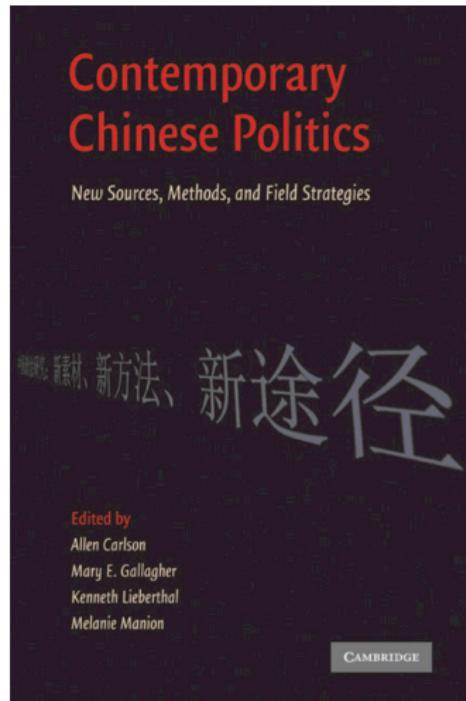


# Using Numbers to Study Contemporary China

Dr Chao-Yo Cheng

(Birkbeck, University of London)

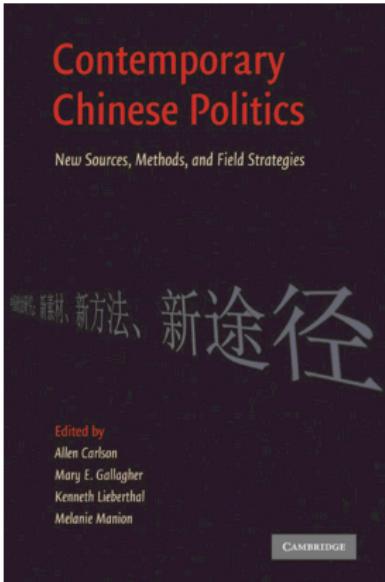
# Using numbers to study China



- ▶ Why "numbers" matter?
- ▶ Where "data" comes from?
- ▶ How should we work with data?

# Why numbers matter?

- ▶ The field of China studies has changed significantly as a result of **data availability** in digital forms and **methodological advances**
  - Quant techniques are essential to **establish patterns** for theory building and policy impact evaluation
  - Using numbers also encourage us to rethink the purpose of China studies – area studies or comparative politics?
  - The use of computational and quantitative data/methods by no means undermine the importance of traditional approaches – **multi-method/data triangulation** is now the norm
- ▶ Changing **geopolitical environment** also prompts the computational or even quantitative turn in the study of contemporary China



**Manion, Melanie (Duke), Kenneth Lieberthal (Michigan), Mary E. Gallagher (Michigan), and Allen Carlson (Cornell).** 2010. *Contemporary Chinese Politics: New Sources, Methods, and Field Strategies*. New York, NY: Cambridge University Press.

# Where “data” comes from

- ▶ Surveys, such as
  - Chinese General Social Survey (CGSS) – Prof Xiaogang Wu (NYU Shanghai)
  - China Family Panel Studies (CFPS) – Prof Yu XIE (Princeton)
  - Asianbarometer or World Values Survey (with caveats)
- ▶ Admin data, including socioeconomic surveys (e.g., statistical yearbooks 统计年鉴) and population census
- ▶ Digital and "big" data, such as
  - Social media posts (e.g., Weibo)
  - Search engine (e.g., Baidu Index) – comparable to Google Trends



北京大学开放研究数据平台

重用数据，再现科研。  
提升引用，跟踪影响。

开放数据  
为您的研究点赞！

[创建我的数据空间](#) 或 [浏览平台数据](#)

<https://opendata.pku.edu.cn/>



## Spatial Data

- China Geo-Explorer II
- China Geo-Explorer I
- China Map Library

## China Statistics

- National Statistics
- Provincial Statistics
- City Statistics
- County Statistics(1997-)
- Industrial Data
- Statistics Datasheets
- Statistics Charts

## Census Data

- Census Maps
- All Census Data
- Economic Census 2004
- Industrial Census 1995
- Census 1982
- Census 1982(10%)
- Census 1990
- Census 1995(1%)
- Province 2000
- County 2000
- Census 2005(1%)
- Census Data Search

## Free China Maps

- 2000 Population Census
- Pop&Env(1990-1999)
- Pop&Env(2000)
- Atlas of Industrial Census

## Workshops and Seminars



### The Summer Leadership Workshop on Spatiotemporal Innovation

1726 Cambridge Street, Cambridge MA 02138  
July 10-14, 2023

2023-02-14

The Summer Leadership Workshop on Spatiotemporal Innovation

More...

## Latest China Statistical News

- Consumer Prices in February 2023(2023/03/20)
- The Gross Imports and Exports Amounts 6.18 Trillion Yuan in the First Two Months of 2023(2023/03/20)
- Total Retail Sales of Consumer Goods from January to February 2023(2023/03/03/20)

<https://www.china-data-online.com/>





<https://index.baidu.com/>

200  
1823-2023

# Qualitative materials for empirical triangulation

- ▶ Gazetteers (通志)
- ▶ Yearbooks (年鉴), published yearly by various Party departments and government offices
- ▶ Culture and history materials (文史资料), published by the Chinese people's consultative conferences (政协)
- ▶ Government work reports (政府工作报告), published by the people's governments
- ▶ ... and so on so forth (e.g., archived historical materials compiled by the government or researchers)

ARTICLE

## In Search of Self-Censorship

Xiaoxiao Shen\* and Rory Truex\* 

Princeton University, Princeton, NJ, USA

\*Corresponding author. E-mail: [xiaoxiao@princeton.edu](mailto:xiaoxiao@princeton.edu) and [rtruex@princeton.edu](mailto:rtruex@princeton.edu)

(Received 27 March 2019; revised 22 August 2019; accepted 11 November 2019; first published online 16 March 2020)

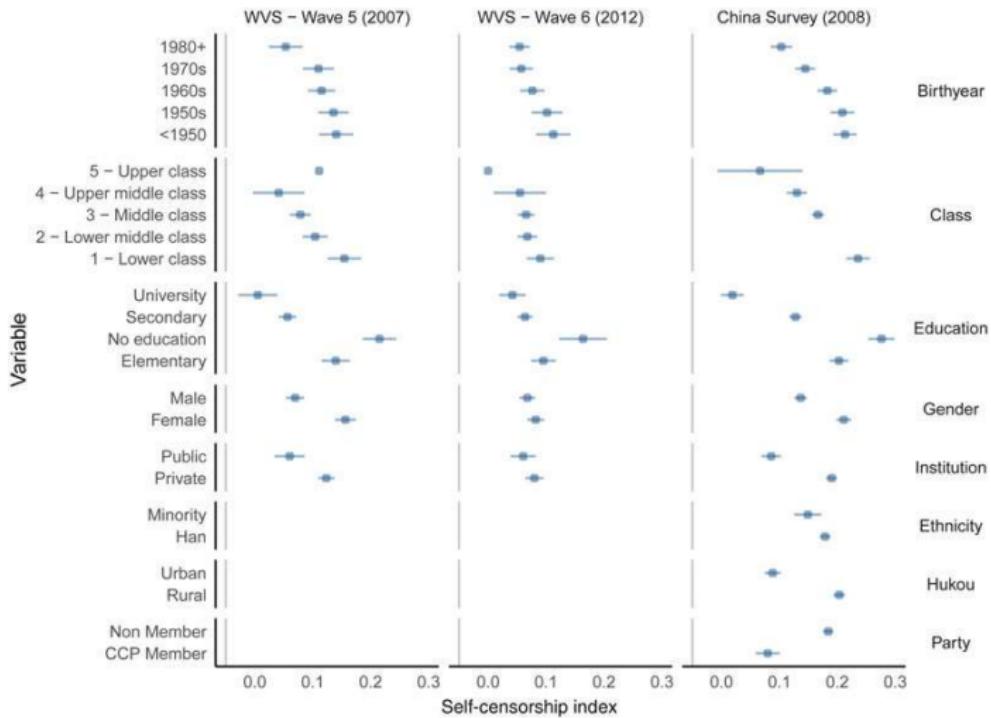
### Abstract

Item nonresponse rates across regime assessment questions and nonsensitive items are used to create a self-censorship index, which can be compared across countries, over time and across population subgroups. For many authoritarian systems, citizens do not display higher rates of item nonresponse on regime assessment questions than their counterparts in democracies. This result suggests such questions may not be that sensitive in many places, which in turn raises doubts that authoritarian citizens are widely feigning positive attitudes towards regimes they secretly despise. Higher levels of self-censorship are found under regimes without electoral competition for the executive.

**Keywords:** preference falsification; item nonresponse; authoritarian; China; self-censorship

**Table 1.** Questions for falsification index construction world values survey – core questionnaire

Regime Assessment Questions	Non sensitive Questions
V115. I am going to name a number of organizations. For each one, could you tell me how much confidence you have in them: is it a great deal of confidence, quite a lot of confidence, not very much confidence or none at all? The government (in your nation's capital) (conf.gov)	V111. I am going to name a number of organizations. For each one, could you tell me how much confidence you have in them: is it a great deal of confidence, quite a lot of confidence, not very much confidence or none at all? Television (conf.tv)
V141. How democratically is this country being governed today? (democracy)	V24. Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people? (gen.trust)
V142. How much respect is there for individual human rights nowadays in this country? (human rights)	V23. All things considered, how satisfied are you with your life as a whole these days? (life.sat)



**Figure 4.** Self-censorship in China by subgroup.

Note: figure shows mean self-censorship index across different population subgroups in the China Survey and China samples of the WVS (Waves 5 and 6). Segments depict 95 per cent confidence intervals.

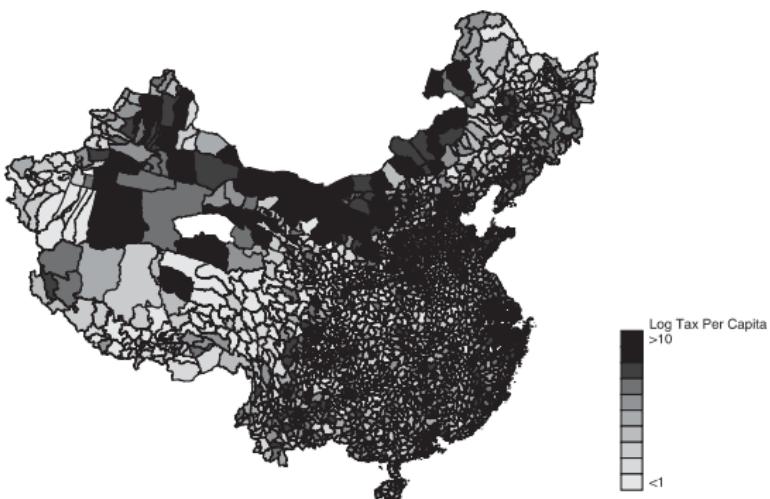
## Show Me the Money: Interjurisdiction Political Competition and Fiscal Extraction in China

XIAOBO LÜ *University of Texas*

PIERRE F. LANDRY *University of Pittsburgh*

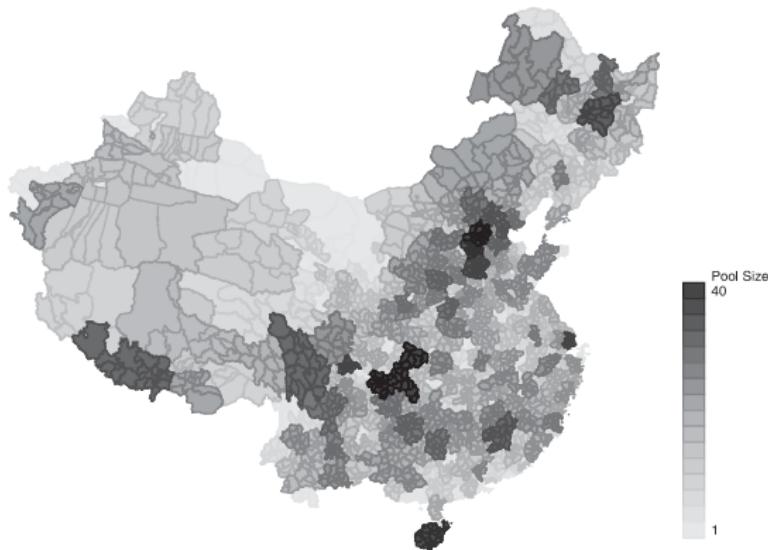
We argue that interjurisdiction competition in authoritarian regimes engenders a specific logic for taxation. Promotion-seeking local officials are incentivized to signal loyalty and competence to their principals through tangible fiscal revenues. The greater the number of officials accountable to the same principal, the more intense political competition is, resulting in higher taxation; however, too many officials accountable to the same principal leads to lower taxation due to shirking by uncompetitive officials and the fear of political instability. Using a panel dataset of all Chinese county-level jurisdictions from 1999–2006, we find strong evidence for an inverse U-shaped relationship between the number of county-level jurisdictions within a prefecture—our proxy for the intensity of political competition—and fiscal revenues in most provinces but not so in politically unstable ethnic minority regions. The results are robust to various alternative specifications, including models that account for heterogeneous county characteristics and spatial interdependence.

**FIGURE 1.** County-level Fiscal Extraction in China (2005)



*Data source:* Authors' Database. This map is drawn on county boundaries.

**FIGURE 3. County-Level Interjurisdiction Political Competition in China (2005)**



Data source: Authors' database. This map is drawn on county boundaries.

**TABLE 1. The Level of Fiscal Revenue Extraction (Prefecture)**

	Log(Tax Per Capita)					
	All	All but Tibet, Xinjiang, Ningxia	Tibet, Xinjiang, Ningxia	All	All but Tibet, Xinjiang, Ningxia	Tibet, Xinjiang, Ningxia
	(1)	(2)	(3)	(4)	(5)	(6)
Pool Size	0.043*** (0.015)	0.050*** (0.015)	-0.007 (0.055)	0.048*** (0.015)	0.058*** (0.015)	-0.004 (0.059)
Pool Size <sup>2</sup>	-0.001** (0.001)	-0.002** (0.001)	-0.002 (0.002)	-0.002** (0.001)	-0.002*** (0.001)	-0.002 (0.002)
Log(Population)	-0.913*** (0.064)	-0.929*** (0.057)	-0.792*** (0.223)	-0.833*** (0.065)	-0.824*** (0.059)	-0.815*** (0.226)
Log(Area Size)	0.038 (0.033)	0.037 (0.034)	-0.014 (0.146)	0.201*** (0.050)	0.246*** (0.055)	0.029 (0.166)
% of Rural Population	-0.011*** (0.002)	-0.010*** (0.002)	-0.016* (0.009)	-0.009*** (0.002)	-0.009*** (0.002)	-0.011 (0.010)
Log(GDP)	0.874*** (0.047)	0.875*** (0.045)	0.871*** (0.214)	0.784*** (0.049)	0.770*** (0.047)	0.802*** (0.213)
Log(Brightness per capita)				0.161*** (0.029)	0.192*** (0.034)	0.087 (0.055)
Minority Prefecture				0.053 (0.069)	0.012 (0.077)	0.107 (0.204)
Minority County				0.066 (0.116)	0.030 (0.119)	-0.061 (0.900)
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Province × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,673	2,465	208	2,661	2,453	208

Note: Clustered standard errors at the prefecture level are reported in the parentheses. We did not report the coefficient estimates of the constant as well as provincial and yearly dummies. \* $p<0.1$ , \*\* $p<0.05$ , \*\*\* $p<0.01$ .

## How Censorship in China Allows Government Criticism but Silences Collective Expression

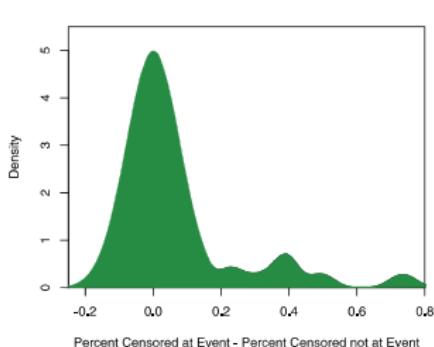
GARY KING *Harvard University*

JENNIFER PAN *Harvard University*

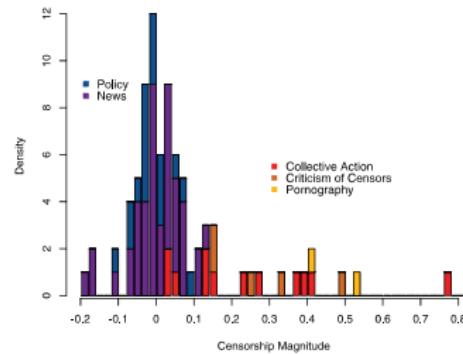
MARGARET E. ROBERTS *Harvard University*

We offer the first large scale, multiple source analysis of the outcome of what may be the most extensive effort to selectively censor human expression ever implemented. To do this, we have devised a system to locate, download, and analyze the content of millions of social media posts originating from nearly 1,400 different social media services all over China before the Chinese government is able to find, evaluate, and censor (i.e., remove from the Internet) the subset they deem objectionable. Using modern computer-assisted text analytic methods that we adapt to and validate in the Chinese language, we compare the substantive content of posts censored to those not censored over time in each of 85 topic areas. Contrary to previous understandings, posts with negative, even vitriolic, criticism of the state, its leaders, and its policies are not more likely to be censored. Instead, we show that the censorship program is aimed at curtailing collective action by silencing comments that represent, reinforce, or spur social mobilization, regardless of content. Censorship is oriented toward attempting to forestall collective activities that are occurring now or may occur in the future—and, as such, seem to clearly expose government intent.

**Figure 3. “Censorship Magnitude,” The Percent of Posts Censored Inside a Volume Burst Minus Outside Volume Bursts.**



(a) Distribution of Censorship Magnitude



(b) Censorship Magnitude by Event Type

## Consultation and Selective Censorship in China

Dimitar D. Gueorguiev and Edmund J. Malesky

PDF

PDF PLUS

Abstract

Full Text

Supplemental Material



### Abstract

Conventional theories of authoritarianism view the need to suppress criticism as a key function of censorship. In a 2013 article, King, Pan, and Roberts challenged this wisdom by arguing that the paramount goal of censorship is defusing collective action, not silencing dissent. After accounting for collective action potential, they argue that criticism has no bearing on censorship. In this research note, we point out that a significant portion of sampled posts in King et al.'s analysis coincided with state-led consultation campaigns that were aimed at soliciting critical public input on policy proposals. This introduces the potential for bias by combining solicited and unsolicited criticism under the generic title of criticism. After reanalyzing King et al.'s aggregate data, studying Chinese censorship directives, and offering a statistically guided thought experiment, we conclude that a more conservative version of their original thesis is in order.

# Searching for Red Songs: The Politics of Revolutionary Nostalgia in Contemporary China

Shuanglong Li<sup>\*</sup> and Fei Yan<sup>†</sup>

## Abstract

Applying a novel approach based on online query volume data, this study provides the first large-scale portrait of revolutionary nostalgia among the Chinese, undertaking an empirical analysis of how the aggregate level of nostalgia is shaped. For each Chinese province, we use the normalized frequency of searches for red songs on Baidu, the most widely used online search engine in China, to quantify the local level of nostalgia. We find that the evolving trends of nostalgia among the provinces are similar but stratified. The results from the dynamic panel data analysis using the Generalized Method of Moments indicate that revolutionary nostalgia is significantly affected by a set of socio-economic determinants, including GDP per capita, income inequality, social development, legal development and the degree of globalization.

**Keywords:** red song; revolutionary nostalgia; China; big data; dynamic panel data analysis

Table 1: The Full List of Red Songs (Search Terms in the Baidu Index)

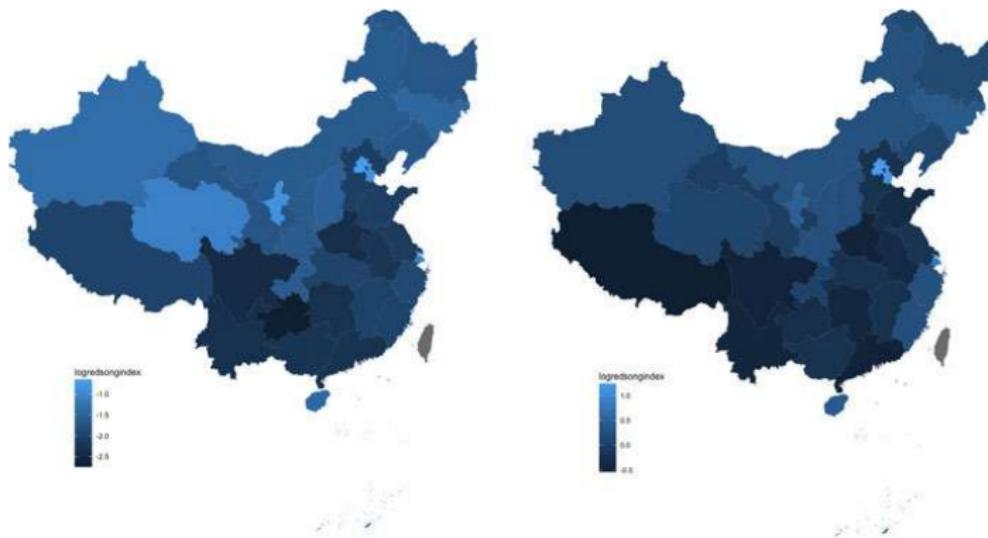
### New Red Songs

最美的歌儿唱给妈妈	阳光路上	情系老百姓	红旗飘飘
走向复兴	小白杨	强军战歌	好男儿就是要当兵
祖国万岁	想家的时候	你是英雄	歌声与微笑
追寻	相亲相爱	南湖菱花开	光荣与梦想
祝福祖国	五星红旗	卢沟谣	共筑中国梦
祝福你亲爱的祖国	我要去延安	两岸一家亲	共和国之恋
中国人	我们众志成城	今天是你的生日	多好啊
咱当兵的人	我们的天空	假如我是你	春天的故事
在灿烂阳光下	我和我的祖国	家乡的味道	把一切献给党
一二三四歌	我的中国心	家的牵挂	把你心交给你
迎风飘扬的旗	为了谁	加油中国	爱中华

### Traditional Red Songs

遵义会议放光辉	我们走在大路上	没有共产党就没有新中国	光荣啊中国共青团
祖国颂	我们要做雷锋式的好少年	毛主席派人来	工农兵联合起来
中国少年先锋队队歌	我的祖国	毛主席的战士最听党的话	革命人永远是年轻
中国人民解放军进行曲	我爱祖国的蓝天	毛主席的书我最爱读	翻身农奴把歌唱
咱们工人有力量	我爱北京天安门	毛主席的话儿记心上	翻身道情
咱们的领袖毛泽东	万岁毛主席	毛主席的光辉	东方红
在太行山上	万泉河水清又清	毛委员和我们在一起	地道战
游击队之歌	听妈妈讲那过去的事情	妈妈教我一支歌	党啊亲爱的妈妈
游击队歌	太阳最红毛主席最亲	敬祝毛主席万寿无疆	当兵的人
英雄赞歌	松花江上	井冈山下种南瓜	弹起我心爱的土琵琶

Figure 3: Geographic Distribution of IRN in 2008 (left) and in 2014 (right) in Surveyed Provinces



## Work with data

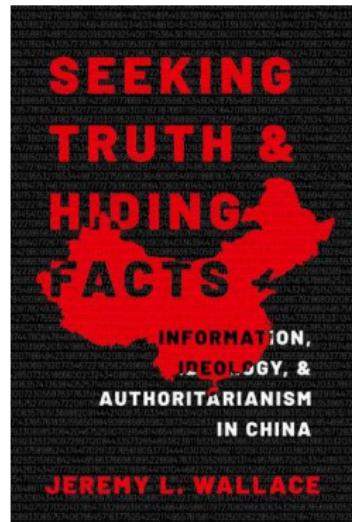
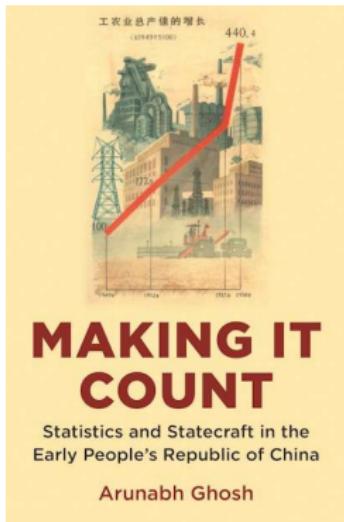
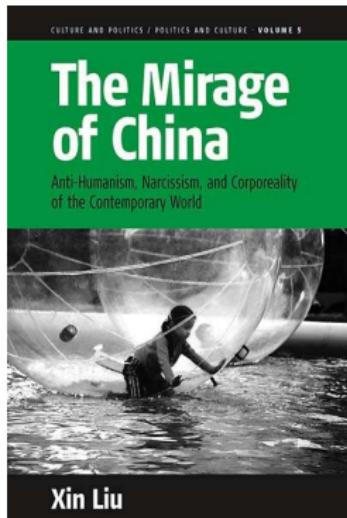
- ▶ Training for different **research methods**
- ▶ A grasp of potential **data sources**
- ▶ Solid and deep **field experiences** and/or **subject knowledge**

- ▶ Training for different **research methods**
  - Qualitative: Case study (Bayesian and machine learning), ethnography and (elite) interviews
  - Quantitative: Econometrics and regression (frequentist and Bayesian), experiments and causal inference
  - Computational: Text-as-data, social network analysis, geographic data science and ChatGPT-assisted research
- ▶ A grasp of potential **data sources**
- ▶ Solid and deep **field experiences** and/or **subject knowledge**

- ▶ Training for different **research methods**
- ▶ A grasp of potential **data sources**
  - Databases
  - Libraries and archives
  - Survey firms and research centres
  - API and web scraping (i.e., coding and programming)
  - Replication packs
- ▶ Solid and deep **field experiences** and/or **subject knowledge**

- ▶ Training for different **research methods**
- ▶ A grasp of potential **data sources**
- ▶ Solid and deep **field experiences** (e.g., politics of data generation and release) and/or **subject knowledge** (e.g., political selection and government finances)

# Data generation process is a political question



Weekly edition

The world in brief

War in the Middle East

War in Ukraine

United States

The world economy

Business

[Graphic detail | Shining light on lies](#)

# A study of lights at night suggests dictators lie about economic growth

Satellite data hints at the scale of their deception

Sep 29th 2022

Share

<https://www.economist.com/graphic-detail/2022/09/29/a-study-of-lights-at-night-suggests-dictators-lie-about-economic-growth>

# Where Are the Missing Dead? How Metrics Management Mitigates Official Data Misreporting in China

Guoer Liu\*

*Department of Political Science, University of Michigan, Ann Arbor, Michigan, USA; guoerliu@umich.edu*

---

## ABSTRACT

A perennial problem in executive politics is that agents charged with carrying out tasks have private information about their performance and corresponding incentives to manipulate this information. Conventional wisdom emphasizes external stakeholders as a source of reliable information. I instead focus on agents with competing interests and argue that when officials are subject to a relative performance evaluation metric, governing outcomes that are easily observable by their peers are more likely to be truthfully reported, while other outcomes are likely to be misreported. Using a difference-in-differences design and the varying reporting cost across sectors, combined with an original dataset on workplace accidents in China, I find that incorporating a “death cap” into the workplace safety metric management system has heterogenous effects on reported accidents. While the counts of casualties and accidents dropped by 33.8% and 29.4% respectively, the decline was entirely driven by sectors in which peer monitoring was not feasible. The results highlight the conditions under which agents are restrained by their peers. In contrast to theories that truthful signals occur when interests are aligned or independent oversight is

# Ingroup Bias in Official Behavior: A National Field Experiment in China\*

Greg Distelhorst<sup>†</sup>

Yue Hou<sup>‡</sup>

May 2014

Forthcoming in *Quarterly Journal of Political Science*

## Abstract

Do ingroup biases distort the behavior of public officials? Recent studies detect large ethnic biases in elite political behavior, but their case selection leaves open the possibility that bias obtains under relatively narrow historical and institutional conditions. We clarify these scope conditions by studying ingroup bias in the radically different political, historical, and ethnic environment of contemporary China. In a national field experiment, local officials were 33% less likely to provide assistance to citizens with ethnic Muslim names than to ethnically-unmarked peers. We find evidence consistent with the ingroup bias interpretation of this finding and detect little role for strategic incentives mediating this effect. This result demonstrates that neither legacies of institutionalized racism nor electoral politics are necessary to produce large ingroup biases in official behavior. It also suggests that ethnically motivated distortions to governance are more prevalent than previously documented.

## Past and current research

- ▶ Governance and development in multiethnic China
  - Rethinking the politics of ethnic local autonomy (book project)
  - Poverty alleviation and local state-building in peripheral provinces (*JJPS* 2021)
  - Ethnic empowerment and environmental governance
  - Duality of ethnic descriptive representation
- ▶ Using AI tools to (re)mapping the information network of political elites
- ▶ Using court records to study administrative litigation, with **Haibo He** (Tsinghua) and **Chao Ma** (UIBE)
- ▶ Multilevel agenda setting in times of crisis: A text-as-data approach, with **Tao Lin** (University of Washington)

## Concluding remarks

- ▶ Find out **who you are** and **what you want to be and do** – your decision may depend on where you are (or going back to history)
- ▶ Curate the skills you need – be **selectively thorough**
- ▶ Find those who can support and collaborate with you – engage in **interdisciplinary knowledge co-production**
- ▶ Read broadly – **go beyond political science and (mainland) China** to gain inspiration and a broader comparative perspective

Thank you!

c.cheng@bbk.ac.uk

# Rethinking the politics of ethnic local autonomy

- ▶ Why do the Communist Party of China grant **ethnic local autonomy**?
- ▶ When and where will the Center designate **ethnic autonomous territories** (EATs)?
- ▶ How can local (policy/administrative) decentralization contribute to nation-state building?

# Rethinking the politics of ethnic local autonomy

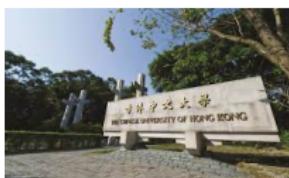
- ▶ **Archival research** in the US, Hong Kong and China
- ▶ **Elite interviews:** Conversations with Han and non-Han cadres at the central and local levels across different provinces
- ▶ **Process-tracing and QCA:** Structured process tracing the designation of ethnic autonomous prefectures



Hoover Institution Archives  
(Palo Alto, California)



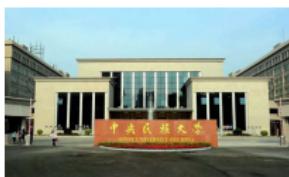
University of California, Berkeley  
(Berkeley, California)



The Chinese University of Hong Kong  
(Hong Kong)



Fudan University  
(Shanghai)



Minzu University of China  
(Beijing)



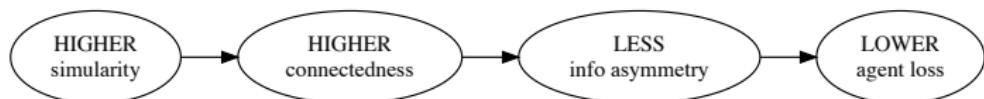
The National Library of China  
(Beijing)



# Rethinking the politics of ethnic local autonomy

- ▶ **Supervised machine learning and social network analysis:** Capturing the *connectedness* and *network structure* among/between central and local cadres based on an original biographical dataset of Chinese political elites (1949-2003)

- Between the pair of cadres, I create the indices of **connectedness** based on their **biographical similarity**



- The same connectedness measure can be used to construct the information network of Politburo members and measure the degree of **central leadership fragmentation** by **community detection**

- ▶ **Multivariate regression:** Modeling the number of active EATs as a function of different network structural indicators

# Rethinking the politics of ethnic local autonomy

- ▶ Ethnic local autonomy acts as an **institution of agent control**

	Results
(1) Central leadership fragmentation (Modularity of the Politburo network)	+ EATs
(2) Local leadership embeddedness (Similarity b/w provincial Party secretary and governor)	+ EATs
(3) Central-local connectedness (Similarity b/w central leader and provincial Party secretary)	– EATs
(4) Central fragmentation + Local embeddedness	+ EATs
(5) Central fragmentation + Central-local connectedness	– EATs
(6) Central-local connectedness + Local embeddedness	– EATs

# Rethinking the politics of ethnic local autonomy

- ▶ EATs act as **enclaves** that constrain the power of provincial elites
  - Sub-provincial EATs are **more** likely to pass and revise autonomous regulations when the provincial elites are potentially defiant
  - Ruling non-Han cadres in sub-provincial EATs are **less** likely to have worked at the provincial level prior to their appointments

# Rethinking the politics of ethnic local autonomy

- ▶ Going **cross-national**: Post-WWII authoritarian regimes are more likely to introduce regional autonomy to ethnic minorities when their ruling power is contested within the inner circle
- ▶ Going back to **history**: Post-civil war *decentralization* and *state power consolidation* in imperial and Republican China
  - Western Han dynasty (202 BC–220 AD): Kingdoms and marquis states
  - The Kuomintang (KMT) regime (1928–1949 AD): Countering recalcitrant provincial warlords through the county-level self-government movement

## (Re)mapping the information network of political elites

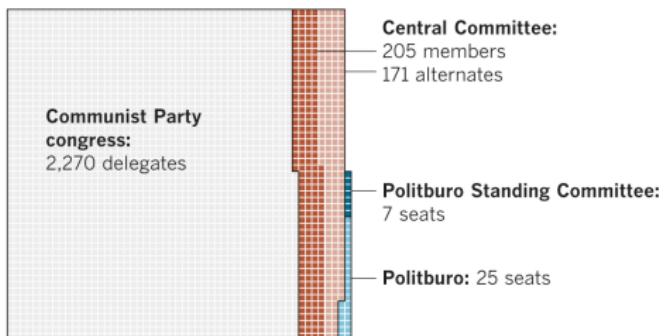
- ▶ The focus on "high-politics" in the literature on Communist regimes: The absence of independent, strong formal institutions has pushed researchers have to focus on prominent political elites in authoritarian regimes
- ▶ **Factionalism** is one of the most researched topics in the Chinese politics literature (e.g., Nathan 1973; Tsou 1976; Pye 1981; Dittmer 2002; Huang 2006; Bo 2007; Shih et al 2012)
- ▶ The power struggles and cooperation between different factions or "informal" groups have been considered as the main driving forces for policy and political changes in the country

## (Re)mapping the information network of political elites

- ▶ Create the timeline of key life events of each cadre using public biographic information
  - Demographic characteristics (e.g., birth year and provincial origin)
  - Educational background (e.g., colleges and universities)
  - Political career (e.g., involvement in key events between 1921 and 1949; Party and government posts after 1949)
- ▶ Create the (decomposed) **attribute vector** of a cadre based on her/her timeline and measure the distance/similarity between a pair of attribute vectors in the multidimensional space
  - Time: Are the cadres of interest both alive in a given year?
  - Location: Are they both at the same location in a given year?
  - Activity: Are they working in the same department/unit in a given year?

# (Re)mapping the information network of political elites

## China's Communist Party congress (as of 2012)



Source: Los Angeles Times reporting

Angelica Quintero / @latimesgraphics

- ▶ Use **variable selection** (elastic net regularization and other methods) to identify the key biographical variables in CC of each year when measuring biographical similarity
- ▶ Use **community detection** to assess the degree to which a Politburo network can be divided

# Using court records to study admin litigation

The release of **China Judgements Online** (<https://wenshu.court.gov.cn/>) made a crucial breakthrough in the empirical legal studies in China (Ma, Yu and He 2016; Liebman et al 2020).



200  
1823-2023

# Using court records to study admin litigation

► **Question:** Are the local courts more likely stand with the citizens following the reforms?

- Railway transport courts (*China Review* 2022)
- Case registration system
- Elevated and off-site trials

► **Extensions**

- Who do the citizens go to the courts? A national **conjoint** experiment
- Local dynamics of judicial decision-making: A **text-as-data** approach