Essays in Environmental Economics

Thomas Stoerk

TESI DOCTORAL UPF / ANY 2017

DIRECTORS DE LA TESI
Dr. Antoine Dechezleprêtre i Dr. Humberto Llavador

DEPARTAMENT D'ECONOMIA I EMPRESA

Universitat Pompeu Fabra Barcelona

Für meine Eltern.

Acknowledgments

So this is it. My thesis. In reality, the thesis is but a small glimpse of what training for a PhD entails. 90% happens behind the scenes. This section is for the 90%.

First and foremost, I am indebted to my advisors Antoine Dechezleprêtre and Humberto Llavador. Apart from their great skills in research, both are very fine human beings and I feel incredibly fortunate to have crossed their paths. Thank you!

My PhD has been split into two parts: After my course stage and initial research at Universitat Pompeu Fabra, I was generously adopted by the fantastic people at the LSE's Grantham Research Institute on Climate Change and the Environment. The research environments at both places have been energetic and enriching. At UPF, I'd like to thank José Apesteguia and Larbi Alaoui for initial conversations, Rosemarie Nagel for teaching me a great deal about transparent research and economic experiments, and to Gianmarco León, Giacomo Ponzetto, Ruben Enikolopov, Maria Petrova, and the members of the Applied group at UPF, as well as to Jeroen van den Bergh at UAB for their feedback on various stages of this dissertation. A special mention goes to Kurt Schmidheiny's truly excellent Econometrics course during the first year in Barcelona. In London, I've learned a great deal from conversations with Simon Dietz, Ralf Martin, Mirabelle Muûls, Roger Fouquet and the many researchers and visitors that pass through the Grantham Research Institute. What unites them all is a burning passion for transparent research on environmental issues.

Experience suggests that all people ever read from a dissertation is this section. So let me give one clear piece of advice: mental health problems in grad school are real, and they are much more frequent than you'd think. I am personally aware of two fellow students who died by suicide in just my own PhD peergroups. My advice to future grad students is to take your own mental health seriously, and to use mental health services at your school. And to faculty: lobby to make such services available at your school.

My PhD was split right in the middle between two great cities: Barcelona and London, and so I am fortunate to have two distinct groups of friends to thank. In Barcelona, a big thank you to Hrvoje, Vicky, Alain, Andrea, Patrick, Tom, Jagdish, Marco, Niklas, María, Cristina, Ana, Darren, Lorenzo, Miguel, Tomás, Ricardo, Francisco and office 20.159. Most importantly to

Michael, who I was very fortunate to call my friend. In London, great times were had with Felix, Victoria, Friedrich, Isa, Davide, Elizabeth, Louise, François, Stephan, Maria, Maria, Martin, Tobi, Ale, Gregor, Hélia, Ashley, Frikk, Greer, Paola, Damien, Ara, Carole, Alice, Baran and everyone at GRI. A special thank you goes to Philipp and Beni. And finally, a heartfelt thank you to the many friends who have hosted me on my nomadic odyssee over the years.

I also want to give credit to the continuous support from Marta Araque and Laura Agustí, without whom finished dissertations at UPF would be few and far between. I am further indebted to Jo Miller and Stuart Rodgers at the Grantham Research Institute for making my stays in London a breeze. Thank you also to the open source communities that maintain LaTex and Python as well as Inkscape, Zotero and wget.

Finalment, a la Núria, per tot.

Abstract

This thesis consists of three chapters that investigate environmental policy questions from an empirical point of view. Chapter 1 examines the trustworthiness of official air pollution data sources for Beijing when compared to similar data from the US Embassy in Beijing. Using a statistical regularity, I find that the official data likely suffered from misreporting until the end of 2012. From 2013 onwards, however, misreporting appears to have stopped. Chapter 2 evaluates China's main air pollution control policy to study the effects of environmental regulation when institutions are weak. I find that the policy was ultimately successful in reducing air pollution, but that those effects only set in once the Chinese government started appropriate air pollution monitoring. Moreover, I quantify the efficiency of different policy instruments to control air pollution in China and find that - in contrast to the United States - a market-based solution and a technology mandate for scrubbers are nearly identical. Finally, Chapter 3 studies whether nudges can help consumers align intention and action when choosing their electricity contract. Using a survey experiment, we find that only a default nudge had a statistically and economically significant effect on consumers' decision to contract renewable energy.

Resum

Aquesta tesi consisteix en tres estudis que investiguen problemes relacionats amb la política de medi ambient des d'un punt de vista empíric. El capítol 1 examina fins a quin punt són fiables les dades sobre la contaminació de l'aire a Beijing quan es comparen amb dades semblants de l'ambaixada dels EUA. Mitjançant l'ús d'una regularitat estadística, proporciono evidència que les dades oficials segurament van ser manipulades fins a finals del 2012. A partir del 2013, però, les dades semblen indicar que es va posar fi a aquesta manipulació. El capítol 2 avalua la política xinesa més important duta a terme per frenar la contaminació de l'aire i pretén estudiar els efectes de la regulació del medi ambient en un context d'institucions febles. Demostro que aquesta política va aconseguir reduir la contaminació, però tan sols després que el govern xinès implementés el monitoreig adequat. A més, quantifico l'eficiència de diferents instruments polítics destinats a controlar la contaminació de l'aire a la Xina. Els resultats indiquen que - en contrast amb els EUA - pràcticament no hi ha diferències entre un instrument de mercat i l'ús prescriptiu de depuradores de gas. Finalment, el capítol 3 analitza si el fet d'introduir canvis en la informació pot ajudar els consumidors a seguir les seves intencions a l'hora de triar un contracte d'electricitat. A través d'un experiment en forma d'enquesta es demostra que només una preselecció té un efecte significatiu en els sentits estadístic i econòmic sobre la presa de decisió dels consumidors a l'hora de contractar energia renovable.

Preface

The main theme of this thesis is air pollution control in China. Economic growth in China has soared after Deng Xiaoping initiated China's economic reforms in 1978, And as has always been the case since the Industrial Revolution, an increase in economic activity has gone hand in hand with an increase in energy consumption (Greenstone, 2016). This relationship holds true for all industrialized economies around the globe; there is no exception. In China, energy consumption has meant the consumption of coal. Coal that was often of a low quality, burned in dated furnaces, and subject to little oversight meant rampant air pollution. As a result, air pollution levels in China in recent decades are comparable to London at the height of the Industrial Revolution.

This situation has started to change recently. In 2003, hydroengineer President Hu and geologist Prime Minister Wen assumed office. The 'scientific development' paradigm, which emphasized environmental protection alongside economic growth, started to substitute for economic growth. At the same time, increased living standards in the Eastern provinces and the availability of smartphones meant that the population started to pay attention to the problem of air pollution. Initially, politicians responded by misreporting improvements in air pollution, often declaring foggy days as *Blue Sky Days*. At the same time, however, the central government in Beijing started to realize the gravity of China's pollution problems. China's policy priorities are set out in Five-Year Plans, and the 11th Five-Year Plan (2006-2010) was the first to include a total emissions control target that had high-level political backing. This thesis analyzes the transition in China's environmental governance for the topic of air pollution.

Chapter 1, *Statistical corruption in Beijing's air quality data has likely ended in 2012*, which I published in *Atmospheric Environment* in 2016, is my point of departure. In this Chapter, I assess the trustworthiness of the official air quality data in Beijing. Previous research had found that the official air quality data of the Beijing Municipal Environmental Protection Bureau often featured bunching around the politically important *Blue Sky Day* threshold: an excess of values just below the threshold and a lack of values just beyond it suggested misreporting for political reasons (Andrews, 2008a). The official response by the authorities was, however, that they could predict air pollution dynamics and control emergency responses with such accuracy as to explain the observed pattern (Cyranoski, 2009). Chapter 1 tests this claim by comparing the official data to air quality data measured by the US Embassy. The data are similar, but not identical, and I use a statistical regularity typically used in forensic economics that enables me to compare patterns in both datasets. I find that until 2012, the official data appears to be misreported. After an update to air quality standards at the end of 2012, however, misreporting of the official data appears to have stopped.

Chapter 2, *Compliance, Efficiency, and Instrument Choice: Evidence from Air Pollution Control in China*, forms the main body of this thesis. I evaluate China's main air pollution control policy to inform a wider question that is of interest not only to environmental economics, but to development economics more generally: how does regulation work when the institutional setting is weak? China is the ideal setting to answer this question, for a number of reasons. Firstly, when China passed its total emissions control target in the 11th Five-Year Plan in 2005, it lacked appropriate monitoring of SO_2 pollution. A Ministry of Environmental Protection did not exist, and SO_2 pollution monitoring relied on aggregate coal use statistics that were collected by the same provincial governments that were subject to the air pollution control targets. Halfway through the compliance period, this situation changed drastically when China installed continuous SO_2 emissions monitoring for the main polluters at the prefecture-level and created the Ministry of Environmental Protection in 2008, which was vested with the political power to collect pollution data without political interference. Secondly, NASA launched a satellite in late 2004 just before the start of the policy which provides an independent measure for SO_2 pollution in China. The satellite data allow me to study the effect of the air pollution control target on both reported and actual pollution. I find that ultimately the pollution control targets work, but only after the Chinese government started to appropriately monitor SO_2 pollution.

In the second part of Chapter 2, I quantify the efficiency of the command and control regulation by constructing a market-based benchmark allocation of reduction targets across provinces. This benchmark allocation is based on detailed SO_2 marginal abatement cost curves for each province in China. I find that despite the initial lack of environmental governance, China's air pollution control targets were surprisingly efficient. Compared to a market-based benchmark, I find an efficiency loss of just below 50%, which is at the lower end of what we know from more developed countries. Furthermore, I use the marginal abatement cost curves to compare the SO_2 control experience of China to the US. Surprisingly, I find that nearly all of the efficiency gains from trade of a market-based policy instrument could have been reaped by a technology mandate on scrubbers. In the US, the opposite is true.

Chapter 3, *From intention to action: Can nudges help consumers to choose renewable energy?*, co-authored with Katharina Momsen of Mannheim University, which we published in *Energy Policy* in 2014, closes the thesis. In this Chapter, we investigate the consumer side of electricity markets in shifting towards renewable energy. Specifically, we ask whether nudges - simple changes in the information structure when making a decision - can help electricity consumers align their intentions with their actions and choose the renewable energy contracts that consumers state they prefer. Using an original survey experiment, we test 9 different nudges from the behavioural economics literature. We find that only a default nudge has an effect on consumer choice whereas all other nudges are ineffective in our setting. A default nudge increases the share of individuals who choose renewable energy by 44.6% compared to the control group.

Contents

List of Figures xvi

List of Tables xvii

1 Statistical corruption in Beijing's air quality data has likely ended in 2012 3

- 1.1 Introduction . . . 4
- 1.2 Data . . . 5
- 1.3 Methods . . . 7
- 1.4 Results . . . 8
  - 1.4.1 An anomaly in the official air pollution data . . . 8
  - 1.4.2 Comparing the BMEPB and the US Embassy data . . . 10
- 1.5 Discussion . . . 13
- 1.6 Concluding Remarks . . . 15
- 1.A Data Sources . . . 17
- 1.B Definitions of the Air Quality Indices (AQIs) used by the Beijing Municipal Environmental Protection Bureau and the US Embassy . . . 19
- 1.C Robustness check: Analysis after dropping BMEPB observations from non-particulate matter AQIs . . . 21
- 1.D Robustness check: Figure 2 using BMEPB AQI breakpoints . . . 23
- 1.E Additional results for the Discussion section . . . 25
- 1.F Benford's Law for a generic air pollution dataset . . . 27

2 Compliance, Efficiency, and Instrument Choice: Evidence from Air Pollution Control in China 29

- 2.1 Introduction . . . 31
- 2.2 Related literature . . . 35
- 2.3 The SO_2 reduction policy in context . . . 39
- 2.4 Data . . . 42
  - 2.4.1 Data on pollution outcomes . . . 43

2.4.2 Data on reactions by the regulated provincial governments . . . 45
2.5 Empirical analysis . . . 47
2.5.1 Empirical strategy . . . 47
2.5.2 The effect of the SO_2 control policy on SO_2 pollution 49
2.5.3 The effect of monitoring . . . 59
2.5.4 Reactions by the regulated agents in response to the targets . . . 63
2.5.5 Interpretation of the empirical findings . . . 70
2.6 Instrument choice and gains from trade . . . 71
2.6.1 Construction of the marginal abatement cost curves . 71
2.6.2 The cost-efficient counterfactual allocation of SO_2 reduction targets . . . 74
2.6.3 A measure for marginal abatement benefits . . . 79
2.6.4 Gains from trade . . . 81
2.7 Concluding remarks . . . 83
2.A Additional figures . . . 87
2.B Marginal SO_2 abatement cost curves for all provinces in China 94

3 From intention to action: Can nudges help consumers to choose renewable energy **101**
3.1 Introduction . . . 102
3.2 Methods: An Original Experiment . . . 103
3.2.1 Setting . . . 103
3.2.2 The Nudges . . . 105
3.2.3 Recorded Data and Randomisation . . . 109
3.3 Results and Discussion . . . 111
3.3.1 Results . . . 111
3.3.2 Discussion . . . 113
3.4 Conclusions and Policy Implications . . . 117
3.A Original decision screens . . . 119
3.B Robustness Check: Probit and Logit Results . . . 123

Bibliography **125**

List of Figures

1.1 Histogram of air pollution levels (BMEPB data) 9
1.2 Histogram of air pollution levels (US Embassy data) 9
1.3 Histogram of aerosol optical density (AERONET data) . . . 10
1.4 Observed frequencies and Benford's Law (BMEPB data) . . 11
1.5 Observed frequencies and Benford's Law (US Embassy data) 11
1.6 Observed frequencies and Benford's Law (AERONET data) 11
1.7 Goodness-of-fit with Benford's Law over time (BMEPB data) 12
1.8 Goodness-of-fit with Benford's Law over time (US Embassy data) 13
1.C.1 Pollutant composition of the BMEPB dataset 21
1.C.2 Histogram of air pollution levels (Restricted BMEPB data) . 22
1.C.3 Observed frequencies and Benford's Law (Restricted BMEPB data) 22
1.D.1 Illustration of the different PM2.5 breakpoints 23
1.D.2 Histograms of air pollution levels (BMEPB data and US Embassy data) 24
1.E.1 The *Blue Sky Day* Anomaly in Beijing and the Comparison Cities over time 26
1.E.2 Histogram of air pollution levels and observed frequencies and Benford's Law (Restricted BMEPB data) 26
1.F.1 Observed frequencies and Benford's Law for a generic air pollution dataset 27
1.F.2 Observed frequencies and Benford's Law for many generic air pollution datasets 28

2.1 The SO_2 emission control targets 41
2.2 SO_2 pollution in China in January 2006 based on NASA satellite data 44
2.3 Example of a Government Work Report 46
2.4 Political attention to air pollution over time 47
2.5 Dynamic treatment effects: Official SO_2 emissions 52
2.6 Dynamic treatment effects: SO_2 satellite data 53

2.7 Dynamic treatment effects: SO_2 satellite data (Monthly) . . 55
2.8 Dynamic treatment effects: rhetorical compliance 64
2.9 Real compliance: Installation of desulfurization devices . . 67
2.10 Real compliance: Timing of desulfurization device installation 68
2.11 Real compliance: Shutdown of small thermal units 69
2.12 Marginal abatement cost curve examples 73
2.13 Consistency of SO_2 marginal abatement cost data and official data 74
2.14 SO_2 marginal abatement cost curve for China 75
2.15 The cost-efficient allocation of SO_2 reduction targets 76
2.16 Relation between MAC and SO_2 reduction targets 78
2.17 Marginal welfare impact of SO_2 abatement 82
2.A.1 The Great Recession and SO_2 reduction targets 87
2.A.2 Likely misreporting over time 88
2.A.3 Dynamic treatment effects: Likely misreporting 89
2.A.4 Political attention to air pollution over time (normalized by report length) 90
2.A.5 Relation between welfare impact and SO_2 reduction targets 91
2.A.6 Difference between cost-efficient and actual allocation . . . 92
2.A.7 Economic growth and air pollution in China 93

3.1 Decision Screen for the Control Group 104
3.2 Raw Choice Data 111
3.A.1 Decision Screen for Priming (Intention) 119
3.A.2 Decision Screen for Priming (Memory) 120
3.A.3 Decision Screen for Priming (Reassemble) 120
3.A.4 Decision Screen for Mental Accounting 120
3.A.5 Decision Screen for Framing (Gains) 121
3.A.6 Decision Screen for Framing (Losses) 121
3.A.7 Decision Screen for Decoy 121
3.A.8 Decision Screen for Social Norms 121
3.A.9 Decision Screen for Default 122

List of Tables

1.B.1 Breakpoints for Different Pollutants Before 31st December 2012 . . . 19
1.B.2 Breakpoints for Different Pollutants From 31st December 2012 Onwards . . . 20

2.1 Descriptive statistics . . . 50
2.2 The effect of the policy for the whole period (2006-2010) . 51
2.3 The effect of the policy for each year of the 11th Five-Year Plan (2006-2010) . . . 54
2.4 Heterogeneous treatment effects depending on initial pollution levels . . . 57
2.5 Controlling for marginal abatement cost and marginal welfare impact . . . 58
2.6 The effect of monitoring: Overview . . . 61
2.7 The effect of monitoring: Evidence . . . 62
2.8 Reactions by the regulated agents in response to the targets 65
2.9 Actual and counterfactual SO_2 target allocations . . . 77
2.10 Efficiency gains from trade . . . 83

3.1 Descriptive Statistics and Randomisation . . . 110
3.2 Average Treatment Effects . . . 112
3.3 Heterogeneous Treatment Effects . . . 115
3.B.1 Average Treatment Effects (Probit Specification) . . . 123
3.B.2 Average Treatment Effects (Logit Specification) . . . 124

Stoerk, T. (2016). Statistical corruption in Beijing's air quality data has likely ended in 2012, *Atmospheric Environment*, 127 (February 2016): 365-371.

Chapter 1

Statistical corruption in Beijing's air quality data has likely ended in 2012

Abstract

This research documents changes in likely misreporting in official air quality data from Beijing for the years 2008 to 2013. It is shown that, consistent with prior research, the official Chinese data report suspiciously few observations that exceed the politically important Blue Sky Day threshold, a particular air pollution level used to evaluate local officials, and an excess of observations just below that threshold. Similar data, measured by the US Embassy in Beijing, do not show this irregularity. To document likely misreporting, this analysis proposes a new way of comparing air quality data via Benford's Law, a statistical regularity known to fit air pollution data. Using this method to compare the official data to the US Embassy data for the first time, I find that the Chinese data fit Benford's Law poorly until a change in air quality measurements at the end of 2012. From 2013 onwards, the Chinese data fit Benford's Law closely. The US Embassy data, by contrast, exhibit no variation over time in the fit with Benford's Law, implying that the underlying pollution processes remain unchanged. These findings suggest that misreporting of air quality data for Beijing has likely ended in 2012.

1.1 Introduction

Ambient air pollution is the cause of a major health crisis in current-day China. Pollution from ambient particulate matter is the fourth most important health burden in China (Yang et al., 2013), and coal heating alone has been attributed to causing a loss of 5 life years (Chen, Ebenstein, et al., 2013). At the same time, political stakes are high: "Fix air pollution or 'bring your head'!" was the unequivocal message that Beijing's mayor received from the central government earlier this year (Chen, 2014). To address public pressure, the Beijing Municipal Environmental Protection Bureau (BMEPB) regularly publishes statistics on the city's air quality. These data, however, have not gained public trust as they often stand in stark contrast to measurements reported by the US Embassy in Beijing (Grammaticas, 2011; Le Monde, 2013; Wong, 2013).

And the public might have a point: the 'leaders make numbers' phenomenon has a tradition in China (Liu and Yang, 2009). Air quality data in China have been found to be suspicious in several cities until 2010 based on both a discontinuity test (Chen, Jin, et al., 2012; Ghanem and Zhang, 2014), and on a detailed analysis of histograms and the location of measurement stations for Beijing until 2007 (Andrews, 2008a). These studies suggest that misreporting is prevalent around the politically important Blue Sky Day threshold, possibly due to the strict enforcement of promotion criteria (Wang, 2013; Zheng and Kahn, 2013). The Blue Sky Day label designates days with an air quality index (AQI) of 100 or less, and the number of Blue Sky Days enters the performance assessment of local officials. The central government has, for instance, used the number of Blue Sky Days to create a national ranking of air quality in Chinese cities (Andrews, 2008b), and as one of the indicators for a city to qualify as a national environmental protection model city (Chen, Jin, et al., 2012). The number of Blue Sky Days in Beijing has also been widely reported by the media due to its salience. By reporting a higher number of Blue Sky Days, local officials can therefore improve their chances for a positive evaluation, while at the same time appeasing the public.

The present research adds to the literature by employing a systematic approach to identifying misreporting that goes beyond the analysis of a discontinuity. This improves on the literature in two aspects: Firstly, my method allows to compare the similar, but slightly different US Embassy data to the BMEPB data for the first time and thus distinguish misreporting from temporary measures such as factory closures on borderline Blue Sky Days. Secondly, I can detect changes in misreporting over time. My contribution additionally lies in a novel application of Benford's Law to compare two similar sets of air quality data to detect misreporting. While Benford's Law is known to fit air pollution in general, it is ex ante difficult to know whether it will fit a particular sample. In my application, I use a control dataset to establish whether Benford's Law fits the particular distribution of air pollution data. Given that the air quality datasets are sufficiently comparable, Benford's Law can then be used to test whether the suspicious dataset conforms to Benford's Law. This application of Benford's Law is particularly useful to detect likely misreporting even when air quality data come from different sources that are not comparable in levels. Furthermore, the method is easily scalable and extends naturally to other settings where two slightly different, yet comparable datasets measure air pollution. The increased availability of aerosol optical density (AOD) data, for example from satellite imagery, offers such settings. To illustrate this use, I show that my analysis holds also when using AOD data rather than the US Embassy data as the control dataset.

1.2 Data

This research uses a new database of Beijing ambient air quality data that was collected and merged from three different sources (see Appendix A for the data sources). The first source is the BMEPB. The BMEPB has been reporting daily air quality measurements from 31st December 2007 until 18th March 2013. The second source comes from the measurements that the Embassy of the United States in Beijing provide via their Twitter account. These data report hourly and run from 8th April 2008 3pm to 31st December 2013 11pm, with a missing period from 6th November 2008 to 7th February 2009. The third source is the AERONET measurement station in Beijing, who report aerosol optical density (AOD) measurements from 26th March 2008 to 31st December 2013, with several measurements during different times of each day.

While following the same calculation procedure, the AQIs of the BMEPB and the US Embassy aggregate different pollutants and use slightly different break points to convert pollutant concentrations into AQI values (see Appendix B for the full definitions). The following paragraphs explain the exact composition of both datasets and provide evidence to show that both data sources are comparable because they reflect the same underlying pollution processes.

The BMPEB measures three pollutants until 30th December 2012 (PM_{10}, NO_2, SO_2) and, following a change in its AQI, six pollutants from 31st December 2012 onwards ($PM_{2.5}$, PM_{10}, NO_2, SO_2, O_3, CO). The US Embassy, by contrast, measures only $PM_{2.5}$ throughout. While the differing definitions might suggest that the datasets might not be sufficiently comparable due to the possibility of the BMEPB measurements being based on pollutants other than particulate matter, in practice this concern is muted by the clear prevalence of particulate matter-driven observations in the BMEPB data. The BMEPB data report the main pollutant for all observations with an AQI exceeding 50, and, as shown in Appendix 1.B, the AQI of a given observation is only determined by the concentration of the main pollutant on that day, that is by the pollutant concentration defined as most harmful to human health compared to the other pollutant concentrations measured on the same day.

Using the information on the main pollutant, I extract the share of particulate matter-driven observations as 96.93% for the full BMEPB sample, and as 98.03% until 30th December 2012 and as 78.72% thereafter (see Appendix 1.C: Fig. 1.C.1). Note that the true share of particulate matter-driven observations after 30th December 2012 is likely higher than 78.72% as 10.85% of the observations fail to specify a main pollutant and it is reasonable to suppose that the main pollutant would have been $PM_{2.5}$ or PM_{10} for part of the observations. The clear majority of the BMEPB observations are thus based on AQIs set by particulate matter[1].

A remaining concern is that the BMEPB measures only PM_{10} until 30th December 2012, whereas the US Embassy measures $PM_{2.5}$. The analysis until 30th December 2012 therefore presupposes a high correlation between PM_{10} and $PM_{2.5}$ in Beijing. A sharp discontinuity in the pollution processes, for instance, would appear in both the BMEPB and the US Embassy data only if this correlation were high.

Several pieces of evidence exist to show that this correlation is, in fact, very high. Firstly, I use the information on the main pollutant for observations with an AQI beyond 50 to convert the BMEPB data into concentrations of PM_{10} and $PM_{2.5}$, and then correlating these concentrations to the US Embassy's $PM_{2.5}$ concentrations for those US Embassy observations that would have yielded an AQI beyond 50 using the BMEPB's AQI definition to ensure comparability. This exercise shows a high correlation between both the inferred BMEPB PM_{10} concentrations and the US Embassy $PM_{2.5}$ concentrations (correlation coefficient: 0.71) and the inferred BMEPB $PM_{2.5}$ concentrations and the US Embassy $PM_{2.5}$ concentrations (correlation coefficient: 0.91).

[1]As a robustness check, Appendix 1.C further shows that the findings drawn from the BMEPB data are unaffected when dropping all observations that report a main pollutant other than $PM_{2.5}$ or PM_{10}.

Secondly, recent research on ambient air quality in Beijing finds similarly high correlations between 0.69 and 0.85 between the BMEPB PM_{10} concentrations and the US Embassy PM2.5 concentrations for a 423 day period between 10 May 2010 and 6 December 2011, which is part of my earlier subsample (Wang, Hu, et al., 2013). Such a high correlation is not specific to Beijing and borne out by research from other parts of the world (for Mumbai, India, see Kumar and Joseph, 2006; for Milan, Italy, see Marcazzan et al., 2001).

Overall, this evidence suggests that both data sources move very closely together, and thus reflect the same underlying pollution processes. In terms of data quality, the US Embassy data is less likely to be influenced by the Blue Sky Day threshold than the BMEPB data because the number of Blue Sky Days in Beijing does not enter the evaluation of US officials. Additionally, I use aerosol optical density (AOD) measurements by the AERONET measurement station in Beijing as a third data source to ensure that the conclusions regarding the true shape of the distribution of air quality in Beijing are not driven by the possibility of the US Embassy to also misreport, and to demonstrate that my method of analysis extends to AOD data. Due to their high accuracy, a typical use AERONET AOD data is verification of satellite-based AOD data (Li, Carlson, and Lacis, 2015; Xue et al., 2014). Interpolated AOD at 550nm correlates well with US Embassy data (64.1%), while the correlation with the BMEPB data is reasonable, but significantly weaker (34.8%).

1.3 Methods

To make the BMEPB and the US Embassy data comparable and find a direct indication for misreporting, this research compares the goodness-of-fit of both the BMEPB and the US Embassy data to a statistical regularity called Benford's Law. Benford's Law is a distribution that closely characterizes the distribution of the first significant digit in many naturally occurring large data sets (Benford, 1938). According to Benford's Law, the frequency for the first non-zero digit is approximately governed by the following distribution:

$$Frequency(i) = log_{10}(1 + \frac{1}{i}) \quad where \quad i = 1, 2, ..., 9. \tag{1.1}$$

The goodness-of-fit with Benford's Law as an indication for fraudulent data has been used both in economics (Abrantes-Metz, Villas-Boas, and Judge, 2011; Michalski and Stoltz, 2013; Varian, 1972) and, less frequently, for ambient air quality data (Brown, 2005; De Marchi and Hamilton, 2006; Dumas and Devine, 2000). While Benford's Law is known to fit air quality data in general[2], it is ex ante difficult to know whether it will fit a particular sample. This analysis therefore uses Benford's Law in a new way: First, graphical evidence on the fit between Benford's Law and relevant subsamples of the datasets is shown to give an indication of whether misreporting occurs in the BMEPB's data and whether it has changed over time. The datasets from the US Embassy and AERONET are then used as a control dataset to verify that air quality data for Beijing fit Benford's Law. Second, the analysis computes a goodness-of-fit measure between the predicted frequency and the empirical frequency for both data sources and plots this measure over time.

The goodness-of-fit measure is the χ^2 statistic, which is an appropriate statistical measure for comparing data with discrete categories to a predicted distribution (Siegel, 1956)[3]. In the case of Benford's Law, there are 9 discrete categories corresponding to the digits 1 to 9. The χ^2 statistic is defined as:

$$\chi^2 = \sum_{i=1}^{9} \frac{(FrequencyObserved_i - FrequencyBenford_i)^2}{FrequencyBenford_i} \tag{1.2}$$

1.4 Results

1.4.1 An anomaly in the official air pollution data

First, I investigate whether the raw data contain suggestive evidence for manipulation of the official air quality data as a result of incentives. The histograms of the Beijing air quality data from the BMEPB show a striking anomaly: there are disproportionally many observations just at and below the politically important Blue Sky Day threshold of 100, and surprisingly few values just above 100 (Fig. 1.1, a). Next, the sample period is split

[2]See Appendix 1.F for an illustration of the fit of Benford's Law with a generic air pollution dataset.

[3]Other measures commonly used to track goodness-of-fit with Benford's Law include the Kolmogorov-Smirnov statistic, the Kuiper statistic, the d (distance) statistic and the statistic m (max) (Morrow, 2014). The conclusions drawn from Figures 5 and 6 are robust to using any of these goodness-of-fit measure because of their high correlation with the χ^2 statistic in both samples: Kolmogorov-Smirnov: 0.94 (BMEPB data), 0.90 (US Embassy data); Kuiper: 0.96 (BMEPB), 0.90 (US Embassy); d (distance): 0.93 (BMEPB), 0.90 (US Embassy); m (max): 0.73 (BMEPB), 0.89 (US Embassy).

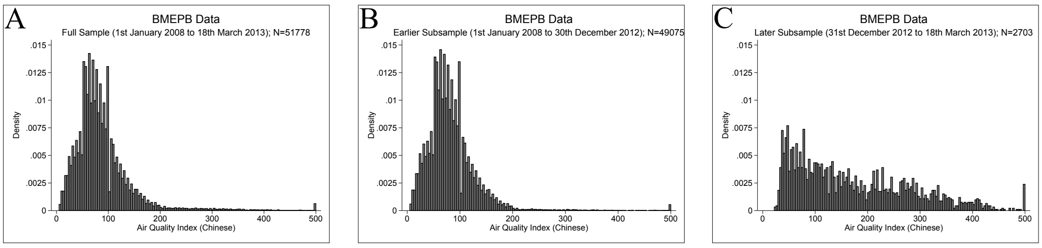


Figure 1.1 *Histogram of air pollution levels (BMEPB data)*. Histograms of the BMEPB data. (a) full sample, (b) earlier subsample, (c) later subsample. AQI values of 100 and less constitute Blue Sky Days.

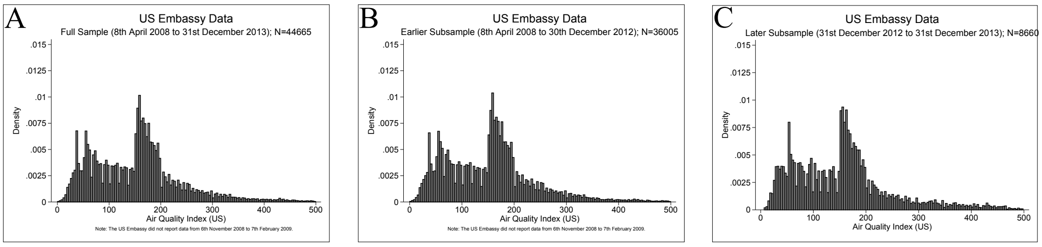


Figure 1.2 *Histogram of air pollution levels (US Embassy data)*. Histograms of the US Embassy data. (a) full sample, (b) earlier subsample, (c) later subsample.

on 31st December 2012, the date on which the BMEPB started to include measurements of $PM_{2.5}$. Strikingly, the missing values above 100 come entirely from the earlier subsample (Fig. 1.1, b). The later subsample does not show this pattern (Fig. 1.1, c).

While misreporting seems to be a likely explanation, and was so interpreted by (Andrews, 2008a) who found the same anomaly until 2007, it is not the only one. Instead, Beijing officials defended themselves by attributing this anomaly to emergency measures such as temporary factory closures on borderline Blue Sky Days (Cyranoski, 2009).

If emergency measures rather than misreporting explained the missing values in the BMEBP data, however, the US Embassy data should exhibit the same anomaly. The histograms for the US Embassy data show that this is not the case, neither for the full sample (Fig. 1.2, a) nor for the different subsamples (Fig. 1.2, b and c). An analysis of the AERONET AOD data confirms these results (Fig. 1.3). This evidence suggests that misreporting by the BMEPB was prevalent until 2012 and then stopped.

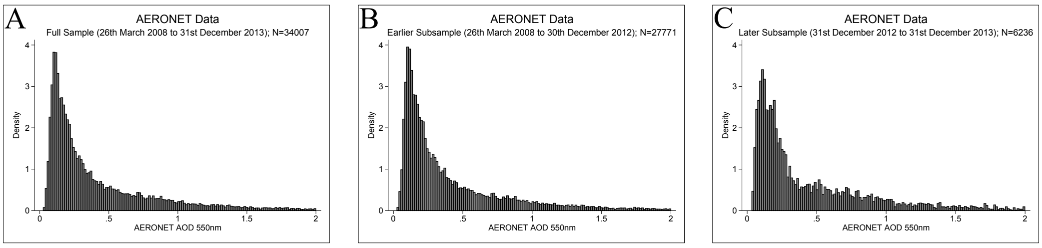


Figure 1.3 *Histogram of aerosol optical density (AERONET data)*. Histograms of the AERONET ground-based AOD data. (a) full sample, (b) earlier subsample, (c) later subsample. A small fraction of AERONET observations beyond 2 are not shown to allow for more detailed representation of observations in the relevant range.

1.4.2 Comparing the BMEPB and the US Embassy data

Yet, an important caveat is that the BMEPB and the US Embassy data are not comparable in levels due to the late inclusion of $PM_{2.5}$ measurements in the BMEBP data and to the use of slightly different breakpoints (see Appendix 1.B)[4]. Benford's Law is therefore used to make both data sources comparable and find a direct indication for misreporting. As can be seen, the BMEPB data match Benford's Law poorly for the full sample (Fig. 1.4, a), suggesting misreporting. This misreporting seems to happen exclusively in the earlier subsample (Fig. 1.4, b). In the later subsample (Fig. 1.4, c), by contrast, the BMEPB data match Benford's Law closely. Misreporting is therefore the likely margin of action. To verify that this improvement in goodness-of-fit is not due to a change in the underlying processes that generate the air pollution, Figure 1.5 shows that the fit of the US Embassy data is good and unchanged throughout all periods. Figure 1.6, shows that the fit of the AERONET AOD data with Benford's Law is good and unchanged through all periods. The AERONET AOD data thus give the same conclusions as the US Embassy data, suggesting that the conclusions are not driven by a bias in the US Embassy data.

Likely misreporting over time. To ensure that the improvements in the goodness-of-fit with Benford's Law for the BMEPB data reflect a general trend rather than being an artifact of the date at which I split the sample, the analysis computes the goodness-of-fit over time. Figure 1.7 shows that the goodness-of-fit between the BMEPB data and Benford's Law markedly

[4]The different breakpoints in the AQI used by the US Embassy also explain the hump-shaped distribution of the US Embassy's histogram (Fig. 1.2, a-c). Appendix 1.D provides further illustration of this point and shows the close comparability of Fig. 1.1c and Fig. 1.2c when using the same breakpoints.

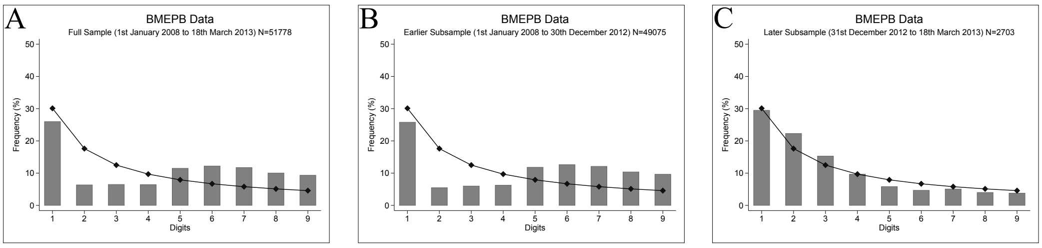


Figure 1.4 *Observed frequencies and Benford's Law (BMEPB data)*. Observed frequencies for the first significant digits (grey bars) against the frequencies predicted by Benford's Law (black connected dots) for the BMEPB data. (a) full sample, (b) earlier subsample, (c) later subsample.

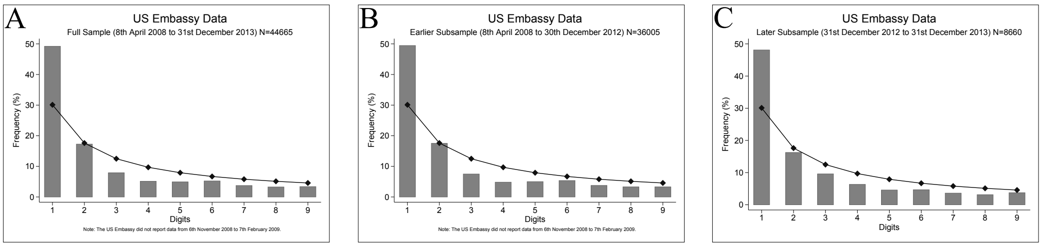


Figure 1.5 *Observed frequencies and Benford's Law (US Embassy data)*. Observed frequencies for the first significant digits (grey bars) against the frequencies predicted by Benford's Law (black connected dots) for the US Embassy data. (a) full sample, (b) earlier subsample, (c) later subsample.

improved after the BMEPB started measuring $PM_{2.5}$ on 31st December 2012. To check that the underlying pollution processes remain the same, Figure 1.8 shows that the goodness-of-fit for the US Embassy data has not changed. The reason for the improved goodness-of-fit for the BMEPB data

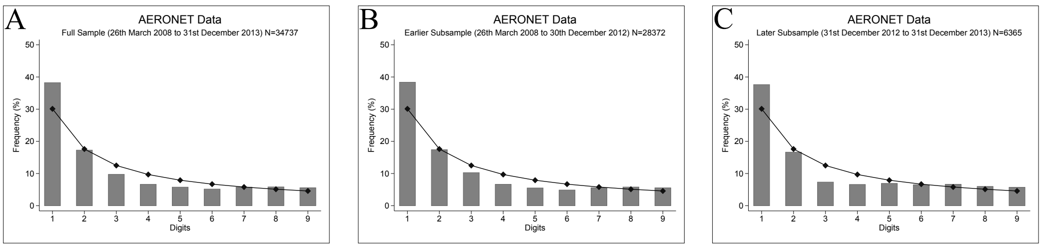


Figure 1.6 *Observed frequencies and Benford's Law (AERONET data)*. Observed frequencies for the first significant digits (grey bars) against the frequencies predicted by Benford's Law (black connected dots) for the AERONET ground-based AOD data. (a) full sample, (b) earlier subsample, (c) later subsample.

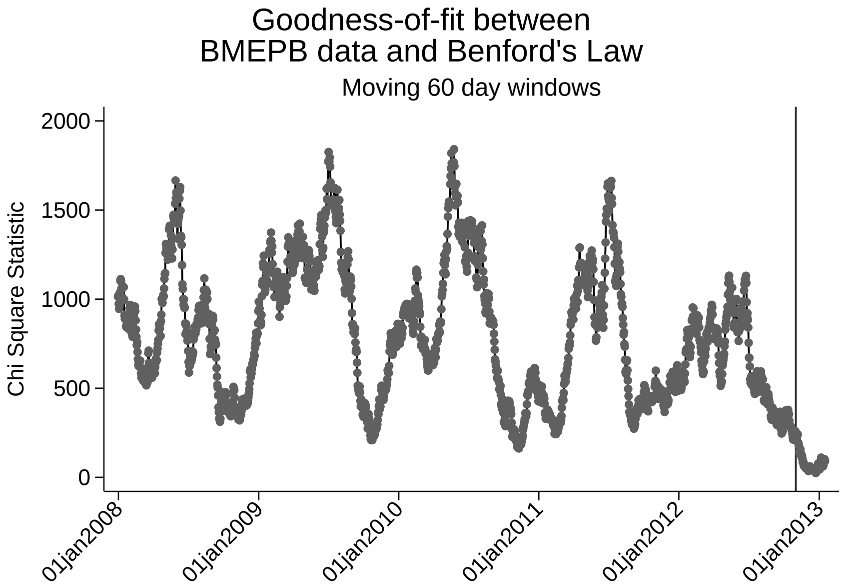


Figure 1.7 *Goodness-of-fit with Benford's Law over time (BMEPB data).* χ^2 statistic comparing the BMEPB data to Benford's Law. Computed over moving 60 day windows. The vertical black line marks the earliest time window during which the BMEPB started measuring PM 2.5 (from 31st December 2012 onwards).

must therefore lie in the Chinese measurements.

In summary, my findings suggest that emergency air control measures are unlikely to explain the anomaly in the BMEPB data, and that the BMEPB likely engaged in misreporting from 2008 to 2012. From 2013 onwards, however, misreporting appears to have stopped. At the same time, my analysis illustrates the usefulness of applying Benford's Law to detect misreporting in settings with independent measurements of air pollution.

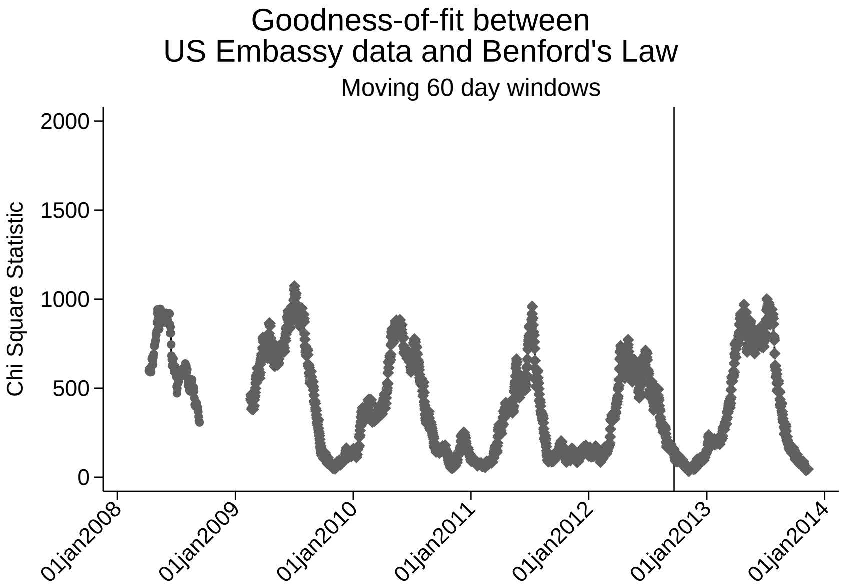


Figure 1.8 *Goodness-of-fit with Benford's Law over time (US Embassy data).* χ^2statistic comparing the US Embassy data to Benford's Law. Computed over moving 60 day windows. The vertical black line marks the earliest time window during which the BMEPB started measuring PM 2.5 (from 31st December 2012 onwards). The US Embassy did not report data between 6th November 2008 and 7th February 2009.

1.5 Discussion

The above results raise two questions. Firstly, is it possible that the air pollution control efforts taken during the Olympic Games 2008 might confound the results? The Olympics lasted from 8-24 August 2008, and emission control measures started on 20 July 2008 (Zhang et al., 2012). Subsequent evaluation based on both the official air pollution data and aerosol optimal depth from satellite imagery has shown that these control measures improved the average air pollution in the area of Beijing (Chen, Jin, et al., 2013). Traffic restrictions likely contributed to these improvements (Wang, Westerdahl, et al., 2009), but so did favourable meteorological conditions in greater Beijing during the time of the Olympic Games (Simonich, 2009; Tang et al., 2009; Wang, Primbs, et al., 2009; Yao et al., 2009; Zhang et al., 2012). The improvements in air quality lasted until about one year after the Olympic Games (Chen, Jin, et al., 2013).

The time period of the Olympic Games 2008 and the associated improvements in air quality is part of my earlier subsample. In this subsample, I find likely misreporting in the BMPEB data. Given that my measure of misreporting is the scarcity of observations at the Blue Sky Day threshold and the excess of values just below it as well as the goodness of fit with Benford's Law, a change in the mean air pollution, as documented in the literature above, does not affect misreporting.

The Olympic Games 2008 drew attention to the problem of air pollution, however, possibly leading to a change in misreporting that is independent of the actual improvements. Appendix 1.E shows that misreporting indeed improved for a short period in the immediate aftermath of the Olympics, only to return to its previous levels from 2009 onwards. Regarding the results of my analysis, the presence of the Olympics in my earlier subsample would go against me finding misreporting in the earlier timeperiod, however[5].

Secondly, given the disappearance of the anomaly in the reported air pollution measurements in the official data for Beijing at the end of 2012, a natural question to ask is what caused this change. To narrow down the reasons for the change, it is important to understand whether this change is Beijing-specific or instead reflects a China-wide trend. To explore this question, I use air pollution data on two major Chinese cities that exhibit a Blue Sky Day anomaly similar to Beijing, but for whom the US Embassy does not measure air pollution. These cities are Anshan, in Liaoning Province, and Jining, Shandong Province. As shown in Figure 1.E.1 in Appendix 1.E, the mass of observations just below the cutoff is considerably smaller for both the comparison group and Beijing during most of the period of analysis (excepting the aftermath of the Olympic Games 2008). Misreporting seems to stop in Beijing at the end of the sample, however, while continuing in the two comparison cities. This indicates that the underlying reason for the likely end of misreporting in the BMEPB data is not the result of a China-wide trend.

Beijing differs in two crucial aspects. Firstly, the presence of the US Embassy has created pressure on misreporting throughout my sample period. Secondly, Beijing was amongst the first cities to be selected for trial of the new air pollution regulation HJ633-2012 at the end of 2012, when the change in misreporting occurred. While this regulation does not contain provisions related to the Blue Sky Day threshold, HJ633-012 introduced $PM_{2.5}$ into the list of pollutants to be measured, thus requiring the installa-

[5] This consideration notwithstanding, Appendix 1.E also repeats the analysis in Figures 1.1 and 1.3 for the BMEPB data for the earlier subsample excluding the data from the start of the control measures for the Olympic Games 2008 (20 July 2008) until one year after their end (24 August 2009). The results are unchanged.

tion and calibration of new technology. It is plausible to presume that the process of setting up the new system might have led to increased inspections by the upper hierarchy. A press release from the Ministry of Environmental Protection dated 31st December 2012 confirms this, quoting MEP Chief Engineer Wan as specifically stressing quality control in the air quality monitoring as part of the implementation, and announcing intensified supervision and checks as part as the next step following the introduction of HJ633-2012 (Ministry of Environmental Protection of the People's Republic of China, 2012). A situation of increased attention from the political hierarchy at the end of 2012, coupled with the presence of the US Embassy's measurements from 2008 onwards, might thus have led to the end of statistical corruption in Beijing's air quality data at the end of 2012.

One important caveat regarding these results is in order. The analysis relies on two pieces of evidence: a sharp discontinuity in the histograms of the raw data, and the goodness-of-fit with Benford's Law. Both pieces of evidence can detect a relabeling of AQI values from just beyond the Blue Sky Day cutoff to just below it. More sophisticated methods of misreporting, however, such as shifts in the entire distribution of AQI values, would go undetected. While this limitation is important in theory, in practice this is less so: previous studies that detected misreporting of air quality data in Beijing until 2007 (Andrews, 2008a) and other cities in China until 2010 (Chen, Jin, et al., 2012; Ghanem and Zhang, 2014) found misreporting to consist of the relabeling practice that the analysis can track. The methodology used in this study is thus accurate in closely following known misreporting over time. To establish that all possible kinds of misreporting have stopped, however, would require further analysis.

1.6 Concluding Remarks

Political pressure to fix air pollution can result in "statistical corruption": government authorities respond by misreporting the desired data. This research makes use of the unique setting in Beijing, where air quality is independently measured by both the Chinese authorities and the US Embassy. Using a novel way of assessing statistical misreporting via Benford's Law that makes both data sources comparable for the first time, this analysis suggests that the authorities in Beijing likely manipulated air quality from 2008 to 2012. My results thus suggest that the anomalies found in (Andrews, 2008a; Chen, Jin, et al., 2012; Ghanem and Zhang, 2014) are unlikely to be explained by emergency control measures as claimed by Beijing officials (Cyranoski, 2009) but instead are likely to show true misreporting. My results furthermore show that misreporting of air pollution data in Beijing extended well beyond 2010 despite the recent increase in attention towards environmental policy making in China. From 2013 onwards, however, this has changed: despite ongoing suspicion regarding the quality of Chinese air pollution data (Ghanem and Zhang, 2014), the 'leaders make numbers' phenomenon seems to have been overcome.

Environmental governance in China is currently at a crossroads. Proposals have called for policies to decouple economic activity from pollution (Xiao, Mujumdar, and Che, 2015), to improve health outcomes by reducing pollution (Hughes, 2012), and to realize the co-benefits of reducing air pollution for climate change mitigation (Schmale et al., 2014). At the same time, research has shown that ambient air pollution in China is significantly impacted by government action and the structure of political incentives (Almond et al., 2009; Jia, 2014), thus highlighting the role for policy. Whichever strategies China may decide to implement, reliable air quality data are needed for successful implementation and evaluation. The findings of the present research are thus a reason for optimism: unlike only a few years ago, statistics on atmospheric pollutants now seem to allow for evidence-based environmental policy making in China.

Appendix

1.A Data Sources

BMEPB web interface: http://www.bjepb.gov.cn/air2008/Air1.aspx. The historical dataset could not be downloaded directly; instead, a webscraping algorithm was used. This web interface stopped reporting in March 2013.

US Embassy web interface: http://www.stateair.net. In conformity with the US Department of State's data use statement, I note that the US Embassy data used in this study are not fully verified or validated and are released with the sole purpose of providing health information to US citizens who are travelling abroad. I furthermore give attribution to the US Department of State for providing the data.

AERONET (Aerosol Robotic Network) data interface for Beijing: http://aeronet.gsfc.nasa.gov. AERONET is a network of ground-based remote sensing aerosol measurement sites. The downloaded data are aerosol optical density (AOD) measurements for the Beijing site that is hosted by the Institute of Atmospheric Physics (PIs: Hongbin Chen and Philippe Goloub) for the time period from 26th March 2008 to 31st December 2013, with several measurements during different times of each day. For this study, I use the quality assured data (Level 2.0), which have a pre- and post-field calibration applied, are automatically cloud cleared, and manually inspected (Holben et al., 2006; Smirnov et al., 2000). There are no AOD measurements for around 20% of the days in this time period, with the missing days spaced equally across years. The reason for the missing values is twofold: these correspond to days during which the AOD measurements had to be discarded due to imprecise measurements, or to days with excessive cloud coverage[6]. Both reasons explain roughly half of the missing values each. The AERONET data for Beijing contain AOD measure-

[6] Personal communication from PI Hongbin Chen of the AERONET Beijing site.

ments for the wavelengths 440nm and 675nm. I interpolate the data for AOD of a wavelength of 550nm, the wavelength that is typically used in the literature (Cesnulyte et al., 2014), using the Ångström exponent, provided by AERONET, by rearranging formula (2) from (Eck et al., 1999):

$$\tau_{550} = \tau_{440}(\frac{550}{440})^{-\alpha}, \tag{1.3}$$

where α is the Ångström exponent, and $\tau\lambda$ is the AOD at wavelength λ[7].

[7]To see this, rearrange $\alpha = -\frac{ln(\frac{\tau_{550}}{\tau_{440}})}{ln(\frac{550}{440})}$ to $-\alpha ln(\frac{550}{440}) = ln(\frac{\tau_{550}}{\tau_{440}})$, which equals $exp\{-\alpha ln(\frac{550}{440})\} = \frac{\tau_{550}}{\tau_{440}}$, which in turn equals $\tau_{440}\{exp(ln(\frac{550}{440})\}(-\alpha) = \tau_{550}$ because $e^{a*b} = (e^a)^b$.

1.B Definitions of the Air Quality Indices (AQIs) used by the Beijing Municipal Environmental Protection Bureau and the US Embassy

The AQI is an index that is used to inform the population about the health effects of different air pollutants. The AQIs used by the BMEPB and the US Embassy are calculated according to the same two-step procedure.

In the first step, the concentration of each pollutant is converted into an individual air quality index value (IAQI) via the following formula:

$$IAQI_p = \frac{IAQI_{High} - IAQI_{Low}}{BP_{High} - BP_{Low}}(C_p - BP_{Low}) + IAQI_{Low} \tag{1.4}$$

where C_p is the measured concentration of pollutant p, BP_{High} is the breakpoint that is higher than or equal to C_p while BP_{Low} is the breakpoint that is lower than or equal to C_p. $IAQI_{High}$ and $IAQI_{Low}$ are the AQI scores that correspond to the BP_{High} and BP_{Low} according to the following tables:

Table 1.B.1 Breakpoints for Different Pollutants Before 31st December 2012

		BMEPB[i]		*USEmbassy*[ii]
AQI[iii]	SO_2	NO_2	PM_{10}	$PM_{2.5}$
0	0	0	0	0
50	50	80	50	15.5
100	150	120	150	40.5
150	*iv*	*iv*	*iv*	65.5
200	800	280	350	150.5
300	1600	565	420	250.5
400	2100	750	500	350.5
500	2620	940	600	500

Note. All breakpoints are for concentrations over a 24 hour period (in $\mu g/m^3$). *i*: GB 3095-1996, SEPA Announcement [2000] No. 1 (Amendment to GB 3095-1996). *ii*: EPA-454/B-12-001. *iii*: The official name for the air quality index used by the BMEPB was "Air Pollution Index" until the 31st of December 2012. To avoid confusion, the name "Air Quality Index" is used throughout this article. *iv*: The BMEPB does not employ a separate breakpoint for an AQI of 150 before 31st December 2012.

Table 1.B.2 Breakpoints for Different Pollutants From 31st December 2012 Onwards

	BMEPB^{i}						*USEmbassy*^{ii}
AQI^{iii}	SO_2	NO_2	CO	O_3	PM_{10}	$PM_{2.5}$	$PM_{2.5}$
0	0	0	0	0	0	0	0
50	50	40	2	100	50	35	15.5
100	150	80	4	160	150	75	40.5
150	475	180	14	215	250	115	65.5
200	800	280	24	265	350	150	150.5
300	1600	565	36	800	420	250	250.5
400	2100	750	48	*iii*	500	350	350.5
500	2620	940	60	*iii*	600	500	500

Note. All breakpoints are for concentrations over a 24 hour period (in $\mu g/m^3$); excepting CO (measured in mg/m^3) and O_3 (measured over an 8 hour period). *i*: HJ 633-2012. *ii*: EPA-454/B-12-001. *iii*: O_3 concentrations beyond 800 $\mu g/m^3$ per 8 hour period do not have a corresponding air quality index.

In the second step, the individual air quality index scores are aggregated by picking the highest amongst the IAQI values. The AQI therefore reflects the pollutant with the highest IAQI.

$$AQI = \max_{p \in \{SO_2, NO_2, CO, O_3, PM_{10}, PM_{2.5}\}} \{IAQI_p\} \tag{1.5}$$

An example to illustrate the above definition. Assume hypothetically that the US Embassy measured a concentration of PM2.5 of $34\mu g/m^3$ over a 24 hour period. Then, the corresponding AQI would be

$$IAQI_{PM_{2.5}} = \frac{100 - 50}{40.5 - 15.5}(34 - 15.5) + 50 = 87 \tag{1.6}$$

1.C Robustness check: Analysis after dropping BMEPB observations from non-particulate matter AQIs

To guard against concerns that the reported findings might be influenced by the minority of BMEPB observations from days for which particulate matter was not the main pollutant, Figures 1.1 and 1.3 are redrawn after excluding observations that are identified as based on neither PM_{10} nor $PM_{2.5}$ as the main pollutant on a given day.

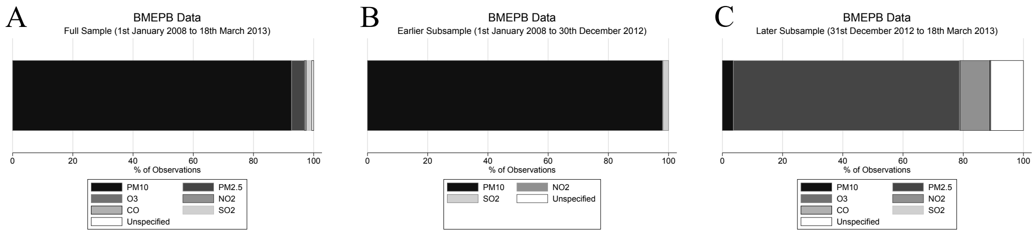


Figure 1.C.1 *Pollutant composition of the BMEPB dataset*. Share of main pollutants for all observations with an AQI beyond 50. (a) full sample, (b) earlier subsample, (c) later subsample.

Figure 1.C.1 shows the share of each pollutant in the BMEPB data based on the main pollutant reported for observations with an AQI exceeding 50 for the full sample and the two subsamples. PM10 and PM2.5 drive 96.96% of observations for the full sample, 98.03% of the observations for the earlier subsample, and 78.72% of the observations in the later subsample.

Excluding these observations is a conservative approach because it is likely that some of the AQI observations that fail to specify a pollutant are in fact based on $PM_{2.5}$ or PM_{10}. Furthermore, dropping the observations identified as based on neither PM_{10} nor $PM_{2.5}$ rather than replacing them with their nearest neighbouring observation based on $PM_{2.5}$ or PM_{10} is likely to decrease the chance of finding a close goodness-of-fit with Benford's Law as the greater part of these observations come from the lower end of the AQI distribution rather than the distribution as a whole.

As can be seen from Figures 1.C.2 and 1.C.3, the conclusions from analysis are not driven by the minority of observations identified as not from particulate matter. As in the main analysis, the anomaly of missing values at the Blue Sky Day threshold is present in the full sample (Fig. 1.C.2, a), comes entirely from the earlier subsample (Fig. 1.C.2, b) and vanishes in the later subsample (Fig. 1.C.2, c).

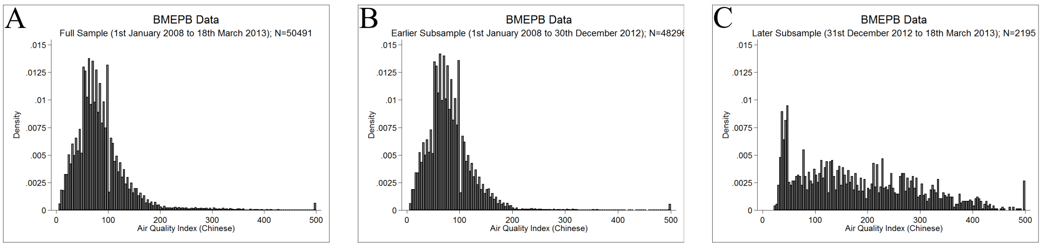


Figure 1.C.2 *Histogram of air pollution levels (Restricted BMEPB data)*. Histograms of the BMEPB data after dropping observations identified as neither from PM10 nor PM2.5. (a) full sample, (b) earlier subsample, (c) later subsample. AQI values of 100 and less constitute Blue Sky Days.

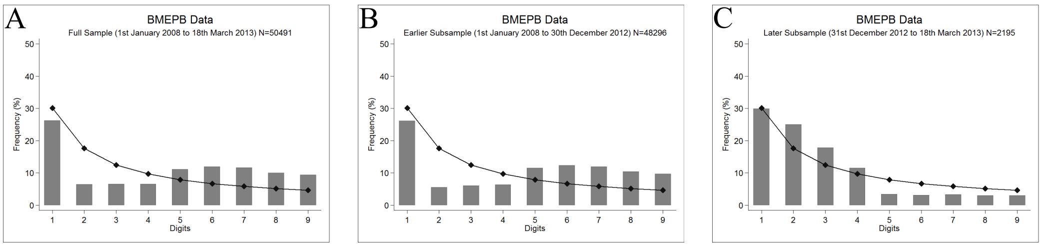


Figure 1.C.3 *Observed frequencies and Benford's Law (Restricted BMEPB data)*. Observed frequencies for the first significant digits (grey bars) against the frequencies predicted by Benford's Law (black connected dots) for the BMEPB data after dropping observations identified as neither from PM10 nor from PM2.5. (a) full sample, (b) earlier subsample, (c) later subsample.

1.D Robustness check: Figure 2 using BMEPB AQI breakpoints

This section explains why the histograms of the AQIs measurements by the BMEPB and the US Embassy appear different when comparing Fig. 1.1c and Fig. 1.2c and shows their close comparability through converting the US Embassy data from the US AQI to the BMEPB AQI.

From Appendix 1.B, recall that the US Embassy and the BMEPB use different breakpoints when converting $PM_{2.5}$ concentrations into AQI values. As illustrated in Figure 1.D.1 below, the US Embassy maps $PM_{2.5}$ concentrations from 65.5 to 150.5 $\mu g/m^3$ to an AQI of 150 to 200, whereas the BMEPB only maps $PM_{2.5}$ concentrations from 115 to 150 $\mu g/m^3$ to an AQI ranged 150 to 200. This explains why the US Embassy AQI observations look bunched between AQI values of 150 and 200 compared to the BMEPB data.

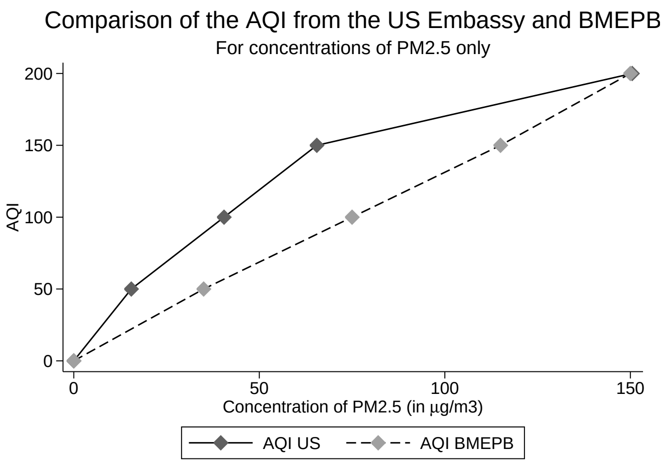


Figure 1.D.1 *Illustration of the different PM2.5 breakpoints*. This graph illustrates the different breakpoints used by the US Embassy and the BMEPB in the later subsample for AQI values up to 200. The figure is based on the AQI definitions reported in Appendix 1.B.

To illustrate this point, the left panel in the composite Figure 1.D.2 below reproduces the original BMEPB histogram for the later subsample (left panel) and compares it to the US Embassy data mapped into the Chinese AQI in the right panel (right panel). The similarity between both histograms provides further evidence that the underlying raw data of both the BMEPB and the US Embassy reflect the same pollution processes and thus offer a good degree of comparability.

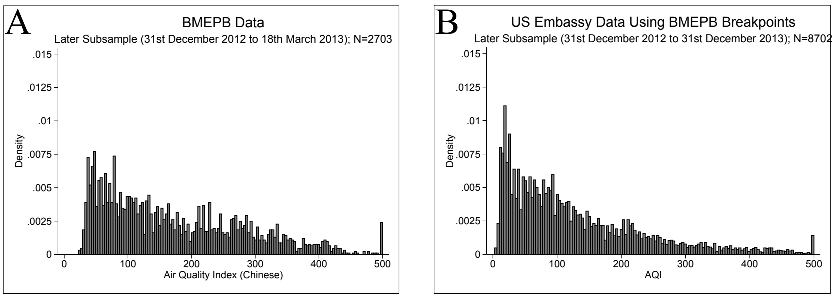


Figure 1.D.2 *Histograms of air pollution levels (BMEPB data and US Embassy data)*. Histograms of the BMEPB data (left panel) and the US Embassy data converted into the AQI used by the BMEPB (right panel).

1.E Additional results for the Discussion section

This section contains additional results related to the Olympic Games 2008. Firstly, to track the development of the Blue Sky Day anomaly over time, I compute an index for the anomaly that I define as the frequency of observations just beyond the Blue Sky Day divided by the frequency of observations just below it. The reasoning behind this choice is that the likelihood of observing an AQI just beyond the threshold should be roughly equal to the likelihood of observing an AQI just below the schedule, giving the index a natural interpretation: index values far below one give an indication for mislabeling of values in this part of the distribution. Formally, I compute the anomaly index as the following ratio:

$$AnomalyIndex = \frac{\#ofobservationsforwhich 100 \leq AQI \leq 105}{\#ofobservationsforwhich 95 \leq AQI < 100} \tag{1.7}$$

Figure 1.E.1 below graphs the time series of this index for both the pair of comparison cities and for Beijing. The data for Anshan and Jining are taken from the webpage of the Ministry for Environmental Protection of the People's Republic of China (http://english.mep.gov.cn/).

Figure 1.E.1 suggests that there was a temporary decrease in likely misreporting in Beijing following the Olympic Games until the end of 2008 , after which misreporting resumed to the previous level. If anything, this temporary end to misreporting makes it less likely for my analysis to detect misreporting for the earlier subsample as a whole.

To nonetheless show the robustness of the analysis to the choice of time period, Figure 1.E.2 shows that my analysis is robust to dropping the entire period from the start of the control measures preceding the Olympic Games from 20 July 2008 onwards to the end of the temporary improvements in air quality to 24 August 2009, one year after the Olympics ended. The results are identical to the results from the earlier subsample based on the full dataset.

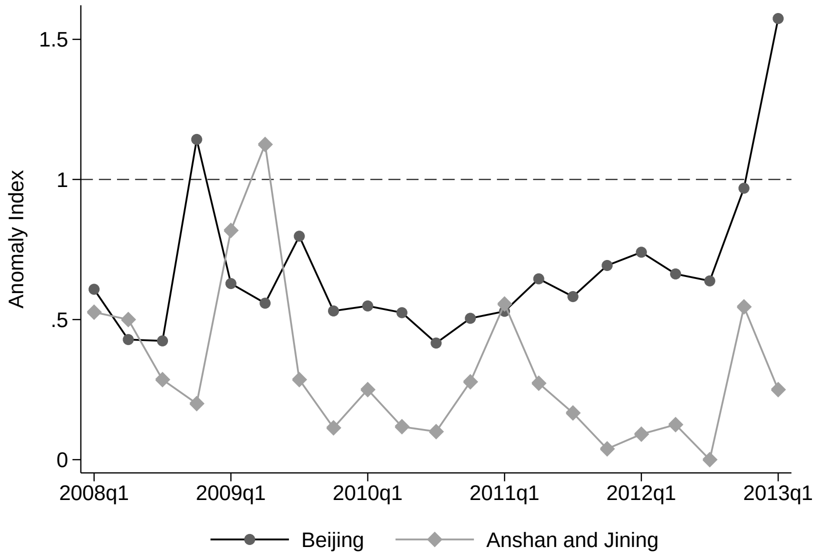


Figure 1.E.1 *The* Blue Sky Day *Anomaly in Beijing and the Comparison Cities over time*. The anomaly index tracks the anomaly over time. One data point represents the average value of the anomaly index for a quarter of the natural year, for either Beijing or the mean of Anshan and Jining.

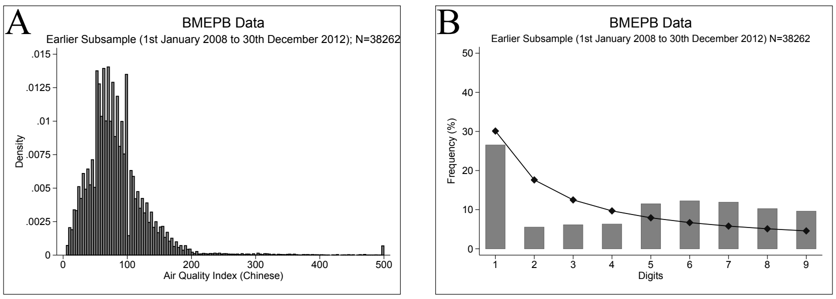


Figure 1.E.2 *Histogram of air pollution levels and observed frequencies and Benford's Law (Restricted BMEPB data)*. Histogram (left panel) and comparison with Benford's Law (right panel) of the BMEPB data for the earlier subsample (1st January 2008 to 30th December 2012) after dropping observations from 20 July 2008 to 24 August 2009, the period of the Olympic Games 2008 and the related air quality improvements.

1.F Benford's Law for a generic air pollution dataset

This section shows that Benford's Law is a good description of a generic air pollution dataset. I use the historical dataset of air pollution in Europe. The source for these data is the full dataset from 1991-2012 from AirBase, a database on European air quality maintained by the European Environmental Agency (www.eea.europa.eu/themes/air/air-quality/map/airbase).

The fit of the air pollution data with Benford's Law is good both when combining the data from all European countries (Fig. 1.F.1) and when looking at each country individually (Fig. 1.F.2).

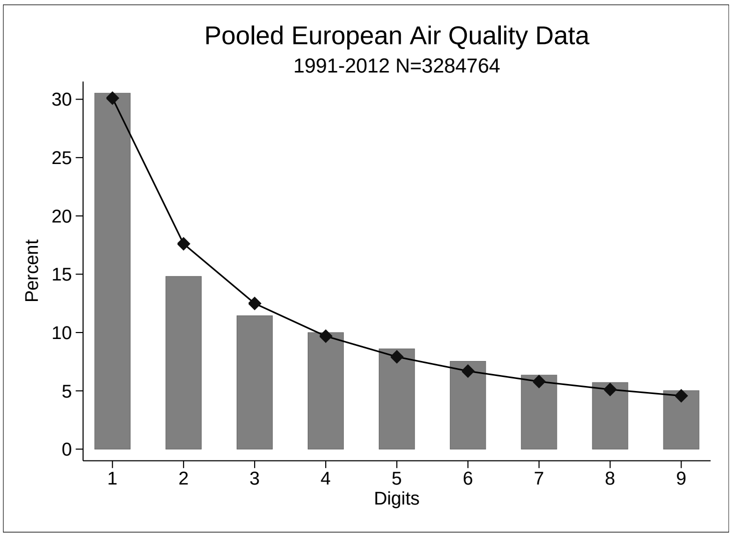


Figure 1.F.1 *Observed frequencies and Benford's Law for a generic air pollution dataset*. Observed frequencies for the first significant digits (grey bars) against the frequencies predicted by Benford's Law (black connected dots) for the European air quality dataset *Air Base*, pooled over all countries for the years 1991-2012.

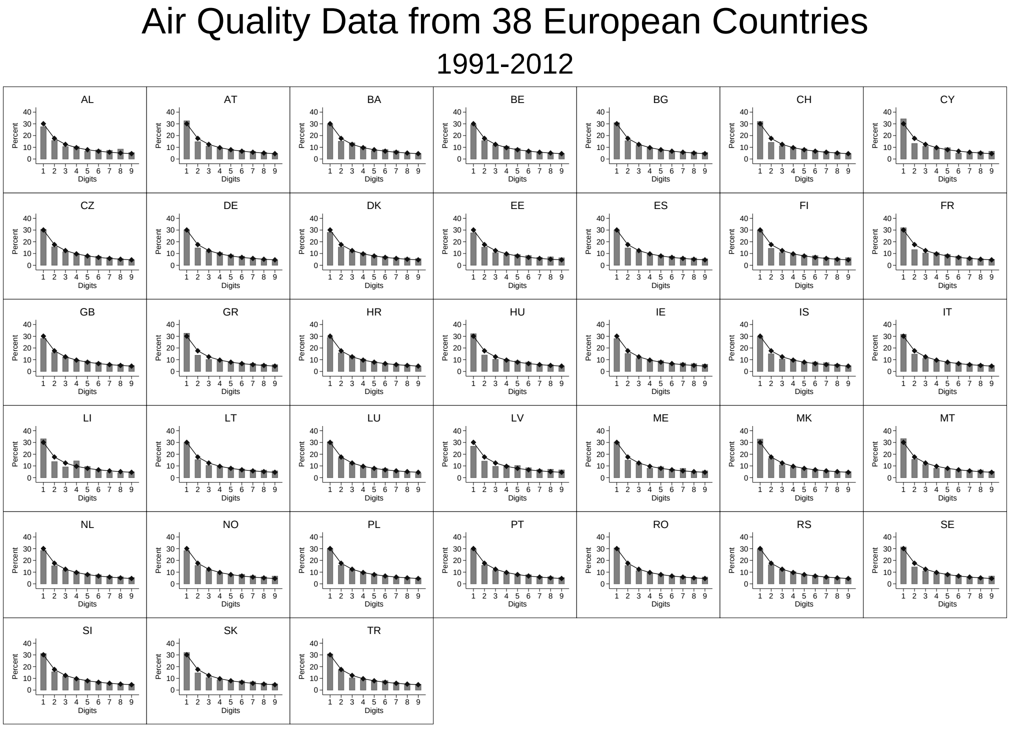


Figure 1.F.2 *Observed frequencies and Benford's Law for many generic air pollution datasets*. Observed frequencies for the first significant digits (grey bars) against the frequencies predicted by Benford's Law (black connected dots) for the European air quality dataset *Air Base*, shown for each individual countriy for the years 1991-2012.

Chapter 2

Compliance, Efficiency, and Instrument Choice: Evidence from Air Pollution Control in China

Abstract

This chapter evaluates China's main air pollution control policy. In 2005, China decided on a 10% SO_2 emissions reduction goal as part of the 11th Five-Year Plan (2006-2010). I study the effect of this policy on pollution outcomes, using both the official, misreporting-prone indicator and independent NASA SO_2 satellite data in a differences-in-differences strategy that exploits variation in target stringency at the province level. I find that results from the official and the satellite data differ initially when the Chinese government lacked the ability to effectively monitor SO_2 pollution. Ultimately, however, the policy worked and reduced air pollution by 11%. The regulated provincial governments react through rhetorical compliance, measured by a unique dataset of quantified political statements, and by shutting down small, inefficient thermal units. Rhetorical compliance increases, especially before the government gained the ability to monitor SO_2 in 2008. Real compliance sets in through the shutdown of small, inefficient thermal units. Next, I compute detailed marginal abatement cost curves for SO_2 for each province in China, thus illustrating the large heterogeneity in abatement cost across provinces. I use those curves to construct the counterfactual cost-efficient allocation of SO_2 reduction targets across provinces. Using this benchmark, I find that the cost-efficient alloca-

tion would increase efficiency by 49% at the margin, by lowering marginal abatement cost from 658€/tSO_2 to 338€/tSO_2. This finding is robust to inclusion of a back-of-the-envelope measure for the marginal benefits of abatement. I conclude that a market-based allocation of SO_2 reduction targets would have doubled the efficiency of China's main air pollution control policy. Contrary to the US experience, I find that a mandate on scrubbers would reap most of the efficiency gains of the market-based allocation.

2.1 Introduction

The effective design and implementation of environmental regulation is crucial for correcting environmental externalities. Traditionally, economists have analyzed environmental regulation in developed countries where technical expertise, appropriate monitoring of pollution and rule of law often allowed for the successful implementation of regulation. Recently, attention turned to environmental regulation in developing countries and how compliance with regulation interacts with imperfect institutions (Duflo et al., 2013; Oliva, 2015). At the same time, developing countries are often more severely affected by the most important environmental externalities such as air pollution. According to the latest WHO estimates, air pollution is responsible for one in eight of total global deaths, which makes 7 million deaths a year (WHO, 2014). One country which is particularly struck by air pollution is China. As development has soared, so has air pollution. Recent air pollution levels in Northern China are as severe as those in London at the height of the Industrial Revolution[1].

This study provides the first empirical evaluation of China's total emissions control target in the 11th Five-Year Plan (FYP) from 2006 to 2010. In an effort to bring down air pollution, the Chinese government decided to limit the total emissions of sulphur dioxide (SO_2) by 10% relative to 2005 baseline levels. The national limit was later split into widely varying reduction targets for each province. This study evaluates the SO_2 reduction policy along three margins: First, did the policy improve SO_2 pollution outcomes? Second, how did the regulated provincial governments comply? Third, how efficient was the policy compared to a counterfactual market-based policy instrument?

The 11th FYP marks a turning point in environmental policy-making in China; it is considered 'the most environmentally ambitious document in the history of the Communist Party' (Watts, 2011). When the policy was passed in 2005, China's regulatory agency was the weak State Environmental Protection Administration (SEPA). SEPA did not have access to reliable SO_2 pollution data in 2005 and had to implement the regulation based on limited information from aggregate SO_2 emission statistics. This situation changed drastically in 2008, when the central government upgraded SEPA to become the Ministry of Environmental Protection (MEP), gave it high-level political backing and allowed it to track SO_2 pollution at the source

[1] The level of total suspended particles (TSP) in London in 1890 was just beyond 600 $\mu g/m^3$ (see Figure 3 in Fouquet, 2011). Similarly, TSP levels North of China's Huai River were just beyond 600 $\mu g/m^3$ up until 1980 (see Figure 2 in Chen, Ebenstein, et al., 2013).

and independent of provincial governments (State Council, 2007).

This empirical setting is insightful for several reasons. While data on the implementation of the SO_2 reduction policy is only available at the province level, my setting is unique because of the availability of real pollution data in the period before the Chinese government could monitor it. This is due to coincidence: in late 2004, just before the start of the policy, NASA launched the *EOS-Aura* satellite that provides an independent and reliable data source for SO_2 pollution in China. Additionally, rich outcome variables are available for China to track the behaviour of provincial governments along multiple margins, including quantified political statements about current and past political priorities at the province-year level. This allows me to study whether regulated agents react differently when monitored appropriately. Lastly, a further advantage to studying air pollution regulation in China is the availability of micro-level data on the cost of SO_2 abatement in each province.

Firstly, I evaluate whether the SO_2 control policy improved pollution outcomes despite the lack of regulatory capacity at the start of the 11th FYP. Exploiting variation across provinces in a differences-in-differences (DID) specification, I recover the causal effect of the SO_2 control policy on both reported pollution, using the official indicator, and on real pollution, measured through NASA satellite data. I then study whether the effect of the SO_2 reduction target differs at the county level according to the initial distribution of pollution within the province. Finally, I investigate the timing of the effects of the policy to study compliance over time.

I find that the policy was a success according to the official SO_2 emissions indicator: a one-standard deviation increase in the stringency of the reduction target leads to a statistically significant 6% decrease in SO_2 emissions. The effect is similar based on the SO_2 NASA satellite data at an effect size of 11% per standard deviation. Within a province, the estimated effect is stronger for counties that were initially more polluted. However, the timing of the estimated differs when comparing the official and the satellite data: according to the official data, provinces with a higher reduction target started to decrease their SO_2 emissions immediately and monotonically after the start of the 11th FYP. The satellite data, by contrast, paint a different picture: according to the satellite data, SO_2 pollution in provinces with higher targets only started to decrease in 2008, after the government gained the ability to monitor SO_2 pollution. A plausible reason for these differences in timing is misreporting due to moral hazard. The evidence, however, is only suggestive because these differences are not statistically different.

Next, I evaluate the mechanisms through which the regulated agents comply with the policy. I study their reactions along two margins: rhetorical and real compliance. Given the initial period of moral hazard when the Chinese government could not monitor SO_2, I first ask whether provincial governments reacted to the SO_2 reduction targets by changing their political rhetoric. I construct a unique province-year dataset of political statements by the governments of each province in China. These statements discuss the political priorities of each provincial government, both for the preceding year and the year to come. This allows me to study not only the content of the provincial governments' rhetoric but also the timing. I use these statements to quantify the importance of air pollution in a province's political rhetoric. To study real compliance, I collect data on the two most economic measures to reduce SO_2 at the province level. These are the installation of desulfurization equipment (scrubbers) in power plants and industrial facilities, and the shutdown of small, inefficient power plants.

I find that provincial governments strongly adjusted their rhetoric to the political goals from the center: a one-standard deviation increase in the province's SO_2 reduction target leads to a 30% increase in political statements on air pollution, mainly driven by mentions of sulfur. A placebo test using other goals from the 11th FYP shows that provinces specifically adapt their rhetoric to the SO_2 targets. Regarding the timing of the political statements, I find that the increase in political statements is most pronounced in statements that praise past work on air pollution. This is particularly visible for the period just before the government could monitor air pollution. In light of the difference between the official and the satellite SO_2 data, this provides further suggestive evidence for strategic misreporting. Ultimately, however, the SO_2 reduction policy worked, and compliance became real. Among the abatement measures available to provinces, I find that real compliance for provinces with higher SO_2 reduction targets relied less on the installation of desulfurization equipment, or on an earlier installation of this equipment. Instead, I find that the main driver for bringing down SO_2 pollution was the shutdown of small, inefficient power plants.

To study the efficiency of the policy design and quantify the gains from trade across different policy instruments, I compute detailed SO_2 marginal abatement cost (MAC) curves for each province in China. These curves show the large heterogeneity in SO_2 abatement cost across the provinces of China. Based on the MAC curves, I find that the Chinese government did not equate marginal abatement cost across space. Instead, the reduction targets favoured coastal provinces in the East where abatement costs are higher. This is consistent with a tale-of-two-cities story, in which China would develop amenity-based consumer cities along the coast, while maintaining a base of polluting manufacturing in its interior (Kahn, 2006; Zheng and Kahn, 2013). Using the MAC curves, I construct the counterfactual cost-efficient allocation of SO_2 reduction targets across provinces needed to achieve the 10% SO_2 reduction target. This allows me to study the gains from trade from moving from a command-and-control regulation to a market-based allocation of SO_2 reduction targets. I find that the market-based allocation would increase efficiency by 49% at the margin, lowering marginal abatement cost from 658€/tSO_2 to 338€/tSO_2. Surprisingly, I find that a technology mandate on scrubber usage would reap most of the efficiency gains of the market-based solution. This finding stands in stark contrast to the US SO_2 control experience, for which Carlson et al. (2000) find large efficiency gains from a market based policy compared to a scrubber mandate.

As a robustness check, I quantify how sensitive these estimates are to the inclusion of a measure for the marginal benefits of reducing air pollution. I use the back-of-the-envelope method employed by Oliva (2015) to construct a measure for the marginal abatement benefits of reducing SO_2 emissions at the province level. This method proceeds in 3 steps: (i) how does the SO_2 control policy change pollutant concentrations?, (ii) what health effects do the changes in pollutant concentrations cause?, and (iii) what is the monetary value of those health effects? I find that the gains from trade from moving to the cost-efficient allocation are robust to including the benefits of air pollution abatement: at the margin, welfare increases are 45%, a similar magnitude as the cost-based efficiency increases.

Overall, my empirical findings show that China's flagship air pollution control policy worked. SO_2 pollution, as measured through independent satellite data, decreased by more than 10% as a result of the policy. This success, however, appears to hinge on the Chinese government's ability to fully monitor SO_2 pollution. Before the 2008 changes in monitoring, compliance is mostly rhetorical. This finding provides support for a theoretical literature pioneered by Jean-Jacques Laffont suggesting the importance of regulatory capacity for the working of even simple policies (for an overview, see Laffont, 2005). Assessing the efficiency of the air pollution control policy, I find that the Chinese government fails to exploit the large heterogeneities in abatement cost across provinces. A more market-based allocation of SO_2 reduction targets, such as through an emissions trading scheme, would double allocative efficiency.

My analysis thus suggests that China currently faces problems in its air pollution control that are similar to the problems that today's developed countries faced in the past. Air pollution control is initially plagued by monitoring difficulties, and works in general, though it is not initially cost-efficient. Compared to environmental regulations in other emerging countries such as India and Mexico that have been evaluated in the literature (Duflo et al., 2013; Oliva, 2015), it is likely that China's better pollution monitoring capacities towards the end of the implementation explain the success of the policy in reducing air pollution.

This chapter is organised as follows: Section 2.2 discusses the related literature. Section 3.2.1 describes the setting and discusses environmental governance in China as well as the 2008 changes in the SO_2 monitoring. Section 2.4 explains my data sources, while Section 2.5 contains the empirical analysis. Section 2.6 constructs detailed marginal abatement cost curves at the province level and, in combination with a marginal benefit measure, computes the gains from trade across different policy instruments. Section 2.7 concludes.

2.2 Related literature

The two main reasons to study air pollution in China are the magnitude of the problem and the possibility to alleviate it through policy. Air pollution in China causes enormous health costs as shown by literatures in both economics and health. Chen, Ebenstein, et al. (2013), for instance, use a natural experiment to find that one coal-subsidy alone led to the loss of 2.5 billion life-years in Northern China. Epidemiological studies summarized in Yang et al. (2013) give the same sense of magnitude: they find air pollution to be the fourth most important health burden in China. In monetized terms, the health cost amount to 1.2 to 3.8% of GDP (World Bank and State Environmental Protection Administration, 2007). Air pollution furthermore induces direct productivity losses Chang et al. (2016). At the same time, Jia (2014) has shown in a convincing causal setting that pollution is a side effect of political incentives. A large literature in urban economics, political economy and environmental law backs this conclusion (Almond et al., 2009; Wang, 2013; Zheng and Kahn, 2013; Zheng, Sun, et al., 2014). It therefore seems possible to mitigate air pollution through the right combination of policies and incentives.

My study is the first to provide a causal empirical evaluation of the total emissions control policy in the 11th FYP. Despite the huge costs of air pollution in China, there has been no empirical evaluation of China's flagship air pollution control policy. Evaluation so far has come in one of two guises: through detailed narrative accounts of the changes (Hao et al., 2007; Schreifels, Fu, and Wilson, 2012) or through model-based studies in atmospheric science (Lu et al., 2010; Wang, Jang, et al., 2010a,b). Additionally, I am amongst the first to evaluate any environmental policy in China. The main other study I am aware of is Kahn, Li, and Zhao (2015), who analyze water pollution regulation. Their findings complement the empirical part of my research and allow for a richer interpretation of the role of pollution monitoring.

This research also contributes to two distinct literatures in environmental economics. The first is a nascent empirical literature that evaluates the design and implementation of environmental policies in imperfect settings, while the second literature studies the efficiency of the design of regulation through comparison of different policy instrument.

In the first literature, Duflo et al. (2013) study how the emissions of regulated industrial firms in India respond to changes in effective monitoring. Using a field experiment, they find that improved auditing leads regulated firms to reduce pollution, most notably through water pollution. Relatedly, Oliva (2015) provides a static snapshot of an air pollution regulation gone wrong. She documents the ineffectiveness of emissions regulation for private cars in Mexico City. Using a statistical test to detect misreporting, she shows that car owners often fail to comply and instead pay a bribe to circumvent regulation, leading to a private benefit to car owners at great damage to the public. Hansman, Hjort, and Leon (2015), by contrast, study the design of a reform aimed at preserving fish stock in Peru. They show that the piecemeal nature of the reform led to unforeseen side effects for air pollution outcomes.

This chapter sheds light on whether a simple command-and-control policy can be effective in a setting in which the government has a very low regulatory capacity initially. Due to the availability of independent satellite data on pollution throughout, I can study the behaviour of regulated agents in terms of likely misreporting, pollution outcomes and rhetorical compliance both before and after the Chinese government gains the ability to monitor pollution in 2008. Understanding how the affected agents comply with an environmental regulation under moral hazard is central to policy implementation and thus of great importance for contexts beyond China. The analysis of a command-and-control regulation in an emerging country such as China contributes furthermore to the existing literature on the effectiveness of air pollution control, which has focused almost exclusively on developed countries such as the US (Auffhammer and Kellogg, 2011; Chay and Greenstone, 2005; Chay, Dobkin, and Greenstone, 2003; Henderson, 1996)[2]. Additionally, while research on air pollution regulation has led to clear findings for pollutants such as TSP, the results for SO_2 are more mixed (Greenstone, 2004; Greenstone and Hanna, 2014).

The second literature to which I contribute studies the design of environmental regulation and the use of different policy instruments. This line of research asks a normative question - what would be the efficient allocation of SO_2 reduction targets? - and uses this benchmark to assess the efficiency losses of actual policy. Most contributions come from analyses of air pollution control regulation in the US, most notably through the US SO_2 mitigation efforts (for an overview, see Schmalensee et al., 1998 and Stavins, 1998), but a small number of studies exist from other countries such as Chile (Montero, Sanchez, and Katz, 2000). My work relates most closely to Carlson et al. (2000) and Oates, Portney, and McGartland (1989).

Carlson et al. (2000) compute marginal abatement cost curves for SO_2 for the electricity sector in the US to quantify the welfare gains from trade of moving from command-and-control regulation to SO_2 emissions trading. While those cost savings are large, they are significantly lower than anticipated ex ante. Carlson et al. (2000) show, however, that the efficiency gains from trading are large if they are compared instead to another policy instrument, a technology mandate in scrubber usage. My analysis follows their questions for the Chinese experience. Instead of starting from the market-based policy instrument and comparing the efficiency gains against different types of hypothetical command-and-control instrument, I start from the actual command-and-control instrument and compute a market-based policy instrument as counterfactual from which to judge efficiency gains. Based on these computations, I can then assess the efficiency loss of a hypothetical technology mandate in scrubber usage. Contrary to the US experience in Carlson et al. (2000), I find that a mandate on scrubber usage is nearly identical to the market-based solution for the national 10% control target. The reason for the difference across the two countries is that in the US, the gains from trade came mainly from differential access to low-sulfur coal, whereas in China the gains from trade are mainly driven by differences in industrial structure across provinces.

Another closely related paper is Oates, Portney, and McGartland (1989), who compare the efficiency of incentive-based regulation against command-and-control regulation to control air pollution in Baltimore. They find

[2]One exception are Greenstone and Hanna (2014), who investigate the effect of regulation on air and water pollution in India.

that a well designed command-and-control regulation can deliver pollution reductions at a welfare cost that can be lower than a comparable incentive-based regulation. This conclusion, however, only applies when that command-and-control regulation is informed by the marginal cost of abatement rather than by political considerations. Taken together, these studies show that while moving from a command-and-control regulation to a market-based regulation is generally seen as increasing the efficiency of the regulation (Schultze, 1977), whether this is so is an empirical question that depends on the particular case of the regulation under consideration. In particular, it will depend on whether the cost of abatement was taken into account when designing the the command-and-control allocation of targets[3].

I estimate detailed marginal abatement cost curves for SO_2 for each province in China in 2005, contributing to the few studies that estimate full marginal abatement cost curves in environmental economics in general (Gollop and Roberts, 1985; Carlson et al., 2000 and Abito, 2012)[4]. In particular, this study is the first to derive comprehensive marginal abatement cost curves at the province level in China. Two earlier contributions by Tu and Shen (2014) and Li, Wu, and Zhang (2015) provide interesting analysis in this direction but both studies rely on modelling in addition to microdata and only compute partial snapshots rather than full MAC curves. I use the marginal abatement cost curves to predict the counterfactual allocation of SO_2 reduction targets across provinces in China. This allows me to quantify the efficiency gains from trade from moving from the actual command-and-control allocation of SO_2 reduction targets to the cost-efficient allocation. Furthermore, I investigate the robustness of this quantification to the inclusion of a measure for the marginal abatement benefits. Overall, I contribute by studying whether the design of air pollution control regulation in a developing country such as China differs from the experience of developed countries like the US, and show possible differences in detail.

[3]Further related studies are Newell and Stavins (2003), who propose a simple way of calculating the efficiency gains from market-based policies in situations with limited data, and show an application to NO_x control in the US, and Antle et al. (2003), who investigate spatial heterogeneity in the abatement cost for carbon sequestration in the US and show that different policy instruments vary in efficiency by a factor of 5.

[4]Partial estimates of marginal abatement cost curves for compliance with air pollution control regulation are further reported in Hartman, Wheeler, and Singh (1997), Becker and Henderson (2001), Keller and Levinson (2002) and Becker (2005).

2.3 The SO_2 reduction policy in context

This section provides the context around the 11th Five-Year Plan (FYP) and China's flagship air pollution control policy that is the focus of this research. Environmental governance in China has undergone a rapid transformation in the last two decades. Until 2005, economic growth was the defining development paradigm. Environmental policies, where they existed, were paper tigers: they lacked political support from the central government and were rarely enforced. 2005 marks the turning point with a Five-Year Plan that 'was the most environmentally ambitious document in the history of the Communist Party' (Watts, 2011). The following paragraphs sketch how this change can best be seen as a change in political will rather than a change in formal laws.

Before the policy Laws regulating SO_2 emissions have existed in China since 1998, when the State Council approved the establishment of the 'Two Control Zones', a policy to address acid rain and SO_2 emissions (McElwee, 2011). Enforcement of this policy was intensified in 2000, but has remained constant since. Implementation of SO_2 policies, however, still encountered great difficulties (Gao et al., 2009), as China's overall development strategy remained firmly rooted in economic growth. Existing environmental policies overall, for instance on energy efficiency, were left underfunded by the central government (Gao et al., 2009; Lin, 2007). And while the 10th Five-Year Plan included a nationwide goal to reduce SO_2 emissions by 10%, it did not have political backing, and failed to induce SO_2 emissions reductions (Schreifels, Fu, and Wilson, 2012), possibly because of a lack of incentives for meeting the targets (Wang, 2013).

First change in the political outlook of the central government came in 2003, when President Hu - a hydroengineer - and Prime Minister Wen - a geologist - took power. The 'scientific development' paradigm, which emphasized environmental protection alongside economic growth, started to substitute for economic growth. Environmental governance, however, was still weak.

2005: The 11th Five-Year Plan (2006-2010) Amidst the increasing political will to implement and enforce environmental policies, the general directions of the 11th FYP started being discussed as early as mid-2003, and probably ended by 2004 (Xu, 2011). During the National People's Congress in March 2006, the 11th FYP was presented in its final form and included emissions control targets for air pollution (SO_2) and water pollution (chemical oxygen demand, or COD) as well as a target on energy efficiency. Concurrently, environmental governance started being taken seriously, when the once powerless SEPA successfully stopped hundreds of billions of Yuan of industrial investment on environmental grounds at the beginning of of 2005. This radical action came as a surprise to Chinese society (Gao et al., 2009). In March 2008, the SEPA's new political authority was formalized when SEPA obtained full rank in the State Council and received ministry status as the MEP (McElwee, 2011). This is the period that I analyze in this research.

The air pollution control target consisted of a 10% SO_2 emissions control target for China as a whole. This reduction target was handed down to the provinces in May 2006 at the latest, when SEPA - with high-level political backing - signed formal, binding reduction targets with the provincial governments (Gao et al., 2009; Xu, 2011)[5]. These reduction targets were given the highest political priority, paralleled only by mandates on growth, social stability and the one-child policy (Wang, 2013). Compliance with environmental targets entered the performance evaluation of local leaders (Moser, 2013). Provincial SO_2 reduction targets, in particular, were made a veto target: failure to comply would nullify all other performance achievements of a provincial leader (Kahn, Li, and Zhao, 2015; Xu, 2011). The reduction targets specified reductions in SO_2 emissions from 2006 to 2010 with 2005 as the baseline. These are the reduction targets that I use in this study. Figure 2.1 shows that these targets vary considerably, mandating reductions from 0% to more than 25%.

SO_2 emissions data in China in 2005 were of bad quality, and misreporting-prone. A province's SO_2 emissions were calculated by aggregating the physical quantity of coal used in a province in a given year, and by then multiplying this quantity with SO_2 emissions factors depending on the sulfur-intensity of the type of coal used. The political authority to compile these data rested with the provincial governments. While this procedure yields only coarse aggregate data at best (Guan et al., 2012), it can also be corrupted to produce the desired data. Anecdotal evidence and extensive field work in several provinces confirm that misreporting of SO_2 emissions data was prevalent in the first 2 years of the SO_2 control policy (Song et al., 2015). The likely reason why the central government decided to use emissions rather than other pollution data is that the network of *in situ* measurement stations was not of sufficiently high quality and quantity in 2005

[5]The 31 province-level administrative units include the four municipalities directly under the central government (Beijing, Chongqing, Shanghai and Tianjin) and the five autonomous regions (Guangxi, Neimongol, Ningxia, Tibet and Xinjiang).

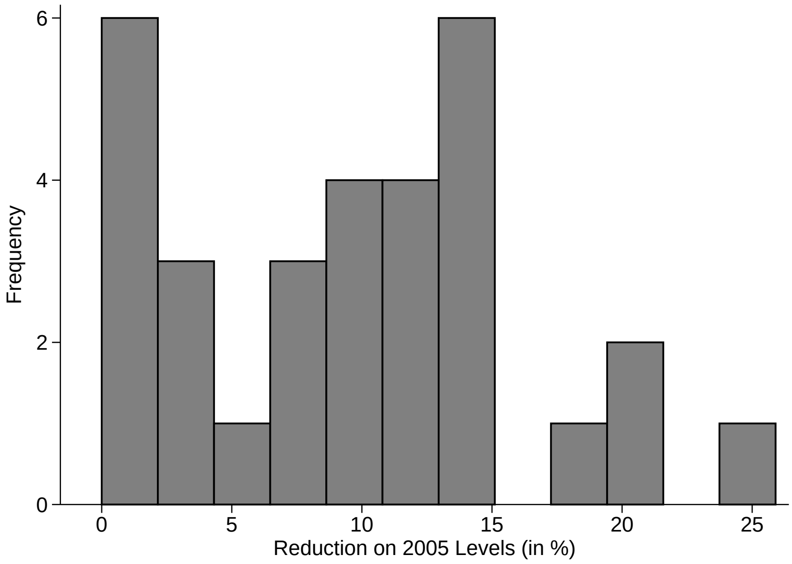


Figure 2.1 *The SO_2 emission control targets*. This figure shows the variation in the SO_2 emissions reduction targets across the 31 provinces. Targets are shown as percentage reduction on 2005 SO_2 emission baselines. The mean of the distribution is 9.4% and the standard deviation is 6.8 percentage points. Data source: State Council (2006).

to track SO_2 pollution outcomes on the ground.

Up to 2005, provincial governments in China had little incentive to control air pollution. A province in China is governed by a pair of provincial leaders, the governor and the party secretary. Provincial leaders are career officials who are appointed in a top-down manner. As career officials, they are often positioned outside their native provinces and move frequently as a results of promotion or demotion. Cadre regulations stipulate a maximum term length of 5 years for both governors and party secretaries (Kahn, Li, and Zhao, 2015). This rule is mostly respected: on average, provincial leaders have held office for just two years, with only a small number exceeding a tenure of 5 years. As a consequence of these frequent changes, provincial leaders are mainly concerned with furthering their own career in the short period that they hold office. Following the regulations in the past, this meant relying on pollution-intensive, quick-and-dirty GDP growth (Chen, Li, and Zhou, 2005; Jia, 2014; Li and Zhou, 2005). Incentives for provincial leaders to fix air pollution are further weakened by vested interests in polluting enterprises, of which local governments are often major shareholders (Gao et al., 2009). As provincial leaders got promoted to posts elsewhere, air pollution therefore remained unfixed.

2008: Changes to SO_2 monitoring In 2007, the State Council passed a law that fundamentally changed the regulator's capacity to monitor SO_2 pollution (the law is known as the 'Reduction of the Three Ways'). The core of the law was twofold: to change the politics of SO_2 data collection to avoid tampering with data at the political level, and to install appropriate monitoring equipment and ensure frequent statistical inspections on the ground (State Council, 2007).

On the political side, reporting was taken from provincial governments and put directly under the political control of the MEP. The MEP, in turn, directly reports to the State Council. On the ground, SO_2 measurement stations were build in pollution hotspots and the number of environmental monitoring officials was increased by 17% (Song et al., 2015). Key industrial polluters for each prefecture had their SO_2 emissions tracked on site. By May 2008, uninterrupted automatic monitoring devices with data feeds directly into the local environmental agency were used for this purpose (McElwee, 2011; Ministry of Environmental Protection of the People's Republic of China, 2008a,b). All changes became effective in July 2008 at the latest (Song et al., 2015). These changes mark a landslide for environmental governance in China: a nearly powerless agency was upgraded into a powerful ministry. A ministry given the tools to ensure it was going to be a force to be reckoned with.

The 2008 changes naturally divide the analysis into three periods: (i) before the policy; (ii) SO_2 reduction targets, but little monitoring; (iii) SO_2 reduction targets, and comprehensive monitoring. These periods will guide my subsequent empirical estimations to evaluate the policy.

2.4 Data

I compile a unique dataset that allows me to study the effect of the SO_2 control policy along two margins: Firstly, to evaluate the effect of the policy on pollution outcomes, I use two different data sources: (i) the official, misreporting prone SO_2emissions indicator and (ii) independent satellite data from NASA. Secondly, I evaluate the reactions by the regulated provincial governments. I divide these reactions into rhetorical compliance and real compliance. To measure rhetorical compliance, I build a novel dataset of political reports, which I quantify. Behaviour related to real compliance is based on official data sources on SO_2 abatement measures. I firstly describe the pollution data, and then the data related to the behaviour of the regulated provincial governments.

2.4.1 Data on pollution outcomes

Official SO_2 data The SO_2 control policy relies on SO_2 *emissions* as the official indicator. Note that this is a proxy indicator for the ultimate goal of reducing pollutant *concentrations*. It is likely that the central government chose this indicator due to a combination of a legacy of state-planning that focused on total emissions control and a lack of suitable SO_2 concentrations data from *in situ* measurement stations. I use the data reported in the *China Energy Databook* (Fridley, Romankiewicz, and Fino-Chen, 2013), who compile the official SO_2 emissions data from the statistical yearbooks.

The official SO_2 emissions data gives at best a noisy picture of true SO_2 emissions (Guan et al., 2012), and anecdotal evidence and research based on fieldwork suggest that the incentives before the 2008 policy changes led to severe misreporting (Song et al., 2015). The literature more generally has also noted misreporting of air pollution data (Andrews, 2008a; Chen, Jin, et al., 2012; Ghanem and Zhang, 2014). However, this kind of misreporting consists of relabeling around the politically sensitive *Blue Sky Day* threshold in air pollution index data. It is thus tangential to misreporting in SO_2 emissions data, affects the mean of the data very little, and there are signs that this particular kind of misreporting has come to an end with the introduction of $PM_{2.5}$ measurements from 2012 onwards (Stoerk, 2016). The potential source of misreporting was therefore removed after the MEP started collecting SO_2 pollution data directly in 2008.

SO_2 satellite data To measure real SO_2 pollution, I make use of the uniqueness of my empirical setting. In August 2004 - just before the start of the SO2 control policy - NASA launched a satellite with the Ozone Monitoring Instrument (OMI). In lay terms, OMI is an instrument that captures images of Earth from space at different wavelengths. Post-processing trough extraction algorithms produces SO_2 vertical columns of high precision that became available in 2014[6]. It is unlikely that the Chinese government would have had access to the satellite data during the policy, an these data were not used in the official evaluation of the policy. Dates with cloud cover can

[6]The new algorithm significantly improved the precision of the extracted vertical column densities and removed a number of biases compared to the earlier data product that was available from OMI (Krotkov et al., 2016). Note further that the OMI data itself have a detection threshold that is two magnitudes smaller than earlier satellite data and can thus enable the detection of SO_2 pollution from human activity in the lowest part of the atmosphere (NASA, 2014).

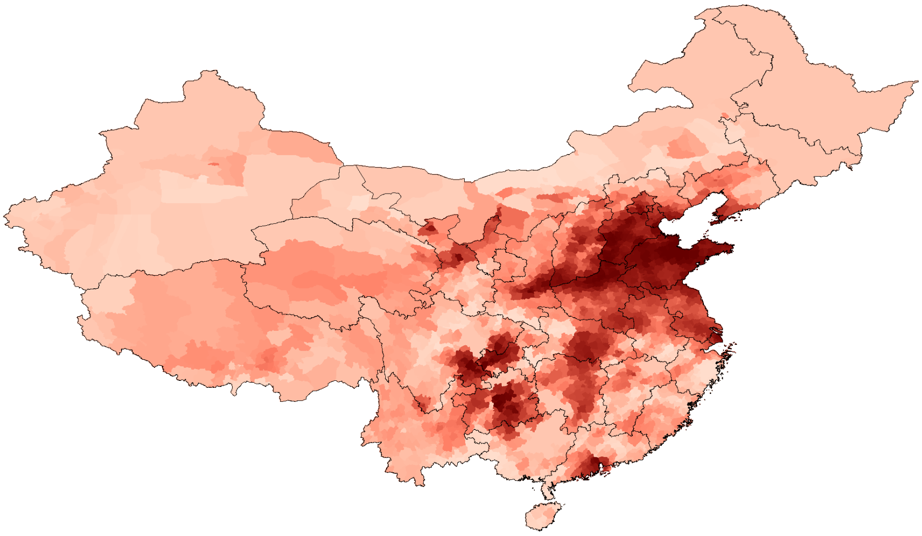


Figure 2.2 *SO_2 pollution in China in January 2006 based on NASA satellite data*. This figure shows the cross-section of the SO_2 satellite data based on the NASA OMI SO_2 data product for January 2006 mapped to the county-level.

lead to missing values over individual pixels, but this is not generally considered a first-order problem (Krotkov et al., 2016)[7]. All in all, the NASA SO_2 satellite data are a very good proxy for ground-level SO_2 emissions. To illustrate, Figure 2.2 shows the cross-section in January 2006.

Relation between both data sources SO_2 emissions decay in a span of 4-36 hours (Fioletov et al., 2015; He, 2012). Since the satellite data capture Earth daily, they represent a snapshot of SO_2 pollution on that day. Given the quick decay, this prevents leakage from confounding the outcome and allows me to capture local rather than transported SO_2 emissions. NASA's retrieval algorithm produces four different data products, each of which corresponds to SO_2 pollution at different levels of altitude in the atmosphere (NASA, 2014). For this analysis, I use the lowest level at an altitude of 900m above ground, for two reasons: firstly, because this is the best proxy for anthropogenic emissions sources, and secondly, because a lower altitude further minimizes transportation. Secondly, the precision of the satellite images is high enough to identify individual sources of pollution that produce as little as 30kt of SO_2 annually (Fioletov et al., 2015). I aggregate these daily cross-sections to the province-month and the province-year

[7]Furthermore, I aggregate daily values to at least monthly, and pixels of not bigger than 0.25 degrees by 0.25 degrees latitude-longitude to counties and provinces, further reducing the magnitude of the potential problem due to clouds.

levels. Annual changes in the SO_2 satellite data can therefore be expected to be mimicked closely in the official statistics[8].

2.4.2 Data on reactions by the regulated provincial governments

This section describes the data sources that I use to study the reactions of the provincial governments along two margins: (i) rhetorical compliance and (ii) real compliance.

Rhetorical compliance I build a comprehensive dataset of political statements by each provincial government to study whether the regulated provincial governments respond to the SO_2 reduction targets by mimicking the central government's rhetoric. In China, each provincial government has the obligation to issue a government work report every year. This report is publicly delivered by one of the two highest ranked officials in the province, the party secretary or the governor. Each report contains information on the provincial government's activities and achievements. These reports are divided into two parts: Part 1 discusses the provincial government's work and achievements in the preceding period, while Part 2 discusses the work in the period to come. This unique setting allows me not only to investigate political rhetoric in general, but to also specifically investigate rhetorical responses relating to past and future achievements in response to the SO_2 reduction targets. Figure 2.3 shows an example of a government work report from Liaoning province in 2003, delivered by then-governor Bo Xilai.

To measure the provincial government's political attention towards air pollution for a given province-year, I scan each province's government work report for the years 2002-2010 and construct a variable that is equal to the number of occurrences of keywords related to air pollution[9]. Keywords were chosen from a technical document on urban air pollution in developing countries (GTZ, 2009) as well as from China-specific air pollution articles, webpage entries and blog posts in March 2014 from *China Daily*, *Global Times*, *Beijing Review* and *Jinyang Yangcheng Evening News*. To further rule out cherry-picking of keywords, I have defined the list of keywords as widely as possible. Figure 2.4 contains the raw count of keywords over time, offering two take-aways. Firstly, there is a distinct increase in air pollution related keywords: Mentions of air pollution increase by more than

[8] A cross-check of known ground-level emission sources with SO_2 OMI satellite data revealed a correlation of 0.91 (Fioletov et al., 2015).

[9] 8 reports are missing in 2002 and 1 report is missing in 2003.


政府工作

当前位置：网站首页 -> 政府信息 -> 政府工作 -> 省政府工作报告

政府工作 The Government Work

【 字体大小：大 | 中 | 小 】 【打印文章】 【关闭窗口】

2003年省政府工作报告

——2003年1月20日在辽宁省第十届人民代表大会第一次会议上

省长 薄熙来

各位代表：

现在，我代表省人民政府，向大会报告工作，请予审议，并请省政协委员和列席同志提出意见。

一、本届政府五年工作的回顾

过去的五年，正处世纪之交。全省各族人民在辽宁省委的领导下，高举邓小平理论伟大旗帜，认真贯彻“三个代表”重要思想，同心协力，艰苦奋斗，开拓进取，克服了种种困难，改革开放和现代化建设取得了显著成就。

——国民经济取得新进展。国内生产总值2002年完成5458亿元，五年年均增长8.9%。社会消费品零售总额2002年实现2258亿元，五年年均增长9.3%；地方财政一般预算收入2002年完成398.8亿元，年均增长12%。固定资产投资五年共完成6449亿元。

Figure 2.3 *Example of a Government Work Report*.The graph shows the beginning of the government work report for Liaoning province in 2003. It was delivered by then-governor Bo Xilai.

400% during the 11th Five-Year Plan (2006-2010). Secondly, one keyword drives this increase: the specific keyword 'sulfur'. This keyword is directly related to the provincial SO_2 reduction targets from the 11th Five-Year Plan. The outcome variable therefore appears to capture relevant political statements[10].

Real compliance Based on my own calculations of the marginal cost of SO_2 emissions abatement (see Section 2.6), the installation of desulfurization devices in existing industrial and power plants (scrubbers), fuel-switching to better quality coal and the shutdown of small, inefficient thermal units are the main margins by which the provincial governments could reduce SO_2 emissions over the relatively short time horizon of 5 years. I collect data on both the timing and the quantity of installation of desulfurization devices at the province level during the 11th Five-Year Plan (2006-2010). Furthermore, I compute the number and capacity of small thermal

[10]One valid concern with the raw count of air pollution keywords is that the length of the government work reports might have changed over time or across provinces. This could introduce a bias in my measure. To show that this is not the case, Figure 2.A.4 in Appendix 2.A divides the number of keywords by the length of the report and finds a very similar pattern. I prefer to use raw data because of its clearer interpretation. All regressions results are robust to using the adjusted number of keywords. These results are available upon request.

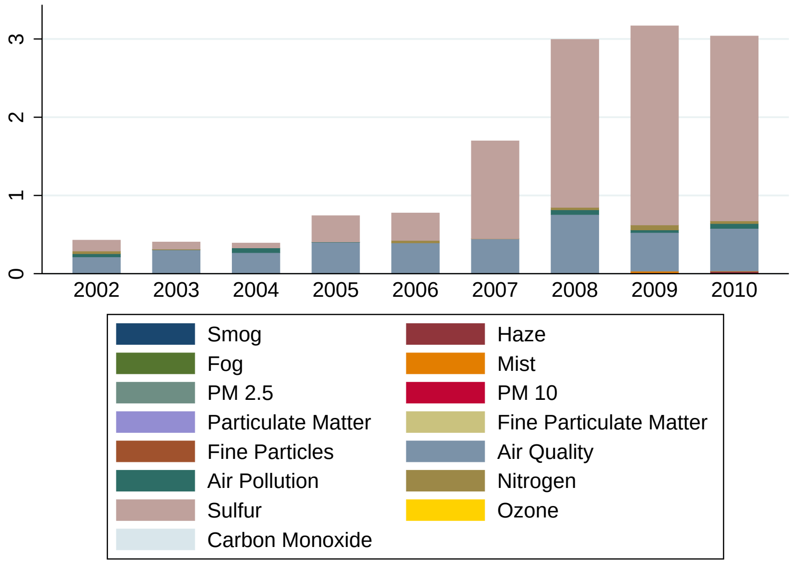


Figure 2.4 *Political attention to air pollution over time*. The graph shows the mean count of keywords related to air pollution for all provincial government work reports in a given year from 2002 to 2010.

units that were shut down by 2010 in each province. The latter are based on a planning document from 18th January 2008, in which the MEP asked the provincial governments to submit a concrete proposal for the thermal units to be shut down over the following two years. Anecdotal reports suggest that these shutdowns did indeed take place. All data were compiled from sources available through the data center of the MEP (datacenter.mep.gov.cn). I do not analyze the sulfur content of coal used at the province-year level since reliable data is unavailable.

2.5 Empirical analysis

2.5.1 Empirical strategy

Baseline specification I use the following difference-in-differences model to investigate whether a higher SO_2 reduction target led to a relatively stronger decrease in SO_2 emissions:

$$y_{pt} = \beta_0 + \beta_1 Reductiontarget_p \times D(Post)_t + \beta_2 Reductiontarget_p + \sum_{t=1}^{T} \beta_{3t}\gamma_t + \alpha_p + u_{pt} \tag{2.1}$$

The outcome y_{pt} for province p at time period t is either the official SO_2 emissions data that was used by the central government to assess the policy or the independent satellite SO_2 data. The variable $Reductiontarget_p$ is the provincial SO_2 reduction target and captures the cross-sectional variation shown in Figure 2.1. Policy variation over time is captured in the indicator $D(Post)$ that takes on the value 1 from 2006 onwards. α_p are province fixed effects. The estimate for β_1 gives the causal effect of an increase in the target stringency by one unit for the whole period of the 11th Five-Year Plan, from 2006 to 2010.

Identification The DID specification exploits cross-province variation in the stringency of the SO_2 reduction target to estimate the causal effect of the pollution control policy. Identification relies on a combination of three factors: (i) common trends in the outcome variables prior to the SO_2 control policy, (ii) a sharp deviation from those trends following the policy changes, and (iii) the absence of forward-looking considerations that would also explain SO_2 abatement efforts by the provincial governments in 2006.

Provincial governments can rely on three main channels to bring down pollution. All of which are quick to implement, in particular for a country like China: fuel-switching to higher quality coal with a lower sulfur content, installation of desulfurisation devices, and the shutdown of small, inefficient thermal units. It is therefore reasonable to expect an immediate effect of the SO_2 reduction policy on pollution outcomes. Lu et al. (2010) provide evidence that abatement measures started immediately: power plants already started to switch to better quality coal in 2005 (with a sulfur content reduced by about 20%compared to the preceding year), and flue-gas desulfurization technology doubled from below 10% to more than 20% of all operating power plant capacity.

Common pre-trends and the timing of effects are empirically testable, and I show below that these conditions are fulfilled. Consideration (iii) is not directly testable, but supporting evidence shows that it is likely fulfilled. While the exact algorithm used by the Chinese government to allocate the targets is unknown, and it is unlikely that the targets were distributed randomly, random allocation is not needed for my identification. Instead, I only need that the allocation of SO_2 reduction targets across provinces was independent of forward-looking considerations that would explain SO_2 pollution reductions by a province independent of the SO_2 reduction targets.

The official statement by the State Council on how the target distribution would have taken place mentions a whole array of factors that were used to determine the allocation of targets for a province (State Council, 2006): (i) environmental quality and environmental capacity, (ii) current pollution levels, (iii) level of economic development, (iv) SO_2 mitigation capabilities and (v) regional differentiation (Eastern, Central, Western). Xu (2011) finds that the allocation of targets does not correlate with either of those factors (barring a correlation of non-power SO_2 emissions divided by the area of a province). Therefore, there is no evidence that any of these factors drove the target allocation, which suggests that the allocation of targets followed guidelines orthogonal to changes in SO_2 emissions at the turn of the 11th FYP. I also rule out the possibility that the allocation of SO_2 reduction targets followed the cost of abatement, both by itself and net of benefits. If that were the case, my empirical strategy would pick up the compound effect of an SO_2 reduction target and a cost advantage. In Section 2.6, I derive detailed marginal abatement cost (MAC) curves and combine those with a measure of marginal abatement benefits. I find that neither the MAC nor the marginal welfare impacts correlate with the target allocation, lending further credibility to my empirical strategy (shown in Figures 2.16 and 2.A.5). Furthermore, I empirically test for the influence of both factors in a robustness check below.

Inference I follow Bertrand, Duflo, and Mullainathan (2004) and compute standard errors that are heteroskedasticity-robust and clustered at the province level. Statistical inference based on these standard errors, however, could still be incorrect if the number of clusters is too small, as the required asymptotic results might not apply. China has 31 provinces, yielding significantly less than 50 clusters, the usual rule of thumb. While reporting the clustered standard errors, I therefore base my statistical inference on p-values derived from the wild bootstrap method described in Cameron, Gelbach, and Miller (2008). This is common practice in applied research on China (see, for instance, Martinez-Bravo et al., 2013).

2.5.2 The effect of the SO_2 control policy on SO_2 pollution

Baseline results Table 2.1 provides the summary statistics for my sample. The first two columns of Table 2.2 show the results from estimation Equation (2.1) for the effect of the policy for the full 11th FYP. A first glance reveals that the policy had a different effect depending on the indicator

Table 2.1 Descriptive statistics

	Mean	Std. Dev.	Min.	Max.	Time Period
SO2 Emissions (kt)	739	471	1	2003	2002-2010
SO2 Satellite (DU)	0.37	0.39	-0.02	1.93	2005-2010
Selected SO2 Sat. (DU)	0.72	0.53	-0.03	2.57	2005-2010
SO2 Reduction Target (% on 2005 Baseline)	9.65	6.71	0	25.9	-

Note. Satellite data is measured in Dobson Units (DU). Selected SO_2 Sat. is the sample of polluted cities. Satellite SO_2 measurements below 0.2 Dobson Units are generally considered as clean air and negative values are likely noise from measurement error. Replacing negative values as either 0s or missing does not change the subsequent results.

used for evaluation. According to the official SO_2 emissions data, the policy was a success: SO_2 emissions decrease in response to the target, with the estimated magnitude being a 5.8% decrease for a one-standard deviation increase in target stringency. The satellite data, by contrast, do not show a significant relationship between the targets and the SO_2 pollution outcomes. The sign of the estimated coefficient is in the same direction as with the official indicator, and the estimated magnitude is even higher, but it is not statistically different from zero.

To improve the precision of the estimates, I increase statistical power by focussing on polluted cities only. Given the same amount of noise in the data, a higher absolute effect can be expected to be more easily detected in this sample. The sample of polluted cities is built by taking the location of each *in situ* measurement stations run by the MEP[11]. The point estimate for the effects of a higher reduction target is nearly identical to the overall sample in percentage terms, but the coefficient turns significant because the higher absolute effect in the polluted sample increases statistical power (Table 2.2). All baseline estimates taken together, I find that that the SO_2 control policy was effective in reducing air pollution over the whole sample period.

[11] I select the 25 nearest pixels up to a distance of at most 25 kilometers around the city centroids, and compute the province-level observations only based on those pixels. Results from this sample are robust to changes in the 25 kilometers cutoff, as most pixels are closer than 15 kilometers from the city centroid.

Table 2.2 The effect of the policy for the whole period (2006-2010)

	SO2 Emissions (Kt)	Satellite SO2 (Dobson Units)	Selected Sat. SO2 (Dobson Units)
Reductiontarget × D(Post)	-6.3***	-0.0057	-0.0115*
	(1.6)	(0.0040)	(0.0050)
	[0.01]	[0.27]	[0.06]
Reductiontarget	0.3	0.0214***	0.0251***
	(1.34)	(0.0067)	(0.0076)
	[0.74]	[0.00]	[0.00]
D(Post)	107.87***	0.0482	0.0838**
	(17.7)	(0.0286)	(0.0386)
	[0.00]	[0.12]	[0.04]
Province FE	✓	✓	✓
Effect size (% of mean/σ)	*5.8%*	*10.5%*	*10.9%*
Observations	279	186	186
Provinces	31	31	31
R^2	0.96	0.90	0.84

Note. Heteroskedasticity-robust standard errors clustered at the province level are shown in parentheses. Wild bootstrap p-values from 1000 repetitions clustered at the province level are shown in square brackets. The effect size gives the estimated coefficient of the interaction term β_1 in Equation (2.1) as percentage of the mean of the dependent variable for a one standard deviation(σ)-increase in the SO_2 reduction target. $***p < 0.01$, $**p < 0.05$, $*p < 0.1$.

Dynamic treatment effects To take a closer look at what might explain these differing findings, I estimate yearly versions of the above DID specification. Instead of collapsing the time periods into before and after the policy as in Equation (2.1), I interact a dummy for each time period with the reduction target. This yields the following equation:

$$y_{pt} = \beta_0 + \sum_{t=1}^{T} \beta_{1t} Reductiontarget_p \times \gamma_t + \beta_2 Reductiontarget_p + \sum_{t=1}^{T} \beta_{3t}\gamma_t + \alpha_p + u_{pt} \tag{2.2}$$

As before, y_{pt} are the SO_2 pollution outcomes for province p in time period t. In this specification, the SO_2 reduction targets are interacted with each time period γ_t to estimate the differential trajectory of SO_2 pollution for

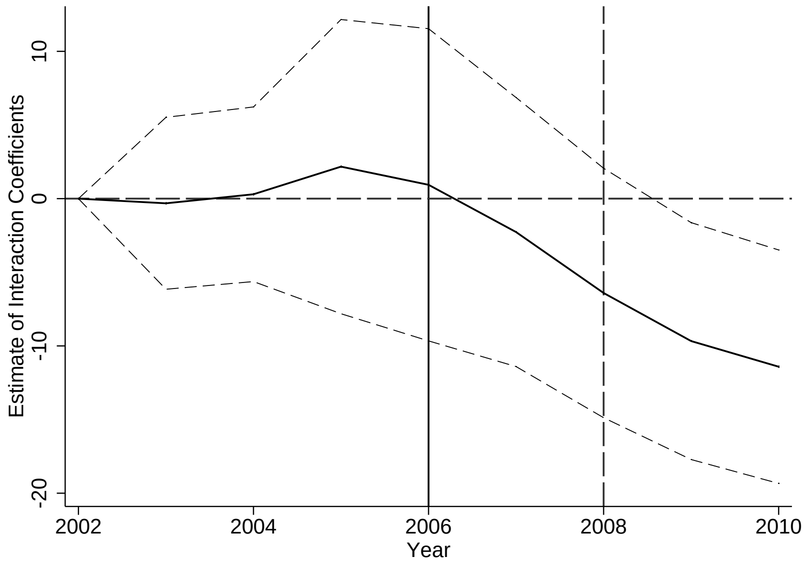


Figure 2.5 *Dynamic treatment effects: Official SO_2 emissions*. The solid line plots the point estimate for yearly coefficient estimates of the interaction coefficients $\sum_{t=1}^{T} \beta_{1t}$ in Equation (2.2) for the official SO_2 emissions data. The dashed lines show the 95% confidence bands, based on heteroskedasticity-robust standard errors clustered at the province-level. The excluded time period $t = 1$ is the year 2002. The vertical line before 2006 marks the start of the policy and the vertical line before 2008 marks the start of SO_2 monitoring.

provinces with different SO_2 reduction targets. These effects are captured in the point estimates for β_{1t} for periods 2 through T, where T is the last observation for 2010.

Figures 2.5 and 2.6 plot the estimate of the interaction coefficients $\sum_{t=1}^{T} \beta_{1t}$ in equation (2.2) and show that the identifying assumption of common pre-trends is satisfied. Both indicators are on a common trend before the start of the policy in 2006. However, there is only one year of satellite data before the start of the policy, because NASA only launched the satellite in late 2004. Figure 2.7 zooms in on the satellite data and plots the estimates for $\sum_{t=1}^{T} \beta_{1t}$ estimated on monthly satellite data. While there is more noise in the monthly data, there are common pre-trends before the start of the policy, as illustrated through the vertical grey lines. Table 2.3 shows the corresponding estimates.

The official SO_2 emissions data also show a clear deviation from the pre-trend at the start of the policy, lending further support to the identification. SO_2 emissions decrease immediately and linearly until the end of the 11th

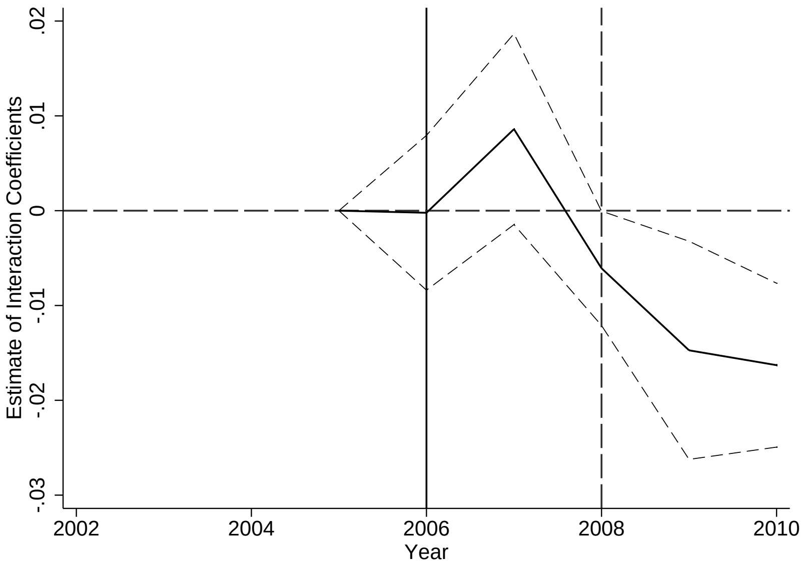


Figure 2.6 *Dynamic treatment effects: SO_2 satellite data*. The solid line plots the point estimate for yearly coefficient estimates of the interaction coefficients $\sum_{t=1}^{T} \beta_{1t}$ in Equation (2.2) for the NASA SO_2 satellite data. The dashed lines show the 95% confidence bands, based on heteroskedasticity-robust standard errors clustered at the province-level. The excluded time period $t = 1$ is the year 2005. The vertical line at 2006 marks the start of the policy and the vertical line at 2008 marks the start of SO_2 monitoring.

FYP. On the other hand, a look at the SO_2 satellite data paints a different picture: SO_2 pollution does not go down for provinces with a higher reduction target when the central government cannot monitor SO_2. After 2008, the satellite data show a distinct drop in SO_2 pollution. Taken together, I find the following: (i) there is possible misreporting of SO_2 emissions during the first two years of the 11th FYP when the Chinese government lacks the ability to properly monitor SO_2 emissions. (ii) once the government gains the ability to monitor, rhetorical compliance turns into real compliance and SO_2 pollution is reduced. Ultimately, the policy worked.

Heterogeneous treatment effects based on initial pollution levels Where do the improvements in air pollution take place? Figure 2.2 shows a large heterogeneity in the location of SO_2 pollution across provinces. Sichuan, for instance, suffers from a pollution hotspot towards the East, and enjoys comparatively lesser pollution in the Western part of the province. The NASA satellite data allow me to exploit this heterogeneity to ask whether

Table 2.3 The effect of the policy for each year of the 11th Five-Year Plan (2006-2010)

	SO2 Emissions (Kt)	Satellite SO2 (Dobson Units)	Selected Sat. SO2 (Dobson Units)
2002× Reductiontarget	Excluded	-	-
2003× Reductiontarget	-0.32	-	-
	(3.93)		
	[0.93]		
2004× Reductiontarget	0.30	-	-
	(3.64)		
	[0.93]		
2005× Reductiontarget	2.17	Excluded	Excluded
	(4.04)		
	[0.66]		
2006× Reductiontarget	0.94	-0.00	0.00
	(4.08)	(0.00)	(0.01)
	[0.86]	[0.93]	[0.94]
2007× Reductiontarget	-2.27	0.01	0.01
	(3.70)	(0.00)	(0.01)
	[0.62]	[0.21]	[0.29]
2008× Reductiontarget	-6.40	-0.01	-0.01***
	(3.69)	(0.00)	(0.01)
	[0.14]	[0.11]	[0.00]
2009× Reductiontarget	-9.67**	-0.01**	-0.02**
	(3.79)	(0.00)	(0.00)
	[0.02]	[0.02]	[0.01]
2010× Reductiontarget	-11.41***	-0.02***	-0.03***
	(3.89)	(0.00)	(0.00)
	[0.00]	[0.00]	[0.00]
Province FE	✓	✓	✓
Mean dep. var.	739	0.37	0.72
Observations	279	186	186
Provinces	31	31	31
R^2	0.98	0.96	0.93

Note. Heteroskedasticity-robust standard errors clustered at the province level are shown in parentheses. Wild bootstrap p-values from 1000 repetitions clustered at the province level are shown in square brackets. The yearly interaction coefficients are estimates for $\sum_{t=1}^{T} \beta_{1t}$ in Equation (2.2), while 'Excluded' is the omitted time period. Note that the SO_2 reduction started in 2006 while government monitoring of SO_2 pollution became effective in 2008. $***p < 0.01$, $**p < 0.05$, $*p < 0.1$.

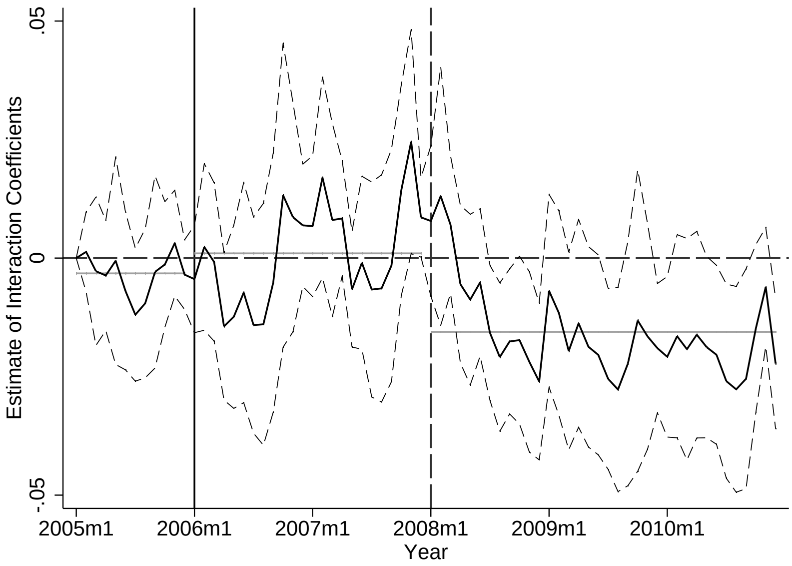


Figure 2.7 *Dynamic treatment effects: SO_2 satellite data (Monthly)*. The solid line plots the point estimate for monthly coefficient estimates of the interaction coefficients $\sum_{t=1}^{T} \beta_{1t}$ in Equation (2.2) for the NASA SO_2 satellite data. The dashed lines show the 95% confidence bands, based on heteroskedasticity-robust standard errors clustered at the province-level. The horizontal grey lines plot the average of the point estimates of the interaction coefficient in the pre-period (2005), the period without monitoring (2006-2007) and the period with monitoring (2008-2010). The excluded time period $t = 1$ is January 2005. The vertical line at 2006 marks the start of the policy and the vertical line at 2008 marks the start of SO_2 monitoring.

the effect of the SO_2 pollution reduction target in the province differs depending on the initial level of pollution. To answer this question, I map the SO_2 satellite data to the county-level, yiedling 2,638 cross-sectional units. For each county, I measure its mean SO_2 pollution for 2005 relative to all other counties within the same province. This information is captured in a variable that takes the value of 1 for counties in the lowest quartile of initial pollution within their province, up to a value of 4 for counties that have the highest initial pollution. I then re-estimate the DID specification from Equation (2.1) at the county level on each subsample along the distribution of initial pollution. As above, statistical inference relies on heteroskedasticity standard errors at the province-level and t-statistics from a wild bootstrap procedure with 1000 repetitions. I expect to find no effects for the lowest quartile, because initial pollution in these counties is below

0.2 Dobson Units on average, making further air quality improvements unlikely. Furthermore, for the same precision of data, a nominally smaller effect is harder to detect statistically, making it more likely to find a significant effect the higher the initial level of pollution.

The results in Table 2.4 confirm that this is the case. As expected, SO_2 reductions in response to the targets mainly take place in the higher two quartiles of the initial distribution, though only the most polluted quartile is statistically significant. This reproduces the baseline results, where the effect only turned statistically significant for the sample of polluted cities due to noise. The effect size for the two highest quartiles is a 9.4% and 9.9% decrease in SO_2 pollution per one standard-deviation increase in the stringency of the reduction target, respectively.

Robustness checks for identification Firstly, I show that the neither the cost nor the welfare impact at the margin correlates with the SO_2 reduction targets. To test this, I use the following empirical specification:

$$\begin{aligned} y_{pt} = \beta_0 &+ \beta_1 Reductiontarget_p \times D(Post)_t + \beta_2 X_p \times D(Post)_t \\ &+ \beta_3 Reductiontarget_p + \beta_4 X_p + D(Post)_t + \alpha_p + u_{pt} \end{aligned} \tag{2.3}$$

where X_p is either a province-level measure of (i) the marginal abatement cost or (ii) the ratio of marginal abatement benefits to marginal abatement cost. Specific details for the construction of those measures are provided in Section 2.6. In a nutshell, my approach is this: I compute detailed marginal abatement cost curves for SO_2 for each province in China, based on a reliable set of micro data on the cost and abatement potential of fine-grained polluting activities. The marginal benefits from reducing air pollution are based on a back-of-the-envelope calculation that follows Oliva (2015) and evaluates health improvements based on the value of a statistical life.

Table 2.4 Heterogeneous treatment effects depending on initial pollution levels

	SO2 Satellite Data			
	Quartile 1	Quartile 2	Quartile 3	Quartile 4
Reductiontarget	0.0004	-0.0032	-0.0082	-0.0119*
× D(Post)	(0.0028)	(0.0030)	(0.0055)	(0.0053)
	[0.93]	[0.41]	[0.26]	[0.07]
Reductiontarget	0.0311***	0.0248***	0.0224***	0.0210***
	(0.0023)	(0.0025)	(0.0046)	(0.0044)
	[0.00]	[0.00]	[0.00]	[0.00]
D(Post)	0.0356*	0.0328	0.0382	0.0293
	(0.0178)	(0.0173)	(0.0274)	(0.0365)
	[0.06]	[0.10]	[0.26]	[0.47]
Province FE	✓	✓	✓	✓
Effect size (% of mean/σ)	-1.5%	5.8%	9.9%	9.4%
Observations	47,376	47,592	47,016	48,096
Provinces	31	31	31	31
R-squared	0.54	0.60	0.59	0.61

Note. The table reports estimates for Equation (2.1) at the county-yearmonth level for different subsamples. Quartile marks the quartile of initial based on its 2005 SO_2 pollution, calculated from SO_2 satellite data, relative to the mean SO_2 pollution within the same province. Heteroskedasticity-robust standard errors clustered at the province level are shown in parentheses. Wild bootstrap p-values from 1000 repetitions clustered at the province level are shown in square brackets. $***p < 0.01$, $**p < 0.05$, $*p < 0.1$.

Table 2.5 Controlling for marginal abatement cost and marginal welfare impact

	SO2 Emissions (Kt)		SO2 Satellite (Dobson Units)		Sel. Satellite SO2 (Dobson Units)	
Reductiontarget	-6.26**	-6.15***	-0.0057	-0.0058	-0.0116*	-0.0117**
× D(Post)	(1.62)	(1.63)	(0.0040)	(0.0041)	(0.0050)	(0.0051)
	[0.01]	[0.00]	[0.23]	[0.23]	[0.06]	[0.04]
Reductiontarget	0.33	1.33	0.0214***	0.0364***	0.0252***	0.0461***
	(1.36)	(1.44)	(0.0067)	(0.0055)	(0.0077)	(0.0062)
	[0.81]	[0.22]	[0.00]	[0.00]	[0.00]	[0.00]
MAC	0.01		-0.0000		-0.0000	
× D(Post)	(0.02)		(0.0000)		(0.0000)	
	[0.55]		[0.90]		[0.19]	
MAC	-0.12***		-0.0003***		-0.0003***	
	(0.02)		(0.0000)		(0.0001)	
	[0.00]		[0.00]		[0.00]	
MAC/MAB		-0.59		0.0001		0.0009
× D(Post)		(0.61)		(0.0015)		(0.0016)
		[0.45]		[0.90]		[0.39]
MAC/MAB		0.71*		0.0053***		0.0068***
		(0.60)		(0.0011)		(0.0012)
		[0.08]		[0.00]		[0.00]
D(Post)	99.97**	113.20***	0.0492	0.0470*	0.0992**	0.0753*
	(21.02)	(20.20)	(0.0316)	(0.0271)	(0.0465)	(0.0395)
	[0.02]	[0.00]	[0.11]	[0.08]	[0.03]	[0.07]
Province FE	✓	✓	✓	✓	✓	✓
Observations	279	279	186	186	186	186
Provinces	31	31	31	31	31	31
R^2	0.96	0.96	0.90	0.90	0.84	0.84

Note. Heteroskedasticity-robust standard errors clustered at the province level are shown in parentheses. Wild bootstrap p-values from 1000 repetitions clustered at the province level are shown in square brackets. The table reports the results from estimating Equation (2.3). MAC refers to the marginal abatement cost given the actual SO_2 reduction target and MAC/MAB is the ratio of the marginal abatement benefits to the marginal abatement cost. $***p < 0.01$, $**p < 0.05$, $*p < 0.1$.

Table 2.5 shows that the estimates for β_2 are 0 and that the estimates for β_1 are nearly identical to those in Table 2.2. Neither the marginal abatement cost by themselves nor the marginal abatement benefits change the estimated effect of an SO_2 reduction target.

Secondly, I find that there is no relationship between the economic downturn in the Great Recession and the stringency of a province's SO_2 reduction targets. If this were the case, my results could be confounded because a slowdown in economic activity could go hand in hand with a reduction in SO_2 pollution. It is important to note that even though China experienced a slowdown in growth in 2009, this was only a decrease in growth rates. Even in 2009, there are only 3 provinces with a growth rate below 5%, and the growth rate for those provinces is still positive. To check whether the 2009 downturn is correlated with the SO_2 reduction targets, I compute the magnitude of the downturn as the deviation in the average of the growth rates for 2002-2007 for each province. I find that there is no relationship between the magnitude of the recession and the reduction targets in a province. This is because different provinces with similar reduction targets experienced rather different deviations from their long term growth rates in 2009, and overall there is no statistically significant correlation between SO_2 reduction targets and the downturn. A linear regression with standard errors clustered at the province level finds a best fit with a p-value of 0.31 (these results are shown in Figure 2.A.1 in Appendix 2.A). The Great Recession can therefore not explain the decrease in SO_2 pollution from 2008 onwards.

2.5.3 The effect of monitoring

This subsection provides suggestive evidence that the difference between the official SO_2 emissions data and the SO_2 satellite data before 2008 is driven by the government's newly-gained ability to monitor SO_2 pollution. Firstly, it is important to note that the difference between the NASA satellite data and the official SO2 data starts growing in 2006 and spikes in 2007. From 2008 on, the difference shrinks again to less than the initial levels and stays there by 2010. Secondly, this spike is related to the SO2 reduction targets: the interaction between the year dummy for 2007 and the reduction target of a province from estimating Equation (2.2) is positive, albeit only significant at p=0.08 for the cluster-robust standard errors or p=0.13 for the wild bootstrap procedure. Both analyses are shown in Figures 2.A.2 and 2.A.3 in Appendix 2.A

I estimate heterogeneous treatment effects by splitting the sample into provinces for which misreporting before 2008 is more likely than for others. Factors that influence the propensity of regulated provincial governments to take advantage of lacking monitoring and to misreport untrue pollution improvements initially can be grouped into: (i) technological factors that determine the cost of abatement; (ii) socioeconomic factors, such as population density that determine the benefits of abatement; and (iii) the career concerns of provincial leaders themselves. I provide evidence for all three categories.

I construct detailed marginal abatement cost curves (Section 2.6) and split provinces into those where the marginal abatement cost is high given the actual reduction targets. Marginal abatement costs range from 170€/tSO_2 (Gansu) to 660€/tSO_2 (Heilongjiang). My hypothesis is that those provinces with higher abatement costs will have a higher incentive to misreport in the first period. Next, I compute the marginal benefits for each province and split the sample into two groups according to their welfare impacts. Finally, I use data on party secretaries from Persson and Zhuravskaya (2016) to identify provinces led by politicians with career concerns. I hypothesize that provinces with leaders in the last term in office or with more local ties are less career concerned will engage in less misreporting when the government cannot monitor real compliance[12].

The following triple-differences specification interactions the reduction target with an indicator variable that takes on the value of one for the subsample with more likely misreporting according to the hypotheses discussed above. In this way, I can test whether the likely reason that explains why compliance becomes real in 2008 is in fact the government's ability to monitor SO_2 pollution outcomes. I restrict the sample to the years until 2008 since this test concerns only the subperiod during which the government cannot monitor SO_2 pollution.

$$\begin{aligned} y_{pt} =& \beta_0 + \beta_1 Reductiontarget_p \times D(Post)_t + \beta_2 Reductiontarget_p + \beta_3 D(Post)_t \\ &+ \beta_4 H_p + \beta_5 Reductiontarget_p \times H_p + \beta_6 H_p \times D(Post)_t \\ &+ \beta_7 Reductiontarget_p \times D(Post)_t \times H_p + \alpha_p + u_{pt} \end{aligned} \tag{2.4}$$

In this specification, H_p is an indicator that takes the value 1 for those provinces more likely to engage in misreporting when the government can-

[12]According to Persson and Zhuravskaya (2016), politicians with local ties are more prone to contributing to local public goods.

Table 2.6 The effect of monitoring: Overview

y_{pt}	Misreporting Unlikely	Misreporting Likely	Hypothesis
SO2 emissions	$\hat{\beta}_1$	$\hat{\beta}_1 + \hat{\beta}_7$	$\hat{\beta}_7 < 0$
SO2 satellite data	$\hat{\beta}_1$	$\hat{\beta}_1 + \hat{\beta}_7$	$\hat{\beta}_7 > 0$
Normalized difference	$\hat{\beta}_1$	$\hat{\beta}_1 + \hat{\beta}_7$	$\hat{\beta}_7 > 0$

Note. This table provides an overview on how to interpret the estimated effect of the SO_2 reduction target for each province from separate estimations of Equation (2.4) for both the official SO_2 emissions indicator and the SO_2 satellite data. The sample runs through the first policy period until 2008 when the government could not monitor SO_2 emissions. The column labelled 'hypothesis' captures the ex ante expected coefficient for the triple interaction term under the hypothesis of misreporting. 'Normalized difference' captures the difference between the SO_2 satellite data and the SO_2 emissions data, both normalized to 2005 levels.

not monitor compliance. The interpretation for the causal effect of a higher reduction target is the estimate for β_1 for the subgroup unlikely to misreport, and the estimate for $\beta_1 + \beta_7$ for the subgroup likely to misreport.

I estimate this specification for both outcome variables and for the normalized difference between both outcome variables. This yields 6 different point estimates for the effect of a higher reduction target (3 for the different outcomes, 2 for each subgroup). Table 2.6 illustrates this. My hypothesis is that for the subsample of provinces more likely to engage in misreporting, the drop in *official* SO_2 *emissions* is greater, which implies that $\hat{\beta}_1 > \hat{\beta}_1 + \hat{\beta}_7$, which is equivalent to $\hat{\beta}_7 < 0$. For real compliance measured by SO_2 *satellite data*, my hypothesis is that the provinces more likely to misreport will reduce pollution less, implying $\hat{\beta}_1 < \hat{\beta}_1 + \hat{\beta}_7$, which is equivalent to $\hat{\beta}_7 > 0$.

I repeat this procedure for all different ways of splitting the sample. The results are shown in Table 2.7 and provide only mixed evidence that provinces with a higher incentive to misreport do so. Judging from the signs of the estimated coefficients, it is possible that provinces with leaders who are subject to career concerns report higher SO_2 emission reductions, but lower real SO_2 improvements. The evidence for misreporting based on the cost and benefits from abatement is more mixed: SO_2 emission

Table 2.7 The effect of monitoring: Evidence

Misreporting likely due to	y_{pt}	$\hat{\beta}_7$	Significance $\hat{\beta}_7$
Career concerns	SO2 emissions	-0.17	p=0.99
	SO2 satellite	0.008	p=0.38
	Normalized difference	0.025	p=0.35
Marginal abatement cost	SO2 emissions	-3.08	p=0.52
	SO2 satellite	-0.009	p=0.29
	Normalized difference	0.008	p=0.34
Marginal welfare impacts	SO2 emissions	7.64	p=0.13
	SO2 satellite	0.002	p=0.80
	Normalized difference	0.002	p=0.86

Note. The table shows the estimated effect of the SO_2 reduction target for each province from separate estimations of Equation (2.4). The outcome variables are the official SO_2 emissions indicator, the SO_2 satellite data as well as the 'Normalized difference', which captures the difference between the SO_2 satellite data and the SO_2 emissions data, both normalized to 2005 levels. The sample is for the first policy period until 2008 when the government could not monitor SO_2 emissions. The significance of $\hat{\beta}_7$ is the p-value using wild bootstrap p-values from 1000 repetitions clustered at the province level.

changes are higher for provinces with higher abatement cost, but lower for provinces with a higher marginal welfare impact. The normalized difference between both data sources is positive for all three hypotheses of misreporting. It is important to note, however, that the interaction coefficient β_5 is never significant due to the low number of cross-sectional units, making these results suggestive rather than definite. Subsection 2.5.5 below therefore compares my empirical findings to recent research on water pollution control in China to better understand the effect of monitoring on pollution outcomes.

2.5.4 Reactions by the regulated agents in response to the targets

This subsection provides evidence into the responses by the regulated agents. Two facts stand out from the empirical analysis above. On the one hand, regulated provincial governments likely misreport the desired emissions data when the central government cannot monitor them. On the other hand, my analysis shows that the SO_2 reduction control policy was ultimately successful in reducing air pollution. Provincial governments thus reacted along two margins. They appeared to bring down air pollution through *rhetorical compliance* initially, and switched to *real compliance* towards the end of the policy.

Rhetorical compliance As shown above in Table 2.3, the effect of the reduction targets has a different sign when comparing the official SO_2 emissions data and the NASA SO_2 satellite data before 2008. Given that the SO_2 satellite data move closely with real SO_2 emissions (see Section 2.4), this suggests initial misreporting of data. This subsection investigates whether this misreporting was accompanied by further political rhetoric by the provincial governments. In particular, I test whether the provincial governments changed their political rhetoric in response to the SO_2 reduction control targets. To do so, I estimate versions of Equation (2.2), using different counts of keywords as outcome variables. Firstly, I use the overall number of keywords related to air pollution in each government work report. Figure 2.8 plots the estimates for the yearly interaction coefficients β_{1t}. It shows that provinces that received a higher SO_2 reduction target show a distinct increase in their political rhetoric on air pollution, both when analyzing the full government work report (left panel) and when looking at the part of the report that contains statements about work done in the preceding period (right panel).

Next, I zoom in and split political statements into those on work done in the preceding year and those related to projects for the year to come. The right panel in Figure 2.8 shows that the increase is most pronounced in the period before the central government had the capacity to monitor. Provincial governments that received a higher SO_2 reduction target claim past work on projects related to air pollution in 2007, thus exploiting the central government's inability to monitor SO_2 initially. That is, a higher SO_2 reduction target induces provinces to claim work on air pollution in 2006. Political statements relative to future work on air pollution also increase with the SO_2 reduction targets, although the evidence is less stark.

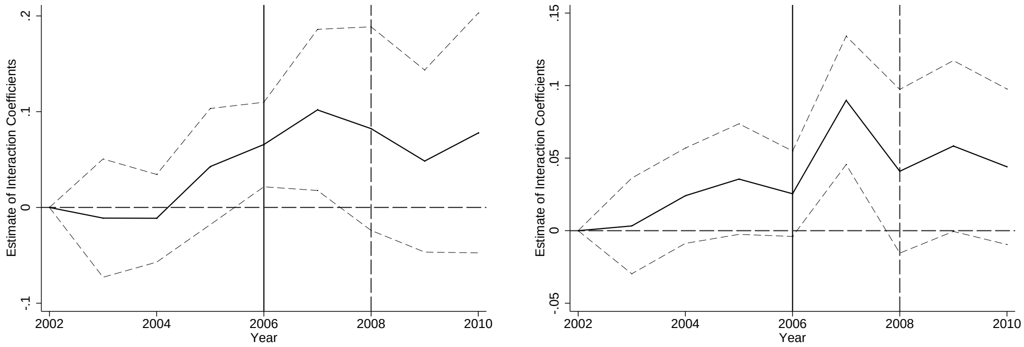


Figure 2.8 *Dynamic treatment effects: rhetorical compliance*. The solid line plots the point estimate for yearly coefficient estimates of the interaction coefficients $\sum_{t=1}^{T} \beta_{1t}$ in Equation (2.2) for the overall number of keywords related to air pollution in a province-year government work report for (i) the full report (*left panel*) and (ii) only for the part of the report that reports on work done in the period preceding the report (*right panel*). The dashed lines show the 95% confidence bands, based on heteroskedasticity-robust standard errors clustered at the province-level. The excluded time period $t = 1$ is the year 2002. The vertical line at 2006 marks the start of the policy and the vertical line at 2008 marks the start of SO_2 monitoring.

Finally, Table 2.8 summarizes these results based on estimating Equation (2.1) on the political attention variables to find that: (i) political attention to air pollution increases with target stringency, and (ii) statements about past work on air pollution peak in the period in which the government could not monitor SO_2[13]. As shown through the comparison of results from the official SO_2 emissions indicator and NASA satellite data, however,

[13]Table 2.8 includes a robustness check to further show that my estimates on rhetorical compliance are meaningful. This test involves estimating the effect of the SO_2 reduction targets on closely related, yet different placebo outcomes that should not be affected by the SO_2 reduction targets. The 11th Five-Year Plan (2006-2010) included goals to increase China's forest cover from 18.2% to 20% and to extend the coverage of rural medical care from 23.5% to 80%. To measure the political attention towards these policies, I use the count of 'forest' and of 'medical care' in the government work reports as dependent variables for the falsification test. Results based on these outcomes using the specification in Equation (2.1) are reported in the columns 'Placebo Outcomes' in Table 2.8. As in the case of the keywords related to air pollution, both keywords are mentioned more often during the 11th Five-Year Plan (2006-2010) than before. Provinces that received a higher SO_2 reduction target, however, do not talk more about either topic. This strongly suggests that governments of provinces with higher reduction targets do not mimic the Central government's political agenda in their own statements in general. Instead, they specifically change their political communications in response to the SO_2 reduction targets.

Table 2.8 Reactions by the regulated agents in response to the targets

	Political attention to air pollution			Placebo outcomes	
	All statements	Past period	Future period	'Forest'	'Medical care'
Reductiontarget	0.07*	0.03**	0.04**	-0.12	0.02
$\times$ D(Post)	(0.02)	(0.01)	(0.02)	(0.07)	(0.07)
	[0.08]	[0.03]	[0.03]	[0.31]	[0.79]
Reductiontarget	-0.01	-0.00	-0.00	-0.07	0.35***
	(0.03)	(0.02)	(0.02)	(0.10)	(0.07)
	[0.74]	[0.66]	[0.78]	[0.40]	[0.00]
D(Post)	1.19***	0.48**	0.71**	3.59**	5.68***
	(0.24)	(0.13)	(0.16)	(0.96)	(0.79)
	[0.00]	[0.01]	[0.01]	[0.04]	[0.00]
Province FE	✓	✓	✓	✓	✓
Effect size	30.9%	29.6%	31.5%	-7.2%	1.2%
Observations	274	273	273	274	274
Provinces	31	31	31	31	31
R^2	0.51	0.42	0.43	0.69	0.52

Note. Heteroskedasticity-robust standard errors clustered at the province level are shown in parentheses. Wild bootstrap p-values from 1000 repetitions clustered at the province level are shown in square brackets. The effect size gives the estimated coefficient of the interaction term β_1 in Equation (2.1) as percentage of the mean of the dependent variable for a one standard deviation(σ)-increase in the SO_2 reduction target. $***p < 0.01$, $**p < 0.05$, $*p < 0.1$.

SO_2 pollution during that period did not improve.

Taken together, my empirical findings suggest that the provincial governments exploit the central government's inability to monitor along two margins: they likely misreport the desired SO_2 emissions data, and they back up the misreporting through rhetorical compliance in their public political statements.

Real compliance As shown above, the SO_2 reduction targets worked in reducing SO_2 pollution significantly from 2008 onwards. Based on my own calculations of the marginal cost of SO_2 emissions abatement (see Section

2.6), the installation of desulfurization devices in existing industrial and power plants, fuel-switching to better quality coal, and the shutdown of small, inefficient thermal units are the main margins by which the provincial governments could reduce SO_2 emissions over relatively short time horizon of 5 years. I provide evidence on the last two channels.

The data on the *installation of scrubbers* include both the capacity of the desulfurization equipment as well as the timing of its installation in each province. Figure 2.9 shows that while provinces with a higher SO_2 reduction targets installed more desulfurization devices on average, that effect is not statistically different from zero. Figure 2.10 computes the skewness of the timing of the installation of the scrubbers between the years 2006 and 2010 and correlates it to the SO_2 reduction targets at the province level. As can be seen, provinces with a higher SO_2 reduction target did not install scrubbers earlier than other provinces. The data on the *shutdown of small, inefficient thermal units* shows a much clearer picture. Figure 2.11 shows that the higher the reduction target, the higher the capacity of small thermal units shut down by 2010. Shanghai and Beijing are the exception to this rule, most likely because they already shut down inefficient plants in the past.

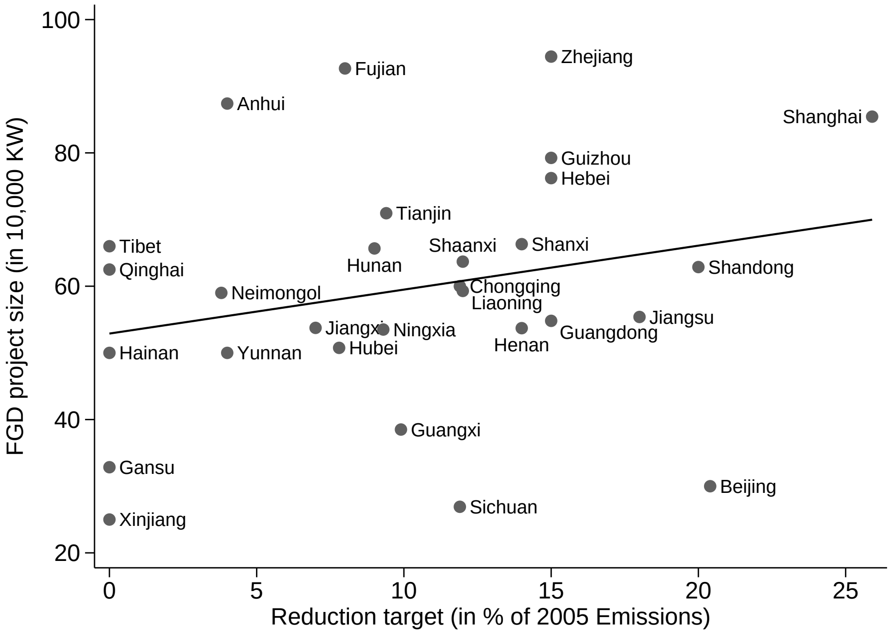


Figure 2.9 *Real compliance: Installation of desulfurization devices*. This graph plots the relationship between the SO_2 reduction target and the planned installation of desulfurization devices at the province level. The solid line fits a linear regression with slope parameter b=0.66 and p=022 computed from standard errors clustered at the province level.

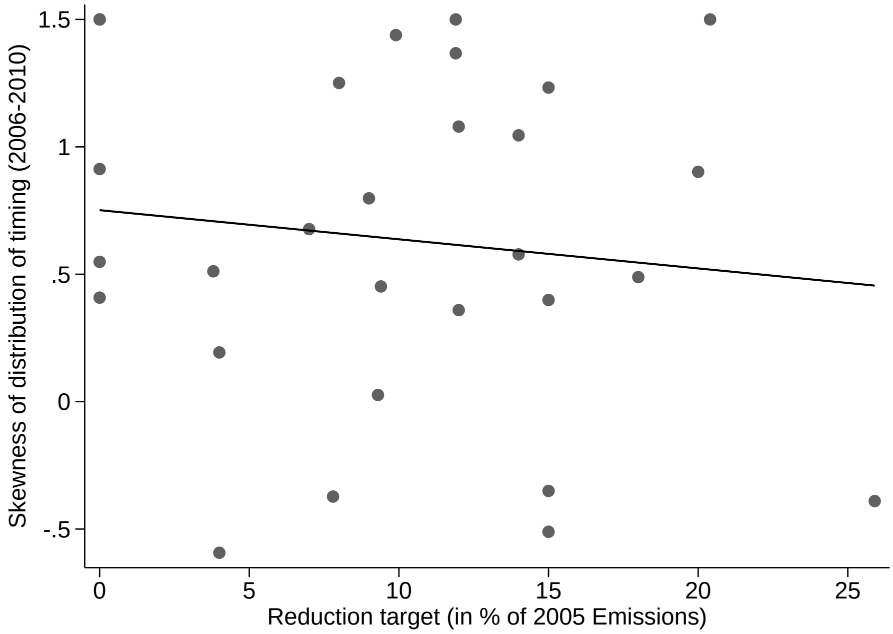


Figure 2.10 *Real compliance: Timing of desulfurization device installation*. This graph tests whether provinces with a higher reduction target installed the planned desulfurization devices earlier. The vertical axis shows the skewness for each province of the 5 yearly observations from 2006-2010, where the weight is the capacity (in 10,000 KW) of planned desulfurization devices in each year. The solid line fits a linear regression with slope parameter b=-0.01 and p=0.56 computed from standard errors clustered at the province level.

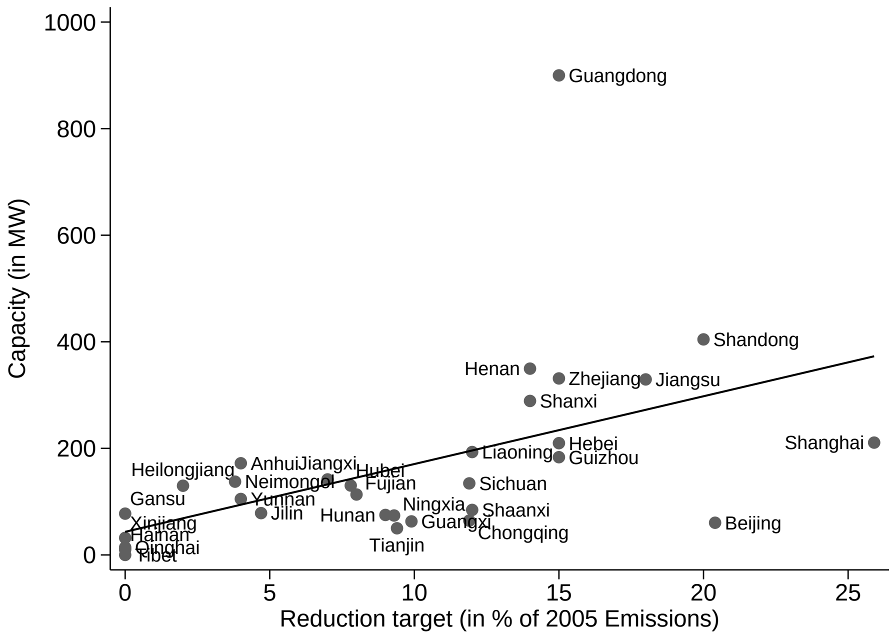


Figure 2.11 *Real compliance: Shutdown of small thermal units*. This graph plots the relationship between the SO_2 reduction target and the decommissioning of small thermal units at the province level. The solid line fits a linear regression with slope parameter b=12.72*** and p=0.004 computed from standard errors clustered at the province level.

2.5.5 Interpretation of the empirical findings

This section puts my empirical findings into context. My findings suggest the importance of the government's monitoring to achieve compliance with the pollution control regulation. As explained above, the 2008 institutional shift in environmental governance can be summarized as bringing about radical improvements in SO_2 monitoring through a number of changes: most importantly, the gathering of air pollution data was taken from the provincial governments and directly done under the auspices of the newly created Ministry of Environmental Protection (State Council, 2007). Moreover, there was a 17% increase in the number of monitoring officials (Song et al., 2015), and continuous, automatic on-site tracking of SO_2 emissions for key polluters was started in each prefecture (Ministry of Environmental Protection of the People's Republic of China, 2008a,b). Those changes hint at the key role for monitoring in achieving compliance.

A comparison to findings from contemporaneous research strengthens this conclusion: Kahn, Li, and Zhao (2015) study water pollution control in China during the 11th Five-Year Plan (2006-2010). That is, they estimate the effect of a pollution control policy for the same country, time period and institutional setting. Their setting differs in one crucial aspect, however: the central government had access to high-quality monitoring data for water pollution from the start. A comparison to their findings can thus shed light onto the effect of monitoring on compliance.

In Table 4 (Panel A), Kahn, Li, and Zhao (2015) estimate yearly treatment effects for the effect of the water pollution control policy. Their findings show a pattern that is very different from my findings along two characteristics: (i) real compliance is immediate and statistically different from zero for all years (excepting 2009) and (ii) the yearly effects are of a similar magnitude throughout. Given that the salient difference to my setting is the government's ability to monitor pollution already in 2006 and 2007, this comparison provides further evidence that information is the likely channel that explains the behaviour by the provincial governments until 2008 and the lack of real compliance[14].

[14]Relatedly, Stavins (1998) mentions that 'the SO_2 program has also brought home the importance of monitoring and enforcement provisions' in the US experience of SO_2 pollution control, lending further support to my empirical findings.

2.6 Instrument choice and gains from trade

In this section I address two questions on the evaluation of policy design: what is the welfare-optimizing allocation of targets across provinces? And: what are the welfare gains from trade associated with moving from the actual allocation to the welfare-optimizing allocation?

As discussed in the Introduction, both types of question have been studied for the US (Carlson et al., 2000; Oates, Portney, and McGartland, 1989). That research has shown that the efficiency of different policy instruments is an empirical question. It will depend, among other things, on whether the command-and-control regulation is designed in an enlightened way that takes the cost of pollution abatement into account. Whether China's flagship air pollution control regulation has been cost-efficient is therefore a question of great interest. It is also of immediate policy-relevance: The Economist hypothesized that China relies too much on command-and-control regulation (The Economist, 2013).

In this section, I construct detailed marginal abatement cost (MAC) curves for SO_2 at the province level for China. I use these SO_2 MAC curves to predict the counterfactual cost-efficient allocation. In this way, I can assess whether the actual command-and-control regulation was enlightened and took into account the cost of abatement. Furthermore, I can quantify the efficiency gains from trade from moving from the actual allocation of reduction targets compared to the cost-efficient allocation. As noted by Stavins (2003), cost might be the best measure to assess policy efficiency. Finally, I compute a back-of-the-envelope measure for the marginal benefits of SO_2 abatement at the province level to study whether the gains from trade based on cost alone are driven by omitting the benefits of reducing air pollution.

2.6.1 Construction of the marginal abatement cost curves

Data I use a rich set of micro data on SO_2 emissions and abatement costs in each province in China at a very detailed level. These data are compiled by the IIASA research institute for use as input into their GAINS model for China (on the model, see IIASA, 2010a, IIASA, 2010b, Klimont et al., 2009). These data rely on a variety of sources of two kinds: *common data*, that are used across different countries and rely on the assumption of free international markets in abatement equipment, and *country-specific data*. Common data include the unit investment cost for technologies, fixed costs of operation, and the amount of input factors needed for some of the variable cost components. These data have been compiled and updated by IIASA for several decades on the basis of expert meetings at the UN (AP EnvEcon, 2010).

Country-specific data, on the other hand, include a detailed breakdown of China's industrial structure: the type and size of polluting installations, the facility operating conditions, national fuel consumption data, prices of inputs (labour, electricity, fuel cost) as well as unabated emission factors and removal efficiencies (AP EnvEcon, 2010). The local data are compiled by IIASA experts in collaboration with local experts from the Chinese Energy Research Institute in Beijing and Tsinghua (Purohit et al., 2010). These data are combined into unit cost estimates per technology, as well as abatement potential[15].

The data give a detailed breakdown of all SO_2 emission sources for each province, split into different sectors (such as the combustion of coal) that use different fuels (such as gas or low-sulfur coal) and each of which has different abatement technologies at its disposal (such as limestone injection). Each of the abatement technologies is characterised by a unit cost of abatement (in €/tSO_2) and its abatement potential (i.e. how much would the emissions factor of the current sector-fuel combination be lowered when switching to the abatement technology). To illustrate, one abatement technology would be the use of limestone-injection (*abatement technology*) in a modern coal-fired power plant (*fuel* and *sector*) at the cost of 515.38€/tSO_2 (*unitcost of abatement*).

Construction of the MAC curves To construct the MAC curve for one province, I follow a two-step procedure. First, I rank each abatement option by the unit cost of abatement within each sector and fuel. Second, I abate SO_2 emissions within each sector and fuel in increasing cost order across the Province. Figure 2.12 shows two examples, Beijing and Sichuan, to illustrate the large heterogeneity in marginal abatement cost across provinces (Appendix 2.B contains the MAC curves for each province). In contrast to many MAC studies, I do not rely on a top-down model but base the cost es-

[15]The data on the cost of abatement are thus not purely engineering estimates, as they rely on the real cost of abatement in different countries. Rather than being a lower bound on the cost of abatement, they are more likely to be an upper bound: cheaper local inputs and China's capacity for scale will likely allow for cheaper abatement. This explains why the cost estimates for China can slightly differ from estimates for the US. The use of MAC estimates from revealed-preference settings such as Meng (Forthcoming) and Gosnell, List, and Metcalfe (2016) would therefore be desirable, but no such estimates are available for China or at the required level of comprehensiveness.

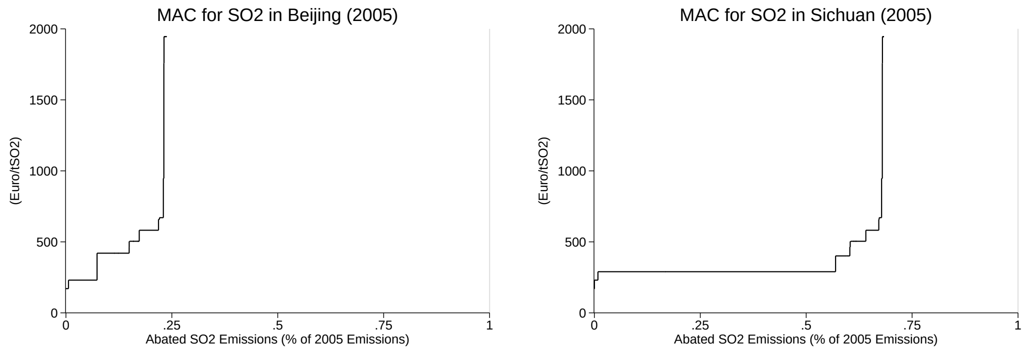


Figure 2.12 *Marginal abatement cost curve examples*. This graph shows the marginal SO_2 abatement cost curves for Beijing and Sichuan in 2005. The horizontal axis lists the abatement intensity relative to the 2005 SO_2 emissions level.

timates entirely on data, and I am the first to provide complete MAC curves on Chinese provinces in this way.

Consistency checks To check that my data represent the official Chinese data well, Figure 2.13 shows the average ratio between the GAINS SO_2 emissions and the official MEP SO_2 emissions as a function of the province's level of emissions: the overall fit between my data and the MEP data is good (correlation: 85.9%) and fairly stable across provinces, except for two outliers with very low emissions: Tibet and Hainan. Overall, the GAINS data report higher emissions for all provinces than do the official data. This is in line with the literature in atmospheric science that has found that GAINS data tend to be more comprehensive and slightly overpredict official SO_2 data sources, which is possibly due to differing assumptions on the distribution of fuel consumption across sectors (Klimont et al., 2009) and the fact that official MEP statistics lack rural pollution sources and biofuels (Lu et al., 2010).

Additionally, there is a drawback in using detailed SO_2 emissions data at the microlevel: the detailed breakdown across sectors does not allow me to compute the cost of moving activity across sectors (such as moving electricity generation from coal-fired power plants to wind turbines). In other words, these data limitations make it necessary to assume scrappage cost and imperfect substitution across sectors in the short run. Given the 5-year horizon of the policy, however, I feel that this is a reasonable assumption to make.

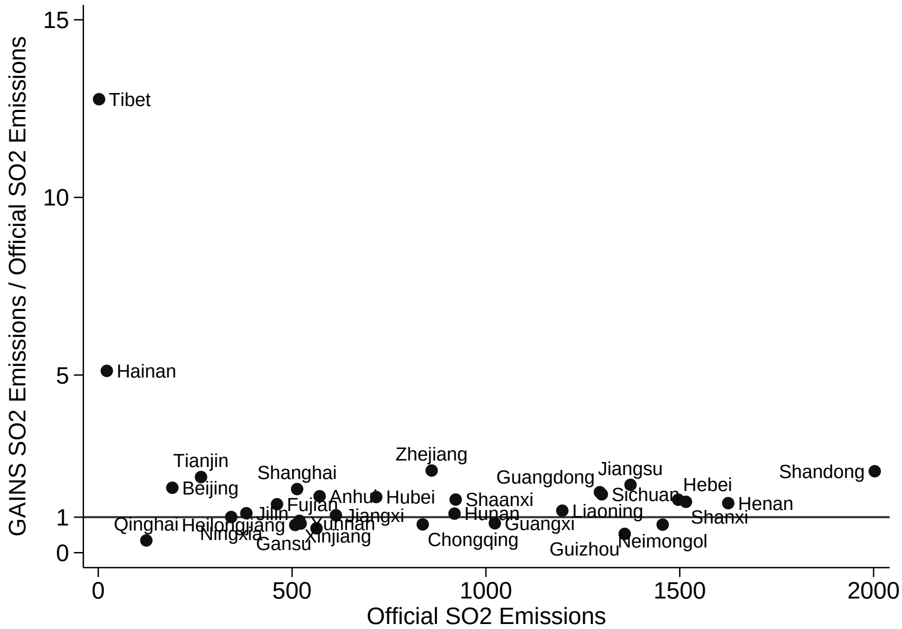


Figure 2.13 *Consistency of SO_2 marginal abatement cost data and official data*. This graph shows the ratio of the SO_2 emissions data underlying the construction of the marginal abatement cost curve (from IIASA's GAINS model) and the official SO_2 emissions data from the Ministry of Environmental Protection for the year 2005 on the vertical axis. The datapoints are ordered according to the official SO_2 emissions level on the horizontal axis. The raw correlation between both data sources is 85.9%.

2.6.2 The cost-efficient counterfactual allocation of SO_2 reduction targets

In this subsection, I compute the counterfactual allocation of SO_2 reduction targets across provinces for the cost-efficient policy. I pool the province-level microdata and construct the SO_2 marginal abatement cost curve at the national level for China. As shown in Figure 2.14, marginal abatement cost is relatively flat until an abatement level of 35% of 2005 SO_2 emissions. Up until this abatement level, the cost of SO_2 abatement at the margin is less than 500€/tSO_2. Beyond ca. 45% of 2005 SO_2 emissions, however, the picture changes and the cost of abatement rises rapidly. Beyond 50%, abatement becomes prohibitively costly.

Figure 2.14 also shows the counterfactual marginal abatement cost for

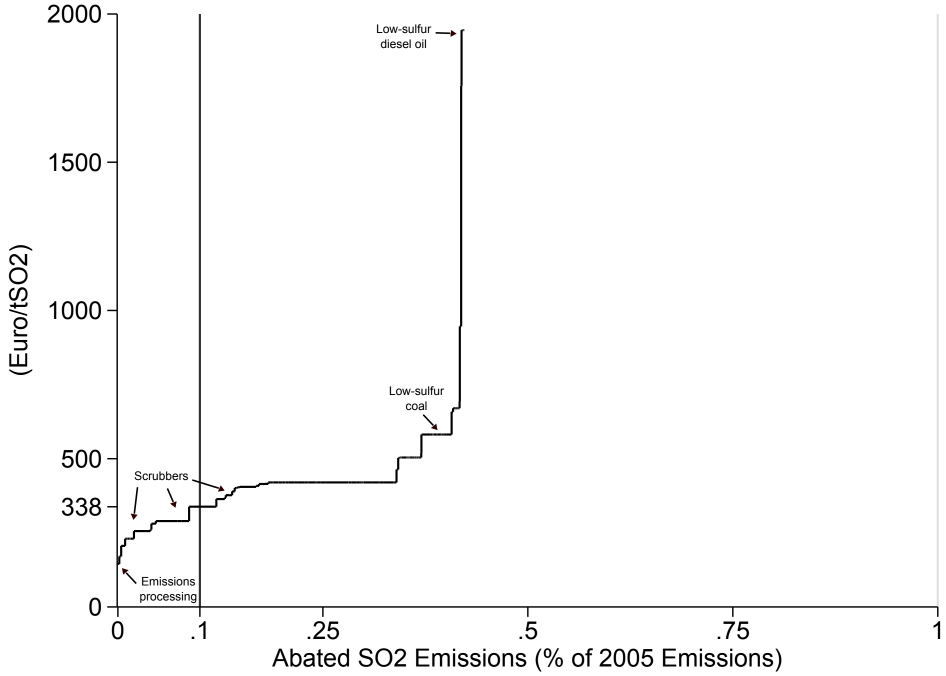


Figure 2.14 *SO_2 marginal abatement cost curve for China*. This graph shows the marginal SO_2 abatement cost curve for China in 2005. The horizontal axis plots the abatement intensity relative to the 2005 SO_2 emissions level. The vertical line at an abatement level of 10% illustrates the 10% national SO_2 emissions control target for China as a whole. Its intersection with the marginal abatement cost curves shows that the counterfactual marginal abatement cost for this abatement level is 338€/tSO_2.

the national SO_2 reduction target of 10%, given by the intersection between the vertical line at 10% and the MAC curve. At an abatement level of 10%, the cost-efficient allocation of SO_2 reduction targets would have led to a marginal abatement cost of 338€/tSO_2. This figure is the counterfactual marginal cost of the Chinese government's SO_2 emissions control strategy had the central government distributed the provincial reduction targets in a cost-optimal way.

The intersection of the marginal abatement cost curve for China as a whole with the 10% national abatement level also produces the cost-efficient allocation of SO_2 reduction targets across provinces. Figure 2.15 shows that this allocation is far more skewed than the distribution of targets that was actually used in the 11th FYP. The actual allocation already ranges from targets of 0% to 25.9%, but is approximately uniformly dis-

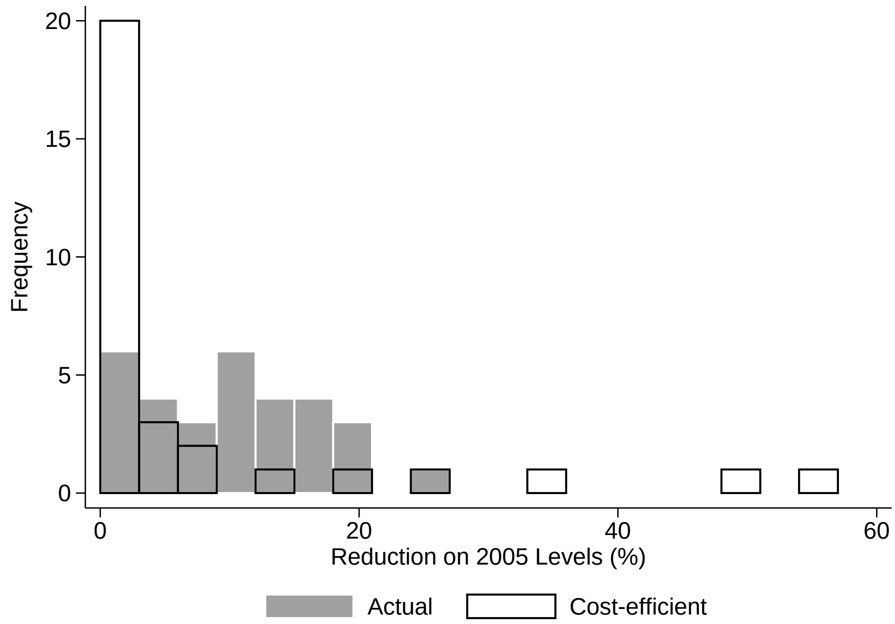


Figure 2.15 *The cost-efficient allocation of SO_2 reduction targets*. This graph shows the counterfactual cost-efficient allocation of SO_2 reduction targets under the 10% SO_2 total control target of the 11th Five-Year Plan (2006-2010), shown in white boxes. The three provinces with the highest targets under the cost-efficient allocation are Sichuan, Shandong and Zhejiang. Solid grey boxes show the actual allocation of reduction targets in comparison.

tributed within this range. The cost-efficient allocation, by contrast, is more unequal. A small number of provinces would bear most of the reductions. These provinces are Sichuan, Shandong, and Zhejiang. The industrial structure of these provinces allows for comparatively cheap installation of wet flue-gas desulfurisation in industry and power plants and the use of more efficient combustion processes in refineries and steel sintering.

Based on the MAC curves, I find that the Chinese government did not equate marginal abatement cost across space. Instead, the reduction targets favoured coastal provinces in the East even though abatement coasts are higher at the margin. These results are shown in Table 2.9 and in Figure 2.A.6 in Appendix 2.A, which depicts the same data on a map that colours the difference between the actual and the cost-efficient allocation. The actual allocation is consistent with a tale-of-two-cities story, in which China would develop amenity-based consumer cities along the coast,

Table 2.9 Actual and counterfactual SO_2 target allocations

	SO2 reduction target under different allocations (in % of 2005 emissions)			Difference between target allocations (in percentage points)		
	Cost-efficient	Optimal	Actual	Actual - cost-efficient	Actual - optimal	Cost-efficient - optimal
Anhui	1.5	8.3	4	2.5	-4.3	-6.8
Beijing	7.3	22.9	20.4	13.1	-2.5	-15.6
Chongqing	50.5	0	11.9	-38.6	11.9	50.5
Fujian	0.7	0.7	8	7.3	7.3	0
Gansu	2.5	0.8	0	-2.5	-0.8	1.7
Guangdong	0.7	48.3	15	14.3	-33.3	-47.6
Guangxi	20	0	9.9	-10.1	9.9	20
Guizhou	5.8	0	15	9.2	15	5.8
Hainan	0	0	0	0	0	0
Hebei	1.7	0.1	15	13.3	14.9	1.6
Heilongjiang	1.7	0	2	0.3	2	1.7
Henan	0.6	0.1	14	13.4	13.9	0.5
Hubei	2.5	35.2	7.8	5.3	-27.4	-32.7
Hunan	1	1	9	8	8	0
Jiangsu	0.9	25.3	18	17.1	-7.3	-24.4
Jiangxi	1.6	1.6	7	5.4	5.4	0
Jilin	1.6	1.6	4.7	3.1	3.1	0
Liaoning	4.2	0	12	7.8	12	4.2
Neimongol	1.1	0	3.8	2.7	3.8	1.1
Ningxia	0.1	0	9.3	9.2	9.3	0.1
Qinghai	3.5	0	0	-3.5	0	3.5
Shaanxi	13.9	0	12	-1.9	12	13.9
Shandong	25.5	0.6	20	-5.5	19.4	24.9
Shanghai	6.2	6.2	25.9	19.7	19.7	0
Shanxi	0.6	0	14	13.4	14	0.6
Sichuan	56.2	0	11.9	-44.3	11.9	56.2
Tianjin	2.2	52.1	9.4	7.2	-42.7	-49.9
Tibet	0	0	0	0	0	0
Xinjiang	1.6	0	0	-1.6	0	1.6
Yunnan	1.3	1.3	4	2.7	2.7	0
Zhejiang	35.4	35.4	15	-20.4	-20.4	0

Note. This table shows the SO_2 emissions reduction targets under each policy regime. 'Cost-efficient' are the counterfactual targets for achieving the cost-efficient allocation given the national 10% SO_2 reduction target, 'Optimal' are the counterfactual targets for achieving the welfare-optimizing allocation given the national 10% SO_2 reduction target and 'Actual' are the provincial reduction targets used in the 11th Five-Year Plan (2006-2010).

while maintaining a base of polluting manufacturing in its interior (Kahn, 2006; Zheng and Kahn, 2013). Shanghai is a prime example for this: under the cost-efficient allocation, Shanghai would have received an SO_2 reduction target of 6.2% on 2005 levels. The actual allocation, however, gave Shanghai a reduction target of 25.9%, or more than four times the cost-efficient target.

My findings show that actual command-and-control regulation that China used to control SO_2 pollution in the 11th FYP was not cost-efficient. As suggested by Oates, Portney, and McGartland (1989), command-and-control regulation will only be efficient if it is designed in an enlightened fashion by keeping an eye on the cost of abatement. Figure 2.16 shows that this was not done by the Chinese government in 2005. There is no statistically significant relationship between the SO_2 reduction target a province received under the 11th FYP and its abatement cost at the margin.

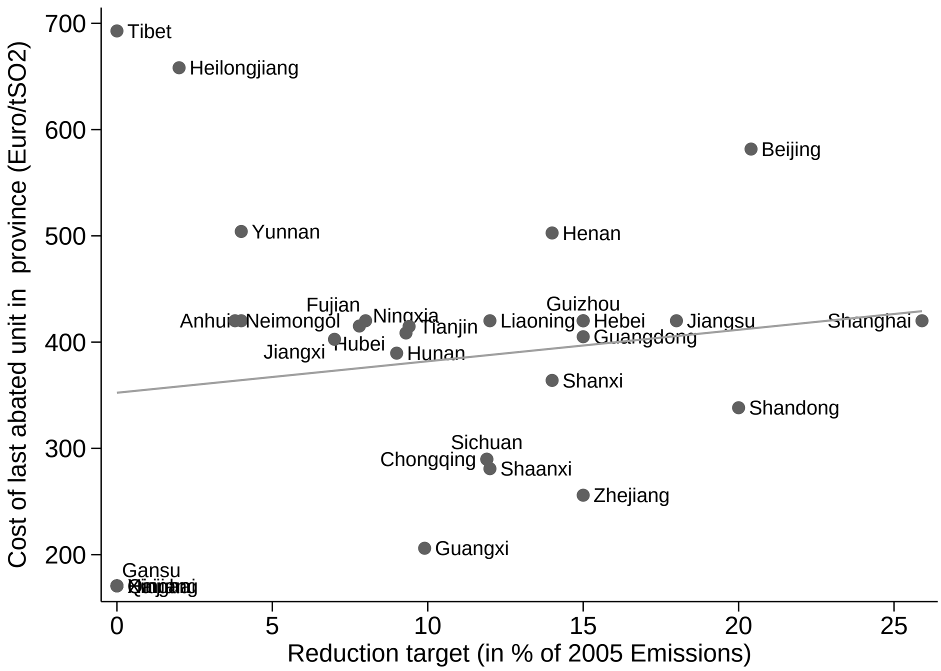


Figure 2.16 *Relation between MAC and SO_2 reduction targets*. This graph shows the lack of correlation between a province's actual SO_2 reduction target in the 11th Five-Year Plan (2006-2010) and its marginal abatement cost at the level of the target. The solid line fits a linear regression with slope parameter b=2.96 and p=0.54 computed from standard errors clustered at the province level.

2.6.3 A measure for marginal abatement benefits

My final aim is to quantify the gains from trade from moving from a command-and-control regulation to a market-based allocation based on abatement cost. Because China's provinces differ markedly with respect to income levels, population densities and initial pollution, it is important to not rely exclusively on a measure of cost. I use the method employed by Oliva (2015) to construct a back-of-the-envelope measure for the marginal abatement benefits of reducing SO_2 pollution at the province level. This method proceeds in 3 steps: (i) how does the SO_2 control policy change pollutant concentrations?, (ii) what health effects do the changes in pollutant concentrations cause? and (iii) what is the monetary value of those health effects?

(i) Changes in pollutant concentrations I use the results from Wang, Jang, et al. (2010b), who use the CMAP modelling system (maintained by the US Environmental Protection Agency) to simulate the ex ante effects of the SO_2 reduction policy on concentrations of SO_2 and $PM_{2.5}$. Their measurements allow me to attribute changes in pollutant concentrations to the SO_2 emissions reduction target of each province.

(ii) Health effects To convert the changes pollutant concentrations to changes in health outcomes, I use dose-response estimates for $PM_{2.5}$ and SO_2 from Bombardini and Li (2016). They estimate an elasticity of infant mortality rates of 0.9 to SO_2 and of 2.2 for $PM_{2.5}$. I combine those estimates with data on SO_2 and $PM_{2.5}$ levels in 2005 from MEP and the *China Energy Databook* (Fridley, Romankiewicz, and Fino-Chen, 2013) to approximate a linear dose-response function for each pollutant.

I use the estimates from Bombardini and Li (2016) for two reasons: firstly, they use an instrumental strategy approach to estimate a dose-response function for the health effects of air pollution, thus correcting downward bias from OLS estimates (due to migration, income effects and avoidance behaviour). Additionally, their study is from China, from a recent period, and includes consistent estimates for both SO_2 and $PM_{2.5}$, which are the main pollutants that are affected by the SO_2 reduction policy. The downside is that this restricts my focus on infant mortality when calculating the benefits from reducing air pollution[16].

[16]Infant mortality, however, is likely to capture a first-order welfare effect. Matus et al. (2012) calculate that 71.4% of all air pollution costs in China are health costs, and that mortality captures over 85% of those health costs. Chen, Ebenstein, et al. (2013), in turn, show that mortality impacts from TSP are strongest in infants. Evidence from Indonesian

(iii) Valuation To convert the health damages into monetary values, I use a baseline value of a statistical life (VSL) of 1 million yuan from World Bank and State Environmental Protection Administration (2007). This value is a midpoint between the VSL estimates from the reviewed studies ranging from 0.24 to 1.7 million yuan[17]. Following Hammitt and Robinson (2011), I account for the income heterogeneity across Chinese provinces by adjusting the central VSL estimate according to the income level in each province using an income elasticity of VSL of one. This yields VSL estimates from 360,000 yuan (Guizhou) to 3,000,000 yuan (Shanghai).

I multiply these VSL estimates by the mortality numbers to compute the benefit of reducing SO_2 emissions by 1 kt for each province. Since I employ a linear approximation, the marginal benefits of abatement is constant[18].

Caveats I rely on an overly conservative measure of benefits by including only infant mortality which may underestimate the true benefits and lead to an underestimation of the welfare-optimizing level of SO_2 abatement for China. However, this concern is muted in practice since even my lower bound benefit measure suggests SO_2 abatement up until the prohibitive marginal abatement cost ranges. Additional abatement benefits at the margin would thus only have a negligible effect on the welfare-optimizing abatement level. The advantage is that the dose-response functions used are more precise for infants because low migration translates into better knowledge of lifetime exposure to pollution, improving the consistency of estimates *across* provinces.

The welfare-optimal SO_2 target allocation In a similar vein to the MAC data, I use the combined data on both marginal cost and benefit to construct marginal welfare impact curves. I combine the marginal abatement benefit data with the marginal abatement cost data by dividing the marginal abatement cost by the marginal abatement benefit to obtain a measure of the marginal welfare impact of abatement. When this ratio is

wildfires in 1997 also points to large infant mortality effects from exposure to particulates, mostly driven by prenatal exposure (Jayachandran, 2009). Greenstone and Hanna (2014) also focus on infant mortality to evaluate the effect of pollution.

[17] VSL estimates for China are available for more recent periods (Ito and Zhang, 2016), but to evaluate a policy from 2005 I prefer to use estimates from that period.

[18] For those provinces with a reduction target of zero, I find that the benefits are 0 or nearly 0, too. This is due to low initial pollution, low VSL estimates, and low population numbers. The exception is Gansu, for which I compute the marginal benefits as the average of its 5 nearest neighbours with respect to initial pollution, VSL estimates, and population numbers.

below one, benefits are larger than cost. Once this ratio exceeds one, costs are higher than benefits. Figure 2.17 pools the province-level data to obtain the marginal welfare impacts for China as a whole. The vertical dashed line at 0.1 marks the 10% national SO_2 emissions control target of the 11th FYP. The marginal benefits of abatement exceed the marginal cost of abatement by more than 14 times at this abatement level[19]. Again, there is no correlation between a province's SO_2 reduction target and the marginal welfare impact of reducing SO_2 emissions (shown in Figure 2.A.5 in Appendix 2.A).

2.6.4 Gains from trade

Finally, I study the gains from trade of the two different policy instruments. Gains from trade will be possible if the actual command-and-control regulation used to allocate the SO_2 reduction targets in the 11th FYP was not done efficiently. As I have shown above, this is the case. This subsection quantifies these efficiency gains in terms of cost-efficiency. To further take into account the heterogeneities across Chinese provinces, I construct a back-of-the-envelope measure for the marginal benefits of air pollution abatement to study the robustness of the findings based on pure cost-efficiency.

Table 2.9 shows that the cost-efficient and the optimal allocation differ strongly for a number of provinces, most notably for Sichuan, Chongqing, Tianjin and Guangdong (right panel). This shows that moving from the actual to the cost-efficient allocation does not necessarily present an efficiency gain ex ante along both efficiency measure. Instead, the correlation between the cost-efficient and the welfare-optimising allocation is slightly negative, while the correlation between the actual and the cost-efficient and optimal allocations is similar (29% vs 28%)[20].

Table 2.10 shows the gains from trade that are possible. Firstly, I follow the literature Stavins (2003) and assess the gains from trade using the cost efficiency measure. I find that moving from the actual allocation to the cost-efficient allocation would decrease abatement cost by 49% at the

[19]It is also possible to predict the welfare-optimal allocation of SO_2 reduction targets across provinces. These results are shown in Table 2.9 for comparison with the counterfactual cost-efficient allocation. The distribution of targets is skewed similar to the cost-efficient allocation. The provinces that would receive a higher reduction targets are different, however. Tianjin, Guangdong, Zhejiang and Hubei would receive the highest reduction targets. The exception is Shanghai, which would maintain its 6.2% reduction target from the cost-efficient allocation.

[20]It is further possible to calculate other allocations such as achieving the 10% reduction through the mandated use of scrubbers or with a 10% target for each provinces. Detailed results from these calculations are available upon request.

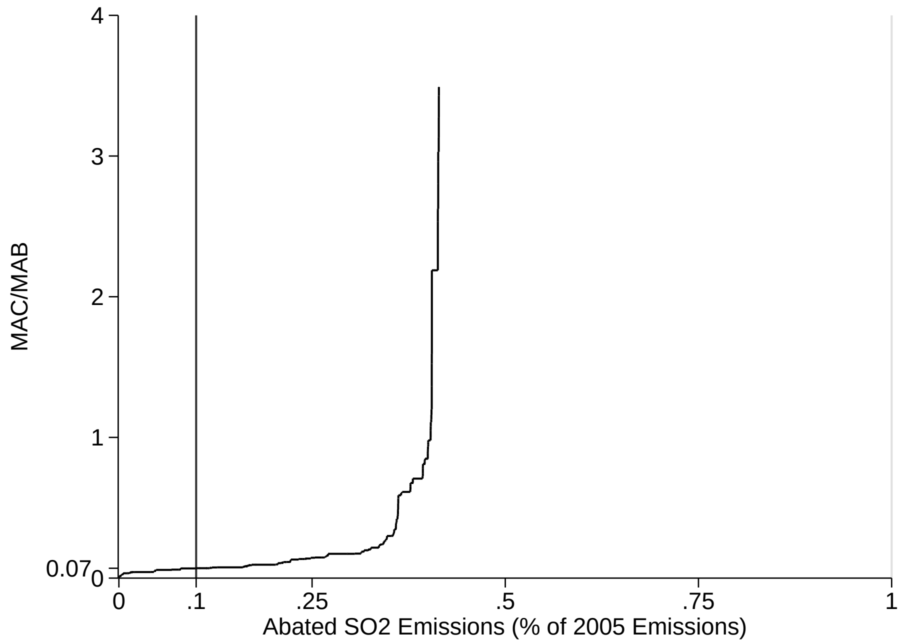


Figure 2.17 *Marginal welfare impact of SO_2 abatement*. This graph shows the ratio of marginal abatement cost (MAC) to marginal abatement benefits (MAB) for SO_2 for China in 2005. The horizontal axis plots the abatement intensity relative to the 2005 SO_2 emissions level. The horizontal, longdashed line at 1 marks the welfare-optimal level of SO_2 abatement (40.3%). The vertical, shortdashed line at 0.1 marks the ratio of marginal abatement cost to marginal abatement benefit for the 10% SO_2 reduction target (0.07). Prohibitive cost ranges beyond a MAC/MAB ratio of 3.5 not shown.

margin (from 658€/tSO_2 to 338€/tSO_2). Secondly, I find that this conclusion is robust to taking into account the benefit side. Adding the benefit measure, I find efficiency improvements of 45% (from a welfare ratio of 2.19 to a welfare ratio of 1.2).

Summarizing the findings, I find that the command-and-control regulation on SO_2 pollution control in the 11th FYP does not take into account cost of abatement. Secondly, I find that a market-based allocation would improve both cost-and welfare-efficiency by 49%. Taken at face value, this finding seems to differ from Carlson et al. (2000) for the US. They find that while an emissions trading scheme led to appreciable, but lower than expected efficiency gains in the US context, these efficiency gains would have been twice as high had the alternative command-and-control policy been

Table 2.10 Efficiency gains from trade

	Efficiency measure	
Allocation	Abatement cost at the margin (in €/tSO2)	Welfare impact at the margin (MAC/MAB)
Actual	658	2.19
Cost-efficient	338	1.2

Note. The efficiency measures report the highest cost of all last abated units across provinces for the actual allocation from the 11th Five-Year Plan (2006-2010) and the counterfactual cost-efficient allocation reported in Table 2.9. Estimates for the cost of the actual allocation exclude outliers. The welfare impact at the margin is the ratio between marginal abatement cost and marginal abatement benefits.

the forced adoption of scrubbers.

In China, those two allocations are nearly identical: Figure 2.14shows that the cost-efficient allocation of SO2 reduction targets relies predominantly on the use of wet flue-gas desulfurization in the electricity sector. Low-sulfur coal, by contrast, only becomes cost-efficient after about 18% of 2005 SO_2 emissions in China are abated, making scrubbers the most economical abatement technology to reduce SO_2 emission levels by 10%. Overall, my findings are therefore surprising compared to Carlson et al. (2000)'s findings from the US: while a market-based policy instrument such as an SO_2 emissions trading scheme with a national cap would double allocative efficiency in China and, at the margin, induce cost savings and increase welfare by 50%, most of those efficiency gains could be reaped by forced installation of scrubbers. The reason for this difference is that the gains from trade in China come mainly from differences in industrial structure across provinces, whereas the in the US experience efficiency gains would have relied on differential access to low sulfur coal due to transport cost.

2.7 Concluding remarks

This research evaluates China's main air pollution control policy. In 2005, China decided on a 10% SO_2 emissions reduction goal as part of the 11th Five-Year Plan (2006-2010). To study the effect of this policy on pollution outcomes, I use both official, misreporting-prone indicator and independent NASA SO_2 satellite data in a differences-in-differences strategy that exploits variation in target stringency at the province level. Overall, my empirical findings show that SO_2 pollution control in the 11th Five-Year Plan (2006-2010) was a success. SO_2 pollution, as measured by independent satellite data, decreased by more than 10% as a result of the policy, most notably through the shutdown of inefficient and small thermal units. Compliance, however, only started when the central government upgraded its pollution monitoring capacity in 2008. I provide suggestive evidence that the increase in monitoring capacity is the likely explanation for this compliance. Before the changes in monitoring, the behaviour by the provincial governments fits the old Chinese adage 'Heaven is High and the Emperor far away': misreporting of pollution data seems pervasive and backed up by misleading rhetorical compliance. This finding is in line with the experiences of environmental regulation summarized in Stavins (2003), who cautions about the importance of monitoring and enforcement. My findings also provide support for a theoretical literature pioneered by Jean-Jacques Laffont suggesting the importance of regulatory capacity for the working of even simple policies (for an overview, see Laffont, 2005).

To analyze the efficiency of the chosen command-and-control regulation, I construct detailed marginal abatement cost curves for each province in China and show the large heterogeneity across provinces in marginal abatement cost. Based on these cost curves, I can quantify the efficiency gain of moving from the actual command-and-control regulation to a counterfactual market-based policy instruments such as a cost-based SO_2 emissions trading scheme. I find that efficiency would have increased by 95% at the margin, lowering marginal abatement cost from 658€/tSO_2 to 338€/tSO_2. This finding is robust to the inclusion of a back-of-the-envelope measure of marginal abatement benefits. Contrary to the US SO_2 control experience, however, most of the efficiency gains of a market-based solution could have been reaped by a technology mandate on scrubbers.

My analysis, then, suggests that China faces similar problems in its air pollution control as more developed countries in the past. Air pollution control is initially plagued by monitoring difficulties, and works in general, though it is not initially cost-efficient. In analyzing environmental policy in China during the 11th Five-Year Plan (2006-2010), I zoom in on a period in recent Chinese history that saw the turning point in environmental governance with the creation of the Ministry of Environmental Protection in 2008. Figure 2.A.7 in Appendix 2.A shows that the change in policy coincides with a change in China's dynamics of pollution and economic growth. My research can thus be seen as a detailed account of how China achieved the turning point on its environmental Kuznets curve (Grossman and Krueger, 1995).

Understanding China's ongoing transition from the planet's biggest dirty industrial powerhouse to a cleaner consumer economy is crucial for China, but also for the planet as a whole. The study of China further adds to the empirical knowledge about the working of regulation in developing countries (Duflo et al., 2013; Oliva, 2015) and shows how appropriate regulatory capacity of the government is a pre-requisite for effective regulation.

Appendix

2.A Additional figures

This appendix contains additional figures that report on a number of robustness checks mentioned in the main body of this chapter. All Figures are self-contained.

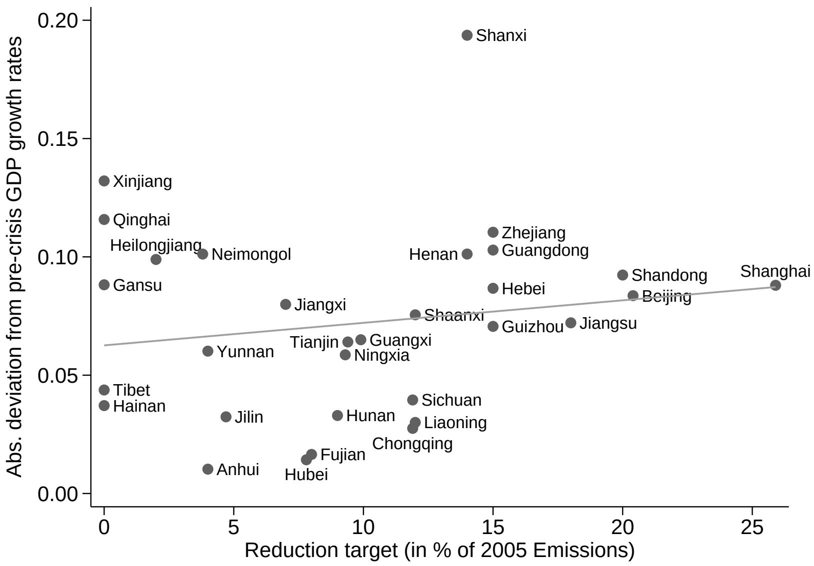


Figure 2.A.1 *The Great Recession and SO_2 reduction targets*. This graph plots the SO_2 reduction targets on the horizontal axis against the absolute deviation from the pre-crisis (2002-2007) GDP growth rate for each province in China in 2009, the year China was struck by the Great Recession. The solid line fits a linear regression with slope parameter b=0.001 and p=0.305 computed from standard errors clustered at the province level.

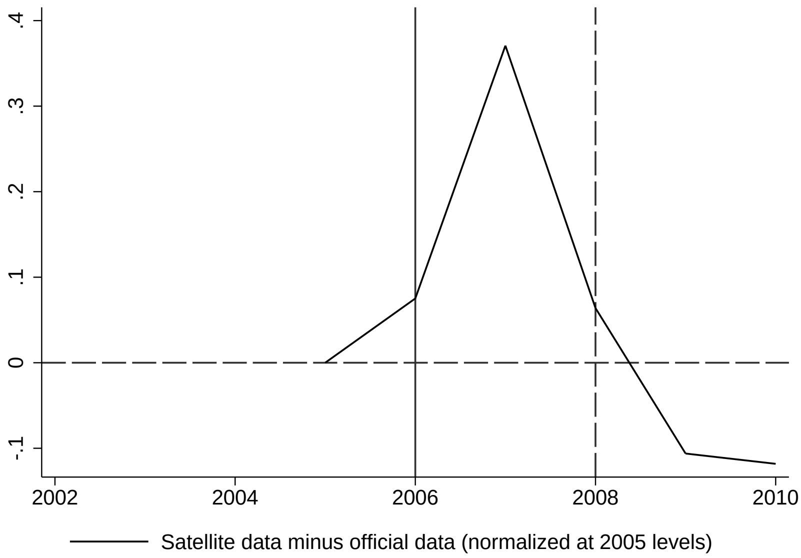


Figure 2.A.2 *Likely misreporting over time*. The graph shows the difference between the SO2 satellite data from NASA and the official SO2 data from the Chinese government. Both variables are normalized to 2005 levels for China.

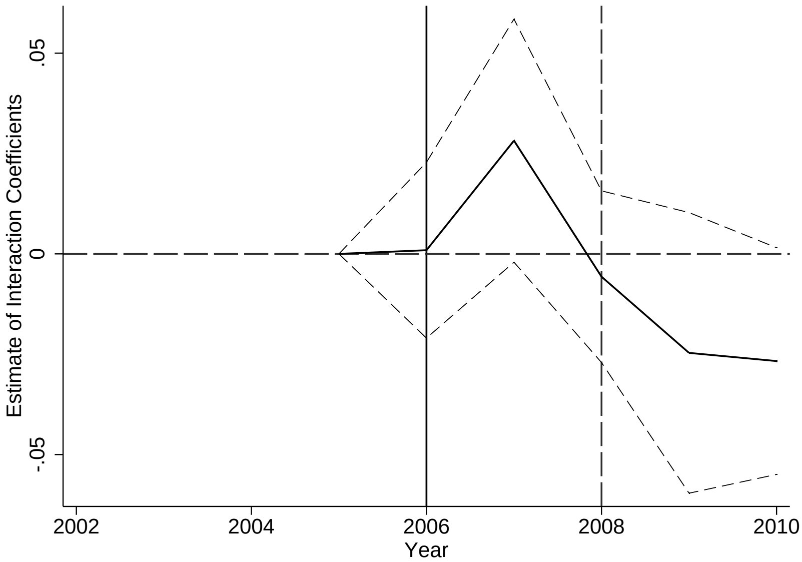


Figure 2.A.3 *Dynamic treatment effects: Likely misreporting*. The solid line plots the point estimate for yearly coefficient estimates of the interaction coefficients $\sum_{t=1}^{T} \beta_{1t}$ in Equation (2.2) for the difference between the SO2 satellite data from NASA and the official SO2 data from the Chinese government. Both variables are normalized to 2005 levels for China. The dashed lines show the 95% confidence bands, based on heteroskedasticity-robust standard errors clustered at the province-level. The excluded time period $t = 1$ is the year 2002. The vertical line at 2006 marks the start of the policy and the vertical line at 2008 marks the start of SO_2 monitoring. The p-value for the 2007 interaction is p=0.13 based on the wild bootstrap procedure with 1000 repetitions or p=0.08 based on the cluster-robust standard errors.

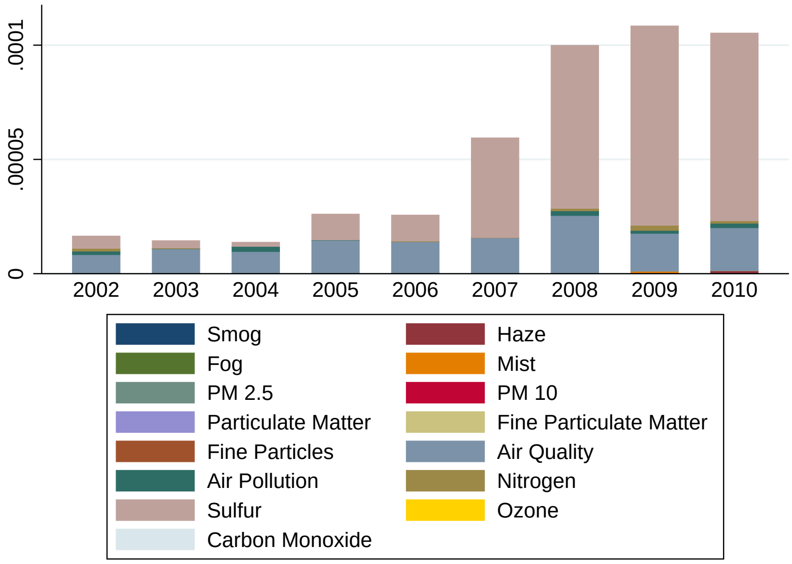


Figure 2.A.4 *Political attention to air pollution over time (normalized by report length)*. The graph shows the mean count of keywords related to air pollution for all provincial government work reports in a given year from 2002 to 2010. The count of keywords is normalized by the length of each government work report.

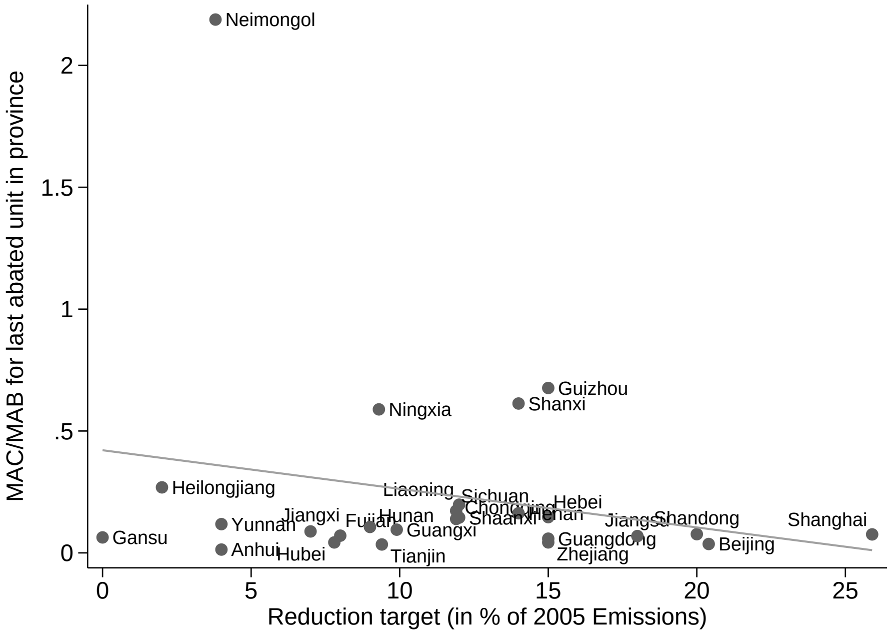


Figure 2.A.5 *Relation between welfare impact and SO_2 reduction targets*. This graph shows the lack of correlation between a province's actual SO_2 reduction target in the 11th Five-Year Plan (2006-2010) and the marginal welfare impact at the level of the target. The solid line fits a linear regression with slope parameter b=-0.16 and p=0.36 computed from standard errors clustered at the province level.

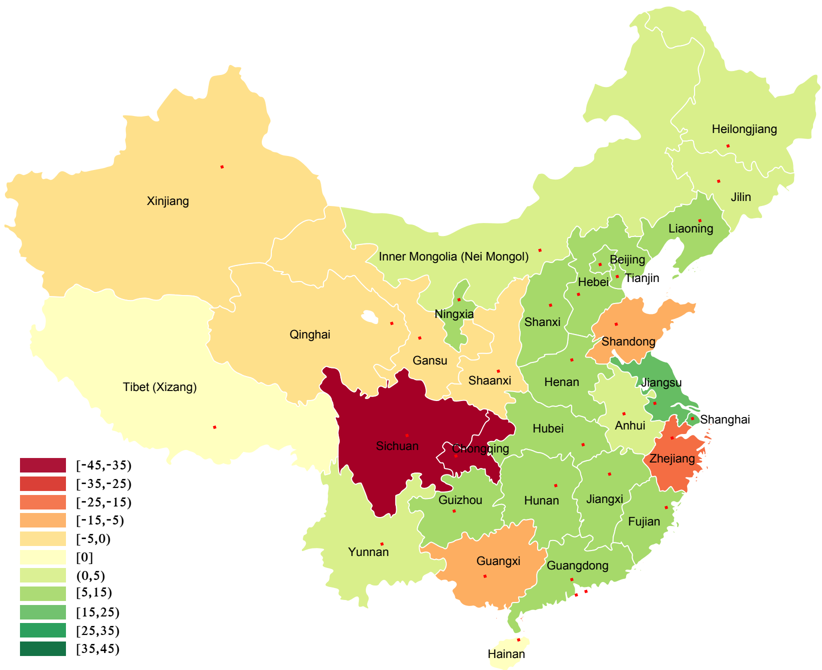


Figure 2.A.6 *Difference between cost-efficient and actual allocation*. The graph shows the difference in percentage points between the SO_2 reduction target of each province under the actual minus the cost-efficient allocation (data from Table 2.9).

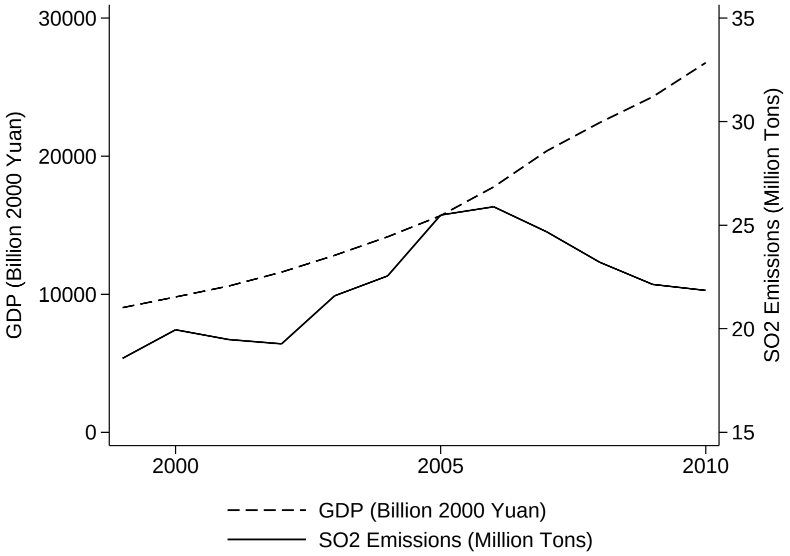
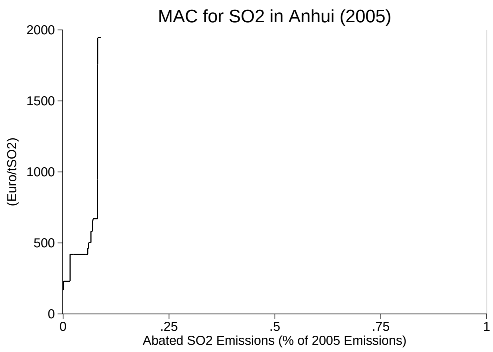
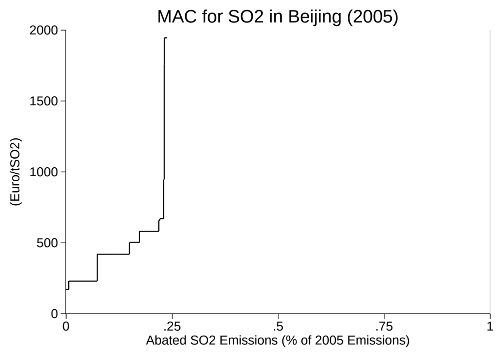


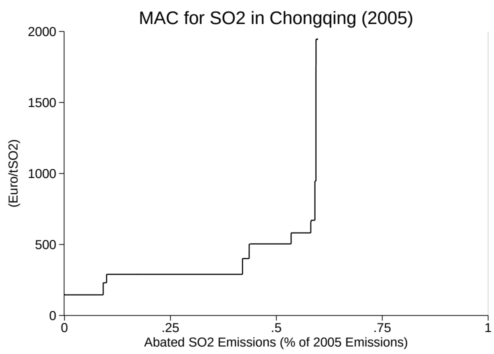
Figure 2.A.7 *Economic growth and air pollution in China*. This figure shows the recent timeseries for GDP (in billion constant 2000 yuan) and for SO_2 emissions (in million tons) for China from 1999 to 2010. Data source: Fridley, Romankiewicz, and Fino-Chen (2013).

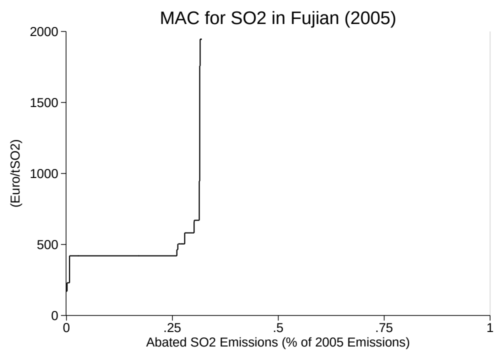
2.B Marginal SO_2 abatement cost curves for all provinces in China

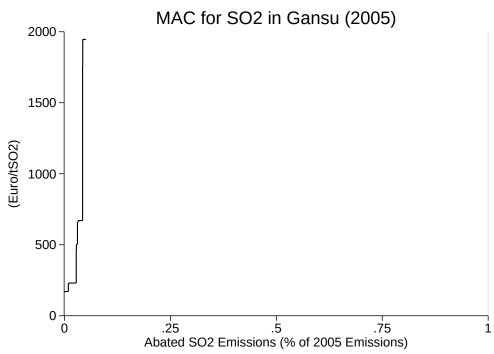
This section shows the marginal SO_2 abatement cost curves for all provinces in 2005. The horizontal axis in each figure lists the abatement intensity in percentage terms relative to the 2005 SO_2 emissions level. Marginal SO_2 abatement cost curve figures that list the abatement intensity in absolute terms (kt) are available upon request.

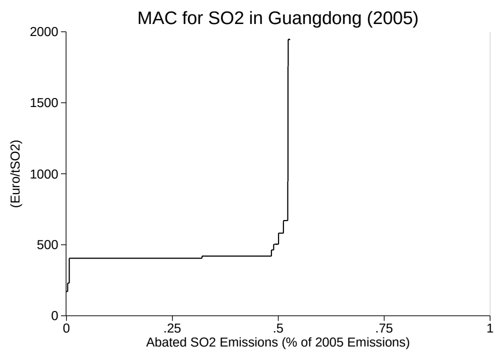


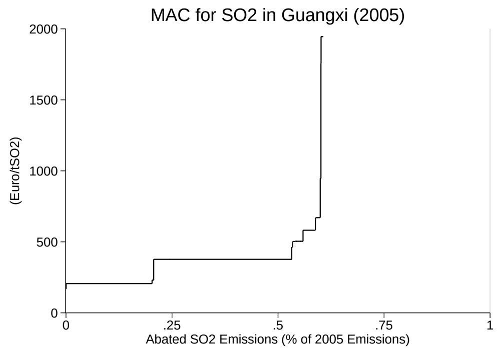


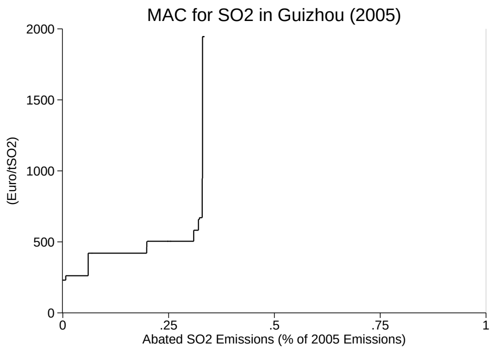


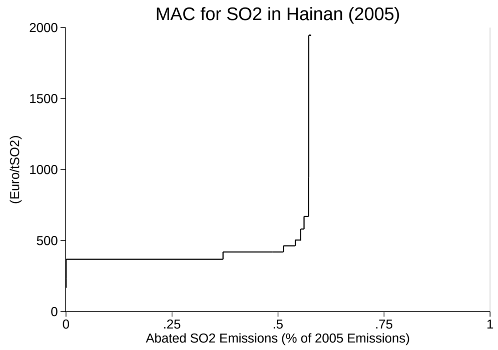


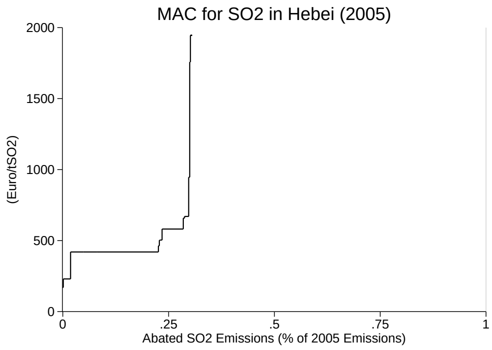


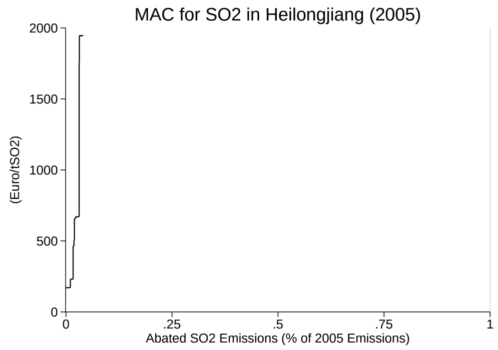


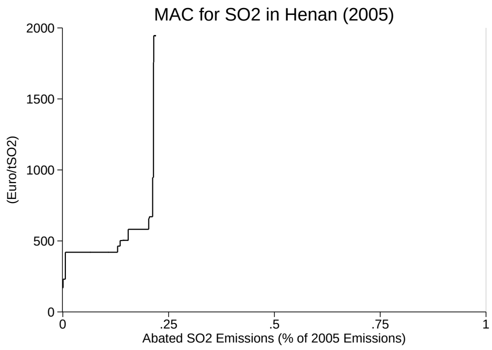

MAC for SO2 in Guangxi (2005)
(Euro/tSO2)
0
500
1000
1500
2000
0
.25
.5
.75
1
Abated SO2 Emissions (% of 2005 Emissions)

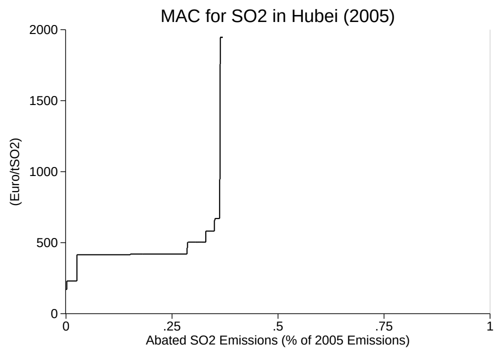

MAC for SO2 in Guizhou (2005)
(Euro/tSO2)
0
500
1000
1500
2000
0
.25
.5
.75
1
Abated SO2 Emissions (% of 2005 Emissions)

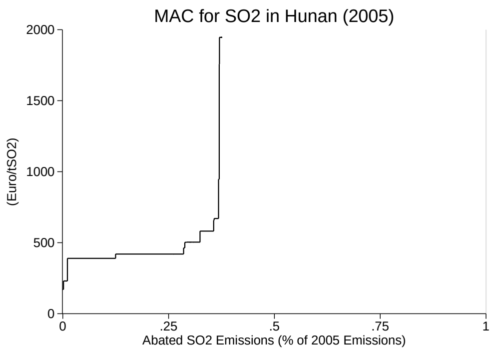

MAC for SO2 in Hainan (2005)
(Euro/tSO2)
0
500
1000
1500
2000
0
.25
.5
.75
1
Abated SO2 Emissions (% of 2005 Emissions)

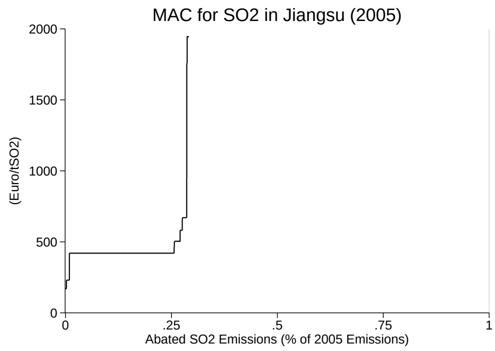

MAC for SO2 in Hebei (2005)
(Euro/tSO2)
0
500
1000
1500
2000
0
.25
.5
.75
1
Abated SO2 Emissions (% of 2005 Emissions)

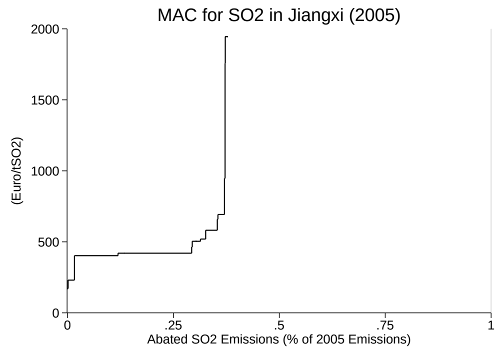

MAC for SO2 in Heilongjiang (2005)
(Euro/tSO2)
0
500
1000
1500
2000
0
.25
.5
.75
1
Abated SO2 Emissions (% of 2005 Emissions)

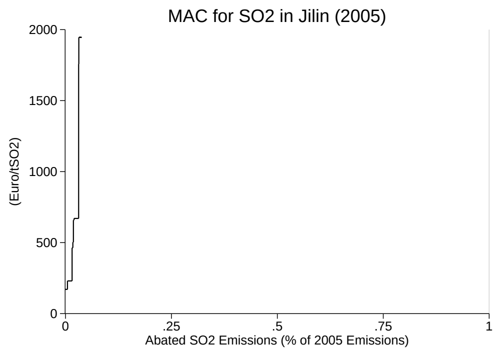

MAC for SO2 in Henan (2005)
(Euro/tSO2)
0
500
1000
1500
2000
0
.25
.5
.75
1
Abated SO2 Emissions (% of 2005 Emissions)

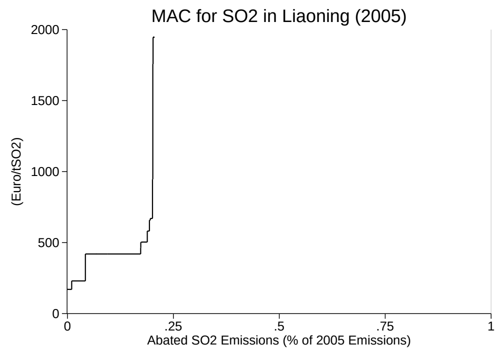

MAC for SO2 in Hubei (2005)
(Euro/tSO2)
0
500
1000
1500
2000
0
.25
.5
.75
1
Abated SO2 Emissions (% of 2005 Emissions)

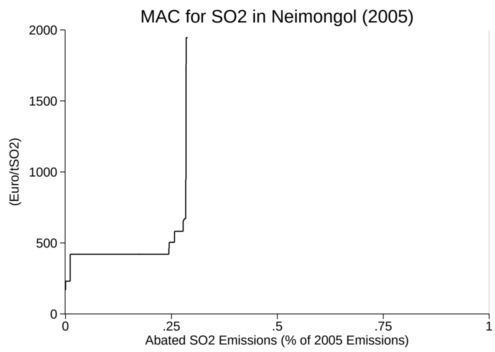

MAC for SO2 in Hunan (2005)
(Euro/tSO2)
0
500
1000
1500
2000
0
.25
.5
.75
1
Abated SO2 Emissions (% of 2005 Emissions)

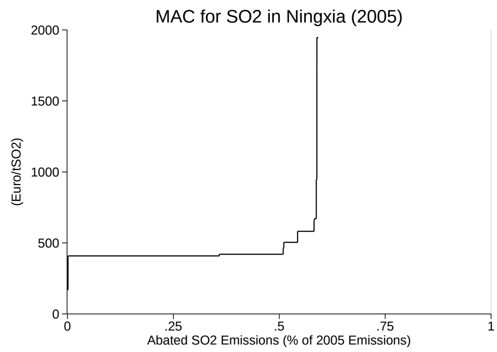

MAC for SO2 in Jiangsu (2005)
(Euro/tSO2)
0
500
1000
1500
2000
0
.25
.5
.75
1
Abated SO2 Emissions (% of 2005 Emissions)

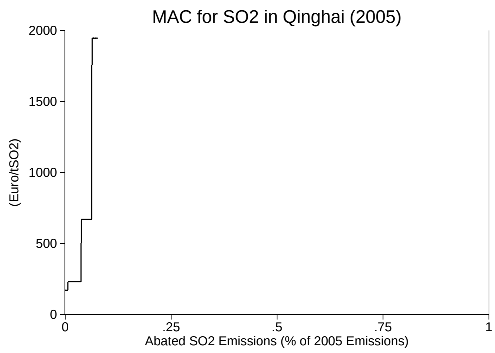

MAC for SO2 in Jiangxi (2005)
(Euro/tSO2)
0
500
1000
1500
2000
0
.25
.5
.75
1
Abated SO2 Emissions (% of 2005 Emissions)

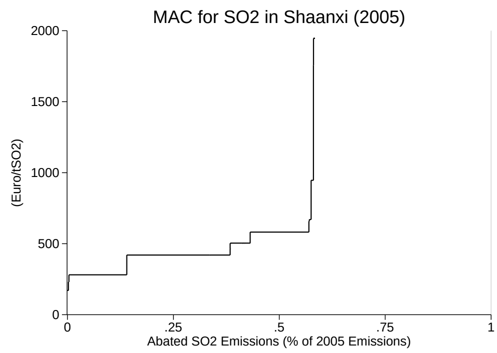

MAC for SO2 in Jilin (2005)
(Euro/tSO2)
0
500
1000
1500
2000
0
.25
.5
.75
1
Abated SO2 Emissions (% of 2005 Emissions)

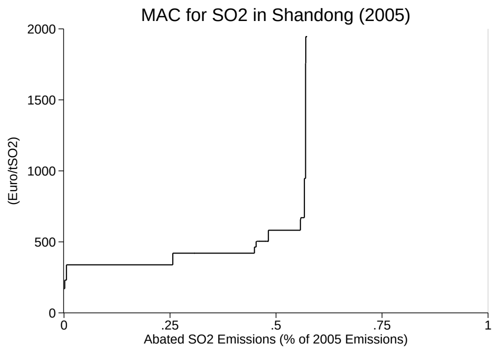

MAC for SO2 in Liaoning (2005)
(Euro/tSO2)
0
500
1000
1500
2000
0
.25
.5
.75
1
Abated SO2 Emissions (% of 2005 Emissions)

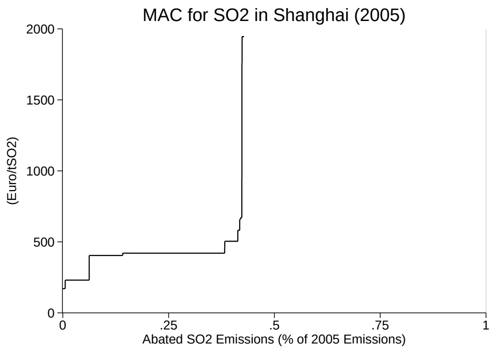

MAC for SO2 in Neimongol (2005)
(Euro/tSO2)
0
500
1000
1500
2000
0
.25
.5
.75
1
Abated SO2 Emissions (% of 2005 Emissions)

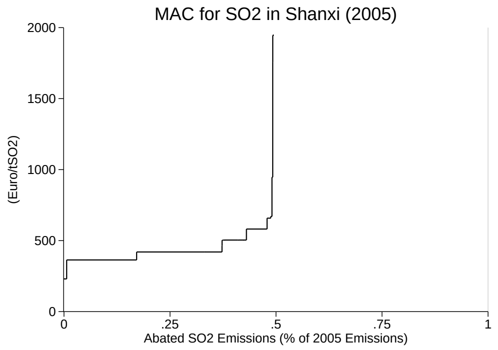

MAC for SO2 in Ningxia (2005)
(Euro/tSO2)
0
500
1000
1500
2000
0
.25
.5
.75
1
Abated SO2 Emissions (% of 2005 Emissions)

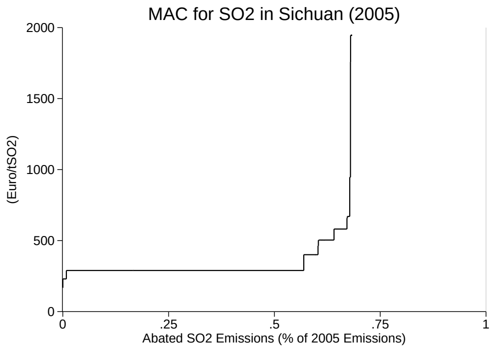

MAC for SO2 in Qinghai (2005)
(Euro/tSO2)
0
500
1000
1500
2000
0
.25
.5
.75
1
Abated SO2 Emissions (% of 2005 Emissions)

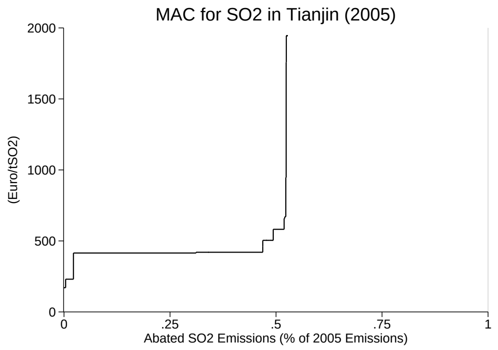

MAC for SO2 in Shaanxi (2005)
(Euro/tSO2)
0
500
1000
1500
2000
0
.25
.5
.75
1
Abated SO2 Emissions (% of 2005 Emissions)

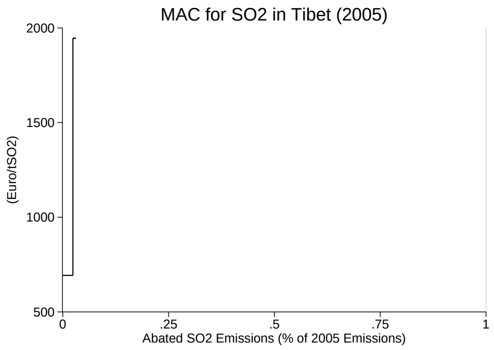

MAC for SO2 in Shandong (2005)
(Euro/tSO2)
0
500
1000
1500
2000
0
.25
.5
.75
1
Abated SO2 Emissions (% of 2005 Emissions)

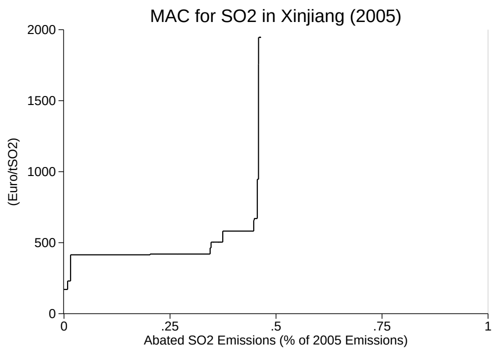

MAC for SO2 in Shanghai (2005)
(Euro/tSO2)
0
500
1000
1500
2000
0
.25
.5
.75
1
Abated SO2 Emissions (% of 2005 Emissions)

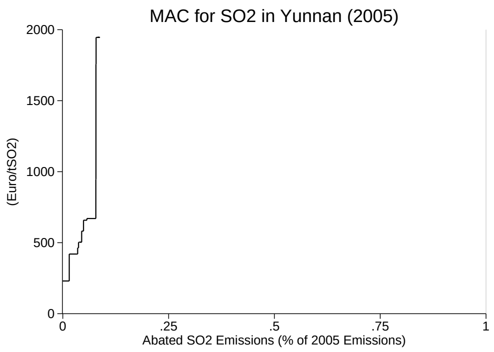

MAC for SO2 in Shanxi (2005)
(Euro/tSO2)
0
500
1000
1500
2000
0
.25
.5
.75
1
Abated SO2 Emissions (% of 2005 Emissions)


MAC for SO2 in Sichuan (2005)
(Euro/tSO2)
0
500
1000
1500
2000
0
.25
.5
.75
1
Abated SO2 Emissions (% of 2005 Emissions)

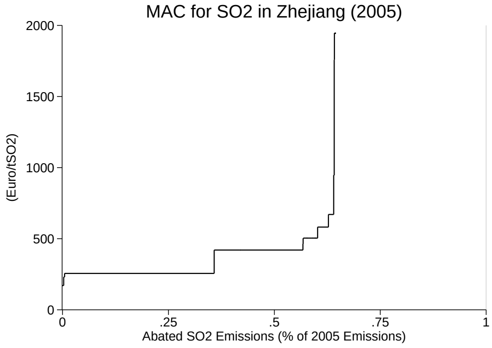

MAC for SO2 in Tianjin (2005)
(Euro/tSO2)
0
500
1000
1500
2000
0
.25
.5
.75
1
Abated SO2 Emissions (% of 2005 Emissions)


MAC for SO2 in Tibet (2005)
(Euro/tSO2)
500
1000
1500
2000
0
.25
.5
.75
1
Abated SO2 Emissions (% of 2005 Emissions)


MAC for SO2 in Xinjiang (2005)
(Euro/tSO2)
0
500
1000
1500
2000
0
.25
.5
.75
1
Abated SO2 Emissions (% of 2005 Emissions)


MAC for SO2 in Yunnan (2005)
(Euro/tSO2)
0
500
1000
1500
2000
0
.25
.5
.75
1
Abated SO2 Emissions (% of 2005 Emissions)

MAC for SO2 in Zhejiang (2005)



Momsen, K. and Stoerk, T. (2014). From intention to action: Can nudges help consumers to choose renewable energy?, *Energy Policy*, 74: 376-382.

Chapter 3

From intention to action: Can nudges help consumers to choose renewable energy

Abstract

In energy consumption, individuals feature a gap between intention and action. Survey data from the US, the UK, and other European countries show that 50-90% of respondents favour energy from renewable sources, even at a small premium. Yet less than 3% actually buy renewable energy. We investigate how nudges - a slight change in the information set that an individual faces when taking a decision - can help individuals align behaviour with intention. We present evidence from an original survey experiment on which nudges affect the choice whether to contract renewable energy or conventional energy. We find that only a default nudge has a significant effect, while all other nudges prove ineffective. In our setting, a default nudge increases the share of individuals who choose renewable energy by 44.6%.

3.1 Introduction

One of the most pressing environmental problems is climate change (Nordhaus, 2013; Stern, 2006). While the energy production overall is the single biggest contributor to greenhouse gas emissions (IPCC, 2007), consuming renewable energy instead of conventional energy reduces these emissions (Shafiei and Salim, 2014). Renewable energy policies that address climate change thus either focus on innovations in technology or changes in behaviour. While policy-making has predominantly relied on the former, we investigate the latter. The following stylized fact shows the potential of our research:

Surveys in various Western countries typically show that 50-90% of respondents favour energy from renewable sources, even at a small premium (Kaenzig, Heinzle, and Wüstenhagen, 2013; Pichert and Katsikopoulos, 2008). Yet, those preferences do not translate into action: Actual users of renewable energy constitute but a tiny fraction of the population, 0.4% in Finland, 0.5% in the UK, 1% in Ireland and Germany, 2% in Switzerland, and 2.8% in the US (Bird, Wüstenhagen, and Aabakken, 2002; Heeter and Nicholas, 2013). The gap between intention and action has only recently been recognised in research on energy behaviour (Allcott and Mullainathan, 2010; Sunstein and Reisch, 2013). A nudge - a slight change in the information set that an individual faces when taking a decision - can help people align intention and action.

The use of nudges as a policy tool has become widespread following Thaler and Sunstein (2008) and Camerer et al. (2003). This literature suggests two complementary rationales for using nudges: Firstly, the gap between intention and action shows that individuals are boundedly rational in the choice between conventional and renewable energy. Due to their limitations in cognitive processes and attention, individuals have difficulties understanding the situation they are in and suffer from an imperfect ability to process new information (Ariely, 2009; Spiegler, 2011; Thaler and Sunstein, 2003). Consequently, they often fail to act upon their long-term intentions (O'Donoghue and Rabin, 1999; Taubinsky, 2013). This is where nudges can help individuals. Nudges are an attractive policy tool: they are cheap and can easily be scaled up. Furthermore, nudges are coercion-free: individuals retain the freedom to pick from the original choice set. Lastly, they are uncontroversial: it is unavoidable to select a way a decision is presented.

Secondly, research on the effectiveness of nudges in energy consumption (Allcott, 2011; Allcott and Mullainathan, 2010; Allcott and Rogers, 2012; Costa and Kahn, 2013) has shown the great effectiveness of using nudges as energy policy instruments. Allcott and Mullainathan (2010), for instance, find that a nudge can lower energy consumption by as much as 2% and at a negative cost. Empirical evidence on the effectiveness of different nudges for the choice between conventional and renewable energy is missing, however. This research paper uses an original survey experiment to test how several nudges affect the choice whether to contract renewable energy or conventional energy. The nudges we implement in our survey each address one or more potential biases in the behaviour of decision makers.

The remainder of this chapter is structured as follows: Section 3.2.1 presents the setting of our experiment, Section 3.2.2 describes each nudge and its implementation, and Section 3.3 presents and discusses the empirical results. Section 3.4 concludes and provides policy recommendations.

3.2 Methods: An Original Experiment

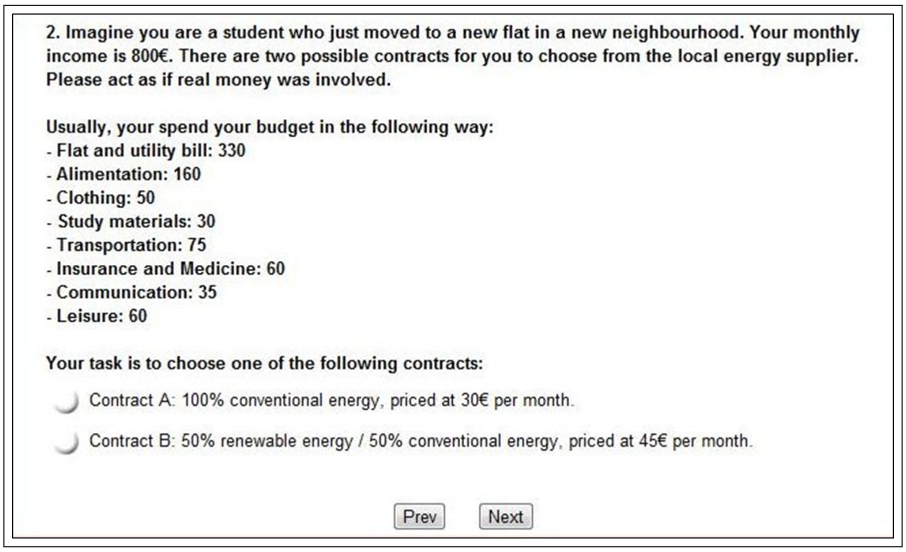
3.2.1 Setting

We provide evidence on which nudges do and which do not work at the time of choosing an energy contract. Our original experiment imitates the situation that a consumer faces when she has just clicked on the website of a utility company and can choose between two different contract offers. To emulate this setting, we implemented the experiment as an online survey (a similar methodology is used by Lillemo (2014)).

Our experiment runs as follows: We ask the subjects to imagine they have moved to a new neighbourhood and need to sign an energy contract. The control group faces two options: buy conventional energy or buy a 50%/50% mix of renewable and conventional energy at a higher cost[1]. The decision for the control group is depicted in Figure 3.1.

Note that we cannot exclude that our subjects were distracted while taking part in our experiment. We consider these potential disturbances a good thing, however, because they add realism to our setting: disturbances also occur when people choose an energy contract in real life.

[1] The choice between a purely conventional energy contract and a contract that offers a 50%/50% mix is due to the offers that were available at the authors' local utility companies at the time of creation of the experiment in early summer 2011.


2. Imagine you are a student who just moved to a new flat in a new neighbourhood. Your monthly income is 800€. There are two possible contracts for you to choose from the local energy supplier. Please act as if real money was involved.

Usually, your spend your budget in the following way:
- Flat and utility bill: 330
- Alimentation: 160
- Clothing: 50
- Study materials: 30
- Transportation: 75
- Insurance and Medicine: 60
- Communication: 35
- Leisure: 60

Your task is to choose one of the following contracts:

Contract A: 100% conventional energy, priced at 30€ per month.

Contract B: 50% renewable energy / 50% conventional energy, priced at 45€ per month.

Prev Next

Figure 3.1 *Decision Screen for the Control Group*. This figure is a caption of the decision screen shown to the subjects assigned to the control group.

The survey was sent to German and international students in June 2011[2]. Since we expected a large share of Germans, the economic choice situation is built on data for income, prices and spending that reflect the typical German student[3]. Note that the default nudge was implemented using a different software. Due to a programming error, we did not obtain any data in 2011. After changing the software, we reran the original default survey without any changes in October 2013 targeting similar subjects[4].

[2]To be precise, we sent it to mailing list subscribers of Club der Ehemaligen e.V., Friedrich-Ebert-Stiftung, Max Weber-Programm Bayern, and Studienstiftung des deutschen Volkes, as well as to graduate students of the Barcelona Graduate School of Economics.

[3]Our main source is a study by the German National Association for Student Affairs (19. Sozialerhebung des Deutschen Studentenwerkes, Kurzfassung (p.22)) that investigated the average budget of German students in 2009. The energy budget comes from a casual survey done among our colleagues at the Barcelona Graduate School of Economics. We round these data for convenience. Our sample features 77% Germans.

[4]We used Surveymonkey for the main part of the experiment, but needed to use GoogleDrive for the default nudge due to that feature not being implementable in Surveymonkey. Unbeknownst to us, GoogleDrive neither recorded the choices nor other data.

3.2.2 The Nudges

We operationalise each nudge in up to three experimental treatments. The original decision screens for all nudges are shown in the Appendix. The following section presents (i) a *review* of the theory and evidence on the working of each nudge and (ii) our *implementation* in the survey experiment.

Priming

Review: Mazar and Ariely (2006) find that having subjects recall the ten commandments decreases cheating. A similar effect can be found in consumption: Morwitz, Johnson, and Schmittlein (1993) find that when asked whether they intend to buy a car in the following six months, consumers' purchase rates increased by 35%. This effect is called "priming" and can be explained by bounded rationality. Tversky and Kahneman (1974), for instance, argue that people assess the probability that an event occurs with the ease by which they can recall examples of it. Following this line of reasoning, Gennaioli and Shleifer (2010) find that a decision maker does not use all available information but relies on what comes to mind. According to Kahneman and Frederick (2005), what comes to mind is shaped by stimulus salience and priming.

Implementation: We implement three different kinds of treatments for priming. *Priming-Intention*: Directly before presenting the actual choice problem, we ask subjects whether they intend to buy renewable energy in the future. *Priming-Memory*: We ask subjects to write down from memory everything they know about the link between climate change and energy production. Thus, they have their own knowledge in mind when taking their decision. *Priming-Reassemble*: Here, we ask subjects to reassemble fragments of sentences about the relationship of energy production and climate change. This revives the subjects' knowledge and makes the negative effects of choosing conventional energy more salient.

Mental Accounting

Review: A lab experiment performed by Mazar and Zhong (2010) shows that individuals who have spent money on green products behave in a less altruistic way in a dictator game than individuals who have spent money on conventional products. The authors cannot fully explain this licensing effect, where a previous ethically favourable action induces subsequent less ethical behaviour. In our view, the above behaviour can be interpreted in the light of mental accounting, according to which individuals classify expenditures into different mental accounts which can act as self-control devices (Thaler, 2004). Consumers hesitate to use the money mentally labelled into one account for a purpose that falls into another.

Interestingly, this reasoning can be applied to interpret the findings of Mazar and Zhong (2010). As established by the authors empirically, the consumption of green products carries ethical cachet compared to conventional consumption. Analogous to the choice between conventional and renewable energy, consumers compare two dimensions: the satisfaction of their consumption needs and the ethical benefit. When deciding how much to offer in a dictator game that follows ethical spending, an individual will give less. By the same token, consumers might attribute the cost of the green product to two different mental accounts, the consumption account and the ethical account.

Implementation: Consumers might tend to choose the ethically favourable renewable energy contract when exposed to a situation that refills their ethical account. We implement *Mental Accounting* by informing subjects that an ethical donation of 15 € has not been successful because the recipient NGO has gone out of business. Notice that 15 € is the price difference between conventional and renewable energy.

Framing

Review: When confronting the individual with the decision which contract to choose, the energy supplier can formulate the choice in different ways. It can inform the consumer about the carbon dioxide emissions she would mitigate with renewable energy, or it can state how much more carbon dioxide is produced when choosing conventional energy. The energy supplier can thus emphasize possible gains or losses, while the outcome is the same.

The study of framing goes back to Tversky and Kahneman (1981), who empirically verified its importance. Their findings can be interpreted from the view of Prospect Theory (Kahneman and Tversky, 1979), according to which individuals evaluate outcomes in terms of deviations from a reference point. The individuals' response when facing a loss is stronger than when experiencing an equivalent gain. In our case, this reference outcome is the option that is not framed in terms of losses or gains. Tversky and Kahneman (1981) conclude that adopting a frame is an ethically significant act that has an effect in the choice process. In our experiment, we want to find out whether individuals can be nudged towards choosing the contract including renewable energies by framing the contracts in a way that they lead to losses or gains.

Implementation: We provide additional information by informing the decision maker about the additional carbon dioxide emissions from Contract A as compared to Contract B. We frame this fact either as gains (*Framing-Gains*) or as losses (*Framing-Losses*).

Decoy

Review: Consumers violate the independence of irrelevant alternatives axiom of von Neumann-Morgenstern expected utility theory (Ariely, 2009; Ariely and Wallsten, 1995). This phenomenon was first studied by Huber, Payne, and Puto (1982): Consider a consumer who has to choose between two products whose attributes differ in various dimensions, and where none of the items is dominant in all the dimensions. An energy contract can differ on essentially two dimensions: price and the percentage of energy produced from renewable sources[5]. The introduction of a third alternative - a decoy - that is clearly dominated by only one of the two alternatives can greatly influence the decision process.

Ariely and Wallsten (1995) analyse this behaviour. The consumer, initially unable to weight the different dimensions, reconstructs the choice space subjectively. By ignoring certain dimensions and giving more weight to others, the consumer ends up with a subjective dominance relationship. The decoy might thus work since it helps the consumer to weight information.

Implementation: In our case, the decoy is an alternative that is weakly dominated by the environmentally friendly contract, as it is equal in the price dimension and dominated in the environmental dimension. We expect this additional information to nudge the individual towards the environmentally friendly contract.

[5]We assume for simplicity that we can pin down environmental impact into this single criterion. This is a simplifying assumption given that nuclear energy is being reconsidered as a low carbon energy source.

Social Norms

Review: Experiments show that people conform to the opinion of others (Sunstein, 2003). Furthermore, evidence from field experiments has shown that an upfront lead donation increases the contributions of potential donors a staggering 44% to 300%, with neutral or positive effects on the response rates (Huck and Rasul, 2007; List and Lucking-Reiley, 2002). In the field of household energy choice, Costa and Kahn (2013) evaluate a randomised field experiment in California, in which information on neighbours' energy consumption is added to households' electricity bills. In particular, each household receives a comparison between its own energy consumption and that of its neighbours plus that of its energy-efficient neighbours. The average treatment effect of this nudge is a 2% decrease in energy consumption. These findings confirm earlier evidence in Allcott and Mullainathan (2010).

Implementation: We add the following sentence to the control group's choice: "From your local energy provider you receive the information that the majority of your neighbours uses an energy mix that features 50% renewable energy".

Default

Review: The fact that people tend to stick to the default option can be explained by inertia. Since decision makers prefer not to change the status quo due to switching costs and loss aversion, they rather decide not to decide (Spiegler, 2011). Rubinstein (2012) refers to this bias of sticking to the chosen option as default tendency.
In the context of organ donations, Johnson and Goldstein (2003) show that the share of organ donors is twice as high when being a donor is the default compared to the situation when not being a donor is the default. Clearly, sticking to the default must yield positive benefits for the decision makers. In the context of energy choice, Pichert and Katsikopoulos (2008) analyse empirically if people stick to the kind of energy that is offered to them as a default contract. Using natural and laboratory experiments, they show that more people end up using renewable energy when this kind of energy is the default contract.

Implementation: We inform our subjects that the default energy contract in their region consists of 50% renewable energy and 50% conventional energy. They can actively choose between this default contract and a contract consisting entirely of conventional energy. If they do not make an active choice, however, they will keep the default contract and use renewable energy.

3.2.3 Recorded Data and Randomisation

Recorded data: The outcome of interest is each subject's choice of energy contract. After that decision, each subject rated their agreement to a number of statements to elicit the subject's preferences for money, the environment, and environmental action in their daily lives. Typical statements are "If I were a little richer, my current life would be more enjoyable", "I am concerned about climate change" and "I am willing to pay higher taxes for improved environmental conservation". The subjects furthermore reported their ecological footprint and their carbon footprint[6]. Additionally, we recorded study major, gender, nationality and age as well as self-reported data on monthly income, rent payment and utility expenses. Table 3.1 reports descriptive statistics for the subjects. Each column belongs to a treatment group or the control group.

Randomisation: We used day of birth as randomisation device. Depending on the stated day, participants were assigned to a different treatment. Note that even if some subjects reported the wrong date of birth, we see no reason to believe that this should have occurred in a systematic way. Table 3.1 shows that randomisation worked: excepting the odd, small-magnitude case, the p-values of a pairwise mean-comparison between each treatment and the control group are insignificant[7] .

Notice that in the main, balancedness also holds for the subjects in the default treatment who were surveyed at a different point in time due to a software error. Default subjects were slightly richer (at a 10% level), favour money more and have a slightly lower ecological footprint. All other covariates are statistically indistinguishable from the control subjects.

[6] We used a tool provided by the World Wide Fund for Nature (WWF), available online http://footprint.wwf.org.uk/ [Last accessed 3rd June 2011 17:00 CET]

[7] As a robustness check, we computed the non-parametric Kruskal-Wallis test to jointly test all treatment groups. We find statistically significant differences in the ecological footprint, the carbon footprint and the preferences for money. As can be seen from Table 3.1, these differences are not large, however.

Table 3.1 Descriptive Statistics and Randomisation

	Priming (Intention)		Priming (Memory)		Priming (Reassemble)		Mental Accounting		Framing (Gains)		Framing (Losses)		Decoy		Social Norms		Default		Control
	Mean [Stand. Dev.]	p-value	Mean [Stand. Dev.]	p-value	Mean [Stand. Dev.]	p-value	Mean [Stand. Dev.]	p-value	Mean [Stand. Dev.]	p-value	Mean [Stand. Dev.]	p-value	Mean [Stand. Dev.]	p-value	Mean [Stand. Dev.]	p-value	Mean [Stand. Dev.]	p-value	Mean [Stand. Dev.]
Demographics																			
Age	24.2	0.71	24.2	0.89	24.1	0.85	24.3	0.55	24.6	0.96	23.4	0.02**	23.2	0.09*	23.8	0.40	25.6	0.55	24.2
	[3.3]		[3.1]		[3.2]		[3.9]		[4.1]		[6.7]		[2.5]		[3.7]		[3.9]		[4.7]
Female	0.56	0.43	0.57	0.59	0.54	0.35	0.48	0.18	0.59	0.57	0.54	0.32	0.58	0.49	0.53	0.30	0.55	0.51	0.65
	[0.50]		[0.51]		[0.51]		[0.51]		[0.50]		[0.51]		[0.50]		[0.51]		[0.52]		[0.48]
German	0.70	0.42	0.79	0.95	0.77	0.93	0.75	0.80	0.79	0.94	0.72	0.60	0.78	0.97	0.83	0.56	0.91	0.33	0.78
	[0.48]		[0.43]		[0.43]		[0.44]		[0.42]		[0.46]		[0.42]		[0.38]		[0.30]		[0.42]
Economist	0.29	0.94	0.18	0.49	0.26	0.83	0.27	0.91	0.370	0.46	0.25	0.74	0.16	0.19	0.27	0.91	0.27	0.93	0.29
	[0.46]		[0.41]		[0.45]		[0.46]		[0.49]		[0.44]		[0.37]		[0.46]		[0.47]		[0.46]
Monthly Budget	1031.2	0.48	823.0	0.22	853.1	0.06*	1052.1	0.42	1128.2	0.86	1048.4	0.55	897.7	0.08*	894.0	0.17	1318.1	0.09*	1037.1
	[619.3]		[376.9]		[487.1]		[935.3]		[827.3]		[679.2]		[540.3]		[576.7]		[555.4]		[497.5]
Monthly Rent	328.3	0.35	281.6	0.21	276.8	0.09*	342.3	0.33	315.1	0.15	349.3	0.52	298.2	0.29	287.7	0.10	420.6	0.45	368.3
	[237.9]		[115.6]		[127.6]		[289.7]		[269.5]		[269.4]		[135.20]		[217.9]		[279.9]		[190.8]
Monthly Utility Bill	157.0	0.55	125.5	0.55	210.0	0.20	327.7	0.07*	320.4	0.51	279.8	0.14	152.8	0.84	182.2	0.87	113.7	0.42	178.2
	[179.7]		[156.5]		[167.9]		[613.7]		[536.0]		[335.7]		[141.8]		[186.2]		[145.0]		[236.2]
Preferences for																			
Environment	5.04	0.40	4.84	0.30	5.17	0.24	4.85	0.46	5.02	0.75	4.85	0.49	4.98	0.93	5.07	0.53	5.09	0.59	4.96
	[0.82]		[0.52]		[0.66]		[0.72]		[0.70]		[0.76]		[0.75]		[0.70]		[0.58]		[0.77]
Money	4.47	0.98	4.48	0.57	4.45	0.64	4.71	0.12	4.39	0.19	4.57	0.96	4.65	0.41	4.38	0.56	5.16	0.00***	4.55
	[0.85]		[0.67]		[0.76]		[0.78]		[0.57]		[0.60]		[0.74]		[0.96]		[0.48]		[0.70]
Environmental Action	4.57	0.14	4.24	0.57	4.59	0.21	4.40	0.69	4.48	0.45	4.25	0.32	4.36	1.00	4.41	0.97	4.41	0.95	4.36
	[0.70]		[0.76]		[0.59]		[0.75]		[0.85]		[0.58]		[0.96]		[0.77]		[0.68]		[0.79]
Behavior																			
Ecological Footprint	2.85	0.20	2.69	0.76	2.47	0.31	2.58	0.66	2.56	0.38	3.08	0.13	3.05	0.098*	2.46	0.29	2.13	0.03**	2.74
	[0.59]		[0.60]		[0.61]		[0.71]		[0.82]		[1.10]		[0.99]		[0.63]		[0.26]		[0.96]
Carbon Footprint	13.55	0.34	11.64	0.83	10.64	0.26	11.38	0.46	11.20	0.33	14.19	0.29	14.64	0.19	10.39	0.14	13.14	0.35	12.88
	[5.508]		[5.50]		[4.22]		[4.88]		[5.04]		[6.46]		[7.01]		[4.64]		[1.69]		[6.68]

Note. This table shows descriptive statistics on the composition of each treatment group along three dimensions of covariates (*Demographics*, *Preferences*, *Behavior*). The p-value is obtained from a parwise Wilcoxon ranksum test of equality of means between each treatment group and the control group. The overall lack of significant differences between treatment and control groups shows that the randomisation of the treatments worked. $***p < 0.01$, $**p < 0.05$, $*p < 0.1$.

3.3 Results and Discussion

3.3.1 Results

In this section, we first present summary statistics for the different treatment groups. Second, we test whether the probability of choosing the renewable contract is significantly different in any of the nudging groups compared to the control group.

Figure 3.2 shows how the 475 subjects chose, according to each treatment group. The result from the no-nudge comparison group is the benchmark against which we compare the effectiveness of each nudge. As can be seen, renewable energy was chosen by 41 out of 85 subjects in this group, or 48,2%. Recall that for all nudges we test the following H_0: $\mu_{control} = \mu_{nudge}$, where μ_{nudge} depicts the mean choice of contract in treatment groups[8].

[8] For the decoy treatment, we coded the dominated decoy contract as "0.5", reflecting the fact that it is an intermediate option between the conventional energy contract and the renewable energy contract.

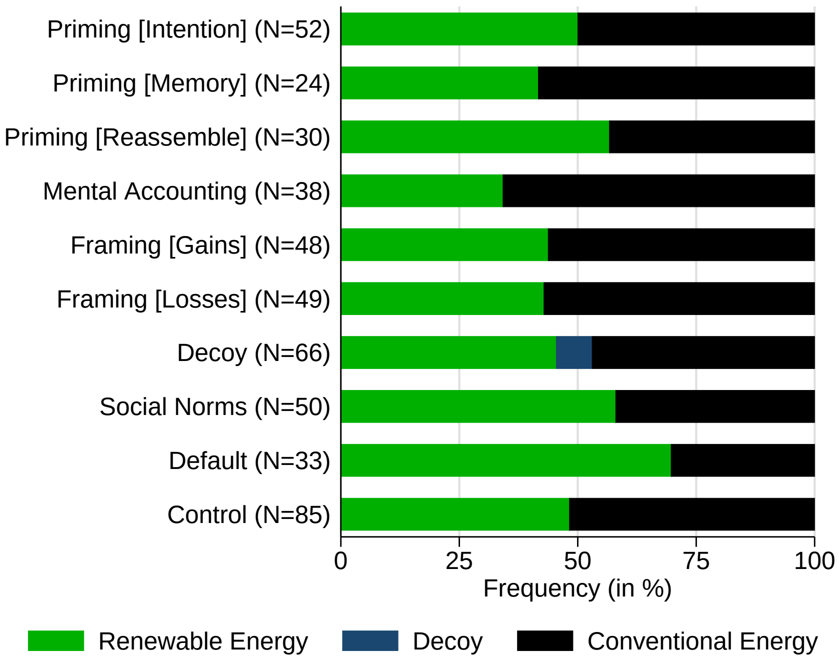


Figure 3.2 *Raw Choice Data*. This figure shows the choices in each treatment group as a share of all choices in that group.

Figure 3.2 suggests that several nudges seem to have a strong effect on the choice of renewable versus conventional energy. Some nudges seem to have worked as expected, such as Priming [Reassemble] (share of choices for renewable: 56.7%), Social Norms (58%) or Default (69.7%). Interestingly, other nudges appear to have a *negative* effect: they seem to increase the share of respondents that choose conventional energy. In particular Priming [Memory] (41,7%) and Mental Accounting (34.2%) appear to have worked in this way. In the next step, we determine whether these graphical differences are statistically significant .

Table 3.2 Average Treatment Effects

	Linear probability model	p-value	t-statistic
Priming (Intention)	0.0176	0.843	0.198
	(0.0889)		
Priming (Memory)	-0.0657	0.570	-0.569
	(0.1155)		
Priming (Reassemble)	0.0843	0.429	0.791
	(0.1066)		
Mental accounting	-0.1402	0.141	-1.474
	(0.0951)		
Framing (Gains)	-0.0449	0.621	-0.494
	(0.0908)		
Framing (Losses)	-0.0538	0.551	-0.597
	(0.0900)		
Decoy	0.0101	0.901	0.124
	(0.0811)		
Social Norms	0.0976	0.275	1.093
	(0.0893)		
Default	0.2146**	0.028	2.197
	(0.0977)		
Observations	475		
R^2	0.028		

Note. Dependent variable equals 1 if the subject has chosen the contract with renewable energy, 0.5 for the decoy contract, and zero otherwise. Model includes a constant. Robust standard errors in parentheses. $***p < 0.01$, $**p < 0.05$, $*p < 0.1$.

Table 3.2 reports the results from a linear probability model with dummies for each treatment group (the omitted category being the control group). Because the treatments were assigned using day of birth, randomization is not conditional on the covariates. This allows us to estimate the average treatment effect for each nudge using a linear probability model without additional controls. The coefficients on the dummies show the average treatment effect of each nudge. As can be seen from Table 3.2, only the default nudge had a significant effect. Following our treatment, the relative number of subjects who selected the renewable energy contract increased by 44.6% relative to the control group (from 48.2% to 69.7%). While we are aware of the fact that the *magnitude* of the effect in our survey experiment does not directly translate into real world applications, we believe that the *direction* of the effect can be trusted: Our results show that the mechanism of the default nudge works convincingly in our decision situation of whether to contract conventional or renewable energy.

We prefer the linear probability model as compared to a Probit or Logit specification due to its easier interpretability. To test the robustness of our findings, however, we estimated the average treatment effects of the same model via both Probit and Logit and find qualitatively identical effects of very similar magnitudes (Appendix 3.B reports the results from those tests). As a further robustness check, we compute the same hypothesis tests via t-tests and non-parametric Wilcoxon rank sum tests. In both cases, only the default nudge is statistically significant up to the 10% level. These results are available upon request.

3.3.2 Discussion

Heterogeneous treatment effects. This section provides a tentative analysis of heterogenous treatment effects for different subgroups for each treatment. This analysis is tentative only and meant only to inform discussion for two reasons: firstly, the subgroups are small. More importantly, however, there is a possible selection bias: Many subjects in the experiment chose not to fill in the questionnaires after they made their choice, and the decision whether or not to fill in the questionnaire is unlikely to be unrelated to subject characteristics. This is in contrast to the main survey: when we recruited the subjects, we were careful not to announce what the survey was about to avoid self-selection of subjects particularly concerned with the environment. For the decision whether to fill in the questionnaire, however, this is not the case: the subjects knew what the study was about, and the decision whether or not to continue might be related to subject characteristics.

We nonetheless find it interesting to report the results from the subgroup analysis in light of the findings from Costa and Kahn (2013), who show that the same nudge can have opposing effects depending on political leaning. To guard against cherry-picking we chose all subgroups beforehand when designing the survey. The procedure for testing is identical to the full sample; that is we test whether the mean of the restricted subgroup in a nudge condition is significantly different from the mean of the corresponding subgroup in the no-nudge comparison group. Given that the subgroups are smaller than the full sample treatment groups, we only use the Wilcoxon rank sum test which has better properties in small samples compared to a t-test (Siegel, 1956). While the smallest of any subgroup contained only 3 observations, the vast majority contained more than 10, which is the usual rule of thumb for the minimum of observations needed.

Table 3.3 reports the p-values from a Wilcoxon ranksum test. For convenience, we mark the ones below the 10% significance level in bold letters and with colours indicating the direction of the effect of the nudge on the corresponding subgroup. Green implies that the nudge worked favourably and increased the probability to choose the renewable contract. Conversely, the brown color marks that the nudge made subjects more likely to choose conventional energy.

As shown in Table 3.3, we where interested in the following subgroups. First, we investigated whether a nudge might work differently depending on the gender. For our experiment, the answer is no (however, not enough male subjects reported their gender to determine the effect on males only for the default nudge). Second, we were interested in whether economics students, who might have theoretical knowledge about nudges, were affected differently than their colleagues from other majors. This seems to have been the case. Economists that were subject to reassembling the text about energy production and climate change were more likely (at the 10% level) to choose the renewable contract than their economist peers in the no-nudge condition. This is particularly interesting since the tendency of this treatment went in the other direction for the full sample. In addition, the nudge Social Norms had a significant effect (10%) only on economists. As expected from the raw choice data, the effect goes into the desired direction.

Next, we split participants into the relatively rich and relatively poor. For the rich subjects, Social Norms proves effective (10%), with an effect that goes into the desired direction. Poor subjects, on the other hand,

Table 3.3 Heterogeneous Treatment Effects

	Priming (Intention)	Priming (Memory)	Priming (Reassem-ble)	Mental Accounting	Framing (Gains)	Framing (Losses)	Decoy	Social Norms	Default
Demographics									
Gender									
Female	0.50	0.90	0.29	0.58	0.64	0.40	0.36	0.30	0.056*
Male	0.29	0.36	0.28	0.29	0.82	0.71	0.35	0.61	*i*
Major									
Economics	0.46	0.44	0,097*	0.19	0.80	0.59	0.55	0,097*	0.19
Non-Economics	0.97	0.59	0.43	0.28	0.64	0.65	0.80	0.62	*i*
Budget									
Above Median	0.96	0.68	0.21	0.22	0.77	0.58	0.59	0,080*	0.037**
Below Median	0.89	0,062*	0.39	0.30	0,091*	0,024**	0.33	0.28	0.65
Rent									
Above Median	0.76	0.74	0.63	0.17	0.62	0.95	0.54	0,081*	0.009***
Below Median	0.56	0.54	0.71	0.45	0.74	0.27	0.33	0.49	0.41
Preferences for									
Environment									
Above Median	0.94	0.34	0.80	0.12	0.13	0.42	0.61	0.61	0.43
Below Median	0.84	0.51	0,064*	0.95	0.17	0.84	0.48	0.23	0.031**
Money									
Above Median	0.56	0.86	0.53	0.27	0.98	0.54	0.49	0,096*	0.025**
Below Median	0.71	0.35	0.89	0.81	0.31	0.83	0.71	0.62	0.24
Environmental Action									
Above Median	0.71	0.45	0.67	0.21	0.15	0.54	0.67	0.31	0.47
Below Median	0.63	0.49	0.15	0.40	0.15	0.37	0.32	0.58	0.001***
Behavior									
Ecological Footprint									
Above Median	0.53	0.95	0.76	0.27	0.49	0.41	0.92	0.17	0.032**
Below Median	0,072*	0.31	0.42	0.38	0.92	0.82	0.63	0.92	0.67
Carbon Footprint									
Above Median	0.99	0.61	0.44	0.39	0.94	0.70	0.80	0.13	0.040**
Below Median	0.69	0.11	0.95	0.18	0.41	0.63	0.97	0.78	0.96

Note. The depicted values are the p-values from Wilcoxon ranksum tests. The color indicates the direction of the treatment effect,with green indicating an increased probability to choose the renewable contract. Conversely, brown indicates a decreased probability of choosing the renewable contract. *i* denotes a lack of subjects in the subgroug. $***p < 0.01$, $**p < 0.05$, $*p < 0.1$.

appear to have suffered adverse effects from Priming (Memory) (10%), Framing (Gains) (5%) and Framing (Losses) (10%). The richer subject effect of Social Norms (5%) repeats itself for subjects that pay a higher rent. They become more likely to choose the renewable contract. Last, those that like the environment less are affected as predicted by Priming-Reassemble (10%) and those that like money more are responsive to Social Norms (10%).

As a next step, we look at subgroups according to environmental behavior, which we recorded with ecological footprint and carbon footprint measures from an external WWF survey. We split the subgroups into those subjects into weakly above and below the median. While investigating subgroups for reported environmental behavior and carbon footprint leads to no significant effects, the ecological footprint subgroup below the median does for Priming (Intention) (10%). It might thus be interesting to investigate whether the carbon footprint and the ecological footprint do not generally move one to one. Finally, we find a number of significant subgroups for the Default treatment but do not want to stress an interpretation, as few subjects in the Default treatment responded to the questionnaire and the significant subgroups simply tend to be those with the larger number of subjects.

Relation of to other studies. Compared to the findings from the field studies on nudges in energy consumption analysed in Allcott and Mullainathan (2010) that find a 2% effect from a nudge, a 44.6% increase in the uptake of renewable energy seems very large. Such a large effect of a default, however, is consistent with the anecdotal evidence reported in (Pichert and Katsikopoulos, 2008), who mention two real-world examples from Germany where the share of households using renewable energy exceeded 90% after the introduction of a default.

Another literature related to our paper is the research on the intention to consume sustainably produced food. Whereas we focus on the gap between intention and action, the literature on sustainable food consumption mainly deals with the difference between attitude and intention: Although many people express a positive attitude towards sustainably produced food, only few people actually plan to consume it (Arvola et al., 2008; Vermeir and Verbeke, 2006, 2008). Using surveys, the mentioned authors try to explain the gap between attitude and behavioural intentions. Our research goes one step further in the process of decision making and investigates the gap between intention and action.

Our experiment was designed to establish the effectiveness of each nudge. To establish through which channels each nudge was effective or not, however, is left for future research. As Thaler and Sunstein (2008) and Sunstein and Reisch (2013) point out, defaults work through various channels such as loss aversion, endorsement and inertia, and choice complexity. Hence, it is likely through one or a combination of these channels that defaults proved effective in our survey experiment. It could well be that it is the multiplicity of channels that made the default nudge superior to the other nudges in the selection of renewable energy. Furthermore, we studied our participants' comments which they were free to give during the survey. However, the comments did not show any pattern that could hint at a reason why only the default nudge was effective.

One final caveat concerning the results is that most of the subjects were German. While our sample includes many international subjects as well, the magnitude of the effect could differ depending on the individual nationalities. We leave this question for future research as our sample is too small to study these subgroups separately.

3.4 Conclusions and Policy Implications

Climate change is a severe problem for the environment. Policies that address climate change either focus on technological innovations or behavioural changes. Whereas policy-making has mainly focused on technology, our research studies behaviour. We investigate how nudges affect an individual's decision to choose between renewable energy and conventional energy. Nudges are an attractive policy tool because they are inexpensive, free of coercion and implementable at scale. Besides, nudges are unavoidable in most situations.

Our research speaks to policy because in many Western countries a clear majority of consumers exhibit a gap between intention and action: they consume conventional energy, although they prefer renewable energy and would be willing to pay a premium (Kaenzig, Heinzle, and Wüstenhagen, 2013; Pichert and Katsikopoulos, 2008). Only a tiny fraction of the population, however, actually uses renewable energy (Bird, Wüstenhagen, and Aabakken, 2002). At the same time, the consumption of renewable energy has been linked to reductions in greenhouse gas emissions (Shafiei and Salim, 2014). A policy that can bridge this gap by helping consumers to follow through on their intention to contract renewable energy would be an effective way of mitigating greenhouse gas emissions.

To inform the design of such policy, we present evidence from an original survey experiment on which nudges have an influence on the choice whether to contract renewable energy or conventional energy. We present and review the literature on six nudges that are known to work in related decision situations and adapt them to the decision on the source of energy. Our empirical results show that only a default nudge has a significant effect, while all other nudges prove ineffective. In our setting, the introduction of a default option increases the share of individuals who chose renewable energy by 44.6%. While we believe that the precise *magnitude* of the effect is not informative outside the context of our experiment, we are convinced that the *direction* of the effect translates to the field setting that is relevant for policy.

We therefore inform policy in two dimensions: Our research can inform private actors such as utility companies about the optimal design of the information that they make available to consumers, for instance through online marketing (Herbers and Ramme, 2014). Public actors, on the other hand, could use our findings to consider the implementation of default renewable energy contracts as an alternative way of promoting renewable energies. For this task, more research is needed to design the most efficient default contract in the field. While default nudges in the choice of renewable versus conventional energy have been introduced in the field by a number of private companies and shown to work in general (see Pichert and Katsikopoulos, 2008; Kaenzig, Heinzle, and Wüstenhagen, 2013), a systematic evaluation is still lacking.

Appendix

3.A Original decision screens

The following figures show the screen captions for each of the treatment groups. We only show the part of the decision screen that differs from the control group.

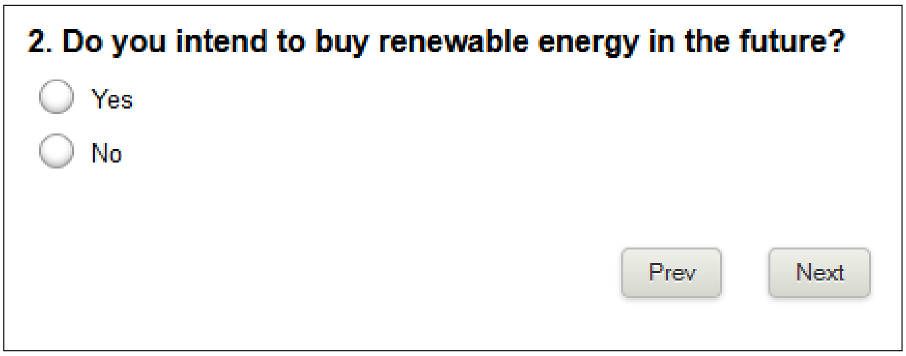
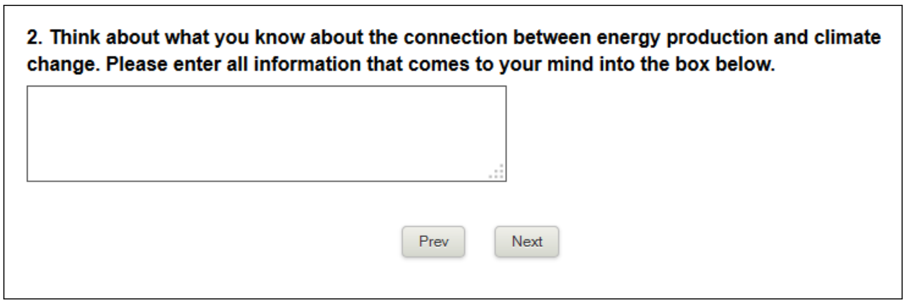


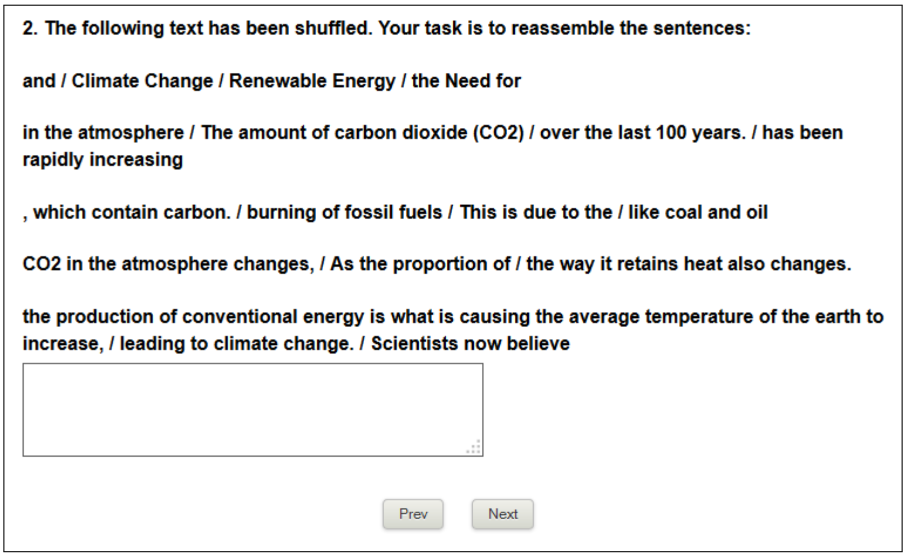
Figure 3.A.1 *Decision Screen for Priming (Intention).*



2. Think about what you know about the connection between energy production and climate change. Please enter all information that comes to your mind into the box below.

Prev Next

Figure 3.A.2 *Decision Screen for Priming (Memory).*



2. The following text has been shuffled. Your task is to reassemble the sentences:

and / Climate Change / Renewable Energy / the Need for

in the atmosphere / The amount of carbon dioxide (CO2) / over the last 100 years. / has been rapidly increasing

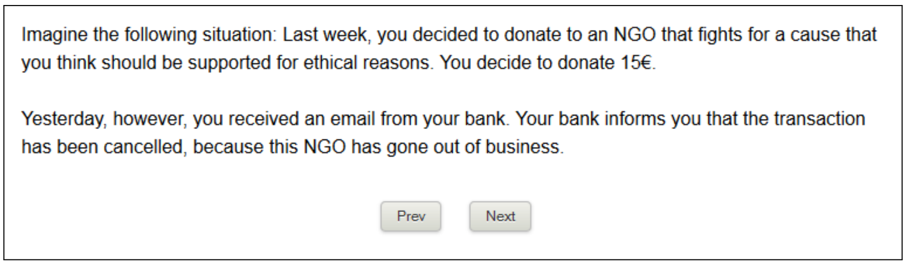
, which contain carbon. / burning of fossil fuels / This is due to the / like coal and oil

CO2 in the atmosphere changes, / As the proportion of / the way it retains heat also changes.

the production of conventional energy is what is causing the average temperature of the earth to increase, / leading to climate change. / Scientists now believe

Prev Next

Figure 3.A.3 *Decision Screen for Priming (Reassemble).*



Imagine the following situation: Last week, you decided to donate to an NGO that fights for a cause that you think should be supported for ethical reasons. You decide to donate 15€.

Yesterday, however, you received an email from your bank. Your bank informs you that the transaction has been cancelled, because this NGO has gone out of business.

Prev Next

Figure 3.A.4 *Decision Screen for Mental Accounting.*

Figure 3.A.5 *Decision Screen for Framing (Gains)*.

Figure 3.A.6 *Decision Screen for Framing (Losses)*.

Figure 3.A.7 *Decision Screen for Decoy*.

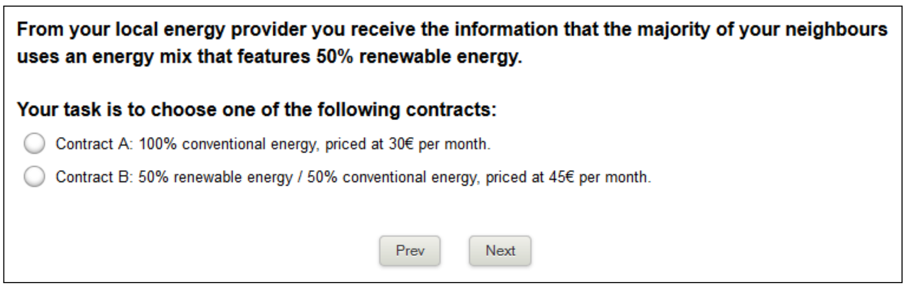


Figure 3.A.8 *Decision Screen for Social Norms*.

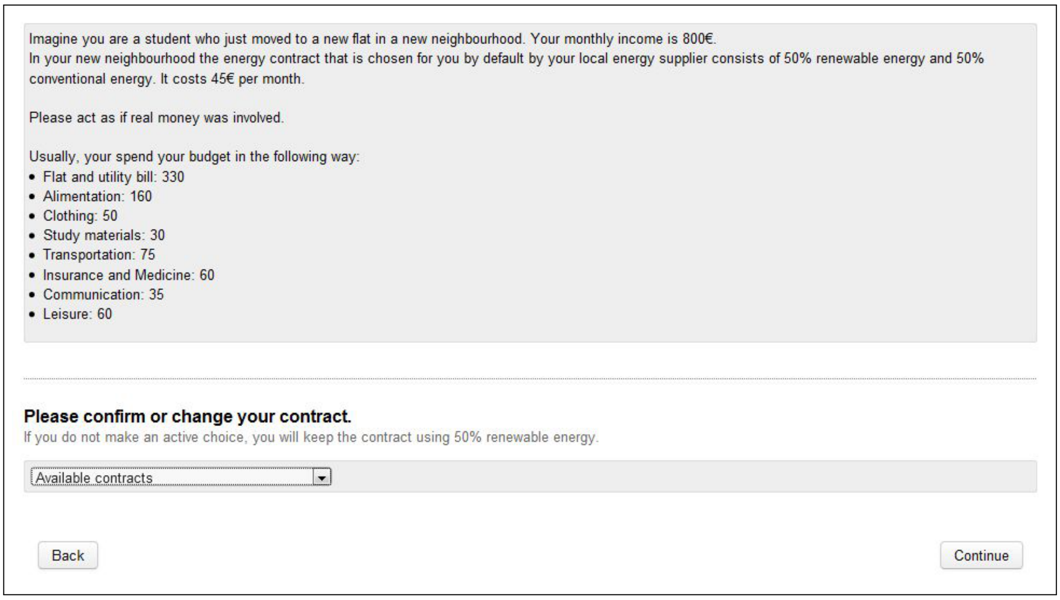


Figure 3.A.9 *Decision Screen for Default*.

3.B Robustness Check: Probit and Logit Results

This section contains results from a robustness check. We estimate the average treatment effect of each treatment using Probit and Logit estimation rather than our preferred Linear Probability Model to show that our findings are robust to the choice of specification. As can be seen in Tables 3.B.1 and 3.B.2, all specifications yield effects that are qualitatively identical and of very similar magnitudes.

Table 3.B.1 Average Treatment Effects (Probit Specification)

	Probit model	p-value	z-statistic
Priming (Intention)	0.0177	0.841	0.20
	(0.088)		
Priming (Memory)	-0.0663	0.569	-0.57
	(0.1163)		
Priming (Reassemble)	0.0846	0.427	0.79
	(0.1066)		
Mental accounting	-0.1446	0.147	-1.45
	(0.0905)		
Framing (Gains)	-0.0451	0.618	-0.50
	(0.0905)		
Framing (Losses)	-0.0542	0.547	-0.60
	(0.0900)		
Decoy	0.0480	0.559	0.58
	(0.0821)		
Social Norms	0.0982	0.273	1.10
	(0.0895)		
Default	0.2234**	0.036	2.10
	(0.1063)		
Observations	475		
Pseudo R^2	0.021		

Note. Dependent variable equals 1 if the subject has chosen the contract with renewable energy, 0.5 for the decoy contract, and zero otherwise. Model includes a constant. Marginal effects calculated at sample mean. Standard errors in parentheses. $***p < 0.01$, $**p < 0.05$, $*p < 0.1$.

Table 3.B.2 Average Treatment Effects (Logit Specification)

	Logit model	p-value	z-statistic
Priming (Intention)	0.0177	0.841	0.20
	(0.088)		
Priming (Memory)	-0.0665	0.570	-0.57
	(0.1169)		
Priming (Reassemble)	0.0847	0.428	0.79
	(0.1069)		
Mental accounting	-0.1458	0.150	-1.44
	(0.0907)		
Framing (Gains)	-0.0452	0.619	-0.50
	(0.0907)		
Framing (Losses)	-0.0543	0.548	-0.58
	(0.0821)		
Decoy	0.0480	0.559	0.58
	(0.0821)		
Social Norms	0.0983	0.274	1.09
	(0.0899)		
Default	0.2259**	0.038	2.07
	(0.1091)		
Observations	475		
Pseudo R^2	0.021		

Note. Dependent variable equals 1 if the subject has chosen the contract with renewable energy, 0.5 for the decoy contract, and zero otherwise. Model includes a constant. Marginal effects calculated at sample mean. Standard errors in parentheses. $***p < 0.01$, $**p < 0.05$, $*p < 0.1$.

Bibliography

Abito, **J. M.** (2012): “Welfare Gains from Optimal Pollution Regulation”, Job Market Paper.

Abrantes-Metz, **R. M.**, **S. B. Villas-Boas**, and **G. Judge** (2011): “Tracking the Libor rate”, *Applied Economics Letters*, 18(10), pp. 893–899.

Allcott, **H.** (2011): “Social norms and energy conservation”, *Journal of Public Economics*, 95(9), pp. 1082–1095.

Allcott, **H.** and **S. Mullainathan** (2010): “Behavioral science and energy policy”, *Science*, 327(5970), pp. 1204–1205.

Allcott, **H.** and **T. Rogers** (2012): *The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation*, tech. rep., National Bureau of Economic Research.

Almond, **D.**, **Y. Chen**, **M. Greenstone**, and **H. Li** (2009): “Winter Heating or Clean Air? Unintended Impacts of China’s Huai River Policy”, *American Economic Review (Papers and Proceedings)*, 99(2), pp. 184–190.

Andrews, **S. Q.** (2008a): “Inconsistencies in air quality metrics: ’Blue Sky’ days and PM10 concentrations in Beijing”, *Environmental Research Letters*, 3, pp. 1–14.

Andrews, **S. Q.** (2008b): “Seeing through the smog: Understanding the limits of Chinese air pollution reporting”, *China Environmental Series*, 10, pp. 5–32.

Antle, **J.**, **S. Capalbo**, **S. Mooney**, **E. Elliott**, and **K. Paustian** (2003): “Spatial heterogeneity, contract design, and the efficiency of carbon sequestration policies for agriculture”, *Journal of Environmental Economics and Management*, 46(2), pp. 231–250.

AP EnvEcon (2010): *GAINS Abatement Cost Guide. Version 1: A synthesis of information relating to the estimation, handling and use of abatement cost variables in the GAINS model framework*.

Ariely, **D.** (2009): *Predictably irrational, revised and expanded edition: The hidden forces that shape our decisions*, HarperCollins.

Ariely, **D.** and **T. S. Wallsten** (1995): “Seeking subjective dominance in multidimensional space: An explanation of the asymmetric dominance effect”, *Organizational Behavior and Human Decision Processes*, 63(3), pp. 223–232.

Arvola, **A.**, **M. Vassallo**, **M. Dean**, **P. Lampila**, **A. Saba**, **L. Lähteenmäki**, and **R. Shepherd** (2008): “Predicting intentions to purchase organic food: The role of affective and moral attitudes in the Theory of Planned Behaviour”, *Appetite*, 50(2), pp. 443–454.

Auffhammer, **M.** and **R. Kellogg** (2011): “Clearing the Air? The Effects of Gasoline Content Regulation on Air Quality”, *American Economic Review*, 101, pp. 2687–2722.

Becker, **R. A.** (2005): “Air pollution abatement costs under the Clean Air Act: evidence from the PACE survey”, *Journal of Environmental Economics and Management*, 50(1), pp. 144–169.

Becker, **R. A.** and **J. V. Henderson** (2001): “Cost of Air Quality Regulation”, *Behavioral and Distributional Effects of Environmental Policy*, ed. by **C. Carraro** and **G. E. Metcalf**, Chicago: University of Chicago Press, pp. 159–186.

Benford, **F.** (1938): “The Law of Anomalous Numbers”, *Proceedings of the American Philosophical Society*, 78(4), pp. 551–572.

Bertrand, **M.**, **E. Duflo**, and **S. Mullainathan** (2004): “How Much Should We Trust Differences-in-Differences Estimates?”, *Quarterly Journal of Economics*, 199(1), pp. 249–275.

Bird, **L.**, **R. Wüstenhagen**, and **J. Aabakken** (2002): “A review of international green power markets: recent experience, trends, and market drivers”, *Renewable and Sustainable Energy Reviews*, 6(6), pp. 513–536.

Bombardini, **M.** and **B. Li** (2016): “Trade, Pollution and Mortality in China”, Mimeograph.

Brown, **R.** (2005): “Benford’s Law and the screening of analytical data: the case of pollutant concentrations in ambient air”, *Analyst*, 130(9), pp. 1280–1285.

Camerer, **C.**, **S. Issacharoff**, **G. Loewenstein**, **T. O’Donoghue**, and **M. Rabin** (2003): “Regulation for Conservatives: Behavioural Economics and the Case for ’Asymmetric Paternalism’”, *University of Pennsylvania Law Review*, 151, pp. 1211–1254.

Cameron, **A. C.**, **J. B. Gelbach**, and **D. L. Miller** (2008): “Bootstrap-Based Improvements for Inference with Clustered Errors”, *Review of Economics and Statistics*, 90(3), pp. 414–427.

Carlson, **C.**, **D. Burtraw**, **M. Cropper**, and **K. L. Palmer** (2000): “Sulfur Dioxide Control by Electric Utilities: What Are the Gains from Trade?”, *Journal of Political Economy*, 108(6), pp. 1292–1326.

Cesnulyte, **V.**, **A. V. Lindfors**, **M. R. A. Pitkanen**, **K. E. J. Lehtinen**, **J. J. Morcrette**, and **A. Arola** (2014): “Comparing ECMWF AOD with AERONET observations at visible and UV wavelengths", *Atmospheric Chemistry and Physics*, 14(2), pp. 593–608.

Chang, **T.**, **J. Graff Zivin**, **T. Gross**, and **M. Neidell** (2016): "The Effect of Pollution on Worker Productivity: Evidence from Call-Center Workers in China", NBER Working Paper No. 22328.

Chay, **K. Y.** and **M. Greenstone** (2005): "Does Air Quality Matter? Evidence from the Housing Market", *Journal of Political Economy*, 113(2), pp. 376–424.

Chay, **K.**, **C. Dobkin**, and **M. Greenstone** (2003): "The Clean Air Act of 1970 and adult mortality", *Journal of Risk and Uncertainty*, 27(3), pp. 279–300.

Chen, **L.** (2014): "Mayor of Beijing Told to Fix the Smog or 'Bring Your Head'", *Epoch Times*.

Chen, **Y.**, **H. Li**, and **L.-A. Zhou** (2005): "Relative performance evaluation and the turnover of provincial leaders in China", *Economics Letters*, 88(3), pp. 421–425.

Chen, **Y.**, **A. Ebenstein**, **M. Greenstone**, and **H. 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*, 110(32), pp. 12936–12941.

Chen, **Y.**, **G. Z. Jin**, **N. Kumar**, and **G. Shi** (2012): "Gaming in Air Pollution Data? Lessons from China", *B.E. Journal of Economic Analysis & Policy*, 12(3).

— (2013): "The promise of Beijing: Evaluating the impact of the 2008 Olympic Games on air quality", *Journal of Environmental Economics and Management*, 66, pp. 424–443.

Costa, **D. L.** and **M. E. Kahn** (2013): "Energy conservation "nudge" and environmentalist ideology: Evidence from a randomized residential electricity field experiment", *Journal of the European Economic Association*, 11(3), pp. 680–702.

Cyranoski, **D.** (2009): "Beijing's clean air claims questioned", *Nature*.

De Marchi, **S.** and **J. T. Hamilton** (2006): "Assessing the accuracy of self-reported data: an evaluation of the toxics release inventory", *Journal of Risk and Uncertainty*, 32(1), pp. 57–76.

Duflo, **E.**, **M. Greenstone**, **R. Pande**, and **N. Ryan** (2013): "Truth-telling by Third-party Auditors and the Response of Polluting Firms: Experimental Evidence from India", *Quarterly Journal of Economics*, 128(4), pp. 1499–1545.

Dumas, **C. F.** and **J. H. Devine** (2000): "Detecting Evidence of Non-Compliance In Self-Reported Pollution Emissions Data: An Application of Benford's Law", Tampa, FL.

Eck, **T. F.**, **B. N. Holben**, **J. S. Reid**, **O. Dubovik**, **A. Smirnov**, **N. T. O'Neill**, **I. Slutsker**, and **S. Kinne** (1999): "Wavelength dependence of the optical depth of biomass burning, urban, and desert dust aerosols", *Journal of Geophysical Research*, 104(D24), pp. 31333–31349.

Fioletov, **V. E.**, **C. A. McLinden**, **N. Krotkov**, and **C. Li** (2015): "Lifetimes and emissions of SO2 from point sources estimated from OMI", *Geophysical Research Letters*, 42(6), pp. 1969–1976.

Fouquet, **R.** (2011): "Long run trends in energy-related external costs", *Ecological Economics*, 70(12), pp. 2380–2389.

Fridley, **D.**, **J. Romankiewicz**, and **C. Fino-Chen** (2013): *China Energy Databook Version 8.0*, Lawrence Berkeley National Laboratory.

Gao, **C.**, **H. Yin**, **N. Ai**, and **Z. Huang** (2009): "Historical Analysis of SO2 Pollution Control Policies in China", *Environmental Management*, 43, pp. 447–457.

Gennaioli, **N.** and **A. Shleifer** (2010): "What comes to mind", *The Quarterly Journal of Economics*, 125(4), pp. 1399–1433.

Ghanem, **D.** and **J. Zhang** (2014): "'Effortless Perfection': Do Chinese cities manipulate air pollution data?", *Journal of Environmental Economics and Management*, 68(2), pp. 203–225.

Gollop, **F. M.** and **M. J. Roberts** (1985): "Cost-Minimizing Regulation of Sulfur Emissions: Regional Gains in Electric Power", *Review of Economics and Statistics*, 67(1), p. 81.

Gosnell, **G. K.**, **J. A. List**, and **R. Metcalfe** (2016): "A New Approach to an Age-Old Problem: Solving Externalities by Incenting Workers Directly", NBER Working Paper No. 22316.

Grammaticas, **D.** (2011): "China's pollution, a toxic issue", *BBC*.

Greenstone, **M.** (2004): "Did the Clean Air Act cause the remarkable decline in sulfur dioxide concentrations?", *Journal of Environmental Economics and Management*, 47(3), pp. 585–611.

— (2016): *The Global Energy Challenge*, 36th Barcelona GSE Lecture, delivered in Barcelona on 17th November 2016.

Greenstone, **M.** and **R. Hanna** (2014): "Environmental Regulations, Air and Water Pollution, and Infant Mortality in India", *American Economic Review*, 104(10), pp. 3038–3072.

Grossman, **G.** and **A. Krueger** (1995): "Economic Growth and the Environment", *Quarterly Journal of Economics*, 110(2), pp. 353–377.

GTZ (2009): "Air Quality Management", *Sustainable Transport: A Sourcebook for Policy-Makers in Developing Cities*, Eschborn, Germany: Gesellschaft fuer Technische Zusammenarbeit.

Guan, **D.**, **Z. Liu**, **Y. Geng**, **S. Lindner**, and **K. Hubacek** (2012): "The gigatonne gap in China's carbon dioxide inventories", *Nature Climate Change*, 2(9), pp. 672–675.

Hammitt, **J. K.** and **L. A. Robinson** (2011): "The Income Elasticity of the Value per Statistical Life: Transferring Estimates between High and Low Income Populations", *Journal of Benefit-Cost Analysis*, 2(1), pp. 1–27.

Hansman, **C.**, **J. Hjort**, and **G. Leon** (2015): "Interlinked Firms and the Consequences of Piecemeal Regulation", Mimeograph.

Hao, **J.**, **K. He**, **L. Duan**, **J. Li**, and **L. Wang** (2007): "Air pollution and its control in China", *Frontiers of Environmental Science & Engineering in China*, 1(2), pp. 129–142.

Hartman, **R. S.**, **D. Wheeler**, and **M. Singh** (1997): "The cost of air pollution abatement", *Applied Economics*, 29(6), pp. 759–774.

He, **H.** (2012): "Air Pollutant Concentrations and Trends over the Eastern U.S. and China: Aircraft Measurements and Numerical Simulations", Doctoral Dissertation, University of Maryland, College Park.

Heeter, **J.** and **T. Nicholas** (2013): "Status and Trends in the US Voluntary Green Power Market (2012 Data)".

Henderson, **V.** (1996): "Effects of Air Quality Regulation", *American Economic Review*, 86(4), pp. 789–813.

Herbers, **C.** and **I. Ramme** (2014): "Online marketing of green electritiy in Germany - A content analysis of providers' websites", *Energy Policy*, 66, pp. 257–266.

Holben, **B. N.**, **T. F. Eck**, **I. Slutsker**, **A. Smirnov**, **A. Sinyuk**, **J. Schafer**, **D. Giles**, and **O. Dubovik** (2006): "AERONET's version 2.0 quality assurance criteria", *Proceedings of the SPIE, the International Society for Optical Engineering*, vol. 14, 64080Q.1–64080Q.14.

Huber, **J.**, **J. W. Payne**, and **C. Puto** (1982): "Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis", *Journal of Consumer Research*, pp. 90–98.

Huck, **S.** and **I. Rasul** (2007): "Comparing charitable fundraising schemes: Evidence from a natural field experiment", *Manuscript, October*.

Hughes, **V.** (2012): "Where there's smoke", *Nature*, 489(7417), S18–S20.

IIASA (2010a): "GAINS Asia: A Tool to Combat Air Pollution and Climate Change Simultaneously".

— (2010b): "GAINS Asia: Scenarios for Cost-Effective Control of Air Pollution and Greenhouse Gases".

IPCC (2007): *Climate Change 2007-Mitigation of Climate Change: Working Group III Contribution to the Fourth Assessment Report of the IPCC*, vol. 4, Cambridge University Press.

Ito, **K.** and **S. Zhang** (2016): “Willingness to Pay for Clean Air: Evidence from Air Purifier Markets”, NBER Working Paper No. 22367.

Jayachandran, **S.** (2009): “Air quality and early-life mortality evidence from Indonesia’s wildfires”, *Journal of Human Resources*, 44(4), pp. 916–954.

Jia, **R.** (2014): “Pollution for Promotion”, Job Market Paper.

Johnson, **E.** and **D. Goldstein** (2003): “Do defaults save lives?”, *Science*, 302, pp. 1338–1339.

Kaenzig, **J.**, **S. L. Heinzle**, and **R. Wüstenhagen** (2013): “Whatever the consumer wants, the consumer gets? Exploring the gap between consumer preferences and default electricity products in Germany”, *Energy Policy*, 53, pp. 311–322.

Kahn, **M. E.** (2006): *Green Cities: Urban Growth and the Environment*, Washington D.C.: Brookings Institution Press.

Kahn, **M. E.**, **P. Li**, and **D. Zhao** (2015): “Water Pollution Progress at Borders: The Role of Changes in China’s Political Promotion Incentives”, *American Economic Journal: Economic Policy*, 7(4), pp. 223–242.

Kahneman, **D.** and **S. Frederick** (2005): “A model of heuristic judgment”, *The Cambridge Handbook of Thinking and Reasoning*, pp. 267–293.

Kahneman, **D.** and **A. Tversky** (1979): “Prospect theory: An analysis of decision under risk”, *Econometrica: Journal of the Econometric Society*, pp. 263–291.

Keller, **W.** and **A. Levinson** (2002): “Pollution abatement costs and foreign direct investment inflows to US states”, *Review of Economics and Statistics*, 84(4), pp. 691–703.

Klimont, **Z.** et al. (2009): “Projections of SO2 , NOx and carbonaceous aerosols emissions in Asia”, *Tellus B*, 61(4), pp. 602–617.

Krotkov, **N. A.** et al. (2016): “Aura OMI observations of regional SO2 and NO2 pollution changes from 2005 to 2015”, *Atmospheric Chemistry and Physics*, 16(7), pp. 4605–4629.

Kumar, **R.** and **A. E. Joseph** (2006): “Air pollution concentrations of PM2. 5, PM10 and NO2 at ambient and kerbsite and their correlation in Metro City-Mumbai”, *Environmental Monitoring and Assessment*, 119(1-3), pp. 191–199.

Laffont, **J.-J.** (2005): *Regulation and Development*, Cambridge, UK: Cambridge University Press.

Le Monde (2013): “Chine: la pollution de l’air pousse Pékin ’à accepter plus de transparence’”.

Li, **H.** and **L.-A. Zhou** (2005): “Political turnover and economic performance: the incentive role of personnel control in China”, *Journal of Public Economics*, 89(9-10), pp. 1743–1762.

Li, **J.**, **B. E. Carlson**, and **A. A. Lacis** (2015): “How well do satellite AOD observations represent the spatial and temporal variability of PM 2.5 concentration for the United States?”, *Atmospheric Environment*, 102, pp. 260–273.

Li, **X.**, **X. Wu**, and **F. Zhang** (2015): “A method for analyzing pollution control policies: Application to SO2 emissions in China”, *Energy Economics*, 49, pp. 451–459.

Lillemo, **S. C.** (2014): “Measuring the effect of procrastination and environmental awareness on households’ energy-saving behaviours: An empirical approach”, *Energy Policy*, 66, pp. 249–256.

Lin, **J.** (2007): “Energy conservation investments: A comparison between China and the US”, *Energy Policy*, 35, pp. 916–924.

List, **J. A.** and **D. Lucking-Reiley** (2002): “The effects of seed money and refunds on charitable giving: Experimental evidence from a university capital campaign”, *Journal of Political Economy*, 110(1), pp. 215–233.

Liu, **J.** and **H. Yang** (2009): “China Fights Against Statistical Corruption”, *Science*, 325(5941), pp. 675–676.

Lu, **Z.**, **D. G. Streets**, **Q. Zhang**, **S. Wang**, **G. R. Carmichael**, **Y. F. Cheng**, **C. Wei**, **M. Chin**, **T. Diehl**, and **Q. Tan** (2010): “Sulfur dioxide emissions in China and sulfur trends in East Asia since 2000”, *Atmospheric Chemistry and Physics*, 10(13), pp. 6311–6331.

Marcazzan, **G. M.**, **S. Vaccaro**, **G. Valli**, and **R. Vecchi** (2001): “Characterisation of PM10 and PM2. 5 particulate matter in the ambient air of Milan (Italy)”, *Atmospheric Environment*, 35(27), pp. 4639–4650.

Martinez-Bravo, **M.**, **G. Padro i Miquel**, **N. Qian**, and **Y. Yao** (2013): “Political Reform in China: Elections, Public Goods and Income Distribution”, Mimeograph.

Matus, **K.**, **K.-M. Nam**, **N. E. Selin**, **L. N. Lamsal**, **J. M. Reilly**, and **S. Paltsev** (2012): “Health damages from air pollution in China”, *Global Environmental Change*, 22(1), pp. 55–66.

Mazar, **N.** and **D. Ariely** (2006): “Dishonesty in everyday life and its policy implications”, *Journal of Public Policy & Marketing*, pp. 117–126.

Mazar, **N.** and **C.-B. Zhong** (2010): “Do green products make us better people?”, *Psychological Science*, 21(4), pp. 494–498.

McElwee, **C. R.** (2011): *Environmental Law in China*, Oxford; New York: Oxford University Press.

Meng, **K. C.** (Forthcoming): “Using a Free Permit Rule to Forecast the Marginal Abatement Cost of Proposed Climate Policy”, *American Economic Review*.

Michalski, **T.** and **G. Stoltz** (2013): "Do Countries Falsify Economic Data Strategically? Some Evidence That They Might", *Review of Economics and Statistics*, 95(2), pp. 591–616.

Ministry of Environmental Protection of the People's Republic of China (2008a): "Issuance of Automatic Pollution Source Monitoring", Original in Chinese.

— (2008b): "Management Measures for the Operation of Automatic Monitoring Installations for Pollution Sources", Original in Chinese.

Ministry of Environmental Protection of the People's Republic of China (2012): *Air quality monitoring of the implementation of the first phase of the new standard task is completed in Beijing, Tianjin and other regions and 74 cities will take effect on January 1, 2013, to carry out surveillance*, tech. rep., Original in Chinese.

Montero, **J.-P.**, **J. M. Sanchez**, and **R. Katz** (2000): "A Market-Based Environmental Policy Experiment in Chile", MIT CEEPR Working Paper 00-005.

Morrow, **J.** (2014): "Benford's Law, families of distributions and a test basis", CEP Discussion Paper No. 1291.

Morwitz, **V. G.**, **E. Johnson**, and **D. Schmittlein** (1993): "Does measuring intent change behavior?", *Journal of Consumer Research*, pp. 46–61.

Moser, **A.** (2013): "The Yangzonghai case: struggling for environmental justice", *China and the Environment. The Green Revolution*, ed. by **S. Geal**, London: Zed Books, pp. 96–135.

NASA (2014): "OMSO2 Readme File v1.2.0 Released Feb 26, 2008 Updated: September 26, 2014".

Newell, **R. G.** and **R. N. Stavins** (2003): "Cost heterogeneity and the potential savings from market-based policies", *Journal of Regulatory Economics*, 23(1), pp. 43–59.

Nordhaus, **W. D.** (2013): *The Climate Casino: Risk, Uncertainty, and Economics for a Warming World*, Yale University Press.

Oates, **W. E.**, **P. R. Portney**, and **A. M. McGartland** (1989): "The net benefits of incentive-based regulation: a case study of environmental standard setting", *American Economic Review*, 79(5), pp. 1233–1242.

O'Donoghue, **T.** and **M. Rabin** (1999): "Doing it now or later", *American Economic Review*, pp. 103–124.

Oliva, **P.** (2015): "Environmental Regulations and Corruption: Automobile Emissions in Mexico City", *Journal of Political Economy*, 123(3), pp. 686–724.

Persson, **P.** and **E. Zhuravskaya** (2016): "The Limits of Career Concerns in Federalism: Evidence from China", *Journal of the European Economic Association*, 14(2), pp. 338–374.

Pichert, **D.** and **K. V. Katsikopoulos** (2008): “Green defaults: Information presentation and pro-environmental behaviour”, *Journal of Environmental Psychology*, 28(1), pp. 63–73.

Purohit, **P.**, **M. Amann**, **R. Mathur**, **I. Gupta**, **S. Marwah**, **V. Verma**, **I. Bertok**, **J. Borken-Kleefeld**, **A. Chambers**, **J. Cofala**, et al. (2010): “Gains Asia: Scenarios for cost-effective control of air pollution and greenhouse gases in India”.

Rubinstein, **A.** (2012): *Lecture notes in microeconomic theory: the economic agent*, Princeton University Press.

Schmale, **J.**, **D. Shindell**, **E. van Schneidemesser**, **I. Chabay**, and **M. Lawrence** (2014): “Clean up our skies”, *Nature*, 515, pp. 335–337.

Schmalensee, **R.**, **P. L. Joskow**, **A. D. Ellerman**, **J. P. Montero**, and **E. M. Bailey** (1998): “An Interim Evaluation of Sulfur Dioxide Emissions Trading”, *Journal of Economic Perspectives*, 12(3), pp. 53–68.

Schreifels, **J. J.**, **Y. Fu**, and **E. J. Wilson** (2012): “Sulfur dioxide control in China”, *Energy Policy*, 48(2012), pp. 779–789.

Schultze, **C. R.** (1977): *The Public Use of Private Interest*, Washington D.C.: The Brookings Institution.

Shafiei, **S.** and **R. A. Salim** (2014): “Non-renewable and renewable energy consumption and CO2 emissions in OECD countries: A comparative analysis”, *Energy Policy*, 66, pp. 547–556.

Siegel, **S.** (1956): *Nonparametric Statistics for the Behavioral Sciences*, New York: McGraw-Hill Book Company, Inc.

Simonich, **S. L. M.** (2009): “Response to Comments on "Atmospheric Particulate Matter Pollution during the 2008 Beijing Olympics"”, *Environmental Science & Technology*, 43, pp. 7590–7591.

Smirnov, **A.**, **B. N. Holben**, **T. F. Eck**, **O. Dubovik**, and **I. Slutsker** (2000): “Cloud-Screening and Quality Control Algorithms for the AERONET Database”, *Remote Sensing of Environment*, 73(3), pp. 337–349.

Song, **X.-l.**, **L.-y. Zhang**, **M.-j. Gong**, **X. Wang**, and **Y. Qi** (2015): “China’s Emission Data Accounting System of the Total Emission Control During the 11th Five-Year Plan Period”, *China Population, Resources and Environment*, 25(1), pp. 46–54.

Spiegler, **R.** (2011): *Bounded rationality and industrial organization*, Oxford University Press.

State Council (2006): “Major Pollutant Emissions Control in the 11th Five-Year Plan Period”, Original in Chinese.

— (2007): “State Council approved the transfer of energy - saving emission reduction statistics. Monitoring and assessment of the implementation of programs and methods of notification”, Original in Chinese.

Stavins, **R. N.** (1998): “What Can We Lean from the Grand Policy Experiment? Lessons from SO2 Allowance Trading”, *Journal of Economic Perspectives*, 12(3), pp. 69–88.

— (2003): “Experience With Market-Based Environmental Policy Instruments”, *Handbook of Environmental Economics*, ed. by **K.-G. Mäler** and **J. Vincent**, vol. Volume 1, Elsevier Science B.V.

Stern, **N.** (2006): “Review on the economics of climate change”, *London HM Treasury*.

Stoerk, **T.** (2016): “Statistical corruption in Beijing’s air quality data has likely ended in 2012”, *Atmospheric Environment*, 127, pp. 365–371.

Sunstein, **C. R.** (2003): *Why societies need dissent*, Harvard University Press.

Sunstein, **C. R.** and **L. Reisch** (2013): “Automatically green: Behavioral economics and environmental protection”, *Available at SSRN 2245657*.

Tang, **X.**, **M. Shao**, **M. Hu**, **Z. Wang**, and **J. Zhang** (2009): “Comment on "Atmospheric Particulate Matter Pollution during the 2008 Beijing Olympics"”, *Environmental Science & Technology*, 43, p. 7588.

Taubinsky, **D.** (2013): “From Intentions to Actions: A Model and Experimental Evidence of Inattentive Choice”, *Working Paper*.

Thaler, **R. H.** (2004): *Mental accounting matters*, Russell Sage Foundation. Princeton, NJ: Princeton University Press.

Thaler, **R. H.** and **C. R. Sunstein** (2003): “Libertarian paternalism”, *The American Economic Review*, 93(2), pp. 175–179.

— (2008): *Nudge: Improving decisions about health, wealth, and happiness*, Yale University Press.

The Economist (2013): “The East is grey”, *The Economist*, (Aug 10th 2013).

Tu, **Z.** and **R. Shen** (2014): “Can China’s Industrial SO2 Emissions Trading Pilot Scheme Reduce Pollution Abatement Costs?”, *Sustainability*, 6(11), pp. 7621–7645.

Tversky, **A.** and **D. Kahneman** (1974): “Judgment under uncertainty: Heuristics and biases”, *Science*, 185(4157), pp. 1124–1131.

— (1981): “The Framing of Decisions and the Psychology of Choice”, *Science*, 211, p. 30.

Varian, **H.** (1972): “Benford’s law”, *American Statisticiain*, 26(3), p. 65.

Vermeir, **I.** and **W. Verbeke** (2006): “Sustainable food consumption: Exploring the consumer "attitude–behavioral intention" gap”, *Journal of Agricultural and Environmental Ethics*, 19(2), pp. 169–194.

— (2008): “Sustainable food consumption among young adults in Belgium: Theory of planned behaviour and the role of confidence and values”, *Ecological Economics*, 64(3), pp. 542–553.

Wang, **A. L.** (2013): "The search for sustainable legitimacy: Environmental law and bureaucracy in China", *Harvard Environmental Law Review*, 37, pp. 367–440.

Wang, **J.-F.**, **M.-G. Hu**, **C.-D. Xu**, **G. Christakos**, and **Y. Zhao** (2013): "Estimation of Citywide Air Pollution in Beijing", *PLOS One*, 8(1), pp. 1–6.

Wang, **L.**, **C. Jang**, et al. (2010a): "Assessment of air quality benefits from national air pollution control policies in China. Part I: Background, emission scenarios and evaluation of meteorological predictions", *Atmospheric Environment*, 44(28), pp. 3442–3448.

— (2010b): "Assessment of air quality benefits from national air pollution control policies in China. Part II: Evaluation of air quality predictions and air quality benefits assessment", *Atmospheric Environment*, 44(28), pp. 3449–3457.

Wang, **W.**, **T. Primbs**, **S. Tao**, and **S. L. M. Simonich** (2009): "Atmospheric particulate matter pollution during the 2008 Beijing Olympics", *Environmental Science & Technology*, 43(14), pp. 5314–5320.

Wang, **X.**, **D. Westerdahl**, **L. C. Chen**, **Y. Wu**, **J. Hao**, **X. Pan**, **X. Guo**, and **K. M. Zhang** (2009): "Evaluating the air quality impacts of the 2008 Beijing Olympic Games: On-road emission factors and black carbon profiles", *Atmospheric Environment*, 43(30), pp. 4535–4543.

Watts, **J.** (2011): *When a Billion Chinese Jump*, London: Faber and Faber Ltd.

WHO (2014): *7 million premature deaths annually linked to air pollution*.

Wong, **E.** (2013): "On Scale of 0 to 500, Beijing's Air Quality Tops 'Crazy Bad' at 755", *New York Times*.

World Bank and **State Environmental Protection Administration** (2007): *Cost of Pollution in China: Economic Estimates of Physical Damages*, Conference Edition.

Xiao, **H.-W.**, **A. Mujumdar**, and **L. Che** (2015): "Pollution: uncouple from economy boom", *Nature*, 517, p. 145.

Xu, **Y.** (2011): "The use of a goal for SO2 mitigation planning and management in China's 11th Five-Year Plan", *Journal of Environmental Planning and Management*, 54(6), pp. 769–783.

Xue, **Y.**, **X. He**, **H. Xu**, **J. Guang**, **J. Guo**, and **L. Mei** (2014): "China Collection 2.0: The aerosol optical depth dataset from the synergetic retrieval of aerosol properties algorithm", *Atmospheric Environment*, 95, pp. 45–58.

Yang, **G.**, **Y. Wang**, **Y. Zeng**, **G. F. Gao**, **X. Liang**, **M. Zhou**, **X. Wan**, **S. Yu**, **Y. Jiang**, **M. Naghavi**, et al. (2013): "Rapid health transition in China, 1990-2010: findings from the Global Burden of Disease Study 2010", *The Lancet*, 381(9882), pp. 1987–2015.

Yao, **X.**, **X. Xu**, **K. Sabaliauskas**, and **M. Fang** (2009): "Comment on "Atmospheric Particulate Matter Pollution during the 2008 Beijing Olympics"", *Environmental Science & Technology*, 43, p. 7589.

Zhang, **J. P.** et al. (2012): "The impact of circulation patterns on regional transport pathways and air quality over Beijing and its surroundings", *Atmospheric Chemistry and Physics*, 12(11), pp. 5031–5053.

Zheng, **S.** and **M. E. Kahn** (2013): "Understanding China's Urban Pollution Dynamics", *Journal of Economic Literature*, 51(3), pp. 731–772.

Zheng, **S.**, **C. Sun**, **Y. Qi**, and **M. E. Kahn** (2014): "The Evolving Geography of China's Industrial Production: Implications for Pollution Dynamics and Urban Quality of Life", *Journal of Economic Surveys*, 28(4), pp. 709–724.