

UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

The Politics of Poor Air Quality in Contemporary China /

Permalink

<https://escholarship.org/uc/item/8sn173wv>

Author

Oliver, Steven

Publication Date

2014

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

The Politics of Poor Air Quality in Contemporary China

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Political Science

by

Steven Oliver

Committee in charge:

Professor Susan L. Shirk, Chair
Professor James H. Fowler
Professor Edmund J. Malesky
Professor Barry J. Naughton
Professor Philip G. Roeder

2014

The Dissertation of Steven Oliver is approved, and it is acceptable in
quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2014

DEDICATION

To my mother, father, and sister,
I wouldn't have achieved anything without you.

TABLE OF CONTENTS

SIGNATURE PAGE	iii
DEDICATION.....	iv
TABLE OF CONTENTS.....	v
LIST OF FIGURES	viii
LIST OF TABLES.....	ix
ACKNOWLEDGEMENTS.....	x
VITA.....	xii
ABSTRACT OF THE DISSERTATION	xiii
CHAPTER 1: INTRODUCTION.....	1
Campbell’s Law and the Cadre Evaluation System.....	4
Air Quality and Agenda Setting.....	6
Making Sense of Manipulation.....	9
CHAPTER 2: CAMPBELL’S LAW AND THE CADRE EVALUATION SYSTEM ...	11
Campbell’s Law and the Cadre Evaluation System.....	15
Poor Air Quality and Chongqing’s Blue Sky Action Campaign	20
Procedure to Detect Evidence of Manipulation and Classify Air Quality Data	25
Data Description and Classifying County-Specific Samples	29
Results and Analysis.....	31

<i>Multivariate Regression Results</i>	32
<i>Instrumental Variables (IV) Regression Results</i>	36
Robustness Checks and Sensitivity Analysis.....	40
<i>Alternative Model Selection and Sensitivity Analysis</i>	41
<i>Robustness Check #1: Tenure of Bo Xilai</i>	42
<i>Robustness Check #2: Complicity and Collusion</i>	44
Discussion and Implications	46
CHAPTER 3: AIR QUALITY AND AGENDA SETTING	48
Agenda Setting under Authoritarianism	52
Air Quality, Public Skepticism, and Agenda Setting in Contemporary China.....	55
Data and Methods Description	61
<i>Data Description</i>	62
<i>Methods Description</i>	66
Results and Analysis.....	69
Discussion and Implications	72
CHAPTER 4: MAKING SENSE OF MANIPULATION.....	74
Debate over Performance and Promotion.....	78
Literature on Manipulation in Finance and Accounting.....	81
Manipulation and the Response of Leaders.....	85
Implications for Tests of the Performance-Promotion Hypothesis	91
Discussion and Implications	93
CHAPTER 5: CONCLUSION	95
APPENDIX A.....	99

APPENDIX B	112
REFERENCES	119

LIST OF FIGURES

Figure A1: Map of Chongqing and Chongqing’s County-Level Jurisdictions.....	99
Figure A2: Histogram of Station-Level Air Quality Data Reported from Beijing	100
Figure A3: Histograms of County-Level API Scores by Target Assignment Status Reported from Chongqing	101
Figure A4: Probability Density Functions and Empirical Cumulative Density Functions of County-Level Z-Scores by Target Assignment Status	102
Figure A5: McCrary (2008) Test for Evidence of Post-Bo Xilai Manipulation.....	103
Figure A6: McCrary (2008) Test for Evidence of Collusion.....	104
Figure B1: Google Trends Score for Keyword ‘Air Quality’	112
Figure B2: Analysis of Structural Breaks in Daily Volume of Weibo Posts from Verified Sina Weibo Accounts.....	113
Figure B3: Analysis of Structural Breaks in Daily Volume of Weibo Posts from Non- Verified Sina Weibo Accounts	114
Figure B4: Analysis of Structural Breaks in Daily Baidu Media Index Scores	115

LIST OF TABLES

Table A1: Summary Statistics on County-Level Air Quality by Target Assignment Status	105
Table A2: Difference in Proportions of Suspect Samples by Target Assignment Status	106
Table A3: Summary Statistics for County-Level Jurisdictions by Target Assignment Status	107
Table A4: Firth Logistic Regression of County-Level Classification Status on Target Assignment Status	108
Table A5: OLS Regression of County-Level Classification Status on Target Assignment Status	109
Table A6: OLS Regression of County-Level Z-Scores on Target Assignment Status ..	110
Table A7: Sensitivity Analysis using Different Classification Thresholds and Z-Scores Calculated using Alternative Bandwidths	111
Table B1: Summary Statistics for Sina Weibo Posts and Baidu Media Index	116
Table B2: Intercept-Only Structural Break Models	117
Table B3: Temporal Order of Breaks in Public and Media Attention	118

ACKNOWLEDGEMENTS

It has been over a decade since I first came to UC San Diego as an undergraduate and nearly as many years since I entered the Department of Political Science. In that time, I have come to rely upon the support and encouragement of countless faculty and peers as well as friends and associates in the UC San Diego community. Given the number of people that have contributed to my development as a scholar as well as my life while a student in San Diego, the task of giving them all proper recognition is rather daunting. I am certain to omit the names of many who deserve a place in these acknowledgements. If I have left you out, know that it is my poor memory rather than any lack of affection towards you and all that you have done for me over the years that is to blame: I know that I would not have achieved what I have without you.

Among the faculty, I must acknowledge my committee members for their incredible patience with my perfectionism and procrastination. I know I have not been an easy student to mentor. Nevertheless, I am eternally grateful for their guidance over the years and they have all left an important mark upon me as a scholar. The support and encouragement provided by my advisor Susan Shirk is the primary reason that I decided to pursue the study of Chinese politics. Barry Naughton's sage insights into Chinese politics and economics have helped me to develop the substantive knowledge necessary to become a half-credible China scholar. The sincere care that Phil Roeder demonstrates for the welfare of all students within the Department as well as his dedication as a scholar is a constant inspiration. The scholarly and life advice that Eddy Malesky and James Fowler have shared over the years have enriched my scholarship as well as my life more generally. Without their assistance, my work and I would both be the lesser.

In addition to the members of my committee, I would also like to acknowledge the guidance of Matthew McCubbins and David Lake in teaching me to think like a social scientist. Lastly, I wish to thank Erik Gartzke for always encouraging me to allow my passion to determine my research agenda.

My peers also deserve great praise. Ryan Jablonski and Laurie Gay, you have both been a source of constant support and encouragement. Rupal Mehta, you have been a wonderful friend for many years and always been a great listener. I am lucky to have a friend like you. Dan Enemark, you have been a great friend, roommate, and advisor. Your counsel is appreciated more than I sometimes let you know. Jonathan Mark, nobody could ask for a better supporter and advocate. However, you are perhaps one of the worst cooks that I know. Dimitar Gueorgiev, Paul Schuler, and Kai Ostwald, thank you for the many criticisms and feedback that you have offered. Ben Tang, Chris Fariss, and Micah Gell-Redman, you all have brilliant minds and I am always grateful whenever you are able to lend them to whatever project I am working. But Micah, get some sleep and do not let your children put you in an early grave. As for everyone else unnamed within the Department of Political Science, I cannot thank you enough for all that you have done for me over the years.

Last but not least, I want to recognize my family. I would not be where I am today without the support of my parents Barbara and Mark and my sister Kendra. You have always gone the extra mile for me and I know that I would not have made it anywhere near as far as I have without you. I am fortunate to have you all and it is with the deepest love that I dedicate this dissertation to you.

VITA

2004-2006	Department of Homeland Security, Undergraduate Scholar
2006	Bachelor of Science, University of California, San Diego
2006-2009	Department of Homeland Security, Graduate Fellow
2009-2010	National Security Education Program Boren Fellow
2010	Master of Arts, University of California, San Diego
2010-2014	Teaching Assistant, Department of Political Science, University of California, San Diego
2014	Doctor of Philosophy, University of California, San Diego

PUBLICATIONS

“The Political Economy of Plunder: Economic Opportunity and Modern Piracy.” *Journal of Conflict Resolution*, 57(4): 682-708.

FIELDS OF STUDY

Major Field: Political Science

Sub-Field: Chinese Politics

Professors Barry J. Naughton and Susan L. Shirk

Sub-Field: Comparative Politics

Professors Edmund J. Malesky, and Philip G. Roeder

Sub-Field: Methods

Professor James H. Fowler

ABSTRACT OF THE DISSERTATION

The Politics of Poor Air Quality in Contemporary China

by

Steven Oliver

Doctor of Philosophy in Political Science

University of California, San Diego, 2014

Professor Susan L. Shirk, Chair

What can the issue of poor air quality in China's rapidly growing urban centers teach scholars about politics in contemporary China and authoritarian regimes more generally? This dissertation argues that the answer is (surprisingly) quite a bit. To this end, the dissertation makes use of a wide range of original data and novel methods to offer insights from a recent campaign to improve ambient air quality in the province-ranked municipality of Chongqing as well as recent episodes of abysmal air quality and resulting public outcry over ambient air quality monitoring and reporting standards into a number of ongoing scholarly debates. For example, do top-down accountability

institutions such as the Chinese Communist Party (CCP) cadre evaluation system through which leaders link the promotion of sub-national officials to quantitative indicators of their performance in office create perverse incentives for officials to manipulate these same indicators? If officials can indeed manipulate indicators, then is it possible for leaders to promote officials based upon their performance in office? Absent bottom-up accountability institutions like competitive elections and where leaders maintain substantial control over the media, who sets the public agenda and can the public agenda influence the policy agenda? The following chapters address these and related questions.

CHAPTER 1: INTRODUCTION

What can the issue of poor air quality teach scholars about politics in the contemporary People's Republic of China? Moreover, what can it teach scholars about politics under authoritarian regimes more generally? These are the broad questions that motivate and unify the subsequent chapters of this dissertation.

Although the thin, grayish-brown band that regularly floats just above the horizon of major urban centers like that of Beijing has long passed unnoticed and uncommented upon by urban residents, this is no longer true. Recent episodes of abysmal air quality and an unexpectedly vocal public response reveal the potential political dimensions of the issue of poor air quality. When dense smog descended on the national capitol of Beijing in October and November 2011, it provoked a significant outpouring of public criticism directed at local officials, skepticism over the reliability of government reported air quality information that grossly understated the severity of air pollution, and calls for leaders to reform national ambient air quality monitoring and reporting standards in line with international standards (Wong 2011).¹ Just over one year later, dense smog returned once again to engulf Beijing as well as urban centers across China throughout January and February 2013 in what the international press has since coined the 'Airpocalypse' (Lim 2013). Public outcry arising from these and related episodes appear to have been the

¹ This was not the case a year prior when the U.S. Embassy in Beijing, in an apparent gaffe, publicly described abysmal air quality in the national capitol on 14 November 2010 as 'crazy-bad'. The statement, based upon air quality monitoring data independently collected by the U.S. Embassy, resulted in immediate protest from Chinese state officials that resulted in a hasty retraction of the original statement (Watts 2010). This episode drew widespread international attention, but appears to have passed unnoticed and uncommented upon by Beijing residents.

catalysts for leaders to put in place public policies to address the menacing haze obscuring the horizon.

By the numbers, the estimated social welfare losses to the Chinese public from poor air quality as well as the drag that these costs impose on China's development are staggering: According to Matus et al. (2012), the social-welfare losses resulting from poor air quality in 2005 were USD 112 billion, or a 5% annual loss of GDP. The authoritative 2010 Global Burden of Disease Study further estimates that poor air quality contributed to 1.2 million premature deaths and the loss of 25 million healthy years of life in 2010 alone (Yang et al. 2013). Although drawing on historical data, Chen et al. (2013) demonstrate that exposure to air pollution resulting from state policies originally put in place during the Mao era to provide public heating to the residents of urban centers in northern China contribute to an estimated loss of 2.5 billion life years of life expectancy, or an average of 5.5 years per person for the 500 million residents of northern China.

While the social welfare losses caused by poor air quality are large and demand public policy solutions to address this issue, what makes the issue of poor air quality political and that increasingly commands the attention of political leaders is the potential for this issue, if left unaddressed, to threaten social stability. On the one hand, it seems implausible that abysmal air quality in Beijing or other major urban centers could bring the public onto the streets in protest. Furthermore, it is unclear that any such protests would threaten the rule of incumbent leaders, or that such protests could destabilize the ruling Chinese Communist Party (CCP) regime.

Yet on the other hand, environmental issues have increasingly become a cause of protest and a potential source of social instability that concerns leaders (Economy 2004,

2007). Aside from anti-Japanese protests, the largest protests by number of participants since the Tiananmen protests of 1989 have been related to public concerns over local environmental conditions. For example, protests against the placement and operation of paraxylene (PX) chemical plants in urban centers like Xiamen, Fujian province and Dalian, Liaoning province in 2007 and August 2011 drew and estimated 8,000-10,000 and 10,000-70,000 protestors (Cody 2007; Watts 2011). Although both sets of protests were largely peaceful, more recent protests against PX plants in Maoming, Guangdong province in April 2014 turned violent and resulted in smaller, sister protests breaking out in major urban centers including Shenzhen and Guangzhou (Li 2014).

More generally, scholarly literature on protests suggests that social scientists have a spectacularly poor record of identifying the causes and predicting the spread of protests that can threaten leaders or destabilize their regimes (Kuran 1991, 1995; Lohmann 1994). Just as social scientists are unable to foresee such threats, leaders are also unlikely to be able to *ex ante* determine the cause of protests that may threaten their rule. Consistent with theories which posit that even the leaders of authoritarian must provide at least some minimum acceptable level of public goods in order to stave off threats to their rule from outside of their regimes (Bueno de Mesquita et al. 2003; Lake and Baum 2001), this dissertation argues that the threat to social instability from leaders failing to address issues such as poor air quality ultimately creates incentives for leaders to respond.

With these points in mind, this introduction describes the three core chapters that make up this dissertation, the puzzles and debates that motivate each chapter, and the contributions they offer to current literature on contemporary Chinese politics and the politics of authoritarian regimes more generally.

Campbell's Law and the Cadre Evaluation System

When an indicator becomes a target, does it cease to be a good measure of the process it is intended to monitor? The answer to this question is important for the efforts of leaders to address the issue of poor air quality as well as scholarly debate over the role of top-down accountability institutions in three decades of spectacular development in China. Whereas some scholars give substantial credit to the CCP cadre evaluation system—a top-down accountability institution through which leaders link the promotion of sub-national officials to quantitative indicators of their performance in office—in allowing leaders to create powerful incentives for officials to implement leaders' desired policies and deliver on leaders' development targets and goals, other scholars remain suspicious that linking promotion to indicators of performance creates perverse incentives for officials to manipulate indicators in order to exaggerate accomplishments and understate failures.

In addition to the obvious problems that manipulation and resulting bias in indicators implies for the ability of leaders to evaluate the impact of public policy interventions carried out by sub-national officials, such behavior also has a number of other important implications. First, if officials manipulate indicators and leaders condition policy decisions on these indicators, then leaders run a greater risk of committing potentially costly policy errors. Second, if officials manipulate indicators that are then used to report improvements in the provision of public goods like clean air to the public, then these indicators may diverge from actual conditions and be seen as less credible by the public. Third, if officials manipulate indicators and leaders do indeed condition promotion decisions on officials' performance vis-à-vis reported indicators,

then this raises uncomfortable questions for performance-based models of promotion used by some scholars to explain China's spectacular development and argue that promotion within the modern party and state hierarchy is essentially meritocratic.

Yet demonstrating evidence of sub-national officials acting on such perverse incentives is extremely challenging: In order for manipulation to be successful, evidence of manipulation must be unobservable to leaders as well as scholars. Although other scholars have attempted to demonstrate evidence of such behavior with respect to economic development indicators like GDP growth (Koech and Wang 2012; Wallace 2014), these approaches suffer from the issue of observational equivalence.² Demonstrating a causal link between institutions like the CCP cadre evaluation system and incentives for such behavior further complicates this task.

Despite these challenges, this chapter offers a clean test of the proposition that linking promotion to indicators creates perverse incentives for officials to manipulate indicators. It does so in the context of a recent campaign to improve ambient air quality in Chongqing, a province-ranked municipality located in China's mountainous, western interior. The chapter adopts a regression discontinuity based procedure proposed by recent environmental studies (Chen et al. 2012; Ghanem and Zhang 2013) that uses the McCrary (2008) test to detect evidence of manipulation in air quality data reported by sub-national officials and that plausibly overcomes the issue of observational equivalence in previous studies. Furthermore, the chapter leverages information from the campaign

² These studies treat economic development indicators like GDP as output and total energy consumption as input, and then attempt to demonstrate evidence of manipulation on the part of sub-national officials by examining anomalous divergences in the sub-national rate of growth between these two indicators. Although manipulation on the part of economic development obsessed officials could indeed account for dissimilar changes in the rate of GDP growth and rate of growth in energy consumption, a host of economic activities could also account for these divergences.

about the assignment of annual air quality improvement targets to specific, county-level jurisdictions within which the promotion of leading, county-level officials was linked to delivering on these targets through the local cadre evaluation system as well as an instrumental variables regression strategy to demonstrate a causal link between the cadre evaluation system and evidence of manipulation.

In doing so, this chapter is able to overcome both the empirical challenge of ruling out observational equivalence and directly linking institutions to evidence of manipulation by sub-national officials. This represents not only an empirical contribution, but also a substantive and theoretical contribution. In demonstrating evidence of manipulation linked to the cadre evaluation system, this chapter not only validates claims made in the substantive literature about the perverse incentives generated by the cadre evaluation system but also raises a series of difficult questions for performance-based models of promotion used to explain China's spectacular development. The third core chapter addresses these questions.

Air Quality and Agenda Setting

What can recent episodes of abysmal air quality and resulting public skepticism over the reliability of government reported ambient air quality information teach us about who sets the public agenda in China? Moreover, what can these episodes teach us about the ability of the public agenda to set the media agenda and influence the policy agenda under an authoritarian regime? The second core chapter moves to address the surprising public response to recent episodes of abysmal air quality in Beijing that appear to have

played a substantial role in pressuring leaders to elevate the importance of policies to combat poor air quality.

Specifically, this chapter argues that recent episodes of abysmal air quality and public outcry offer a unique opportunity to observe the process by which actors outside of the party and state such as well-known celebrities, entrepreneurs, and public intellectuals play a growing role in setting the public agenda in contemporary China. This chapter further argues that despite leaders' control over what issues are placed on the media agenda, such actors nevertheless appear to be able to set the media agenda by exerting pressure on China's highly competitive, commercial media to report on issues of public concern. Moreover, this chapter argues and demonstrates evidence that these actors also appear able to influence the policy agenda, pressuring leaders to respond to both the public and media agendas despite leaders' monopoly agenda control and the absence of bottom-up accountability institutions like competitive and multi-party elections.

This chapter demonstrates this in the context of widespread public concern and skepticism over the reliability of government reported ambient air quality information that unexpectedly emerged in Beijing in late 2011. This episode saw clear demands from the public for leaders to reform existing ambient air quality monitoring and reporting standards as well as to require sub-national officials to publicly report information on the concentration of fine particulate matter, also known as PM_{2.5}. Facing substantial pressure from both the public and questions in the media, leaders responded by announcing reforms to monitoring and reporting standards with proposed reforms specifically reflecting public demands. Moreover, leaders have followed through on these reforms, rapidly implementing new national monitoring and reporting standards and

exerting substantial pressure on officials throughout China to implement policies designed to combat poor air quality.

In order to demonstrate evidence of the public agenda setting the media agenda and even influencing the policy agenda, this chapter makes use of data and methods novel to the study of agenda setting including data on the daily volume of micro-blog posts that explicitly mention air quality made by users of China's most popular micro-blogging platform, Sina Weibo, and corresponding data on media coverage. Furthermore, the chapter uses methods for detecting and dating structural breaks in time-series data to demonstrate a sudden and sustained increase in the level of public attention to the issue of poor air quality in mid-October 2011, two weeks prior to a similar shift in media coverage and attention, and nearly a month prior to leaders' initial announcements of their intent to reform standards. In taking this methodological approach, this chapter seeks to establish temporal precedence of changes in the public, media, and policy agendas implied by recent changes to agenda setting theory.

In addition to the above substantive and empirical contributions, this chapter also makes a theoretical contribution to literature on politics in authoritarian regimes more generally. In particular, this chapter demonstrates evidence of a non-institutional channel through which actors traditionally excluded from the policy-making process can influence leaders and shift the policy agenda in their favor by setting the public agenda. While responsiveness to the public agenda on the part of leaders neither implies their accountability to the public nor that public demands fundamentally challenge leaders' monopoly over the policy agenda, it does support the observation of an increasingly

pluralized policy-making process despite the absence of any formal moves by leaders towards political liberalization.

Making Sense of Manipulation

The third core chapter of the dissertation returns to the question of top-down accountability institutions and manipulation, and is motivated by two puzzles that emerge from empirical findings in the first core chapter. These two puzzles relate to an ongoing debate central to the study of contemporary Chinese politics over whether factional affiliation or performance determines the promotion of sub-national officials within the party and state hierarchy: First, if factional affiliation rather than performance determines the promotion of sub-national officials, then what explains evidence of manipulation of indicators that leaders use to monitor officials' performance in office? Second, if officials can manipulate indicators used to monitor their performance, then is performance-based promotion actually possible?

Curiously, evidence of manipulation demonstrated in the first core chapter both supports the contention that performance matters for the promotion of sub-national officials but also raises difficult questions for existing theories that explain promotion in terms of performance. However, this chapter offers a potential solution. It argues that leaders are not naïve to the perverse incentives created for officials by linking promotion to indicators. Instead, it argues that leaders adjust indicators that they suspect officials have manipulated by discounting them based upon conjectured bias and reduce the weight that they assign to adjusted indicators (i.e. their informativeness) in making promotion decisions. This implies leaders neither uncritically use nor reject suspect

indicators. Rather, this implies that leaders incorporate information from manipulated indicators when evaluating the performance of officials. However, they do so in a more complicated way than is implicitly assumed in scholarly literature.

In making this argument, this chapter seeks to make a contribution to the ongoing debate described above. Empirical tests of the *performance-promotion hypothesis*—the hypothesis that the performance of sub-national officials in delivering rapid economic development determines their chances of promotion upwards within the party and state hierarchy that is key to performance-based models of promotion—have thus far failed to produce robust and consistent findings in support this implication. Scholars skeptical of performance-based promotion have seized upon null and weaker than predicted results to argue that alternative factors including factional affiliation instead determine promotion.

Yet null and weaker than predicted results in support of the hypothesized relationship between performance and promotion might also be explained by leaders' response to manipulation on the part of officials. If leaders account for manipulation in the manner argued above, then this should bias towards type II error in existing tests of the hypothesis. By treating indicators as less informative and reducing the weight assigned to indicators in promotion decisions, leaders reduce the sensitivity of promotion outcomes to indicators and thus reduce the strength of the observed relationship between 'performance' as measured by indicators and promotion.

CHAPTER 2: CAMPBELL'S LAW AND THE CADRE EVALUATION SYSTEM

*The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor*³

When an indicator becomes a target, does it cease to be a good measure of the process it is intended to monitor? The answer to this question is important for a recent debate in economics and political science over the role of top-down accountability institutions in China's spectacular development. On the one hand, many of scholars argue that China's spectacular development in recent decades is the result of incentives created by the ruling Chinese Communist Party's (CCP) cadre evaluation system. Specifically, these scholars argue that leaders use this political institution to create incentives for sub-national officials to actively pursue leaders' development targets and goals by linking the promotion of officials to quantitative indicators of their performance in office. Yet on the other hand, a small group of scholars argues that by linking promotion to indicators, leaders also create perverse incentives for officials to ultimately confound leaders' efforts to assess performance by manipulating those same indicators. Demonstrating a relationship between the cadre evaluation system and evidence of manipulation by sub-national officials thus raises potentially difficult questions for scholars in the former

³ The above epigram is by Donald T. Campbell and is drawn from Campbell (1976). The quote represents the classic statement of *Campbell's Law*.

camp who argue that performance strongly determines promotion in contemporary Chinese politics.

To this end, this chapter argues that leaders do indeed create such perverse incentives for officials by linking promotion to indicators through top-down accountability institutions such as the cadre evaluation system. Moreover, this chapter provides a first direct empirical test of this proposition in the context of a recent intervention to improve ambient air quality in Chongqing, a province-ranked municipality located in China's mountainous, western interior. During Chongqing's *Blue Sky Action* campaign (2005-2012), municipal leaders assigned air quality improvement targets to certain, county-level jurisdictions within the municipality. Municipal leaders then used the local cadre evaluation system to link the promotion of leading county officials in those jurisdictions to delivering a set, annual proportion of so-called Blue Sky days—days in which jurisdictions recorded Air Pollution Index (API) scores of 100 or less—in order to incentivize officials to enforce emissions abatement policies. This chapter demonstrates that jurisdictions that were assigned targets and where promotion was linked to the reported, annual proportion of Blue Sky days were nineteen times more likely to report air quality data indicative of manipulation aimed at inflating this indicator than jurisdictions that were not assigned targets.

In demonstrating this result, this chapter adopts a regression discontinuity based procedure proposed by recent environmental studies (Chen et al. 2012; Ghanem and Zhang 2013) that uses the McCrary (2008) test to detect and measure evidence of manipulation resulting from leaders' use of the annual proportion of Blue Sky days as an indicator of officials' performance in improving air quality. The chapter uses daily,

county-level API scores reported during the campaign to detect evidence of manipulation among jurisdictions and classify data reported by individual jurisdictions as either suspect or non-suspect. The chapter then uses an instrumental variables estimation strategy that leverages a national-level environmental monitoring regulation and Chongqing's mountainous geography to instrument for target assignment status. The chapter thus provides the first empirical demonstration a clear, causal relationship between the cadre evaluation system and evidence of manipulation by sub-national officials in China.

This chapter makes a number of contributions. First, the chapter makes an empirical contribution to by offering a test of the implications of *Campbell's Law* with respect to the cadre evaluation system.⁴ *Campbell's Law*—stated in the above epigram—is intuitive and perhaps even a truism, but its implications are difficult to empirically demonstrate. Manipulation, if intended to mislead leaders, can only be successful if evidence is unobservable to leaders or is observationally equivalent to reasonable, alternative explanations. Recent studies which claim to demonstrate evidence of manipulation in sudden divergences between sub-national per capita GDP growth and underlying energy consumption in China (Koech and Wang 2012; Wallace 2014) are unable to rule out reasonable, alternative explanations for such anomalous movements. The procedure used to detect evidence of manipulation in air quality data allows this chapter to plausibly overcome criticisms regarding observational equivalence. Furthermore, this chapter also makes a contribution to existing environmental studies that

⁴ An alternative statement of the same principle is attributed to economist Charles Goodhart and known Goodhart's Law states that any statistical regularity will tend to collapse once pressure is placed upon it for control purposes (Goodhart 1975). However, this chapter refers to the principle as Campbell's Law rather than Goodhart's Law as use of the later term has generally been constrained to use in discussions of monetary policy and inflation targeting.

are institutions-free by demonstrating how the cadre evaluation system explains variation in evidence of manipulation.

Second, the chapter makes a substantive contribution to literature on the cadre evaluation system. While not denying that the cadre evaluation system allows leaders to create incentives for officials to deliver on leaders' development targets, perverse incentives for officials to manipulate indicators that leaders use to monitor officials' performance is troublesome for the prevailing literature. Manipulation, whether carried out by officials through the falsification of data used to generate indicators or selective enforcement of policies whose impacts are short-term and largely cosmetic, has non-trivial implications for leaders: Leaders must rely upon indicators of questionable validity not only to make policy decisions but also in evaluating the performance of officials as well as decisions regarding the promotion of subordinate officials into positions of greater authority within the party and state hierarchy, and ultimately regime leadership.⁵ Evidence of a causal relationship between cadre evaluation system and manipulation cannot help but raise questions about whether the promotion of officials in China can in fact be performance-based and meritocratic if leaders cannot trust indicators used to monitor performance.

Lastly, the chapter makes a theoretical contribution to the broader literature on authoritarian institutions. Spectacular development in China is taken as evidence by some that the leaders of the ruling CCP have somehow *gotten it right* with respect to rule through top-down accountability institutions while their peers in the Soviet Union and

⁵ This chapter uses validity in the sense of measurement theory such that the validity of a given indicator is defined as the degree that it measures the phenomena that it is intended to measure. This is distinct from reliability and that instead describes the degree to which the indicator measures the phenomena with consistency across multiple observations.

Eastern Europe did not. Moreover, leaders have achieved this outcome while largely eschewing risky, bottom-up institutions like elections or media freedom. Yet while leaders may indeed be able to create powerful incentives for officials to deliver on leaders' development targets and goals, evidence of manipulation suggests leaders certainly have not somehow overcome problems of moral hazard inherent to rule through top-down accountability institutions.

The remainder of the chapter is divided as follows: First, the chapter discusses Campbell's Law and the cadre evaluation system. Second, the chapter describes leaders' interests in improving air quality as well as officials' incentives under the *Blue Sky Action* campaign; Third, the chapter details the procedure used to detect evidence of manipulation and classify samples of air quality data; Fourth, the chapter describes the data used to construct the dependent variable; Fifth, the chapter presents results and analyses, also addressing the potential endogeneity of target assignment status; Sixth, the chapter provides sensitivity analysis of its main results and defends the interpretation of its results by offering a series of robustness checks; Lastly, it concludes with a brief restatement of this chapter's findings and an explanation of how findings presented in this chapter raises additional puzzles for a core debate in the literature on contemporary Chinese politics.

Campbell's Law and the Cadre Evaluation System

The essence of *Campbell's Law* is that when policy-makers attempt to shape the behavior of actors through policy based upon quantitative indicators, they cannot help but to create perverse incentives for these same actors to engage in hidden action to influence

indicators in a manner contrary to policy-makers' interests. Actors may attempt to directly falsify data used to generate the indicator. Alternatively, actors may alter their behavior and distort how they allocate resources in order to influence the underlying process an indicator is intended to monitor. Regardless of substantive differences, this chapter refers to such action on the part of officials as manipulation.

Within principal-agent relationships, such manipulation on the part of agents represents an instance of moral hazard and its consequences represent agency losses to principals. The implications of Campbell's Law and agency losses are by no means unique to politics much less democratic or non-democratic politics. However, they represent an acute problem for the leaders of authoritarian regimes who rely heavily upon top-down accountability institutions to ensure their rule. As principals, leaders must rely on subordinate officials to implement policy in support of regime stability as well as their own survival in office. As agents, subordinate officials may prefer to shirk their duties rather than implement leaders' desired policies and then report back leaders' desired but ultimately unrealized policy outcomes. To the degree that monitoring is costly and the leaders of authoritarian regimes are constrained in their ability to employ relatively low-cost "fire-alarm" versus high-cost "police-patrol" type monitoring institutions (McCubbins & Schwartz 1984; Hart 2010), officials in authoritarian regimes may thus be better able to mislead leaders as to their performance in office than are their peers in democratic regimes.

Despite of these concerns, recent literature in economics and political science that seeks to explain three decades of spectacular development in China strongly credits such top-down accountability institutions with creating incentives for sub-national officials to

implement policies to deliver on leaders' development targets and goals. Specifically, this literature attributes spectacular development to the Chinese Communist Party's cadre evaluation system. Scholars argue that leaders use this institution to create incentives for officials to implement policy in support of regime stability as well as leaders' own survival in office by linking rewards for officials to quantitative indicators of their performance in office.

Embedded within China's multi-level, geography-based administrative system, the cadre evaluation system allows leaders to create these incentives by linking rewards for officials serving at the next-lowest administrative level to indicators that are used to monitor officials' performance.⁶ For example, qualitative studies conducted at the grassroots describe how county and township leaders formally link the remuneration of grassroots officials to indicators like the growth rate of local tax receipts and the maintenance of local birth rates in line with national family planning targets (Edin 2003; Whiting 2001). Consensus is that leaders create the most powerful incentives through control over the *nomenklatura* (Burns 1989)—the list of offices over which leaders possess appointment power—and many scholars argue that leaders exercise of this power by linking promotion to performance vis-à-vis development indicators like GDP growth (Chen et al. 2005; Li & Zhou 2005).

Taking this argument even further, some scholars argue that China's geography-based administrative system assists leaders in generating information about the otherwise hidden actions of officials. In particular, these scholars argue the M-Form (multi-

⁶ China's administrative system formally consists of the following five separate levels, in descending order: national, provincial, municipal/prefectural, county, and township. Officials at each level are formally subordinate to leaders at the next highest administrative level.

divisional) organizational structure of China's administrative system, whereby control over administrative functions is grouped within geographically rather than functionally defined administrative units, allows leaders to compare the relative performance of officials serving in similar administrative units. Comparison of the relative performance of officials allows leaders to generate information and infer whether or not individual officials are shirking their duties if they fail to deliver on targets and goals (Qian and Xu 1993; Maskin, Qian and Xu 2000; Qian, Roland and Xu 2006; 2007).⁷ These scholars argue that leaders create tournament-like competition among officials by conditioning promotion on this information, thereby reducing officials' incentives to shirk. Altogether, this literature gives the impression that China's leaders have somehow *gotten it right* in addressing moral hazard inherent to rule through top-down accountability institutions while minimizing their need to employ risky, bottom-up institutions to monitor and motivate officials.

Yet while leaders may use the cadre evaluation system to generate information about officials' performance in office, substantive literature suggests that linking rewards to indicators also leads to the consequences implied by *Campbell's Law* of creating perverse incentives for officials to manipulate these same indicators. Studies suggest that such perverse incentives arise in a range of policy areas including rural development and taxation (Odgaard 1990; Cai 2000; Berntsen and Lu 2003; Tsai 2008), family planning (Merli 1998; Merli and Rafferty 2000), energy management (2001), legal and judicial

⁷ This is in contrast to the U-Form (unitary) organizational structure that characterized former Leninist regimes in the Soviet Union and Eastern Europe in which control and responsibility for administrative functions was grouped under highly specialized bureaucracies. To offer a parallel example to the one above, it was particularly difficult for leaders of the Soviet Union to compare the performance of an official serving in the Ministry of Railways to another official serving in the Ministry of Agriculture or Industry.

affairs (Minzner 2011), and environmental protection (Wang 2013). There has also been a notable focus of this literature has been on suspected manipulation of economic development indicators such as GDP growth (Rawski 2000; Koech & Wang 2012, Fernald, Malkin and Spiegel 2013; Wallace 2014).

Whatever the policy area in which a given indicator is used to monitor performance, such behavior on the part of sub-national officials represents an instance of moral hazard. For example, Bernstein and Lü (2003) discuss how the falsification of villager incomes by grassroots officials seeking to in meeting tax-collection targets from above contributed to near rebellion in many rural areas in the 1990s. Similarly, Cai (2004) and Guo (2009) argue that in order to impress superiors with their performance in fostering economic development and improve their chances at promotion, county-level officials engage in inefficient and cyclic spending to finance the construction of *zhengji* or political prestige projects. In doing so, these officials generate local debt that must ultimately be guaranteed by leaders at higher administrative levels. Whether or not the manipulation of indicators involves the falsification of data used to calculate indicators or the implementation of short-term and largely cosmetic policies to mislead leaders, the potential agency losses for leaders are substantial.

However, this literature has in many ways been peripheral to discussions of the cadre evaluation system. Although observers of Chinese politics have long suspected systematic manipulation of development indicators by officials in response to perverse incentives created by the cadre evaluation system, the difficulty of demonstrating an empirical relationship between the cadre evaluation system and evidence of manipulation has prevented discussion beyond anecdote and under-theorizing of its implications for

contemporary debates. The following section of this chapter thus focuses on introducing the case of a recent intervention to improve ambient air quality in Chongqing in which this proposition can be tested.

Poor Air Quality and Chongqing's Blue Sky Action Campaign

Recent episodes of abysmal air quality have brought both international and public attention to the issue of poor air quality in China's urban centers. Leaders have clear interests in addressing this issue as the costs resulting from poor air quality represent a substantial drag on China's development. For example, Matus et al. (2012) estimate social-welfare losses resulting from poor air quality in 2005 at USD 112 billion, or a 5% annual loss of GDP. The 2010 Global Burden of Disease Study further estimates poor air quality contributed to 1.2 million premature deaths and the loss of 25 million healthy years of life in 2010 alone (Yang et al. 2013).

Moreover, criticism resulting from these episodes reveals public concerns over the health effects poor air quality as well as skepticism of government reported air quality indicators (Wong 2011), and has heightened leaders' concerns about the adverse impact of poor air quality on social stability. The response of leaders to these crises has involved updating national air quality monitoring and reporting standards (Xinhua 2012), committing substantial resources to regional improvement efforts (Wee 2013), and exerting unprecedented pressure on officials across China to improve air quality by elevating the priority assigned to air quality indicators within the cadre evaluation system (MEP 2014).

However, this is not the first instance in which leaders have employed the cadre evaluation system to create incentives for officials to improve environmental conditions. Beginning with the Social and Economic Development Guideline of the national-level 11th Five Year Plan (2006-2010), leaders set development targets and goals with respect to improving environmental conditions.⁸ Furthermore, leaders explicitly defined these targets as ‘binding’ (*yueshu xing*), unequivocally indicating these targets had assumed veto status within the cadre evaluation system and that failure to deliver on environmental targets would be linked to promotion.⁹

Among the environmental targets set by national leaders and included in the Environmental Protection Guideline of the 11th FYP was that each of China’s 113 Key Environmental Protection municipalities was to deliver a given number of days per annum on which municipal air quality met the national Ministry of Environmental Protection’s (MEP) Grade I or Grade II standards.¹⁰ Days that met these standards—also known as *Blue Sky* days—were defined as days on which a jurisdiction recorded an Air Pollution Index (API) score of 100 or less. The target set by national leaders was such that sub-national officials were required to deliver an annual proportion of Blue Sky days in excess of 80% by the conclusion of the 11th FYP.

⁸ China’s Five Year Plans are a vestige of central planning from the pre-Reform period.

⁹ Environmental targets contained in the Social and Economic Development Guideline of the 11th FYP and that were defined as binding included a 10% reduction in SO₂ and COD emissions, 20% reduction in energy intensity per unit of GDP, 30% reduction in water use by industrial value added, 1.8% increase in forest coverage, 0.3% reduction in land under cultivation. Targets for the annual proportion of Blue Sky days for the 113 key environmental protection cities was not included in the Social and Economic Development Guideline but rather the Environmental Protection Guideline of the national-level 11th FYP. However, the target was included in the Social and Economic Development Guideline of Chongqing’s own municipal-level 11th FYP, reflecting the priority assigned to this target, specifically in Chongqing.

¹⁰ The 113 Key Environmental Protection municipalities are the most developed, most densely populated, and also the most politically important municipalities in China. They include all province-ranked municipalities, sub-provincial municipalities, and provincial capitals.

The province-ranked municipality of Chongqing located in China's western interior and indicated in the left panel of Figure A1 was at this time one of the poorest performers among the 113 key municipalities. In the MEP's 2004 Quantified Assessment of Urban Environmental Improvement, a performance evaluation conducted on an annual basis since 1989 and widely known as the *Chengkao*, Chongqing ranked below all other municipalities of comparable size and rank in terms of air quality. During the 10th FYP (2001-2005) period, Chongqing also recorded the highest per capita rate and highest per capita growth rate for environmental petitions related to poor air quality of China's four province-ranked municipalities.¹¹ Even the state-led media reported news of Chongqing's poor performance in improving air quality, indicating displeasure by national leaders with underperformance by officials in this politically and economically important municipality.

In response to pressure from national leaders and in apparent anticipation of environmental targets in the upcoming 11th FYP, municipal leaders launched Chongqing's *Blue Sky Action* campaign in early 2005. Consonant with targets in the 11th FYP, the focus of the campaign was to deliver an annual proportion of Blue Sky days in excess of 80% by the end of 2010.¹² The timeline of the campaign was extended in 2008 and the targeted, annual proportion of Blue Sky days was raised to 85% by the end of 2012. Notably, 2012 was the year in which the 18th CCP National Party Congress was to

¹¹ China's four province-ranked municipalities include Beijing, Tianjin, Shanghai, and Chongqing. Chongqing first became a province-ranked municipality in 1997 by an act of the State Council. The per capita rate of environmental petitions related to air quality is based upon the author's calculation using data on the total number of environmental petitions reported in annual China Environmental Protection Yearbooks scaled by the total urban population of each municipality taken from the China Statistical Yearbook from corresponding years.

¹² Municipal leaders later codified this target in the Social and Economic Development Guideline of the municipal-level 11th FYP.

be held and when promotions into the national party leadership are announced for officials serving in province-ranked jurisdictions.

The substance of the campaign is described in detail by in-depth policy studies (Li and Chan 2009; Zhang and Jiang 2009) as well as primary documents including the 2005 and 2008 *Blue Sky Action* Enforcement Plans. These enforcement plans were circulated by Chongqing's municipal-level government, describe a wide range of abatement policies aimed at reducing particulate emissions, and explicitly assign responsibility for implementation to municipal-level line departments and county-level jurisdictions, where both sets of units are subordinate to municipal leaders.¹³ Given China's geography-based administrative system and the dispersed nature of emissions sources for PM₁₀—one of the three monitored pollutants used to calculate API scores and the pollutant primarily responsible for days failing to qualify as Blue Sky days—the success of the campaign hinged on creating incentives for leading county officials—party secretaries and government executives—to implement abatement policies described in the enforcement plans.

For example, a baseline particulate inventory collected by Chongqing's municipal-level Environmental Protection Bureau (EPB) and the Chongqing Academy of Environmental Sciences identified dispersed emissions sources such as quarries, construction sites, small-scale factories, freight traffic, and residential use of coal for cooking and heating purposes as major sources of PM₁₀ emissions (Zhang and Jiang 2009). Abatement policies described in the enforcement plans and designed to limit

¹³ In addition, the preambles of both the 2005 and 2008 versions of the Blue Sky Action Enforcement Plan explicitly reference the Chongqing's poor ranking vis-à-vis the *Chengkao* and their desire to raise the municipality's ranking as a justification for the launch of the campaign.

emissions from these dispersed sources included ensuring the closure of quarries, the proper sheeting of trucks entering and leaving construction sites as well as the spraying down of construction sites to suppress dust, and the restrictions on operation of small-scale factories amongst others. The organization of county-level line departments such as traffic and construction bureaus under the command of leading county officials rather than municipal-level line departments meant successful implementation of such abatement policies was dependent upon the cooperation of county officials.

In order to create incentives for officials to direct costly resources and attention towards implementation, municipal leaders assigned targets vis-à-vis the county-level, annual proportion of Blue Sky days to county-level jurisdictions and used the local cadre evaluation system to link the promotion of officials to delivering on these targets. According to Zhang and Jiang (2009), county officials were compelled to personally sign responsibility contracts to deliver on these targets. Li and Chan (2009) further indicate that interview evidence suggests officials widely understood that failure to deliver would negatively affect promotion.

However, municipal leaders assigned targets and linked the promotion of county officials to delivering on these targets only in a subset of Chongqing's forty county-level jurisdictions.¹⁴ Only nine of forty jurisdictions were assigned targets in either version of the enforcement plan. The remaining thirty-one jurisdictions were neither explicitly named nor were they assigned targets in either version of the enforcement plan. Analysis of county-level Social and Economic Development Guidelines for the 11th and 12th FYPs

¹⁴ County-level jurisdictions that were assigned targets under the campaign include Yuzhong, Dadukou, Jiangbei, Shapingba, Jiulongpo, Nan'an, Ba'nian, Beibei, and Yubei Districts. Under the revised 2008 version, Beibei New District was also assigned targets.

confirms the difference in priority given to the annual proportion of Blue Sky days for these two sets of jurisdictions.¹⁵

Jurisdictions that were assigned targets are indicated in grey in the right panel of Figure A1. These jurisdictions are geographically concentrated as they make up Chongqing's urban core and immediate suburbs. Generally speaking, these jurisdictions are more developed, more densely populated as well as more polluted than jurisdictions that were not assigned targets. Nevertheless, this chapter hypothesizes that by linking the promotion of leading county-level officials in these jurisdictions to delivering on targets, leaders also created greater incentives for these officials to manipulate air quality data reported from their jurisdictions than their peers in jurisdictions that were not assigned targets. The following section describes the methods used to detect evidence of manipulation on a jurisdiction-by-jurisdiction basis as well as test the association between the cadre evaluation system and evidence of manipulation under the campaign.

Procedure to Detect Evidence of Manipulation and Classify Air Quality Data

In order to detect evidence of manipulation and classify samples of reported air quality data as either *suspect* or *non-suspect*, this study adopts a procedure used in environmental studies (Chen et al. 2012; Ghanem and Zhang 2013) for detecting an anomaly observed in the distributions of reported API scores. Recent environmental studies (Andrews 2008; Chen et al. 2012; Ghanem and Zhang 2013) argue that this anomaly indicates the manipulation of ambient air quality data by sub-national officials.

¹⁵ County-level 11th and 12th FYPs for jurisdictions that were assigned targets include the annual proportion of Blue Sky days as a development target, with 12th FYPs designating these targets as “binding”. In contrast, neither the 11th FYPs nor the 12th FYPs of jurisdictions that were not assigned targets make any mention of the annual proportion of Blue Sky days as a target.

Although these studies demonstrate convincing and widespread evidence of manipulation, they do not delve deeply into the institutional roots of manipulation. The procedure uses the McCrary (2008) test to detect a discontinuity in the relative density of daily API scores that occurs at the Blue Sky cutoff. Figure A2 in the appendix uses API scores reported from Beijing to illustrate this discontinuity.

The top panel in Figure A2 depicts a histogram constructed using daily, station-level API scores reported from Beijing from 2008 through 2011 whereas the bottom panel depicts the same data transformed into inferred PM10 concentrations that underlie reported API scores. The horizontal axis indicates reported API score (PM10 concentration) and the vertical axis indicates the relative density of observations. The vertical red line indicates the location of the *Blue Sky* cutoff. Observations with scores to the left of the cutoff (i.e. API score ≤ 100) qualify as Blue Sky days while observations to the right of the cutoff (i.e. API score > 100) do not.

The discontinuity is clearly visible in the sudden increase and then dramatic decrease in the relative density of observations with scores around the *Blue Sky* cutoff. While a similarly discontinuity is visible in the histogram of API scores at 50 in the top panel, this is an artifact of the piecewise linear construction of API scores from underlying PM10 concentrations. The bottom panel demonstrates how after transforming API scores back into their original values that this discontinuity disappears while the discontinuity at the Blue Sky cutoff persists.

Environmental studies (Andrews 2008; Chen et al. 2012; Ghanem and Zhang 2013) argue the anomaly located at the Blue Sky cutoff and the resulting discontinuity in the probability density function of the distribution are man-made and suggest that the

presence of this discontinuity indicates manipulation on the part of officials. Indeed, there is no reason to believe this anomaly is the result of a natural process as the cutoff is arbitrary with respect to any such process that might systematically influence scores. Furthermore, the increase in the relative density of observations immediately to the left of the cutoff has the practical effect of inflating the reported, annual proportion of Blue Sky days recorded for a given jurisdiction, systematically biasing this indicator upward.¹⁶

Visual analysis of data presented in Figure A2 alone suggests manipulation. However, both Chen et al. (2012) and Ghanem and Zhang (2013) propose a procedure to detect evidence of manipulation by testing for the presence of the resulting discontinuity using the McCrary (2008) test. The McCrary (2008) test is a diagnostic tool used in regression discontinuity (RD) designs to test for a violation of the assumption that agents cannot influence their own treatment status by manipulating the running variable used to assign treatment.¹⁷ In the context of RD designs, violation of this assumption is observable in a discontinuity in the relative density of observations that co-occurs with the known cutoff for assigning treatment. The test is thus designed to detect a discontinuity analogous to the one observed in reported air quality data and that is also

¹⁶ Andrews (2008) as well as Ghanem and Zhang (2013) provide discussion as to the mechanisms that sub-national officials can use to manipulate air quality. While neither explicitly states that the almost perfect sorting of scores around the cutoff indicates data falsification on the part of officials, both studies strongly hint at this possibility. This chapter offers three reasons to support this assertion: Firstly, emissions are inherently stochastic (Brännlund and Löfgren 1996) and leaders as well as officials cannot perfectly predict daily API scores. A second and related point is that neither leaders nor officials possess *ex ante* knowledge of the amount of resources that would be necessary to selectively enforce emissions controls in order to game the cutoff. Thirdly, emissions controls available to leaders as well as officials are imprecise and unlikely to generate a sharp, almost perfect discontinuity.

¹⁷To draw an example from educational testing, consider an abundance of test scores observed immediately above the cutoff necessary to receive a passing mark is taken as evidence that test takers can manipulate testing outcomes such as by lobbying administrators for a ‘mercy pass’.

indicative of manipulation. The McCrary (2008) test therefore offers a means to detect evidence of manipulation in samples of reported air quality data.

To briefly describe, the McCrary (2008) test itself involves two stages and generates two statistics, the later of which can be used for classification. These two stages involve (1) constructing a first-stage histogram by grouping observations into bins using a pre-selected bin size (b) and (2) estimating the height of the left and right limits of the probability density function at the cutoff with a local linear estimator that uses a pre-selected bandwidth (h).¹⁸ The first statistic is the log difference in the height of the left and right limits, that is to say the estimated height of the discontinuity, and the second is the Z-Score for the log difference in height.

While the log difference in height is sensitive to variance of the underlying distribution from which it is estimated, the corresponding Z-Score is standardized for this variance. Moreover, Z-Scores are similar to t-Statistics in that they can be used to generate p-values upon which to base the decision of whether to reject the null hypothesis that no discontinuity exists at the cutoff in a given sample. Z-Scores generated by running the McCrary (2008) test on samples of reported air quality data thus can provide evidence of manipulation by sub-national officials.

In terms of classifying samples of reported air quality data as *suspect* or *non-suspect*, where resulting Z-Scores fall below the lower bound of the 95% CI of a standard normal curve ($Z\text{-Score} < -1.960$) or an alternatively specified significance threshold, researchers can reject the null hypothesis and conclude a discontinuity is present in the

¹⁸ McCrary (2008) advises that the bandwidth to bin size ratio, or alpha, should always exceed ten. Results and analysis presented in this study are equal to or exceed ten.

density distribution of the sample. This implies that the sample should be classified as *suspect*. Alternatively, researchers should fail to reject the null when Z-Scores are above this lower bound. The sample should then be classified as *non-suspect*. Z-Scores thus offer a basis for generating an easy to interpret, binary dependent variable.¹⁹

Data Description and Classifying County-Specific Samples

Using the procedure described above as well as available county-level API scores reported from each of Chongqing's forty county-level jurisdictions during the *Blue Sky Action* campaign, this section describes the construction the binary dependent variable used in subsequent analysis. Data were collected from a public database accessible from the official website of the Chongqing Environmental Monitoring Center (CQEMC). The database contains daily, county-level API scores reported for each of Chongqing's forty county-level jurisdictions with coverage beginning in mid-2008. Summary statistics for county-level data grouped by whether the reporting jurisdiction was assigned targets or not can be seen below in Table A1 in the appendix.²⁰

Figure A3 in the appendix presents histograms constructed from two separate and mutually exclusive samples of observations (i.e. daily, county-level API scores) drawn from the database. The histogram in the left panel is constructed using the sample of

¹⁹ It is important to note that the Z-Score does not directly indicate the magnitude of the estimated discontinuity. Although a larger discontinuity will generate a smaller corresponding Z-Score, this study does not claim to make any inferences about the magnitude on the basis of Z-Scores.

²⁰ The CQEMC is the government-sponsored organization responsible for collecting and reporting all environmental monitoring data within the municipality. The CQEMC website and database can be accessed by the following URL: <http://www.cqemc.cn/> Limited temporal coverage of the database is a potential issue. The CQEMC database does not includes daily, county-level API Scores reported prior to the database being made public in mid-2008. This prevents subsequent analysis from using air quality data reported during the pre-campaign period to establish a baseline vis-à-vis the dependent variable. However, establishing such a baseline is unnecessary so long as the assumption holds that any discontinuity at the cutoff is man-made and not the consequence of any natural and unobserved process.

observations reported from the nine jurisdictions that were assigned targets. Alternatively, the histogram in the right panel is constructed using the sample reported from the thirty-one jurisdictions that were not assigned targets. Similar to the top panel in Figure A2, the horizontal axes of both panels indicate reported API score. The vertical axes indicate the relative density of observations binned according to reported score. The vertical red lines indicate the location of the Blue Sky cutoff.

Visual analysis of the left panel of Figure A3 indicates clear evidence of a discontinuity that is indicative of manipulation in the sample of observations reported from jurisdictions that were assigned targets. In contrast, the right panel offers limited evidence of a discontinuity in the sample of observations reported from jurisdictions that were not assigned targets. Moreover, the noticeable height of the discontinuity in the left panel suggests a strong association between the target assignment status and jurisdictions reporting suspect data.

In order to generate county-level Z-Scores and to classify samples of observations as either suspect or non-suspect, I divide available observations into 40 county-specific samples and conduct the McCrary (2008) test separately on each sample where bin-size is set to one. Figure A4 in the appendix graphically depicts the distributions of resulting Z-Scores for multiple levels of *alpha*.

Figure A4 presents plots of both the probability density functions (PDFs) and corresponding empirical cumulative density functions (ECDFs) for resulting, county-level Z-Scores, divided by target assignment status. The top row contains plots that compare the PDFs of resulting Z-Scores by target assignment status and the bottom row contains plots that compare the corresponding ECDFs. Plots are stacked according to the

alpha used to conduct the McCrary (2008) test and to generate scores. The horizontal axes of all plots indicate county-level Z-Score and the vertical axes indicate the relative density of county-level Z-Scores. The PDFs and ECDFs of scores from jurisdictions that were assigned targets are indicated in red while the PDFs and ECDFs of scores from jurisdictions that were not assigned targets are indicated in blue. The dotted vertical lines (Z-Score = -2) approximate the 95% CI threshold whereby Z-Scores located to the left of the threshold can be classified as suspect and scores to the right classified as non-suspect.

Comparison of the PDFs and ECDFs by target assignment status indicates greater evidence of manipulation in samples reported from jurisdictions that were assigned targets versus those that were not assigned targets. For all levels of *alpha*, Z-Scores from jurisdictions that were assigned targets are all less than the 95% CI threshold. This indicates greater evidence of manipulation and suggests that jurisdictions that were assigned targets uniformly reported data that can be classified as suspect. Kolmogorov-Smirnov and Anderson-Darling tests further suggest that regardless of the *alpha* used to run the McCrary (2008) test, the null hypothesis of equivalence should be rejected ($p\text{-value} < 0.001$). Given the abstract nature of Z-Scores and in the interest of simplifying results and analysis, the chapter classifies county-specific samples according to their Z-Scores generated using an *alpha* of fifteen as either suspect or non-suspect, the binary dependent variable.

Results and Analysis

This section presents results from difference-in-proportions tests for the proportion of county-specific samples of API scores classified as suspect according to the

underlying target assignment status of the jurisdiction from which these samples were reported. Results are presented in Table A2 in the appendix for a variety of thresholds.

Table A2 contains three separate, two-by-two sub-tables where each sub-table reports the proportion of samples classified as suspect or non-suspect according to the classification threshold indicated in the far left column. Within sub-tables, rows indicate the target assignment status of the jurisdiction that reported the sample and columns indicate the classification status. Each cell of the two-by-two sub-table reports the count and row-percentages. Lastly, the far right column contains the estimated difference-in-proportions by target assignment status as well as p-values.

Results presented in Table A2 support the hypothesized relationship. County-level jurisdictions that were assigned targets under the *Blue Sky Action* campaign report samples that are uniformly classified as suspect at the 95% CI threshold. In contrast, only 38.7% of jurisdictions that were not assigned targets are classified as suspect report samples classified as suspect. The difference-in-proportions is substantial and grows with the application of higher thresholds. Regardless of the threshold used, the difference-in-proportions is significant (p-value < 0.001).

Multivariate Regression Results

It is possible that results from bivariate difference-in-proportions tests are subject to omitted variable bias. Unobserved factors correlated with a county-level jurisdiction's target assignment status such as population density, level of development, and level of pollution could independently affect the incentives of leading county officials to manipulate and confound results. Moreover, there may be concerns that target assignment

status is endogenous to the propensity of officials to manipulate and thereby report suspect data. I use multivariate and IV regression analysis to put these concerns to rest.

If officials derive incentives to manipulate from outside the cadre evaluation system, then omitted variable bias may confound results presented in Table A2. For example, it may be argued that officials in more densely populated jurisdictions are more likely to face protest or unrest stemming from poor environmental conditions as higher population density lowers the costs of organizing collective action. It may also be that officials in more developed jurisdictions face greater pressure from wealthier as well as presumably more mobile, tax-paying residents to demonstrate improvements in conditions. Lastly, it may be that leaders simply assigned targets to the most polluted jurisdictions and that the apparent effect of target assignment status is instead attributable to a jurisdiction's underlying level of pollution.

Summary statistics presented in Table A3 in the appendix suggest that these concerns are not unjustified: County-level jurisdictions that were assigned targets were more densely populated, more developed, and more polluted than jurisdictions that were not assigned targets. Table A3 reports summary statistics for annual, county-level observations of population density in 10,000 persons per sq. km and summary statistics for annual, county-level observations of per capita GDP in 10,000 yuan adjusted for inflation to year 2000 values. Table 2 also reports summary statistics for annual, county-level (average) concentration of monitored particulates used to construct API scores—

particulate matter of ten-micrometer radial diameter (PM_{10}), sulfur dioxide (SO_2), and nitrogen dioxide (NO_2)—in milligrams per cubic meter.²¹

While summary statistics indicate that overlap in the distributions of these variables, they also indicate jurisdictions that were assigned targets were indeed more densely populated, more developed, and more polluted. Balance tests suggest that leaders' decision to assign targets to certain jurisdictions was not balanced on these factors. Imbalance in terms of population density, level of development, and level of pollution further suggests that tests should at the very least control for these potential confounds by including the above variables as covariates.

Using the binary dependent variable of classification status, this section thus models the hypothesized relationship using logistic regression and controls for these potential confounds. However, the small sample size ($n = 40$) as well as near perfect association between the classification status of county-specific samples of air quality data as suspect and target assignment status demonstrated in Table A2 raises non-trivial issues in maximum likelihood estimation of non-linear models. To be specific, maximum likelihood estimation of logistic regression models using small samples with imbalanced and highly predictive variables often results in complete or quasi-complete separation.

²¹ Data were collected for each year of the *Blue Sky Action* campaign (2005-2012). Data used to calculate population density and per capita GDP were collected from Chongqing Statistical Yearbooks while data on the concentration of the three monitored pollutants was collected from Municipal State of the Environment reports published on an annual basis by the Chongqing Environmental Protection Bureau (EPB). Per capita GDP was inflation adjusted to its year 2000 nominal value using a province-level deflator from China Data Online. The reduced number of annual, county-level observation for pollutant concentrations is due to the omission of this data from the 2005 and 2006 State of the Environment reports.

Separation produces infinite parameter values that result in the failure of non-linear models to converge during estimation.²²

Subsequent analyses use Firth logistic regression to overcome the issue of separation. Heinze and Schemper (2002) propose Firth logistic regression as a method to address separation when estimating logistic regression models using small samples with imbalanced and highly predictive variables. Firth logistic regression corrects for small sample bias by means of penalized maximum likelihood estimation based upon a procedure proposed by Firth (1993) to reduce small-sample bias in maximum likelihood estimates more generally. Heinze and Schemper (2002) apply this idea to logistic regression. Their estimation procedure addresses separation, produces finite parameter estimates, and allows for model convergence. Their estimation procedure is implemented in the R programming environment by way of the *logistf* R-package (Heinze and Ploner 2004).

Covariate values were generated for population density, per capita GDP, and pollutant concentrations by calculating the mean of (available) annual, county-level observations of each variable. Table A4 in the appendix presents the results from regression of classification status on target assignment status and covariates where population density and per capita GDP are both logged and pollutant concentrations are scaled to micrograms per cubic meter of atmosphere for ease of presentation. All models report estimated coefficients rather than odds ratios.

²² A basic logistic regression of classification status on target assignment status produces a coefficient with an estimated variance in excess of 5000, indicating separation.

Model 1 shows baseline results while Models 2 through 4 include covariates for population density, per capita GDP, and pollutant concentrations. The estimated coefficients for target assignment status are positive, large, and significant (p-value < 0.01) in all models. This supports the hypothesized relationship between institutional incentives and manipulation as target assignment status is positively related to the propensity of jurisdictions to report suspect data.

In contrast, the estimated coefficients for all covariates fail to reach standard levels of significance in all models. This suggests that although target assignment is imbalanced on population density, level of development, and level of pollution, these factors do not appear to have an independent effect on officials' incentives to manipulate air quality data outside of institutions. This finding is consistent with the argument of O'Brien and Li (1999) that the cadre evaluation system not only serves to direct officials' attention towards the implementation of leaders' desired policies, but also serves to insulate officials from outside pressures.

Instrumental Variables (IV) Regression Results

One concern about these results is that in attempting to bias and inflate the annual proportion of Blue Sky days recorded in their respective jurisdictions, officials also manipulate data on the concentrations of the three monitored pollutants used to calculate API scores. Presumably, this should systematically bias pollutant concentrations downward in jurisdictions that were assigned targets and bias the estimated coefficients for pollutant concentrations. Manipulation of PM₁₀ concentration data is of particular

concern as PM₁₀ was the monitored pollutant primarily responsible for failure to report API scores that met the Blue Sky cutoff.

Model 5 instruments for PM₁₀ concentration in order to address this concern. First stage regression results as well as a detailed explanation of the logic for this approach are provided in the appendix. To briefly summarize, Model 5 instruments for PM₁₀ concentration using the ratio of annual, county-level gross output of the construction sector to GDP as a proxy for the construction intensity. According to the results from the baseline particulate inventory collected at the start of the campaign, construction and building activity was a leading source of PM₁₀ emissions. By regressing reported PM₁₀ concentrations on PM₁₀ sources and using predicted values in the second stage regression presented in Model 5, this should help allay concerns about unobserved bias. Similar to Model 4, the estimated coefficient for instrumented PM₁₀ concentration in Model 5 for does not reach standard levels of significance and further supports the assertion that incentives for officials to report suspect data flow through the cadre evaluation system.

Another concern regards the potential endogeneity of target assignment status. If leaders were more likely to assign targets to jurisdictions administered by officials who were more likely to manipulate, then results presented thus far may be misleading. While it seems unlikely that leaders can *ex ante* predict which officials are more prone to manipulate and thus which jurisdictions are more likely to report suspect data, perhaps factors absent from the model but correlated with officials' propensity to manipulate influenced leaders' decision regarding target assignment.

Models 6 and Models 7 in Table A3 instrument for target assignment status using two, binary instruments in order to addresses these concerns: First, whether a county-

level jurisdiction contains a *nationally regulated* air quality monitoring station and second, whether a jurisdiction is located within the river valley that contains the urban core of the municipality. The logic underlying the use of these instruments and their suitability vis-à-vis the exclusion restriction follows from the logic behind municipal leaders' assignment of targets under the campaign.

Facing pressure to improve municipal air quality, municipal leaders sought to create incentives for county officials to implement abatement policies. However, the performance of municipal leaders in achieving improved municipal air quality—measured by the annual proportion of Blue Sky days recorded for the municipality as a whole—was based only on air quality data collected from so-called *nationally regulated* air quality monitoring stations. These stations are maintained by local environmental officials, but are regulated in the sense that their placement and operation is subject to a national environmental regulation that effectively fixes their location in the more densely populated, more developed, and more polluted jurisdictions in Chongqing's urban core and immediate suburbs.²³

This regulation also effectively links municipal leaders' performance vis-à-vis the annual proportion of Blue Sky days recorded for the municipality as a whole to improvements in air quality in specific, county-level jurisdictions that contain such stations but not in jurisdictions where such stations are absent. The placement of *nationally regulated* stations by regulation therefore offers a source of variation that

²³ The specific regulation is HJB 193-2005. This regulation effectively fixes the location of monitoring stations in more densely populated, more developed, and more polluted jurisdictions in Chongqing's urban core so that municipal leaders cannot 'improve' air quality by shifting these stations to outlying jurisdictions with lower levels of pollution.

predicts target assignment status, plausibly meets the exclusion restriction, and is also (presumably) exogenous to county officials' propensity to manipulate.

In addition, Chongqing's geographic topology also played an important role in determining the target assignment status of jurisdictions. Chongqing is known as the "Mountain City" with its urban core situated in the river valley that contains the confluence of the Yangtze and Jialing rivers. This valley and the urban core of Chongqing, are separated from much of the immediate suburbs as well as outlying jurisdictions by the line of hills that make up the Huaying Mountain range to its east and west that rise between 800-1000 meters above sea level.²⁴ This mountainous geography traps pollutants in the valley. Moreover, spillovers between neighboring jurisdictions located within the valley mean that in order to show improvements in air quality in jurisdictions that contained *nationally regulated* stations, municipal leaders also needed to assign targets to neighboring jurisdictions within the valley where such stations were absent. First stage regression results as well as a discussion of these instruments are provided in the supplementary appendix.

Model 6 and Model 7 in Table A4 report the second stage regression results after instrumenting for the presence of *nationally regulated* stations and geography. Model 6 replicates the baseline regression reported in Model 1. Model 7 replicates the regression including covariates reported in Model 4. The estimated coefficients for target assignment status in both Model 6 and 7 are in the predicted direction, are substantively similar to the estimated coefficients in the non-instrumented models, and are significant

²⁴ The average elevation of the municipality is 237 meters above sea level.

(p -value < 0.001). Again, these results support the hypothesized relationship between institutional incentives and manipulation.

Transforming the estimated coefficients into odds ratios provides a basis to interpret the substantive effect of target assignment status on the likelihood of jurisdictions reporting suspect data. Given minimal evidence that any of the added covariates explain variation in the propensity for jurisdictions to report suspect data and based upon the minimum Bayesian Information Criteria (BIC) for Model 1, this chapter uses coefficients from Model 1 to calculate odds ratios. Based upon Model 1, *county-level jurisdictions that were assigned targets were 19.01 times more likely to report suspect data than jurisdictions that were not assigned targets.*

Robustness Checks and Sensitivity Analysis

Results presented in the previous section support the hypothesized relationship between the cadre evaluation system and incentives to manipulate. However, there are a number of reasons why readers might question these results. First, readers might question whether results reported above are a product of the choice of models or are sensitive to changes in the significance threshold used to distinguish suspect from non-suspect samples. Second, readers familiar with contemporary Chinese politics might question whether evidence of manipulation during the Blue Sky Action can at all be separated from the context in which it occurred: Chongqing during the tenure in office of the now-disgraced CCP Municipal Secretary and former Politburo-member Bo Xilai. Third, the logic of target assignment described suggests that municipal leaders had incentives to remain complicit and perhaps even collude with leading county officials in jurisdictions

that contained *nationally regulated* stations, thus raising questions about interpretation of these results offered thus far. In order to address these concerns, this section offers sensitivity analysis as well as a series of robustness checks and discussion.

Alternative Model Selection and Sensitivity Analysis

A first set of concerns is empirical in nature: Are results presented above an artifact of model selection? Alternatively, are results an artifact of the specification of the significance threshold used to code the dependent variable from county-level Z-Scores or the bandwidth selection used to calculate underlying Z-Scores?

In order to address the question of model selection, Table A5 in the appendix reports OLS regression results for Models 1 through Model 8 from Table A4. Results are consistent across all models with a positive and significant association between target assignment status and jurisdictions reporting suspect data.

In order to jointly address concerns regarding model selection and specification of the threshold used to code the dependent variable, Table A6 in the appendix reports OLS regression results for Model 1 through Model 8 from Table A4 using underlying Z-Scores rather than the binary dependent variable. In interpreting results presented in Table A6, it is important to note that the sign of all coefficients should reverse as negative Z-Scores reflect evidence of manipulation. Results reported in Table A6 are consistent with results reported in Table A4, with a negative association between target assignment status and Z-Scores.

To address further concerns about the sensitivity of results to specification of the threshold used to code binary dependent variable *as well as* the bandwidth used to

generate the underlying county-level Z-Scores, Table A7 in the appendix reports the coefficients on target assignment status rerunning Model 1 using a series of different specifications for the binary dependent variable. Coefficients reported in each column are estimated based upon county-level Z-Scores generated using the listed bandwidth to bin-size ratio (i.e. *alpha*) and where rows indicate the threshold then used to classify data as suspect or non-suspect.

The estimated coefficients for target assignment status are significant and in the predicted direction for all specifications of the binary dependent variable. This offers strong evidence that results presented in the prior section are neither a product of model selection nor specification of the dependent variable.

Robustness Check #1: Tenure of Bo Xilai

A second concern is substantive in nature: Is what appears to be clear evidence of manipulation a product of the context in which it occurred rather than of institutional incentives more generally? The period of observation in this study overlaps with the tenure in office of now-disgraced former Politburo member Bo Xilai as Chongqing's CCP Municipal Secretary. Bo was appointed to this office in November 2007 and held office until he was purged from the CCP Politburo and removed from all party and state posts on 15 March 2012.²⁵ Information revealed since Bo was purged the leadership suggests that Bo's term in office was characterized by widespread corruption and the

²⁵ Bo's purge was catalyzed by the attempted defection of his former lieutenant and Chongqing's vice-mayor Wang Lijun at the U.S. Consulate in Chengdu, Sichuan on 6 February 2012. For further details, the reader can consult the now voluminous body of largely speculative and journalistic accounts regarding the political reasons for Bo's removal. To my knowledge, none refer to the manipulation of air quality data.

abuse of power. Simply, this suggests results may be the product of a particularly corrupt municipal leadership.

Ruling out the idiosyncratic effects of Bo's term in office is difficult due to data limitations. Although the Blue Sky Action campaign was initiated two years prior to Bo's appointment to office, county-level data is not available for this earlier period. However, there is a simple test to rule out the alternative explanation that manipulation was solely a product of Bo's corrupt municipal leadership that involves rerunning the McCrary (2008) test on the subset of daily, county-level API scores reported after Bo's removal. If manipulation was solely a function of Bo's leadership, then evidence of manipulation should disappear in data reported from jurisdictions that were assigned targets following Bo's removal.

Figure A5 in Appendix A depicts the distribution of the subset of daily, county-level API scores reported from jurisdictions that were assigned from 16 March 2012. The dramatic discontinuity in the relative density of observations around the Blue Sky cutoff appears to persist despite Bo's removal from office, and the discontinuity is significant ($p\text{-value} < 0.001$).

It is thus possible to rule out the alternative explanation that manipulation was simply a product of Bo's tenure in office. One speculative critique of this interpretation is that upon appointment to office, Bo may have altered institutional incentives for leading, county-level officials in some unobservable way such that they derived incentives to manipulate where none had existed before. The 2008 revision of the *Blue Sky Action* campaign's enforcement plan did occur almost immediately after Bo appointment. However, comparison of the original 2005 version and the revised 2008 version of the

enforcement plan reveals no observable changes vis-à-vis the institutions used to manage leading, county-level officials. Evidence of manipulation even after Bo's removal from office suggests that to the degree that Bo altered institutional incentives in even an unobservable way, those incentives were robust to removal from office and the end of his political career.

Robustness Check #2: Complicity and Collusion

A final concern is both substantive and theoretical in nature. As noted above, the logic of target assignment described above in the previous section suggests that municipal leaders had incentives to remain complicit and perhaps even collude with leading county officials to manipulate in jurisdictions that contained *nationally regulated* stations. This in turn suggests an alternative interpretation of the seemingly at odds with the interpretation offered thus far: Rather than indicating that officials faced perverse incentives to manipulate, target assignment status instead indicates that officials perhaps enjoyed the tacit approval of leaders.

To be clear, leaders did stand to benefit from manipulation in jurisdictions that contained *nationally regulated* monitoring stations. Manipulation in these jurisdictions not only served to lower county-level API scores, but also municipal-level API scores used to calculate the annual proportion of Blue Sky days for the entire municipality and upon which municipal leaders were evaluated by their own superiors. To the degree that manipulation lowered municipal-level API scores on days where the municipality would otherwise report a score that did not meet the cutoff and that failed to qualify as a Blue Sky day, leaders had incentives to overlook manipulation by officials on such days.

Substantive literature on central-local relations in China not only suggests that such behavior is common among leaders at sub-national administrative levels, but that such leaders often actively collude with subordinate officials to mislead their own superiors at higher administrative levels (Zhou 2010). Indeed, according to Li and Chan (2009) implementation of abatement policies under the *Blue Sky Action* campaign by leading county officials was to be coordinated by a vice-mayor. This suggests municipal leaders not only possessed incentives but also the means to coordinate manipulation by officials when doing so was to leaders' benefit.

Although it is not possible to rule out complicity or collusion with officials on the part of leaders, it is possible to test whether jurisdictions that were assigned targets report suspect data *even when leaders do not benefit* from manipulation by officials. Demonstrating such evidence validates the interpretation offered thus far by showing that whether officials engage in manipulation is not strictly contingent on the tacit approval of leaders. Moreover, this also suggests that regardless of tacit approval or active efforts to coordinate the manipulation when it is to leaders' benefit, officials are still able to do so when it is solely for their own benefit.

In order to test whether jurisdictions that were assigned targets report suspect data even when leaders do not benefit, this section examines the distribution of daily, county-level API scores reported from jurisdictions that were assigned targets on days when the recorded, municipal-level API score exceeded the Blue Sky cutoff (i.e. municipal-level API score > 100). By definition, leaders do not benefit from manipulation on such days: While manipulation may improve daily, municipal-level API scores, it does not increase the annual proportion of Blue Sky days recorded for the municipality. Consequently,

evidence of manipulation—observable in a discontinuity in the relative density of daily, county-level API scores at the Blue Sky cutoff—indicates manipulation even when leaders do not benefit.

Figure A6 in the appendix depicts the density plot generated by running the McCrary (2008) test on the subset of daily, county-level API scores reported from jurisdictions that were assigned targets and on days when the recorded, municipal-level API score was between 111 and 120.²⁶ The horizontal axis indicates county-level API score. The vertical axis indicates the relative density of observations (i.e. daily, county-level API scores) binned according to reported score where bin-size equals one. The vertical red line indicates the location of the Blue Sky cutoff.

The discontinuity in the density plot presented in Figure A6 indicates that jurisdictions that were assigned targets appear to report suspect data even when leaders do not benefit. Despite restrictive conditions in the selection of observations used to plot the first-stage histogram and conduct the McCrary (2008) test, results indicate the presence of the predicted discontinuity at the cutoff (p-value < 0.001). This supports the hypothesized relationship and the interpretation that the cadre evaluation system creates perverse incentives for officials to manipulate indicators.

Discussion and Implications

To briefly summarize, this chapter argues that as argued by *Campbell's Law*, when the leaders link rewards like promotion to indicators of officials' performance in

²⁶ By examining county-level data reported on days when the municipality missed the *Blue Sky* cutoff by a substantial margin, I weigh against the possibility that I am mistaken on the ex post nature of falsification.

office through the cadre evaluation system, they also create perverse incentives for officials to manipulate indicators. This chapter clearly demonstrates the relationship between the cadre evaluation system and evidence of manipulation in the context of Chongqing's *Blue Sky Action* campaign: County-level jurisdictions that were assigned air quality improvement targets and where leading county officials' promotion was linked to delivering on targets were at least nineteen times more likely to report data indicative of manipulation aimed at inflating the reported, annual proportion of Blue Sky days than jurisdictions where leaders did not establish this link. Moreover, these results are robust to alternative specifications and a number of alternative interpretations.

Beyond the contributions listed earlier, results and analyses presented in this chapter ultimately raise questions for a core debate in current literature on Chinese politics about the determinants of political promotion in contemporary China. Participants in this debate are roughly divided into two camps: Those who argue that leaders link the promotion of sub-national officials to their performance vis-à-vis development indicators such as GDP growth and those who argue promotion is instead based upon factors such as an officials' factional affiliation. Clear evidence of manipulation in response to institutional incentives raises puzzles for both sides of this debate: For scholars who claim that factional affiliation matters, does evidence of manipulation not indicate that officials themselves believe that performance ultimately does matter for promotion? For scholars who claim that performance matters, how can leaders promote officials based upon their performance when the indicators of their performance are inherently suspect? The final chapter of this dissertation attempts to offer answers to these questions.

CHAPTER 3: AIR QUALITY AND AGENDA SETTING

*Please, everyone pay close attention to the following new word: PM2.5.*²⁷

What can recent episodes of abysmal air quality and resulting public skepticism over the reliability of government reported ambient air quality information teach us about who sets the public agenda in contemporary China? Moreover, what can these episodes teach us about the ability of the public agenda to set the media agenda as well as influence the policy agenda under authoritarianism?

This chapter argues that recent episodes of abysmal air quality offer a unique opportunity to observe the process by which actors outside of the party and state such as well-known celebrities, entrepreneurs, and public intellectuals play a growing role in setting the public agenda in contemporary China. This chapter further argues that despite leaders' control over what issues are placed on the media agenda, such actors nevertheless appear to be able to set the media agenda by exerting pressure on China's highly competitive, commercial media to report on issues of public concern. Moreover, this chapter argues and demonstrates evidence that these actors also appear able to influence the policy agenda, pressuring leaders to respond to both the public and media agendas despite leaders' monopoly agenda control and the absence of formal, bottom-up accountability institutions like competitive and multi-party elections.

²⁷ This message was posted by Pan Shiyi to his personal Sina Weibo weblog on 5 November 2011. Pan is a real-estate developer and popular online commentator with over sixteen million followers on Sina Weibo.

Specifically, this chapter demonstrates the ability such actors to set the public and media agendas by focusing on widespread public concern and skepticism over the reliability of government reported ambient air quality information that unexpectedly emerged in late 2011. This chapter argues that this episode marks an important change in the public agenda with a sudden and dramatic increase in public attention to the issue of poor air quality in China's rapidly expanding urban centers. Moreover, the episode saw the emergence of clearly articulated demands for leaders to reform ambient air quality monitoring and reporting standards as well as to require local officials to publicly report information on the concentration of fine particulate matter, also known as PM2.5. Facing substantial pressure from both the public and questions in the commercial media, leaders responded by announcing their intent to reform monitoring and reporting standards with proposed reforms specifically reflecting demands for the public reporting of information on PM2.5 concentrations.

In order to demonstrate evidence of the public agenda setting the media agenda and even influencing the policy agenda, this chapter makes use of data and methods novel to the study of agenda setting. In particular, the chapter uses data on the daily volume of micro-blog posts that explicitly mention air quality made by users of China's most popular micro-blogging platform, Sina Weibo, as well as methods for detecting structural breaks in time-series data to demonstrate a sudden and sustained increase in the level of public attention to the issue of poor air quality in mid-October 2011. Treating this change as indicative of a change in the public agenda, the chapter also uses these methods and data to model daily media coverage related to air quality to demonstrate a similarly sudden and sustained increase in media attention that indicates change in the media

agenda. The chapter then demonstrates that the observed increase in media attention lags the increase in public attention, occurring at the very end of October 2011. Lastly, the chapter demonstrates that these changes in both the public and media agendas precede initial public announcements made by national-level leaders of their intent to reform existing monitoring and reporting standards to address public concerns.

In taking this methodological approach, this chapter seeks to establish temporal precedence of changes in the public, media, and policy agendas implied by recent changes to agenda setting theory. Agenda setting theory is causal in nature and implies a clear causal ordering of change in agendas. However, empirical studies of propositions drawn from agenda setting theory often fail to establish unambiguous causal order of change implied by theory (Rogers and Dearing 1988; Kosicki 1993). While this chapter is not the first to overcome this challenge, it is nonetheless the first to do so using a methodological approach which treats estimating the date of change in public and media agendas as a problem of detecting structural breaks in time-series data. This innovation represents a methodological contribution to broader research on agenda setting theory.

This chapter also seeks to make a contribution to the application of agenda setting theory in non-democratic contexts. Although agenda setting theory has been applied to numerous contexts to explain the relationship between the public, media, and policy agendas, the understanding of who sets the public agenda and the potential for the public agenda to influence the policy agenda in non-democratic contexts is substantially altered from the context of democratic politics. Whereas traditional agenda setting theory privileges the role of the media agenda in setting the public agenda, leaders' control over the media agenda through selective censorship and control of content as well as the

absence of bottom-up accountability institutions through which the public can plausibly hold leaders to account in non-democracies suggests an interpretation that privileges the role of leaders in setting the public agenda.

Yet substantive literature on agenda setting in contemporary China suggests that while leaders maintain monopoly control over the policy agenda and substantial control over what issues are placed on the media agenda, they are not entirely insulated from the influence of the public agenda. In particular, this chapter demonstrates evidence of a non-institutional channel through which actors traditionally left outside of the policy-making process can influence leaders and shift the policy agenda in their favor by first setting the public agenda. While responsiveness to the public agenda on the part of leaders neither implies their accountability to the public nor that public demands fundamentally challenge leaders' monopoly over the policy agenda, it does support the observation of an increasingly pluralized policy-making process despite the absence of any formal moves by leaders towards political liberalization (Wang 2008; Mertha 2009).

The remainder of the chapter is as follows: First, the chapter provides a review of agenda setting theory and its application to non-democracies, focusing specifically on contemporary China; Second, it provides an account of the crisis mentioned above that occurred during the final months of 2011 and then develops hypotheses regarding the order of change in the public, media, and policy agendas with respect to the issue of poor air quality; Third, the chapter offers a description of data and methods used to establish the order of agenda change; Fourth, the chapter offers its results and analysis. The chapter concludes with a discussion of findings and their connections to substantive literature.

Agenda Setting under Authoritarianism

Who sets the public agenda? Agenda setting theory privileges the role of the media in setting the public agenda in open societies (McCombs and Shaw 1972; McCombs 2004). The underlying argument is that the media draws public attention to specific issues by placing them on the media agenda, thus creating greater public awareness, and leading the public to perceive these issues as more important than alternatives. In the presence of bottom-up accountability institutions like competitive and multi-party elections through which the public can hold leaders to account, this implies the public agenda will influence the policy agenda (Page and Shapiro 1983).

However, this understanding of who sets the public agenda and its potential to influence the policy agenda hinges upon conditions underlying democratic politics. In non-democracies where bottom-up accountability institutions through which the public can plausibly hold leaders to account are absent and where leaders can directly control what issues are placed on the media agenda through selective censorship and control of media content, the relationship between agendas posited by traditional agenda setting theory is open to question. In particular, there is reason to question whether congruence between agendas is due to a bottom-up effect of the public or media agendas on the policy agenda or a top-down process with leaders setting the public and media agendas to reflect their own desired policy agenda.

Contemporary China would seem to offer a clear case in which leaders do indeed exercise significant control over what issues are placed on the media agenda and actively use these controls to set the public agenda. For example, Chan (2007) argues that efforts by Chinese leaders to guide public opinion through propaganda work and reporting by

the state-led media have shifted since the early 1990s towards attempting to shape ‘what people think about’ as opposed to ‘what people think’. Nevertheless, substantive literature suggests that explanations of who sets the public agenda that ignore the impact of the Internet as well as media commercialization risk overlooking critical changes in what actors besides leaders can set the public agenda in contemporary China as well as their ability to use the public agenda to pressure leaders and influence the policy agenda.

Scholars writing on the impact of the Internet point to its role as a quasi-public sphere in which members of the public can raise and openly discuss issues of public concern (Lagerkvist 2005; Qian and Bandurski 2011). In making this argument, scholars like Xiao (2011a; 2011b) focus on the emergence and the role of online ‘opinion leaders’ such as well-known celebrities, entrepreneurs, and public intellectuals in placing issues on the public agenda. Given that these actors are identifiable, can reach large public audiences, can quickly and widely disseminate information, and have incentives to cultivate a positive reputation with the public, they play an important role in placing issues of personal concern and that resonate with the broader public onto the public agenda. In doing so, these actors draw public attention and creating greater awareness of these issues, thereby setting the public agenda independently of the traditional media and leaders.

In addition to the emergence of actors outside of the party and state who can set the public agenda, scholars writing in the substantive literature also point to the impact of media commercialization in discussing how the public agenda may also undermine leaders’ controls over the media agenda. Although leaders possess a range of mechanisms with which to restrict what issues traditional media outlets place on the media agenda, the

need for media outlets to compete over public audiences and advertising revenues creates countervailing incentives (Shirk 2011; Qian and Bandurski 2011). While media outlets can be heavily fined or find their operating licenses revoked for breaking restrictions on reporting about issues that leaders would rather see suppressed, they also face significant pressure to circumvent restrictions and report on issues of widespread public concern. To the degree that the Internet and online discussion substitute as a source of information for audiences actively seeking information, this creates strong incentives for individual media outlets to break restrictions in order to capture audiences. Under these conditions, the public agenda as expressed on the Internet may thus set the media agenda and further draw public attention to a given issue.

The ability of the public and media agendas to influence the policy agenda ultimately lies in the threat of collective action if leaders fail to respond to public concerns. Admittedly, the possibility of collective action in response to poor air quality, food safety, corruption, or any other single issue of public concern that outside actors or ‘opinion leaders’ might place on the public agenda appears remote. The possibility that collective action linked to any such issue might produce direct challenges to the existing political system or threaten leaders’ monopoly over the policy agenda appears even more remote. Indeed, recent research on online censorship by King, Pan, and Roberts (2013) suggests censors (and by extension leaders) allow criticism but strictly censor calls for collective action.

Yet while the vigilance of censors reduces the possibility of collective action, such behavior reveals leaders’ underlying concerns with the potential for collective action, regardless of its proximate cause. Almost none predicted that the proximate cause

of the collective actions that precipitated the collapse of communist regimes in Eastern Europe in 1989 would lead in such a direction (Kuran 1991). As Lohmann (1994) points out, leaders' response or non-response to public concerns may also reveal information about the nature of leaders' preferences to the public. Even the remote possibility of collective action thus creates some incentive for leaders to respond by adjusting the policy agenda to more closely reflect the public agenda and give the appearance of responsiveness to public concerns. The following section describes one such recent instance in which this chapter argues that leaders did indeed choose to respond and develops hypotheses regarding the order of change in public, media, and policy agendas implied by this process.

Air Quality, Public Skepticism, and Agenda Setting in Contemporary China

Recent episodes of abysmal air quality and public skepticism over the reliability of government reported ambient air quality information provide an ideal case in which to illustrate and test propositions about the ability of actors outside of the party and state to set the public and media agendas as well as to pressure leaders and influence the policy agenda. In particular, this chapter focuses on the first of the now seemingly perennial episodes of poor air quality that occurred in Beijing in the final months of 2011. Though subsequent episodes of poor air quality such as the episode which occurred in January and February 2013 and which the international press dubbed Beijing's 'Airpocalypse' have captured international attention, it is the episode that occurred in late 2011 that brought the issue of poor air quality in China's rapidly expanding urban centers onto the public agenda.

Figure B1 in Appendix B above graphically illustrates this point using online search query data from Google Trends. Figure B1 depicts the relative volume of search queries by Google Search users that include the keywords ‘Air Quality’ in Chinese and that originate from IP addresses registered in the People’s Republic of China.²⁸ Observations are reported on a weekly basis and coverage extends from 2009-2013. A clear pattern of bursting in the relative volume of search queries—direct evidence of members of the public actively seeking information on air quality—is observable starting with a first burst in late 2011 and that falls within the period of the first episode. While the burst’s magnitude is lesser than subsequent bursts, it nevertheless marks the initiation of a series of subsequent episodes related to the issue of poor air quality that shows few signs of abating.

The narrative surrounding the first episode is as follows: Early arrival of a particularly harsh winter in 2011 and the extensive use of coal for interior heating in northern China contributed to a sharp seasonal decline in ambient air quality in the national capitol of Beijing and surrounding provinces. Yet while air quality and visibility had reached equally low levels during the immediately preceding winter (Associated Press 2010; Watts 2010), what differentiated 2011 from prior years was the degree to which public attention came to focus on discrepancies between ambient air quality information reported by municipal environmental officials in Beijing and alternative information reported by the US Embassy in Beijing.

²⁸ For a detailed description of how Google Trends calculates, standardizes, and reports the relative volume of search queries for a given keyword, consult the methodology section of the official Wiki for Google Trends.

Beginning in 2008, officials at the US Embassy began continuous monitoring of ambient air quality in Beijing from the grounds of the US Embassy compound in the municipality's Chaoyang District. US Embassy officials also began reporting this information in terms of Air Quality Index (AQI) scores along with health advisories via the public Twitter account *BeijingAir*. This move was not welcomed by Beijing officials and resulted in heated exchange behind closed doors between representatives from the Ministry of Foreign Affairs (MFA) and US Embassy officials on 7 July 2009. The excerpt below contains the US Embassy's own summary of the exchange was made public by way of the Wikileaks leak of U.S. diplomatic cables:

At the request of the Ministry of Foreign Affairs (MFA), [US Embassy officials] met on July 7 with Mr. Wang Shuai of MFA's Office of U.S. Affairs to respond to MFA's concerns about recent publicity in international and local press surrounding an air quality monitor installed on the Embassy compound. MFA registered complaints on behalf of the Beijing Environmental Protection Bureau (EPB) and the Chinese Ministry of Environmental Protection (MEP), saying that making this data (which in their view "conflicts" with "official" data posted by the Beijing EPB) available to the general public through an Embassy-operated Twitter site has caused "confusion" and undesirable "social consequences" among the Chinese public. MFA asked [the US Embassy] to consider either limiting access to the air quality data only to American citizens, or otherwise identify a suitable compromise.²⁹

The standard used by the US Embassy to monitor and report information on ambient air quality and that provoked concerns from the MFA is the same standard used by the US Environmental Protection Agency (EPA) to monitor and report levels of pollutants regulated by the US Clean Air Act, including fine particulate matter, also known as PM_{2.5}. Alternatively, the existing monitoring and reporting standards used by the Chinese Ministry of Environmental Protection (MEP) and municipal environmental

²⁹ The cable was sent on 10 July 2009 and its Wikileaks reference ID is 09BEIJING1945.

officials in Beijing did not include PM_{2.5} but rather included respirable suspended particles, also known as PM₁₀.³⁰ Public health literature finds that while short-term and long-term exposure to elevated concentrations of both PM_{2.5} and PM₁₀ are hazardous to human health, exposure to PM_{2.5} is more strongly associated with negative health outcomes including cancer, heart failure, and premature death.

While few members of the public likely understood the distinction between PM_{2.5} and PM₁₀ prior to this episode, the use of different monitoring and reporting standards by municipal and US Embassy officials led to divergent metrics for reporting ambient air quality information to the public. For example, PM_{2.5} concentrations on 9 October 2011 rose to such high levels that monitoring equipment operated by the US Embassy registered concentrations that exceeded the maximum allowable for calculating AQI scores. The US Embassy subsequently reported that conditions were ‘Beyond Index’ having exceeded ‘Hazardous’ levels, and issued a corresponding health advisory for US citizens and city residents to remain indoors. On the same day, municipal officials reported an Air Pollution Index (API) score of 121 based upon PM₁₀ concentrations, reporting conditions as only ‘Lightly Polluted’.

Although differences between monitoring standards can largely account for this discrepancy, public skepticism over the reliability of air quality information fell strongly on municipal environmental officials rather than the US Embassy. This chapter argues that common knowledge of alternative information reported by the US Embassy as well as resulting skepticism can be strongly attributed to the activities of what Xiao (2011a;

³⁰ The difference between PM_{2.5} and PM₁₀ is in the size of the particulates being measured. PM₁₀ measures particulates of 10 micrometers radial diameter or less while PM_{2.5} measures particulates of 2.5 micrometers radial diameter or less.

2011b) refers to as online ‘opinion leaders’. As poor air quality persisted through the final months of 2011, well-known online commentators including billionaire real-estate developer Pan Shiyi whose Sina Weibo account registers more than sixteen million followers, as well as environmental advocate Ma Jun, and *Beijing News* editor Yu Ping, played active roles in drawing attention to the existence of information reported by the US Embassy, the discrepancy with information reported by municipal environmental officials, and advocating for reform of existing monitoring and reporting standards.

For example, on 6 November 2011 Pan Shiyi posted an online poll entitled “Appeal to the MEP to Adopt a Mandatory PM2.5 Standard,” to his personal Sina Weibo weblog asking followers to respond to the following message:

Experts say PM2.5 pollution is very harmful to human health. Only when the state has adopted a mandatory standard will each city implement such a standard. Knowing the grave nature of the problem, all people must consciously work to prevent air pollution and change their own unhealthy lifestyles and habits. Ask your friends to vote and to forward this message. In one week, I will place the results of the vote in a letter that I will present to the Minister of Environmental Protection.

Users then were offered the choice of voting for the MEP to adopt such a standard in the current year, the following year, or not do adopt such a standard. While the poll was unofficial, within a week it had drawn over 40,000 participants with 98.3% voting to adopt a mandatory PM2.5 monitoring and reporting standard similar to that used by the US Embassy.

With online discussion and rapidly increasing public attention to the issue of air quality becoming newsworthy, commercial media outlets soon turned their attention to actively reporting on the issue of poor air quality and the impenetrable fog that hanging

over Beijing. Consistent with discussion of the relationship between the public and media agendas in the prior section, this implies a first observable implication regarding the order of change in agendas:

Implication #1: If the public agenda sets the media agenda, then an increase in public attention will precede increase in media attention.

Facing a fast rising tide of public attention, growing questions in the media, and little sign of abatement in poor air quality, national-level leaders chose to respond by publicly announcing their intent to reform monitoring and reporting standards. Premier Wen Jiabao delivered what is now considered the initial announcement of this change at the annual conference of the China Council for International Cooperation on Environment and Development (CCICED), a high-level government advisory body, on 15 November 2011. In summarizing remarks made by Premier Wen, the state-led media reported Wen to have emphasized need for the, “improvement of environmental monitoring and environmental standards, bringing them gradually into line with international standards, so that the results of environmental monitoring and peoples’ perceptions of the environment are closer” (CCICED 2012). Consistent with discussion of the ability of the public and media agendas to influence the policy agenda in the prior section, this suggests a second observable implication:

Implication #2: If the public and media agendas can influence the policy agenda, then an increase in both public and media attention will precede leaders’ announcements

Of course, the initial announcement by Premier Wen did not specifically mention the issue of air quality or publicly commit leaders to the inclusion of PM2.5 into

reformed standards. However, Vice Premier Li Keqiang and Minister of Environmental Protection Zhou Shengxian made further public announcements in mid-December on the heels of another bout of particularly poor air quality in early December that drafting of reformed national monitoring and reporting standards that included PM_{2.5} had been fast-tracked. Ultimately, this process culminated in the passage of the reformed standards by Premier Wen at a special executive council meeting of the State Council on 29 February 2012, immediately prior to the opening of the 2012 annual plenary session of the National People's Congress (NPC).

The reformed standards involved a number of substantial changes to previous ambient air quality monitoring and reporting standards passed in 2000. In addition to tightening limits on pollutants monitored under the previous set of standards including SO₂, NO₂, and NO_x, the new standards included PM_{2.5}. Reforms also included a clear plan and timetable for establishing a national network of automated monitoring stations. The network was targeted to cover 72 municipalities and begin operation by no later than 1 January 2013. The plan ultimately called for the network to cover all of China's municipalities by the close of 2016. The timetable for the rollout of this network was even integrated as a key project in the national 12th FYP. Taken altogether, the policy response was surprisingly swift and exceeded the expectations of many observers.

Data and Methods Description

In order to test for the hypothesized order of changes in the public, media, and policy agendas, this chapter makes use of data on public and media attention to the issue

of poor air quality as well as leaders' announcements of their intent to reform ambient air quality monitoring and reporting standards. Although estimating the date of change in the policy agenda based upon leaders' announcements is relatively straightforward, estimating the date of change in the public and media agendas is more complicated. What constitutes evidence of a change in the public or media agenda? Moreover, how can researchers go about estimating the date of change in one agenda for comparison with change in another?

This chapter treats the estimation of the date of change in the public and media agendas as a problem of identifying structural change in public and media attention. Specifically, this chapter argues sudden and dramatic increases in the level of public and media attention to the issue of poor air quality are indicative of the introduction of this issue onto the public and media agendas, respectively. The hypothesized order of changes to the public, media, and policy agendas described in the prior section implies that public attention should experience such a level-shift prior to a similar level-shift in media attention, and that level-shifts in both public and media attention should precede leaders' announcements.

Data Description

The chapter measure public attention using data on the daily volume of posts by users of the Sina Weibo microblogging service, specifically posts containing the keyword 'PM2.5'. Sina Weibo is China's most popular microblogging platform and is considered by media scholars to offer a space in which members of the public have the opportunity to raise and discuss a range of issues not explicitly prohibited by censors. As previously

mentioned, ‘PM2.5’ also emerged as a salient term with respect to concerns over the issue of poor air quality and a focal point for concerns the reliability of government reported, ambient air quality information. Moreover, public demands for leaders to reform existing monitoring and reporting standards focused specifically on the inclusion of PM2.5 into a new standard.

Data on the daily volume of posts was collected from Weibo Search, a service offered by the Chinese online media company Sina Corp. that owns and operates Sina Weibo.³¹ Weibo Search allows users to query the daily volume of posts to Sina Weibo that contain a user-specified keyword and that also meet user-specified classifications regarding the type of account from which the post originated. For example, users of Weibo Search may query the volume of posts made from verified versus non-verified accounts. Weibo users with verified accounts have verified their offline identity with administrators of Sina Weibo while users with non-verified accounts have not gone through this process.

The distinction between verified and non-verified users is not a trivial one vis-à-vis agenda setting theory and earlier discussion of online ‘opinion leaders’. Verified users include well-known celebrities, entrepreneurs, and public intellectuals that have the ability to quickly and widely distribute information and their opinions to a wide audience of other users. So-called “Big-Vs” like as Pan Shiyi can have millions of followers and are considered to possess the ability to influence which issues receive attention from both verified and non-verified users. Furthermore, verified Weibo users are identifiable as their accounts reflect their offline identity and they have a greater stake in maintaining

³¹ Weibo Search can be accessed at www.s.weibo.com.

their reputation among followers regardless of whether their followers are verified or non-verified users. Consequently, level-shift in the volume of posts from verified accounts is more indicative of change in the public agenda than is level-shift in the volume of posts from non-verified or all accounts.

Alternatively, the chapter measures media attention using data on the daily volume of media coverage as reported by Baidu Index.³² Baidu Index is a service provided by the Chinese online media company Baidu Corp. that owns and operates Baidu, China's most popular search engine. Similar to Google Trends, Baidu Index reports the daily volume of search queries by Baidu users according to keyword in the form of Baidu Search Index (BSI) Scores. In addition to BSI Scores, Baidu Index also reports information on the daily volume of online media coverage based upon user-specified keywords in the form of daily Baidu Media Index (BMI) scores.

According to the official Wiki for Baidu Index, daily BMI scores report the daily volume of articles containing user-specified keyword posted to major online news portals such as Sohu.com and Baidu News. Although it is unclear as to whether BMI Scores reflect the volume of original articles or re-postings of the identical articles across multiple portals, the distinction is of questionable importance in measuring media attention: Re-posting of articles across multiple portals indicates greater media attention to an issue regardless of whether content is original.

Rather than daily BMI scores generated using the keyword 'PM2.5' to measure media attention, this chapter instead uses daily BMI scores generated using the keyword 'Air Quality' in Chinese. The reason underlying this choice is simple: Daily BMI scores

³² Baidu Index can be accessed at www.index.baidu.com.

generated using the keyword ‘PM2.5’ do not vary or assume a value greater than zero prior to 2 November 2011. This indicates that prior to 2 November 2011, the media does not refer to PM2.5 when discussing the issue of poor air quality. Although a trivial point, the sudden introduction of the term PM2.5 into media coverage of air quality when the term was previously absent from coverage itself represents a structural change in media attention and strongly suggests the date of change in the media agenda. Furthermore, estimating the date of level-shift in media attention requires some variation in the measure of media attention prior to the probable date of change. Alternatively, BMI scores generated using the keyword ‘Air Quality’ exhibit variation across the entire sample window.

Data on the daily volume of Weibo posts containing the keyword ‘PM2.5’ from both verified and non-verified accounts as well as daily BMI scores generated using the keyword ‘Air Quality’ in Chinese were collected for a sample window of eight months starting on 1 July 2011 and ending on 29 February 2012. Summary statistics are presented in Table B1 in Appendix B.

Summary statistics on the range and variance of each variable suggests that all variables should be scaled to address outliers. Table B1 also includes logged values of each variable where logged values are calculated by taking the natural logarithm of one plus the original values to account for the presence of zeros in the unlogged variables. All results and analysis presented in the subsequent section are generated using log values.

Methods Description

This chapter uses two complimentary methodological approaches to detect structural change in public and media attention: The supremum F-Test for detecting endogenous or unknown structural breaks in time-series data and the information criteria based method for detecting multiple, unknown breaks suggested by Bai (1997a; 1997b) as well as Bai and Perron (1998; 2003). The intuition underlying both approaches is relatively straightforward. In the context of time-series analysis, both approaches detect structural change or instability in time-series by modeling the series in both a piecewise and non-piecewise fashion, and then compare the fit of resulting models where the null is non-piecewise.

Given the hypothesized nature of structural change in both public and media attention, this paper models the data using simple, intercept-only models. Although this ignores the time-series structure of the data, the hypothesized increase in public and media attention to the issue of poor air quality beginning in late 2011 should be observable in a sudden increase but persistently elevated level in both the daily volume of Weibo posts as well as daily BMI scores. Testing for the presence of a level-shift and estimating the date of change does not require modeling the time-series nature of the data.

The first approach to detecting level-shift involves the supremum F-Test, an extension of the Chow test. Originally proposed by Chow (1960) in order to test for structural change or breaks in the relationship between variables in linear regressions, the Chow test is also used to detect breaks at a known date in time-series data. The test involves fitting a model to a sub-sample of data prior to a known date, fitting the same

model to the sub-sample of data after the break date, and then comparing the fit of the two models by calculating an F-Statistic. Under the null hypothesis (i.e. no structural change at the suspected date), the F-Statistic should converge to zero. Alternatively, the null hypothesis can be rejected if the value of the F-Statistic deviates upward and away from zero, suggesting the presence of structural change at the suspect date.

Although the Chow test is one of the most basic tests for structural change, it requires a researcher to *ex ante* specify the date of the break. While the exact dates and number of breaks are sometimes known, this chapter treats the dates of the hypothesized changes in public and media attention as unknown—also known as endogenous breaks. Under such conditions, the supremum F-Test can be used to estimate the dates and number of endogenous breaks. Rather than require pre-specification of a date, the supremum F-test treats each date within the sample as a potential break and then calculates separate models and F-Statistics for all possible breaks. Resulting F-Statistics can then be plotted against time with the supremum or maximum value of resulting F-Statistics providing a consistent estimate of the date of a break date, if any, in the time-series. Multiple peaks in the resulting plot of F-Statistics against time suggest multiple breaks.

In addition to the supremum F-Test, the chapter also uses the information criteria-based approach for detecting multiple and unknown structural breaks suggested by Bai (1997a; 1997b) as well as Bai and Perron (1998). While this approach is similar to the supremum F-test in that it involves the estimation the same model on all possible sub-samples or segments, it differs in that it involves the estimation of models on multiple

sub-samples of observations in a time-series, subject to constraints on the total possible number of breaks in the data.

The algorithm underlying this approach is described in Bai and Perron (2003) and is implemented in the R package *strucchange* as described by Zeileis et al. (2003). For estimating the date of a break in a single or one break model, the algorithm involves estimating all potential models that contain one break, calculating the resulting residual sum of squares (RSS) for each potential model, and selecting the model with the overall minimum RSS. For estimating the dates of breaks in a two-break model, the algorithm takes all potential one-break models, inserts an additional break, re-estimates and saves the resulting RSS of each potential model, and then selects the model with the minimum RSS. This process is repeated until the maximum possible number of breaks is reached.

The resulting set of models with one-or-more breaks can then be directly compared on the basis of their resulting Bayesian Information Criteria (BIC). The model with the minimum BIC is then selected and confidence intervals can be constructed around the estimated date or dates of structural breaks. While the resulting Akaike Information Criteria (AIC) can also be used to select the model with the best fit, Bai and Perron (2003) suggest that using the AIC to guide model selection will lead to overestimates in the optimal number of breaks. The benefit of this approach is that it allows for the estimation of confidence intervals around estimated breaks. The resulting estimated dates and confidence intervals can then be compared between time-series to test the hypothesized order of level-shifts in public and media attention and, by extension, the appropriate causal order of changes in public, media, and policy agendas.

Results and Analysis

Does the order of increase in the levels of public and media attention to the issue of poor air quality support the hypothesized order of structural changes in the public and media agendas? Moreover, do these increases precede announcements by leaders? Empirical results below support the hypothesized order of structural changes with the estimated dates for increase in the daily volume of Weibo posts containing the keyword ‘PM2.5’ preceding the estimated date for increase in daily BMI scores. Furthermore, both changes appear to precede the initial public announcement by Premier Wen Jiabao on 15 November 2011 of leaders’ intent to reform ambient air quality monitoring and reporting standards to more closely reflect public perceptions and address public concerns.

Results for the optimal number of breaks and the estimated date of level-shifts in the daily volume of Weibo posts can be seen in Figures B2 and B3 in Appendix B. The left panel of Figure B2 depicts the trend as well as the estimated date of level-shifts in the volume of posts by the users of verified accounts whereas the middle panel depicts the results from the supremum F-tests and the right panel depicts the results for optimal number of breaks according to the minimum BIC. The left, middle, and right panels of Figure B3 depict the same information based upon the volume of posts by users of non-verified accounts.

Visual analysis of the trends in the left panels of both Figure B2 and B3 suggest clear level-shift in the daily volume of posts from both verified and non-verified users between the months of September and November 2011. The single peak in the plot of the F-statistics in middle panels of both Figure B2 and B3 indicates a one-break model is appropriate for modeling level-shift in the trends. However, the minimum BICs depicted

in the right panels of both Figure B2 and B3 suggest instead using a two break model. Estimated coefficients for intercept-only models using the daily volume of posts from verified accounts with zero, one, and two breaks are reported below in Model 1, Model 2, and Model 3 in Table B2, respectively. Estimated coefficients for intercept-only models using the daily volume of posts from non-verified accounts with zero, one, and two breaks are reported in Model 4, Model 5, and Model 6 in Table B2, respectively.

Using zero, one, and two-break models, the left panels of both Figure B2 and B3 depict the estimated intercept-only models of the data by plotting the fitted values for each model. Fitted values for models with zero, one, and two breaks are depicted in blue, red, and green, respectively. Changes in level of intercepts correspond to the estimated date of level-shifts in the trend. For models estimated with a single break, the estimated date of level-shift in posts from verified accounts is 19 October 2011 while the date of level-shift in posts from non-verified accounts is 8 October 2011. For models estimated with two breaks, the estimated date of the first and smallest level-shift is on 5 September 2011 for posts from both verified and non-verified accounts while the date of the second and largest level-shift is on 19 October 2011, consistent with the model with one break for posts from verified accounts.

The optimal number of breaks and estimated date of level-shift in daily BMI scores can be seen in Figure B4 in Appendix B. Similar to Figures B2 and B3, the left panel of Figure B4 depicts the trend as well as the estimated level-shifts in daily BMI scores whereas the middle panel depicts the results from the supremum F-test for structural change and the right panel depicts the results for optimal number of breaks according to the minimum BIC.

Visual inspection of the trend in the left panel of Figure B4 strongly suggests a level-shift between late October and early November. The single and sharp peak in F-Statistics in the middle panel as well as the minimum BIC in the right panel both suggest a single break model provides an appropriate basis for modeling the trend. Similar to Figures B2 and B3, the left panel of Figure B4 depicts both zero and single break models by plotting fitted values. Fitted values for models with zero and one break are depicted in blue and red, respectively. Estimated coefficients for intercept-only models with zero and one break are reported in Models 7 and 8 in Table B2, respectively. Changes in level of the model intercepts correspond to the estimated date of level-shift in the trend.

For the single break model, the estimated date of level-shift in the daily BMI scores is 31 October 2011. This is consistent with the hypothesized sequence of changes if the public agenda is indeed capable of setting the media agenda. Moreover, change in the media agenda indicated by the level-shift in BMI scores also precedes the first public announcement of leaders' intent to reform ambient air quality monitoring and reporting standards, thereby indicating change in the policy agenda. Table B3 reports the estimated dates of change in the public and media agendas, their corresponding 95% confidence intervals, and the date of the public announcement by Premier Wen in chronological order.

The confidence intervals for estimated dates of level-shifts in the daily volume of Weibo posts precede and do not overlap with the estimated dates of level-shift in daily BMI scores. Furthermore, the confidence intervals both temporally precede the initial date on which daily BMI Scores generated using the keyword 'PM2.5' rise above zero, thus indicating the first time the media uses the term PM2.5 in reporting on the issue of

air quality and, by definition, a level-shift in the trend. Importantly, the confidence interval surrounding the level-shift in BMI scores also precedes and does not overlap with the first announcement of leaders' intent to reform ambient air quality monitoring and reporting standards. Taken altogether, this supports the hypothesized order of changes in the public, media, and policy agendas through which the public agenda can influence the policy agenda.

Discussion and Implications

To briefly conclude, results and analysis presented above appear to confirm the hypothesized order of change in agendas. By demonstrating the appropriate causal order of increase in the level of public and media attention, results and analysis provide support for the ability of actors outside of the party and state to set the public as well as media agendas. The date of leaders' announcements following increases in public and media attention further supports the contention that these actors, formally outside of the policy-making process, are nevertheless capable of pressuring leaders and influencing the policy agenda. The responsiveness of leaders thus suggests a non-institutional channel through which actors formally outside the policy-making process can influence the policy agenda by first setting the public agenda.

The use of this strategy has not gone unnoticed in the substantive literature. For example, both Wang (2008) and Mertha (2009) note similar instances of actors outside of the party and state who have traditionally been excluded from the policy-making process attempting to influence the policy agenda by first setting the public agenda. These scholars further point to the use of this strategy as well as its occasional effectiveness in

provoking the response of leaders as evidence of increasing pluralization of the policy-making process despite the absence of moves towards political liberalization.

These signs of the pluralization of the policy-making process and the responsiveness by leaders to the public agenda are worth note. However, it is also worth noting that responsiveness on the part of leaders does not necessarily imply that they are accountable to these actors, the public more broadly, or that leaders are without the means to limit the ability of these actors to influence the policy agenda. For example, an anti-rumor campaign launched by online censors and state authorities in August 2013 targeted so-called “Big-Vs” as well as lesser-known Sina Weibo users. Charles Xue, an American investor of Chinese origin and popular online commentator with over twelve million online followers known for railing against official corruption and venality, was arrested in Beijing on 23 August 2013 for soliciting prostitution. The arrest and subsequent public shaming of Xue on state television was widely seen as politically motivated and intended to impress the point that even so-called “Big-Vs” are not beyond the reach of leaders (Buckley 2013). Nevertheless, while the strategic interaction between such actors and policymakers is worth expanding upon, this chapter leaves this task for future work.

CHAPTER 4: MAKING SENSE OF MANIPULATION

GDP figures are 'man-made' and therefore unreliable... When evaluating Liaoning's economy, [Li Keqiang] focuses on three figures: (1) electricity consumption; (2) volume of rail cargo; and (3) the amount of loans disbursed... other figures, especially GDP statistics, are 'for reference only,' he said smiling³³

How can scholars make sense of manipulation? Empirical results presented in the first chapter demonstrate evidence that sub-national officials respond to institutional incentives by manipulating indicators of their performance in office. The result of strategic behavior by officials, evidence of manipulation raises two puzzles for an important debate in literature on contemporary Chinese politics.

This debate concerns what influences the promotion of sub-national officials within the modern party and state hierarchy and pits two groups of scholars against one another: Scholars who argue that performance influences officials' subsequent promotion prospects and scholars who instead argue that factional affiliation influences their promotion prospects. The two puzzles raised by the first chapter are as follows: First, if performance does not influence the promotion of sub-national officials, then what explains clear evidence of the manipulation of indicators that leaders use to monitor performance? Second, if officials can manipulate and thereby systematically bias

³³ The excerpt was drawn from a now well-known U.S. Diplomatic cable leaked by Wikileaks and originally sent from the U.S. Embassy in Beijing. The full cable summarizes the contents of a meeting between U.S. Ambassador Clark J. Randt and current State Premier Li Keqiang in 2007. At the time of the meeting, Li was serving as Liaoning CCP Provincial Secretary. Liaoning is a developed coastal province located in northeastern China.

indicators, then is it actually possible for leaders to promote officials based upon their performance in office?

This chapter offers a potential solution to these puzzles. Drawing on models of manipulation presented in a literature on finance and accounting as well as often overlooked evidence from substantive literature on the quality of GDP-based macroeconomic indicators, this chapter argues that leaders are aware of the perverse incentives created by top-down accountability institutions like the cadre evaluation system. Leaders neither naively accept indicators as unbiased nor do they uncritically incorporate information gleaned from indicators into subsequent decisions. Though only occasionally mentioned in the state-led press or discussed in the substantive literature, leaders correct for systematic bias introduced through manipulation by discounting bias concealed in indicators—a process referred to as ‘wringing out water content’—based upon rational expectations of the magnitude of bias introduced by officials. In this sense, leaders attempt to ‘back out’ bias and generate unbiased and valid indicators of performance.

Yet while this chapter argues that leaders attempt to generate unbiased indicators and can *on average* correct for bias, it argues they cannot perfectly correct for bias on a case-by-case basis.³⁴ The reason for this is leaders’ uncertainty over the reporting

³⁴ To briefly explain why this distinction is potentially rather important, consider the problem of optimally allocating limited government resources and assistance across a number of jurisdictions following a natural disaster. Leaders develop rational expectations that all officials will overstate the needs of their jurisdiction for assistance as well as the degree to which officials will (on average) overstate need. Leaders subsequently adjust their allocation in light of these expectations. However, leaders’ uncertainty over the degree to which officials overstate need inevitably leads to under-allocation of assistance to some jurisdictions and over-allocation to others. Greater uncertainty does not mean that leaders under-allocate or over-allocate on average, but that the degree to which they under-allocate or over-allocate in individual cases is greater. According to Amartya Sen, failure to optimally allocate assistance in the wake of natural disasters in this way is the proximate cause of man-made disasters such as ensuing famines.

objective of officials: Leaders may develop rational expectations that officials bias indicators by a certain magnitude but are uncertain at any time about the exact costs and benefits to officials from biasing indicators.

Uncertainty over the reporting objective implies that leaders correct for bias on average but also under-estimate bias in the case of some officials and over-estimate bias in the case of others. This implies that after wringing out water content, corrected indicators offer a noisier signal of the performance of officials than is assumed by scholars who ignore manipulation. Greater uncertainty over the reporting objective of officials implies an even noisier signal. As the epigraph above suggests, leaders are aware of this and consequently treat indicators as less informative or, in the words of Premier Li Keqiang, ‘for reference only’ when evaluating sub-national conditions as well as the performance of officials.

This substantive and theoretical point has critical implications for recent empirical studies of the *performance-promotion hypothesis*. Contrary to predictions drawn from current theories of performance-based promotion, studies that attempt to directly test the conjecture that leaders will be more likely to promote officials who perform by delivering rapid economic development in their assigned jurisdictions produce null and weak results. Scholars who argue against theories of performance-based promotion have subsequently seized upon these results as evidence that falsifies theories advanced by their opponents and as evidence that supports their own, namely those that emphasize factional affiliation.

However, the argument at the heart of this chapter implies that both groups of scholars have overlooked the role of manipulation and the response of leaders to

manipulation in generating these results. Uncertainty about the reporting objective of officials makes corrected indicators noisier and less informative to leaders when making decisions about whom to promote by lowering the signal-to-noise ratio of indicators as a measure of true performance. This implies that when making decisions about whom to promote, leaders will be less sensitive to differences in indicators as these are more likely to be the product of noise than differences in true performance. The implication for tests of the performance-promotion hypothesis is that leaders' decision to grant rewards (i.e. promotions) are less sensitive to differences in indicators used to proxy for performance. In terms of hypothesis testing, this argument implies that tests of the performance-promotion hypothesis will be biased towards type II error and produce results that lead to incorrect failure to reject the null hypothesis that performance does not influence promotion.

The contribution offered by this chapter is thus not just to make sense of evidence of manipulation but also to offer an alternative explanation for one of the most vexing empirical puzzles in literature on contemporary Chinese politics. This explanation is unique in that it provides a theoretical reasoning for this puzzling result that incorporates substantive knowledge about perverse institutional incentives. Moreover, it suggests limits to top-down accountability institutions like the cadre evaluation system and the argument that the leaders of the Chinese Communist Party (CCP) regime have somehow *gotten it right*.

The remainder of this chapter is as follows: First, the chapter summarizes debate over the role of performance in the promotion of sub-national officials in China. Second, the chapter briefly detours from this debate to introduce a literature from finance and

accounting about the manipulation of accounting indicators that offers much needed answers. Third, the chapter presents its argument regarding how leaders respond to the manipulation of indicators. Fourth, the chapter reviews the puzzling results of recent empirical studies that seek to test the performance-promotion hypothesis and articulates how the argument developed in the previous sections explains these results. Lastly, the chapter offers a brief conclusion as well as discussion of implications.

Debate over Performance and Promotion

The role of performance in the promotion of sub-national officials is the subject of a major debate in literature on contemporary Chinese politics. On the one hand, a first group of scholars argues that officials' performance is critical to their subsequent prospects for promotion upwards within the modern party and state hierarchy. The argument advanced by these scholars is that leaders use top-down accountability institutions like the cadre evaluation system to monitor officials as well as to generate information about their performance in office. Leaders then condition promotion decisions upon this information.

In particular, a handful of these scholars argue that China's geography-based administrative system allows leaders to generate information about officials' performance relative to their peers. By rank ordering officials based upon their relative performance vis-à-vis economic development indicators like per capita GDP growth rates and by conditioning promotion decisions on these rankings, leaders not only create tournament-like competition amongst officials to deliver on leaders' development targets and goals but also select the best performing and most competent officials for promotion (Qian and

Xu 1993; Maskin, Qian and Xu 2000; Qian, Roland and Xu 2006; 2007; Xu 2011). This argument underlies claims advanced by pundits that promotion within the modern party and state hierarchy is meritocratic (Bell 2012; Li 2013) and that leaders have somehow gotten it right with respect to rule through top-down accountability institutions where their peers in the Soviet Union and Eastern Europe failed.

On the other hand, a second group of scholars argues that this reasoning is deeply flawed. These scholars question why leaders would necessarily wish to promote the top performing and the most competent officials based upon the nature of politics in authoritarian regimes. In particular, these scholars question why leaders would promote officials on the basis of performance and their perceived competence rather than based upon their personal loyalty to leaders.

Literature on authoritarian regimes suggests that leaders do indeed have much to fear from subordinates. For example, Egorov and Sonin (2011) argue that leaders face a tradeoff between loyalty and competence when selecting subordinates. They argue that leaders who are especially vulnerable to being replaced select subordinates who are loyal rather than competent. The reason for this is that competent subordinates are more capable of removing leaders from office whether through coups or alternative means. To this end, Boix and Svolik (2013) demonstrate that coups—whether by fellow leaders or subordinates—represent the primary manner of exit from office for the leaders in authoritarian regimes. Though the leaders of regimes that institutionalize politics through parties or legislatures are substantially less likely to face the threat of removal by coup, coups nevertheless remain the modal manner by which leaders exit from office.

Substantive literature on informal and personalistic factions in Chinese leadership politics further supports the contention that loyalty as well as concerns over the factional balance of power within elite leadership bodies also influence promotion decisions (Dittmer 1995; Dittmer and Wu 1995; Nathan 1973; Nathan and Tsai 1995). Although the factional affiliation of officials is difficult to assess given the informal nature of factions, scholars maintain that observables like shared work and educational experience can be used to proxy for factional affiliation. For example, Shih et al. (2010) use biographical information to proxy for factional affiliation and examine the membership composition of CCP Central Committees by factional affiliation. Despite the argument that elite politics have become more institutionalized and characterized by norms of intra-elite civility since the Cultural Revolution, they find evidence to suggest that leaders still attempt to pack this top decision-making body with their own factional supporters rather than elevate the supporters of rival leaders.

Of course, there is reason to believe that performance and factional affiliation are not always rival. For example, leaders would presumably prefer to promote officials who are both loyal and competent. However, the current literature generally pits these two arguments against one another and leaves little room to maneuver. Three decades of spectacular development in China and the absence of outward signs of major factional conflict within the ruling Chinese Communist Party have further led scholars who emphasize the importance of performance to assume a dominant position within the debate.

Nevertheless, findings in support of a core proposition drawn of theories of performance-based promotion are surprisingly fragile. Specifically, empirical studies that

directly test the performance-promotion hypothesis—the conjecture that leaders will be more likely to promote sub-national officials who ‘perform’ by delivering rapid economic development—tend to generate null results or results that are substantially weaker than predicted by theory (Shih et al. 2012; Su et al. 2013). It is with these puzzling results in mind that this chapter briefly detours from the above debate to lay the foundation for a possible alternative explanation that draws on a scholarly literature in finance and accounting that may offer an answer.

Literature on Manipulation in Finance and Accounting

What can explain null or weak results in support of the performance-promotion hypothesis yet evidence that sub-national officials nevertheless manipulate indicators of their performance in office? This section makes the case for why an analogous problem discussed in a literature on finance and accounting may help explain this puzzle. Specifically, this section argues that a weaker than predicted relationship observed between promotion and performance is analogous to weaker than predicted pay-for-performance sensitivity of compensation packages offered by boards of directors to the managers of firms observed in the literature on finance and accounting.

To begin, standard principal agent theory argues that in order to closely align the interests of managers with their own, boards offer managers compensation packages with high pay-for-performance sensitivities that closely link managers’ compensation to the performance of firms.³⁵ However, the pay-for-performance sensitivity of compensation

³⁵For example, standard principal-agent theory would prescribe that boards of directors offer compensation packages that give managers equity in the firm like stock or options in order to align their interests with those of boards and shareholders more generally

packages actually offered to managers is substantially lower than predicted by theory (Jensen and Murphy 1990). Scholars working in the fields of finance and accounting have subsequently developed a literature to explain this seemingly puzzling behavior on the part of boards of directors.

Beginning with Stein (1989), one branch of this literature focuses on how equity-based compensation can induce managers to manipulate accounting indicators such as earnings used to monitor performance and firm value. In the model proposed in Stein (1989), managers manipulate earnings by making myopic investments that boost short-term earnings at the expense of the long-term firm value. Models proposed by Fischer and Verrechia (2000) as well as Goldman and Sleazak (2006) allow managers to bias earnings reports through faulty accounting practices. Regardless of the method used by managers to manipulate earnings, the common thread that connects these models is that by closely linking compensation to indicators of performance through packages with high pay-for-performance sensitivities, boards create incentives for managers to manipulate indicators.

The class of models that underlies this literature is so-called signal-jamming models.³⁶ Originally proposed by Holmstrom (1999), signal-jamming models are the compliment of signaling models: Whereas actors in signaling models reveal private information through personally costly action, actors in signal-jamming models conceal private information through personally costly action. The interpretation of these models in the context of manipulation by managers is that they must expend personally costly

³⁶ Fudenberg and Tirole (1986) were the first to term these models ‘signal-jamming’ although Holmstrom (1982) offers the first example of the model.

effort to bias earnings reports or oversee myopic investments to manipulate indicators and conceal true performance, thereby jamming the signal of performance that would otherwise be observed by boards.

In equilibrium, boards are not fooled: Boards rationally expect managers to manipulate and thus boards discount indicators for manipulation based upon their expectations and knowledge of the cost and benefits to managers from manipulation. Despite an inability to fool boards that are capable of ‘backing out’ and correcting for bias, managers nevertheless continue to manipulate. Managers cannot credibly commit to refrain from engaging in manipulation, thus boards discount indicators and correct for suspected bias regardless of whether or not managers actually manipulate. Though managers do not benefit from manipulation, they consequently have no choice but to manipulate.

However, this result hinges on two critical assumptions: Boards (1) possess rational expectations of the magnitude of bias introduced by managers and (2) possess perfect information with regards to the reporting objective of managers, where the reporting objective refers to managers’ optimization problem of how to balance the costs and benefits from manipulation. Fischer and Verrechia (2000) loosen the second assumption in their proposed model to demonstrate that even when boards possess rational expectations regarding the magnitude of bias introduced by managers, uncertainty over the reporting objective undermines their ability to perfectly correct for bias in the case of individual managers. While rational expectations allow boards to correct for bias on average, uncertainty prevents boards from perfectly correcting for bias on a case-by-case basis.

The practical implications of this result are important for understanding how informative indicators are to boards as a measure of the performance of managers. Whereas boards correct for bias on average, uncertainty means that they over-estimate the true magnitude of bias in the case of some managers and under-estimate in the case of others. Greater uncertainty means boards under-estimate and over-estimate the true magnitude of bias by a greater margin in individual cases. While such errors cancel out in the aggregate, they nevertheless lower the signal-to-noise ratio of a given indicator as a measure of performance.

This is important in that noisier indicators are poorer measures of performance. In the case of firms, boards of directors end up over-compensating some managers and under-compensating others. In the extreme case where boards are entirely uncertain with regards to the reporting objective of managers, noise approaches infinity, the signal-to-noise ratio of corrected indicators approaches zero, and compensating managers based upon indicators is no better than compensating managers at random. Under such conditions, performance-based pay no longer makes sense given that boards cannot glean information from corrected indicators.

Although such conditions are unlikely to arise, they serve to illustrate the point that increasing uncertainty over the reporting objective reduces how informative indicators are to boards in determining the performance and making decisions regarding the compensation of managers. One of the implications of this relationship is that in order to control managers' perverse incentives to manipulate, boards offer managers compensation packages with lower pay-for-performance sensitivities. Alternatively stated, boards offer packages in which a smaller portion of compensation comes in the

form of equity. The empirical implication of this strategy is that the observed relationship between compensation and differences in indicators of performance will be weaker than predicted by standard models that do not incorporate manipulation or uncertainty over the reporting objective.

There thus appears to be a potential analogue for the weaker than predicted relationship between promotion and performance in contemporary China in the lower than predicted pay-for-performance sensitivities of compensation packages offered to the managers of firms. The following section develops this argument by considering evidence that leaders are aware of manipulation as well as systematic bias within indicators, and how leaders respond.

Manipulation and the Response of Leaders

Although the analogy of boards and managers with leaders and officials follows from the principal-agent relationship that exists between each pair of actors, the application of models originally used to explain manipulation in a firm setting to sub-national governance in China requires some explication. The goal of managers in manipulating indicators such as earnings is to influence assessments of their firm's value to boost the firm's stock price. The potential benefits to managers from manipulation are straightforward: Managers whose compensation packages contain equity-based incentives can increase their own compensation by manipulating and inflating stock prices. Analogously, sub-national officials face incentives to manipulate indicators similar to those faced by managers. Officials desire to increase their own compensation,

where the promise of promotion within the party and state hierarchy generates ‘high-powered’ incentives similar to equity-based compensation (Whiting 2004).

Despite being absent from discussion in scholarly literature on promotion, the manipulation of indicators by officials in response to institutional incentives as well as resulting systematic bias in indicators is common knowledge. However, significantly less is understood about the degree to which leaders are aware of manipulation and how they respond to resulting bias in indicators. Though evidence of how leaders actually correct for bias and incorporate information gleaned from corrected indicators is sparse, this section suggests that the process appears to be analogous to the process described in literature on finance and accounting through which boards of directors correct for bias in accounting indicators.

Sub-national officials’ incentives to manipulate indicators are common knowledge in China. The Chinese financial press regularly carries discussion of the chronic problem of ‘water content’ (*shuifen*) in macroeconomic indicators such as sub-national GDP growth rates. Reported provincial growth rates over the past decade have regularly exceeded national GDP growth rates, an anomaly that is commonly attributed to manipulation by career-interested officials who see delivering rapid GDP growth as a means to promotion. This and other statistical anomalies in GDP-based indicators are taken by some to suggest widespread manipulation. Moreover, this has raised extensive debate amongst economists over the quality of China’s GDP statistics since the late 1990s (Chow 2006; Holz 2002, 2003, 2004, 2006, 2008, 2013; Maddison 2006; Rawski 2001).

Although officials' incentives to manipulate are common knowledge, it is only through the occasional hint dropped in speeches or statements that leaders acknowledge they are aware of systematic bias in indicators or how they correct for systematic bias. The methods through which leaders 'wring out water content' (*ji shuifen*) and correct for systematic bias are not well understood.³⁷ Yet although sparse, information does exist to suggest that leaders discount indicators that they perceive to be exaggerated or biased and have done so for decades.

For example, the passage below is drawn from a report submitted by Vice Premier Tan Zhelin and Minister of Agriculture Liao Luyan to the Central Committee on 16 November 1958 and suggests that leaders were aware of officials' perverse incentives to manipulate even during the Great Leap Forward (1958-1962) and corrected for bias in agricultural production figures by discounting indicators:

The total crop yield for 1958 will reach 425 billion kilos. This is after the provinces, cities, and autonomous regions reduced the figures provided by the counties by 10 to 30 percent...The figure of 425 billion kilos is relatively reliable...the total will not be lower than 375 billion kilos, and can be published as such. This is...a very great leap forward. This point must be affirmed; this overall total cannot be cast into doubt by a small minority of false or exaggerated reports.³⁸

According to MacFarquhar (1983), CCP Chairman Mao Zedong himself echoed the contents of this report two weeks later in a speech to the 6th Plenum of the 8th Central Committee in Wuhan, Hubei province on 30 November 1958. In the speech, Mao referred to false reporting as a problem, repeated the corrected and lower-bound estimate

³⁷ If leaders' strategy were, then sub-national officials could presumably adjust their behavior in light of this information to optimize, potentially rendering the strategy of leaders sub-optimal.

³⁸ The translation of the passage is taken directly from Yang (2012; pg. 330). The original report from which the passage was taken is entitled, "Central Committee's Approval of 'The Overall Situation, Problems, and Views Regarding Agricultural Production and the Rural People's Communes,' by Comrades Tan Zhelin and Liao Luyan."

of 375 billion kilos, and advocated that the published indicator should be further reduced to 350 billion kilos, “for safety’s sake.”³⁹

Regardless of the fact that leaders’ expectations turned out to be inaccurate and leaders grossly under-estimated the magnitude of bias, this passage reveals that leaders were nonetheless aware of officials’ perverse incentives to manipulate. Moreover, it reveals that leaders and sought to correct for resulting bias in light of these expectations by discounting indicators, even during the Great Leap Forward.⁴⁰

The epigraph at the beginning of this chapter suggests that contemporary leaders are similarly aware of perverse incentives and even suggests the methods that they employ correct for bias. Drawing on comments made by former National Bureau of Statistics (NBS) Commissioner Liu Hong in a 2000 interview, Holz (2014) confirms that leaders contrast reported provincial GDP data with economic data collected through sample surveys in each province as well as by comparing the sub-national growth rates of variables related to GDP—such as energy consumption and rail cargo—under the

³⁹ Mao again repeated these comments in a second speech to the Central Committee on 9 December 1958, but instead advocated that the publication of the estimate of 375 billion kilos as the upper-bound estimate had since risen to over 430 billion kilos. The discussion of Mao’s comments at the 6th Plenum of the 8th Central Committee can be found in MacFarquhar (1983; pg. 120-121).

⁴⁰ The inaccuracy of leaders’ estimates of the magnitude of the bias introduced by officials during the Great Leap Forward period as well as the inability of other leaders to openly question Mao or agricultural production figures that falsely indicated the success of the collectivization of agriculture raises an important question: Could leaders’ expectations during this episode be described as rational? Unfortunately, this question cannot be answered in this chapter. However, one promising direction for research is suggested by Bolton et al. (2006) who propose a model of short-term behavior by the firm managers in speculative markets. The proposed model suggests that optimal incentive contracts under such conditions may induce myopic behavior by managers at the expense of the long-term value of the firm as speculative investors with short time horizons determine the incentive contract offered to the manager. This is analogous to arguments made in leadership histories of the period that emphasize the link between the Great Leap Forward’s success and Mao’s continued hold on power. Simply put, Mao could not afford to let the Great Leap Forward—a crash development program designed to industrialize China within five years rather than the decades it had taken the Soviet Union—fail else risk being sidelined within the CCP leadership, an outcome which was realized when it spectacularly failed. With Mao’s short-time horizon and the ability to set (at least implicitly) the incentive contract for sub-national officials, it is little wonder that officials were willing to overstate grain production figures and over-extract grain from the countryside even when they knew this was resulting in death from famine on a massive scale, undermining the regime.

assumption that GDP growth rates cannot outpace the growth rates of variables related to GDP. When reported provincial GDP growth rates are found to outstrip these variables, leaders presumably correct GDP according to variables they suspect officials have not manipulated.

Whatever the method used by leaders to discount indicators, the result is the same: When leaders suspect an indicator is biased and overstated, they correct the indicator downward in line with their expectations about the magnitude of the bias. Alternatively, when they suspect an indicator is understated, they correct the indicator upward. Leaders then use the resulting corrected indicator, now presumably free of bias, to evaluate local conditions or the performance of officials. The discounting behavior of leaders to correct for bias is consistent with the discounting behavior of boards discussed in the finance and accounting literature.

Yet while leaders discount indicators and may develop rational expectations about the magnitude of bias on average, uncertainty on the part of leaders regarding the reporting objective of officials is a reasonable assumption. In the context of tournament-like competition amongst officials for limited promotion opportunities within the party and state hierarchy, leaders are unlikely to possess perfect knowledge of the potential costs and benefits to officials from biasing indicators. For example, leaders do not perfectly know the effort and resources required for officials to achieve a workable scheme to manipulate indicators. Falsifying data used to calculate indicators may require officials spend limited political capital in the form of promises or favors to subordinates in order to assure their compliance in bias schemes. Leaders similarly do not know how competitive officials perceive themselves to be vis-à-vis peers in competing for limited

promotion opportunities or whether officials themselves believe that they may soon be considered for promotion. Although leaders might anticipate greater incentives for officials to manipulate at some times and under certain conditions, they are unlikely to perfectly understand the exact costs and benefits.⁴¹

Whatever the case, it is not possible to interview leaders and confirm their uncertainty about the costs and benefits to officials from manipulation, or whether they consider corrected indicators to be less informative than indicators in the counterfactual case where officials do not have the option to manipulate. However, Premier Li Keqiang's suggestion that leaders use GDP-based indicators for 'reference only' supports the contention that much like boards, leaders consider corrected indicators less informative than indicators in the counterfactual case.

Less informative and noisier indicators complicate theories of performance-based promotion, including those that posit tournament-like competition. Under and over-estimation of bias on a case-by-case basis due to uncertainty over the reporting objective creates noisier rankings: Some officials rank higher than they should while some rank lower due to noise in corrected indicators rather than true differences in performance. In the extreme case where leaders are entirely uncertain about the reporting objective, leaders cannot glean information from corrected indicators and any rank ordering generated based on corrected indicators is random with respect to true performance.

⁴¹ Two exogenous factors that may affect the costs and benefits from manipulation faced by officials include the number of peers and their tenure in office. Under tournament-like competition, a higher number of peers—potential competitors—should increase the likelihood of officials engaging in manipulation. Similarly, officials who expect to be considered for promotion in the present and who might benefit from manipulation have a greater incentive to manipulate. Both Guo (2009) and Wallace (2014) exploit the later condition to demonstrate evidence of perverse incentives.

Conditioning promotion on noisy rankings in the extreme case is thus no different than promoting officials at random.

Although the extreme case is unlikely to arise, it is illustrative in that it demonstrates how greater uncertainty over the reporting objective makes corrected indicators less informative to leaders. In order to control the perverse incentives of officials to manipulate, leaders derive similar incentives to those faced by boards of directors to reduce the pay-for-performance sensitivity of compensation packages offered to managers. In the case of leaders, they can do this by reducing the weight they assign to indicators or rankings generated from indicators in making promotion decisions. Similar to the relationship between compensation and performance discussed in the finance and accounting literature, the relationship between promotion and performance in the presence of uncertainty over the reporting objective will be weaker than predicted in the absence of manipulation.

Implications for Tests of the Performance-Promotion Hypothesis

Manipulation and uncertainty over the reporting objective of officials has important implications for empirical studies that attempt to test the performance-promotion hypothesis. According to scholars who support theories of performance-based promotion, leaders condition the promotion of officials on performance vis-à-vis economic development indicators. Empirical studies attempt to test this hypothesis by demonstrating that officials who deliver rapid economic development—usually measured using sub-national per capita GDP growth rates—are more likely to receive promotions within the party and state hierarchy.

However, tests of this proposition yield inconsistent results that fail to meet standard thresholds for significance. While some scholars claim to demonstrate a significant and positive relationship between per capita GDP growth rates and the likelihood of promotion at various administrative levels (Li & Zhou 2005; Chen et al. 2005; Landry 2008), others have had difficulty replicating these results or demonstrate results that fall below standard levels of significance (Shih et al. 2012; Su et al. 2013). This has led some to question dominant theories of performance-based promotion. Moreover, this has led some scholars to claim that null or weaker than predicted results support alternative theories.⁴²

Rather than taking sides in this debate, this chapter offers an alternative explanation for null and weaker than predicted results. It follows from the argument developed in previous sections that manipulation and uncertainty over the reporting objective of officials reduces how informative indicators like per capita GDP growth rates are to leaders in evaluating the performance of officials. In response to less informative indicators, leaders reduce the weight they assign to indicators or rankings generated from indicators in making promotion decisions. The observable implication of this argument for empirical tests is a weaker than predicted relationship between promotion and indicators of performance such as reported per capita GDP growth rates.

Manipulation and uncertainty over the reporting objective therefore implies that indicators like per capita GDP growth rates are less reliable as measures of the underlying

⁴² For example, some scholars argue that the reason for null or weak results in support of the performance-promotion hypothesis is that leaders use tax-revenue growth rates rather than per capita GDP growth rates as a key indicator of performance (Su et al. 2012). The explanation advanced above does not take sides in such debates as it may apply regardless of the indicator or dimension of performance that leaders might emphasize as important to subordinate officials.

construct of performance than is implicitly assumed in current studies. As uncertainty over the reporting objective increases, the observed relationship between promotion and indicators of performance should become increasingly attenuated to reflect increased noise and a lower signal-to-noise ratio in corrected indicators as a measure of the true performance of officials. This issue does not arise in the absence of the option to manipulate. This issue also does not arise in the presence of manipulation but absent uncertainty over the reporting objective.

By ignoring manipulation and the uncertainty of leaders over the reporting objective of officials, scholars overlook an important construct validity issue in empirical studies of the performance-promotion hypothesis. In the presence of uncertainty over the reporting objective, manipulation makes indicators worse measures of performance and thus a poor operationalization of the underlying theoretical construct of performance. Moreover, this construct validity issue suggests that tests of the performance-promotion hypothesis will be biased towards type II error—incorrect failure to reject the null hypothesis—and will improperly conclude ‘performance’ does not influence the likelihood of promotion in error.

Discussion and Implications

To briefly summarize, this chapter offers an argument to make sense of manipulation. Manipulation of indicators is a rational and strategic response by sub-national officials to incentives created by top-down accountability institutions. Leaders are aware of these incentives and respond by correcting indicators based upon rational expectations about the magnitude of bias introduced by officials. However, leaders’

uncertainty over the reporting objective of officials reduces how informative indicators are to leaders by increasing noise in corrected indicators and lowering the signal-to-noise ratio of indicators as a measure of officials' true performance in office. As a consequence, indicators are less informative to leaders and leaders will reduce the weight that they assign to corrected indicators in making decisions such as whom to promote within the party and state hierarchy.

In terms of debate and the two puzzles in the introduction, this chapter suggests that performance does indeed matter but that perverse incentives for officials to manipulate indicators limit the degree to which leaders can promote officials based upon indicators of their performance. Manipulation is therefore not trivial. Although officials cannot mislead leaders with rational expectations on average, manipulation nonetheless undermines the utility of indicators to leaders in generating information about the relative performance of officials by making them less reliable. Seemingly puzzling results of empirical studies of the performance-promotion hypothesis that proxy for performance using indicators like per-capita GDP growth rates are thus less puzzling than at first glance.

CHAPTER 5: CONCLUSION

To conclude, poor air quality can teach scholars quite a bit about politics in the contemporary People's Republic of China as well as about politics under authoritarian regimes more generally. While the social welfare costs of poor air quality in China's rapidly expanding urban centers are substantial, until recently very few observers expected that improving poor air quality would become an important political issue for the leaders of the ruling Chinese Communist Party regime. The core chapters of this dissertation exploit this opportunity to offer a number of potential contributions to literature on contemporary Chinese politics as well as to literature on authoritarian politics more broadly.

The first core chapter's contribution to the literature on Chinese politics is to empirically confirm long-standing claims regarding perverse incentives created by top-down accountability institutions like the CCP cadre evaluation system. Although other scholars have attempted to empirically confirm a causal link between institutions and manipulation, their tests produce either null results or suffer from issues of observational equivalence. This chapter plausibly overcomes these issues of observational equivalence and is able to link evidence of manipulation back to variation in institutional incentives, something that has not yet been achieved in the empirical literature. This is an important piece of information for scholars seeking to understand one of the key institutions that underpins continued CCP rule.

Confirming a causal link also allows this chapter to open the door to further research on the relationship between top-down accountability institutions and manipulation. In particular, this chapter raises a number of puzzling questions for prevailing theories regarding the relationship between performance and promotion in Chinese politics. These questions matter for the broader literature on authoritarian politics in that three decades of spectacular development in China have led some to argue that China's leaders have somehow gotten it right with respect to rule through top-down accountability institutions. This is in contrast to the failure of their peers in the Soviet Union and Eastern Europe. However, this argument hinges upon the assumption that leaders can and do promote officials based upon performance vis-à-vis indicators. The ability of officials to act upon perverse incentives, to manipulate indicators, and thereby potentially confound leaders' efforts to generate information about their performance raises uncomfortable questions for scholars and pundits who make such sweeping claims.

The second core chapter's contribution to the literature on Chinese politics is to demonstrate the ability of actors outside of the party and state to set the public agenda. In doing so, these actors are also able to set the media agenda as well as influence leaders and the policy agenda. Scholars of Chinese politics increasingly recognize the ability of actors outside of the party and state to influence the policy agenda and the policy-making process using a variety of strategies (Wang 2008; Mertha 2009). However, it is difficult to demonstrate a causal effect of their actions on the policy agenda. Although this chapter cannot demonstrate causality, it is able to demonstrate a necessary condition for causality in the temporal order of changes in the public, media, and policy agendas implied by agenda setting theory.

This chapter's contribution to literature on authoritarian politics is to suggest a non-institutional channel through which actors formally excluded from the policy-making process under an authoritarian regime can shift the policy agenda in their favor. Of course, responsiveness on the part of leaders is not evidence of an alternative, bottom-up accountability channel. It is ultimately in the threat of potential civil unrest and instability that leaders respond to the demands actors outside of the party and state by making policy concessions and not through non-electoral and informal bottom-up accountability institution such as solidary groups (Tsai 2007) which leaders derive incentives to respond by adjusting the policy agenda. Nevertheless, it is worth noting that actors outside of the regime attempt to shift the policy agenda in their favor by first setting the public agenda, suggesting the importance of public opinion even under authoritarianism.

The third core chapter's contribution to the literature on Chinese politics is to offer a substantively and theoretically motivated solution to the puzzling results produced by direct tests of the performance-promotion hypothesis. Unlike arguments offered by both scholars who support theories of performance-based promotion and promotion based upon the factional affiliation of individual officials, the argument at the heart of this chapter embraces substantive knowledge of manipulation by officials to explain these results. This explanation not only takes into account leaders' awareness of manipulation, but also available information about how leaders respond to manipulation and incorporate information gleaned from indicators into their decisions. The key result is that as uncertainty over the reporting objective of officials increases, indicators become less informative to leaders as a measure of the performance of officials. Subsequently, leaders reduce the weight that they assign to indicators in promotion decisions thus reducing the

strength of the relationship between ‘performance’ as measured by indicators and promotion. In terms of the current literature, this is a unique explanation for the null and weak results in support of this critical hypothesis.

The third core chapter’s contribution to literature on authoritarian politics thus trails the initial chapter by answering the questions raised its empirical results. It suggests that performance does indeed matter and that leaders are still able to incorporate information gleaned from indicators into promotion decisions. Yet it also suggests the rarely recognized limits of top-down accountability institutions. If leaders are uncertain about the reporting objective of officials, indicators will be less reliable as measures of performance than is often assumed. Scholars and pundits who argue that China’s leaders somehow got it right and that promotion within the modern party and state hierarchy is meritocratic fail to recognize the implications of uncertainty for how meritocratic these institutions can actually be. For example, if ‘objective’ indicators of performance are unreliable, then do leaders not have to rely more upon subjective indicators or assessments to evaluate the performance of officials? Further developing and exploring the implications of uncertainty provides a potentially fruitful avenue for research into the problems of rule top-down accountability institutions. However, this dissertation leaves such tasks to future work.

APPENDIX A

Appendix A contains all Figures and Tables mentioned in the text of Chapter 2.

Figures and Tables are organized sequentially beginning with Figures and then followed by Tables.



Figure A1: Map of Chongqing and Chongqing's County-Level Jurisdictions

Note: The left panel indicates Chongqing municipality's location within the People's Republic of China in grey. The right panel indicates jurisdictions that were assigned targets under the Blue Sky Action campaign in grey. The english names of these jurisdictions are as follows: Yuzhong District, Dadukou District, Jiangbei District, Shapingba District, Jiulongpo District, Nan'an District, Ba'nán District, Beibei District, and Yubei District. Additionally, the Beibei New District was assigned targets following the 2008 revision of the campaign's enforcement plan.

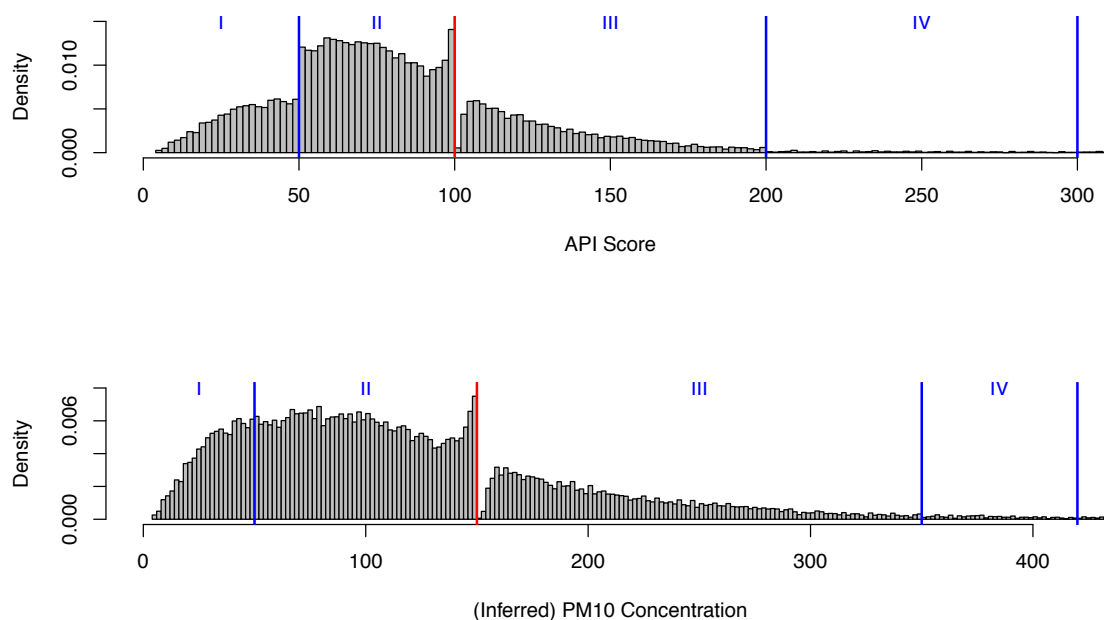


Figure A2: Histogram of Station-Level Air Quality Data Reported from Beijing

Note: The upper panel depicts the histogram of daily, station-level API Scores reported by the Beijing Environmental Protection Bureau for 2008-2011. The lower panel depicts the histogram of inferred PM10 concentrations after transforming API Scores above into their underlying PM10 equivalents. The vertical blue lines indicate where the relationship between PM10 concentrations and API Scores experiences a piecewise linear transformation. The vertical red line indicates the location of the Blue Sky cutoff whereby observations to the left of the cutoff qualify as Blue Sky days and observations to the right of the cutoff do not qualify.

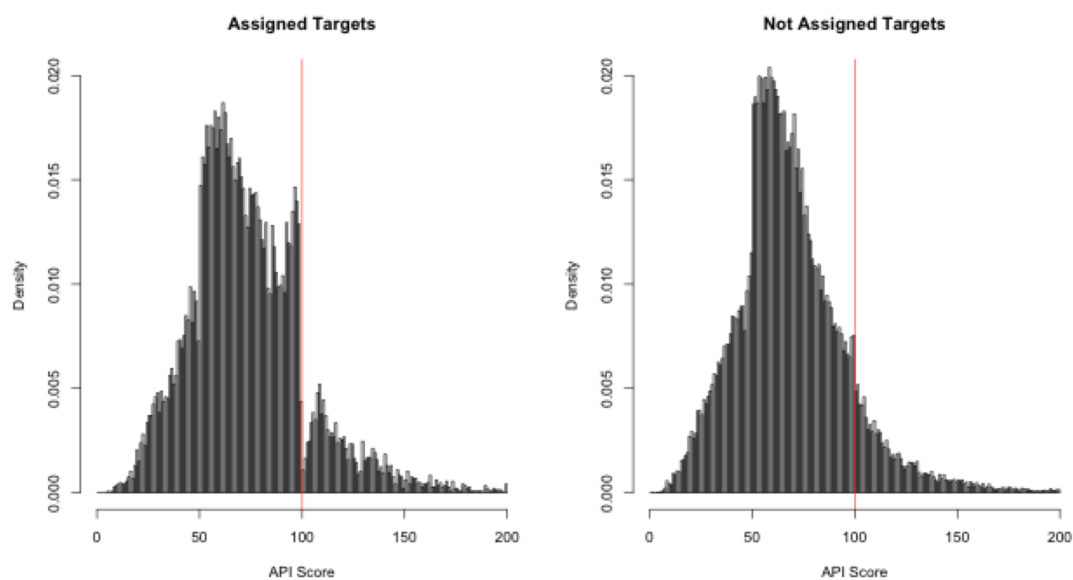


Figure A3: Histograms of County-Level API Scores by Target Assignment Status Reported from Chongqing

Note: Histograms are plotted using available county-level, daily API scores for all 40 county-level jurisdictions in Chongqing using bin-size = 1. Data covers 2008 through 2012 and was collected directly from the official website of the Chongqing Environmental Monitoring Center (CQEMC).

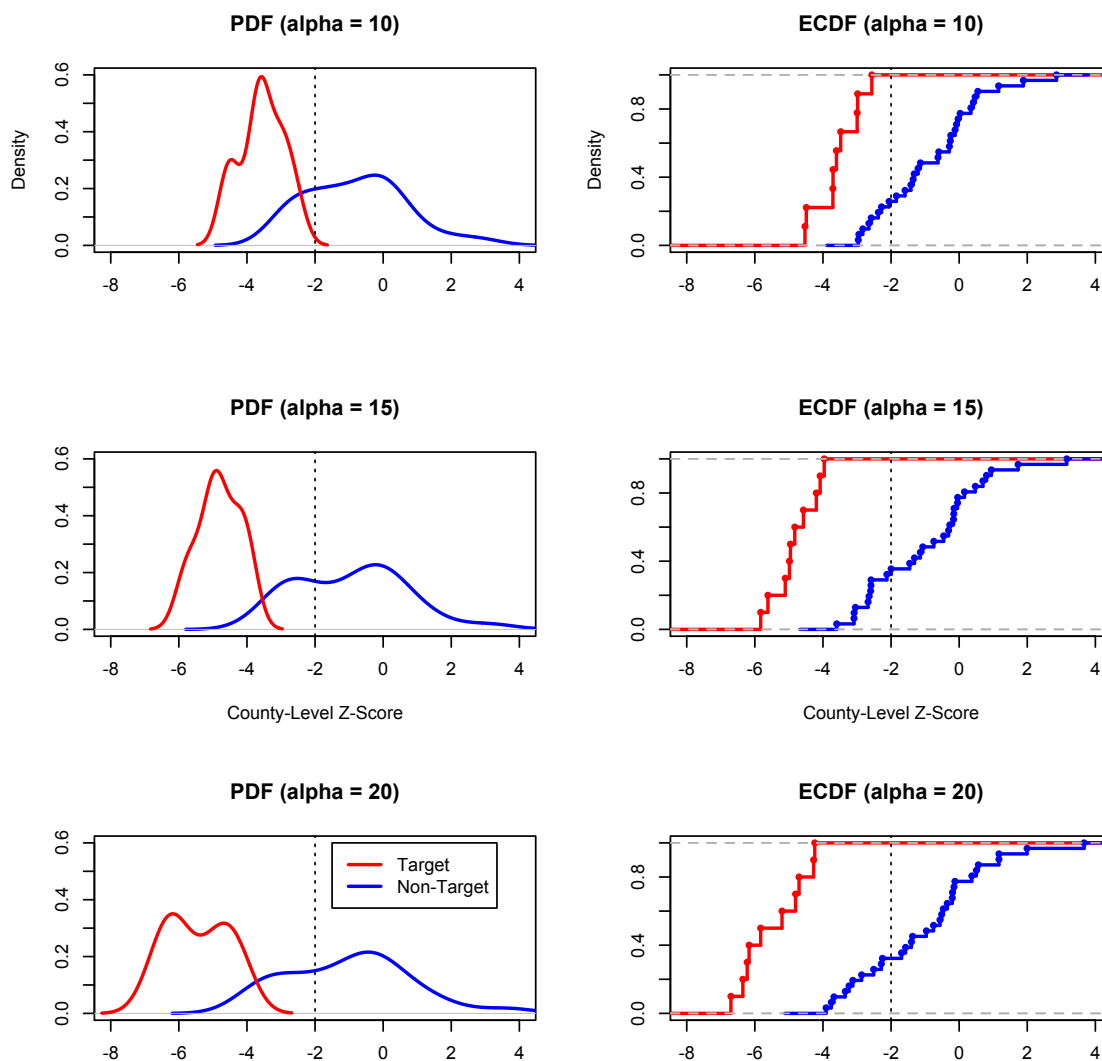


Figure A4: Probability Density Functions and Empirical Cumulative Density Functions of County-Level Z-Scores by Target Assignment Status

Note: Plots contained in the first column depict the probability density functions (PDFs) of county-level Z-Scores according to target assignment status. Plots contained in the second column depict the corresponding empirical cumulative density functions (ECDFs) of county-level Z-Scores according to target assignment status. Plots are ordered top to bottom in the figure according to the alpha used to estimate Z-Scores.

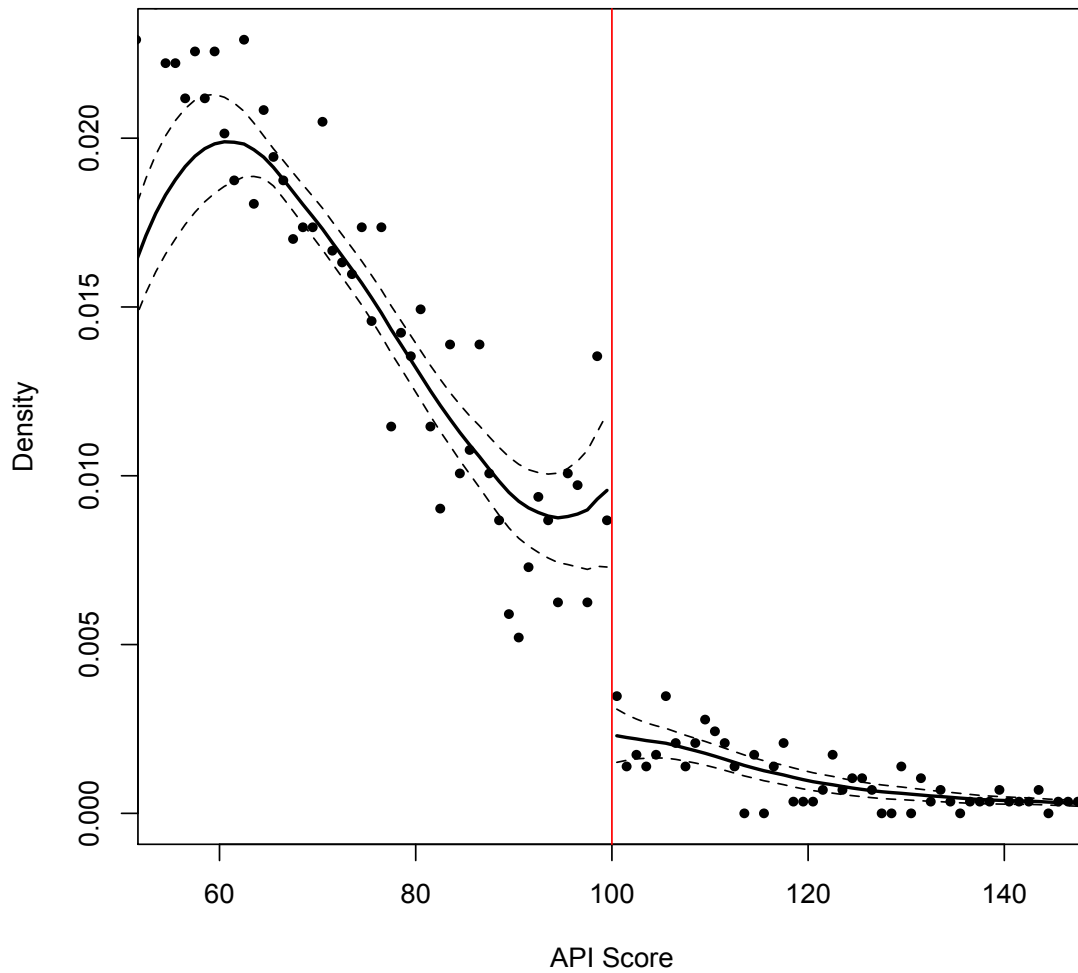


Figure A5: McCrary (2008) Test for Evidence of Post-Bo Xilai Manipulation

Note: The frequency distribution is constructed using a subset of county-level, daily API scores reported from county-level jurisdictions that were assigned targets and that were reported between the day following the purge of Bo Xilai (16 March 2012) and the last calendar day of the Blue Sky Action campaign (31 December 2012). The McCrary (2008) test was conducted using bin-size = 1 and bandwidth = 15 (i.e. $\alpha = 15$). The estimated log difference in height is -1.441 with a standard error of 0.244. The Z-Score is -5.907 (p-value < 0.001).

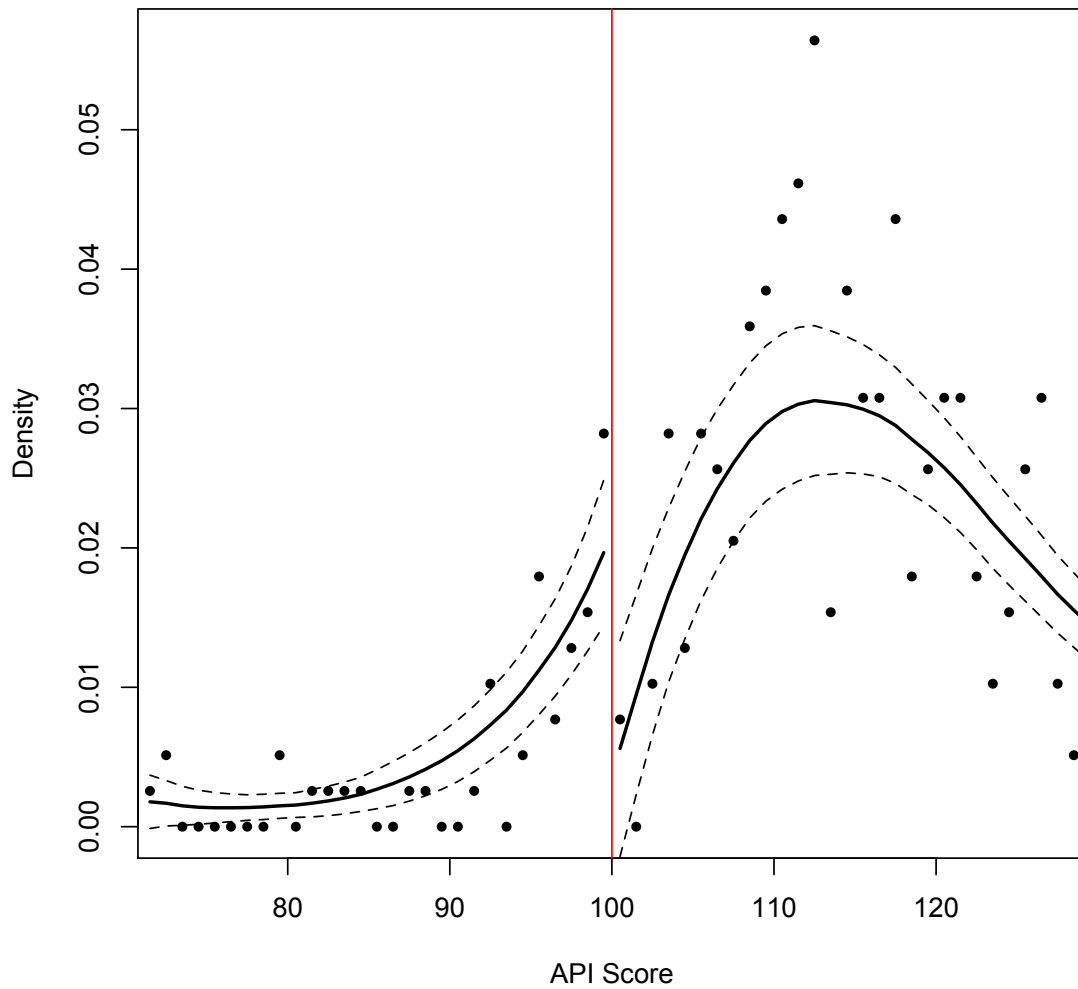


Figure A6: McCrary (2008) Test for Evidence of Collusion

Note: The frequency distribution is constructed using a subset of county-level, daily API scores reported from county-level jurisdictions that were assigned targets. The subset was drawn based from days in which the reported municipal-level, daily API score was greater than 111 and less than 120, thereby failing to meet the Blue Sky cutoff by at least 10 points. The McCrary (2008) test was conducted using bin-size = 1 and bandwidth = 15 (i.e. $\alpha = 15$). The estimated log difference in height is -1.757 with a standard error of 0.517. The Z-Score is -3.399 (p-value < 0.001).

Table A1: Summary Statistics on County-Level Air Quality by Target Assignment Status

	Assigned Targets		Not Assigned Targets	
	Mean (SD)	Valid Obs.	Mean (SD)	Valid Obs.
Daily API Score	73.445 (28.143)	14620	69.155 (27.074)	42329
Type	1.000	14620	1.000	42329
I	0.176	2580	0.208	8814
II	0.702	10256	0.687	29071
III1	0.105	1542	0.092	3913
III2	0.016	227	0.012	487
IV1	0.001	12	0.001	22
IV2	0.000	3	0.000	17
V	0.000	0	0.000	5
Predominant Monitored Pollutant*	1.000	12040	1.000	33515
PM ₁₀	0.966	11635	0.877	29377
SO ₂	0.033	397	0.123	4133
NO ₂	0.001	8	0.000	5
PM ₁₀ Concentration ($\mu\text{g}/\text{m}^3$)	113.444 (49.316)	11635	107.033 (47.811)	29377

Note: Information on the predominant monitored pollutant responsible for daily API scores is only reported for Type II and above days (API score > 50). PM10 Concentration is inferred based upon the subset of observations which both qualify as Type II or above (API score > 50) as well as the predominant monitored pollutant is PM10.

Table A2: Difference in Proportions of Suspect Samples by Target Assignment Status

Classification Threshold	Target Assignment Status	Classification Status		Difference-in-Proportions
		Suspect	Non-Suspect	
95% (Z-Score ≤ -1.960)	Target	9 (100.0%)	0 (0.0%)	Diff = 0.66 p-value < 0.001
	Non-Target	11 (35.5%)	20 (64.5%)	
99% (Z-Score ≤ -2.576)	Target	9 (100.0%)	0 (0.0%)	Diff = 0.71 p-value < 0.001
	Non-Target	9 (29.0%)	22 (71.0%)	
99.9% (Z-Score ≤ -3.291)	Target	8 (89.0%)	1 (11.0%)	Diff = 0.73 p-value < 0.001
	Non-Target	2 (6.5%)	29 (93.5%)	

Note: Thresholds for classifying individual, county-level Z-Scores as suspect of non-suspect are based upon the lower bound of the 95%, 99%, and 99.9% CI for a standard normal curve, respectively. The far right column contains p-values generated by a simple difference in proportions test for the proportion of samples classified as suspect by target assignment status.

Table A3: Summary Statistics for County-Level Jurisdictions by Target Assignment Status

Variable	Target Assignment Status									
	Assigned Targets					Not Assigned Targets				
	Mean	Min	Max	SD	n	Mean	Min	Max	SD	n
Per Capita GDP	3.076	1.152	7.921	1.351	54	1.261	0.321	7.211	0.879	186
Population Density	0.521	0.046	3.265	0.948	54	0.037	0.006	0.135	0.024	186
PM10 Concentration	0.100	0.064	0.133	0.015	54	0.082	0.027	0.124	0.017	186
SO2 Concentration	0.051	0.031	0.087	0.014	54	0.039	0.005	0.112	0.021	186
NO2 Concentration	0.040	0.016	0.072	0.012	54	0.027	0.004	0.052	0.010	185

Note: Annual, per capita GDP is inflation adjusted and expressed in 10,000 Yuan at year 2000 value. Population density is calculated by dividing the resident population by administrative area. Data used to calculate annual, per capita GDP and population density were collected from Chongqing Statistical Yearbooks (2005-2012). Annual, average PM10, SO2, and NO2 concentrations were collected from State of the Environment Reports published annually by the Chongqing EPB. The 2005 and 2006 reports omit this data, resulting in the lower n.

Table A4: Firth Logistic Regression of County-Level Classification Status on Target Assignment Status

Variables	Classification Status				Instrumenting for PM10	Instrumenting for Target Assignment Status	
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Target Assignment Status	3.523*** (1.575)	5.418*** (2.309)	5.278*** (2.129)	4.825** (2.046)	4.686** (1.938)	3.968*** (1.964)	5.317*** (2.506)
Log Population Density		-0.761 (0.527)	-0.337 (0.686)	-0.589 (0.858)	-0.501 (0.769)		-0.515 (0.929)
Log Per-Capita GDP			-1.059 (1.228)	-1.466 (1.688)	-1.254 (1.534)		-1.331 (1.699)
PM10 Concentration				0.001 (0.047)	-0.032 (0.068)		-0.005 (0.047)
SO2 Concentration				0.004 (0.027)	0.000 (0.027)		0.002 (0.027)
NO2 Concentration				0.089 (0.074)	0.091 (0.067)		0.080 (0.071)
Constant	-0.578 (0.374)	-3.278† (1.932)	-1.671 (2.542)	-5.140 (5.614)	-1.939 (6.530)	-0.631† (0.382)	-4.141 (5.578)
AIC	-10.458	-10.068	-8.617	-4.103	-3.97	-5.135	1.172
BIC	-8.769	-6.69	-3.55	6.03	6.163	-3.446	11.306
Likelihood Ratio Test	12.458	14.068	14.617	16.103	15.97	11.613	14.587
p-value	0.000	0.001	0.002	0.013	0.014	0.001	0.024
Valid Obs.	40	40	40	40	40	40	40

Note: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, † $p < 0.1$. Classification status assumes a value of one where county-level jurisdictions report air quality data classified as suspect and a value of zero where such data is classified as non-suspect.

Table A5: OLS Regression of County-Level Classification Status on Target Assignment Status

Variables	Classification Status				Instrumenting for PM10	Instrumenting for Target Assignment Status	
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Target Assignment Status	0.645*** (0.164)	0.860*** (0.236)	0.928*** (0.240)	0.871** (0.250)	0.933*** (0.257)	0.645*** (0.175)	0.871** (0.283)
Log Population Density		-0.107 (0.085)	-0.010 (0.114)	-0.054 (0.135)	-0.072 (0.125)		-0.050 (0.140)
Log Per-Capita GDP			-0.274 (0.216)	-0.036 (0.273)	-0.310 (0.251)		-0.339 (0.282)
PM10 Concentration				-0.001 (0.008)	-0.008 (0.010)		-0.002 (0.009)
SO2 Concentration				0.000 (0.005)	-0.001 (0.005)		0.000 (0.005)
NO2 Concentration				0.019 (0.013)	0.020 (0.011)		0.019 (0.013)
Constant	0.355*** (0.078)	-0.022 (0.310)	0.354 (0.427)	-0.178 (0.892)	0.315 (0.962)	0.355*** (0.080)	-0.104 (0.922)
AIC	50.35	50.68	50.94	53.57	52.82	51.82	55.99
BIC	55.41	57.43	59.38	67.08	66.33	56.88	69.5
Log Likelihood	-22.17	-21.34	-20.47	-18.79	-18.41	-22.91	-20
Valid. Obs.	40	40	40	40	40	40	40

Note: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, † $p < 0.1$. Classification status assumes a value of one where county-level jurisdictions report air quality data classified as suspect and a value of zero where such data is classified as non-suspect.

Table A6: OLS Regression of County-Level Z-Scores on Target Assignment Status

Variables	Z-Scores				Instrumenting for PM10	Instrumenting for Target Assignment Status	
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Target Assignment Status	-3.375*** (0.471)	-4.104*** (0.896)	-4.270*** (0.924)	-3.955*** (0.933)	-4.218*** (0.962)	-3.399*** (0.669)	-3.937*** (1.072)
Log Population Density		0.364 (0.324)	0.125 (0.440)	0.529 (0.505)	0.538 (0.467)		0.509 (0.530)
Log Per-Capita GDP			0.672 (0.833)	0.501 (1.019)	0.464 (0.937)		0.422 (1.068)
PM10 Concentration				-0.003 (0.031)	0.027 (0.037)		0.000 (0.033)
SO2 Concentration				0.017 (0.018)	0.019 (0.018)		0.016 (0.019)
NO2 Concentration				-0.084 [†] (0.048)	-0.092* (0.040)		-0.084 (0.050)
Constant	-1.038** (0.294)	0.241 (1.178)	-0.680 (1.644)	2.646 (3.330)	0.297 (3.594)	-1.033** (0.306)	2.314 (3.495)
AIC	156.78	157.44	158.72	158.91	158.29	159.17	162.59
BIC	161.84	164.2	167.17	172.42	171.8	164.23	176.1
Log Likelihood	-75.38	-74.72	-74.36	-71.45	-71.14	-76.58	-73.3
Valid Obs.	40	40	40	40	40	40	40

Note: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, [†] $p < 0.1$. County-level Z-Scores with negative and smaller values indicate a greater likelihood of the discontinuity and evidence of manipulation. Where Z-Scores are less than -1.96, they can be classified as ‘suspect’ at the 95% CI.

Table A7: Sensitivity Analysis using Different Classification Thresholds and Z-Scores Calculated using Alternative Bandwidths

Classification Threshold	API Scores			Inferred PM ₁₀ Concentrations		
	<i>alpha</i> = 10	<i>alpha</i> = 15	<i>alpha</i> = 20	<i>alpha</i> = 10	<i>alpha</i> = 15	<i>alpha</i> = 20
95% (Z-Score ≤ -1.960)	3.850*** (1.596)	3.523*** (1.575)	3.661*** (1.577)	3.695*** (1.593)	3.389*** (1.573)	3.661*** (1.577)
99% (Z-Score ≤ -2.576)	2.976*** (1.048)	3.807*** (1.579)	4.128*** (1.587)	2.793*** (1.039)	3.807*** (1.579)	3.961*** (1.583)
99.9% (Z-Score ≤ -3.291)	2.468*** (0.974)	4.203*** (1.149)	3.307*** (1.048)	2.468*** (0.974)	4.203*** (1.149)	3.307*** (1.048)

Note: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, † $p < 0.1$; The dependent variable for models 1-3 is the county-level Z-Score produced by the McCrary (2008) test using API scores, conducted at the Blue Sky cutoff (API score > 100). In all models, bin size = 1 and bandwidth varies accordingly. The dependent variable for models 4-6 are county-level Z-Scores generated by the McCrary (2008) test using inferred PM10 concentrations, conducted at the Blue Sky cutoff (PM10 Concentration > 150 microgram / cubic meter). In all models, bin size = 2 and bandwidth varies accordingly. Covariates for Log Per Capita GDP, Log Population Density, as well as logged concentrations of monitored pollutants were included in regressions but suppressed for presentation. The regression contains 40 observations due to the exclusion of Beibei New District for lack of covariates.

APPENDIX B

Appendix B contains all Figures and Tables mentioned in the text of Chapter 3.

Figures and Tables are organized sequentially beginning with Figures and then followed by Tables.

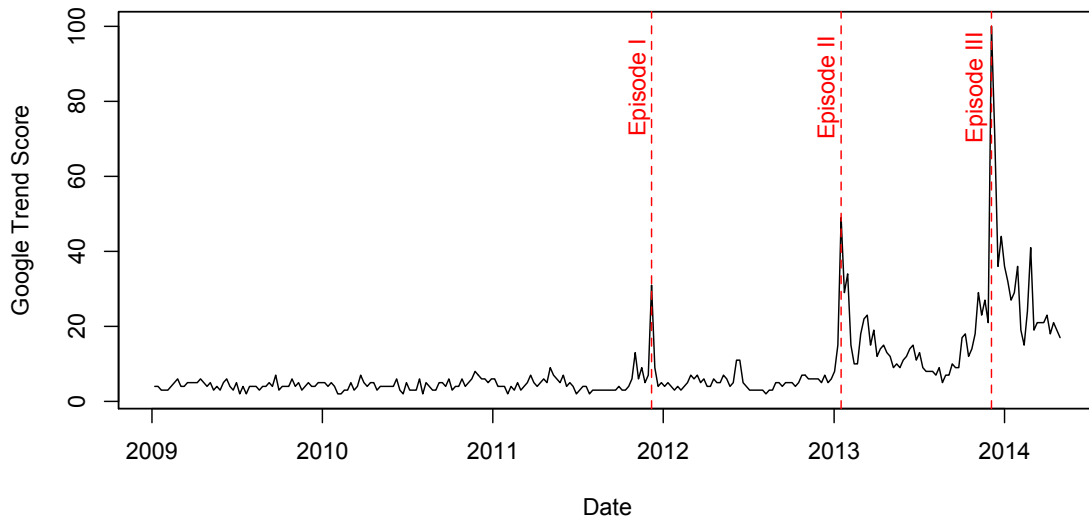


Figure B1: Google Trends Score for Keyword 'Air Quality'

Note: Google Trends Scores depicted in the above figure are based on the relative volume of searches for the keyword 'air quality' in Chinese and by users of Google Search with IP addresses registered in the People's Republic of China. Google Trends Scores are standardized, range in value between 0 and 100, and are reported on a weekly level for the period from 2009 through the beginning of 2014.

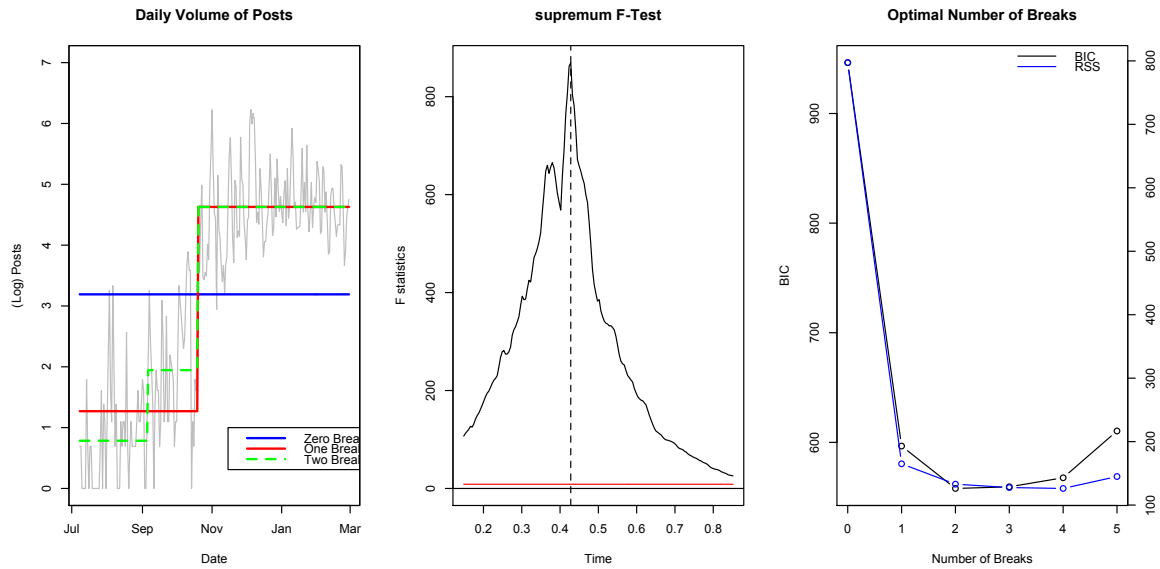


Figure B2: Analysis of Structural Breaks in Daily Volume of Weibo Posts from Verified Sina Weibo Accounts

Note: The left panel depicts the daily volume of Weibo posts from verified Sina Weibo accounts that contain the keyword ‘PM2.5’ as well as the estimated intercept-only models for zero, one, and two breaks. The center panel depicts the results of the supremum F-Test for structural breaks where the supremum of calculated F-Statistics indicates a likely break. The right panel depicts the results from the Information Criteria based test for the optimal number of breaks where the minimum BIC indicates the optimal number of breaks.

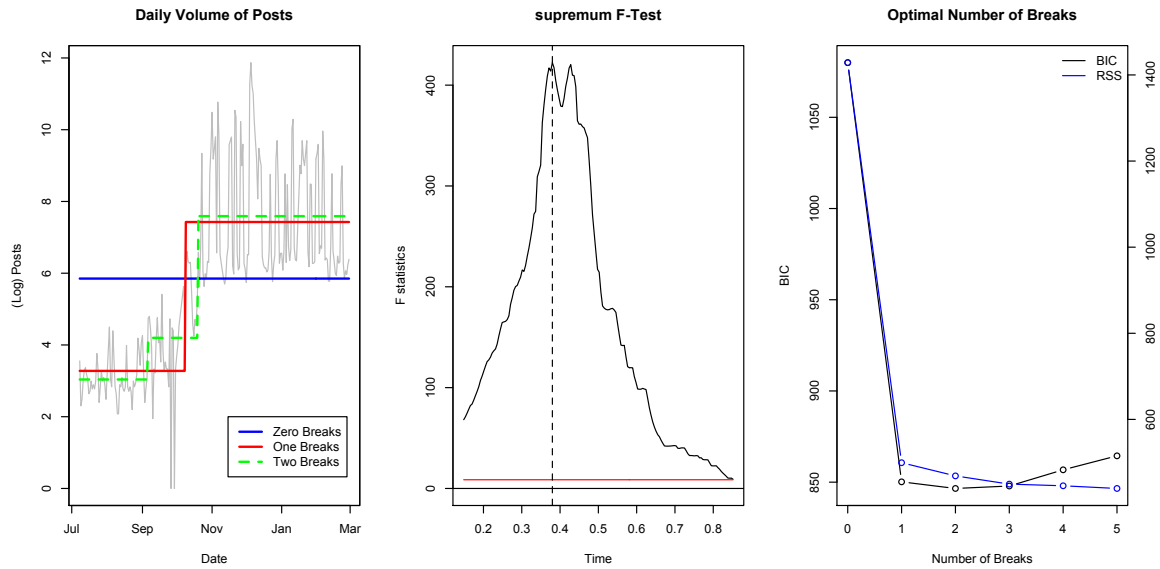


Figure B3: Analysis of Structural Breaks in Daily Volume of Weibo Posts from Non-Verified Sina Weibo Accounts

Note: The left panel depicts the daily volume of Weibo posts from non-verified Sina Weibo accounts that contain the keyword 'PM2.5' as well as the estimated intercept-only models for zero, one, and two breaks. The center panel depicts the results of the supremum F-Test for structural breaks where the supremum of calculated F-Statistics indicates a likely break. The right panel depicts the results from the Information Criteria based test for the optimal number of breaks where the minimum BIC indicates the optimal number of breaks.

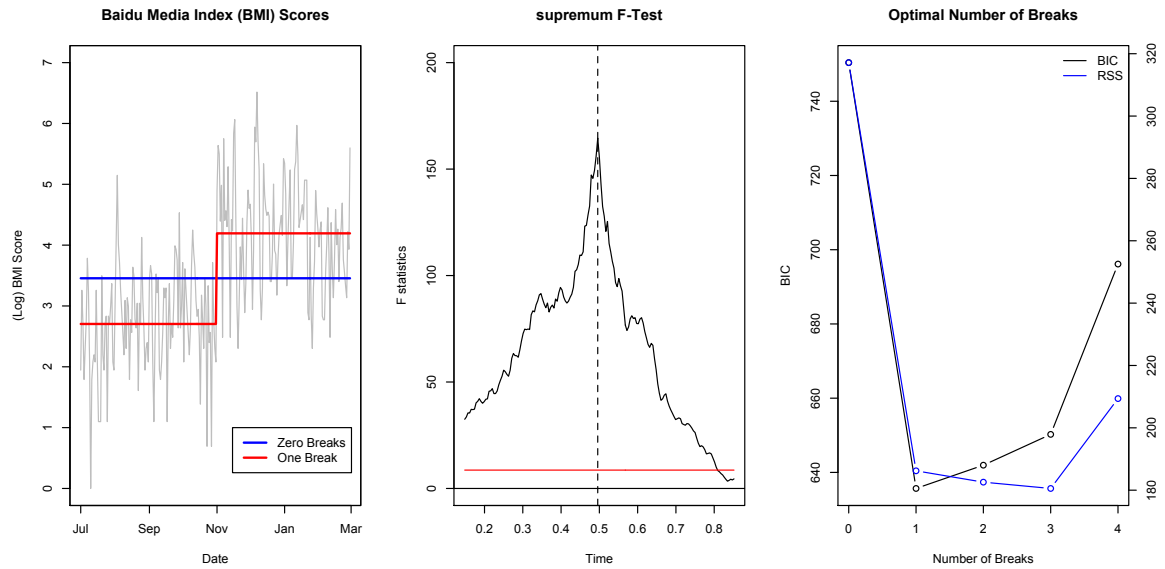


Figure B4: Analysis of Structural Breaks in Daily Baidu Media Index Scores

Note: The left panel depicts the daily Baidu Media Index Score for the keyword ‘Air Quality’ as well as the estimated intercept-only models for zero and one break. The center panel depicts the results of the supremum F-Test for structural breaks where the supremum of calculated F-Statistics indicates a likely break. The right panel depicts the results from the Information Criteria based test for the optimal number of breaks where the minimum BIC indicates the optimal number of breaks.

Table B1: Summary Statistics for Sina Weibo Posts and Baidu Media Index

Variable	Mean	SD	Min	Max	Valid Obs.
Sina Weibo Posts (Verified)	71.55	94.32	0	506	229
Sina Weibo Posts (Non-Verified)	4543.78	13630.39	0	142870	229
Baidu Media Index	58.26	83.95	0	675	229
(log) Sina Weibo Posts (Verified)	3.09	1.88	0	6.23	229
(log) Sina Weibo Posts (Non-Verified)	5.71	2.50	0	11.87	229
(log) Baidu Media Index	3.46	1.16	0	6.52	229

Note: The daily volume of Sina Weibo Posts is the daily count of posts containing the keyword 'PM2.5' for series collected from both verified and non-verified accounts. Baidu Media Index (BMI) are generated based upon articles containing the keywords 'Air Quality'. In order to account for zero values in the original time series, logged values for each series were calculated by taking the natural logarithm of one plus the unscaled daily value of the series.

Table B2: Intercept-Only Structural Break Models

Segment	(Log) Posts from Verified Accounts			(Log) Posts from Non-Verified Accounts			(Log) BMI Score	
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Intercept	3.191*** (0.124)	1.133*** (0.101)	0.785*** (0.102)	5.850*** (0.165)	3.277*** (0.159)	3.038*** (0.191)	3.455*** (0.174)	2.704*** (0.099)
Segment 2		3.319*** (0.129)	1.161*** (0.157)		4.150*** (0.202)	1.161*** (0.295)		1.490*** (0.152)
Segment 3			3.844*** (0.122)			4.554*** (0.229)		
AIC	939.54	628.08	533.33	1073.11	834.43	821.94	743.48	619.82
BIC	946.41	638.38	547.07	1079.97	844.74	835.68	750.41	630.21
logLik	-467.77	-311.04	-262.67	-534.55	-414.22	-406.97	-369.74	-306.91
Valid Obs.	229	229	229	229	229	229	229	229

Note: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table B3: Temporal Order of Breaks in Public and Media Attention

Agenda	Operationalization	Estimated Break	Confidence Interval	
			Lower Bound	Upper Bound
Public Agenda	<i>(Log) Sina Weibo Posts (Non-Verified)</i> Keyword: 'PM2.5'	8 October 2011	2 October 2011	9 October 2011
	<i>(Log) Sina Weibo Posts (Verified)</i> Keyword: 'PM 2.5'	19 October 2011	18 October 2011	21 October 2011
Media Agenda	<i>(Log) Baidu Media Index Scores</i> Keyword: 'Air Quality'	31 October 2011	26 October 2011	11 November 2011
	<i>(Log) Baidu Media Index Scores</i> Keyword: 'PM2.5'	2 November 2011	—	—
Policy Agenda	First Announcement of Reforms by Premier Wen Jiabao	15 November 2011	—	—

Note: Confidence intervals for breaks are estimated at the 95% CI level using robust standard errors. Estimated breaks for the (log) daily volume of Weibo posts from both verified and non-verified Sina Weibo accounts are based upon intercept-only models using one break. The estimated break for the (log) daily Baidu Media Index score is similarly based upon an intercept-only model using one break.

REFERENCES

- Andrews, Steven Q. 2008. "Inconsistencies in Air Quality Metrics: 'Blue Sky' days and PM10 Concentrations in Beijing." *Environmental Research Letters*, 3: 1-14.
- Associated Press. 2010. "US Embassy: Beijing Air Quality is 'Crazy Bad'." Associated Press, 19 November 2010.
- Bai, Jushan. 1997a. "Estimation of a Change Point in Multiple Regression Models." *Review of Economics and Statistics* 79:821-856.
- Bai, Jushan. 1997b. "Estimating multiple breaks at one time." *Econometric Theory* 13: 315-352.
- Bai, Jushan and Pierre Perron. 1998. "Estimating and Testing Linear Models with Multiple Structural Changes." *Econometrica*, 66(1): 47-78.
- Bai, Jushan and Pierre Perron. 2003. "Computation and Analysis of Multiple Structural Change Models." *Journal of Applied Econometrics*, 18(1): 1-22.
- Brännlund, Runar and Karl-Gustaf Löfgren. 1996. "Emissions Standards and Stochastic Waste Load." *Land Economics*, 72(2): 218-230.
- Bell, Daniel A. 2012. "What America's Flawed Democracy Could Learn from China's One-Party Rule." *The Christian Science Monitor*, 24 July 2012.
- Bernstein, Thomas P. and Xiabo Lü. 2003. *Taxation Without Representation in Contemporary Rural China*. Cambridge: Cambridge University Press.
- Boix, Carles and Milan Svolik. 2013. "The Foundations of Limited Authoritarian Government: Institutions, Commitment, and Power-Sharing in Dictatorships." *Journal of Politics*, 75(2): 300-316.
- Bolton, Patrick, Jose Scheinkman, and Wei Xiong. 2006. "Executive Compensation and Short-Termist Behavior in Speculative Markets." *Review of Economic Studies*, 73(3): 577-610.
- Buckley, Chris. 2013. "Crackdown on Bloggers Is Mounted by China." *The New York Times*, 10 September 2013.
- Bueno de Mesquita, Bruce, Alastair Smith, Randolph M. Siverson, James D. Morrow. 2003. *The Logic of Political Survival*, MIT Press.

- Burns, J.P. 1989. *The Chinese Communist Party's Nomenklatura System: A Documentary Study of Party Control of Leadership Selection, 1979-1984*. New York: M.E. Sharpe.
- Cai, Yongshun. 2000. "Between State and Peasant: Local Cadres and Statistical Reporting in Rural China." *The China Quarterly*, 163(3): 783-805.
- Cai, Yongshun. 2004. "Irresponsible State: Local Cadres and Image Building in China." *Journal of Communist Studies and Transition Politics*, 20(4): 20-41.
- Campbell, Donald T. 1976. "Assessing the Impact of Planned Social Change." The Public Affairs Center, Dartmouth College. Occasional Paper Series, Paper #8.
- Chan, Alex. 2007. "Guiding Public Opinion through Social Agenda-Setting: China's media policy since the 1990s." *Journal of Contemporary China*, 16(53): 547-559.
- Chen, Ye, Hongbin Li and Li-an Zhou. 2005. "Relative Performance Evaluation and the Turnover of Provincial Leaders in China." *Economics Letters*, 88: 421-425.
- Chen, Yuyu, Avraham Ebenstein, Michael Greenstone, and Hongbin Li. 2013. "Evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River policy." *Proceedings of the National Academy of Sciences*, 113(32).
- Chen, Yuyu, Ginger Zhe Jin, Naresh Kumar, and Guang Shi. 2012. "Gaming in Air Pollution Data? Lessons from China." *The B.E. Journal of Economic Analysis and Policy*, 12(3).
- China Council for International Cooperation on Environment and Development. 2012. "CCICED 2011 Annual Report." *CCICED Annual Reports*.
- Chow, Gregory. 2006. "Are Chinese Official Statistics Reliable." *CESifo Economic Studies*, 52(2): 396-414.
- Chow, Gregory. 1960. "Tests of Equality Between Sets of Coefficients in Two Linear Regressions." *Econometrica*, 28(3): 591-605.
- Cody, Edward. 2007. "Text Messages Giving Voice to Chinese." *The Washington Post*, 28 June 2007.
- Dearing, J.W. and Everett M. Rogers. 1988. "Agenda Setting Research: Where Has It Been, Where Is It Going?" *Communications Yearbook* 11: 555-594.
- Dittmer, Lowell. 1995. "Chinese Informal Politics." *The China Journal*, 34: 1-34.
- Dittmer, Lowell, and Yushan Wu. 1995. "The Modernization of Factionalism in Chinese Politics." *World Politics*, 47(4): 467-494.

- Economy, Elizabeth C. 2007. "The Great Leap Backward? The Costs of China's Environmental Crisis." *Foreign Affairs*, 86(5): 38-59.
- Economy, Elizabeth C. 2004. *The River Runs Black: The Environmental Challenge to China's Future*. Cornell University Press: Ithaca, NY.
- Edin, Maria. 2003. "State Capacity and Local Agent Control in China: CCP Cadre Management from a Township Perspective." *The China Quarterly*, 173: 35-52.
- Egorov, Georgy, Sergei Guriev and Konstantin Sonin. 2009. "Why Resource-poor Dictators Allow Freer Media: A Theory and Evidence from Panel Data." *American Political Science Review*, 103(4): 645-668.
- Egorov, Georgy, and Konstantin Sonin. 2011. "Dictators and Their Viziers: Endogenizing the Loyalty-Competence Trade-Off." *Journal of the European Economic Association*, 9(5): 903-930.
- Fernald, John, Israel Malkin, and Mark Spiegel. 2013. "On the Reliability of Chinese Output Figures." *Federal Reserve Bank of San Francisco Economics Letter*, 8: 1-4.
- Fewsmith, Joseph. 1994. *Dilemmas of Reform in China: Political and Economic Debate*. Armonk, NY: M.E. Sharpe.
- Firth, David. 1993. "Bias reduction of maximum likelihood estimates." *Biometrika*, 80(1): 27-38.
- Fischer, Paul E. and Robert E. Verrechia. 2000. "Reporting Bias." *The Accounting Review*, 75(2): 229-245.
- Fudenberg, Drew and Jean Tirole. 1986. "A 'Signal-Jamming' Theory of Predation." *The Rand Journal of Economics*, 17(3): 366-376.
- Goldman, Eitan, and Steve L. Slezak. 2006. "An Equilibrium Model of Incentive Contracts in the Presence of Information Manipulation." *Journal of Financial Economics*, 80(3): 603-26.
- Guo Gang. 2009. "China's Local Political Budget Cycles." *American Journal of Political Science*, 53(3): 621-632.
- Heinze, Georg and Michael Schemper. 2002. "A Solution to the Problem of Separation in Logistic Regression." *Statistics in Medicine*, 21: 2409-2419.
- Heinze, Georg and Meinhard Ploner. 2004. "A SAS macro, S-PLUS library and R package to perform logistic regression without convergence problems." Section of Clinical Biometrics, Department of Medical Computer Sciences, Medical University of Vienna. Technical Report 2/2004.

- Holmstrom, Bengt. 1999. "Managerial Incentive Problems: A Dynamic Perspective." *The Review of Economic Studies*, 66(1): 169-182.
- Holz, Carsten A. 2014. "The Quality of China's GDP Statistics." *China Economic Review*, In Press.
- Holz, Carsten A. 2008. "China's 2004 Economic Census and 2006 Benchmark Revision of GDP Statistics: More Questions than Answers?" *The China Quarterly*, 193: 150-163.
- Holz, Carsten A. 2006. "China's Reform Period Economic Growth: How Reliable Are Angus Maddison's Estimates?" *The Review of Income and Wealth*, 52(1): 85-119.
- Holz, Carsten A. 2004. "China's Statistical System in Transition: Challenges, Data Problems, and Institutional Innovations." *The Review of Income and Wealth*, 50(3): 381-409.
- Holz, Carsten A. 2003. "'Fast, Clear and Accurate': How Reliable Are Chinese Output and Economic Growth Statistics?" *The China Quarterly*, 173: 122-163.
- Holz, Carsten A. 2002. "Institutional Constraints on the Quality of Statistics in a Developing and Transitional Economy: The Case of China." *China Information*, 16(1): 25-67.
- Jensen, Michael C. and Kevin J. Murphy. 1990. "Performance Pay and Top-Management Incentives." *Journal of Political Economy*, 98(2): 225-264.
- Kiewiet, D. Roderick and Matthew D. McCubbins. 1991. *The Logic of Delegation: Congressional Parties and the Appropriations Process*. Chicago: The University of Chicago Press.
- 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(2): 326-343.
- Koech, Janet and Jian Wang. 2012. "China's Slowdown May Be Worse Than Official Data Suggest." *Federal Reserve Bank of Dallas Economic Letter*, 7(8): 1-4.
- Kosicki, Gerald M. 1993. "Problems and Opportunities in Agenda-Setting Research." *Journal of Communication* 43(2): 100-127.
- Kung, James Kai-Sing and Shuo Chen. 2011. "The Tragedy of the Nomenklatura: Career Incentives and Political Radicalism during China's Great Leap Famine." *American Political Science Review*, February: 1-19.
- Kuran, Timur. 1991. "Now out of Never: The Element of Surprise in the Eastern European Revolutions." *World Politics*, 44(1): 7-48.

- Lagerkvist, Johan. 2005. "The Rise of Online Public Opinion in the People's Republic of China." *China: An International Journal*, 3(1): 119-130.
- Lake, David A. and Matthew A. Baum. 2001. "The Invisible Hand of Democracy: Political Control and the Provision of Public Services." *Comparative Political Studies*, 34(6): 587-621.
- Landry, Pierre F. 2008. *Decentralized Authoritarianism in China: The Communist Party's Control of Local Elites in the Post-Mao Era*. Cambridge: Cambridge University Press.
- Li, Eric X. 2013. "The Life of the Party: The Democratic Future Begins in China." *Foreign Affairs*, January/February 2013.
- 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: 1743-1762.
- Li Jing. 2014. "People in Maoming and Shenzhen stage fresh protests over planned PX plant." *South China Morning Post*, 5 April 2014.
- Li, Wanxi and Hon S. Chan. 2009. "Clean Air in Urban China: The Case of Inter-Agency Coordination in Chongqing's Blue Sky Program." *Public Administration and Development*, 29: 55-67.
- Lim, Louisa. 2013. "Beijing's 'Airpocalypse' Spurs Pollution Controls, Public Pressure." *National Public Radio*
- Lohmann, Susanne. 1994. "The Dynamics of Informational Cascades: The Monday Demonstrations in Leipzig, East Germany, 1989-1991." *World Politics*, 47(1): 42-101.
- Lorentzen, Peter. 2014. "China's Strategic Censorship." *American Journal of Political Science*, 58(2): 402-414.
- Lorentzen, Peter. 2013. "Regularized Rioting: Permitting Public Protest in an Authoritarian Regime." *Quarterly Journal of Political Science*, 8(2): 127-158.
- Maddison, Angus. 2006. "Do Official Statistics Exaggerate China's GDP Growth? A Reply to Carsten Holz." *The Review of Income and Wealth*, 52(1): 121-126.
- Maskin, Eric, Yingyi Qian, and Chenggang Xu. 2000. "Incentives, Information and Organizational Form." *Review of Economic Studies*, 67: 359-378.
- Matus, Kira, Kyung-Min Nam, Noelle E. Selin, Lok N. Lamsal, John M. Reilly, and Sergey Paltsev. 2012. "Health Damages from Air Pollution in China." *Global Environmental Change*, 22(1): 55-66.

- McCombs, Maxwell and Donald Shaw. 1972. "The Agenda Setting Function of the Mass Media." *Public Opinion Quarterly*, 36: 176-187.
- McCombs, Maxwell. 2004. *Setting the Agenda: Mass Media and Public Opinion*. Cambridge University Press: New York.
- McCrary, Justin. 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." *Journal of Econometrics*, 142: 698-714.
- McCubbins, Matthew D. and Thomas Schwartz. 1984. "Congressional Oversight Overlooked: Police Patrols versus Fire Alarms." *American Journal of Political Science*, 28(1): 165-179.
- Merli, M. Giovanna. 1998. "Underreporting of Births and Infant Deaths in Rural China: Evidence from Field Research in One County in Northern China." *The China Quarterly*, 155(3): 637-655.
- Merli, M. Giovanna and Adrian E. Raferty. 2000. "Are Births Underreported in Rural China? Manipulation of Statistical Records in Response to China's Population Policies." *Demography*, 37(1): 109-126.
- Mertha, Andrew. 2009. "'Fragmented Authoritarianism 2.0': Political Pluralization in the Policy Process." *China Quarterly*, 200(): 995-1012.
- Ministry of Environmental Protection of the People's Republic of China. 2014. "Ministry of Environmental Protection and 31 Provinces sign 'Air Pollution Control Responsibility Contracts'." MEP News Release. Available at: http://www.mep.gov.cn/gkml/hbb/qt/201401/t20140107_266123.htm
- Minzner, Carl F. "China's Turn Against Law." *American Journal of Comparative Law*, 59(4): 935-984.
- Nathan, Andrew J. 1973. "A Factionalism Model for CCP Politics." *The China Quarterly*, 53: 34-66.
- Nathan, Andrew J. and Kellee S. Tsai. 1995. "Factionalism: A New Institutional Restatement." *The China Journal*, 34: 157-192.
- Odgaard, Ole. 1990. "Inadequate and Inaccurate Chinese Statistics: The Case of Rural Private Enterprises." *China Information*, 5(3): 29-38.
- Orlik, Tom. 2014. "Reform at China's National Bureau of Statistics under Ma Jiantang 2008-2013." *China Economic Review*, In Press.
- Qian, Gang and David Bandurski. 2011. "China's Emerging Public Sphere: The Impact of Media Commercialization, Professionalism, and the Internet in an Era of

- Transition.” In *Changing Media, Changing China*, Ed. Susan L. Shirk. New York: Oxford University Press.
- Qian, Yingyi, Barry Weingast, and Gabriella Montinola. 1995. “Federalism, Chinese Style: The Political Basis for Economic Success in China.” *World Politics*, 48(1): 50-81.
- Qian, Yingyi, Gerard Roland, and Chenggang Xu. 2006. “Coordination and Experimentation in M-Form and U-Form Organizations.” *Journal of Political Economy*, 114(2): 366-402.
- Qian, Yingyi, Gerard Roland, and Chenggang Xu. 2007. “Coordinating Changes in Transition Economies.” In *The Economics of Transition: The Fifth Nobel Symposium in Economics*, Eds. Erik Bergloff and Gerard Roland, New York: Palgrave MacMillan, pp. 518-546.
- Qian, Yingyi, and Chenggang Xu. 1993. “Why China’s Economic Reforms Differ: The M-Form Hierarchy and Entry/Expansion of the Non-state Sector.” *Economics of Transition*, 1(2): 135-170.
- Rawski, Thomas G. 2001. “What is Happening to China’s GDP Statistics?” *China Economic Review*, 12: 347-354.
- Rock, Michael T. 2002. “Improving Environmental Performance of China’s Cities.” In *Pollution Control in East Asia: Lessons from Newly Industrializing Economies* by Michael T. Rock, pp. 82-114.
- Shih, Victor, Christopher Adolph, and Mingxing Liu. 2012. “Getting Ahead in the Communist Party: Explaining the Advancement of Central Committee Members in China.” *American Political Science Review*, 106(1): 166-187.
- Shih, Victor, Wei Shan, and Mingxing Liu. 2010. “Gauging the Elite Political Equilibrium in the CCP: A Quantitative Approach Using Biographical Data.” *The China Quarterly*, 201: 79-103.
- Shirk, Susan. 1993. *The Political Logic of Economic Reform in China*. Berkeley, CA: University of California Press.
- Shirk, Susan L. 2011. “Changing Media, Changing China.” In *Changing Media, Changing China*, Eds Susan L. Shirk. New York: Oxford University Press.
- Sinton, Jonathan E. 2001. “Accuracy and Reliability of China’s Energy Statistics.” *China Economic Review*, 12(4): 373-383.
- Stein, Jeremy C. 1989. “Efficient Capital Markets, Inefficient Firms: A Model of Myopic Corporate Behavior.” *The Quarterly Journal of Economics*, 104(4): 655-669.

- Su, Fubing, Ran Tao, Lu Xi, Ming Li. 2012. "Local Officials' Incentives and China's Economic Growth: Tournament Thesis Reexamined and Alternative Explanatory Framework." *China and World Economy*, 20(4): 1-18.
- Svolik, Milan W. 2012. *The Politics of Authoritarian Rule*. Cambridge: Cambridge University Press.
- Tsai, Lily L. 2008. "Understanding the Falsification of Village Income Statistics." *The China Quarterly*, 106: 805-826.
- Tsai, Lily L. 2007. "Solidary Groups, Informal Accountability, and Local Public Goods Provision in Rural China." *American Political Science Review*, 101(2): 355-372.
- Travers, S. Lee. 1982. "Bias in Chinese Economic Statistics: The Case of a Typical Investigation." *The China Quarterly*, 91(3): 478-485.
- Wallace, Jeremy. 2014. "Juking the Stats? Authoritarian Information Problems in China." *British Journal of Political Science*.
- Wang, Alex. 2013. "The Search for Sustainable Legitimacy: Environmental Law and Bureaucracy in China." *Harvard Environmental Law Review*, 37: 365-440.
- Wang, Shaoguang. 2008. "Changing Models of China's Policy Agenda Setting." *Modern China* 34(1): 56-87.
- Watts, Jonathan. 2010. "Twitter Gaffe: US Embassy Announces 'Crazy Bad' Beijing Air Pollution." *The Guardian Environment Blog*, 19 November 2010.
- Watts, Jonathan. 2011. "Tens of Thousands Protest Against Chemical Plant in Northern China." *The Guardian*, 14 August 2011.
- Wee, Sui-Lee. 2013. "China to Invest \$277 Billion to Curb Air Pollution: State Media." Reuters. Published on 24 July 2013.
- Whiting, Susan H. 2004. "The Cadre Evaluation System at the Grass Roots: The Paradox of Party Rule." In *Holding China Together: Diversity and National Integration in the Post-Deng Era*, Eds. Barry Naughton and Dali Yang.
- Whiting, Susan H. 2001. *Power and Wealth in Rural China: The Political Economy of Institutional Change*. Cambridge University Press: NY.
- Randt, Clark T. 2007. "Fifth Generation Star Li Keqiang Discusses Domestic Challenges, Trade Relations with Ambassador." WikiLeaks. WikiLeaks cable: 07BEIJING1760. Retrieved 2014-03-14.
- Wong, Edward. 2011. "Outrage Grows over Air Pollution and China's Response." *New York Times*, 6 December 2011.

- Xiao, Kezhou and Brantly Womack. 2014. "Distortion and Credibility within China's Internal Information System." *Journal of Contemporary China*, 23(88): 680-697.
- Xiao, Qiang. 2011a. "The Battle for the Chinese Internet." *Journal of Democracy*, 22(2): 47-61.
- Xiao, Qiang. 2011b. "The Rise of Online Public Opinion and its Political Impact." In *Changing Media, Changing China*, Ed. Susan L. Shirk. New York: Oxford University Press.
- Xinhua. 2012. "China Revises Air Quality Standards, Including PM 2.5." Xinhua Net. Accessed http://news.xinhuanet.com/english/china/2012-02/29/c_131438402.htm
- Xu, Chenggang. 2011. "The Fundamental Institutions of China's Reforms and Development." *Journal of Economic Literature*, 49(4): 1076-1151.
- Xu, Chenggang. 2012. "Institutional Foundations of China's Structural Problems." In *The Chinese Economy: A New Transition*. eds. Masahiko Aoki and Jinglian Wu. London: Palgrave MacMillan. pp. 272-289.
- Yang, Gonghuan, Yu Wang, Yixin Zheng, George F. Gao, Xiaofeng Liang, Maigeng Zhou, Xia Wan, Shicheng Yu, Yuhong Jiang, Moshen Naghavi, Theo Vos, Haidong Wang, Alan D. Lopez, Christopher J.L. Murray. 2013. "Rapid Health Transition in China, 1990-2010: Findings from the Global Burden of Disease Study 2010." *The Lancet*, 381(9882): 1987-2015.
- Yang, Jisheng. 2012. *Tombstone: The Great Leap Famine, 1958-1962*. New York: Farrar, Straus and Giroux.
- Zeileis, Achim, Freidrich Leisch, Kurt Hornik, and Christian Kleiber. 2002. "strucchange: An R package for testing for structural change in linear regression models." *Journal of Statistical Software*, 7(2): 1-38.
- Zhou, Xuegang. 2010. "The Institutional Logic of Collusion among Local Governments in China." *Modern China*, 36(1): 47-78.