20)	(1.21)	(1.21)
Fraction of own-province newspapers reporting about scandals of...					
...High-level officials from same province		14.8*			
		(8.76)			
...Low-level officials from same province		17.2***			
		(3.55)			
...High-level officials from other provinces		4.31**			
		(2.12)			
...Low-level officials from other provinces		0.068			
		(2.26)			
Fraction of own-province newspapers publishing articles with...					
...references to citizen complaints			6.46*		
			(3.79)		
...“Corruption” in headline				6.88**	
				(2.98)	
...“Anti-Corruption Campaign” in headline					4.47
					(4.50)
No. of obs.	45,198	20,246	20,246	20,246	20,246
R-squared	0.48	0.55	0.55	0.55	0.55
Mean dep. var.	24.8	28.8	28.8	28.8	28.8

Notes: This table shows results from regressions of the number of daily Baidu searches in each province for non-anonymous complaints regressed on the number of high- and low-level officials from the same province who were placed under investigation on the same day. The subsequent columns add the following regressors based on the post-scandal publication behaviour of provincial newspapers from the same province in the Wisenews sample: fraction of newspapers publishing at least one article about high- and low-level officials from the same and other provinces (column 2), fraction of newspapers publishing articles whose text contains references to citizen complaints (column 3), whose headlines contain mentions of corruption (column 4) and the anti-corruption campaign (column 5). See also Table 1.3. All regressions include province and day fixed effects. Observations are at the province-day level. Column 1 includes all provinces, columns 2 to 5 include the 14 provinces in the Wisenews sample. Robust standard errors are in parentheses. Significance at the 1%, 5% and 10% is denoted by ***, ** and *, respectively.

There are more internet searches for filing complaints about officials when officials from the same province are investigated for corruption and when local newspapers report about these corruption cases.

Table 1.5: Comments on Sina Weibo Posts about High-Level Corrupt Officials

	Number of comments		Time to first comment	
	(1)	(2)	(3)	(4)
Post about high-level officials from same province	0.42**	0.42*	-48.2	-11.8
	(0.21)	(0.21)	(65.0)	(48.3)
Newspaper from same province	1.19*	1.19*	-85.9*	-25.6
	(0.60)	(0.61)	(46.2)	(20.5)
Seasonal controls	✓		✓	
User province FE	✓	✓	✓	✓
Newspaper FE	✓		✓	
Official FE	✓		✓	
Weibo post FE		✓		✓
No. of obs.	74,803	74,803	6,024	6,024
R-squared	0.092	0.34	0.24	0.79
Mean dep. var.	0.41	0.41	67.0	67.0

Notes: This table shows results from regressions of the number of Sina Weibo comments by users' province (columns 1 and 2) and time to the first comment in hours (columns 3 and 4) on a dummy variable for whether a high-level officials from the same province is mentioned and a dummy variable for whether the newspaper is from the same province. All regressions include fixed effects for the commenters' province and either official, newspaper and province fixed effects and seasonal controls (from Table 1.2 or Sina Weibo post fixed effects. The sample is restricted to users who report to be from mainland China. Results restricted to only users from the 14 provinces that are represented in the Wisenews newspaper sample are similar and are reported in Appendix Table 1.D2. Observations are at the Sina Weibo post - (user) province level. Standard errors in parentheses are clustered by newspaper. Significance at the 1%, 5% and 10% is denoted by ***, ** and *, respectively.

Sina Weibo users are significantly more likely to comment on posts by newspapers from their own province and posts about officials from their own province.

Table 1.6: Coefficients for All Newspaper and High-level Official Provinces

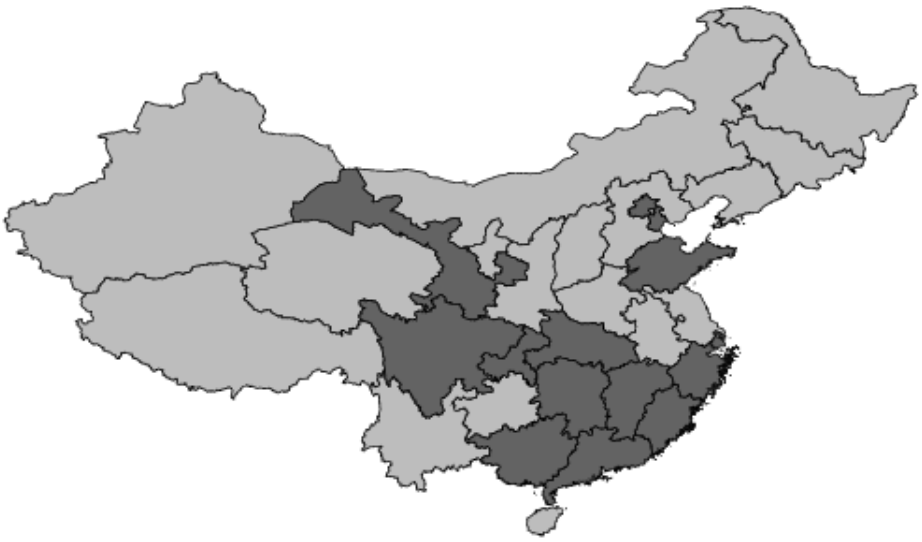
Official province	Newspaper province									
	Centre	Sichuan	Tianjin	Shandong	Guangdong	Guangxi	Jiangxi	Hubei	Hunan	Chongqing
Centre	0.008***	0.009***	0.011***	0.014***	0.014***	0.015***	0.017***	0.014***	0.014**	0.014***
	(0.003)	(0.002)	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)	(0.003)	(0.006)	(0.003)
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)

Notes: This table shows results from the regression of a dummy for whether an article is published on the interaction between a post-investigation dummy and indicators for each newspaper's (in columns) and high-level official's province (in rows), for provinces with at least one high-level official caught during the campaign. 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 1.2. There are 3,458,778 observations in this regression, the R-squared is 0.076 and the mean dependent variable is 0.006. Observations are at the official-newspaper-day level. Standard errors in parentheses are clustered by newspaper and official. Significance at the 1%, 5% and 10% is denoted by ***, ** and *, respectively.
Only the coefficients on the diagonal are negative, implying that only investigations of high-level officials from the same province are underreported.

Appendix

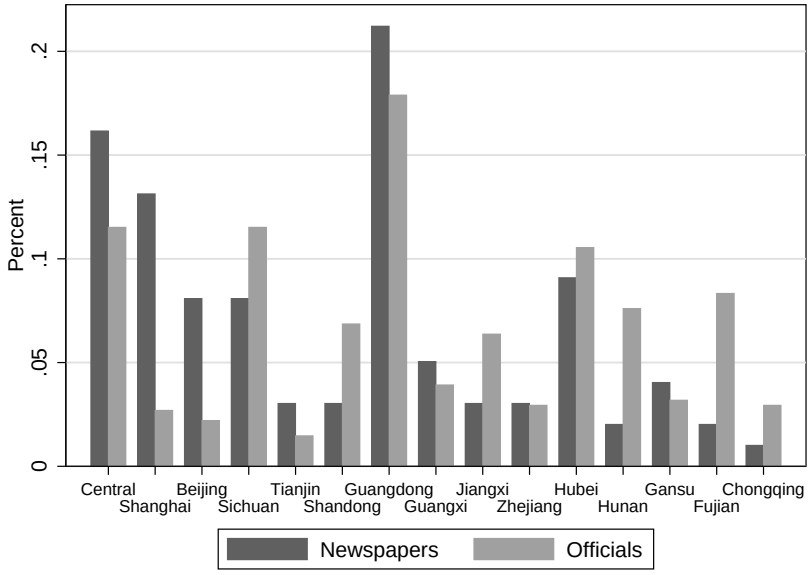
1.A Detailed Sample Description

Figure 1.A1: Map of Provinces in Sample



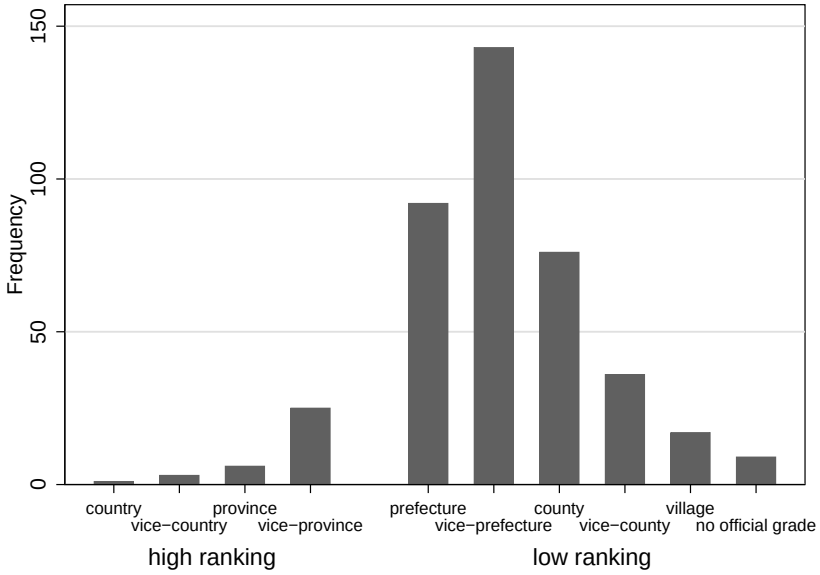
Notes: This map shows provinces and province-level municipalities in the newspaper sample in dark grey. Source: GADM version 3.6. See `www.gadm.org`. WGS84 projection.

Figure 1.A2: Geographic Distribution of Officials and Newspapers



Notes: This chart shows the distribution of officials and newspapers in the sample across provinces.

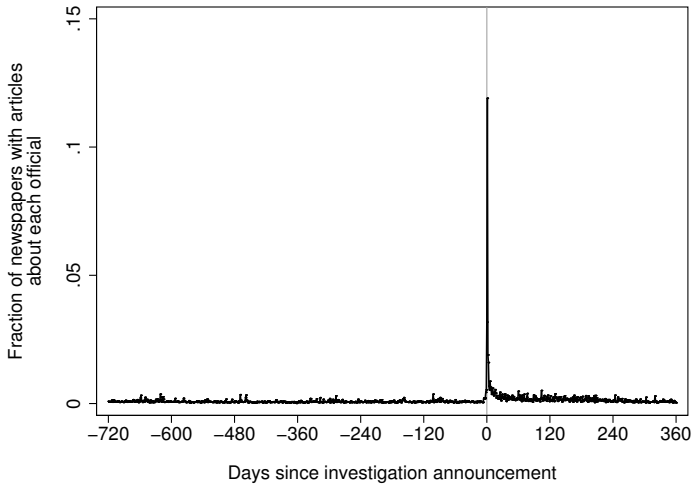
Figure 1.A3: Distribution of Officials' Ranks



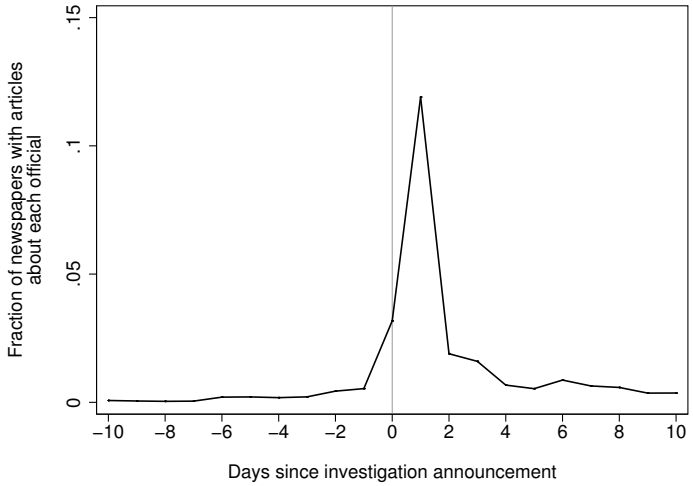
Notes: This chart shows the distribution of official ranks in the sample.

Figure 1.A4: Fraction of Newspaper with Articles about each Official

A. Full sample



B. Ten days before and after investigation announcement



Notes: These charts show the average fraction of newspapers that publish at least one article about each investigated official. The x-axis shows days before and after the announcement of an investigation into an official.

Table 1.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

Notes: 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 1.A2: Officials with No Articles or Only Articles Pre- or Post-Investigation

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

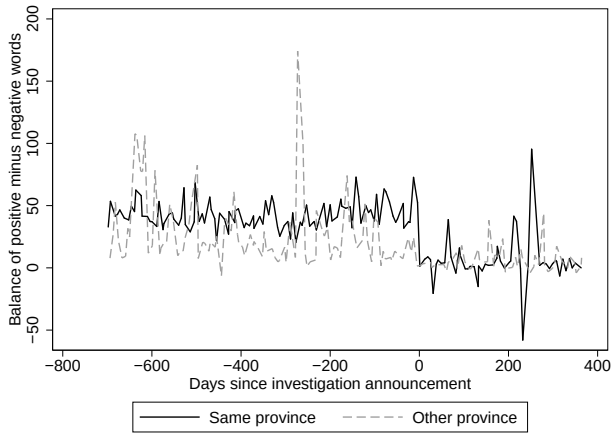
Notes: There are 12 officials with no articles in the sample, 124 officials with only articles following the corruption scandal and 17 officials with only articles before the corruption scandal. This table shows the probability an official belonging to each of these three cases regressed on official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

1.B Descriptive Text Analysis of Newspaper Articles

1.B.1 Sentiment Analysis

I adopt a lexicon approach to classify the sentiment expressed in each article. After cleaning and segmenting the text, I count number of positive and negative expressions in each article using the National Taiwan University Sentiment Dictionary (NTUSD) in simplified Chinese (Ku and Chen, 2007).[52] NTUSD contains 5,830 negative and 2,330 positive words and expressions and is widely used in the Chinese natural language processing literature. I use the difference between the number of positive and negative words in each article as my main measure of sentiment. Appendix Figure 1.B1 shows that the sentiment of articles about an official turn more negative after the official is investigated for corruption. This is the sentiment measure used in Column 2 of Table 1.3.

Figure 1.B1: Evolution of Sentiment Expressed in Articles



Notes: This chart shows the average balance of positive relative to negative words in articles on the y-axis and the days since an official is investigated on the x-axis. The solid black line shows sentiment in articles by newspapers from the same province as an official and the dashed grey line shows the sentiment in newspapers from other provinces. After an official's investigation is announced, articles about that official become more negative.

[52] I tokenise the text using Jieba (a Chinese text segmenter in Python) and remove non-Chinese characters and stopwords from the Baidu search engine stopwords list.

1.B.2 Content Analysis

I use an unsupervised generative machine learning model to extract the two main topics from the sample of newspaper articles. Specifically, I implement a latent Dirichlet allocation (LDA) model based on Blei, Ng and Jordan (2003), which is a commonly used model in text analysis (see, e.g. Blei, 2012; Gentzkow, Kelly and Taddy, 2019).[53] The resulting topic model clearly identifies two topics in the data: the first relates to development and the second to corruption. Appendix Figure 1.B2 shows the 30 words which are most closely associated with each topic. Figure 1.2 shows how the probability of articles being classified as "corruption" changes when an official is investigated.

Figure 1.B2: Word cloud of topic model terms

Topic 1: "Development"

Topic 2: "Corruption"

Notes: These charts show the 30 words which are most associated with each topic from a LDA model, translated into English. The keywords are scaled by their estimated probability of belonging to each topic.

[53] I use Python's Gensim LDA package which is based on the algorithm by Hoffman, Bach and Blei (2010). The text is first tokenised using Jieba (a Chinese text segmenter in Python), after which I remove non-Chinese characters and stopwords from the Baidu search engine stopwords list.

1.C Sina Weibo

I identify the official Sina Weibo account of each newspaper in the Wisenews archive. Out of 99 newspapers, 71 had an official Sina Weibo account and 61 posted continuously over the sample period. I use Sina Weibo's search function to search for the name of each high-level official in the posts by each newspaper and manually clean the data of false matches. The data were collected between 27 July 2019 to 5 August 2019. Unlike newspaper articles from the archive, social media posts can be deleted by newspapers at a later date. It is possible that newspapers delete previous posts they wrote about officials who were subsequently placed under investigation. In the following analysis, I restrict the sample to posts which were posted about officials after they were investigated, by newspapers that post continuously over the sample period. In total, there were 2,407 posts about 35 officials by 61 newspapers. I have information about the text of the post, the time of posting, the number of likes, forwards and reposts. Appendix Table 1.C1 shows summary statistics of newspapers' Sina Weibo accounts in Panel A and their posts in Panel B.

Table 1.C1: Summary Statistics of Sina Weibo Sample

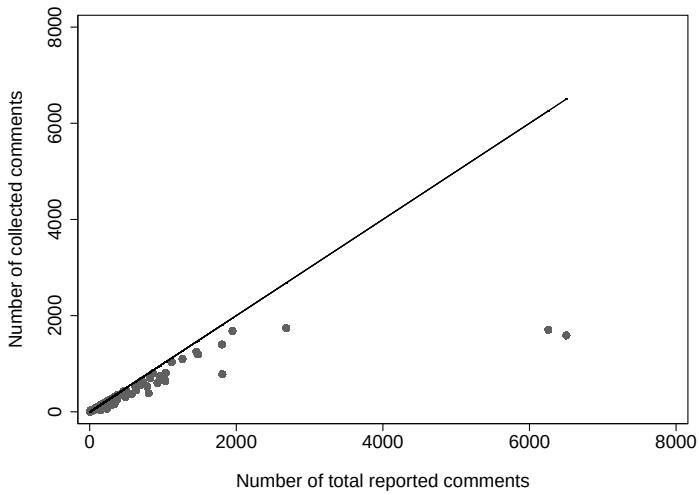
	mean	sd	min	max	obs
A. Newspapers' Sina Weibo Accounts					
Party NP	0.23	0.42	0	1	61
Govt-owned commercial NP	0.28	0.45	0	1	61
Subsidiary NP	0.49	0.50	0	1	61
Central NP	0.16	0.37	0	1	61
Provincial NP	0.52	0.50	0	1	61
Sub-provincial NP	0.31	0.47	0	1	61
Number of friends	1,201	1,259	41	8,259	61
Number of followers	8,737,136	14,309,986	10,240	92,136,679	61
Number of posts	73,830	38,476	1,978	182,801	61
B. Newspapers' Sina Weibo posts					
Number of comments	28.05	157.68	0	6258	2407
Number of forwards	75.35	452.60	0	16656	2407
Number of likes	25.31	119.20	0	2447	2407
C. Commenters on Sina Weibo posts					
Number of followers	83,088	2,156,174	0	92,023,217	30,430
Number of friends	574	874	0	20,000	30,430
Number of posts	6,104	17,128	0	697,415	30,430
Gender	0.78	0.42	0	1	30,430
Verified	0.10	0.30	0	1	30,430

Notes: This table reports summary statistics for newspapers' Sina Weibo accounts (Panels A), their posts about high-level corrupt officials (Panel B) and the users who comment on these posts (Panel C).

I collect the comments made on each post using Sina Weibo's API.[54] For each comment, I know the text of the comment and the time of commenting, as well as the username of the commenter, their self-reported province and city of residence, gender, whether the account was verified by Sina Weibo (generally indicating a public interest account), the date the account was created and the number of friends, followers and posts of that user. After removing suspected spam comments, there are a total of more than 30,000 comments on all posts, see Appendix Table 1.C1 for summary statistics.

Depending on users' privacy settings, the Sina Weibo API does not allow access to all comments. The API also has a rate limit of maximum 2000 comments per post, which is not binding for any of the posts in the sample. The number of comments reported on Sina Weibo and the number of comments I was able to collect are highly correlated, with a correlation coefficient of 0.82. Appendix Figure 1.C1 shows the correlation between both measures in graphical form. The discrepancy between both measures is largest for a small number of posts that have a very large number of comments.

Figure 1.C1: Reported versus Collected Comments on Sina Weibo



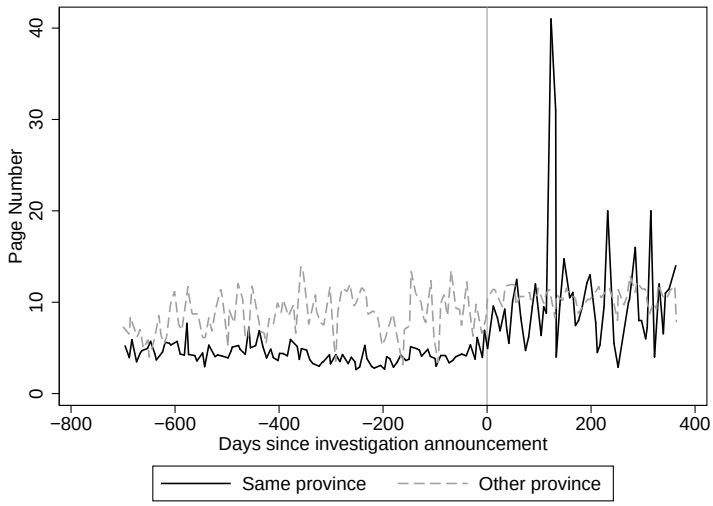
Notes: This chart shows a scatter plot of each Sina Weibo post with the number of comments I collected on each post on the y-axis and the number of comments that are reported for each post on the x-axis. The solid black line shows the 45 degree line where the numbers of comments collected and reported are the same.

[54] I access the API using the sinaweibopy Python interface.

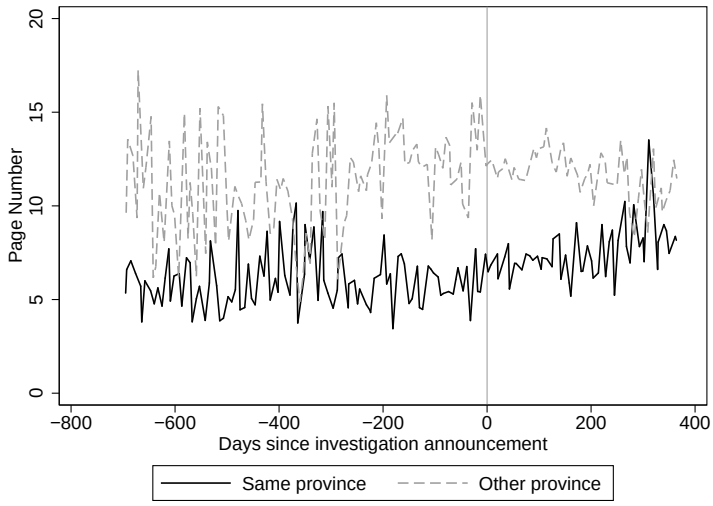
1.D Additional Figures and Tables

Figure 1.D1: Page Number of Articles

A. High-level officials

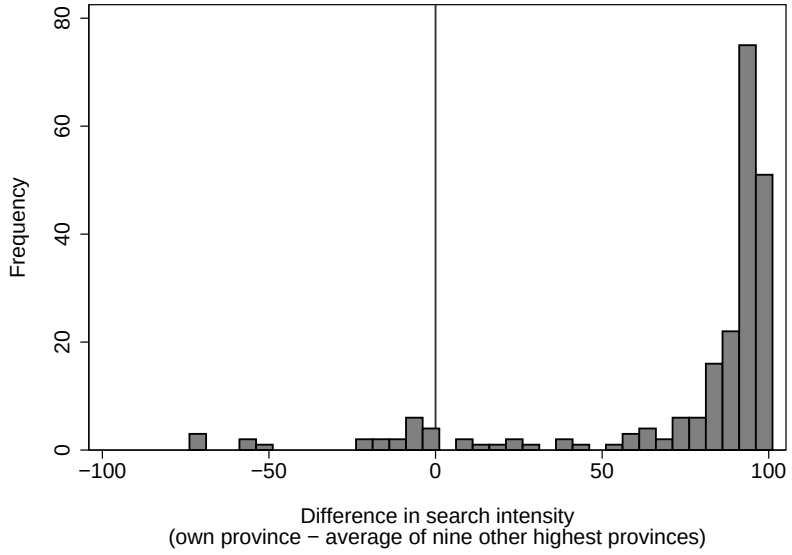


B. Low-level officials



Notes: These charts show the average page number of articles on the y-axis and the days before and after the investigation on the x-axis.
Before a high-level official is investigated for corruption, articles about them are placed further forward in their home-province's newspapers than in newspapers from other provinces. This relative gap in page numbers closes once a high-level official is under investigation. A similar, but less pronounced pattern holds for low-level officials.

Figure 1.D2: Search Intensity for Officials under Investigation on Baidu



Notes: This chart shows a histogram of the difference in Baidu search volume for an official in their own province relative to the average of the nine highest other provinces in the year following their investigation announcement (x-axis). A number greater than zero implies that there were on average 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 for the post-announcement year (data only available since September 2013). See Appendix Table 1.D1 for a regression analysis.
Individuals are more likely to search for details about officials from their own province who were investigated during the anti-corruption campaign.

Table 1.D1: Baidu Search Volume for Corrupt Officials by Province

	Full sample		Wisenews newspaper sample	
	(1)	(2)	(3)	(4)
High-level official from same province	83.5***	83.9***	83.7***	83.8***
	(1.94)	(5.69)	(1.96)	(6.10)
Low-level official from same province	78.8***	79.9***	79.0***	80.0***
	(2.70)	(2.67)	(2.74)	(2.72)
User province FE		✓		✓
Official FE		✓		✓
No. of obs.	2,417	2,417	1,723	1,723
R-squared	0.57	0.84	0.63	0.85
Mean dep. var.	23.2	23.2	26.1	26.1

Notes: This table shows results from regressions of an index of Baidu search volume by users' province on dummy variables for whether the high or low-level official is from the same province in the year after the official was investigated for corruption. Columns 2 and 4 include fixed effects for each official and province of Baidu user. Columns 1 and 2 show results for the full sample, columns 3 and 4 show results only for the 14 provinces that are represented in the Wisenews newspapers sample. Baidu search index was normalised to 100 for the province with the most searches. Data for the ten provinces with the largest search volumes are available. Observations are at the official-province level. Standard errors in parentheses are clustered by official. Significance at the 1%, 5% and 10% is denoted by ***, ** and *, respectively.

Users are more likely to search for officials from their own province. This effect is larger for high-level officials than low-level officials, but this difference is not statistically significant.

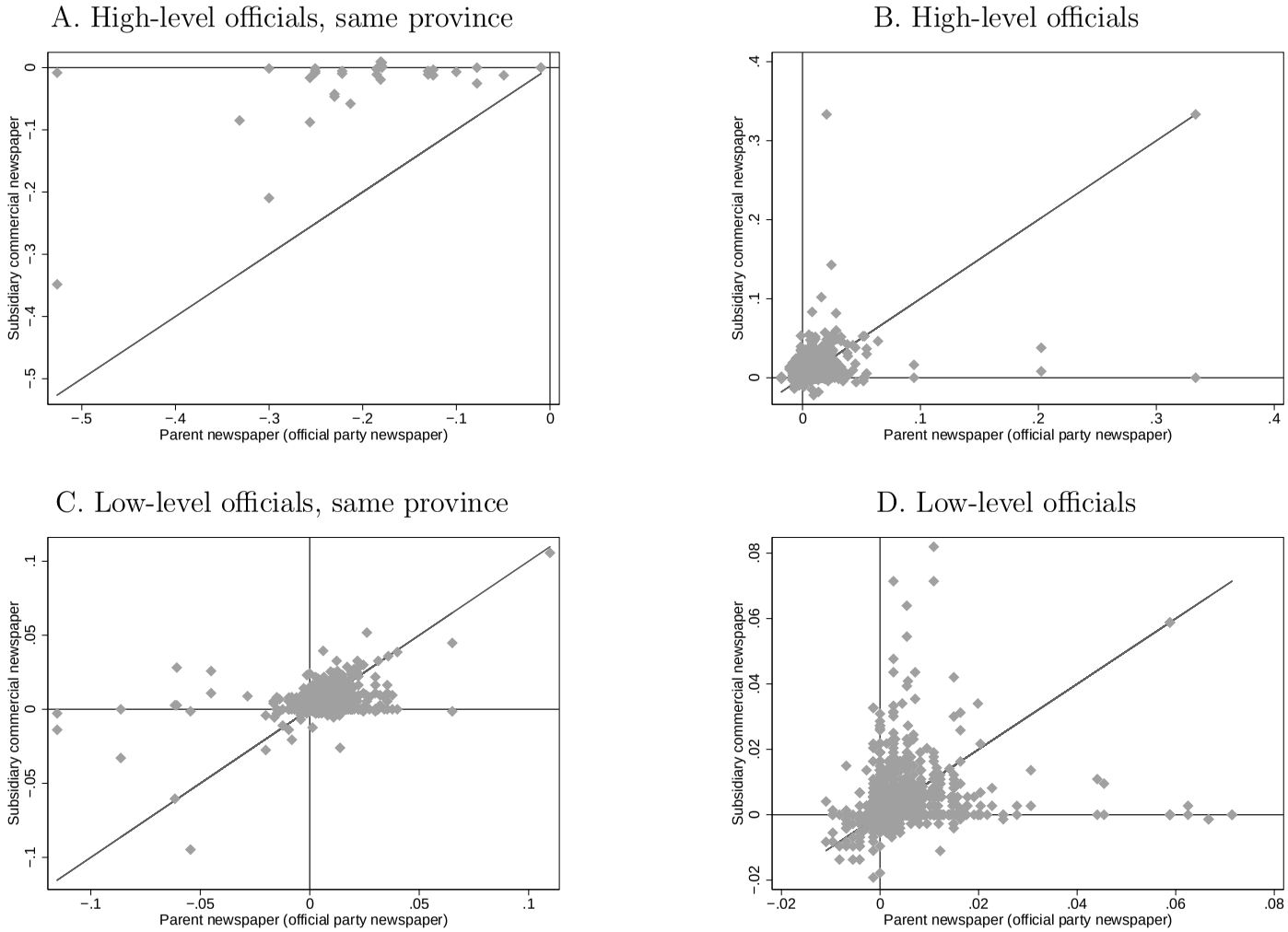
Table 1.D2: Comments on Sina Weibo Posts about High-Level Corrupt Officials: Restricted Sample of WiseNews Provinces

	Number of comments		Time to first comment	
	(1)	(2)	(3)	(4)
Post about high-level officials from same province	0.51**	0.50*	-23.0	-7.89
	(0.25)	(0.25)	(58.4)	(39.4)
Newspaper from same province	1.33**	1.33**	-64.3*	-22.9
	(0.54)	(0.56)	(37.5)	(20.0)
Seasonal controls	✓		✓	
User province FE	✓	✓	✓	✓
Newspaper FE	✓		✓	
Official FE	✓		✓	
Weibo post FE		✓		✓
No. of obs.	33,782	33,782	3,610	3,610
R-squared	0.11	0.44	0.25	0.71
Mean dep. var.	0.67	0.67	42.1	42.1

Notes: This table shows results from regressions of the number of Sina Weibo comments by users' province (columns 1 and 2) and time to the first comment in hours (columns 3 and 4) on a dummy variable for whether a high-level officials from the same province is mentioned and a dummy variable for whether the newspaper is from the same province. All regressions include fixed effects for the commenters' province and either official, newspaper and province fixed effects and seasonal controls (from Table 1.2 or Sina Weibo post fixed effects. All regressions include fixed effects for the commenters' province and either official, newspaper and province fixed effects, as well as seasonal controls from column 2 in Table 1.2 or Sina Weibo post fixed effects. The sample in this table is restricted to include only users who report to be resident in one of the 14 provinces that are represented in the Wisenews newspapers sample. Observations are at the Sina Weibo post - (user) province level. Standard errors in parentheses are clustered by newspaper. Significance at the 1%, 5% and 10% is denoted by ***, ** and *, respectively.

Sina Weibo users are significantly more likely to comment on posts by newspapers from their own province and posts about officials from their own province.

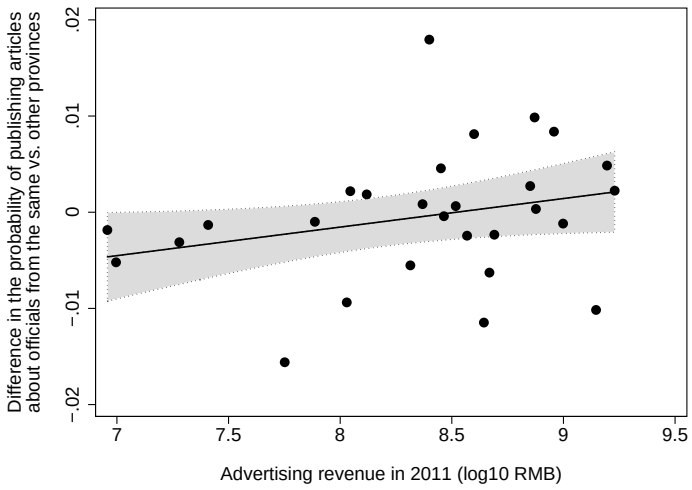
Figure 1.D3: Subsidiaries and their Parent Newspapers



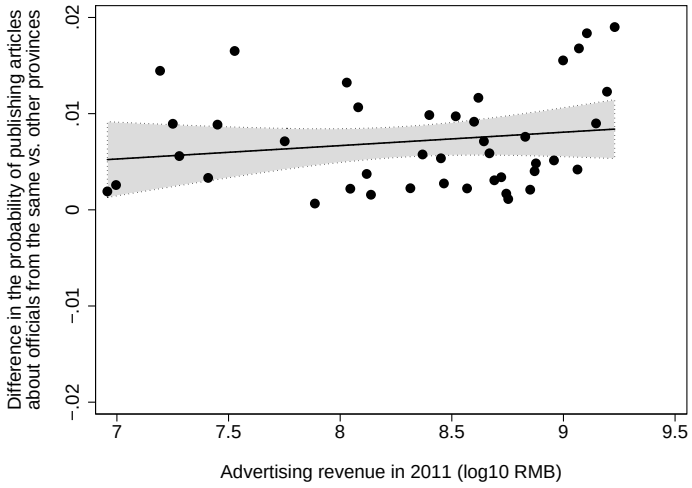
Notes: These charts show the difference in the mean daily probability of reporting about a given official before and after their corruption investigation announcement for a subsidiary newspaper (y-axis) relative to their parent newspaper (x-axis). Each point is a subsidiary-parent newspaper pair. A negative number means that reporting is lower after the announcement than before. The 45 degree line is shown for comparison.
All subsidiary newspapers underreport strictly less than their parent newspaper when high-level officials are investigated for corruption (all points lie above the 45 degree line). There are no clear systematic patterns when reporting about low-level officials from the same province or officials from other provinces.

Figure 1.D4: Advertising revenue and reporting about corruption scandals

A. High-level officials



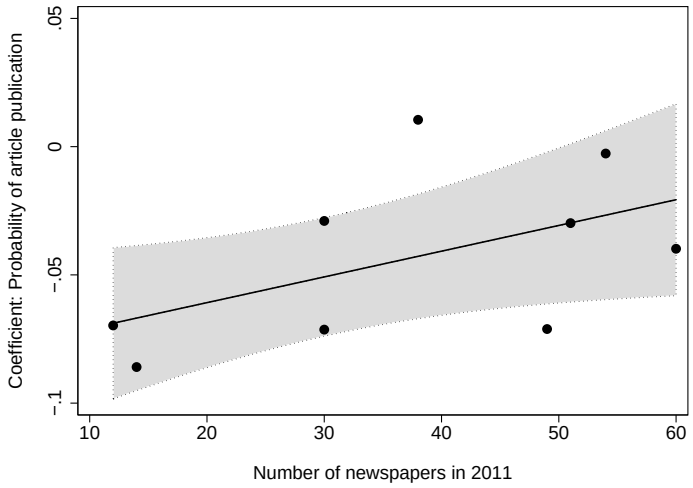
B. Low-level officials



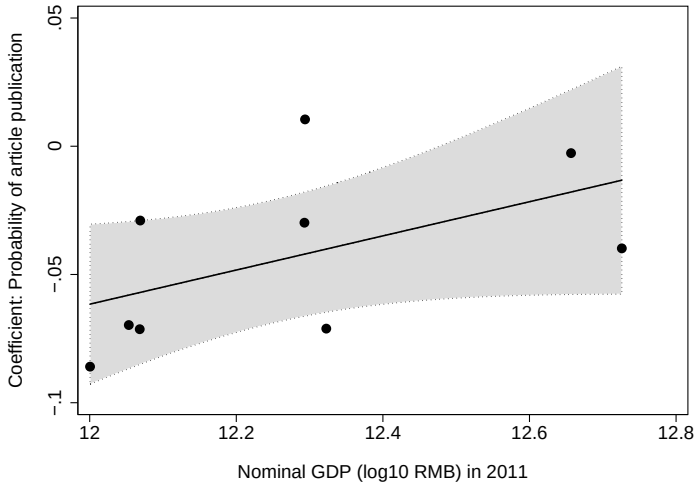
Notes: These charts show the difference between the average daily probability that a newspaper publishes an article about investigations of officials from their own province relative to officials from other provinces on the y-axis and the newspaper's advertising revenue in 2011 (in log10 RMB, using data from Qin, Strömberg and Wu (2018). The black line shows the fitted values from a linear regression with the 95% confidence interval implied by robust standard errors shown in grey.
Newspapers with higher advertising revenues underreport corruption scandals involving high-level officials from their own province relative to those from other provinces. There is no relationship between advertising revenue and relative reporting for low-level officials.

Figure 1.D5: Heterogeneity in underreporting of high-level corruption scandals across province

A. Number of newspapers

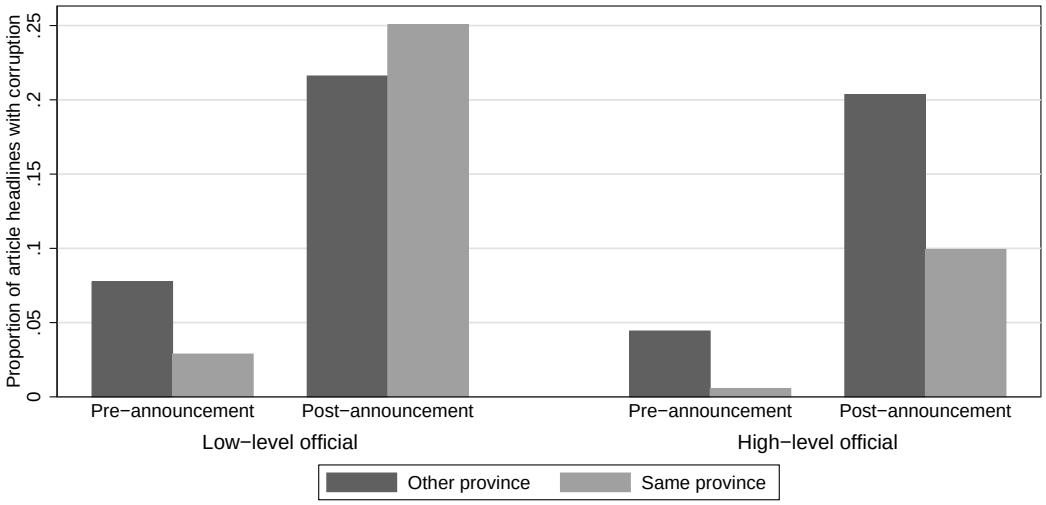


B. GDP



Notes: These charts show on the y-axis the estimated coefficients on the interaction between a post-investigation indicator and a dummy for whether a newspaper and a high-level official are from the same province separately by province. The regression uses a dummy variable for whether an article is published as the dependent variable and controls for seasonal and official-newspaper pair fixed effects. On the x-axis are the number of newspapers in each province in 2011 using data from Qin, Strömberg and Wu (2018) in Panel A and the nominal GDP of each province (in log10 RMB) using data from the National Bureau of Statistics in 2011 in Panel B. The black line shows the fitted values from a linear regression with the 95% confidence interval implied by robust standard errors shown in grey.
Newspapers are underreport corruption scandals involving high-level officials from their own province less in provinces with more newspapers and higher GDP.

Figure 1.D6: Share of Articles with "Corruption" in the Headline



Notes: This chart shows the fraction of articles that have a headline which includes the word 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.
Pre-announcement, fewer articles' headlines include the word corruption when they were published by newspapers from the same province as an official. Newspapers did not appear to systematically uncover corruption scandals of officials from their own province before an anti-corruption campaign investigation.

1.E Additional Robustness Check Tables

Table 1.E1: 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

Notes: The first two columns of this table show results from regressions of number of articles published on interactions between dummies for the post-investigation period, whether a newspaper and official are from the same province, whether an official is high- or low-ranking using the Poisson QML estimator (see Section 1.7). 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 1.2. Observations are at the official-newspaper-day level. Standard errors in parentheses are clustered by newspaper and official. Significance at the 1%, 5% and 10% is denoted by ***, ** and *, respectively.

The main result that investigations of high-level officials are underreported in newspapers from the same province remains robust to using a Poisson model.

Table 1.E2: Underreporting of Local Corruption Scandals: Excluding Cases of Potential Reverse Causality

	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

Notes: This table shows results from a regression 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- or low-ranking. The sample excludes 42 officials with at least one article published in a same-province newspaper before the investigation, whose the headline includes the word corruption.The regression includes official-newspaper-pair fixed effects and all seasonal controls as in Table 1.2. Observations are at the official-newspaper-day level. Standard errors in parentheses are clustered by newspaper and official. Significance at the 1%, 5% and 10% is denoted by ***, ** and *, respectively.

The main result that investigations of high-level officials are underreported in newspapers from the same province remains robust to excluding officials whose investigations were potentially uncovered by newspapers from the same province before the anti-corruption campaign investigation.

Table 1.E3: Underreporting of Local Corruption Scandals: ChinaFile Definition of High- versus 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

Notes: This table shows results from a regression 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- or low-ranking using an alternative definition according to ChinaFile. The regression includes official-newspaper-pair fixed effects and all seasonal controls as in Table 1.2. Observations are at the official-newspaper-day level. Standard errors in parentheses are clustered by newspaper and official. Significance at the 1%, 5% and 10% is denoted by ***, ** and *, respectively.

The main result that investigations of high-level officials are underreported in newspapers from the same province remains robust to using the alternative ChinaFile definition of high- and low-ranking officials.

1.F Regression Tables for Figures in Main Text

Table 1.F1: Municipality-Level Underreporting

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

Notes: This table shows results from a regression of a dummy for whether an article is published on an indicator for the post-scandal period ($Post_{o,t}$) and its interactions with two dummy variables that take the value of 1 if an official and a newspaper are from the same city ($Post_{o,t} \times SameCity_{o,n}$) and if an official and a newspaper are from different cities of the same province ($Post_{o,t} \times DifferentCity_{o,n}$), which are partly shown in Figure 1.3. The sample is restricted to city-level officials and newspapers. The p-value of a Wald-test for equality between the coefficients on $Post_{o,t} \times SameCity_{o,n}$ and $Post_{o,t} \times DifferentCity_{o,n}$ is reported at the bottom of the table. The regression includes official-newspaper-pair fixed effects and all controls from column 2 in Table 1.2. Observations are at the official-newspaper-day level. Standard errors in parentheses are clustered by newspaper and official. Significance at the 1%, 5% and 10% is denoted by ***, ** and *, respectively.

City-level newspapers underreport corruption scandals involving officials from the same city, but not those involving officials from other cities in the same province.

Table 1.F2: Party Newspapers versus Commercial Newspapers

	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

Notes: 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) and dummies for the post-scandal period, whether a newspaper and official are from the same province and whether an official is high- or low-ranking, which are partly shown in Figure 1.5. NP stands for newspaper. 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 1.2. Observations are at the official-newspaper-day level. Standard errors in parentheses are clustered by newspaper and official. Significance at the 1%, 5% and 10% is denoted by ***, ** and *, respectively.
Provincial newspapers are underreport investigations into high-level officials from the provincial government to a greater extent than sub-provincial newspapers from the same province. Provincial newspapers publish more articles about low-level provincial officials than sub-provincial newspapers.

Table 1.F3: Provincial Newspapers versus Sub-Provincial Newspapers

	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

Notes: 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 1.4. The sample is restricted to provincial government officials and local newspapers. NP stands for newspaper. 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 1.2. Observations are at the official-newspaper-day level. Standard errors in parentheses are clustered by newspaper and official. Significance at the 1%, 5% and 10% is denoted by ***, ** and *, respectively.

Provincial newspapers are underreport investigations into high-level officials from the provincial government to a greater extent than sub-provincial newspapers from the same province. Provincial newspapers publish more articles about low-level provincial officials than sub-provincial newspapers.

Table 1.F4: Proximity to the Head of the Province

	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

Notes: 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 has served on the party standing committee with the incumbent party secretary of the province (served versus not served) and dummies for the post-investigation period, whether a newspaper and official are from the same province and whether an official is high-ranking, which are partly shown in Figure 1.6. 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 1.2. Observations are at the official-newspaper-day level. Standard errors in parentheses are clustered by newspaper and official. Significance at the 1%, 5% and 10% is denoted by ***, ** and *, respectively.

Newspapers underreport corruption investigations involving high-level officials who are connected to the top official of their province by twice as much as officials who are of the same administrative rank, but less well connected.

Table 1.F5: Case Study: Wan Qingliang's Network

	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

Notes: This table shows results from a regression of a dummy for whether an article is published on an indicator for the period after each official's corruption investigation ($Post_{o,t}$) and before Wan Qingliang's investigation announcement (Panel A) or after his investigation announcement (Panel B), interacted with two sets of dummy variables: 1) whether a newspaper is from Guangzhou city (Guangzhou NP) or from Guangdong province (other NPs) and 2) whether an official is connected to Wan Qingliang (Wan OF) or other Guangdong officials of the same ranks (other OF), as shown in Figure 1.7. 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 1.2. Observations are at the official-newspaper-day level. Standard errors in parentheses are clustered by newspaper and official. Significance at the 1%, 5% and 10% is denoted by ***, ** and *, respectively.

Before Wan Qingliang was investigated, newspapers from his city underreported investigations of his associates relative to other comparable officials and relative to other newspapers in the province. Following Wan Qingliang's own investigation, there were no longer any differences in how his city's newspapers report investigations of connected and unconnected officials and how they versus other newspapers report about connected officials.

Table 1.F6: 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)
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)

Table continued on next page.

Table 1.F6: Coefficient estimates of dynamic model

Month	High-level official x Same province	Low-level official x Same province	High-level official	Low-level official
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)

Notes: 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- or low-ranking and whether a newspaper and official are from the same province, see Figure 1.8. 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 1.2 except for month fixed effects due to collinearity. Observations are at the official-newspaper-day level. 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. Investigation announcements do not appear to be anticipated by newspapers. The difference in reporting about officials from the same versus other provinces is only significant in the month of the announcement for low-level officials, but persists for the entire year post-announcement period for high-level officials.

Table 1.F6: 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)
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)

Table continued on next page.

Table 1.F6: Coefficient estimates of dynamic model

Month	High-level official x Same province	Low-level official x Same province	High-level official	Low-level official
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)

Notes: 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- or low-ranking and whether a newspaper and official are from the same province, see Figure 1.8. 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 1.2 except for month fixed effects due to collinearity. Observations are at the official-newspaper-day level. 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. Investigation announcements do not appear to be anticipated by newspapers. The difference in reporting about officials from the same versus other provinces is only significant in the month of the announcement for low-level officials, but persists for the entire year post-announcement period for high-level officials.

Chapter 2

Anti-Japanese Protests, Social Media Hate Speech and Television Shows in China

I am deeply grateful to Ekaterina Zhuravskaya for her advice and guidance. I also wish to thank Matthew Gentzkow, Emeric Henry, Philipp Ketz, Jonathan Lehne, Hillel Rapoport, David Strömberg, Liam Wren-Lewis and participants at the Paris School of Economics, CES North America Conference 2018, Stanford University, CEPREMAP India-China Conference, Rencontres d'Aussois, 4th Economics of Media Bias Workshop,Toulouse School of Economics, Barcelona GSE Summer Forum and TILEC-GovReg Workshop on Governance of Big Data and AI for valuable comments. I gratefully acknowledge the funding provided by Labex OSE, G-MonD and a French government subsidy managed by the Agence Nationale de la Recherche under the framework of the Investissements d'avenir programme reference ANR-17-EURE-001.

Abstract

Can stereotypes in entertainment media explain the recent rise in negative sentiment against foreigners? Despite close economic ties, anti-Japanese sentiment is high in China. I assemble detailed information on Chinese TV broadcasts during 2012 and document that around 20 percent of all TV shows aired during prime time were historical TV dramas set during the Japanese occupation of China during World War II. To identify the causal effect of media on sentiment, I exploit high-frequency data and exogenous variation in the likelihood of viewing Sino-Japanese war dramas due to channel positions and substitution between similar programmes. I show that exposure to these TV shows lead to a significant increase in anti-Japanese protests and anti-Japanese hate speech on social media across China. These effects are driven by privately rather than state-produced TV shows.

Keywords: Protests, Social Media, Ethnic Prejudice, Television

JEL Codes: L82, C55, D74.

2.1 Introduction

Media has long been used to influence individuals' beliefs and behaviour (see, e.g., DellaVigna and Gentzkow, 2010). Partisan news coverage can affect voting outcomes (see, e.g., DellaVigna and Kaplan, 2007; Enikolopov, Petrova and Zhuravskaya, 2011). State-sponsored propaganda can strengthen regime support and lead to violence against specific groups (see, e.g., Adena et al., 2015; Yanagizawa-Drott, 2014). And while not typically designed for this purpose, entertainment media can have impacts on a range of social and political outcomes (see, e.g., La Ferrara, Chong and Duryea, 2012; Kearney and Levine, 2015; Durante, Pinotti and Tesei, 2019). Stereotypical and biased portrayals of particular groups are common across entertainment products and I show in this paper that they can have harmful consequences.

I study the effect of entertainment media on ethnic prejudice by showing how TV soaps increase anti-Japanese protests and hate speech on social media in China. Anti-Japanese sentiment in China has its roots in the Japanese occupation of China during World War II and is widespread even among younger generations. While China and Japan have strong economic ties today, this resentment continues to adversely affect economic exchange (see, e.g., Fisman, Hamao and Wang, 2014; Che et al., 2015). In recent years, territorial disputes over the Senkaku/Diaoyu islands in the East China Sea have led to numerous anti-Japanese protests in China and incidents around the islands involving both countries' military. I collect detailed data on the programming schedule of all major Chinese TV channels

in 2012 and data on TV soap content and producers from official approval forms for TV show production. Around one fifth of all TV soaps aired in 2012 were dramas set during the Sino-Japanese war with highly negative depictions of Japanese soldiers. The share of distribution licenses awarded to this type of shows (as a fraction of all domestically produced TV shows) increased by approximately eight fold from 2004 to 2011. This increase mirrors trends in anti-Japanese sentiment in China according to survey data (see Figure 2.1).

These television dramas are intended for entertainment rather than as government propaganda. More than three quarters of these dramas are produced by private companies rather than the government. As all provincial TV channels are available throughout the whole country, there is no evidence that they specifically target their programming to their local audience. In particular, I show that broadcasts of Sino-Japanese war dramas are uncorrelated with provinces' experience during the war, economic ties to Japan, TV and internet usage and a wide range of other local factors.

I identify the effect of these Sino-Japanese war TV dramas on anti-Japanese sentiment by combining exogenous geographic variation in the intensity of exposure to these shows with a high-frequency empirical analysis. I use two sources of variation in likelihood of viewing a historical drama: the order of TV channels (following Martin and Yurukoglu, 2017, for the US) and the number of TV shows that are on neighbouring channels, which captures a substitution effect between TV shows. This approach takes into account selection of audiences into different TV programmes and the estimated effects are based on those marginal viewers that have been moved

to watch Sino-Japanese war dramas due to factors that are orthogonal to their prior anti-Japanese sentiment.

I use high-frequency data on the occurrence of anti-Japanese protests and anti-Japanese hate speech on Sina Weibo (the most popular Chinese micro-blogging platform). This rules out any concerns that the war dramas could be strategically scheduled to coincide with unexpectedly high anti-Japanese sentiment in the population, as TV schedules are set at least one week in advance. I account for expected increases in tensions, by controlling for important Sino-Japanese war anniversaries and the intensity of the Senkaku/Diaoyu island conflict by using the number of articles about the islands in the People's Daily (the main Chinese Communist Party newspaper).

Greater exposure to TV dramas set during the Sino-Japanese war increases the likelihood of anti-Japanese protests in China. Using data from Wallace and Weiss (2015), I construct a province-day panel of anti-Japanese protests and find that a one standard deviation increase in predicted historical TV drama viewership the night before, increases the probability of an anti-Japanese protest in the province by around 0.5 percentage points, which is a 40 percent increase in the mean probability of protests. Privately produced war TV dramas represent almost 80 percent of all war dramas in the sample and drive the effects on anti-Japanese protests. Areas that were occupied and suffered more civilian casualties during the Sino-Japanese war respond more when being exposed to these TV dramas.

Using a random sample of posts provided by Hong Kong University's

Weiboscope project (Fu, Chan and Chau, 2013), I show that higher predicted viewership of Sino-Japanese war TV dramas significantly increases anti-Japanese hate speech on social media. This increase is not purely driven by users directly discussing these TV show. Exposure to these TV shows lead to an increase in nationalist sentiment expressed on social media, more discussion of the Senkaku/Diaoyu island conflict and calls for boycott of Japanese goods. The effect of TV shows on anti-Japanese and nationalist sentiment is driven by users writing new posts rather than reposting or forwarding content by other users. Similar to the protest results, the effect is driven by privately rather than state produced TV shows.

As a placebo test, I shows that there are no effects of future predicted historical TV drama viewership on protests and social media hate speech. All results remain robust to excluding broadcasts by a province's home television station. I also present an event study type analysis of the social media data at a 15-minute frequency.

This paper contributes to a literature on the different social and political impacts of the media (for a review of the literature, see DellaVigna and La Ferrara, 2015). Media content can be explicitly designed to influence people's attitudes and behaviours, such as in the case of partisan news and government propaganda. Martin and Yurukoglu (2017); DellaVigna and Kaplan (2007); Enikolopov, Petrova and Zhuravskaya (2011) show how partisan TV channels affected electoral outcomes in the US and Russia. In this paper, I focus on ethnic prejudice and protests as main outcomes. Yanagizawa-Drott (2014); Adena et al. (2015) show that propaganda can be effective in inciting ethnic resentment and violence, in the context of the

Rwandan Genocide and Nazi Germany, while nationalistic radio broadcasts can trigger a backlash among unintended recipients (DellaVigna et al., 2014).

This paper studies the effects of entertainment media. While the primary purpose of entertainment media is profit, studies have shown that it can have other unintended outcomes, such as on divorce (Chong and La Ferrara, 2009), fertility (La Ferrara, Chong and Duryea, 2012), teenage pregnancy (Kearney and Levine, 2015) and political attitudes (Durante, Pinotti and Tesei, 2019). I show how negative depictions of particular ethnic groups in entertainment TV programmes can lead to increased ethnic prejudice and protests. This type of bias is far from unique to Chinese historical TV dramas. One methodological contribution of this paper is to exploit the intensive margin variation in TV consumption beyond coverage.

This paper is also related to a growing literature on social media (see, e.g. Zhuravskaya, Petrova and Enikolopov, 2020, for a recent review). In developed democracies, much of the focus rests on how social media might promote populism, in particular with the help of fake news (see, e.g., Allcott and Gentzkow, 2017; Allcott, Gentzkow and Yu, 2019; Vosoughi, Roy and Aral, 2018). The subject of this paper is hate speech on social media, in the form of expressions of xenophobia. Indeed, papers, such as Bursztyn et al. (2019); Müller and Schwarz (2019), find an empirical link between social media use and hate crimes. In an autocracy like China, the government intervenes in social media, by censoring sensitive posts (see, e.g., King, Pan and Roberts, 2013; Fu, Chan and Chau, 2013) and creating fake posts (see, e.g., King, Pan and Roberts, 2017), but can also use social media actively to

monitor its citizens (see, e.g. Qin, Strömberg and Wu, 2017). In contrast, anti-Japanese sentiment is topic that generally does not attract attention from the government.

This paper also contributes to a literature about the determinants and persistence of historical prejudices (e.g., Voigtländer and Voth, 2012) and how history can be made more salient by current events (e.g., Fouka and Voth, 2016; Ochsner and Roesel, 2017). The effects of nationalism in economic exchange have been previously examined by Pandya and Venkatesan (2016); Guiso, Sapienza and Zingales (2009), while Fisman, Hamao and Wang (2014); Che et al. (2015) studied the effects of the Sino-Japanese conflict on current economic outcomes. The findings of this paper suggest that the media can help explain the persistence of anti-Japanese sentiment in China.

In the next sections of the paper, I provide a brief background on Sino-Japanese relations and the media environment in China, followed by a description of the data in Section 2.3. Section 2.4 presents the empirical strategy and Section 5 the results of the paper. Section 2.7 concludes.

2.2 Background

2.2.1 Relationship between China and Japan

Brief history Despite close economic ties, political relations between China and Japan remain tense. The two countries have fought two wars

in recent history. The First Sino-Japanese War (1894-1895) ended in defeat for the Qing dynasty of China, which was forced to pay reparations and cede multiple territories (most notably, Taiwan) to Japan. The Second Sino-Japanese War (1937-1945) or the Pacific War of World War II began after Japan invaded Beijing from Japanese-occupied Manchuria in northeastern China and ended with Japan's surrender in 1945. The war cost China an estimated 20 million casualties and a number of Japanese soldiers were subsequently convicted for war crimes, including the mass murder and rape of civilians and the use of chemical weapons. Between 1949 and 1972, Communist China was internationally isolated and had no official relationship with Japan. In the following decades, as China underwent economic reforms, large sums of Japanese overseas development aid and investment flowed into China. Today, the two countries' economies are heavily dependent on each other.

Anti-Japanese sentiment in China According to the 2013 PEW survey, 78 percent of Chinese people feel that Japan has not apologised sufficiently for its actions during the war. In recent years, anti-Japanese protests broke out in China over new Japanese textbooks that were seen as whitewashing Japanese war crimes and Japanese officials' visits to the Yasukuni Shrine, which commemorates soldiers who fought in the war. The persistent war memory has significant negative economic consequences even today. Che et al. (2015) find that Chinese regions with higher casualties during the war trade less with Japan and receive less FDI from Japan. Fisman, Hamao and Wang (2014) show that Chinese and Japanese firms with high exposure to the other country suffered large stock market losses

at the time of anti-Japanese protests in China in 2005 and 2010.

Senkaku/Diaoyu island conflict A critical issue for the relationship between the two countries is the question of sovereignty over eight uninhabited islands in the East China Sea, called Diaoyu in Chinese and Senkaku in Japanese. These islands are currently under Japanese control, but are also claimed by China (and Taiwan). This dispute has simmered relatively quietly for a number of years, but tensions between the two countries have been on the rise over the past decade. In 2010, the collision of a Chinese trawler and the Japanese coastguard in the waters near the disputed islands led to public protests in both countries. Two years later marked another high point in tensions. In April 2012, the right-wing Governor of Tokyo announced his decision to purchase the islands with public money from their private owner – a move that was interpreted by China and Taiwan as a provocation. In July, Taiwanese and Japanese coastguards came to a standoff near the Senkaku/Diaoyu islands. In a bid to resolve the situation, the Japanese prime minister announced that the Japanese government will purchase the islands, which was seen by the Chinese side as an effort to nationalise the disputed islands. Activists from Hong Kong landed on the islands on August 15 and were detained by the Japanese. On September 11, Japan officially purchased the islands, China sent in patrol ships. Since then, neither side has backed down and continued provocations (often using military means) have led to increasing fears of a military escalation.

Anti-Japanese protests and boycott of Japanese goods in China The Senkaku/ Diaoyu island conflict sparked extensive anti-Japanese demon-

strations across China, boycotts of Japanese goods and closures of Japanese businesses in China when the protests turned increasingly violent. Over two thirds of Chinese cities saw at least one anti-Japanese protest during August and September and 75 of these protests had more than 1000 participants (Wallace and Weiss, 2015). According to estimates by the Japanese government, the direct cost of these protests reached $126 million. More economically damaging were the concurrent consumer boycotts of Japanese goods, which are estimated to have lowered Japanese exports by $3.5 billion (Heilmann, 2016). Social media potentially played an important role in the coordination of these boycotts. Calls for boycotts and blacklists of Japanese brand names were circulated widely during the summer, see Figure 2.A3.

The anti-Japanese protests in China did not appear to have a clear unified aim. According to anecdotal accounts and social media posts, many protesters participated to vent their anger and frustrations at Japan's actions over the Senkaku/Diaoyu islands and to express their own patriotism (see, for instance, the model by Passarelli and Tabellini, 2017). Calls for protests and boycotts also emphasise the importance of drawing attention to the issue and putting (economic) pressure on Japan to change its policies. Zhou and Wang (2016) finds that belief in instrumental reasons increase willingness to participate in future protests, but are not correlated with reported participation in past protests.[1] Some of the protests were also aimed at the Chinese government with demands that the government take

[1] A significant predictor of reported participation in past and future protests is knowing someone who has participated in an anti-Japanese demonstration, suggesting an important network and coordination aspect.

a harder stance with Japan and support the anti-Japanese protests.[2]

Official position of the Chinese government The Chinese Communist Party (CCP) legitimises its rule in part by appealing to nationalist sentiments. The official narrative emphasises how a weak imperial China and the corrupt Nationalist Party (Kuomingtang or KMT) allowed China to be carved up by foreign powers during the "century of humiliation", while the CCP defended China from Japan during the Sino-Japanese war and made China into an important global power. However, as a rule, the Chinese government discourages all forms of demonstrations and anti-Japanese protests are no exception. Foley, Wallace and Weiss (2018) find that party leaders of cities where spontaneous protests took place in 2012 were subsequently less likely to be promoted. In a survey of Chinese university students, Zhou and Wang (2016) find that members of the CCP were less likely to participate in these protests. Throughout the entire protest period, the People's Daily (official newspaper of the CCP) only featured one article about anti-Japanese protests held by the Chinese diaspora overseas. According to leaked guidelines from the State Council Information Office in September 2012, websites were requested to censor content related to "mobilizing anti-Japan demonstrations, stirring up excitement, rioting and looting".[3] Similarly, during a wave of anti-Japanese protests in 2005, Stockmann (2010) found that the Chinese propaganda department banned coverage of the protests and restricted reporting about Japan to follow the more positive government line.

[2]This is in line with survey findings by Weiss (2019) that the Chinese public are more hawkish than dovish when it comes to foreign policy.

[3]See `https://chinadigitaltimes.net/2012/09/ministry-of-truth-anti-japan-protests`.

2.2.2 Media environment in China

Television

Television set ownership was close to universal in China in 2012 and cable television was the most popular subscription option. On average, people watched close to 3 hours of television per day and peak viewing times were between 7pm and 10pm, when over 40 percent of the population were estimated to watch television. Urban households had access to on average 64 TV channels (Wang, 2012). The availability of TV channels is determined by local cable providers and varies across and in some cases within provinces. All major TV stations in China are owned by the government at different levels (central, provincial, prefecture or county) and can operate a number of TV channels. While the government provides a token subsidy to its TV stations, they receive the vast majority of their revenue from advertising (Zhao and Guo, 2005). The 14 TV channels of the central government's TV station (China Central Television or CCTV) are generally available across China, as are the general-interest channels of the 31 provincial TV stations. TV channels from sub-provincial TV stations are usually only broadcast in their own or neighbouring locality, with the exception of Xiamen and Shenzhen TV.[4] In addition, there are educational and other specialised TV channels, which often require additional subscription fees. The default order in which different TV channels are shown differs across provinces (and sometimes cities) and rarely changes over time. Typically, the first channels are CCTV channels, followed by local provincial, and in

[4]These types of local channels can also face stricter content restrictions and only be allowed to show content from CCTV channels.

some case city, TV channels and then by other provinces' TV channels.

Television dramas TV dramas are the most popular type of broadcast on TV, with a viewership rate of over 30 percent, more than double that of news programmes (Wang, 2012). A typical Chinese TV drama is 30 to 45 episodes long, each episode lasts around 45 minutes to an hour across one season. Two or three episodes of the show are broadcast every consecutive evening after the news for around two weeks until the drama ends.[5] While TV soaps, like all other media products in China, are subject to government control, regulations are comparatively laxer. Following the commercialisation of China's television market, the majority of TV dramas have become privately funded and produced by the end of the 1990s. Nevertheless, producers require a government permit and must report the topic of each proposed drama to the State Administration of Radio, Film and Television (SARFT, now renamed National Radio and Television Administration). After production, each TV show needs to receive a distribution license from the same ministry before it can be sold to one of the government-owned TV stations for broadcast. Broadcasts are subject to censorship and politically problematic programmes can be removed (Donald, Hong and Keane, 2014).

Sino-Japanese war dramas While TV dramas about the Sino-Japanese war follow the official version of history, they should not be regarded as official government propaganda. Around 70 percent of these shows on air

[5]Some channels change their schedule on weekends.

in the evenings in 2012 were privately produced.[6] According to media reports and interviews with members of the industry, profit margins for these types of shows are large, partly because they were less politically sensitive and easier to get past the censors than dramas about modern day China.[7] The early 2010s were a high point for historical TV dramas. In 2011, around a quarter of all distribution licenses for domestically produced TV shows were issued to dramas set in 1930s and 1940s China, see Figure 2.1. One of the main Chinese TV show filming locations hired 300,000 actors per year, 60 percent of whom have played Japanese soldiers. In the space of two months, more than 10,000 "Japanese" soldiers would "die" during filming.[8] Beginning in 2013, the central government started to limit the production and broadcast of this type of TV drama, especially during peak TV viewing times. According to an opinion piece in the People's Daily, these Sino-Japanese war dramas have left behind the "burden of propaganda and education" to become pure entertainment.[9]

Social media

At the end of 2012, there were 564 million internet users in China, which is roughly half of the population. Internet users were mostly concentrated in urban areas and more developed regions and spent an average of 20.5 hours per week online. Close to 55 percent of internet users reported using micro-blogging platforms (CNNIC, 2013). Sina Weibo is the most popular of such platforms. It is similar in many respects to Twitter, which is

[6]Section 2.5.2 shows that it is these privately produced shows that drive the results.
[7]See `http://culture.people.com.cn/n/2013/0429/c22219-21324676.html`.
[8]See, e.g., `http://www.chinanews.com/yl/2013/04-12/4723834.shtml`.
[9]See `http://opinion.people.com.cn/n/2013/0329/c1003-20958865.html`.

currently banned in China.[10]

The Chinese government is known to influence online discussions by both censoring and fabricating posts on social media platforms. However, an empirical analysis of over 40,000 known government-sponsored posts by King, Pan and Roberts (2017) found that zero percent of these posts are aimed at disparaging foreign countries. Studies about censorship on the Chinese internet, such as King, Pan and Roberts (2013); Fu, Chan and Chau (2013), also found little evidence that these types of posts are censored.

2.3 Data

2.3.1 TV data

Broadcast schedule I collected detailed TV schedules from the website of China's central government TV station (China Central Television or CCTV). The sample covers the peak viewing hours from 6pm to 12am for 2012 and 51 TV channels that broadcast TV dramas.[11] Apart from 9 local channels that only broadcast in their own province, the rest of the channels are in principle available throughout the country.[12] There are four CCTV and four education channels that show TV dramas. The rest of the

[10]There is a 140 character limit per post. Users can follow other users and comment on each other's posts.

[11]Schedules are available for a total of 74 TV channels, which include foreign language channels and specialised TV channels, such as sport, music or news channels.

[12]In reality, availability may depend on the local cable provider.

TV channels are provincial TV stations.[13] I use official sources and Baidu Wenku (an online depository of documents) to collect information on the channel order for cable TV in each province.

I identify 918 unique TV dramas from the TV schedules and classify them into different genres on the basis of official approval forms from the State Administration of Press, Publication, Radio, Film and Television (SAPPRFT) of China and entries on Baidu Baike, China's most popular online encyclopedia. The main theme of 155 TV dramas (17 percent) was fighting against the Japanese occupation of China, while an additional 37 (4 percent) mentioned the Japanese occupation. Table 2.1 shows that on average each channel shows around 30 minutes of anti-Japanese TV soaps per day, compared to around 100 minutes of other types of TV soaps.

TV audience data For a subset of TV shows, I purchased TV audience data for all Chinese provincial capitals and two other major cities (Xiamen and Shenzhen) from CSM Media Research, the main Chinese TV audience data provider. TV ratings are collected using either a diary or meter and are representative of the population over the age of four in each city.

2.3.2 Social media data

I use open access data collected from Sina Weibo in 2012 from Weiboscope, a data collection and visualisation project by the University of Hong

[13] Two sub-provincial level cities (Xiamen and Shenzhen) also have TV channels, which are widely available.

Kong's Journalism and Media Studies Centre (Fu, Chan and Chau, 2013). For most of the analysis, I use the random sample described in Fu and Chau (2013), which was constructed by randomly generating user account numbers within the range of all possible account numbers.[14] The dataset includes the text and timestamp of 3.8 million posts from the timelines of randomly selected users and more than 117,000 unique user IDs. For each user, I have information on their self-reported gender, province of residence and whether or not they have a verified account (similar to Twitter's verification process).[15]

I classify the content and sentiment expressed in each post using context-specific keywords. For example, a common derogatory term for Japanese people in Chinese is "Japanese devil", which is only used for Japanese people and is always meant negatively. I also classify whether a post expresses nationalist sentiment, refers to the Senkaku/Diaoyu island conflict or calls for the boycott of Japanese goods. See Appendix 2.A for more details about the classification of anti-Japanese hate speech on social media. I construct a province-15 minute panel of the social media data. Panel B of Table 2.1 shows that on average the sample contains around four posts from a given province in a 15-minute time interval, however, there is substantial variance in the number of posts. Of these posts, on average around 0.14 percent contain anti-Japanese hate speech.

[14]This is an approximately 1/9000 sample of all valid Sina Weibo users at that time.

[15]Using the 2 percent of posts with a geolocation tag, I verify that 80 percent of these geotagged posts originate from the same province as reported by the user. A survey of Sina Weibo users in Beijing finds few significant differences between users who use geotags and those that do not (Chang and Manion, 2017). A real-name registration system required microblogging users to reveal their real identity to the service provider and government starting in 2012 (see, e.g. Fu, Chan and Chau, 2013).

2.3.3 Anti-Japanese protests

Data on anti-Japanese protests in 2012 and measures of Japanese occupation during the war at the prefecture level were provided by Wallace and Weiss (2015).[16] Figure 2.B2 shows a map of protest incidence across China. I construct a province-day panel using these data, with descriptive statistics reported in Panel C of Table 2.1. The probability of one reported anti-Japanese protest in a province on a given day is one percent. I also report the number of prefectures within a province that had an anti-Japanese protest and distinguish between protest incidence and onset (defined as the first day of protests in a prefecture of one or more consecutive days of protests).

2.3.4 Other data

Weather I use daily temperature and precipitation data from over 200 weather stations in China from the National Oceanic and Atmospheric Administration (NOAA) and use values from weather stations closest to each provincial capital to construct a province-day panel. I also calculate daily sunset and sunrise times for each provincial capital using the formula provided by NOAA Solar Calculations.

[16]Protests data are collected through Chinese-language internet searches and rely on media reports, activists' websites and social media. To be counted as a protest, there had to be at least five participants and two independent sources.

2.4 Empirical Strategy

Most studies that estimate the causal impact of media consumption on different outcomes exploit exogenous variation in the availability of media and compare outcomes in areas where a particular type of media is available to areas where it is not (see, e.g., Yanagizawa-Drott, 2014; Chong and La Ferrara, 2009; DellaVigna and Kaplan, 2007). This strategy is impractical when studying the effect of Sino-Japanese war TV dramas in China today, as most of the major channels are available throughout the country and the share of historical TV dramas is large. This implies that these types of dramas are almost always available to viewers. Instead, I focus on changes in the intensity of treatment, i.e., the likelihood of viewers in different provinces watching a given historical TV drama. The equation of interest is the following

$$y_{p,t} = \beta \times H_{p,t} + \mathbf{X'_{p,t}}\mu + \theta_t + \gamma_p + \epsilon_{p,t} \tag{2.1}$$

with the indices denoting province (p) and time interval (t), respectively. The dependent variable $y_{p,t}$ is a measure of anti-Japanese sentiment, either as expressed on social media or through real-life protests. $H_{p,t}$ is the viewership of TV dramas set during the Sino-Japanese war. The main coefficient of interest is β. A positively estimated coefficient implies that historical TV drama viewership increases anti-Japanese sentiment in China. $\mathbf{X'_{p,t}}$ are province-time controls, such as the total number of posts in province p at time t. θ_t includes time controls and γ_p are province fixed effects. $\epsilon_{p,t}$ are

clustered at the province-day level.

There are two main identification concerns with this approach. The first is strategic scheduling of Sino-Japanese war dramas in response to changes in anti-Japanese sentiment in the population. The second is that audiences will select into different TV programmes based on their pre-existing biases. My identification strategy exploits the high-frequency nature of the data and exogenous variation in audience numbers across TV channels, time and province. The estimated effects are based on those marginal TV show viewers that have been moved to watch historical TV dramas due to factors that are orthogonal to their prior anti-Japanese sentiment.

2.4.1 Endogenous scheduling

One potential endogeneity concern is strategic scheduling of historical TV dramas when anti-Japanese sentiment in the population is especially high. It is highly unlikely that TV channels could change their scheduling in time to benefit from unexpected temporary increases in anti-Japanese sentiment in their target population, as TV schedules are set at least one week in advance.[17] Additionally, these soap operas usually last for 30 episodes and are typically aired continuously with two or three episodes per evening for around two weeks. I estimate the main results at the day (for anti-Japanese protests) and 15-minute level (for social media outcomes) to rule out this type of reverse causality.

[17]Local newspapers usually post the TV schedule for the upcoming week on the weekend and internet TV schedule websites provide detailed schedules for the week and less detailed schedules for two weeks ahead. In order to attract viewers and advertisers, TV channels often advertise new shows months in advance.

I also take into account potential spikes in anti-Japanese sentiment that could have been anticipated, by controlling for the official memorial dates of the Second Sino-Japanese war and the intensity of the island conflict in θ_t.[18] The intensity of the island conflict is measured by the number of articles mentioning the Diaoyu islands in the People's Daily, the official newspaper of the CCP. This measure is highly correlated with Google trends for the Senkaku islands, which measures the search intensity for terms related to the islands on Google from China, see Appendix Figure 2.B1.

Figure 2.2 shows the scheduled airtime for historical TV dramas across all main TV channels and the number of articles about the Senkaku/Diaoyu island conflict in the People's Daily. After accounting for month fixed effects, there is no correlation between these two measures.

2.4.2 Selection of TV audiences

Given the prevalence of this genre and the national coverage of most TV channels, there is little meaningful variation in the *availability* of Sino-Japanese war dramas. Using actual TV audience figures for historical TV dramas (see equation 2.1) also includes a selection effect, as viewers select TV programmes based on their tastes and, in this case, priors about Japan. A second difficulty is that TV audience data are collected by private TV ratings companies (e.g. Nielsen in the US and CSM Media Research in China) and are very costly to obtain. To solve these issues, I adopt a

[18] This is a conservative measure, as most of the events surrounding the conflict were also unexpected. The memorial dates are by law the 7th of July and 3rd and 18th of September.

three-step approach. In the first stage, I use a subsample of TV audience data to estimate changes in audience numbers across time, channels and provinces due to reasons that are unrelated to underlying anti-Japanese sentiment, using an equation of the following form:

$$V_{c,p,t} = \alpha_1 \times channelorder_{c,p} + \alpha_2 \times tvshows_{c,p,t} + \delta_c + \theta_t + \gamma_p + \epsilon_{c,p,t} \quad (2.2)$$

with the indices denoting channel (c), province (p) and 15-minute time interval (t), respectively. The dependent variable $V_{c,p,t}$ is the viewership of channel c at time t in province p from CSM Media Research.[19] $channelorder_{c,p}$ is the default position of channel c in province p. $tvshows_{c,p,t}$ denote the number of TV shows on air at time t on the five channels before and after channel c in province p. θ_t are time effects, which include hour, day-of-week and month fixed effects, as well as controls for important World War II anniversaries and the number of articles in the People's Daily that mention the Diaoyu/Senkaku island conflict. γ_p are province fixed effects and $\epsilon_{c,p,t}$ are clustered at the level of the TV audience data.

Default channel positions are decided by provincial cable providers and change very infrequently over time. Figure 2.3 shows an example of channel positions in Shanghai relative to Beijing and its two neighbouring provinces. One systematic feature of the channel ordering is that a province's own TV channel tends to be placed ahead of provincial TV channels.[20]

[19] Viewership is defined as the percentage of people watching a given channel at a given time of the total survey population in a given provincial capital.

[20] Table 2.C6 shows that results are robust to excluding the own province's TV channel.

While provincial TV channels are owned by the provincial governments, they broadcast to the entire country and compete with each other and central channels for viewers and advertisement. There are few incentives for these TV channels to specifically target audience from their own province instead of trying to appeal more broadly to the whole country. In Table 2.2 I show that the share of Sino-Japanese war dramas on a province's TV channel is not correlated with that province's experience during the Sino-Japanese war, as measured by civilian casualties and number of occupied cities, number of cities with war memorials, anti-Japanese protests in 2010 and its economic ties with Japan. Neither is this type of broadcast systematically different for provinces depending on the prevalence of television and internet use (see also Appendix Table 2.C2 for additional indicators). Sino-Japanese war drama broadcasts are also uncorrelated with a range of macroeconomic indicators, a province's degree of openness and average educational attainment in a province (see Appendix Tables 2.C1 and 2.C2).[21]

Table 2.3 shows the estimation results for predicting TV audiences (equation 2.2). In line with the findings of Martin and Yurukoglu (2017) for the US, audience numbers are decreasing in channel order after taking into account time controls. There is evidence of a substitution effect, as audience numbers for a given show decrease with the number of other TV shows on neighbouring channels. These two variables alone capture 12 percent of the variation in TV audience numbers and the full regression model has an F-statistic of 588.

[21] Out of 39 variables tested, three are significant at the 10 percent level: share of male population, number of visits by foreign tourists and number of mobile phones per 100 rural households.

In the second step, I calculate the total predicted viewership for historical TV dramas in each province over a time period by aggregating predicted viewership across channels and interacting it with each channel's TV schedule:

$$\widehat{H_{p,t}} = \sum_{c} \widehat{V_{c,p,t}} \times \mathbb{1} SinoJapaneseTV_{c,p,t} \tag{2.3}$$

In the final step, I regress expressions of anti-Japanese sentiment on predicted viewership for historical TV dramas:

$$y_{p,t} = \beta \times \widehat{H_{p,t}} + \mathbf{X'_{p,t}}\mu + \theta_t + \gamma_p + \epsilon_{p,t} \tag{2.4}$$

with the indices denoting province (p) and day (t), respectively. The dependent variable $y_{p,t}$ is the incidence or onset of an anti-Japanese protest or the fraction of social media posts containing anti-Japanese hate speech in province p on day t. $\mathbf{X'_{p,t}}$ are province-time controls, such as rainfall, extreme temperature and the total number of social media posts in province p at time t. θ_t are day-of-week and month fixed effects and γ_p are province fixed effects. $\epsilon_{p,t}$ are clustered at the province level.

2.5 Results

2.5.1 Anti-Japanese Protests

Table 2.4 shows how higher predicted historical TV drama viewership the evening before increases the likelihood of an anti-Japanese protest occurring (columns 1 and 3) or starting (columns 2 and 4), measured either as a dummy variable (columns 1 and 3) or the fraction of cities within a province (columns 2 and 4). A one standard deviation increase in historical TV drama viewership on the previous evening (0.01) increases the probability of protests by approximately 40 percent of the mean probability (1.2 percent, column 1). The effect on new protests is very similar (column 3).

While entertainment media played a role in the wave of anti-Japanese protests in the summer of 2012 across China, the effect is significantly smaller than those of Sino-Japanese war anniversaries and developments in the Senkaku/Diaoyu island conflict. A one standard deviation increase in the number of People's Daily articles about the Senkaku/Diaoyu islands is associated with a greater than four time increase in the mean probability of protests. Sino-Japanese war anniversary dates are associated with a more than 20-fold increase in protest likelihood. Consistent with the literature on protests, adverse weather conditions reduce the likelihood of protests. Rain reduces the likelihood of protests by around 1 percentage point, while uncomfortably high temperature reduces protests by around 0.3 percentage points.[22]

[22]Extreme temperature is defined as a dummy variable that is equal to one if the daily maximum temperature in a province is more than one standard deviation away from its

Private vs. State TV Shows

Table 2.5 shows how the effects of historical TV drama viewership differs by producer type. While privately produced TV shows have a positive and significant effect on protests, state-produced TV shows (including those produced by the government, party or military and state-owned companies) have a negative and significant effect on protests. The difference between both coefficients is significant at the 5 percent level.[23]

77 percent of all historical TV dramas aired in 2012 were produced by private companies. Predicted viewership of privately produced TV shows is higher than that of state produced TV shows. This effect is mainly due to differences in the TV channels where these shows air, state produced TV shows are disproportionately broadcast on TV channels with lower viewership. One explanation is that less successful TV stations are unable to afford more expensive privately produced TV shows which are generally seen to be of better quality and more attractive to viewers.

Legacy of Sino-Japanese War

Table 2.6 shows heterogeneity of the effects across provinces. The effect of historical TV dramas on protests is stronger in provinces that were more affected by the Sino-Japanese war, in terms of Japanese occupation and civilian casualties. Conversely, there is no significant additional effect in provinces that have erected more war memorials or experienced an

mean.

[23]Other producers include joint production between private and state producers or shows where information about the producer was not available.

anti-Japanese protest in 2010. These results suggest that one mechanism through which TV shows affect protests could be through the activation of collective historical grievances. These grievances are already more salient in areas with memorials or recent protests, so that media exposure has no additional impact.

2.5.2 Anti-Japanese Hate Speech on Social Media

Exposure to depictions of Japanese soldiers on TV shows also increase anti-Japanese hate speech on social media. Table 2.7 shows that the fraction of social media posts containing with derogatory terms for Japanese people (column 1), nationalist sentiment (column 2), references to the Senkaku/Diaoyu island conflict (column 3) and calls for boycotts of Japanese goods (column 4) increase significantly with predicted historical TV show viewership. A one standard deviation increase in exposure to historical TV dramas leads to a 60 percent increase in the fraction of posts containing anti-Japanese hate speech, relative to its mean. This is unlikely to be due to an increase in posts discussing historical TV shows' contents using the same derogatory language from the show's dialogue. References to the current territorial conflict and posts containing nationalist sentiment also increase by around one third of the mean and calls for boycotts of Japanese goods increase by around one half of its mean (for a one standard deviation increase in predicted historical TV drama viewership). These results suggest that these TV dramas inflame broader anti-Japanese sentiment in the population.

Appendix Table 2.C3 shows additional social media outcomes. Column 1 shows that overall references to Japan increase significantly with higher exposure to historical TV dramas. These are likely to be more negative posts about Japan that are not captured by the specific anti-Japanese keywords used in the main analysis. Column 3 shows that there is no effect on posts about Japanese food, which is a popular topic among posts referencing Japan. Column 2 shows that there is no effect on references to the Sino-Japanese war. This is consistent with the interpretation that these historical TV dramas are mainly entertainment products with little informational content about the war and that their effect on social media is not driven by posts responding directly to the dramas themselves. Columns 4 and 5 show that there is no statistically significant effect of TV dramas on posts referencing anti-Japanese or other types of protests and demonstrations. One possible explanation is that social media posts about protests are politically sensitive, so that users are more likely to avoid using keywords that would readily flag these posts to censors.

Appendix Tables 2.C4 and 2.C5 show that there are no heterogeneous effects of historical TV drama viewership on anti-Japanese sentiment on social media depending on a province's experience during the Sino-Japanese war.

Private vs. State TV Shows

In Table 2.8, I distinguish between the effects of viewing privately compared to state produced historical TV dramas. As with anti-Japanese protests,

privately produced historical TV dramas have a significant positive effect on social media posts containing anti-Japanese and nationalist sentiment which is significantly different from the effects of state produced TV dramas. The effect of privately produced TV dramas on posts referring to the island conflict and calling for boycotts is also positive, but not statistically different from that of state produced TV dramas.

Original Posts vs. Reposts

Table 2.9 shows the effect of historical TV drama exposure on original social media posts (Panel A) and reposted or forwarded social media posts (Panel B). I find that TV exposure significantly increases the fraction of original social media posts expressing anti-Japanese and nationalist sentiment (columns 1 and 2 in Panel A), while the effect on reposts is much weaker. TV shows have a somewhat greater effect on reposts about the current conflict and anti-Japanese boycott compared to original posts. These results are suggestive that these historical TV dramas increase anti-foreigner sentiment in China.[24]

High-Frequency Event Studies

To more precisely pinpoint the effect of TV shows on social media, the data can be disaggregated further into 15-minute time intervals. Appendix Fig-

[24] Another interpretation is that these TV shows encourage the expression of existing anti-foreigner sentiment on social media, which may influence other social media users' opinions or encourage them to voice or act on their prejudices by normalising and destigmatising hate speech.

ure 2.C1 shows how on average predicted historical TV drama viewership evolves over an evening. The biggest differences in predicted viewership are between 7.30 pm and 9 pm, when the majority of TV stations schedule TV dramas and viewership is highest. Figure 2.4 shows the raw social media data (adjusted for province fixed effects) as a type of event study. In the black line is the average percentage of social media posts containing anti-Japanese sentiment when predicted historical TV drama viewership is high and in the grey line when viewership is low. There is a larger fraction of social media posts with derogatory references to Japanese people and the Senkaku/Diaoyu island conflict on days and in provinces when predicted historical TV drama viewership is high, between 7.30 pm and 9 pm. While the data on social media posts with nationalist sentiment and calls for anti-Japanese boycotts is overall noisier, the overall pattern holds.

2.6 Robustness

2.6.1 Leads and Lags of Historical TV Drama Exposure

Figure 2.5 shows the effect of leads and lags of predicted historical TV drama viewership on a dummy variable for anti-Japanese protest incidence (see Table 2.4 column 1). Viewership of historical TV dramas significantly predict protests one day to five days on, while there is no predictive power

of future TV drama viewership on protests.[25]

Figure 2.6 shows the effect of leads and lags of predicted historical TV drama viewership on the fraction of social media posts containing different aspects of anti-Japanese hate speech (see also Table 2.7) in 15-minute time intervals. The pattern is similar to the protest results. Past or current exposure to historical TV dramas significantly increase social media posts containing nationalist sentiment, references to the island conflict and calls for boycotts of Japanese goods. Future exposure to these dramas has no effect on social media posts, with the exception of posts containing derogatory terms for Japanese people. This result is likely due to the auto-correlation in TV schedules and the noisiness of the social media data at the 15-minute frequency.

In addition to a persuasion effect, historical TV dramas could also lead to a shift in the timing of anti-Japanese protests and social media hate speech, act as a coordination device for potential protesters or lower the perceived costs of expressing anti-Japanese sentiment. The pattern of lead and lagged effects appears less consistent with a pure time-shifting effect, while the larger effect of TV dramas around four days before a protest could point to a time delay due to coordinating and organising a protest. These alternative mechanisms remain relevant for understanding the effect of entertainment media on anti-foreigner sentiment and behaviour, as long as media changes public perceptions of the prevalence and social desirability of these types of sentiments and behaviours.

[25]Historical TV drama viewership on the same day has a positive and significant (at the 10 percent level) effect on anti-Japanese protests. This could be due to protests taking place late in the day or some residual auto-correlation in TV schedules.

2.6.2 Excluding TV Shows By Own-Province TV Channel

Table 2.C6 shows the main protest and social media results using predicted historical TV drama viewership on channels other than a province's own TV channel. The results remain robust to excluding own-province channels, although the coefficients become smaller in magnitude and less precisely estimated. The coefficient on the percentage of social media posts referring to the disputed islands turns marginally insignificant with a p-value of 0.103.

2.7 Conclusion

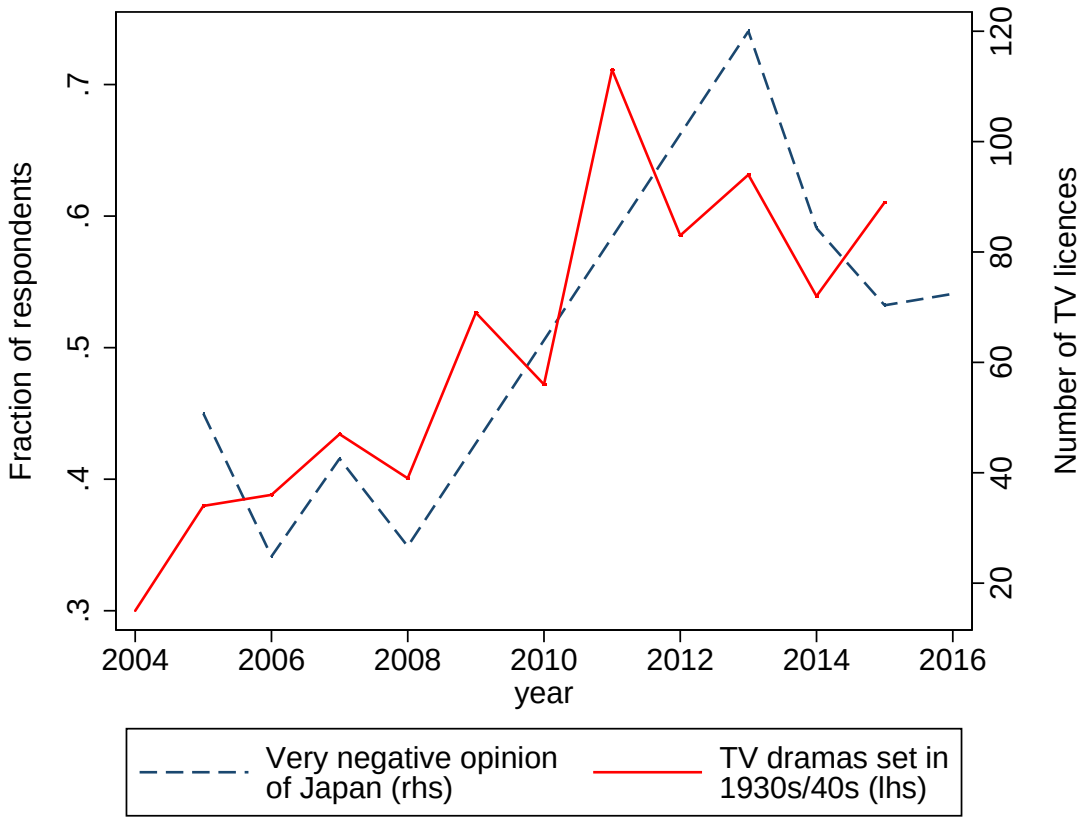
This paper studies the effect of entertainment media on nationalist and anti-foreign sentiment in China. Using detailed information about TV broadcast schedules and TV show content, I show how historical TV dramas set during the Japanese occupation of China 80 years ago affect anti-Japanese sentiment in China today. This type of TV show has become increasingly important over the course of the 2000s and constituted close to 20 percent of all TV shows broadcast by the main Chinese TV channels in the evenings of 2012. I identify the causal effect of these TV dramas by using high-frequency outcome variables and constructing an exogenous measure of predicted viewership based on channel order and the number of rival TV shows. I find that exposure to historical war TV lead to more anti-Japanese protests and more anti-Japanese hate speech on social me-

dia. The effect are driven by privately produced TV shows rather than government programmes.

The results of this paper highlight the importance of studying stereotypes in entertainment media. Entertainment programmes attract a wide audience and while they are not necessarily designed to change opinions and beliefs, they can have unintended and far-reaching influence. As in this paper, stereotypes that appear in films and TV shows are often based on wide-spread beliefs in society, but seeing these beliefs reflected back in the media can reaffirm them and can in some cases lead to protests and social media hate speech.

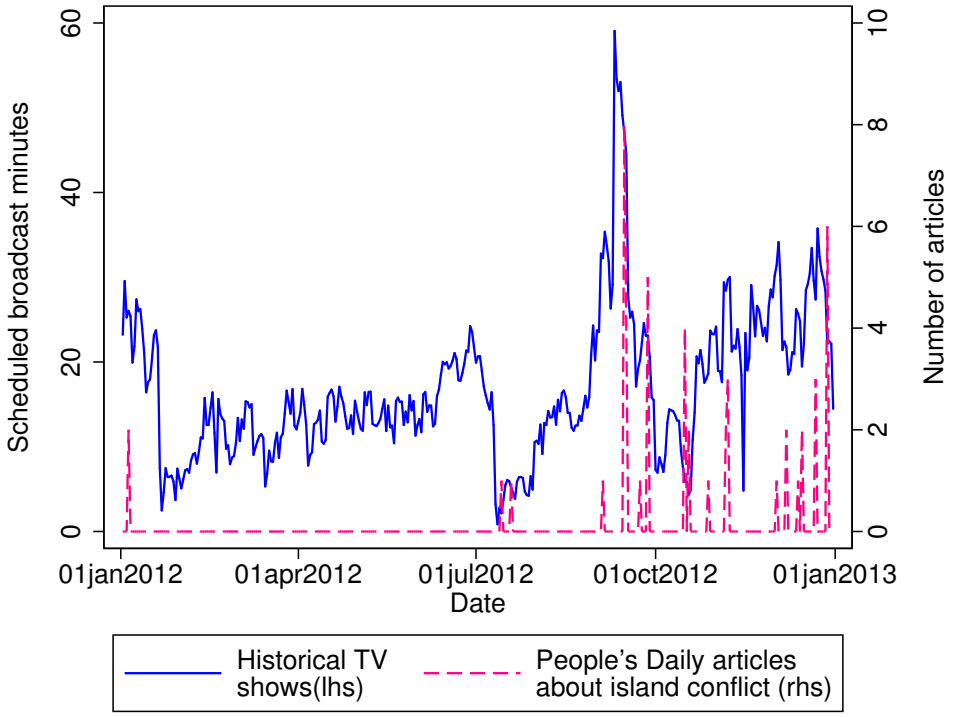
2.8 Figures

Figure 2.1: TV show distribution licenses and anti-Japanese sentiment in China



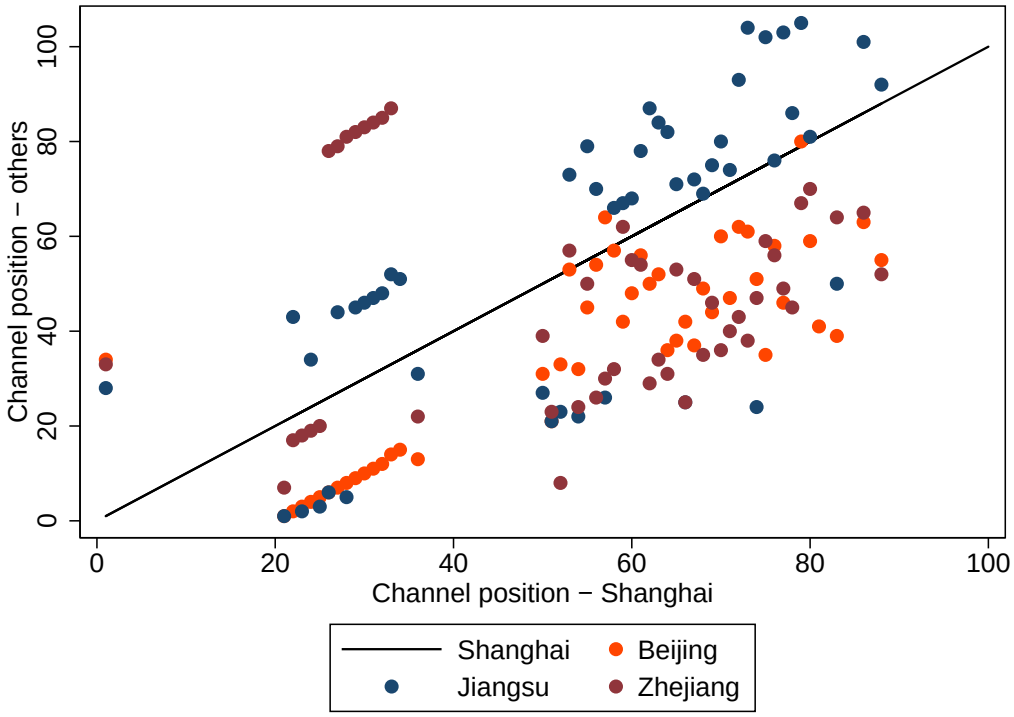
Notes: This chart shows the number of distribution licenses awarded by the State Administration of Film and Television to TV shows set during the 1930s and 1940s in China in red (right-hand axis) and the fraction of Chinese respondents to the PEW Global Attitudes surveys who report having a very negative opinion of Japan in the dashed blue line (left-hand side axis).

Figure 2.2: Historical TV drama scheduling and island conflict



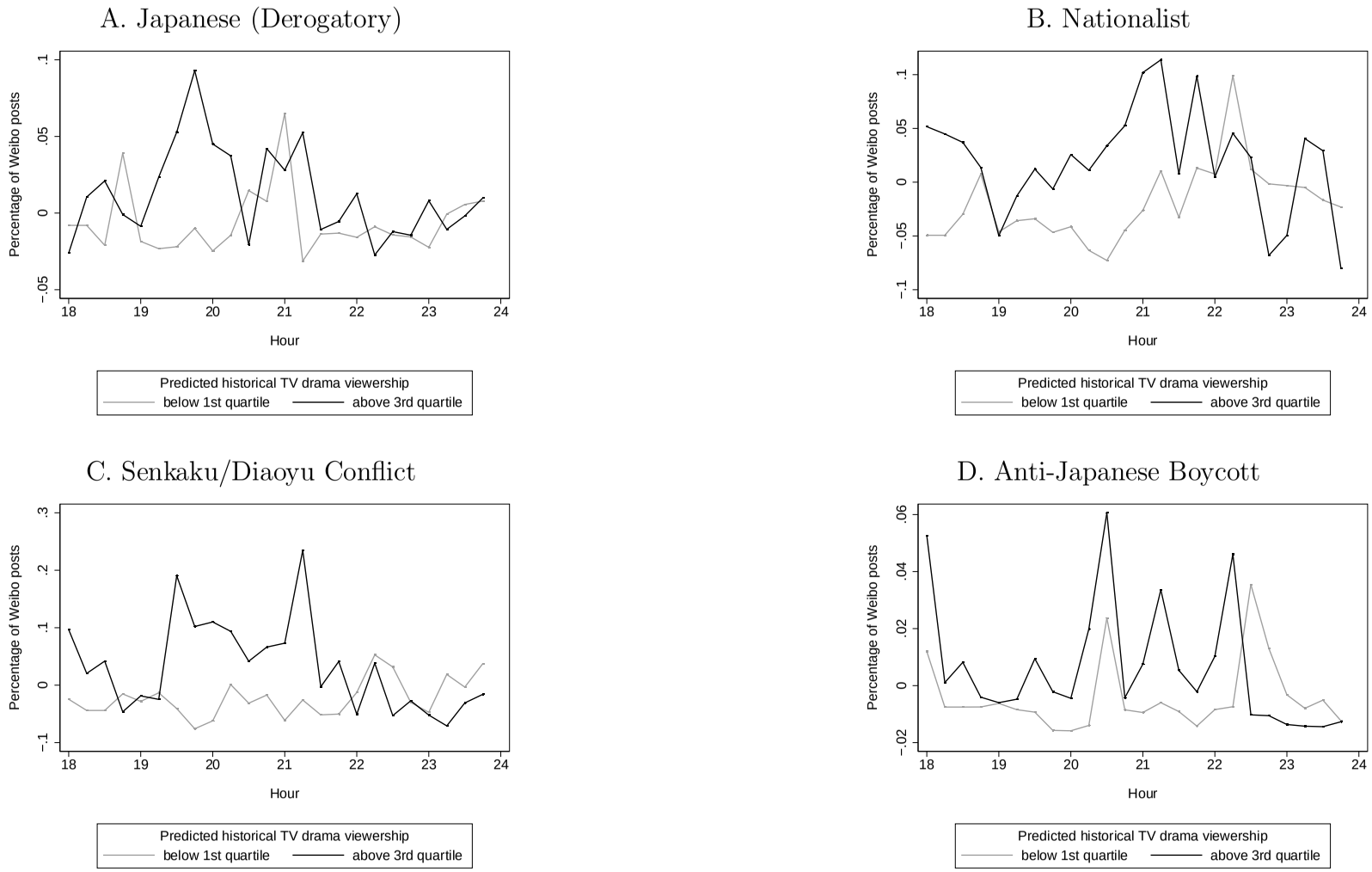
Notes: This chart shows the scheduled airtime for historical TV dramas set during the Sino-Japanese war across all main TV channels (average number of minutes between 6pm and 12am) on the left-hand axis and the number of articles about the Senkaku/Diaoyu island conflict in the People's Daily on the right-hand axis.

Figure 2.3: Channel positions in Shanghai



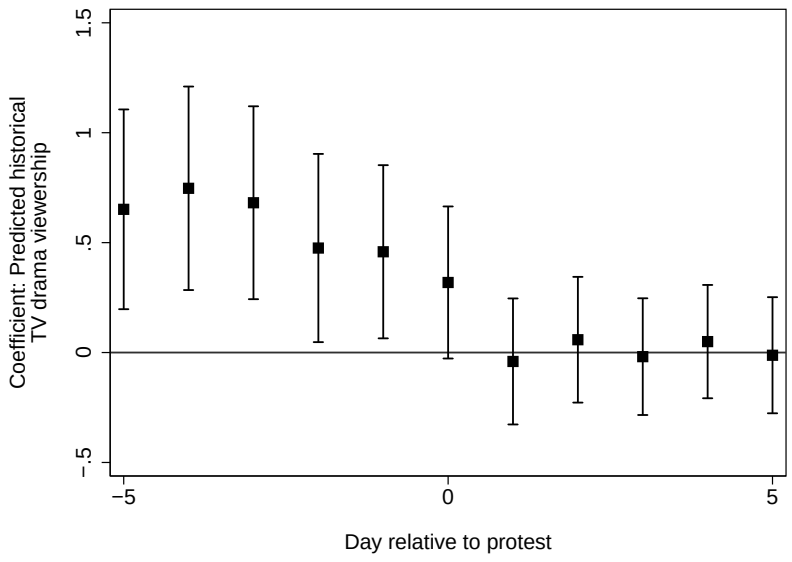
Notes: This chart plots the default channel positions of the main Chinese TV channels for Shanghai on the x-axis against channel positions in Beijing and the neighbouring Jiangsu and Zhejiang provinces on the y-axis.

Figure 2.4: Event Study of Social Media Posts



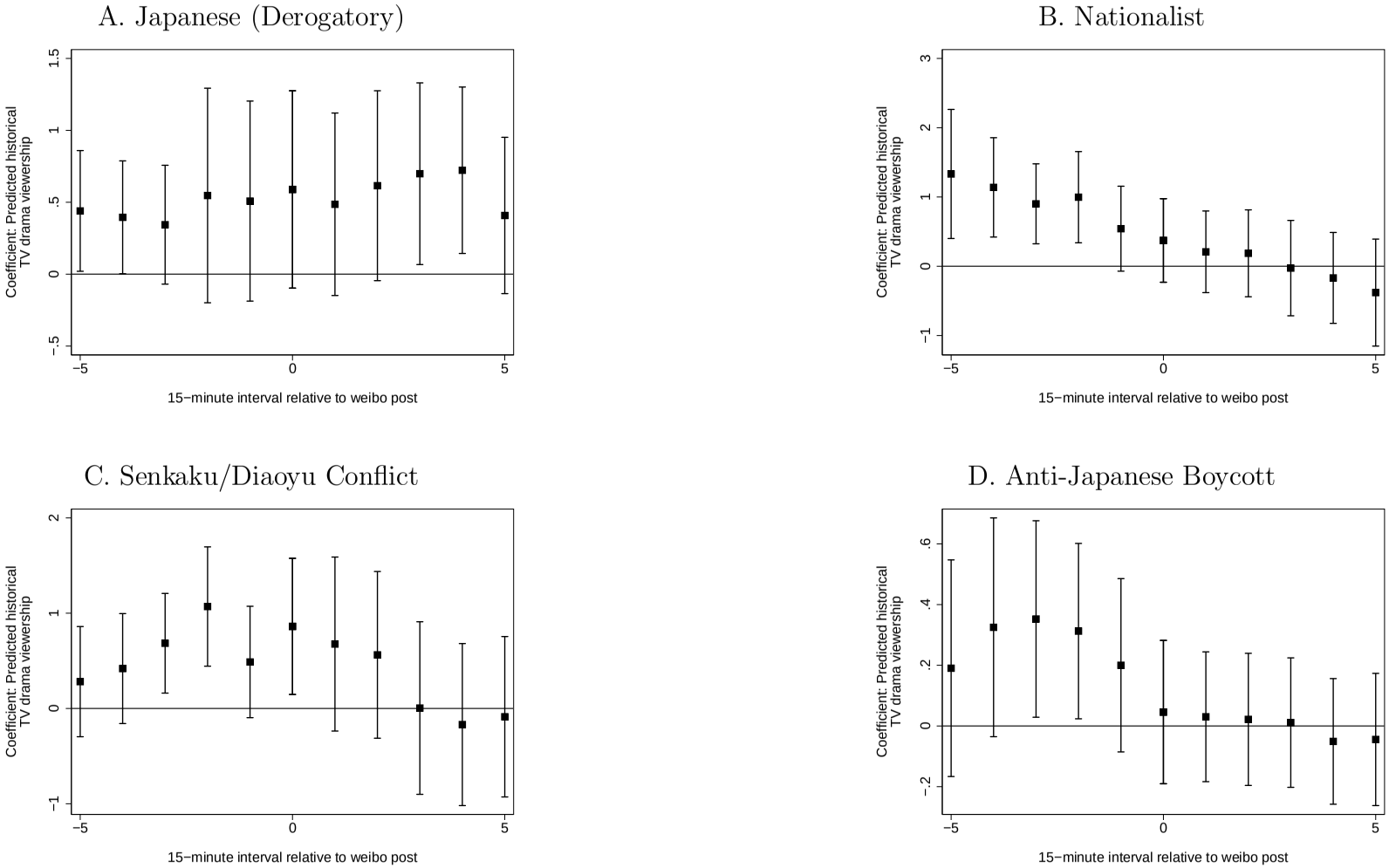
Notes: These charts shows how on average predicted historical TV drama viewership and social media hate speech change over an evening. On the y-axis are the average fraction of social media posts with derogatory terms for Japanese people (Panel A), nationalist sentiment (Panel B), references to the Senkaku/Diaoyu island conflict (Panel C) and calls for boycotts of Japanese goods (Panel D), respectively. The x-axis is the hour of the day. The black line shows the outcome variable when historical TV drama viewership exceeded the third quartile and the grey line when viewership fell below the first quartile. All variables are adjusted for province fixed effects.

Figure 2.5: Anti-Japanese Protest Incidence and Leads and Lags of Historical TV Drama Viewership



Notes: This chart shows the estimated effect of leads and lags of predicted historical TV drama viewership (in days) on an indicator for whether at least one anti-Japanese protest occurs in a given province. On the y-axis are the coefficient estimates and their 95% confidence intervals and on the x-axis is the day relative to an anti-Japanese protest. Controls, fixed effects and standard errors are as in Table 2.4.

Figure 2.6: Social Media Posts and Leads and Lags of Historical TV Drama Viewership



Notes: These charts shows the estimated effect of leads and lags of predicted historical TV drama viewership (in 15-minute time intervals) on the fraction of social media posts with derogatory terms for Japanese people (Panel A), nationalist sentiment (Panel B), references to the Senkaku/Diaoyu island conflict (Panel C) and calls for boycotts of Japanese goods (Panel D), respectively. On the y-axis are the coefficient estimates and their 95% confidence intervals and on the x-axis is the 15-minute time interval relative to the social media posts. Controls, fixed effects and standard errors are as in Table 2.7.

2.9 Tables

Table 2.1: Summary statistics

	mean	sd	min	max	obs
A. TV soap broadcast (in minutes)					
Sino-Japanese war TV drama (explicit)	19.6	49.6	0	551	21,914
Sino-Japanese war TV drama (implicit)	3.70	22.2	0	267	21,914
Other type of TV drama	70.0	76.0	0	551	21,914
B. Social media posts					
Total posts	85.1	138.0	0	1451	11,315
Percentage of posts with...					
...derogatory terms for Japanese	0.052	0.64	0	33	11,315
...references to island conflict	0.11	0.81	0	29	11,315
...nationalist sentiment	0.13	1.09	0	50	11,315
...calls for anti-Japanese boycott	0.020	0.29	0	13	11,315
C. Anti-Japanese protests					
Incidence					
- dummy variable	0.013	0.11	0	1	10,950
- fraction of prefectures	0.0047	0.056	0	1	10,950
Onset					
- dummy variable	0.012	0.11	0	1	10,950
- fraction of prefectures	0.0034	0.044	0	1	10,950

Notes: This table reports summary statistics for TV soap broadcasts (in Panel A, observations are at the channel-day level), social media posts (in Panel B, observations are the province-day level) and anti-Japanese protests (in Panel C, observations are at the province-day level).

Table 2.2: Correlates of Provincial TV Channels' Broadcasts of Sino-Japanese War Dramas

A. Relationship with Japan

Regressor (unit)	Japan. occupation during WWII (number of prefectures) (1)	WWII memorials (number of prefectures) (2)	Civilian casualties during WWII (percent of 1936 population) (3)	Anti-Japan. protests in 2010 (number of prefectures) (4)	Imports from Japan in 2001 (log USD) (5)	Japan. firms in 2001 (log number) (6)	Japan. FDI cum. until 2001 (log USD) (7)
WWII dramas (share of TV soaps)	0.0013	0.0074	0.0059	-0.0041	0.0015	0.019	0.0078
	(0.0037)	(0.0092)	(0.012)	(0.036)	(0.013)	(0.017)	(0.013)
No. of obs.	31	31	28	30	28	27	27
R-squared	0.0039	0.015	0.0062	0.00051	0.00047	0.041	0.013
Mean dep. var.	0.21	0.21	0.22	0.21	0.22	0.21	0.21

B. Television and Internet

Regressor (unit)	TV coverage (percent of population) (1)	Cable TV users (log households) (2)	Digital TV users (log households) (3)	Paid TV users (log households) (4)	Internet penetration (percent of population) (5)	Internet users (log number) (6)	Broadband connection (log households) (7)
WWII dramas (share of TV soaps)	-0.011	0.027	0.016	0.0096	-0.0019	0.026	0.022
	(0.014)	(0.029)	(0.029)	(0.015)	(0.0016)	(0.034)	(0.029)
No. of obs.	31	31	31	30	31	31	31
R-squared	0.018	0.042	0.016	0.014	0.027	0.029	0.027
Mean dep. var.	0.21	0.21	0.21	0.22	0.21	0.21	0.21

Notes: This table shows the correlation between the share of Sino-Japanese war dramas (as share of total TV soap broadcast time) broadcast by a province's main television channel in the evenings of 2012 on different province-level characteristics. In Panel A Relationship with Japan, the regressors are (1) number of prefectures occupied by Japan in WWII, (2) number of prefectures with Sino-Japanese war memorials, (3) civilian casualties during WWII (as percentage of 1936 population), (4) number of prefectures with anti-Japanese protests in 2010, (5) value of imports from Japan, (6) number of Japanese-invested firms, (7) value of Japanese investment. Columns 1, 2 and 4 use data from Wallace and Weiss (2015) (aggregated to the province level and extended to Tibet) and the other columns use data from Che et al. (2015). In Panel B Television and Internet, the regressors are (1) share of population with TV access, (2) number of cable TV users, (3) number of digital TV users, (4) number of paid TV users, (5) share of population with internet access, (6) number of internet users, (7) number of broadband users. The data are from the NBS for 2012. Robust standard errors are in parentheses. Significance at the 1%, 5% and 10% is denoted by ***, ** and *, respectively.

Table 2.3: Predicting TV audience numbers

	Dependent variable: TV viewership rate (percent)		
	(1)	(2)	(3)
Channel position	0.00380***	-0.0239***	-0.0223***
	(0.0000764)	(0.000299)	(0.000568)
Number of TV shows on 5 neighbouring channels	0.0448***	-0.0634***	-0.0646***
	(0.000688)	(0.00142)	(0.00167)
Hour		✓	✓
Day of week		✓	✓
Month		✓	✓
Province			✓
Channel			✓
F-statistic	13,568	1,480	558
No. of obs.	586,836	586,836	586,836
R-squared	0.12	0.46	0.53
Mean dep. var.	0.64	0.64	0.64

Notes: This table shows the estimated coefficients of the first stage regression (equation 2.2) predicting TV viewership of historical TV dramas. The dependent variable is the TV viewership rate, i.e., the number of TV viewers divided by the population over age four. Observations are at the province-15 minute-channel level. Standard errors are clustered at the level of the TV audience data observation, which is approximately the province-90 minute-channel level.

Table 2.4: Historical TV Dramas and Anti-Japanese Protests

	Anti-Japanese Protests			
	Protest incidence		Protest onset	
	Dummy variable	Fraction of prefectures	Dummy variable	Fraction of prefectures
	(1)	(2)	(3)	(4)
Predicted historical TV drama viewership (lagged)	0.458**	0.206*	0.408**	0.152**
	(0.193)	(0.110)	(0.175)	(0.0634)
Island conflict in People's Daily	0.0689***	0.0297***	0.0622***	0.0216***
	(0.00612)	(0.00424)	(0.00677)	(0.00268)
Sino-Jap. War Anniversary	0.251***	0.155***	0.236***	0.135***
	(0.0174)	(0.0176)	(0.0214)	(0.0184)
Rain	-0.0107***	-0.00515***	-0.00956***	-0.00426***
	(0.00171)	(0.000883)	(0.00170)	(0.000674)
Extreme temperature	-0.00299*	-0.00197**	-0.00179	-0.000971**
	(0.00165)	(0.000728)	(0.00154)	(0.000462)
No. of obs.	11,284	11,284	11,284	11,284
R-squared	0.14	0.13	0.13	0.14
Mean dep. var.	0.012	0.0046	0.011	0.0033

Notes: This table shows the estimated coefficients for the second stage regression (equation 2.4) of the effects of predicted historical TV drama viewership on anti-Japanese protests. The dependent variable is either a dummy variable for (columns 1 and 3) or the fraction of prefectures within a province (columns 2 and 4) where a protest occurred (columns 1 and 2) or a new protest started (columns 3 and 4). The main regressor is the predicted fraction of the population over 4 who watched a historical TV drama between 6pm and 12am the day before. Observations are at the province-day level. All regressions include day-of-week, month and province fixed effects. Standard errors (in parentheses) are clustered at the province level. Significance at the 1%, 5% and 10% is denoted by ***, ** and *, respectively.

Table 2.5: Historical TV Dramas and Anti-Japanese Protests: Privately vs. State-Produced Shows

	Anti-Japanese Protest	
	Incidence Dummy	Onset Dummy
	(1)	(2)
Predicted historical TV drama viewership (lagged)...		
...privately produced	0.653**	0.583**
	(0.300)	(0.275)
...state produced	-0.297*	-0.286*
	(0.164)	(0.156)
...other	0.736***	0.705***
	(0.139)	(0.149)
Island conflict in People's Daily	0.0686***	0.0620***
	(0.00601)	(0.00668)
Sino-Jap. War Anniversary	0.251***	0.236***
	(0.0174)	(0.0214)
Rain	-0.0107***	-0.00964***
	(0.00174)	(0.00173)
Extreme temperature	-0.00279	-0.00160
	(0.00165)	(0.00153)
p-value (privately produced= state produced)	0.039	0.041
No. of obs.	11,284	11,284
R-squared	0.14	0.13
Mean dep. var.	0.012	0.011

Notes: This table shows the estimated coefficients for the second stage regression (equation 2.4) of the effects of predicted historical TV drama viewership on anti-Japanese protests. The dependent variable is either a dummy variable for whether a protest occurred (column 1) or a new protest started (column 2). The main regressor is the predicted fraction of the population over 4 who watched a historical TV drama between 6pm and 12am the day before, that is privately produced (producer is a private company), state produced (producer is a central or local government agency, military division or state-owned company) or other (joint productions between private companies and government or no available information about producer). Observations are at the province-day level. All regressions include day-of-week, month and province fixed effects. Standard errors (in parentheses) are clustered at the province level. Significance at the 1%, 5% and 10% is denoted by ***, ** and *, respectively.

Table 2.6: Historical TV Dramas and Anti-Japanese Protests: Heterogeneity across Provinces

	(1)	(2)	(3)	(4)
Interaction with	Japanese occupation during WWII	Civilian casualties during WWII	WWII memorials	Anti-Japanese protests in 2010
Mean of interacted variable	0.64	1.65	0.22	0.14
	A. Dependent Variable: Indicator for Incidence of Anti-Japanese Protest			
Predicted historical TV drama viewership (lagged)	0.131	0.313	0.534**	0.433**
	(0.213)	(0.235)	(0.243)	(0.192)
Interaction term	0.617**	0.0990	-0.282	0.234
	(0.291)	(0.0606)	(0.395)	(0.755)
No. of obs.	11,284	10,192	11,284	11,284
R-squared	0.14	0.14	0.14	0.14
Mean dep. var.	0.012	0.013	0.012	0.012
	B. Dependent Variable: Indicator for Onset of Anti-Japanese Protest			
Predicted historical TV drama viewership (lagged)	0.0516	0.253	0.431*	0.460**
	(0.164)	(0.200)	(0.217)	(0.193)
Interaction term	0.670**	0.105**	-0.0861	-0.478
	(0.247)	(0.0507)	(0.339)	(0.329)
No. of obs.	11,284	10,192	11,284	11,284
R-squared	0.13	0.13	0.13	0.13
Mean dep. var.	0.011	0.011	0.011	0.011

Notes: This table shows the estimated coefficients for the second stage regression (equation 2.4) of the effects of predicted historical TV drama viewership on anti-Japanese protests. The dependent variable is a dummy variable for whether a protest occurred (Panel A) or a new protest started (Panel B). The main regressors are the predicted fraction of the population over 4 who watched a historical TV drama between 6pm and 12am the day before and its interaction with the fraction of prefecture-level cities within a province that 1) were occupied by Japan during WWII, 3) have a Sino-Japanese war memorial and 4) had an anti-Japanese protest in 2010 and 2) civilian casualties during WWII (as percentage of 1936 population). Columns 1, 3 and 4 use data from Wallace and Weiss (2015) (aggregated to the province level and extended to Tibet) and column 2 uses data from Che et al. (2015). Observations are at the province-day level. All regressions include day-of-week, month and province fixed effects and control for rain and extreme temperature. Standard errors (in parentheses) are clustered at the province level. Significance at the 1%, 5% and 10% is denoted by ***, ** and *, respectively.

Table 2.7: Historical TV Dramas and Anti-Japanese Hate Speech on Social Media

	Percentage of social media posts with ...			
	Derogatory terms for Japanese	Nationalist sentiment	References to island conflict	Calls for anti-Japanese boycott
	(1)	(2)	(3)	(4)
Predicted historical TV drama viewership	3.144*	3.881**	2.742**	0.992**
	(1.659)	(1.851)	(1.199)	(0.438)
Island conflict in People's Daily	0.000775	0.285***	0.0390	0.0586**
	(0.0346)	(0.0463)	(0.0602)	(0.0219)
Sino-Jap. War Anniversary	0.0984	0.325	0.274	0.0582
	(0.119)	(0.198)	(0.184)	(0.0720)
Rain	0.0166	-0.0322	-0.0482***	-0.00839
	(0.0141)	(0.0197)	(0.0140)	(0.00592)
Extreme temperature	-0.0160	-0.0116	-0.0182	0.00258
	(0.0176)	(0.0209)	(0.0143)	(0.00619)
Total posts	0.0000196	0.000168*	-0.0000459	-0.0000472
	(0.0000521)	(0.0000884)	(0.0000491)	(0.0000282)
No. of obs.	11,315	11,315	11,315	11,315
R-squared	0.015	0.029	0.085	0.030
Mean dep. var.	0.052	0.13	0.11	0.020

Notes: This table shows the estimated coefficients for the second stage regression (equation 2.4) of the effects of predicted historical TV drama viewership on social media posts. The dependent variable is the percentage of Sina Weibo posts which include derogatory terms for Japanese people (column 1), nationalist sentiment (column 2), references to the Senkaku/Diaoyu island conflict (column 3) and calls for boycotts of Japanese goods (column 4). The main regressor is the predicted fraction of the population over 4 who watched a historical TV drama between 6pm and 12am the day before. Observations are at the province-day level. All regressions include day-of-week, month and province fixed effects. Standard errors (in parentheses) are clustered at the province level. Significance at the 1%, 5% and 10% is denoted by ***, ** and *, respectively.

Table 2.8: Historical TV Dramas and Anti-Japanese Hate Speech on Social Media: Privately vs. State-Produced Shows

	Percentage of social media posts with ...			
	Derogatory terms for Japanese	Nationalist sentiment	References to island conflict	Calls for anti-Japanese boycott
	(1)	(2)	(3)	(4)
Predicted historical TV drama viewership...				
...privately produced	4.703**	5.649**	3.793**	1.137*
	(2.274)	(2.579)	(1.598)	(0.643)
...state produced	-0.0768	-0.542	0.152	0.774
	(0.882)	(2.221)	(1.727)	(0.461)
...other	-3.340	-0.624	-0.0718	0.0845
	(2.313)	(1.434)	(2.997)	(0.358)
Island conflict in People's Daily	-0.00256	0.282***	0.0374	0.0581**
	(0.0355)	(0.0461)	(0.0599)	(0.0218)
Sino-Jap. War Anniversary	0.0915	0.317	0.269	0.0575
	(0.120)	(0.199)	(0.186)	(0.0724)
p-value (privately produced= state produced)	0.048	0.078	0.17	0.68
No. of obs.	11,315	11,315	11,315	11,315
R-squared	0.015	0.030	0.085	0.030
Mean dep. var.	0.052	0.13	0.11	0.020

Notes: This table shows the estimated coefficients for the second stage regression (equation 2.4) of the effects of predicted historical TV drama viewership on social media posts. The dependent variable is the percentage of Sina Weibo posts which include derogatory terms for Japanese people (column 1), nationalist sentiment (column 2), references to the Senkaku/Diaoyu island conflict (column 3) and calls for boycotts of Japanese goods (column 4). The main regressor is the predicted fraction of the population over 4 who watched a historical TV drama between 6pm and 12am the day before, that is privately produced (producer is a private company), state produced (producer is a central or local government agency, military division or state-owned company) or other (joint productions between private companies and government or no available information about producer). Observations are at the province-day level. Observations are at the province-day level. All regressions include day-of-week, month and province fixed effects and control for rain, extreme temperature and total social media posts. Standard errors (in parentheses) are clustered at the province level. Significance at the 1%, 5% and 10% is denoted by ***, ** and *, respectively.

Table 2.9: Historical TV Dramas and Anti-Japanese Hate Speech on Social Media: Original Posts versus Reposts

	Percentage of social media posts with ...			
	Derogatory terms for Japanese	Nationalist sentiment	References to island conflict	Calls for anti-Japanese boycott
		A. Original Posts		
	(1)	(2)	(3)	(4)
Predicted historical TV drama viewership	9.116**	10.26*	1.152	1.088
	(4.299)	(5.731)	(4.802)	(1.168)
No. of obs.	11,315	11,315	11,315	11,315
R-squared	0.0062	0.0088	0.018	0.0042
Mean dep. var.	0.11	0.16	0.21	0.020
		B. Reposts		
	(1)	(2)	(3)	(4)
Predicted historical TV drama viewership	1.817	3.565**	1.814	1.140***
	(1.253)	(1.547)	(1.157)	(0.385)
No. of obs.	11,315	11,315	11,315	11,315
R-squared	0.0059	0.020	0.026	0.022
Mean dep. var.	0.028	0.10	0.055	0.014

Notes: This table shows the estimated coefficients for the second stage regression (equation 2.4) of the effects of predicted historical TV drama viewership on social media posts. The dependent variable is the percentage of original Sina Weibo posts (Panel A) or reposts (Panel B) which include derogatory terms for Japanese people (column 1), nationalist sentiment (column 2), references to the Senkaku/Diaoyu island conflict (column 3) and calls for boycotts of Japanese goods (column 4). The main regressor is the predicted fraction of the population over 4 who watched a historical TV drama between 6pm and 12am the day before. Observations are at the province-day level. All regressions include day-of-week, month and province fixed effects and control for the number of People's Daily articles about the Senkaku/Diaoyu islands, Sino-Japanese war anniversaries, rain, extreme temperature and total social media posts. Standard errors (in parentheses) are clustered at the province level. Significance at the 1%, 5% and 10% is denoted by ***, ** and *, respectively.

Appendix

2.A Social Media

Figure 2.A1: Examples of anti-Japanese hate speech

Japanese devils

There is no difference between bandits and *Japanese devils*! ... - 14 June 2012

Slaughter *Japanese devils* from dog Japan! - 16 August 2012

Senkaku/Diaoyu islands

... boycotting Japanese goods is the best way of saving the *Diaoyu islands* and resisting Japan. - 19 Aug 2012

#*Diaoyu islands* are China's #Down with Japanese imperialism, let's chase the Japanese from the islands. - 11 Sep 2012

Notes: These examples of anti-Japanese hate speech are taken from the random sample of Sina Weibo posts in 2012.

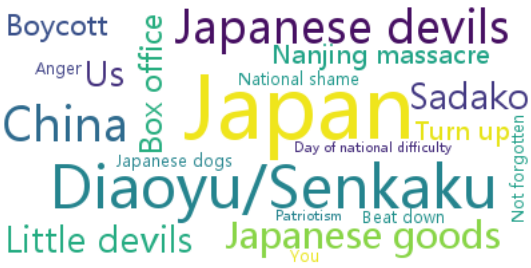
Table 2.A1: Keywords for social media content classification

Original	English translation
Derogatory terms for Japanese people	
日本鬼子	Japanese devil
小鬼子	Little devil
日寇 or 倭寇 or 倭奴	Japanese bandit or pirate
小日本 or 日本仔	Little Japanese
日本恶魔	Japanese monster
Nationalism	
汉奸	Traitor to the Chinese people
爱国	Patriotism
同胞	Fellow citizen
反华	Anti-Chinese
崇洋 or 媚外	Idolising the West/foreign countries
干涉内政	Intervention in internal (Chinese affairs)
境外敌对势力	Hostile forces beyond borders
Senkaku/Diaoyu island conflict	
钓鱼岛	Diaoyu islands
保钓	Protect Diaoyu (islands)
购岛	(Diaoyu) Island purchase
Anti-Japanese boycott	
抵制日货	Boycott Japanese goods
拒绝日货	Refuse Japanese goods

Notes: This table lists the keywords used for classifying the contents of social media posts in the original Chinese and translated into English.

Figure 2.A2: Keywords in Anti-Japanese Weibo posts

Panel A: Japanese devils



Panel B: Senkaku/Diaoyu islands

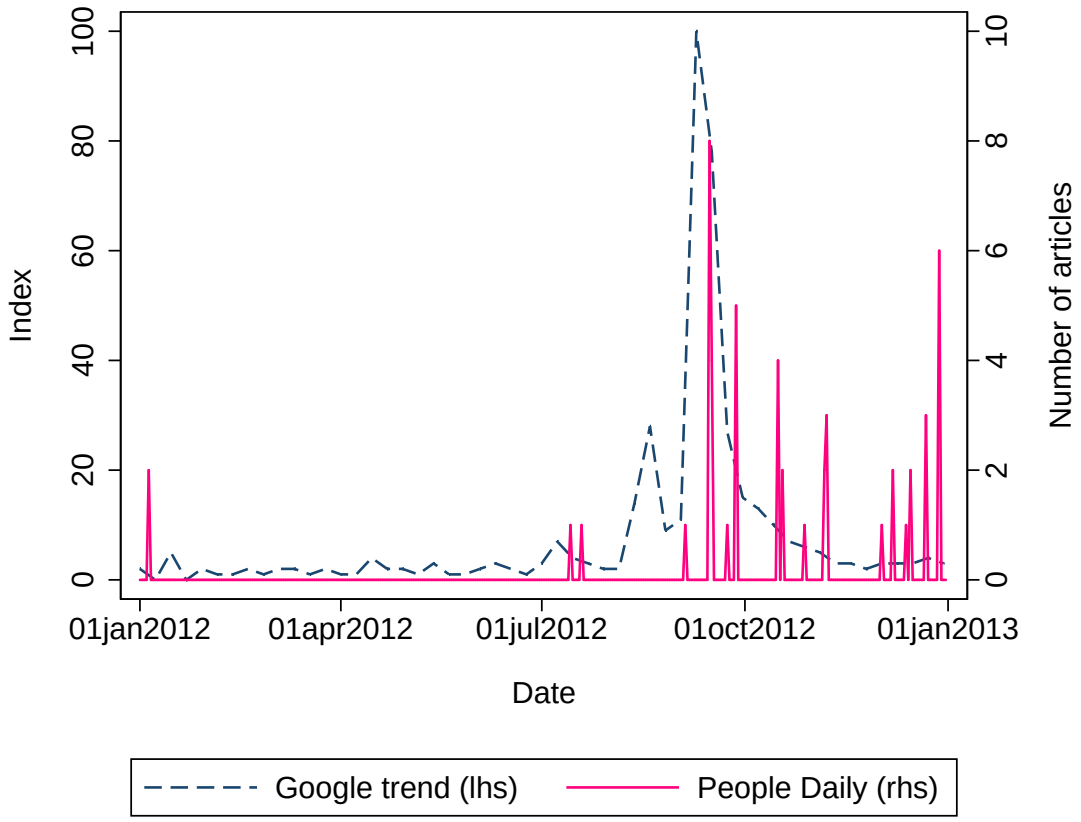
Notes: These charts show English translations of the top 20 keywords for posts containing references to "Japanese devils" (Panel A) and "Senkaku/Diaoyu islands" (Panel B). The Chinese text is segmented using jieba and keywords are defined by their term frequency–inverse document frequency.

Figure 2.A3: Social media flyers calling for boycott

Notes: These figures show social media flyers calling for boycott of Japanese goods, showing common Japanese brand names and their Chinese translations. Source: Heilmann (2016).

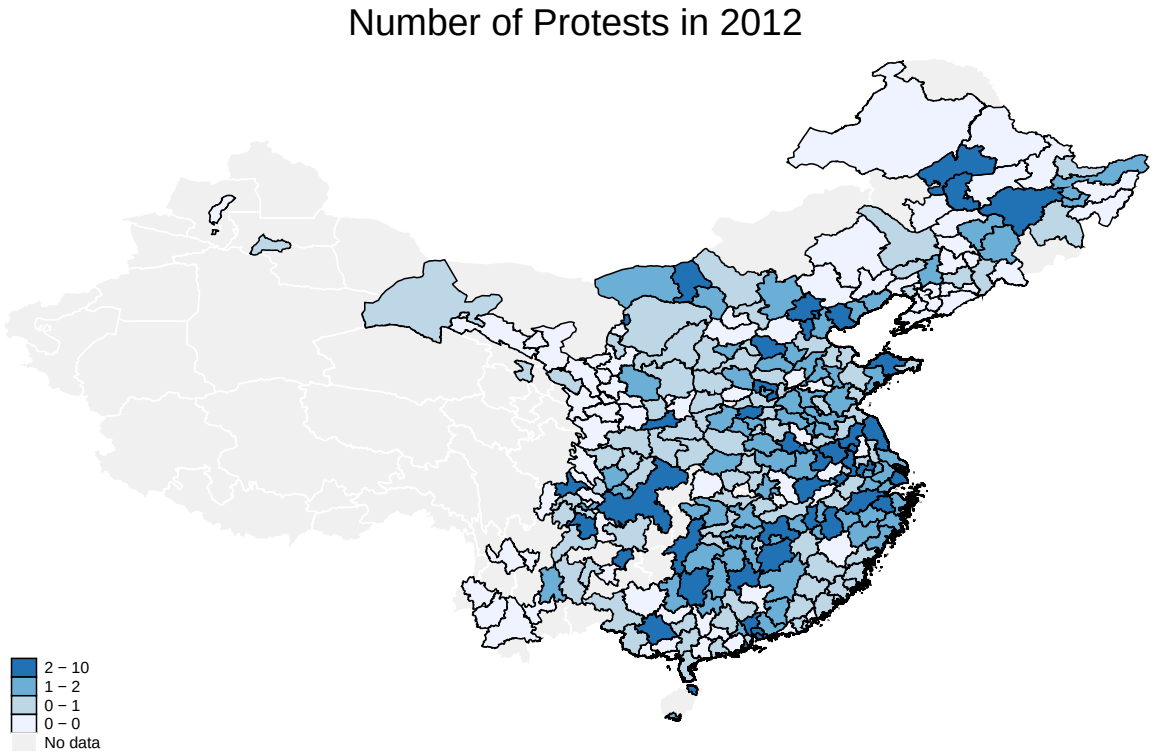
2.B Additional Descriptive Statistics

Figure 2.B1: Salience of island conflict



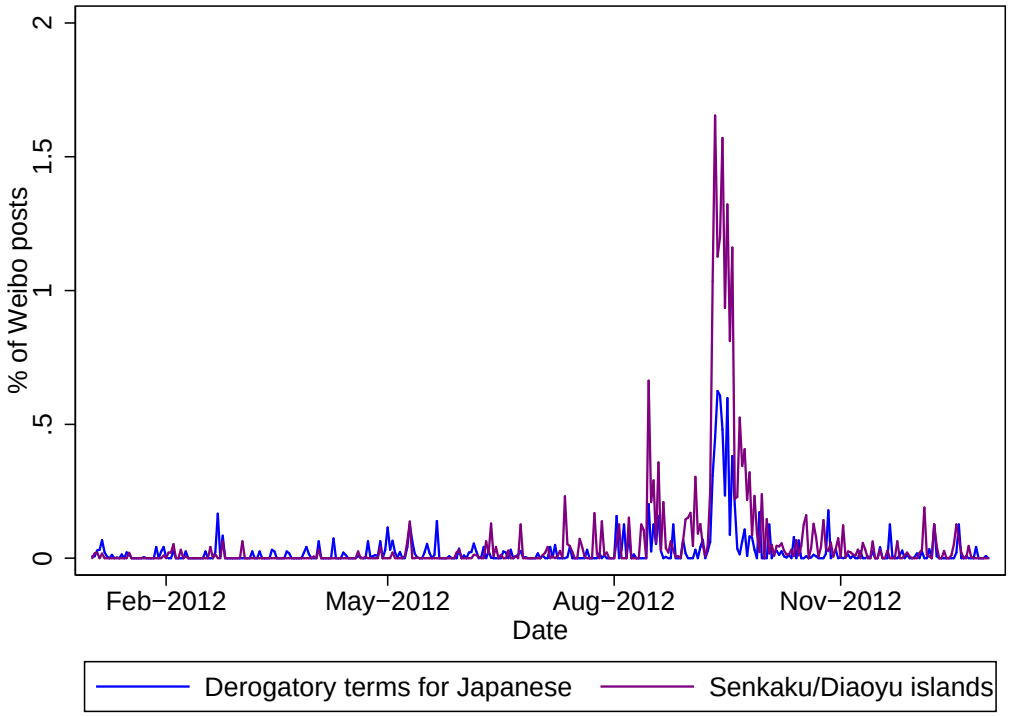
Notes: This chart shows the number of articles in the People's Daily about the Senkaku/Diaoyu islands per day on the right-hand axis and the Google trend index for searches in mainland China related to the Senkaku/Diaoyu islands per week on the left-hand axis.

Figure 2.B2: Map of anti-Japanese protests in China



Notes: This map shows the number of anti-Japanese protests in August and September 2012 across prefectures in China. Source: Wallace and Weiss (2015).

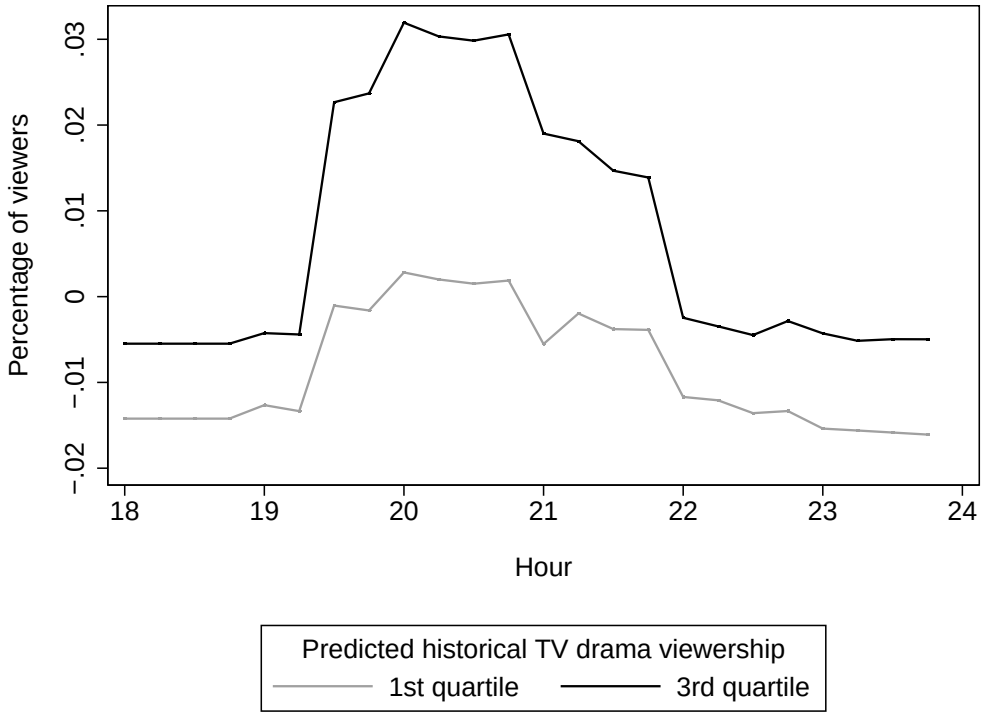
Figure 2.B3: Social Media Hate Speech over Time



Notes: This chart shows the percentage of Sina Weibo posts that contain derogatory terms for Japanese people in the blue line and references to the Senkaku/Diaoyu islands in the purple line.

2.C Additional Results and Robustness Tables

Figure 2.C1: Event Study: Predicted Historical TV Drama Viewership



Notes: This chart shows how on average predicted historical TV drama viewership change over an evening. On the y-axis is the average predicted historical TV drama viewership and on the x-axis is the hour of the day. The black line shows the third quartile of predicted historical TV drama viewership and the grey line shows the first quartile. All variables are adjusted for province fixed effects.

Table 2.C1: Additional Correlates of Provincial TV Channels' Broadcasts of Sino-Japanese War Dramas I

A. Macroeconomic Variables

Regressor	Prefecture-level cities	GDP	Average salary (urban)	Unemployment rate (urban)	Population	Urban	Male
(unit)	(number)	(log RMB)	(log RMB)	(percent)	(log number)	(share of population)	(share of population)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
WWII dramas (share of TV soaps)	0.0028	0.021	0.021	0.025	0.026	-0.000075	-5.29**
	(0.0052)	(0.028)	(0.095)	(0.037)	(0.033)	(0.00013)	(2.25)
No. of obs.	27	31	31	31	31	31	31
R-squared	0.014	0.027	0.0011	0.015	0.032	0.010	0.13
Mean dep. var.	0.20	0.20	0.20	0.20	0.20	0.20	0.20

B. Openness

Regressor	Foreign residents	Imports	Exports	Foreign enterprises	Foreign investment	Intl tourism income	Foreign tourists
(unit)	(share of population)	(log USD)	(log USD)	(log number)	(log USD)	(log USD)	(log visits)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
WWII dramas (share of TV soaps)	0.0015	0.0051	0.011	0.0058	0.0031	0.019	0.024*
	(0.0093)	(0.013)	(0.015)	(0.017)	(0.016)	(0.011)	(0.013)
No. of obs.	31	31	31	31	31	31	31
R-squared	0.00031	0.0058	0.018	0.0042	0.0014	0.067	0.080
Mean dep. var.	0.20	0.20	0.20	0.20	0.20	0.20	0.20

Notes: This table shows the correlation between the share of Sino-Japanese war dramas (as share of total TV soap broadcast time) broadcast by a province's main television channel in the evenings of 2012 on different province-level characteristics. In Panel A Macroeconomic Variables, the regressors are (1) number of prefecture-level cities, (2) nominal GDP, (3) average salary for urban employees, (4) unemployment rate in urban areas, (5) total population, (6) share of urban population, (7) share of male population. In Panel B Openness, the regressors are (1) share of foreign residents, (2) value of imports, (3) value of exports, (4) number of foreign-invested firms, (5) value of foreign investment, (6) income from international tourism, (7) number of visits by foreign tourists. The data are from the NBS for 2012. Robust standard errors are in parentheses. Significance at the 1%, 5% and 10% is denoted by ***, ** and *, respectively.

Table 2.C2: Additional Correlates of Provincial TV Channels' Broadcasts of Sino-Japanese War Dramas II

A. Educational attainment

Regressor	No schooling	Primary school	Junior middle school	Senior middle school	Tertiary education
(unit)	(educational attainment of population over 5 as share of total population over 5)				
	(1)	(2)	(3)	(4)	(5)
WWII dramas (share of TV soaps)	0.31	0.13	-0.31	-0.40	0.077
	(0.26)	(0.35)	(0.29)	(0.62)	(0.29)
No. of obs.	31	31	31	31	31
R-squared	0.017	0.0052	0.025	0.012	0.0013
Mean dep. var.	0.21	0.21	0.21	0.21	0.21

B. Ownership of TVs, Computers and Mobile Phones

Regressor	Colour TVs (per 100 households)		Computers (per 100 households)		Mobile phones (per 100 households)	
(unit)	urban	rural	urban	rural	urban	rural
	(1)	(2)	(3)	(4)	(5)	(6)
WWII dramas (share of TV soaps)	-0.000023	-0.00080	0.000043	0.00015	0.00100	-0.0016**
	(0.00083)	(0.00085)	(0.0011)	(0.0015)	(0.0014)	(0.00059)
No. of obs.	31	31	31	31	31	31
R-squared	0.000016	0.014	0.000039	0.00029	0.015	0.095
Mean dep. var.	0.21	0.21	0.21	0.21	0.21	0.21

Notes: This table shows the correlation between the share of Sino-Japanese war dramas (as share of total TV soap broadcast time) broadcast by a province's main television channel in the evenings of 2012 on different province-level characteristics. In Panel A Educational attainment, the regressors are the share of the population over 5 with (1) no schooling, (2) primary school education (around 6 years of schooling), (3) junior middle school education (around 9 years of schooling), (4) senior middle school education (around 12 years of schooling), (5) tertiary education (more than 15 years of schooling). In Panel B Ownership of TVs, Computers and Mobile Phones, the regressors are the average number of (1, 2) colour TVs, (3, 4) computers, (5, 6) mobile phones owned by 100 households in urban (columns 1, 3, 5) and rural areas (columns 2, 4, 6). The data are from the NBS for 2012. Robust standard errors are in parentheses. Significance at the 1%, 5% and 10% is denoted by ***, ** and *, respectively.

Table 2.C3: Historical TV Dramas and Social Media: Additional Variables

	Percentage of social media posts with references to ...				
	Japan (1)	Sino-Japanese war (2)	Japanese food (3)	Any protests (4)	Anti-Japanese protests (5)
Predicted historical TV drama viewership	9.274***	-0.162	0.147	0.208	0.0392
	(2.752)	(0.656)	(0.437)	(0.681)	(0.123)
Island conflict in People's Daily	0.0646	0.0256	-0.0120	0.0269	0.0290**
	(0.0672)	(0.0229)	(0.0129)	(0.0284)	(0.0124)
Sino-Jap. War Anniversary	-0.0691	0.151**	0.00899	0.127	0.0740
	(0.136)	(0.0707)	(0.0337)	(0.0907)	(0.0522)
Rain	-0.00398	0.0114	0.00490	-0.0125	-0.00413**
	(0.0373)	(0.0114)	(0.00536)	(0.0147)	(0.00171)
Extreme temperature	0.0126	0.0101	0.0140	-0.0130	0.000154
	(0.0293)	(0.00986)	(0.00851)	(0.0106)	(0.00168)
Total posts	0.000178	-0.0000160	-0.00000135	-0.0000320	-0.00000600
	(0.000140)	(0.0000503)	(0.0000458)	(0.0000346)	(0.00000712)
No. of obs.	11,315	11,315	11,315	11,315	11,315
R-squared	0.037	0.0029	0.0023	0.015	0.018
Mean dep. var.	0.30	0.031	0.029	0.045	0.0038

Notes: This table shows the estimated coefficients for the second stage regression (equation 2.4) of the effects of predicted historical TV drama viewership on social media posts with day-of-week, month and province fixed effects. The dependent variable is the percentage of Sina Weibo posts which include references to Japan (column 1), Sino-Japanese war (column 2), Japanese food (column 3), any type of protest or demonstration (column 4) and anti-Japanese protests (column 5). The main regressor is the predicted fraction of the population over 4 who watched a historical TV drama between 6pm and 12am the day before. Observations are at the province-day level. All regressions include day-of-week, month and province fixed effects. Standard errors (in parentheses) are clustered at the province level. Significance at the 1%, 5% and 10% is denoted by ***, ** and *, respectively.

Table 2.C4: Historical TV Dramas and Anti-Japanese Hate Speech: Heterogeneity across Provinces I

	(1)	(2)	(3)	(4)
Interaction with	Japanese occupation during WWII	Civilian casualties during WWII	WWII memorials	Anti-Japanese protests in 2010
Mean of interacted variable	0.64	1.65	0.22	0.14
	A. Dependent Variable: Percentage of social media posts with derogatory terms for Japanese			
Predicted historical TV drama viewership (lagged)	5.111	4.360	3.924	2.966
	(3.398)	(2.666)	(2.484)	(1.877)
Interaction term	-3.696	-0.613	-2.920	1.625
	(3.413)	(0.752)	(3.296)	(3.270)
No. of obs.	11,315	10,220	11,315	11,315
R-squared	0.015	0.015	0.015	0.015
Mean dep. var.	0.052	0.053	0.052	0.052
	B. Dependent Variable: Percentage of social media posts with nationalist sentiment			
Predicted historical TV drama viewership (lagged)	3.682	3.232	4.229	3.481*
	(4.023)	(3.155)	(2.795)	(2.049)
Interaction term	0.373	0.645	-1.304	3.639
	(4.374)	(0.992)	(3.976)	(2.943)
No. of obs.	11,315	10,220	11,315	11,315
R-squared	0.029	0.029	0.029	0.029
Mean dep. var.	0.13	0.14	0.13	0.13

Notes: This table shows the estimated coefficients for the second stage regression (equation 2.4) of the effects of predicted historical TV drama viewership on anti-Japanese content on social media. The dependent variable is the percentage of social media posts containing derogatory terms for Japanese people (Panel A) or nationalist sentiment (Panel B). The main regressors are the predicted fraction of the population over 4 who watched a historical TV drama between 6pm and 12am the day before and its interaction with the fraction of prefecture-level cities within a province that 1) were occupied by Japan during WWII, 3) have a Sino-Japanese war memorial and 4) had an anti-Japanese protest in 2010 and 2) civilian casualties during WWII (as percentage of 1936 population). Columns 1, 3 and 4 use data from Wallace and Weiss (2015) (aggregated to the province level and extended to Tibet) and column 2 uses data from Che et al. (2015). Observations are at the province-day level. All regressions include day-of-week, month and province fixed effects and control for rain, extreme temperature and total social media posts. Standard errors (in parentheses) are clustered at the province level. Significance at the 1%, 5% and 10% is denoted by ***, ** and *, respectively.

Table 2.C5: Historical TV Dramas and Anti-Japanese Hate Speech: Heterogeneity across Provinces II

	(1)	(2)	(3)	(4)
Interaction with	Japanese occupation during WWII	Civilian casualties during WWII	WWII memorials	Anti-Japanese protests in 2010
Mean of interacted variable	0.64	1.65	0.22	0.14
	A. Dependent Variable: Percentage of social media posts with references to island conflict			
Predicted historical TV drama viewership (lagged)	2.641*	3.653**	2.847*	2.759**
	(1.508)	(1.352)	(1.428)	(1.295)
Interaction term	0.191	-0.725	-0.391	-0.151
	(2.187)	(0.793)	(2.165)	(2.829)
No. of obs.	11,315	10,220	11,315	11,315
R-squared	0.085	0.086	0.085	0.085
Mean dep. var.	0.11	0.11	0.11	0.11
	B. Dependent Variable: Percentage of social media posts with calls for anti-Japanese boycott			
Predicted historical TV drama viewership (lagged)	0.871	0.678	0.898	0.969*
	(0.712)	(0.571)	(0.560)	(0.475)
Interaction term	0.227	0.203	0.354	0.209
	(0.723)	(0.176)	(0.726)	(0.642)
No. of obs.	11,315	10,220	11,315	11,315
R-squared	0.030	0.033	0.030	0.030
Mean dep. var.	0.020	0.021	0.020	0.020

Notes: This table shows the estimated coefficients for the second stage regression (equation 2.4) of the effects of predicted historical TV drama viewership on anti-Japanese content on social media. The dependent variable is the percentage of social media posts containing references to the Senkaku/Diaoyu island conflict (Panel A) or calls for boycotts of Japanese goods (Panel B). The main regressors are the predicted fraction of the population over 4 who watched a historical TV drama between 6pm and 12am the day before and its interaction with the fraction of prefecture-level cities within a province that 1) were occupied by Japan during WWII, 3) have a Sino-Japanese war memorial and 4) had an anti-Japanese protest in 2010 and 2) civilian casualties during WWII (as percentage of 1936 population). Columns 1, 3 and 4 use data from Wallace and Weiss (2015) (aggregated to the province level and extended to Tibet) and column 2 uses data from Che et al. (2015). Observations are at the province-day level. All regressions include day-of-week, month and province fixed effects and control for rain, extreme temperature and total social media posts. Standard errors (in parentheses) are clustered at the province level. Significance at the 1%, 5% and 10% is denoted by ***, ** and *, respectively.

Table 2.C6: Historical TV Dramas and Anti-Japanese Protests and Social Media Hate Speech: Robustness to excluding TV Channels from the Same Province

	(1)	(2)	(3)	(4)	(5)	(6)
	Anti-Japanese Protests		Percentage of Social Media Posts with ...			
	Incidence	Onset	Japanese (derogatory)	Nationalist sentiment	Island conflict	Anti-Japanese boycott
	(1)	(2)	(3)	(4)	(5)	(6)
Predicted historical TV drama viewership (lagged)	0.463**	0.416**				
	(0.223)	(0.194)				
Predicted historical TV drama viewership			3.068*	4.352*	2.494	0.967*
			(1.729)	(2.468)	(1.484)	(0.501)
Island conflict in People's Daily	0.0689***	0.0623***	0.00102	0.285***	0.0393	0.0586**
	(0.00613)	(0.00678)	(0.0345)	(0.0463)	(0.0602)	(0.0219)
Sino-Jap. War Anniversary	0.251***	0.236***	0.1000	0.325	0.276	0.0587
	(0.0174)	(0.0214)	(0.119)	(0.199)	(0.184)	(0.0720)
No. of obs.	11,284	11,284	11,315	11,315	11,315	11,315
R-squared	0.14	0.13	0.015	0.029	0.085	0.030
Mean dep. var.	0.012	0.011	0.052	0.13	0.11	0.020

Notes: This table shows the estimated coefficients for the second stage regression (equation 2.4) of the effects of predicted historical TV drama viewership on anti-Japanese protests and social media hate speech. The dependent variable is a dummy variable for whether a protest occurred (column 1) or a new protest started (column 2) or the percentage of social media posts between 6pm and 12am containing derogatory terms for Japanese people (column 3), nationalist sentiment (column 4), references to the island conflict (column 5) and calls for boycott of Japanese goods. The main regressors are the predicted fraction of the population over 4 who watched a historical TV drama between 6pm and 12am the day before (columns 1 and 2) and the same day (columns 3 to 6), excluding a province's own TV channel. Observations are at the province-day level. All regressions include day-of-week, month and province fixed effects and control for rain and extreme temperature, columns 3 to 6 also control for total social media posts. Standard errors (in parentheses) are clustered at the province level. Significance at the 1%, 5% and 10% is denoted by ***, ** and *, respectively.

Chapter 3

Economic performance, land expropriation and bureaucrat promotion in China

with Paul Dutronc-Postel (PSE, IPP)

We wish to thank Simone Bertoli, Sylvie Démurger, Matthew Gentzkow, Marc Gurgand, Emeric Henry, Philipp Ketz, Jonathan Lehne, Nancy Qian, Hillel Rapoport, David Strömberg, Ekaterina Zhuravskaya and participants at the Paris School of Economics, CEPREMAP India-China Conference, 2019 PEDD Conference, 2019 AFSE meeting, 2019 SIOE Annual Conference and 5th DIAL Development Conference for valuable comments. We also thank Cheuk-Ting (Regina) Hung for excellent research assistance. We gratefully acknowledge funding provided by the PSE Development Group and from the OSE LabEX.

Abstract

How effective are career incentives in motivating bureaucrats? We collect the career histories of the top bureaucrats of all Chinese prefectures between 1996 and 2014 and identify the causal effect of career incentives by exploiting variation in the ex ante competitiveness of promotions. Bureaucrats with a smaller starting cohort have a greater likelihood of promotion. This incentivises them to adopt a strategy that relies on real estate investment and rural land expropriation, resulting in faster growth in construction and GDP. We present suggestive evidence that the same incentives result in lower investment in education, public transport and health. We corroborate our findings using survey and remote sensing data, and show that land expropriations are associated with adverse outcomes for expropriated individuals, with subsequent arrests of local officials, and with the emergence of "ghost cities".

Keywords: Personnel Management, Bureaucracy, Economic Performance, Land Expropriation

JEL Codes: D73, H11, O43, P26

3.1 Introduction

Bureaucrats around the world are responsible for implementing social and economic policies. Yet it is difficult to motivate bureaucrat performance, as their chances of being demoted or fired are low and pay follows seniority-based rules. Moral hazard problems in bureaucracies can be potentially mitigated through performance-based promotion criteria. A number of papers have cited yardstick competition between local government officials as key for China's fast growth in recent decades and showed a positive relationship between performance and subsequent promotions for bureaucrats (see, for example, Maskin, Qian and Xu, 2000; Li and Zhou, 2005; Jia, Kudamatsu and Seim, 2015).

In this paper, we identify the causal effect of ex ante promotion incentives on a bureaucrat's performance and policy choices during their term. Different starting cohort sizes lead to variation in the competitive pressure for individual officials. A smaller cohort increases the likelihood of promotion, while being uncorrelated with a range of individual and prefecture-level characteristics. We find that career incentives push bureaucrats to choose a policy mix that leads to faster growth through increased construction and real estate investment, but at the cost of increased expropriation of rural land. We also find suggestive evidence that these incentives negatively affect public good provision. The welfare effects of performance-based promotions can thus be ambiguous.

Using both administrative and online sources, we identify the Chinese Communist Party (CCP) secretaries of all prefectures in China from 1996

and 2014. We collect data on their personal characteristics, as well as their past and future career path. This information allows us to assess whether a bureaucrat was promoted, demoted, transferred laterally, or retired at the end of their term.

Identifying the causal effect of promotion incentives on policy choices is subject to many potential endogeneity concerns. A bureaucrat's unobserved personal characteristics may jointly determine his performance and his ability to advance in the hierarchy. For instance, prefecture party secretaries with personal connections might be placed in higher growth locations. In this paper, we use variation in an individual bureaucrat's competitive environment as exogenous shocks to promotion incentives.

Our measure of competition intensity for each CCP secretary is the size of his starting cohort, that is, the number of other prefecture CCP secretaries who start their term at the same time in the same province. The size of a prefecture CCP secretary's starting cohort is unrelated to his own age, ethnicity, education and whether or not he is a native of the province. Starting cohort size also does not affect a bureaucrat's assignment to particular prefectures. However, the probability that a CCP secretary of a given prefecture is promoted after his term depends negatively on the size of his starting cohort. A prefecture party secretary who starts with four other party secretaries is 10 percentage points more likely to be promoted than a party secretary with twice as many competitors.[1]

The effect of having more competitors on performance is ex ante am-

[1]We also find some evidence of non-linearities in the data. When there is very little competition, an increase in cohort size increases promotion likelihood.

biguous. While increased competition could incentivise bureaucrats to exert more effort, it could also have the opposite effect of discouraging effort. Using a simple theoretical model, we show that the CCP promotion system generates incentives akin to contests between a varying number of players for a fixed number of prizes. A smaller number of competitors increases a bureaucrat's incentives to provide effort, as a given amount of effort translates into higher chances of getting promoted.

We find that fewer competitors increases GDP growth rates over a prefecture party secretary's term: a one standard deviation decrease in the size of a CCP secretary's starting cohort (or 2.4 fewer competitors) increases annual nominal GDP growth by 0.8 percentage points and real GDP growth by 0.3 percentage points. These faster growth rates are driven by higher real estate investment and faster growth in construction employment. We corroborate our findings using satellite data and find significant increases in both nightlight intensity and growth of urban areas due to stronger promotion incentives.

This type construction-led growth strategy relies on increased expropriations of rural land, which we measure directly using retrospective land history data from the China Health and Retirement Longitudinal Survey (CHARLS). Bureaucrats can resell the user rights of expropriated land to private developers in order to boost government revenue, investment and GDP figures, as well as undertake large-scale infrastructure projects. A one standard-deviation decrease in the size of a CCP secretary's cohort increases the probability of expropriation by 14 percent for an individual living during this secretary's term.

The Chinese central government estimated that over 20 percent of land was illegally expropriated by local officials.[2] We find that bureaucrats who expropriated more land are subsequently more likely to be arrested during the central anti-corruption campaign. Prefectures, where more rural land was expropriated, are also more likely to become "ghost cities". For the farmers losing their lands, we find that being expropriated is associated with adverse outcomes later in life. According to news reports, inadequate compensation and forced eviction of farmers trigger two thirds of all social protests in China.[3]

Higher-powered promotion incentives could also shift a bureaucrat's attention away from policies that are less visible. We find suggestive evidence that party secretaries with fewer competitors underinvest in education, transport and health, as measured by the number of teachers per capita, the number of buses per capita, and the number of doctors and hospital beds per capita in the prefecture.

Our paper contributes to a wider literature on understanding the impact of bureaucracies on economic performance (see, for example, Finan, Olken and Pande, 2017, for a recent review). We use the institutional context to identify exogenous variation in the intensity of promotion incentives that a bureaucrat faces. In this respect, our paper is similar to Bertrand et al. (2020) who use a bureaucrat's age at entry into the Indian civil service as a measure for career incentives and find that stronger incentives result in better performance. While our results also show that career incentives can

[2]See `http://en.people.cn/90001/90778/6272123.html`.

[3]`https://www.reuters.com/article/us-china-land/china-ministry-urges-end-to-forcible-land-requisition-paper-idUSBRE94E04320130515`

lead to better headline measures, we also highlight the potential detrimental effects of performance targets on less easily observable outcomes.

This connects our work to the theoretical literature on multitasking problems in principal-agent relationships (see Holmstrom and Milgrom, 1991). In a context where a principal wants to incentivise the completion of several tasks by an agent, the desirability of providing incentives for any one activity decreases with the difficulty of measuring performance in other activities that make competing demands on the agent's time and attention.

We build on the economics and political science literature on Chinese bureaucrats (see Section 3.2 for more details). While previous papers largely focus on the question of whether higher growth leads to promotions, in this paper we ask the reverse, that is, how ex ante promotion incentives affect policy choices. Our paper provides a framework for understanding the trade-offs inherent in the Chinese bureaucratic promotion system between spurring fast growth (see, e.g., Li and Zhou, 2005; Jia, Kudamatsu and Seim, 2015) and its socially undesirable by products (see, e.g., Fisman and Wang, 2017; Jia, 2017).

Our paper also relates to the literature on the importance of land security in development. Chen and Kung (2019) show how Chinese bureaucrats give land price discounts to firms that are connected to higher-ranking officials in return for better promotion prospects. Our findings suggest that this is made possible by local officials' ability to expropriate rural land. Insecure land rights are at the core of many aspects of agricultural and economic development in China (see, e.g., Li, Rozelle and Huang, 2000;

De La Rupelle et al., 2009) and elsewhere (see, e.g., de Janvry et al., 2015; Field, 2007; Goldstein and Udry, 2008).

Section 3.2 provides a brief description of the Chinese bureaucracy and land market. Section 3.4 describes the data. Section 3.5 discusses our empirical strategy and identifying assumptions. Section 3.6 presents our results, and Section 3.7 discusses their robustness. Section 3.8 concludes.

3.2 Background

3.2.1 Chinese bureaucratic system

China is geographically divided into several administrative levels: provinces, prefectures, counties and villages. At each level, the Chinese Communist Party (CCP) secretary is the de facto highest-ranking official. In this paper, we focus on the party secretaries of China's 334 prefectures and 15 vice-provincial cities, which are ranked between a prefecture and a province. Promotions of prefecture party secretaries are typically decided by their immediate superiors, that is, the provincial party standing committee.[4] Therefore, the main competitors for a prefecture party secretary are other prefecture party secretaries within their province.

The criteria for bureaucrat promotions in China have been subject to debate in the literature. Chen, Li and Zhou (2005); Li and Zhou (2005)

[4]Most of these officials will have spent all of their careers within one province. While there are some exchanges of officials across provinces, these are mostly short-term exchanges or lateral transfers

show that high GDP growth is positively correlated with subsequent promotions of provincial-level officials, while Persson and Zhuravskaya (2016) and Jia, Kudamatsu and Seim (2015) find that connections to the local elites and patrons in the central government also play an important role. In contrast, Landry, Lü and Duan (2018) find no relationship between economic growth and promotions at the province and prefecture levels and a positive relationship at the county level. Chen and Kung (2016) show that higher land revenue weakens the link between economic growth and promotions for county-level officials.

Bureaucrats face an age limit for promotions which varies with their rank. Prefecture party secretaries who are 55 or older are in theory no longer eligible for promotions, although we find that this rule is not always enforced. Officially, a bureaucrat's term should be five years long and bureaucrats should not be promoted before spending at least three years in a given position. However, we show in Appendix 3.C that terms tend to be significantly shorter in reality. This is in part a response to the age limits for promotions. Kou and Tsai (2014) show that officials who graduated from university and spent five years in each position will never be eligible for promotion beyond prefecture level. Thus, an optimal career path of a CCP official is a succession of short terms or "sprinting with small steps".[5]

Older bureaucrats rarely retire outright when their term ends, but are instead moved to consultative positions with little real power outside of the CCP and government. Demotions and dismissals of CCP officials were rare before Xi Jinping's anti-corruption campaign in 2012. Since then, over

[5]小步快跑, *xiǎo bù kuài pǎo*. See, for instance, *Nanfang Ribao*, 15 July 2011.

1.34 million officials have been punished in Xi's first five years in office.[6]

3.2.2 Land markets in China

In China land use is strictly governed and all land is divided into rural or urban land. Urban land is owned by local governments, who can auction user rights to developers for real estate or industrial use. In contrast, most rural land is owned by village collectives and allocated to registered rural residents for their housing needs and agricultural purposes. In recent years, some urban areas of China have seen a real estate boom and a resulting large disparity in the value of rural and urban land. This generates strong incentives for local governments to develop farm land, despite the central government's objective to preserve China's farmlands to ensure food security. Local governments are able to "retake" land from farming collectives and are only obliged to compensate farmers for the agricultural productivity of the land rather than its market value. The revenue from such land conversions accrue to the local government and is exempt from the tax-sharing agreement with the centre. These types of land seizures have become a major source of social unrest in China.[7]

[6] See `http://www.bbc.com/news/world-asia-china-41670162`.

[7] See, e.g. `https://www.nytimes.com/2013/09/09/world/asia/as-chinese-farmers-fight-for-homes-suicide-is-ultimate-protest.html`

3.3 Theoretical Framework

We conceptualise our setting as a contest between multiple players, who compete for a fixed number of prizes by expending effort across potentially multiple tasks. (See Appendix 3.B for a formal model.) Prefecture CCP secretaries compete against other members of their starting cohort for promotion to a limited number of higher offices. Promotions are decided by a principal (the provincial party standing committee) on the basis of measured GDP growth outcomes during an agent's (prefecture CCP secretary's) term. A prefecture CCP secretary chooses a policy over his term to increase his chances of promotion, but effort expended on implementing different policies is costly. We distinguish between two types of policies. The first is based on expropriating rural farm land, which results in higher government tax revenue and construction-led growth. According to Yew (2012), "local governments [...] seek to outdo each other in 'place-making', both to attract investments and to conjure up highly visible trophy projects". The second focusses on public goods provision, which might be desirable for the local population (and the bureaucrat himself), but does not lead to higher growth outcomes over a party secretary's term. The first type of effort increases an official's chances of promotion, while the second does not.[8] This set up incorporates a multi-tasking principal-agent problem (Holmstrom and Milgrom, 1991) in a rent-seeking game (Pérez-Castrillo and Verdier, 1992).

Our model generates the following predictions. The optimal amount

[8]This is consistent with the findings of Persson and Zhuravskaya (2016) for provincial leaders in China.

of effort officials spend on promotion-relevant policies decreases with the number of competitors, while their effort on promotion-irrelevant policies increases. In equilibrium, the probability of promotion of any official is lower when there are more competitors.[9] The intuition is the following: when there are more competitors, officials know that their chance to be promoted is lower, therefore they spend less effort on promotion-related activities and more on other pro-social policies. In our setting, we predict that a smaller starting cohort increases prefecture party secretaries' chances of promotion, increases land expropriation, construction and GDP growth and lowers pro-social spending.

3.4 Data and descriptive statistics

3.4.1 Prefecture CCP secretaries

We assemble the full history of CCP secretaries of all 334 prefecture-level administrative units of China from 1996 to 2014 using official Provincial Yearbooks.[10] For each CCP secretary, we collect their birth dates, other individual characteristics (such as education, gender, ethnicity) and full career history using official sources, Wikipedia and Baidu Baike (a Chinese

[9]In the model, we characterise a symmetric equilibrium. CCP secretaries are aware of their competitors and all behave in the same way. Nitzan (1994) show that in an asymmetric equilibrium, where the highest effort wins the contest with certainty, the expectation of mixed strategies can also be a decreasing function of the number of competitors. The negative relationship between effort and competitors may no longer hold when there are sufficiently large asymmetries in candidates' cost of effort or characteristics that affect promotions.

[10]Since the precise date of nomination is not always available, we use the name of the official in office at the end of the year

online encyclopedia similar to Wikipedia).

We identify promotion, lateral transfer, retirement and demotion or dismissal at the end of each official's term. An official is promoted if they attain a position with a higher official administrative rank after the end of their term. We account for the practice of moving older officials into honorary positions outside of the party and government by coding these moves as retirement, even when these positions are theoretically higher ranked.[11] Additional details on how we code bureaucrats' careers are given in Appendix 3.C.

A prefecture party secretary in China has a 49 percent chance of being promoted at the end of their term over the period from 1996 to 2014 (Table 3.C1), with considerable variation across prefectures and years. The average size of a starting cohort for a prefecture party secretary is 4 (with a standard deviation of 2.5) and also varies across time and space.

We identify party secretaries who were investigated or arrested under Xi Jinping's anti-corruption campaign by searching the Baidu Baike entry of each official for a set of keywords.[12] We manually check the correctness of our coding and cross-validate our findings with data from ChinaFile.[13]

[11]This is a similar approach to Li and Zhou (2005), who consider provincial CCP secretaries.

[12]We identify these by looking through the records of a given CCP secretary for a match on any of the following keywords: 'Expulsion from public office' (开除公职), 'Expulsion from Party membership' (开除党籍), 'Corruption' (腐败), 'Bribery' (受贿), 'Legal inspection' (依法审查), 'Suspended for inspection' (停职检查), 'Double designation' (双规; this term is specific to the Chinese Communist Party and refers, as 双开, to a type of internal investigation, that is generally kept secret), 'Illegal' (违法), 'Violation of regulations' (违规), 'Violation of rules' (违纪), 'Crime' (犯罪), 'Imprisonment' (有期徒刑). We then manually check the entries that match those keywords.

[13]See `http://www.chinafile.com/infographics/visualizing-chinas-anti-corruption-campaign`.

Of 1310 prefecture party secretaries, 165 (13 percent) were investigated and faced disciplinary sanctions. According to Chen and Kung (2019), the campaign singled out land-related crimes as one of its main targets.

3.4.2 Land expropriations

We use the 2014 China Health and Retirement Survey for data on rural land expropriation. The survey is representative of the Chinese population aged 45 and above. It contains 11,942 individuals and covers 450 communities across 120 prefectures and 28 provinces of China.[14] CHARLS contains the full retrospective life history of surveyed individuals, as well as their personal characteristics. Instances of land expropriation are documented, with the corresponding date, acreage and potential compensation and we assign each instance to the prefecture where the respondent is registered at the time. Figure 3.C3 shows that the majority of expropriation events occurred after 1996, corresponding to the start of our sample.

We construct a balanced panel of 11,146 individuals over 18 years for whom we have the complete residency history and an unbalanced sample of all individual-year observations where the prefecture of residence is not missing, as in 95.1 percent of observations. We compute the expropriation rate of a given prefecture in a given year as the number of individuals who were expropriated in that year over the total number of rural residents in that prefecture in that year. The average annual expropriation probability is 0.6 percent. 11 percent of respondent have been expropriated at least

[14] As the sampling is at the individual level, we discard information from respondents' spouses.

once over their lifetime.

3.4.3 Macroeconomic data

We use prefecture-level macroeconomic data from the Chinese City Statistical Yearbooks compiled by the National Bureau of Statistics. We rely on the following measures, which are consistently measured throughout the period: nominal and real GDP growth, sectoral nominal GDP growth and employment and real estate investment. As proxies for public goods provision, we use the number of primary school and middle school teachers and the number of doctors and hospital beds in public hospitals and clinics.

Summary statistics are presented in Table 3.C2. The average prefecture over the sample period has 4 million inhabitants and measures 23,000 km^2, with an average nominal GDP growth of 16 percent and average real GDP growth of 13 percent.[15] Real estate investment is an important component of GDP and accounts for on average 6 percent of its value.

The reliability of official Chinese statistics have been widely questioned in the media, although the academic literature on this topic has not always been conclusive.[16] We address this concern in several ways. First, we

[15]These values, even when weighted for the size of prefectures, are larger than the official national-level average GDP growth rates. This is due to the City Statistical Yearbooks' better coverage of richer and more urban prefecture-level cities and worse methodology than aggregate-level data, which are corrected using centrally implemented surveys (see Holz, 2014).

[16]A diplomatic telegram from the ambassador of the United States of America to China, addressed to the USA Secretary of State in 2007 and released by Wikileaks, states for instance: "GDP figures are 'man-made' and therefore unreliable, [then-executive vice premier] Li [Keqiang] said. [...] When evaluating Liaoning's economy, he focuses on three figures: 1) electricity consumption, which was up 10 percent in Liaoning last year [...] All other figures, especially GDP statistics, are 'for reference only,' he said smiling."

find consistent results on macroeconomic outcomes which are not as politically sensitive as GDP. Second, we corroborate our findings using remote sensing data. Third, we analyse the presence and distribution of decimals in reported growth figures, a common strategy in the literature on data manipulation (see, for instance Holz, 2014).

3.4.4 Remote sensing data

Night light intensity Following Henderson, Storeygard and Weil (2011, 2012), we use the intensity of night lights in a given prefecture as a proxy for GDP growth. Between 1996 and 2014, night light intensity over China was measured by six different satellite systems, which introduces structural breaks in the raw data. To deal with this issue, we use the smoothed time series data compiled by Zhang, Pandey and Seto (2016) and focus on the percentile rank of each prefecture in the distribution of growth rates rather than absolute growth as our outcome variable. We compare how well the raw night light data and the processed data approximate aggregate Chinese GDP in Appendix 3.D.

Urban expansion We use satellite data processed by Yao et al. (2018) to measure the extent to which urban areas in each prefecture expand.[17] For similar reasons to those pertaining to night light data, we use the percentile rank of urban area growth as our outcome variable.

https://wikileaks.org/plusd/cables/07BEIJING1760_a.html.

[17] Our results are robust to using alternative data from He, Huang and Ye (2014).

"Ghost cities" Rapid (and potentially excessive) urban expansion have led to the emergence of so-called "ghost cities" in China with particularly high vacancy rates.[18] Jin et al. (2017) use remote sensing data and machine learning to identify these "ghost cities" through discrepancies in activity between older and newer urban settlements at a localised level.[19] We use three lists of "ghost cities" provided by Jin et al. (2017) to assess the efficiency of land expropriation over the sample period in China.

3.5 Empirical strategy

We test the empirical predictions of our model using the size of a prefecture party secretary's starting cohort as the number of relevant competitors. We argue that this variation is plausibly exogenous, allowing us to estimate the causal effect of promotion incentives on bureaucrat performance.

3.5.1 Effect of competition on promotion

We estimate the following equation to show that our measure of competition intensity has an impact on the promotion likelihood of CCP secretaries:

$$P_{c,t} = \zeta Z_{c,t} + \alpha_c + \delta_t + \eta_{c,t} \tag{3.1}$$

[18] One of the examples that has received the most media attention is Ordos (鄂尔多斯) in Inner Mongolia.

[19] Activity is measured using, for instance, data on Internet usage and local points of interest.

where c denotes prefecture and t denotes year. $P_{c,t}$ is a dummy variable that takes the value of 1 in each year t of a prefecture c's party secretary's term if he is promoted at the end of his term. $Z_{c,t}$ is the size of a prefecture party secretary's starting cohort, that is the number of prefecture party secretaries, who start their first year of term in the same year as said party secretary. $Z_{c,t}$ takes the same value for each year of a CCP secretary's term. α_c and δ_t are prefecture and year fixed effects, respectively. As $Z_{c,t}$ is symmetric for all prefecture party secretaries who start their term in the same year in the same province, we cluster the standard errors $\eta_{c,t}$ at the province - start year level.

Prefecture fixed effects account for differences in the number of prefectures across provinces, as well as other time-invariant prefecture-level characteristics. Year fixed effects allow us to rule out China-wide macroeconomic shocks, for example, changes in the central CCP leadership which generate increased turnover at lower levels and other shocks to the supply of bureaucrat candidates. The residual variation in starting cohort size after controlling for location and time fixed effects is primarily driven by variation in the number of their predecessors who exit each province each year, which depends on their age and seniority. A potential identification challenge remains if particularly able or motivated bureaucrats could choose their prefecture and start year so as to minimise competition. We show in the next sections that $Z_{c,t}$ is not correlated with a range of individual party secretary characteristics and that all our results are robust to controlling for CCP secretary characteristics. The size of a bureaucrat's starting cohort is not correlated with trends in the bureaucrat's prefecture *prior* to the bureaucrat's term.

3.5.2 Effect of competition on policy choices

To identify the impact of promotion incentives on officials' policy choices, we estimate the following relationship between the number of competitors of a CCP secretary and different policy outcomes:

$$Y_{c,t} = \beta Z_{c,t} + \alpha_c + \delta_t + \epsilon_{c,t} \tag{3.2}$$

where $Y_{c,t}$ denotes an outcome in prefecture c in year t, such as, GDP growth, construction investment, land expropriation and public goods provision. The rest of the notation is as before and standard errors are clustered at the province-start year level. This model implicitly assumes that β is invariant over an official's term and weights longer terms more than shorter ones. In Section 3.7, we show that our results are robust to relaxing both assumptions. We also report results on the effects of competition on land expropriation using individual-level data from the CHARLS survey.

3.6 Results

3.6.1 Promotions and the number of competitors

Table 3.1 and Figure 3.1 shows how competition, in terms of the size of the starting cohort, affects the promotion likelihood of CCP secretaries. An increase by 1 standard deviation (or 2.4) in the number of competitors in a CCP secretary's cohort decreases his likelihood of promotion by 6.5 percentage points (or 13 percent of a standard deviation), relative to a

mean probability of promotion of 48 percent. This effect is statistically significant at the 1 percent level.

Figure 3.1 also points to non-linearities in the relationship between promotion and the number of competitors. When starting cohort size is very small, the average probability of promotion increases until a starting cohort size of three and then declines almost linearly until a cohort size of eight. Approximately two thirds of the observations lie in this downward sloping region and drive the empirical finding that on promotions are inversely related to the number of competitors (see Appendix Figure 3.C1).

The age at which a secretary takes office also has a significant and substantial negative effect on his probability to be promoted. This appears consistent with anecdotes that well-connected bureaucrats (for example, heirs of CCP leaders or "Princelings") systematically take office at a young age and are quickly promoted. Age at entry is not correlated with the size of an official's cohort (Figure 3.2). We control for age in all our subsequent regression to increase precision.

These results are consistent with our model predictions that a larger number of competitors weakens the promotion incentives facing each bureaucrat and lowers their realised chances of promotion. In order to identify how this change in promotion incentives affects bureaucrat's policy choices, we have to assume that the size of the starting cohort only affects policy outcomes during a bureaucrat's term through its effect on promotion incentives.

A broader threat to identification is the potential correlation between

the size of the starting cohort and characteristics of CCP secretaries, which may impact both their career path and their policy choices during their term. Data on CCP secretary characteristics, such as their gender, ethnicity and education level are only partially available, but we can show in Table 3.2 that there are no systematic correlations between the size of the starting cohort and individual CCP party secretary or pre-determined prefecture characteristics. We show in Section 3.7 that our main results remain robust to controlling for these variables.[20]

The size of a CCP secretary's starting cohort is determined by events *before* he enters office, such as the the average age of prefecture-level secretaries in the province in the preceding year and the initial structure of terms in 1996 (see Section 3.7).

3.6.2 Economic growth and number of competitors

In Table 3.3, we show that GDP growth under a prefecture party secretary is higher when they face stronger promotion incentives, that is, when they enter offce with fewer competitors. A one standard deviation (2.4 competitors) decrease in the size of the starting cohort increases nominal GDP growth by 0.8 percentage points on average (5 percent of the sample average of 15.8 percent), or 9 percent of a standard deviation. This effect is significant at the 1 percent level, even when adjusting for multiple hypothesis testing. We find a similar result for real GDP growth: a one standard

[20]The coefficient on the dummy variable for whether a prefecture party secretary graduated from a Top A class university is negative and significant at the 10 percent level. This is consistent with chance. And we show that the results are robust to controlling for this variable.

deviation decrease in the number of competitors is associated with a 0.3 percentage point increase in real GDP growth or 2.5 percent of the mean growth rate of 12.7 percent. In Tables 3.A13 and 3.A12 we show that this growth is accompanied by increases in employment and mainly driven by the secondary sector.

We find consistent results using remote sensing data, the percentile rank of nightlight growth of a prefecture within a province is positively associated with a smaller number of competitors (Table 3.3, column 3). This relationship is, however, only marginally statistically significant at conventional levels, which could be due to the the measure's noisiness (see Appendix 3.D).

3.6.3 Construction-led growth

How are officials able to generate additional growth? Our model predicts that higher promotion incentives should lead to increased effort by officials on more visible and promotion-relevant activities. The Chinese economic growth model has been characterised by high levels of investment, notably in infrastructure and construction. In Table 3.4 we show evidence of a shift towards a construction-led growth strategy in response to local officials' promotion concerns. We find that the ratio of real estate investment to GDP is on average higher when the CCP secretary has fewer competitors. A decrease by one standard deviation in a CCP secretary's cohort size also increases the real estate investment to GDP ratio of the prefecture by 0.3 percentage points, or 5 percent relative to a 6.6 percent sample average.

This is in large part due to investment in residential real estate, which increases by 0.2 percentage points of GDP (or 3.5 percent of the sample mean). We also find a large and positive, but not statistically significant at conventional levels, effect on employment growth in the construction sector.

We further find that this strategy leads to an overall expansion of urban areas through expropriation of neighbouring farmland. Using satellite data, we show in column 5 of Table 3.4 that prefectures whose party secretaries face less competition experience faster growth in their urban area. In column 4 of Table 3.4, we find that a decrease in the number of competitors of a CCP secretary increases the average prefecture-level expropriation rate based on the CHARLS survey. This effect is large: a one standard deviation increase in a CCP secretary's starting cohort size induces a 0.1 percentage points (or 15 percent) increase in expropriations.

3.6.4 Consequences of land expropriations

A last contribution of our analysis is to shed light on the potential welfare implications of rural land expropriation. We address this in three steps.

First, we use the CHARLS survey to analyse the consequences of expropriations on expropriated individual's life trajectories (Table 3.5). Being expropriated is negatively correlated with the probability changing residence: having ever been expropriated is associated with a 3 percentage points (8 percent) lower probability to have changed residence at least

once. This is likely a result of individuals who have experienced expropriation being less confident in the security of their land usage rights. As a consequence, they became less inclined to migrate and take advantage of employment opportunities elsewhere (see, e.g., De La Rupelle et al., 2009). Having been expropriated is also associated with adverse health outcomes later in life, such as having suffered injuries (2 percentage points, or 22 percent of the mean value) or being hospitalized (2.8 percentage points, or 22 percent of the mean value). It also reduces access to employment in rural collectives or local state-owned enterprises, and results in lower access to pensions, and fewer assets, such as heating. These relationships are robust to the inclusion of time-invariant individual controls and location specific fixed effects.

Second, we look at the correlation between a CCP secretary's history of expropriations, and his probability of being caught in the nation-wide anti-corruption campaign. We find a positive relationship, in line with Chen and Kung (2019) findings of corrupt land market practices by local Chinese officials.

Third, we examine the potential implications for urban development that is made possible through rural land expropriation. We show in Table 3.7 a positive correlation between a prefecture's average expropriation rate throughout the period and the probability that it -or part of it- is categorised as a ghost city according to three different definitions listed in Jin et al. (2017). This is suggestive that some of the extra development at the expense of expropriated farmers leads to inefficient real estate development, in so far as high vacancy rates are evidence of resource misallocation.

3.6.5 Public goods provision

Our model suggests a trade-off for bureaucrats between construction-driven growth policies and other policies which are less visible and matter less for promotions. In Table 3.8 we consider the impact of promotion incentives on a range of different public goods, such as the provision of health services (as measured by the number of hospital beds per capita and the number of doctors per capita), transport services (buses per capita) and education (primary school teachers per capita and middle school teachers per capita), while controlling for confounding factors such as population size and student body size. We normalise these 5 measures and aggregate them into an average z-score.[21] The effect of the size of the competitor pool is positive and significant at the 5 percent level. An increase by one standard deviation in the number of competitors at the start of a CCP secretary's term increases this measure of public good provision by 1.5 percent of a standard deviation. Measured separately, all coefficients on different public goods outcomes are positive and relatively large in magnitude, but none are statistically significant at conventional levels.

We interpret this as suggestive evidence that some substitution might be at play. The presence of many competitors decreases each secretary's chance of being promoted, and thereby dilutes the incentives to provide effort towards promotion, via highly visible investment- and real estate-driven growth. Local officials may instead decide to foster public good provision, for instance if they derive utility from their prefecture's overall

[21] Each variable is normalised by subtracting the annual mean of the variable from each observation and dividing by its annual standard deviation. The aggregate z-score is the average of the 5 normalised indices.

welfare or rely on the patronage of local elites Persson and Zhuravskaya (2016), in addition to their personal promotion chances.

3.7 Robustness checks

3.7.1 Explaining variation in starting cohort size

Our identification relies on the fact that the size of a prefecture party secretary's starting cohort is determined before that particular official enters office and is therefore unlikely to affect his policies other than through promotion incentives.

Figure 3.3 and Table 3.A1 show that the initial allocation of terms in 1996 is predictive of the starting cohort size, especially in the earlier parts of the sample. The five-year term rule was not followed strictly and the remaining variation in starting cohort size is driven to a similar extent by previous prefecture party secretaries retiring or being promoted (Table 3.A2). One important factor that drives the starting cohort size is the seniority rule for promotion. As prefecture party secretaries who are older than 55 theoretically become ineligible for promotion, there is increased turnover among officials who reach this critical age and an increase in the starting cohort size the following year.[22]

[22]Bureaucrats who approach the critical age threshold are often evaluated first for promotion.

Predecessors' age We exploit variation in the starting cohort size due to the seniority rule, by using the number of prefecture-level secretaries in the same province who are over 55 in the year before a given prefecture party secretary enters office as an alternative regressor. Table 3.A3 shows that the results are broadly in line with our main results, notably with regards to nominal and real GDP growth, and expropriation rates.

Controlling for predecessors' promotion probability Similarly, we show in Table 3.A4 that controlling for the promotion rate of prefecture party secretaries in the same province in the year before a party secretary enters office does not alter significantly the magnitude or the precision of our estimates.[23] This suggests that our findings are not driven by serial correlation in promotion probabilities.

Placebo using lagged cohort size We also perform a placebo test, replacing the value of the cohort size by the value it would have taken had the bureaucrat taken office one year earlier. We present the results in Table 3.A5. The coefficient of the placebo cohort size is never precisely estimated and does not show a pattern that matches our main results.

3.7.2 Identification checks

A potential threat to our identification strategy are endogenous placements of bureaucrats, for instance, if local bureaucrats with connections

[23] As it does reduce the sample size (because of the availability of promotion data on predecessors), we do not include this variable in our main specifications.

can choose where and when they become prefecture-level CCP secretaries.

Controlling for additional CCP secretary characteristics We show in Table 3.2 that the size of a CCP secretary's starting cohort is unrelated to a range of observable characteristics, such as ethnicity and local origin. In Table 3.A6, we include these variables in our main specification (see Equation 3.2). This reduces our sample size considerably, from 306 to 260 clusters and from 2577 to 1826 observations. Yet, all of our results are robust in magnitude and sign to this new specification, and most are robust in statistical significance.

Controlling for pre-term prefecture outcomes Table 3.A7 shows that our results are robust to controlling for prefecture-level outcomes in the year before a party secretary's term begins. This alleviates, for instance, the concern that bureaucrats with higher expected promotion likelihood are placed with fewer competitors in areas where greater performance has just been observed.

3.7.3 Alternative specifications

Term-level specification An alternative specification for estimating the effect of promotion incentives on a prefecture party secretary's behaviour during his term is to collapse our data at the term level. The variation on competition (starting cohort size) is indeed constant at the level of a term. We average the various outcomes that we test in the previous section at

the term level and reproduce our results.[24]

We display our results in Table 3.A8. Overall, the results are in line with those obtained with our main specification. Effects on nominal GDP growth, share of real estate investment in GDP, growth of urban areas are of similar magnitude and statistical significance. Some effects lose statistical precision, such as real GDP growth, but the magnitude of these effects is indistinguishable from those obtained in the panel.

Weighted panel specification Collapsing the data at the term level does not allow us to control for annual variations in these variables. As an alternative, we also estimate Equation 3.2 using the inverse term length as a weight for each observation, to counterbalance the fact that longer terms carry more weight in our original specification than shorter ones. We display our results in Table 3.A9 and find that the magnitude, sign and significance of our results are largely unchanged.

Individual-level specification for land expropriation In order to leverage all individual-level information available in the CHARLS survey, we also estimate individual-level regressions, where the outcome variable

[24]We estimate the following relationship between starting cohort size and promotion likelihood:

$$P_{c,t_0} = \xi Z_{c,t_0} + \alpha_c + \mu_{t_0} + \eta_{c,t_0} \tag{3.3}$$

and our main specification on the effect of competition on policy outcomes becomes:

$$Y_{c,t_0} = \pi Z_{c,t_0} + \alpha_c + \mu_{t_0} + \eta_{c,t_0} \tag{3.4}$$

where t_0 denotes the start year of each term, μ_{t_0} are start-year fixed effects and the rest of the notation is as in equation 3.1.

is a dummy variable taking value 1 when the CHARLS respondent declared having been expropriated in a given prefecture in a given year.[25] Table 3.A10 shows that a smaller starting cohort is associated with a higher probability of being expropriated for individuals who live in said CCP secretary's prefecture during his term. The magnitude of the effect of a one standard deviation increase in cohort size of a CCP secretary ranges from a 0.09 percentage points (12 percent) to a 0.11 percentage points (14 percent) increase in the yearly expropriation probability of a CHARLS respondent. These effects are statistically significant at the 5 percent level.

3.7.4 Data quality concerns

Sectoral GDP and employment As headline GDP figures could be susceptible to manipulation, we show that competition also affects less politically sensitive measures, such as sectoral GDP growth and employment. Table 3.A12 shows that less competition increases GDP growth in particularly the secondary sector, consistent with a construction-led growth strategy. Employment growth also responds to promotion incentives (Table 3.A13).

[25] We estimate the following equation:

$$E_{i,c,t} = \beta Z_{c,t} + \mathbf{X'_{i,t}}\gamma + \alpha_c + \delta_t + \mu_{i,c,t} \tag{3.5}$$

where i denotes an individual. $E_{i,c,t}$ is a dummy variable that equals to 1 if individual i was expropriated in prefecture c in year t. $\mathbf{X'_{i,t}}$ are controls for a range time-varying and time-invariant individual characteristics, such as age, age squared, gender, ethnicity and a full set of educational attainment dummies or individual fixed effects. All other notation is as before and standard errors are clustered at the province-start year level.

Sensitivity to outliers The raw figures reported in statistical yearbooks at times exhibit a very high variance, we therefore trim our data to exclude the top and bottom outlying values of each variable.[26] Table 3.A14 shows that our results on GDP growth rates are all statistically indistinguishable from the ones obtained on non-trimmed data.

Evidence on data manipulation The distribution of second decimal digits in reported real GDP growth data of Chinese prefectures does not seem to follow Benford's law, as was previously noted by Ji (2019) (See Figure 3.A1).[27] This in itself is not necessarily proof that the underlying data are manipulated.[28] However, we show that the absence or presence of some specific digits is associated with the size of a CCP secretary's cohort (Table 3.A15). A larger size of a secretary's cohort decreases the likelihood that the real GDP growth rate ends with a 2 or 5, and increases the probability that it ends with an 8, and these effects are statistically significant at the 5 percent level: the distribution of decimal digits differs more from Benford's law when the number of competitors of a CCP secretary is small.[29] We treat these results as suggestive evidence that *in addition* to real effects, promotion incentives may also increase official data manipulation at the local level. However, our consistent results using several different data

[26]We keep all values that satisfy:

$$\text{abs}\left(\frac{Y_{it} - \bar{Y}_t}{\sigma_{Y_t}}\right) \leq 2.15$$

where $\bar{Y}_t$ is the average of Y and σ_{Y_t} its standard deviation across prefectures in year t.

[27]Benford's law, also named 'first digit law', states that "the frequency distribution of first digits in many—but not all—real world data is not distributed uniformly, but according to the widths of gridlines on a logarithmic scale" (Holz, 2014). This also applies to further digits following the first one (Smith, 1997).

[28]Rounding errors can, for instance, explain the over-representation of zeros and fives.

[29]Adjusted for multiple hypotheses testing, the results on 2 and 8 remain significant.

sources make us confident that promotion incentives have a real effect on bureaucrat's policy.

Sensitivity to CHARLS sample definition As the CHARLS survey is retrospective and some respondents have migrated prior to being surveyed in 2014, the average prefecture-level expropriation rates are based on a varying number of individuals. In order to test the robustness of our results to sample selection, we vary our sample in Table 3.A11 to either include all individuals and years where we have information (column 3) or only prefectures and years where we observe at least 20 individuals (our baseline specification, in column 1) or 50 individuals (column 2). These changes do not affect the sign, the magnitude nor the statistical significance of our results.

3.8 Conclusion

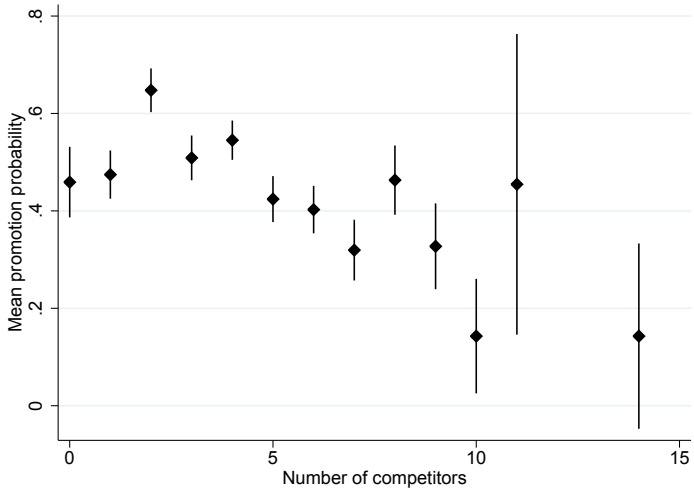
In this paper we investigate the relationship between promotion incentives, economic performance and land expropriation practices by prefecture-level CCP secretaries in China in the past 25 years. We use exogenous variation in the competitiveness of promotions based on the structure of local competition. We find that more competitors reduce the probability of any given party secretary being promoted to higher office at the end of his term. We build a simple theoretical model to show that less competition for promotions incentivises bureaucrats to engage in highly visible activities, such as construction investment. Using administrative, survey and

satellite data, we show that less competition leads to higher GDP growth, real estate investment, land expropriation and growth of the urban area of a prefecture. We also find suggestive evidence that there is a trade-off between construction-led growth and the wider provision of public goods. Finally, we document that land expropriations are linked to negative outcomes for the individuals who face them, corruption charges for the local officials who engage in them, and excess capacity for the cities where they take place.

Our paper contributes to understanding performance and policy incentives of a competitive bureaucracy. Similar to a tournament, Chinese bureaucrats see their chances of promotion increase when there is less competition. This increases the marginal return of effort in only the dimensions that matter for their evaluation. The results are local officials who rely on a model of economic growth driven by real estate investment and land expropriations, at potentially high social cost.

3.9 Figures

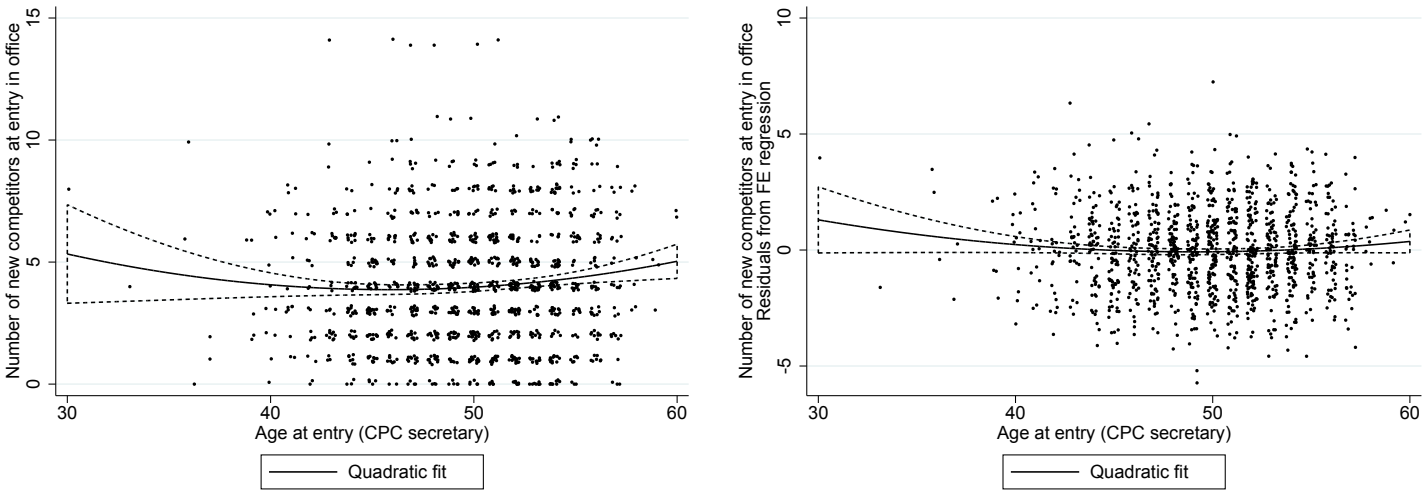
Figure 3.1: Promotion probability and size of starting cohort



Note: This graph shows the probability that average probability of a prefecture party secretary being promoted at the end of his term on the y-axis (95% confidence intervals in bars) against the size of his starting cohort on the x-axis. The size of the starting cohort is defined as the number of prefecture CCP secretaries who start their term in the same year and same province.

Figure 3.2: Size of starting cohort and CCP secretary's age

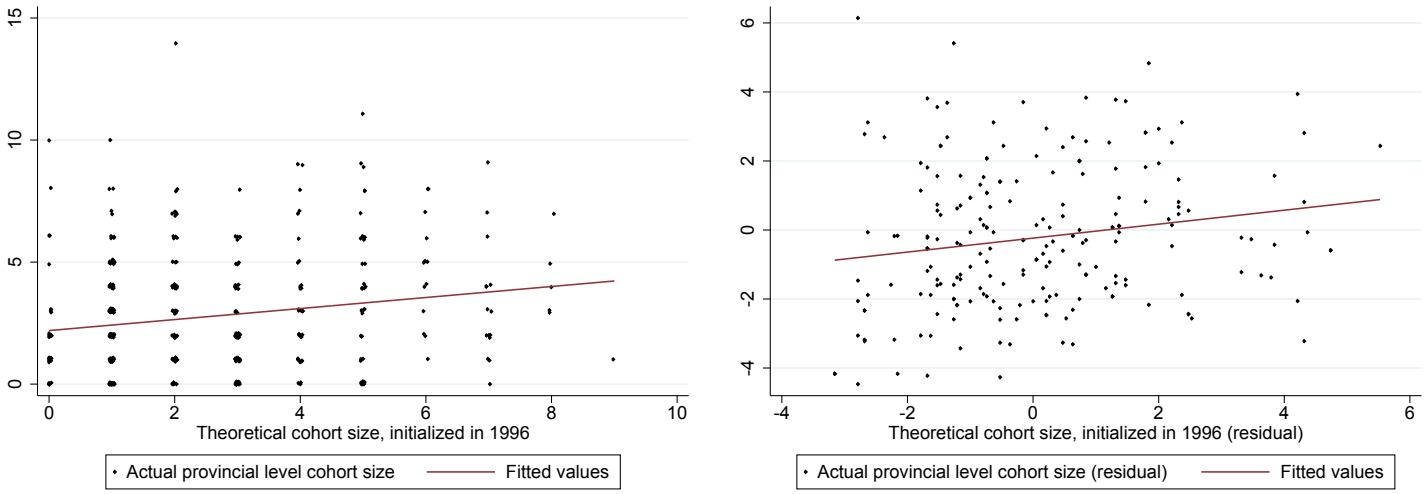
(a) Raw correlation (b) Correlation with FE residuals



Note: These graphs plot the value of the size of the starting cohort, against the age of the CCP secretary when he takes office. The size of the starting cohort is the number of other prefecture-level CCP secretaries who start their term in the province in which a given prefecture CCP secretary starts his term. Panel (a) plots the raw correlation and a quadratic fit. Panel (b) plots the Residuals of the starting cohort size regressed on prefecture-level and starting-year-level fixed effects. The data are from our database on 1310 terms of CCP prefecture-level secretaries.

Figure 3.3: Correlation between theoretical and actual starting cohort size

(a) Raw correlation (b) Correlation of residuals



Note: These graphs plot the value of the actual size of the starting cohort (an observation is thus a province × year), against the theoretical value of the starting cohort size, if starting 1996, all terms had been exactly 5 years long (as per the theoretical administrative rule). The size of the starting cohort is the number of other prefecture-level CCP secretaries who start their term in a given province in a given year. Panel (a) plots the raw correlation and a linear fit. Panel (b) plots the Residuals of the (actual and theoretical) starting cohort sizes regressed on province-level and starting-year-level fixed effects. The data are from our database on 1310 terms of CCP prefecture-level secretaries.

3.10 Tables

Table 3.1: Promotion probability and number of competitors (prefecture panel)

	Promotion likelihood
Size of starting cohort (Normalized)	-0.0647***
	(0.0188)
Age at entry	-0.0353***
	(0.00448)
Nb. of obs.	2680
Nb. of clusters	306
Mean of dep. var.	0.48

Observations are at the prefecture-year level. The sample only includes observations during terms that start at the earliest in 1996 and finish before 2014 and for which we observe the prefecture party secretary's promotion and age, as well as all macroeconomic variables of interest. The dependent variable is a dummy variable for whether the CCP party secretary is promoted at the end of his term. The regressor of interest is the size of the starting cohort, i.e., the number of other prefecture-level CCP secretaries in the province who start their terms in the same year as a given prefectural CCP secretary. Controls include the CCP secretary's age in the year he takes office (reported in the table), as well as prefecture and calendar year fixed effects. Standard errors are clustered at the province-start year level with significance at the 1%, 5% and 10% denoted by ***, ** and *, respectively.

Table 3.2: Correlation of starting cohort size with party secretary and prefecture characteristics

Panel A: Individual characteristics of prefecture party secretary

			Education				
	Ethnic minority	Native of the province	Top university (A) graduate	Top university (B) graduate	Top specialist university graduate	Party school graduate	Teaching college graduate
Size of starting cohort (Normalized)	-0.0182	0.00807	-0.0558*	-0.00388	-0.00121	-0.0315	-0.0149
	(0.0122)	(0.0314)	(0.0296)	(0.0105)	(0.0199)	(0.0332)	(0.0210)
Nb. of obs.	575	549	431	431	431	431	431
Nb. of clusters	244	235	212	212	212	212	212

Panel B: Prefecture-level characteristics before beginning of term

	GDP growth (nominal)	GDP growth (real)	GDP growth rank % GDP	Residential investment	Expropriation rate rank	Urban expansion Z-score	Public good provision
Size of starting cohort	0.000472	-0.00130	-0.0581	-0.00133	-0.000552	-0.0345	0.0133
	(0.00605)	(0.00193)	(0.145)	(0.00108)	(0.000740)	(0.152)	(0.0113)
Nb. of obs.	732	733	770	705	452	732	486
Nb. of clusters	300	300	306	285	237	301	209

Observations are at the term level. The dependent variables are characteristics of individual prefecture party secretaries in Panel A: non-Han ethnicity dummy (column 1), an indicator for whether the prefecture party secretary was born in the same province (column 2), the prefecture party secretary's education level is characterised by a set of dummy variables for his reported alma mater (columns 3 to 7). Top universities are those named as "Double Firsts" by the Chinese Ministry of education in 2015. The dependent variables are prefecture characteristics in the year before a party secretary's term starts in Panel B: nominal GDP growth (column 1), real GDP growth (column 2), GDP growth rank from satellite data (column 3), residential investment as a share of GDP (column 4), expropriation rate from survey data (column 5), urban area expansion rank from satellite data (column 6) and an average z-score for public good provision (column 7). The regressor of interest is the size of the starting cohort (i.e., the number of other prefecture-level CCP secretaries in the province who start their terms in the same year as said prefectural CCP secretary). Controls include the CCP secretary's age in the year he takes office, as well as prefecture and calendar year fixed effects (not reported in table). Standard errors are clustered at the province-start year level with significance at the 1%, 5% and 10% denoted by ***, ** and *, respectively.

Table 3.3: GDP growth and promotion incentives

Variables	GDP growth (nominal)	GDP growth (real)	GDP growth rank
Source	NBS	NBS	Satellite
	(1)	(2)	(3)
Size of starting cohort (Normalised)	-0.00835***	-0.00333***	-0.0923*
	(0.00244)	(0.00122)	(0.0548)
	[0.004]	[0.040]	[0.560]
Age at entry	0.000535	0.000235	-0.00792
	(0.000453)	(0.000227)	(0.0147)
	[1.000]	[1.000]	[1.000]
Nb. obs	2678	2677	2678
Nb. clusters	306	306	306
Mean of dep. var.	0.158	0.127	5.160

Observations are at the prefecture-year level. The sample only includes observations during terms that start at the earliest in 1996 and finish before 2014 and for which we observe the prefecture party secretary's promotion and age, as well as all macroeconomic variables of interest. The regressor of interest is the number of other prefecture-level CCP secretaries in the province who start their terms in the same year as said prefectural CCP secretary. Controls include the CCP secretary's age in the year he takes office (reported in the table), as well as prefecture and calendar year fixed effects. Standard errors are clustered at the province-start year level with significance at the 1%, 5% and 10% denoted by ***, ** and *, respectively. P-values adjusted for family-wise multiple hypothesis testing are reported in square brackets.

Table 3.4: Expropriations, real estate and promotion incentives

Variables	Employment Growth Construction sec.	Real Estate investment % GDP	Residential RE investment % GDP	Expropriation rate	Urban expansion rank
Source	NBS	NBS	NBS	Survey	Satellite
	(1)	(2)	(3)	(4)	(5)
Size of starting cohort (Normalized)	-0.0825	-0.00264*	-0.00178*	-0.00114**	-0.105**
	(0.0645)	(0.00138)	(0.000928)	(0.000550)	(0.0526)
	[1.000]	[0.279]	[0.283]	[0.200]	[0.232]
Age at entry	0.0102	-0.000172	-0.000249	-0.0000829	-0.0118
	(0.0125)	(0.000312)	(0.000241)	(0.000123)	(0.0159)
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
Sample definition					
Prefectures	All	All	All	$N_{expro} > 20$	All
Nb. obs	2678	2678	2648	1324	3054
Nb. clusters	306	306	305	218	341
Mean of dep. var	1.050	0.0656	0.0518	0.00778	5.195

Observations are at the prefecture-year level. The sample only includes observations during terms that start at the earliest in 1996 and finish before 2014 and for which we observe the prefecture party secretary's promotion and age, as well as all macroeconomic variables of interest. In column 4, we restrict the sample to the 120 sampling prefectures of the CHARLS survey, where the annual prefecture-level expropriation rate is estimated on at least 20 observations. In column 5, the dependent variable is the prefecture's annual national rank in the growth of total urban area based on satellite data. The regressor of interest is the number of other prefecture-level CCP secretaries in the province who start their terms in the same year as said prefectural CCP secretary. Controls include the CCP secretary's age in the year he takes office (reported in the table), as well as prefecture and calendar year fixed effects. Standard errors are clustered at the province-start year level with significance at the 1%, 5% and 10% denoted by ***, ** and *, respectively. P-values adjusted for family-wise multiple hypothesis testing are reported in square brackets.

Table 3.5: Land expropriation and individual outcomes

	Mobility		Health			Job History			Social security			Assets
	Change of residence (1)	Hukou ever converted to Urban (2)	Ever Suffered Injury (3)	Ever Confined in Bed (4)	Ever Hospitalized (5)	Job in SOE (6)	Job in Rural Collective (7)	No job in either (8)	Has Any Pension (9)	Has any Medical Coverage (10)	Has any Access to Healthcare (11)	Has Heating (12)
Ever expropriated	-0.0285**	0.0112	0.0213**	0.0255**	0.0284***	-0.0801***	-0.526***	0.559***	-0.0149*	0.000479	-0.00649	-0.0291***
	(0.0130)	(0.00887)	(0.00881)	(0.0111)	(0.0101)	(0.00689)	(0.0110)	(0.0114)	(0.00862)	(0.00951)	(0.00815)	(0.00972)
Male	0.00374	0.00179	0.0615***	0.0199**	0.0474***	0.0427***	0.00318	-0.0172**	0.0344***	0.0544***	0.00239	-0.0143**
	(0.00939)	(0.00587)	(0.00611)	(0.00799)	(0.00714)	(0.00582)	(0.00783)	(0.00787)	(0.00596)	(0.00645)	(0.00636)	(0.00706)
Ethnic minority	-0.0140	-0.0166	0.0120	0.0346*	0.0260	0.00770	0.00749	-0.0112	0.000484	0.00349	0.00905	0.0110
	(0.0243)	(0.0143)	(0.0154)	(0.0205)	(0.0179)	(0.0170)	(0.0193)	(0.0195)	(0.0151)	(0.0161)	(0.0181)	(0.0132)
Education level												
No schooling	-0.0481***	-0.0291***	0.0141*	0.00422	0.0158	-0.0433***	0.0375***	-0.0237**	-0.0374***	-0.0310***	0.0149	-0.0176*
	(0.0124)	(0.00671)	(0.00830)	(0.0113)	(0.0102)	(0.00652)	(0.0102)	(0.0100)	(0.00663)	(0.00739)	(0.00973)	(0.0101)
Primary school	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
Middle School	0.0315***	0.0544***	-0.000696	-0.0116	-0.00629	0.0581***	-0.0749***	0.0435***	0.0581***	0.0560***	-0.00914	0.0152*
	(0.0118)	(0.00733)	(0.00749)	(0.00942)	(0.00844)	(0.00776)	(0.00997)	(0.00985)	(0.00771)	(0.00827)	(0.00737)	(0.00855)
High School	0.0540***	0.129***	-0.0346***	-0.0172	-0.00973	0.174***	-0.159***	0.0439***	0.168***	0.141***	-0.0245**	0.0361***
	(0.0193)	(0.0147)	(0.0107)	(0.0143)	(0.0132)	(0.0162)	(0.0165)	(0.0167)	(0.0159)	(0.0158)	(0.0104)	(0.0124)
More than High School	0.137***	0.422***	-0.0358***	-0.00761	0.00262	0.503***	-0.260***	-0.120***	0.548***	0.510***	-0.0123	0.0746***
	(0.0239)	(0.0224)	(0.0130)	(0.0184)	(0.0176)	(0.0229)	(0.0216)	(0.0206)	(0.0218)	(0.0218)	(0.0141)	(0.0154)
Nb. Obs	10510	10510	10502	10500	10502	10510	10510	10510	10510	10510	10510	10510
Mean dep. var.	0.34	0.11	0.09	0.17	0.13	0.13	0.54	0.38	0.19	0.20	0.11	0.73

The sample in all columns consists in all 10,627 individuals in the CHARLS sample for whom land expropriation is available, except in column (2) where it is restricted to individuals whose first hukou is of rural type. Education level is included as a set of dummy variables, of which the reference level is primary school. Additional control variables include prefecture-level and year-of-birth fixed effects, as well as health status as child (5 dummy variables) and type of the first hukou. The main regressor of interest is a dummy variable taking value 1 if the individual has ever been expropriated. All columns correspond to linear models estimated using ordinary least-squares. Standard errors are clustered at the prefecture level with significance at the 1%, 5% and 10% denoted by ***, ** and *, respectively.

Table 3.6: Expropriation and ex-post arrest probability

Variables	$\mathbb{1}\{\text{Arrest}\}$ (1)	$\mathbb{1}\{\text{Arrest}\}$ (2)
Expropriation rate	3.997*	1.935
	(2.402)	(2.238)
Province FE	Yes	No
Prefecture FE	No	Yes
Nb. of obs.	380	380
Mean of dep. var.	0.15	0.15

Observations are at the prefecture-level. The sample only includes the CHARLS survey sampling prefectures, and years during terms that start at the earliest in 1996 and finish before 2014 and for which we observe the prefecture party secretary's promotion. The dependent variable is a dummy for whether a CCP secretary is caught in an anti-corruption campaign over the course of his career. The regressor of interest is the average expropriation rate in the prefecture and years where the CCP secretary was in office. Controls include the CCP secretary's age in the year he takes office (reported in the table), as well as province (columns 1) or prefecture (column 2) and start-year fixed-effects. Standard errors are robust to heteroskedasticity, with significance at the 1%, 5% and 10% denoted by ***, ** and *, respectively.

Table 3.7: Ex-post ghost city status

	P(Ghost city) (1)	P(Ghost city) (2)	P(Ghost city) (3)
Panel A: Weighted regression			
Prefecture-level average expropriation rate	0.0866*	0.131**	0.112*
	(1.77)	(2.16)	(1.72)
N	121	121	121
Panel B: Unweighted regression			
Prefecture-level average expropriation rate	0.0665	0.101	0.102
	(1.36)	(1.61)	(1.60)
N	121	121	121
Panel C: Panel A + Compensation rates			
Prefecture-level average expropriation rate	0.0931*	0.135*	0.120*
	(1.69)	(1.98)	(1.67)
Prefecture-level average compensation rate	-0.00102	-0.00136	-0.00221
	(-0.91)	(-0.97)	(-1.50)
N	114	114	114

The dependent variable in column 1 is a dummy variable indicating whether (all or part of the) prefecture has been identified as a ghost city by Jin et al. (2017). The dependent variable in column 2 is a dummy variable that further includes ghost cities identified by the BiaoZhun study (cited in Jin et al. 2017). The dependent variable in column 3 is a dummy variable that further includes ghost cities identified by Chi et al. (2016). The weights used in panels A, and C are the average population of the prefecture. In all regressions, the expropriation rate for a given prefecture is computed as the number of expropriations that occured between 1996 and 2014 in the prefecture divided by the number of survey respondents registered in said prefecture (all data reported from the CHARLS survey). In all regressions, the sample is restricted to the 121 endline prefectures (excluding prefectures where the sample size is very small or only represented by migrations). In Panel C, the compensation rate for a given prefecture is the share of expropriations that are reported to have been at least partly compensated. 7 prefectures where the expropriation rate is 0 in all years are dropped from the sample. Standard errors are robust to heteroskedasticity, with significance at the 1%, 5% and 10% denoted by ***, ** and *, respectively.

Table 3.8: Public good provision and promotion incentives

	Average Z-score	Variables				
		# Hospital beds per capita	# Doctors per capita	# Buses per capita	# Teachers per c. (primary)	# Teachers per c. (middle)
Source	NBS	NBS	NBS	NBS	NBS	NBS
	(1)	(2)	(3)	(4)	(5)	(6)
Size of starting cohort (Normalized)	0.0155**	0.190	87.16	0.146**	0.171	0.112
	(0.00748)	(0.147)	(59.50)	(0.0644)	(0.162)	(0.174)
		[0.995]	[0.721]	[0.123]	[1.000]	[1.000]
Age at entry	0.000400	0.000783	4.605	-0.0156	0.0373	0.000322
	(0.00167)	(0.0333)	(11.80)	(0.0171)	(0.0368)	(0.0298)
		[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
Nb. obs	2487	2601	2553	2489	2644	2625
Nb. clusters	303	303	302	303	305	306
Mean of dep. var.	-0.0577	11671.8	6571.8	6.465	18267.8	15629.9

Observations are at the prefecture-year level. The sample only includes observations during terms that start at the earliest in 1996 and finish before 2014 and for which we observe the prefecture party secretary's promotion and age, as well as all macroeconomic variables of interest. In column 1, the dependent variable is the average Z-score of the 5 variables in columns 2 to 6, where each variable is normalized by its yearly mean and standard deviation. The regressor of interest is the number of other prefecture-level CCP secretaries in the province who start their terms in the same year as said prefectural CCP secretary. Controls include the CCP secretary's age in the year he takes office (reported in the table), as well as prefecture and calendar year fixed effects. Standard errors are clustered at the province-start year level with significance at the 1%, 5% and 10% denoted by ***, ** and *, respectively. P-values adjusted for family-wise multiple hypothesis testing are reported in square brackets.

Appendix

3.A Robustness Tables and Figures

Table 3.A1: Theoretical cohort size and actual cohort size

	(1) All	(2) Before 2004	(3) Before 2000
Theoretical cohort size, initialized in 1996	0.0996*	0.262***	0.488***
	(0.0594)	(0.0782)	(0.106)
Observations	390	215	118

Observations are at the province-start year level. The dependent variable is the starting cohort size, defined as the number of prefecture-level party secretaries who start their term in the province in a given year. The regressor of interest is the theoretical starting cohort size as computed by extrapolating from the 1996 distribution of term lengths, and applying a fixed length of 5 years (following the theoretical length of a term). Controls include province level and start year level fixed effects. Standard errors are reported in parentheses with significance at the 1%, 5% and 10% denoted by ***, ** and *, respectively.

Table 3.A2: Starting cohort size and composition of predecessors

	(1)	(2)	(3)	(4)
Nb of t-1 cohort secretaries who are older than 55	0.116**			0.197***
	(0.0469)			(0.0449)
Nb of t-1 cohort secretaries who were promoted		0.972***		0.924***
		(0.0805)		(0.0461)
Nb of t-1 cohort secretaries who retired			0.942***	0.880***
			(0.0791)	(0.0460)
Nb. clusters	364	364	364	364

Observations are at the prefecture-year level. The sample only includes terms that start at the earliest in 1996 and finish before 2014 and for which we observe the prefecture party secretary's promotion and age, as well as all macroeconomic variables of interest. In each column, the dependent variable is the starting cohort size in a given province and year. Standard errors are clustered at the province-start year level with significance at the 1%, 5% and 10% denoted by ***, ** and *, respectively. P-values adjusted for family-wise multiple hypothesis testing are reported in square brackets.

Table 3.A3: Prefecture outcomes and promotion incentives:
Age of predecessors

Panel A: Economic growth

	GDP growth (nominal)	GDP growth (real)	GDP growth rank (Satellite)
Nb of t-1 cohort secretaries who are older than 55	-0.00350***	-0.00174***	-0.000692
	(0.00116)	(0.000583)	(0.00124)
Nb. clusters	291	291	291

Panel B: Real estate and expropriation

	Employment Growth Construction sec.	Real Estate investment % GDP	Residential RE investment % GDP	Expropriation rate	Urban expansion rank
Nb of t-1 cohort secretaries who are older than 55	-0.0502	-0.000944	0.000220	-0.000845**	0.00760
	(0.0401)	(0.000652)	(0.000514)	(0.000425)	(0.0288)
Nb. clusters	291	291	290	235	291

Panel C: Public good provision

	Average Z-score	# Hospital beds per capita	# Doctors per capita	# Buses per capita	# Teachers per c. (primary)	# Teachers per c. (middle)
Nb of t-1 cohort secretaries who are older than 55	0.000602	-0.0845	65.42**	0.0386	0.0862	-0.241***
	(0.00317)	(0.0742)	(28.34)	(0.0357)	(0.0687)	(0.0848)
Nb. clusters	284	289	287	288	290	291

Observations are at the prefecture-year level. The sample only includes terms that start at the earliest in 1996 and finish before 2014 and for which we observe the prefecture party secretary's promotion and age, as well as all macroeconomic variables of interest. In column 4 of Panel C, we restrict the sample to the 120 sampling prefectures of the CHARLS survey, where the annual prefecture-level expropriation rate is estimated on at least 20 observations. In column 5 of Panel C, the dependent variable is the prefecture's annual national rank in the growth of total urban area based on satellite data. The regressor of interest is the number of other prefecture-level CCP secretaries in the province who were older than 55 in the previous year. Controls include the CCP secretary's age in the year he takes office, as well as prefecture and calendar year fixed effects (not reported in table). Standard errors are clustered at the province-start year level with significance at the 1%, 5% and 10% denoted by ***, ** and *, respectively.

Table 3.A4: Prefecture outcomes and promotion incentives: Controlling for promotion rate before term

Panel A: Economic growth

	GDP growth (nominal)	GDP growth (real)	GDP growth rank (Satellite)
Size of starting cohort	-0.0109***	-0.00335**	0.000893
	(0.00318)	(0.00150)	(0.00334)
Promotion rate of year preceding term	-0.0267***	-0.00661	0.00151
	(0.00901)	(0.00412)	(0.00915)
Nb. clusters	232	232	232

Panel B: Real estate and expropriation

	Employment Growth Construction sec.	Real Estate investment % GDP	Residential RE investment % GDP	Expropriation rate	Urban expansion rank
Size of starting cohort	-0.160*	-0.00521**	-0.00253**	-0.0000548	-0.0965
	(0.0896)	(0.00218)	(0.00127)	(0.000829)	(0.0755)
Promotion rate of year preceding term	0.137	0.00394	0.00744**	0.00305	0.0658
	(0.256)	(0.00401)	(0.00329)	(0.00408)	(0.218)
Nb. clusters	232	232	231	193	232

Panel C: Public good provision

	Average Z-score	# Hospital beds per capita	# Doctors per capita	# Buses per capita	# Teachers per c. (primary)	# Teachers per c. (middle)
Size of starting cohort	0.00960	0.267	46.29	0.140*	0.179	-0.143
	(0.00833)	(0.192)	(77.53)	(0.0796)	(0.198)	(0.216)
Promotion rate of year preceding term	0.0342	1.618***	350.6*	0.384*	0.243	-0.225
	(0.0220)	(0.399)	(202.7)	(0.225)	(0.511)	(0.446)
Nb. clusters	231	231	232	231	231	232

See notes for Table 3.A3. All regressions control for the promotion rate of prefecture party secretaries from the same province the year before a party secretary's term begins.

Table 3.A5: Prefecture outcomes and promotion incentives:
Placebo test for lagged cohort size

Panel A: Economic growth

	GDP growth (nominal)	GDP growth (real)	GDP growth rank (Satellite)
Lagged starting cohort size	0.000916	0.000520	-0.00231
	(0.00235)	(0.00111)	(0.00229)
Nb. clusters	281	281	281

Panel B: Real estate and expropriation

	Employment Growth Construction sec.	Real Estate investment % GDP	Residential RE investment % GDP	Expropriation rate	Urban expansion rank
Lagged starting cohort size	-0.0134	0.00211*	0.000620	0.000350	-0.0626
	(0.0653)	(0.00121)	(0.000820)	(0.00141)	(0.0616)
Nb. clusters	281	281	280	226	281

Panel C: Public good provision

	Average Z-score	# Hospital beds per capita	# Doctors per capita	# Buses per capita	# Teachers per c. (primary)	# Teachers per c. (middle)
Lagged starting cohort size	-0.00427	-0.110	70.33	-0.0566	0.0342	-0.120
	(0.00559)	(0.104)	(52.35)	(0.0642)	(0.140)	(0.149)
Nb. clusters	278	279	277	278	280	281

See notes for Table 3.A3. All regressions control for the one year lagged value of the starting cohort size (the value it would have taken had the bureaucrat started one year earlier).

Table 3.A6: Prefecture outcomes and promotion incentives:
Controlling for CCP secretary characteristics

Panel A: Economic growth						
	GDP growth (nominal)	GDP growth (real)	GDP growth rank (Satellite)			
Size of starting cohort (Normalized)	-0.00952***	-0.00362***	-0.0203			
	(0.00225)	(0.00127)	(0.0762)			
Mean of dep. var.	0.16	0.13	5.16			
Panel B: Real estate and expropriation						
	Employment Growth Construction sec.	Real Estate investment % GDP	Residential RE investment % GDP	Expropriation rate	Urban expansion rank	
Size of starting cohort (Normalized)	-0.0463	-0.00253	-0.00126	-0.00189***	-0.103	
	(0.0873)	(0.00192)	(0.00114)	(0.000481)	(0.0699)	
Mean of dep. var.	1.05	0.066	0.052	0.0073	5.20	
Panel C: Public good provision						
	Average Z-score	# Hospital beds per capita	# Doctors per capita	# Buses per capita	# Teachers per c. (primary)	# Teachers per c. (middle)
Size of starting cohort (Normalized)	0.0169**	0.0234	0.0945	0.199**	0.135	0.216
	(0.00741)	(0.162)	(0.138)	(0.0894)	(0.180)	(0.162)
Mean of dep. var.	-0	11672	6572	6	18268	15630
Nb. of obs.	1717	1768	1764	1720	1797	1786
Nb. of clusters	252	257	255	252	257	258

See notes for Table 3.A3. All regressions control for the following CCP secretary characteristics: ethnic minority (a dummy variable for non-Han ethnicities), native of the province (a dummy variable that takes the value of 1 if the prefecture party secretary was born in the same province), education level (a set of dummy variables for his reported alma mater, see more details in 3.2).

Table 3.A7: Prefecture outcomes and promotion incentives: controlling for pre-term outcomes

Panel A: Economic growth

	GDP growth (nominal)	GDP growth (real)	GDP growth rank
Size of starting cohort	-0.00375***	-0.00105**	-0.0378*
	(0.000967)	(0.000471)	(0.0209)
Dep. var. before term	-0.00499	0.121***	-0.143***
	(0.0179)	(0.0333)	(0.0193)
Nb. of clusters	303	301	306

Panel B: Real estate and expropriation

	Employment Growth Construction sec.	Real Estate investment % GDP	Residential RE investment % GDP	Expropriation rate	Urban expansion rank
Size of starting cohort	-0.0620*	-0.000695*	0.000127	-0.000453	-0.0163
	(0.0328)	(0.000418)	(0.000300)	(0.000341)	(0.0234)
Dep. var. before term	-0.0563*	0.447***	0.349***	0.000208	-0.100***
	(0.0324)	(0.0817)	(0.0352)	(0.0958)	(0.0249)
Nb. of clusters	250	291	267	238	301

Panel C: Public good provision

	Average Z-score	# Hospital beds per capita	# Doctors per capita	# Buses per capita	# Teachers per c. (primary)	# Teachers per c. (middle)
Size of starting cohort	0.00642**	0.0391	23.58	0.0480*	0.0681	0.0282
	(0.00272)	(0.0490)	(21.29)	(0.0254)	(0.0615)	(0.0677)
Dep. var. before term	0.367***	0.404***	0.416***	0.399***	0.154***	0.183***
	(0.0490)	(0.0411)	(0.0684)	(0.0514)	(0.0378)	(0.0409)
Nb. of clusters	216	295	297	217	298	298

See notes for Table 3.A3. Each regression controls for the value of the dependent variable in the year before a prefecture party secretary's term begins.

Table 3.A8: Prefecture outcomes and promotion incentives:
Term-level regression

Panel A: Economic growth

	GDP growth (nominal)	GDP growth (real)	GDP growth rank (Satellite)
Size of starting cohort (Normalized)	-0.0145**	-0.00302	-0.111
	(0.00698)	(0.00373)	(0.0767)
Mean of dep. var.	0.16	0.13	5.16

Panel B: Real estate and expropriation

	Employment Growth Construction sec.	Real Estate investment % GDP	Residential RE investment % GDP	Expropriation rate	Urban expansion rank
Size of starting cohort (Normalized)	-0.108	-0.00464**	-0.00265*	-0.00128	-0.168**
	(0.0835)	(0.00212)	(0.00145)	(0.000897)	(0.0771)
Mean of dep. var.	0.91	0.063	0.051	0.0080	5.25

Panel C: Public good provision

	Average Z-score	# Hospital beds per capita	# Doctors per capita	# Buses per capita	# Teachers per c. (primary)	# Teachers per c. (middle)
Size of starting cohort (Normalized)	0.0225**	0.256	0.0722	0.127	0.311	0.216
	(0.0102)	(0.202)	(0.161)	(0.0918)	(0.213)	(0.236)
Mean of dep. var.	-0.04	11775.10	6663.30	6.50	18331.94	15472.92
Nb. of obs.	751	749	753	754	762	760
Nb. of clusters	304	303	302	304	306	306

Observations are at the term level. See notes for Table 3.A3 for details on samples, controls and standard errors.

Table 3.A9: Promotion incentives and Prefecture outcomes:
Weighting observations by the inverse of the term length

Panel A: Economic growth						
	GDP growth (nominal)	GDP growth (real)	GDP growth rank (Satellite)			
Size of starting cohort (Normalized)	-0.00820***	-0.00260**	-0.0740			
	(0.00232)	(0.00118)	(0.0636)			
Mean of dep. var.	0.16	0.13	5.16			
Panel B: Real estate and expropriation						
	Employment Growth Construction sec.	Real Estate investment % GDP	Residential RE investment % GDP	Expropriation rate	Urban expansion rank	
Size of starting cohort (Normalized)	-0.0887	-0.00273**	-0.00168*	-0.00109**	-0.103	
	(0.0606)	(0.00138)	(0.000906)	(0.000550)	(0.0642)	
Mean of dep. var.	1.05	0.066	0.052	0.0078	5.20	
Panel C: Public good provision						
	Average Z-score	# Hospital beds per capita	# Doctors per capita	# Buses per capita	# Teachers per c. (primary)	# Teachers per c. (middle)
Size of starting cohort (Normalized)	0.0187***	0.167	82.84	0.122*	0.205	0.103
	(0.00706)	(0.137)	(56.93)	(0.0619)	(0.158)	(0.178)
Mean of dep. var.	-0	11672	6572	6	18268	15630
Nb. of obs.	2487.0	2601.0	2553.0	2489.0	2644.0	2625.0
Nb. of clusters	303.000	303.000	302.000	303.000	305.000	306.000

See notes for Table 3.A3. All observations are weighted by the inverse of the number of years of a prefecture party secretary's term.

Table 3.A10: Expropriation probability and promotion incentives: Individual-level results

Variables	$\mathbb{1}_{it}\{\text{expro}\}$ (1)	$\mathbb{1}_{it}\{\text{expro}\}$ (2)	$\mathbb{1}_{it}\{\text{expro}\}$ (3)	$\mathbb{1}_{it}\{\text{expro}\}$ (4)	$\mathbb{1}_{it}\{\text{expro}\}$ (5)
Size of starting cohort (Normalized)	-0.00107***	-0.00110***	-0.000948*	-0.00107***	-0.00107***
	(0.000384)	(0.000395)	(0.000522)	(0.000390)	(0.000399)
Control variables					
Start of term year FE	Yes	Yes	No	Yes	Yes
Calendar year FE	No	No	Yes	No	No
Individual FE	No	No	No	No	Yes
Sample definition					
Prefectures	$N > 20$	$N > 20$	$N > 20$	$N > 20$	All
Individuals	All	Rur. h.	All	All	All
CPC Secretaries	All	All	All	$\text{Age}_{t_0} < 55$	All
Nb. Obs	140608	132332	140608	130035	140603
Nb. individuals	10203	9608	10203	10200	10198
Nb. clusters	256	256	256	239	256

Observations are at the individual-year level. The sample only includes observations during terms that start at the earliest in 1996 and finish before 2014 and for which we observe the prefecture party secretary's promotion and age, as well as all macroeconomic variables of interest. The regressor of interest is the number of other prefecture-level CCP secretaries in the province who start their terms in the same year as said prefectural CCP secretary. Controls include the CCP secretary's age in the year he takes office (reported in the table), as well as prefecture and calendar year fixed effects. Standard errors are clustered at the province-start year level with significance at the 1%, 5% and 10% denoted by ***, ** and *, respectively.

Table 3.A11: Individual-level results:
Robustness to CHARLS sample definition

Variables	$\mathbb{1}_{it}\{\text{expro}\}$ (1)	$\mathbb{1}_{it}\{\text{expro}\}$ (2)	$\mathbb{1}_{it}\{\text{expro}\}$ (3)
Size of starting cohort (Normalized)	-0.00102***	-0.00107***	-0.00111***
	(0.000380)	(0.000384)	(0.000382)
	[0.0241]	[0.0188]	[0.0129]
Sample definition			
Prefectures	All	$N > 20$	$N > 50$
Nb. Obs	142453	140608	138486
Nb. individuals	10333	10203	10050
Nb. clusters	323	256	249

Observations are at the individual-year level. The sample only includes observations during terms that start at the earliest in 1996 and finish before 2014 and for which we observe the prefecture party secretary's promotion and age, as well as all macroeconomic variables of interest. The regressor of interest is the number of other prefecture-level CCP secretaries in the province who start their terms in the same year as said prefectural CCP secretary. Controls include the CCP secretary's age in the year he takes office (reported in the table), as well as prefecture and calendar year fixed effects. Standard errors are clustered at the province-start year level with significance at the 1%, 5% and 10% denoted by ***, ** and *, respectively. P-values adjusted for family-wise multiple hypothesis testing are reported in square brackets.

Table 3.A12: Sectoral GDP growth and promotion incentives

Variables	GDP growth Primary sector (nominal)	GDP growth Secondary sector (nominal)	GDP growth rank Tertiary sector (nominal)
Source	NBS	NBS	NBS
	(1)	(2)	(3)
Size of starting cohort (Normalized)	-0.00552*	-0.0119***	-0.00485**
	(0.00303)	(0.00411)	(0.00239)
	[0.419]	[0.024]	[0.262]
Age at entry	0.000976	0.000441	0.000359
	(0.000745)	(0.000776)	(0.000501)
	[1.000]	[1.000]	[1.000]
Nb. obs	2678	2678	2678
Nb. clusters	306	306	306
Mean of dep. var.	0.0996	0.183	0.161

See notes for Table 3.A3.

Table 3.A13: Employment growth and promotion incentives

Variables	Employment growth (Total)	Employment growth (Primary sec.)	Employment growth (Secondary sec.)	Employment growth (Tertiary sec.)
Source	NBS	NBS	NBS	NBS
	(1)	(2)	(3)	(4)
Size of starting cohort (Normalized)	-0.0118**	-0.0697**	-0.0126***	-0.00874**
	(0.00470)	(0.0310)	(0.00453)	(0.00413)
Age at entry	-0.00262	-0.0162	-0.00196*	-0.00185
	(0.00190)	(0.0323)	(0.00112)	(0.00159)
Nb. obs	2678	2671	2678	2678
Nb. clusters	306	306	306	306
Mean of dep. var.	0.00227	0.0436	0.0122	0.00159

See notes for Table 3.A3.

Table 3.A14: Prefecture outcomes and promotion incentives: Robustness to outliers

Panel A: Economic growth						
	GDP growth (nominal)	GDP growth (real)	GDP growth rank	GDP growth Primary sector (nominal)	GDP growth Secondary sector (nominal)	GDP growth Tertiary sector (nominal)
Size of starting cohort (Normalized)	-0.00689***	-0.00283***	-0.0923*	-0.00286	-0.00865**	-0.00335
	(0.00230)	(0.000990)	(0.0548)	(0.00204)	(0.00364)	(0.00221)
Nb. clusters	306	305	306	305	306	306
Panel B: Real estate and expropriation						
	Employment Growth Construction sec.	Real Estate investment % GDP	Residential RE investment % GDP	Expropriation rate	Urban expansion rank	
Size of starting cohort (Normalized)	-0.107*	-0.000789	-0.00105	-0.000644*	-0.105**	
	(0.0550)	(0.000880)	(0.000784)	(0.000376)	(0.0526)	
Nb. clusters	304	299	299	217	341	
Panel C: Public good provision						
	Average Z-score	# Hospital beds per capita	# Doctors per capita	# Buses per capita	# Teachers per c. (primary)	# Teachers per c. (middle)
Size of starting cohort (Normalized)	0.0155**	0.183	0.172	0.145**	0.171	0.112
	(0.00748)	(0.145)	(0.112)	(0.0648)	(0.162)	(0.174)
Nb. clusters	303	303	301	303	305	306

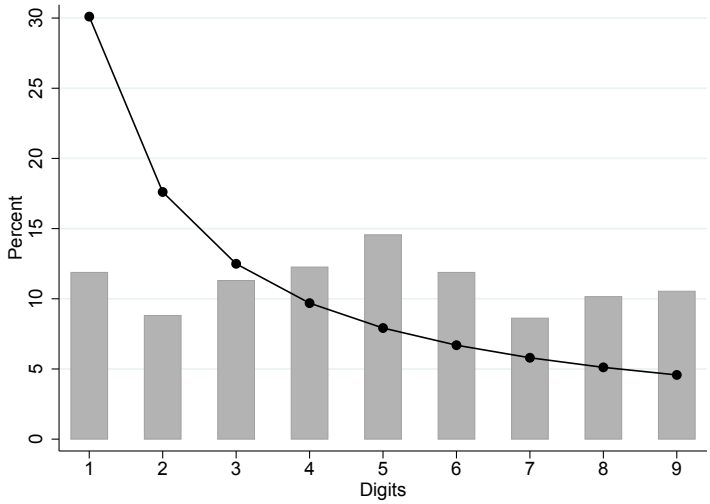
See notes for Table 3.A3. Data are trimmed to remove outliers at the top and bottom of the distribution.

Table 3.A15: GDP data manipulation - Reported real GDP growth rate: Distribution of decimal digits

Panel A: first decimal digit	$D1 = 1$	$D1 = 2$	$D1 = 3$	$D1 = 4$	$D1 = 5$	$D1 = 6$	$D1 = 7$	$D1 = 8$	$D1 = 9$
Size of starting cohort (Normalized)	0.00204	-0.00256	0.00231	0.00230	-0.00249	0.00534	0.00264	-0.00482	-0.00568
	(0.00860)	(0.00847)	(0.00689)	(0.00581)	(0.00738)	(0.00748)	(0.00701)	(0.00654)	(0.00612)
Mean of dep. var.	0.14	0.11	0.091	0.066	0.11	0.096	0.071	0.096	0.059
Nb. of clusters	306	306	306	306	306	306	306	306	306
Panel B: second decimal digit	$D2 = 1$	$D2 = 2$	$D2 = 3$	$D2 = 4$	$D2 = 5$	$D2 = 6$	$D2 = 7$	$D2 = 8$	$D2 = 9$
Size of starting cohort (Normalized)	0.00171	-0.00809***	-0.00188	-0.00110	-0.00739**	-0.00409	0.000937	0.00855***	-0.000825
	(0.00254)	(0.00277)	(0.00285)	(0.00224)	(0.00335)	(0.00331)	(0.00264)	(0.00280)	(0.00224)
Mean of dep. var.	0.013	0.012	0.012	0.012	0.016	0.015	0.0090	0.010	0.0097
Nb. of clusters	306	306	306	306	306	306	306	306	306
Nb. of observations	2677	2677	2677	2677	2677	2677	2677	2677	2677

See notes for Table 3.A3. In each column, the dependent variable is a dummy equal to 1 if the first (Panel A) or second (Panel B) decimal digit of the reported real growth rate of a prefecture in a given year is equal to 1, 2, ..., 9.

Figure 3.A1: Second decimal digit distribution



Note: This graph plots the actual distribution (grey bars) of the second decimal digits in reported real GDP growth rates; and the theoretical distribution (black connected dots) one should observe if said distribution followed Benford's law. The underlying data come from the NBS City Statistical Yearbooks, 1996-2013.

3.B Model

In this appendix we propose a simple model to show how promotion incentives can affect bureaucrats' policy choices. We combine a rent-seeking model (see, e.g. Tullock, 1980; Pérez-Castrillo and Verdier, 1992) with a multi-tasking framework following Holmstrom and Milgrom (1991).

Set up N bureaucrats compete for one position, which yields utility S. Each bureaucrat spends his total available effort e on either task 1 or two, which cost $c_1(e_{1,i})$ and $c_2(e_{2,i})$. The likelihood that bureaucrat i obtains the position is given by the following contest success function (CSF):

$$p_i = \frac{e_{1,i}}{\sum_{k\in[0,N]} e_{1,k}} \tag{3.6}$$

Only effort on the first task enters into the contest function, while effort on the second task yields a private benefit to the bureaucrat of $B(e_{2,i})$.

Maximisation problem Bureaucrats are risk-neutral and maximise their utility

$$\max_{e_{1,i},e_{2,i}} U_i = S * p_i(e_{1,i}, E_{1,-i}) + B(e_{2,i}) - c_1(e_{1,i}) - c_2(e_{2,i})$$

$$\text{subject to } e_{1,i} + e_{2,i} \leq e_i = e$$

$$\text{Or: } \max_{e_{1,i}} U_i = S * p_i(e_{1,i}, E_{1,-i}) + B(e - e_{1,i}) - c_1(e_{1,i}) - c_2(e - e_{1,i})$$

The first order condition is:

$$\begin{aligned} \frac{\partial p_i(e_{1,i}, E_{1,-i})}{\partial e_{1,i}} &= \frac{\partial c_1(e_{1,i})}{\partial e_{1,i}} + \frac{\partial B(e - e_{1,i})}{\partial e_{1,i}} - \frac{\partial c_2(e - e_{1,i})}{\partial e_{1,i}} \\ \frac{\partial p_i(e_{1,i}, E_{1,-i})}{\partial e_{1,i}} &= \frac{\partial \phi(e_{1,i})}{\partial e_{1,i}} \end{aligned} \tag{3.7}$$

where ϕ represents the costs of effort e_1 net of the opportunity cost of not providing $e_2 = e - e_1$. Equation (3.7) defines a "best response curve" to the vector of $N-1$ strategies by the other bureaucrats, $E_{1,-i}$.

Equilibrium A symmetric equilibrium ($e_{1,j} = e_{1,k} = e_1, \quad \forall j \neq k$) exists for this game, where:

$$\begin{aligned}
\frac{\partial p_i(e_1, E_1)}{\partial e_1} &= \frac{\partial \phi(e_1)}{\partial e_1} \\
\frac{(e_1 + E_1) - e_1}{(e_{1,i} + E_1)^2} &= \frac{\phi'(e_1)}{S} \\
\frac{(N-1)e_1}{N^2 e_1^2} &= \frac{\phi'(e_1)}{S} \\
e_1 \phi'(e_1) &= \frac{N-1}{N^2} S
\end{aligned} \tag{3.8}$$

And equilibrium promotion probabilities are equal to:

$$\begin{aligned}
p_i &= \frac{e_{1,i}}{\sum_{k \in [0,N]} e_{1,k}} \\
&= \frac{1}{N}
\end{aligned} \tag{3.9}$$

One can show (as in Pérez-Castrillo and Verdier, 1992) that non-symmetric equilibria do not exist if all agents have the same costs and valuations.

The right-hand side of equation 3.8 is decreasing in the number of competitors N, meaning that the equilibrium level of contest-related effort e_1 will be a decreasing or increasing function of N depending on the shape of $e_1 \phi'(e_1)$.

Recall that $\phi = c_1 - (c_2 - B)$. Thus ϕ' is increasing when the marginal cost of contest-relevant effort increases faster than the marginal cost of private effort, net of marginal private benefit. This is likely to be the case when the level of contest-relevant effort is high, such as when the expected benefit of the context is large relative to the private benefit. In this case, the equilibrium level of contest-relevant effort e_1 is a decreasing function

of N: when facing many competitors are discouraged from providing promotion-relevant effort. Depending on the net private benefit $B - c_2$, bureaucrats may substitute towards private effort.

Equation 3.9 implies that promotion probabilities are also a decreasing function of the number competitors.

Discussion and model predictions We think this model is appropriate for the setting of our paper and delivers a number of testable predictions. In China's meritocratic bureaucrat promotion system, the principal (here, the provincial party standing committee) evaluates a number of agents (here, prefecture party secretaries) for promotion to higher office. In the empirical analysis, the relevant number of competitors N is the size of the starting cohort. S is large, as a prefecture party secretary who is promoted gains substantially in terms of power and compensation.[30] As the principal can observe GDP growth and infrastructure projects, effort spent on increasing growth e_1 are rewarded. Effort on other pro-social policies e_2 are less easily observable to the principal and thus do not result in greater likelihood of promotion. e_2 may, nevertheless, yield private benefits to the bureaucrat. For instance, the bureaucrat may intrinsically care about the local population or local elites and these policies may benefit the bureaucrat himself after retirement.

In equilibrium, a smaller number of competitors increases the likelihood that any given prefecture party secretary is promoted. Prefecture party secretaries spend more effort on promotion-related activities, such as construction and infrastructure projects, and less effort on other policies, such as public goods provision.

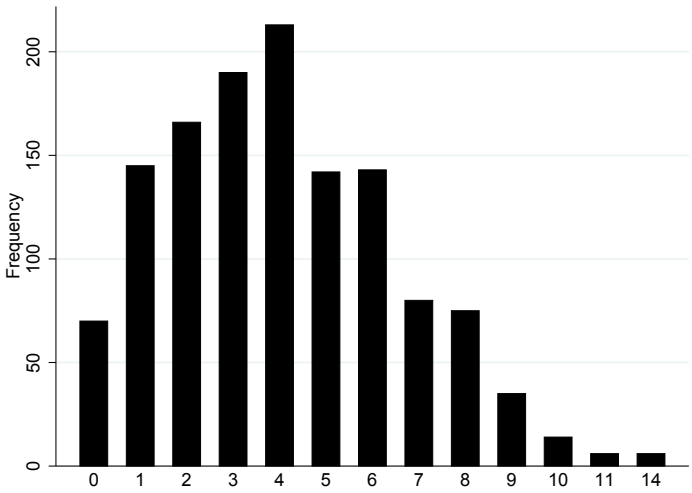
[30]The difference between prefecture and vice-provincial grades is a difference between a low- and a high-official with different compensation schemes.

3.C Detailed sample description

3.C.1 Bureaucrats

Table 3.C1 shows descriptive statistics of prefecture party secretaries in China from 1996 to 2014. The average prefecture party secretary enters office when he is 50 years old with three other prefecture party secretaries from the same province. Figure 3.C1 shows the distribution of the size of the starting cohort. Despite an official term limit of five years, the average term is less than three years long, see Figure 3.C2 for the distribution of term lengths.

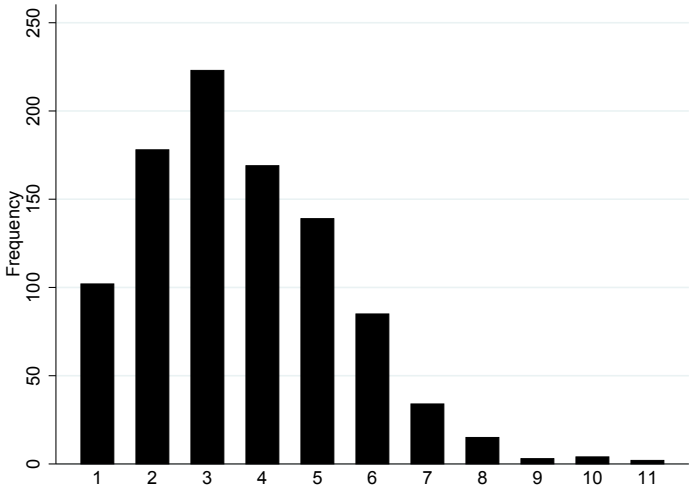
Figure 3.C1: Size of starting cohort



Note: This graph shows the distribution of our measure of the number of competitors in our sample. This measure is defined as the number of other CCP secretaries from the same province as a given CCP secretary, who start their term in the same year. The data are from our database on 1310 terms of CCP prefecture-level secretaries.

For each bureaucrat, we gather data on his career path in the years following his term. We define a time window relative to the end of the term, and collect information on the positions the bureaucrat holds within this time frame. We then attribute ranks to these positions based on official administrative rules. We assign to the bureaucrat the highest rank he reaches within two years of ending his term as a prefecture CCP secretary. Importantly, we distinguish between sub-provincial and regular prefecture-level units. Sub-provincial cities rank one level higher than regular units, so that a move

Figure 3.C2: Distribution of term lengths



Note: This graph shows the distribution of term lengths in our sample. The length of a term is equal to the number of consecutive years at the end of which a bureaucrat is observed as being in office in a given prefecture. The theoretical length of a term is therefore 5 in this graph. The data are from our database on 1310 terms of CCP prefecture-level secretaries.

from a regular prefecture-level unit to a subprovincial city is coded as a promotion.[31]

Many of the transfers are in fact to 'retirement'positions, which are official positions that wield little political power, but are still remunerated and can be officially ranked higher than prefecture-level party secretaries. We use qualitative data from media sources to identify such positions, and re-code transfers to such positions as retirement rather than promotions.[32] This is a similar approach to Li and Zhou (2005), but at the prefecture- rather than provincial level. Prior to the anti-corruption campaign, outright dismissals and demotions were rare and were sometimes masked as early retirement.

Table 3.C1 shows that the likelihood of a prefecture party secretary being promoted at the end of his term is around 49 percent. 26 percent of party secretaries moved into retirement after their term ended. We are able to identify the size of a party secretary's starting cohort in over 98 percent and the career path in 73 percent of all terms.

[31]For almost all secretaries where we collect career information, we are able to rank at least one position.

[32]These positions are at the national-level, provincial-level and prefectural-level, (i) positions in the Chinese People's Political Consultative Conference, often abbreviated as CPPCC, or in Chinese, 政治协商会议委员会, abbreviated 政协, and (ii) positions in to the Standing Commitee of the People's Congress, or in Chinese, 人民代表大会常务委员会, abbreviated 人大.

Table 3.C1: Term-level descriptive statistics

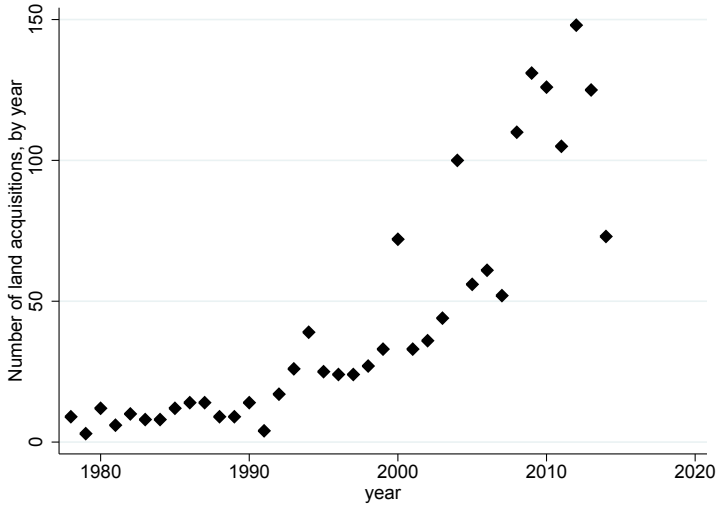
Variables	Mean	Std. deviation	Min	Max
Length of term	2.635	1.778	0	10
Age at entry	49.75	4.231	30	60
Size of starting cohort	4.072	2.524	0	14
Promotion (official rank)	0.526	0.500	0	1
Promotion (actual)	0.487	0.500	0	1
Retirement	0.263	0.440	0	1
Anti-corruption campaign arrest	0.126	0.332	0	1
Data availability				
Age data available	0.946			
Competitor's data available	0.981			
Promotion data available	0.732			
All data available	0.726			

The statistics in this table are computed on the 1310 terms of prefecture-level CCP secretaries that start no earlier than 1996 and end in 2014 or before. The length of one's term is equal to 0 if the secretary is observed in office once at the end of a year, 1 if he is observed in office at the end of two consecutive years, etc. Officially, terms are 5 years long, but this rule is little enforced, so that we concatenate consecutive 'official' terms if any. The age at entry of a secretary is the age of the secretary when he takes office. The variable 'Competitors data available' is a dummy variable equal to 1 if data on the seniority and age of all other secretaries in the province at the time a given secretary takes office is available. Promotion data is based on the positions that a given secretary occupies following his term. We consider all positions for which the starting date is between the last year at the end of which a secretary is in office and two years after. For most individuals this corresponds to a duration of less than one year, and this is by definition strictly less than 2 years. Promotion is officially defined by ranks in the hierarchical structure of the CCP. Some of these promotions are to 'retirement positions'. Some secretaries are nominated to several positions, which explains why the retirement rate and actual promotion rate do not add up to the total promotion rate. The size of the starting cohort corresponds to the number of competitors of a given secretary, ie: the number of secretaries in the same province who start their term in the same year as a given secretary.

3.C.2 Policy outcomes

Figure 3.C3 shows the number of land expropriations recorded each year from the CHARLS survey. Expropriations were increasing rapidly through our sample period between 1996 and 2014.

Figure 3.C3: Land expropriations, CHARLS survey



Note: This graph plots the absolute number of expropriation events that are recorded for each calendar year. In our analysis, we restrict the data to the 1996-2014 period, which is the period where most expropriations occur, and for which we collect CCP secretaries career data. The data come from the 2014 Life History wave of the CHARLS survey.

Table 3.C2 shows summary statistics for the prefecture-year panel used in the main analysis.

Table 3.C2: Summary statistics - macroeconomics variables

Variables	Mean	Std. Deviation	Min	Max
Population (in millions)	4.158	2.334	0.160	12.50
Provincial capita	0.0903	0.287	0	1
Total prefecture area (thousands of km^2)	22.96	45.87	0.978	496.3
Nominal GDP (billion RMB)	75.85	110.5	0.702	1542.0
Nominal GDP growth	0.159	0.0964	-0.736	1.462
Real GDP growth	0.135	0.233	-0.488	7.908
Nominal GDP growth (Primary sector)	0.145	2.187	-0.638	125.0
Nominal GDP growth (Secondary sector)	0.188	0.206	-0.761	6.068
Nominal GDP growth (Tertiary sector)	0.172	0.270	-0.885	11.20
Employment (thousands)	440.4	508.4	40.50	5852.9
Employment growth	0.00200	0.532	-0.957	26.06
Employment share (Primary sector)	0.0524	0.102	0.01	0.740
Employment share (Secondary sector)	0.425	0.137	0.0446	0.844
Employment share (Tertiary sector)	0.523	0.134	0.0991	0.948
Employment growth (Construction sector)	1.022	4.368	-1	91.21
Real estate investment (% GDP)	6.352	5.967	0	91.22
Residential RE investment (% GDP)	5.189	4.353	0.00887	61.50
Expropriation rate	0.00711	0.0430	0	1
National rank in urban area growth	177.2	101.9	1	334
Number of hospital beds	11444.3	7952.1	860	66721
Number of doctors	6473.3	4802.0	377	63193
Number of buses per capita	6.402	6.519	0	115
Number of teachers (primary school)	17756.6	10187.0	595	57887
Number of teachers (middle school)	15062.0	8634.5	400	46287
Number of teachers (higher education)	3095.6	6135.0	0	55416
Observations	2899			

Data included in this table are from the National Bureau of Statistics City Statistical Yearbooks, except for the expropriation rate, which we compute from the CHARLS survey, and the total prefecture area, which is measured using satellite data from the DMSP/OLS. The statistics are computed on our main estimation sample, the 2899 prefecture-year observations where macroeconomic data and CCP secretary characteristics are available.

3.D Measurement of GDP growth using night light data

In this appendix we present a simple comparison of raw night light data from the National Oceanic and Atmospheric Administration (NOAA)'s Defense Meteorological Program Operational Line-Scan System (DMSP-OLS) and nominal GDP figures from the NBS. This reproduces the comparison in Zhang, Pandey and Seto (2016) for China. As shown in Panel A of Figure 3.D1, the DMSP-OLS data are available in different series which partially overlap. The absolute levels of night light intensity (on the left axis) for a given year can be substantially different depending on the series. Zhang, Pandey and Seto (2016) use machine learning algorithms to smooth the global night light data and obtain the processed data in Panel B. This procedure is aimed less at improving the overall fit of the night light data to GDP than in smoothing out differences across series of satellite data.

Figure 3.D1: Night lights and GDP levels

Note: These figures show the aggregate GDP of China in nominal terms (dashed red line, left axis) and the total night time light intensity from satellite data over China's territory (grey solid lines, right axis). GDP data are from the NBS and night time light intensity data are from the DMSP-OLS (Panel A) or Zhang, Pandey and Seto (2016) (Panel B). The different solid grey lines correspond to different satellite systems.

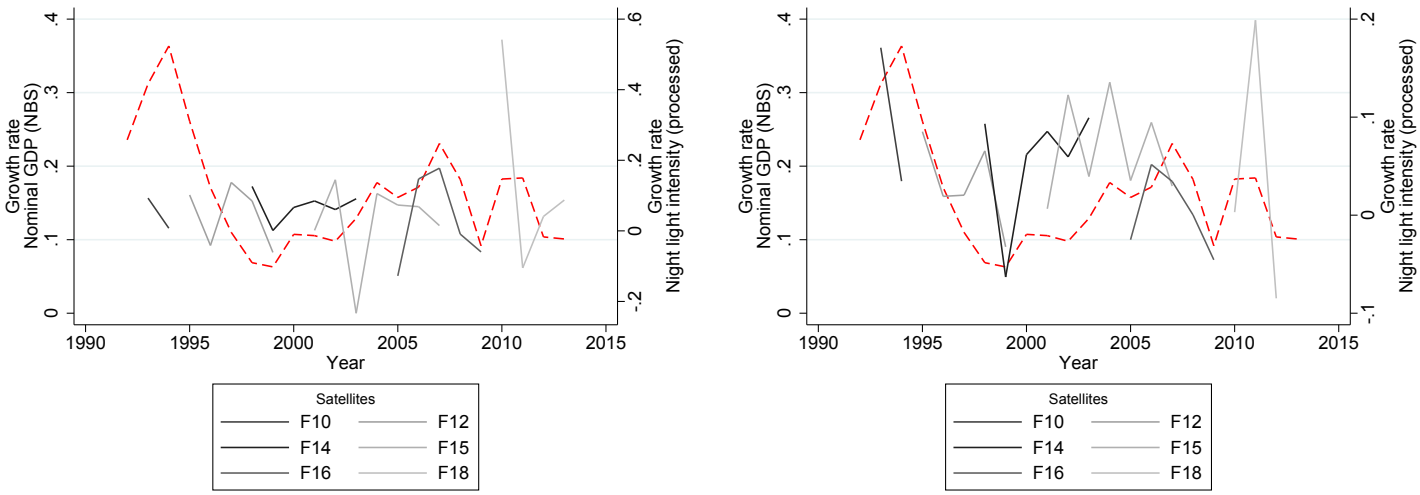
Aside from discrepancies in the levels of night light intensities recorded by different satellite systems, the most problematic aspect of nightlight data in our context is that

two satellite systems of different sensitivity yield year-on-year variations that sometimes go in opposite directions. In Panel A of Figure 3.D2 this is the case for the F14 and F15 systems for the years 2002 and 2003 or the the F15 and F16 systems in 2005.

For this reason, we prefer to use the percentile rank of prefectures rather than the absolute levels of growth rates implied by the satellite data and always report estimates using prefecture and year fixed effects.

Figure 3.D2: Growth of Night light intensity and Growth of GDP

A Raw data B Processed data



Note: These figures show the nominal growth of aggregate Chinese GDP (dashed red line, left axis), and the annual growth of total night time light intensity from satellite data over China's territory (grey solid lines, right axis). GDP data are from the NBS, and night time light intensity data are from the DMSP-OLS (Panel A) or Zhang, Pandey and Seto (2016) (Panel B). The different solid grey lines correspond to different satellite systems.

Bibliography

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.

Allcott, Hunt, and Matthew Gentzkow. 2017. "Social Media and Fake News in the 2016 Election." *Journal of Economic Perspectives*, 31(2): 211–36.

Allcott, Hunt, Matthew Gentzkow, and Chuan Yu. 2019. "Trends in the Diffusion of Misinformation on Social Media." *Research & Politics*, 6(2): 2053168019848554.

Bertrand, Marianne, Robin Burgess, Arunish Chawla, and Guo Xu. 2020. "The Glittering Prizes: Career Incentives and Bureaucrat Performance." *Review of Economic Studies*, 87(2): 626–655.

Bursztyn, Leonardo, Georgy Egorov, Ruben Enikolopov, and Maria Petrova. 2019. "Social Media and Xenophobia: Evidence from Russia." *National Bureau of Economic Research Working Paper*, 26567.

Chang, Charles, and Melanie Manion. 2017. "Political Self-Censorship in an Authoritarian State: A New Theory and Evidence."

Chen, Ting, and James Kai-sing Kung. 2019. "Busting the

"Princelings": The Campaign Against Corruption in China's Primary Land Market." *Quarterly Journal of Economics*, 134(1): 185–226.

Chen, Ting, and J. K. S. Kung. 2016. "Do land revenue windfalls create a political resource curse? Evidence from China." *Journal of Development Economics*, 123: 86–106.

Chen, Ye, Hongbin Li, and Li-An Zhou. 2005. "Relative Performance Evaluation and the Turnover of Provincial Leaders in China." *Economics Letters*, 88(3): 421–425.

Che, Yi, Julan Du, Yi Lu, and Zhigang Tao. 2015. "Once an Enemy, Forever an Enemy? The Long-Run Impact of the Japanese Invasion of China from 1937 to 1945 on Trade and Investment." *Journal of International Economics*, 96(1): 182–198.

Chong, Alberto, and Eliana La Ferrara. 2009. "Television and Divorce: Evidence from Brazilian Novelas." *Journal of the European Economic Association*, 7(2-3): 458–468.

CNNIC. 2013. *The 31 Statistical Report on Internet Development in China.* China Internet Network Information Centre.

de Janvry, Alain, Kyle Emerick, Marco Gonzalez-Navarro, and Elisabeth Sadoulet. 2015. "Delinking Land Rights from Land Use: Certification and Migration in Mexico." *American Economic Review*, 105(10): 3125–3149.

De La Rupelle, Maëlys, Quheng Deng, Li Shi, and Thomas Vendryes. 2009. "Land Rights Insecurity and Temporary Migration in Rural China."

DellaVigna, Stefano, and Eliana La Ferrara. 2015. "Economic and Social Impacts of the Media." In *Handbook of Media Economics.* Vol. 1,

723–768. Elsevier.

DellaVigna, Stefano, and Ethan Kaplan. 2007. "The Fox News Effect: Media Bias and Voting." *Quarterly Journal of Economics*, 122(3): 1187–1234.

DellaVigna, Stefano, and Matthew Gentzkow. 2010. "Persuasion: Empirical Evidence." *Annual Review of Economics*, 2(1): 643–669.

DellaVigna, Stefano, Ruben Enikolopov, Vera Mironova, Maria Petrova, and Ekaterina Zhuravskaya. 2014. "Cross-Border Media and Nationalism: Evidence from Serbian Radio in Croatia." *American Economic Journal: Applied Economics*, 6(3): 103–132.

Donald, Stephanie Hemelryk, Yin Hong, and Michael Keane. 2014. *Media in China: Consumption, Content and Crisis.* Routledge.

Durante, Ruben, Paolo Pinotti, and Andrea Tesei. 2019. "The Political Legacy of Entertainment TV." *American Economic Review*, 109(7): 2497–2530.

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

Field, Erica. 2007. "Entitled to Work: Urban Property Rights and Labor Supply in Peru." *Quarterly Journal of Economics*, 122(4): 1561–1602.

Finan, F., B. A. Olken, and R. Pande. 2017. "The Personnel Economics of the Developing State." In *Handbook of Economic Field Experiments*. Vol. 2 of *Handbook of Economic Field Experiments*, , ed. Abhijit Vinayak Banerjee and Esther Duflo, 467–514. North-Holland.

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.

Fisman, Raymond, Yasushi Hamao, and Yongxiang Wang. 2014. "Nationalism and Economic Exchange: Evidence from Shocks to Sino-Japanese Relations." *Review of Financial Studies*, 27(9): 2626–2660.

Foley, Kevin, Jeremy L. Wallace, and Jessica Chen Weiss. 2018. "The Political and Economic Consequences of Nationalist Protest in China: The 2012 Anti-Japanese Demonstrations." *China Quarterly*, 236: 1131–1153.

Fouka, Vasiliki, and Hans-Joachim Voth. 2016. "Reprisals Remembered: German-Greek Conflict and Car Sales During the Euro Crisis."

Fu, King-wa, and Michael Chau. 2013. "Reality Check for the Chinese Microblog Space: A Random Sampling Approach." *PLOS ONE*, 8(3): e58356.

Fu, King-wa, Chung-hong Chan, and Michael Chau. 2013. "Assessing Censorship on Microblogs in China: Discriminatory Keyword Analysis and the Real-Name Registration Policy." *IEEE Internet Computing*, 17(3): 42–50.

Fukuyama, Francis. 1989. "The End of History?" *National Interest*, 16: 3–18.

Goldstein, Markus, and Christopher Udry. 2008. "The Profits of Power: Land Rights and Agricultural Investment in Ghana." *Journal of Political Economy*, 116(6): 981–1022.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2009. "Cultural Biases in Economic Exchange?" *Quarterly Journal of Economics*, 124(3): 1095–1131.

He, Canfei, Zhiji Huang, and Xinyue Ye. 2014. "Spatial Heterogeneity

of Economic Development and Industrial Pollution in Urban China." *Stochastic Environmental Research and Risk Assessment*, 28(4): 767–781.

Heilmann, Kilian. 2016. "Does Political Conflict Hurt Trade? Evidence from Consumer Boycotts." *Journal of International Economics*, 99: 179–191.

Henderson, J. Vernon, Adam Storeygard, and David N. Weil. 2012. "Measuring Economic Growth from Outer Space." *American Economic Review*, 102(2): 994–1028.

Henderson, Vernon, Adam Storeygard, and David N Weil. 2011. "A Bright Idea for Measuring Economic Growth." *American Economic Review*, 101(3): 194–199.

Holmstrom, Bengt, and Paul Milgrom. 1991. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics, & Organization*, 7: 24–52.

Holz, Carsten A. 2014. "The Quality of China's GDP Statistics." *China Economic Review*, 30: 309–338.

Jia, Ruixue. 2017. "Pollution for Promotion." *21st Century China Center Research Paper.*

Jia, Ruixue, Masayuki Kudamatsu, and David Seim. 2015. "Political Selection in China: The Complementary Roles of Connections and Performance." *Journal of the European Economic Association*, 13(4): 631–668.

Ji, Dongxing. 2019. "Unusual Patterns in China's Prefectural GDP Growth Rates." *Applied Economics Letters*, 26(4): 331–334.

Jin, Xiaobin, Ying Long, Wei Sun, Yuying Lu, Xuhong Yang, and Jingxian Tang. 2017. "Evaluating Cities' Vitality and Identifying

Ghost Cities in China with Emerging Geographical Data." *Cities*, 63: 98–109.

Kearney, Melissa S., and Phillip B. Levine. 2015. "Media Influences on Social Outcomes: The Impact of MTV's 16 and Pregnant on Teen Childbearing." *American Economic Review*, 105(12): 3597–3632.

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. 2017. "How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument." *American Political Science Review*, 111(3): 484–501.

Kou, Chien-wen, and Wen-Hsuan Tsai. 2014. ""Sprinting with Small Steps" Towards Promotion: Solutions for the Age Dilemma in the CCP Cadre Appointment System." *China Journal*, 71: 153–171.

La Ferrara, Eliana, Alberto Chong, and Suzanne Duryea. 2012. "Soap Operas and Fertility: Evidence from Brazil." *American Economic Journal: Applied Economics*, 4(4): 1–31.

Landry, Pierre F., Xiaobo Lü, and Haiyan Duan. 2018. "Does Performance Matter? Evaluating Political Selection Along the Chinese Administrative Ladder." *Comparative Political Studies*, 51(8): 1074–1105.

Li, Guo, Scott Rozelle, and Jikun Huang. 2000. "Land Rights, Farmer Investment Incentives, and Agricultural Production in China." Department of Agricultural and Resource Economics, University of California Davis Working Paper No. 00-024.

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(9-10): 1743–1762.

Martin, Gregory J., and Ali Yurukoglu. 2017. "Bias in Cable News: Persuasion and Polarization." *American Economic Review*, 107(9): 2565–2599.

Maskin, Eric, Yingyi Qian, and Chenggang Xu. 2000. "Incentives, Information, and Organizational Form." *Review of Economic Studies*, 67(2): 359–378.

Müller, Karsten, and Carlo Schwarz. 2019. "Fanning the Flames of Hate: Social Media and Hate Crime." *Available at SSRN 3082972*.

Nitzan, Shmuel. 1994. "Modelling Rent-Seeking Contests." *European Journal of Political Economy*, 20.

Ochsner, Christian, and Felix Roesel. 2017. "Activated History-The Case of the Turkish Sieges of Vienna."

Pandya, Sonal S., and Rajkumar Venkatesan. 2016. "French Roast: Consumer Response to International Conflict—Evidence from Supermarket Scanner Data." *Review of Economics and Statistics*, 98(1): 42–56.

Passarelli, Francesco, and Guido Tabellini. 2017. "Emotions and Political Unrest." *Journal of Political Economy*, 125(3): 903–946.

Pérez-Castrillo, J David, and Thierry Verdier. 1992. "A General Analysis of Rent-Seeking Games." *Public Choice*, 73(3): 17.

Persson, Petra, and Ekaterina Zhuravskaya. 2016. "The Limits of Career Concerns in Federalism: Evidence from China." *Journal of the European Economic Association*, 14(2): 338–374.

Qin, Bei, David Strömberg, and Yanhui Wu. 2017. "Why Does China Allow Freer Social Media? Protests versus Surveillance and Pro-

paganda." *Journal of Economic Perspectives*, 31(1): 117–140.

Smith, Steven W. 1997. "Explaining Benford's Law." In *The Scientist and Engineer's Guide to Digital Signal Processing.*

Stockmann, Daniela. 2010. "Who Believes Propaganda? Media Effects during the Anti-Japanese Protests in Beijing." *China Quarterly*, 202: 269–289.

Tullock, Gordon. 1980. "Efficient Rent-Seeking." In *Toward a Theory of the Rent-Seeking Society.* , ed. Buchanan, J. M., Tollison, R. D. and Tullock, G. Texas A & M Univ Press.

Voigtländer, Nico, and Hans-Joachim Voth. 2012. "Persecution Perpetuated: The Medieval Origins of Anti-Semitic Violence in Nazi Germany." *Quarterly Journal of Economics*, 127(3): 1339–1392.

Vosoughi, Soroush, Deb Roy, and Sinan Aral. 2018. "The Spread of True and False News Online." *Science*, 359(6380): 1146–1151.

Wallace, Jeremy L., and Jessica Chen Weiss. 2015. "The Political Geography of Nationalist Protest in China: Cities and the 2012 Anti-Japanese Protests." *China Quarterly*, 222: 403–429.

Wang, Lanzhu. 2012. *China TV Rating Yearbook.* Press of Chinese Communication University.

Weiss, Jessica Chen. 2019. "How Hawkish Is the Chinese Public? Another Look at "Rising Nationalism" and Chinese Foreign Policy." *Journal of Contemporary China*, 28(119): 679–695.

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

Yao, Yao, Dongsheng Chen, Le Chen, Huan Wang, and Qingfeng

Guan. 2018. "A Time Series of Urban Extent in China Using DSMP/OLS Nighttime Light Data." *PLoS ONE*, 13(5).

Yew, Chiew Ping. 2012. "Pseudo-Urbanization? Competitive Government Behavior and Urban Sprawl in China." *Journal of Contemporary China*, 21(74): 281–298.

Zhang, Qingling, Bhartendu Pandey, and Karen C. Seto. 2016. "A Robust Method to Generate a Consistent Time Series From DMSP/OLS Nighttime Light Data." *IEEE Transactions on Geoscience and Remote Sensing*, 54(10): 5821–5831.

Zhao, Yuezhi, and Zhenzhi Guo. 2005. "Television in China: History, Political Economy, and Ideology." *A Companion to Television*, 521–539.

Zhou, Min, and Hanning Wang. 2016. "Participation in Anti-Japanese Demonstrations in China: Evidence from a Survey on Three Elite Universities in Beijing." *Journal of East Asian Studies*, 16(3): 391–413.

Zhuravskaya, Ekaterina, Maria Petrova, and Ruben Enikolopov. 2020. "Political Effects of the Internet and Social Media." *Annual Review of Economics*, 12(1): 415–438.

List of Figures

1.1 Newspaper Articles versus Internet Searches 48
1.2 Evolution of "Corruption" Articles 49
1.3 Municipality-Level Underreporting 50
1.4 Provincial Newspapers versus Sub-Provincial Newspapers 51
1.5 Party Newspapers versus Commercial Newspapers . 52
1.6 Proximity to the Head of the Province 53
1.7 Case Study: Wan Qingliang's Network 54
1.8 Time Profile of Estimated Coefficients 55
1.A1 Map of Provinces in Sample 62
1.A2 Geographic Distribution of Officials and Newspapers 63
1.A3 Distribution of Officials' Ranks 64
1.A4 Fraction of Newspaper with Articles about each Official 65
1.B1 Evolution of Sentiment Expressed in Articles 68
1.B2 Word cloud of topic model terms 69
1.C1 Reported versus Collected Comments on Sina Weibo 71
1.D1 Page Number of Articles 72
1.D2 Search Intensity for Officials under Investigation on Baidu 73
1.D3 Subsidiaries and their Parent Newspapers 76
1.D4 Advertising revenue and reporting about corruption scandals 77
1.D5 Heterogeneity in underreporting of high-level corruption scandals across province 78
1.D6 Share of Articles with "Corruption" in the Headline 79

2.1 TV show distribution licenses and anti-Japanese sentiment in China 124
2.2 Historical TV drama scheduling and island conflict . 125
2.3 Channel positions in Shanghai 126
2.4 Event Study of Social Media Posts 127
2.5 Anti-Japanese Protest Incidence and Leads and Lags of Historical TV Drama Viewership 128

2.6 Social Media Posts and Leads and Lags of Historical TV Drama Viewership 129
2.A1 Examples of anti-Japanese hate speech 139
2.A2 Keywords in Anti-Japanese Weibo posts 141
2.A3 Social media flyers calling for boycott 142
2.B1 Salience of island conflict 143
2.B2 Map of anti-Japanese protests in China 144
2.B3 Social Media Hate Speech over Time 145
2.C1 Event Study: Predicted Historical TV Drama Viewership 146

3.1 Promotion probability and size of starting cohort . . 186
3.2 Size of starting cohort and CCP secretary's age . . . 187
3.3 Correlation between theoretical and actual starting cohort size 188
3.A1 Second decimal digit distribution 212
3.C1 Size of starting cohort 216
3.C2 Distribution of term lengths 217
3.C3 Land expropriations, CHARLS survey 219
3.D1 Night lights and GDP levels 221
3.D2 Growth of Night light intensity and Growth of GDP 222

List of Tables

1.1 Extent of Underreporting of High-level Corruption Scandals by Newspapers from the Same Province Relative to Newspapers from Other Provinces 56
1.2 Underreporting of Local Corruption Scandals 57
1.3 Framing of Articles Post-Investigation Announcement 58
1.4 Internet Searches for Complaints against Officials . . 59
1.5 Comments on Sina Weibo Posts about High-Level Corrupt Officials 60
1.6 Coefficients for All Newspaper and High-level Official Provinces 61
1.A1 Summary Statistics 66
1.A2 Officials with No Articles or Only Articles Pre- or Post-Investigation 67
1.C1 Summary Statistics of Sina Weibo Sample 70
1.D1 Baidu Search Volume for Corrupt Officials by Province 74
1.D2 Comments on Sina Weibo Posts about High-Level Corrupt Officials: Restricted Sample of WiseNews Provinces 75
1.E1 Poisson and OLS Comparison 80
1.E2 Underreporting of Local Corruption Scandals: Excluding Cases of Potential Reverse Causality 81
1.E3 Underreporting of Local Corruption Scandals: ChinaFile Definition of High- versus Low-Ranking Officials 82
1.F1 Municipality-Level Underreporting 83
1.F2 Party Newspapers versus Commercial Newspapers . 84
1.F3 Provincial Newspapers versus Sub-Provincial Newspapers 85
1.F4 Proximity to the Head of the Province 86
1.F5 Case Study: Wan Qingliang's Network 87
1.F6 Coefficient estimates of dynamic model 88
1.F6 Coefficient estimates of dynamic model 89

2.1 Summary statistics 130

2.2 Correlates of Provincial TV Channels' Broadcasts of Sino-Japanese War Dramas 131
2.3 Predicting TV audience numbers 132
2.4 Historical TV Dramas and Anti-Japanese Protests . 133
2.5 Historical TV Dramas and Anti-Japanese Protests: Privately vs. State-Produced Shows 134
2.6 Historical TV Dramas and Anti-Japanese Protests: Heterogeneity across Provinces 135
2.7 Historical TV Dramas and Anti-Japanese Hate Speech on Social Media 136
2.8 Historical TV Dramas and Anti-Japanese Hate Speech on Social Media: Privately vs. State-Produced Shows 137
2.9 Historical TV Dramas and Anti-Japanese Hate Speech on Social Media: Original Posts versus Reposts . . . 138
2.A1 Keywords for social media content classification . . 140
2.C1 Additional Correlates of Provincial TV Channels' Broadcasts of Sino-Japanese War Dramas I 147
2.C2 Additional Correlates of Provincial TV Channels' Broadcasts of Sino-Japanese War Dramas II 148
2.C3 Historical TV Dramas and Social Media: Additional Variables 149
2.C4 Historical TV Dramas and Anti-Japanese Hate Speech: Heterogeneity across Provinces I 150
2.C5 Historical TV Dramas and Anti-Japanese Hate Speech: Heterogeneity across Provinces II 151
2.C6 Historical TV Dramas and Anti-Japanese Protests and Social Media Hate Speech: Robustness to excluding TV Channels from the Same Province . . . 152

3.1 Promotion probability and number of competitors (prefecture panel) 189
3.2 Correlation of starting cohort size with party secretary and prefecture characteristics 190
3.3 GDP growth and promotion incentives 191
3.4 Expropriations, real estate and promotion incentives 192
3.5 Land expropriation and individual outcomes 193
3.6 Expropriation and ex-post arrest probability 194
3.7 Ex-post ghost city status 195
3.8 Public good provision and promotion incentives . . 196
3.A1 Theoretical cohort size and actual cohort size 197
3.A2 Starting cohort size and composition of predecessors 198
3.A3 Prefecture outcomes and promotion incentives: Age of predecessors 199

3.A4 Prefecture outcomes and promotion incentives: Controlling for promotion rate before term 200
3.A5 Prefecture outcomes and promotion incentives: Placebo test for lagged cohort size 201
3.A6 Prefecture outcomes and promotion incentives: Controlling for CCP secretary characteristics 202
3.A7 Prefecture outcomes and promotion incentives: controlling for pre-term outcomes 203
3.A8 Prefecture outcomes and promotion incentives: Term-level regression 204
3.A9 Promotion incentives and Prefecture outcomes: Weighting observations by the inverse of the term length . 205
3.A10 Expropriation probability and promotion incentives: Individual-level results 206
3.A11 Individual-level results: Robustness to CHARLS sample definition 207
3.A12 Sectoral GDP growth and promotion incentives . . . 208
3.A13 Employment growth and promotion incentives . . . 209
3.A14 Prefecture outcomes and promotion incentives: Robustness to outliers 210
3.A15 GDP data manipulation - Reported real GDP growth rate: Distribution of decimal digits 211
3.C1 Term-level descriptive statistics 218
3.C2 Summary statistics - macroeconomics variables . . . 220