Journal of Public Economics 164 (2018) 70–90

Contents lists available at ScienceDirect

Journal of Public Economics

journal homepage: www.elsevier.com/locate/jpube

Does central supervision enhance local environmental enforcement? Quasi-experimental evidence from China

Bing Zhang[a], Xiaolan Chen[b], Huanxiu Guo[c,*]

[a] State Key Laboratory of Pollution Control & Resource Reuse, School of Environment, Nanjing University, Nanjing, China
[b] School of Economics, Sichuan University, Chengdu, China
[c] Institute of Economics and Finance, Nanjing Audit University, Nanjing, China

ARTICLE INFO

Article history:
Received 20 September 2016
Received in revised form 7 February 2018
Accepted 15 May 2018
Available online 7 June 2018

JEL classification:
D22
Q28
Q58
O53

Keywords:
Environmental federalism
Central supervision
Water pollution
Regression discontinuity design
China

ABSTRACT

This paper draws on a natural experiment generated by the National Specially Monitored Firms (NSMF) program in China to evaluate the effectiveness of central supervision at improving local environmental enforcement. We explore a unique firm-level Chinese Environmental Statistics dataset and utilize a regression discontinuity design to assess the impact of central supervision through the NSMF program on an industrial firm's chemical oxygen demand (COD) emissions. The results suggest that central supervision significantly reduces industrial COD emissions by at least 26.8%. These results highlight the substantial room for improvement in Chinese environmental regulations via central supervision. A more flexible environmental decentralization regime and comprehensive central supervision are thus recommended for future reforms.

1. Introduction

Countries with multiple levels of government face the important question of which level of government should undertake specific regulatory responsibilities (Oates, 2001). In this paper, we propose a new model of decentralized environmental regulation whereby the national government plays a key role in information collection and supervision to complement local environmental regulation. We provide strong empirical evidence by evaluating the National Specially Monitored Firms pilot program (hereafter called the NSMF program), which aimed to reform environmental decentralization in China.

Although decentralization is a mainstream political arrangement for the provision and governance of public goods in most countries (WorldBank, 2000), the decentralization of environmental regulation remains a topic of debate among researchers and policymakers. Many economists advocate decentralization as a more efficient method of providing local public goods. For instance, Tiebout (1956) argues that in a decentralized context, interjurisdictional competition arises and improves the allocation of local public services because voters can choose their preferred localities by "voting with their feet." Oates (1972) further proposes a theoretical framework of interjurisdictional competition and proffers the Decentralization Theorem (Oates, 1972, p.54). In the context of environmental management, a decentralized administration system for environmental regulation is known as environmental federalism (Anderson and Hill, 1997). Critical opinions posit that environmental federalism loses efficiency due to a number of externalities such as free-riding behavior and cross-boundary environmental pollution (Engel, 1996), competition between localities for mobile and polluting capital (Kunce and Shogren, 2005; Levinson, 1997; Markusen et al., 1995; McAusland, 2003), misestimated environmental costs and benefits at the local level (Revesz, 1997), and corruption (Fan et al., 2009).

Empirical evidence on environmental federalism has primarily been based on the experiences of developed countries. On the one hand, an abundance of studies suggest that competition between

* Corresponding author.
E-mail addresses: zhangb@nju.edu.cn (B. Zhang), chenxiaolan@scu.edu.cn (X. Chen), hxguo@nau.edu.cn (H. Guo).

https://doi.org/10.1016/j.jpubeco.2018.05.009
0047-2727/

localities may lead to a "race to the top" in environmental standards rather than a "race to the bottom" (Fredriksson and Millimet, 2002; Konisky, 2007; Levinson, 2003; List and Gerking, 2000). On the other hand, numerous studies provide evidence for interjurisdictional externalities and freeriding behavior (Fredriksson et al., 2006; Gray and Shadbegian, 2004; Konisky and Woods, 2012; Sigman, 2001, 2005). More recently, beyond the debates on the pros and cons of environmental federalism, increasing attention has been paid to alternative policy designs to improve the performance of environmental federalism (see Millimet, 2014 for a review).

China has adopted a different environmental regulation regime from most developed countries. While constitutionally organized as a unitary sovereign, China has a complex system of formal and informal divisions of authority between the central government and various levels of local and regional governments. A number of studies on Chinese central-local relations recognize that the decentralization resulting from economic reforms starting in 1978 has endowed China's local governments with substantial authority over local economic development (Huang, 1996; Jin et al., 2005; White and Landry, 2010). Following this national decentralization trend, China's environmental regulation framework was elaborated at the beginning of the reform era. Although the central government maintains its political authority over environmental planning, most fundamental enforcement decision making and responsibilities have been allocated to local and regional governments. This pragmatic regime is thus characterized by de facto environmental decentralization, i.e., a combination of the central design and local enforcement of environmental regulations (Zheng, 2007).

Previous studies suggest that in the framework of environmental decentralization, local enforcement remains far from adequate and effective in China (Lo et al., 2006; Van Rooij, 2006; Wang et al., 2003). This failure is primarily because local governments can take advantage of enforcement discretion to protect polluting firms (Jia and Nie, 2017; Wang et al., 2003), which may compromise environmental protection in favor of local economic interests (Dasgupta et al., 1997). Moreover, the severe lack of resources and the limited monitoring and inspection capacity of local regulators hamper the effectiveness of local environmental enforcement (Van Rooij, 2006). Ongoing reforms to environmental decentralization have sought to remedy these perverse incentives and enhance local environmental enforcement. Although the issue of environmental decentralization has garnered growing interest, few studies to date have investigated the impact of these reforms.

To bridge this gap in the literature, we study the reform of environmental decentralization in China under the NSMF program. The program's essence is to enhance the central government's role in information collection and supervision by means of automatic real-time monitoring and frequent inspections while maintaining environmental enforcement at the local level. It thus entails a combination of central supervision and environmental decentralization to assess the extent to which the former can enhance local environmental enforcement. In a framework consisting of a principal-agent model with asymmetric information, we first demonstrate that given environmental enforcement at the local level, direct central supervision reduces asymmetric information at both central and local levels and thus reduce firms' optimal pollution emissions by increasing local environmental enforcement. As empirical evidence, we collect detailed water pollution data for 20,607 industrial polluting firms from an administrative dataset of Chinese Environmental Statistics (CES) and set a threshold of 65% of total water pollution emissions, which determines the designation of NSM firms, to conduct a regression discontinuity (RD) assessment design.

Given the alarming extent of environmental destruction in China, our study is urgently needed to clarify the role of the central government and to enlighten future reforms of the country's environmental decentralization. Moreover, China's case is globally relevant as it broadens the geographic scope of the literature on environmental federalism beyond the US and EU models (Vogel et al., 2010). Unlike previous macro-level studies (Fredriksson and Wollscheid, 2014; Sigman, 2007; Sjöberg, 2016), our study is the first to use firm-level data to provide microeconomic evidence on environmental decentralization outcomes. In terms of its methodological contribution, our research uses quasi-experimental methods to address concerns regarding endogenous environmental regulation and provides robust empirical evidence in response. We build on the growing literature on environmental policy evaluation taking an experimental or quasi-experimental approach, which offers credible evidence that can be used in novel policy designs for the effective implementation of environmental regulations in developing countries (Duflo et al., 2013, 2014; Hanna and Oliva, 2010).

Our results show that direct central supervision has had a substantial environmental impact in the short-term by reducing industrial water pollution by at least 26.8%. This reduction is primarily achieved through end-of-pipe treatment at the firm level without affecting production. Our results highlight the role of central supervision in enhancing local environmental enforcement and suggest that more flexible environmental decentralization whereby comprehensive central supervision complements local environmental regulation should be at the core of future reforms of environmental decentralization in China.

The rest of the paper is organized as follows. Section 2 provides background information on China's decentralized system of environmental regulation and the reforms in the NSMF program. Section 3 introduces our theoretical model. Section 4 discusses the RD design of our evaluation. Section 5 introduces the dataset, and Section 6 presents the main results. Finally, Section 7 discusses policy implications and concludes the paper.

2. Background

2.1. Decentralized system of environmental regulation in China

China first elaborated an environmental regulation system at the beginning of the reform era (OECD, 2006). In 1978, the National People's Congress (NPC) added Article 11, Section 1 to the Chinese constitution, stating that "the state protects and improves the living environment and the ecological environment and prevents and controls pollution and other public hazards." On the ground, the Environmental Protection Law (EPL) was passed the same year, requiring the central government as well as local and regional governments at all levels (provincial, prefectural, county and township) to establish environmental institutions. The EPL (1989) stipulated that "the local people's governments at various levels shall be responsible for the environment quality of areas under their jurisdiction and take measures to improve the environment quality."[1] This fundamental environmental legislation thereby provides a legal framework for the decentralized system of environmental regulation in China.

Since the reform era, as environmental damage has become increasingly severe, Chinese environmental institutions have undergone significant expansion. At the central level, the Environmental Protection Leadership Group (EPLG) of the State Council was upgraded from a department within a ministry to the State Environmental Protection Administration (SEPA) in 1998 and finally to the Ministry of Environmental Protection (MEP) in 2008. This gradual increase in ministerial status endowed greater powers

[1] See Article 16, Chapter 3 of the EPL (1989) for more details. http://www.lawinfochina.com/display.aspx?id=1208&lib=law&SearchKeyword=Environmental%20Protection%252%200Law&SearchCKeyword=. (Last consulted 29 Feb 2016).

to environmental institutions. At the local level, a regime of jurisdictional or territorial management was also introduced and promoted to hold local governments responsible for environmental regulation within their jurisdictions. The main principle of this regime is to hold local governments accountable for the implementation of central environmental policies. In practice, over 3000 Environmental Protection Bureaus (EPB) with more than 50,000 staff members have been established and subordinated to local governments at all levels. These EPBs are primary environmental regulators that conduct daily regulatory activities such as monitoring and inspecting polluting firms, analyzing environmental complaints, suing for environmental damage and enforcing sanctions, all of which determine the stringency of environmental regulations at the local level. Unlike in the US, where regional environmental enforcement is under the direct control of the federal Environmental Protection Agency (EPA), Chinese local EPBs respond directly to local governments in addition to the environmental administration hierarchy. The MEP only provides guidance to provincial and sub-provincial regulatory administrations (Yang, 2017). Considering that the environmental policies instituted by the MEP are generally vague and aspirational, local governments are endowed considerable discretion over environmental regulation.

2.2. Imperfect local enforcement of environmental regulations

However, the decentralization of environmental regulation did not translate into improved environmental enforcement as intended (Lo et al., 2006; Wang et al., 2003). Instead, conflicts of interests between central and local governments emerged. Driven by the previous system of cadre evaluation,[2] local officials tended to prioritize economic growth at the expense of the environment (Golding, 2011). Large polluting firms such as state-owned petroleum and petrochemical enterprises, which generate substantial tax revenue and employment for the local economy, often have significant bargaining power *vis-à-vis* local governments, enabling them to influence environmental regulations through personal connections, bribes, or favors and obtain exemptions from sanctions for violations (Jia and Nie, 2017; Wang et al., 2003). Another severe problem that hampers local environmental enforcement is a lack of resources. In the decentralized system, EPBs depend primarily on local government for financial resources, the rest of which stem from the collection of pollution levies from firms (Ma and Ortolano, 2000). According to a survey conducted in Sichuan and Yunnan, local EPBs face chronic shortages of staff and materials, which prevent them from implementing regular and reactive monitoring and inspection activities for all polluting firms (Van Rooij, 2006).[3]

Due to conflicting interests, collusion and limited resources, Chinese environmental decentralization is a typical principal-agent problem with asymmetric information at both central and local levels. Imperfect information on pollution emissions from firms and the performance of local regulators has resulted in imperfect local environmental enforcement and the failure of central environmental policies. For instance, in its 11th Five-Year Plan (FYP, 2006–2010), the central government announced a goal of reducing the amount of key pollutants (COD[4] and sulfur dioxide (SO_2 actually rose by 1.2% and 1.8%, falling short of planned targets. Then-Premier Wen Jiabao characterized the 11th FYP targets for environmental protection as "a very serious matter that cannot be changed and must be unswervingly achieved".[5] However, measures to effectively strengthen central supervision and enhance local environmental enforcement must be taken for these targets to be met.

2.3. The NSMF program

In 2007, the central government launched the NSMF program, which placed key industrial polluters under special monitoring at the national level. The program addressed 3115 water-polluting firms, 3592 air-polluting firms, and 658 sewage treatment plants as NSM firms when it took effect. By 2016, the number of NSM firms had grown to 14,312, including 2660 water-polluting firms, 3281 air-polluting firms, 3812 sewage treatment plants, 21 large-scale livestock and poultry firms, 2901 heavy metals-polluting firms, and 1637 hazardous waste-polluting firms.

According to the official document provided by the MEP, the initial NSM firms were selected on the basis of their pollutant emissions over the last two years.[6] First, using the water-polluting firms on the 2007 NSM list as an example, all polluting firms with records in the CES were ranked by their COD and NH3-N emissions in 2005. The top-ranked firms, which accounted for 65% of total emissions of either COD or NH3-N, were designated as NSM water-polluting firms. The list of designated NSM firms was verified with local governments and finally published on the MEP website.[7]

The essence of the NSMF program is to collect more reliable and accurate information on NSM firms' pollution emissions through direct central supervision. In practice, all NSM firms are required to install automatic monitoring systems to transmit real-time emissions data to the national monitoring network. To ensure the accuracy of the data, the monitoring center conducts monthly supervisory checks on them to identify any abnormalities or inconsistencies, which are recorded and investigated. Moreover, on-site inspections are required at least once a month to ensure the proper functioning of the automatic monitoring systems and the firms' compliance. The verified information and inspection results are passed on to local governments to jointly determine an NSM firm's pollution levy (MEP, 2007). Through comprehensive technical and institutional reforms, the central government sought to significantly reduce the asymmetric information in the decentralized system and enhance the local enforcement of environmental regulation. However, the reform did not reverse the system of decentralized environmental regulation to a centralized system because the fundamental central-local relationship remained unchanged. In the new system, central government did not take over the role of local regulators but merely provided oversight to enhance their regulatory capacity and accountability. Local governments remain responsible for administering local environmental regulations, albeit in a more transparent and supervised environment.

3. Theoretical framework

To understand how central supervision can reduce emissions in China, we elaborate a model to analyze firms' pollution behavior

[2] Local officials were evaluated and promoted in the political hierarchy primarily based on the economic performance and social stability of the area under their administration.

[3] Polluting firms are aware of the resource shortages of EPBs and can strategically avoid monitoring and inspection by discharging excess pollution at night.

[4] COD is a commonly used indicator of water pollution in environmental chemistry.

[5] In the Primer annual report to the National People's Congress in March 2007.

[6] The initial rules and the NSM list did not change until 2009. In 2010, the MEP first added heavy metal pollution for consideration. The screening process for the NSM list has become increasingly complicated since 2010. As of 2016, it consists of six itemized lists for water-polluting firms, air-polluting firms, sewage treatment plants, large-scale livestock and poultry firms, heavy metal-polluting firms, and hazardous waste-polluting firms. For simplicity, we focus on only the initial NSM list from 2007 and restrict the sample to the period prior to 2010 in this study.

[7] See Table A1 in the appendix for detailed figures for NSM polluting firms by province in 2007.

under a system of decentralized command and control regulation.[8] This simple theoretical analysis considers central government's limited information on firms' emissions when local regulators enforce environmental regulations and clarifies the mechanism through which central supervision reduces emissions.

3.1. Model setting

Under environmental decentralization, traditional command and control regulation is implemented in a principal-agent structure. Three types of players are in the game: the central government (principal), local regulators (agents), and polluting firms (agents). On society's behavior, the central government establishes an initial pollution level e^0 as the environmental standard and signs an environmental regulatory contract with the local regulators. These regulators then implement environmental regulations as stipulated in the contract, with a fixed contract payment w. The firms set their production level y and pollution emissions e subject to local regulation. Finally, the central government assesses the local regulator's execution of the contract to ensure effective environmental regulation. For simplicity and without loss of generality, our model assumes one central government, one local regulator and one firm. All players are considered risk-neutral.

First, consider the firm's problem. At any production level y, The firm seeks to maximize its profit π, which is determined by the revenue and cost. We denote firm's revenue from production as $R(y)$ and assume that $R_y > 0$, and $R_{yy} < 0$. The firm emits pollution as a byproduct of its production; in its most general form, the pollution emissions level is a function of the production level and the amount of effort devoted to reducing pollution. For simplicity, we ignore the connection between production and pollution in this model setting. Thus pollution emissions are determined solely by the firm's efforts to abate pollution, which are costly. The extra cost of pollution abatement $C(v)$ is a function of excessive pollution emissions $v = e - e^0$. C is assumed to be twice continuously differentiable and strictly concave, i.e., $C_v > 0$, and $C_{vv} < 0$. Therefore, the firm's profit in compliance with environmental regulation is given by:

$$\pi^c = R(y^0) - C(v) \tag{1}$$

Whereas a firm in noncompliance with environmental regulation does not consider the abatement cost $C(v)$, and its profit equals $R(y) > R(y^0)$.

Under command and control regulations, the local regulator inspects the firm's compliance with the initial environmental standard. If the firm is found in noncompliance, the local regulator imposes a penalty $h^f(v)$ according to its excessive pollution emissions level $v = e - e^0$. For simplicity, we assume that the penalty is a linear function of the firm's excessive pollution emissions level (i.e., $h^f_e > 0$ and $h^f_{ee} = 0$ for any given e^0). In reality, the local regulator cannot always identify and punish violations due to asymmetric information and the regulator's limited resources and capacity to inspect and monitor. Logically, the probability $\rho(\phi_{bf})$ of the firm being caught in violation increases with the local regulator's knowledge of the firm's emissions ϕ_{bf}, i.e., $\rho_{\phi_{bf}} > 0$. Once the firm is caught in noncompliance, it can either accept its punishment by paying a penalty $h^f(v)$ or attempt to collude with local regulator by transferring a bribe $B < h^f(v)$ to the local regulator.

The local regulator earns a fixed wage w according to the contract with the central government. Beyond that, it also receives numerous benefits from local economic development. For instance, it receives tax payments from firms. Moreover, it may be promoted politically to higher ranked position based on the local economic performance. We model these benefits as $\lambda R(y)$, where $\lambda > 0$ is exogenously given for the local regulator. Under this situation, the local regulator has the incentive to collude with firms, possibly resulting in a breach of contract. According to the previous model setting, the central government will supervise the local regulator's execution of the contract. Here, we assume that the probability of the central government detecting collusion is $\sigma(\phi_{cf})$. Similarly, the probability function increases with the central government's knowledge of the firm's emissions level ϕ_{cf}, i.e., $\sigma_{\phi_{cf}} > 0$. If the collusion is found, the firm must pay its penalty $h^f(v)$ in addition to its paid bribe, and the local regulator is also punished for breach of contract. This sanction includes a withdrawal of the original wage w and an additional fine $h^b(v)$. Again, we assume that $h^b_e > 0$ and $h^b_{ee} = 0$ for any given e^0).

In this setting, the firm's payoff varies in different circumstances. The payoff matrix is summed up in Table 1:

From Table 1, we can calculate the firm's expected payoff from pollution under different scenarios:

Scenario A:

$$EU_A = R(y^0) - C(v) \tag{2}$$

Scenario B:

$$\begin{aligned} EU_B =& (1-\rho) \times R(y) + (1-\sigma)\rho \times [R(y) - h^f(v)] \\ &+ \rho\sigma \times [R(y) - h^f(v)] \\ =& R(y) - \rho h^f(v) \end{aligned} \tag{3}$$

Scenario C:

$$\begin{aligned} EU_C =& (1-\rho) \times R(y) + (1-\sigma)\rho \times [R(y) - B] \\ &+ \rho\sigma \times [R(y) - B - h^f(v)] \\ =& R(y) - \rho[B + \sigma h^f(v)] \end{aligned} \tag{4}$$

The firm chooses its best strategy and determines the optimal level of pollution emissions corresponding to $EU_i \subset MAX(EU_A, EU_B, EU_C), i = A, B, C$.

3.2. Model analysis

First consider the case without collusion between firm and local regulator. Under the original decentralized system of environmental regulation, as shown on the left side of Fig. 1, the interaction exists only between the central government and the local regulator and between the local regulator and the firm. The firm chooses between strategies of compliance and violation and determines its production

Table 1
Matrix of firms' payoffs in different scenarios.

	Firm not caught by regulator	Firm caught by regulator; regulator not detected by central government	Firm caught by regulator; regulator detected by central government
Probability	$1-\rho$	$(1-\sigma)\rho$	$\sigma\rho$
Scenario A: Compliance	$R(y^0) - C(v)$	$R(y^0) - C(v)$	$R(y^0) - C(v)$
Scenario B: Violate/Pay penalty	$R(y)$	$R(y) - h^f(v)$	$R(y) - h^f(v)$
Scenario C: Violate/Collusion	$R(y)$	$R(y) - B$	$R(y) - [B + h^f(v)]$

[8] We consider a command and control regime in this paper because it is the primary form of environmental regulation in China.

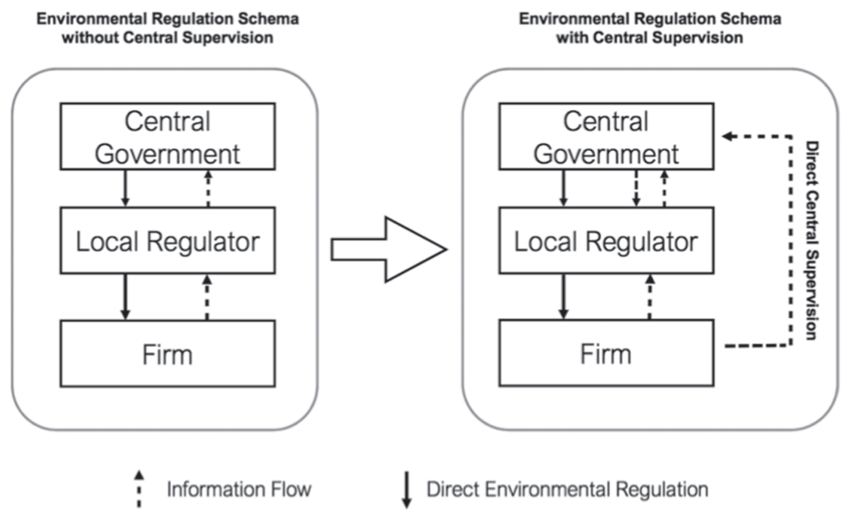

Fig. 1. Different models of decentralized environmental regulation.

and pollution emissions level. In the absence of collusion, the firm considers only its own payoff under regulation. The problem can be solved by maximizing the firm's expected gain due to a change from scenario A to scenario B.

$$\max_{y,e}\Omega = EU_B - EU_A = R(y) - \rho h^f(v) - [R(y^0) - C(v)] \tag{5}$$

The First-Order Condition (FOC) of firm's pollution emissions is as follows:

$$C_e(v) - \rho h_e^f(v) = 0$$
$$\Rightarrow C_e(v) = \rho h_e^f(v) \tag{6}$$

The FOC provides an important implication for firm's optimal pollution emissions in the equilibrium. That is the firm's optimal pollution emissions $e^\star$ corresponds to a point when its marginal cost of pollution abatement equals the marginal cost of punishment if it is caught by the local regulator. Therefore, an increase in either the probability of being caught or the severity of punishment by the local regulator will increase the firm's marginal abatement cost. Because $C_{ee} < 0$ for any given e^0, that is the marginal abatement cost is a decreasing function of e. The increasing marginal abatement cost thus implies the decreasing optimal pollution emissions $e^\star$ of firm. The result is summarized as Proposition 1.

Proposition 1. *In the absence of collusion between the firm and the local regulator, an increase in either the probability of detection by the local regulator or the amount of punishment reduces the firm's optimal pollution emissions.*

Proposition 1 is just similar to the basic ideas of crime and punishment set forth by Becker (1968). More importantly, the proposition gives a rational alternative to the original decentralization system. In a different model of decentralized environmental regulation with central supervision (right side of Fig. 1), central supervision over the firm's pollution emissions can help the local regulator, who has limited resources and capacity constraint, to better inspect and monitor polluting firms. As the real-time information about firm's pollution emissions flows directly to the central government and is shared with the local regulator, the knowledge of a firm's emissions ϕ_{bf} will increase. Since $\rho_{\phi_{bf}} > 0$, the probability of detection will increase and firm's pollution emissions will reduce.

The case is more complicated if the firm attempts to collude with local regulator. Here, a firm will need to consider the local regulator's payoff in addition to its own payoff in its decision making. Under the environmental decentralization, if the local regulator strictly enforces its environmental regulation contract with the central government, its expected payoff is $w + \lambda R(y^0)$. If the local regulator colludes with the firm, it risks being detected by the central government with a probability of σ. In this case, its payoff is $\sigma[\lambda R(y) + B - h^b(v)]$. Otherwise, its payoff is $(1-\sigma)[w + \lambda R(y) + B]$. The local regulator's expected gain from collusion with firm can be expressed as:

$$\begin{aligned} ENETI =& (1-\sigma)(w + \lambda R(y) + B) + \sigma[\lambda R(y) + B - h^b(v)] \\ & - [w + \lambda R(y^0)] \\ =& \lambda R(y) - \lambda R(y^0) + B - \sigma[w + h^b(v)] \end{aligned} \tag{7}$$

The firm's expected payoff from collusion can be calculated as:

$$ER = EU_C - EU_A = R(y) - \rho B - \rho\sigma h^f(v) - R(y^0) + C(v) \tag{8}$$

The local regulator and the firm are now in a cooperative game, where their strategies are interdependent. A solution of the Nash bargaining equilibrium can be found by maximizing the product of the firm and the local regulator's gains from collusion. The overall maximization problem becomes:

$$\max_{y,e}\Omega = ER \times ENETI \tag{9}$$

The FOC of the firm's pollution emissions can be simplified and is expressed as follows:

$$C_e(v) - \rho\sigma h_e^f(v) + \frac{1}{\lambda}\sigma h_e^b(v) = 0$$
$$\Rightarrow C_e(v) = \sigma\left[\rho h_e^f(v) - \frac{1}{\lambda}h_e^b(v)\right] \tag{10}$$

Compare Eq. (10) with Eq. (6). Since $\lambda > 0, h_e^b(v) > 0$ and $0 < \sigma < 1$, the marginal abatement cost for a firm's excessive pollution emissions in the case of collusion is lower than that in the case of non-collusion. This result is because the local regulator derives benefits from the firm's pollution at a cost of potential punishment h^b, which offsets the marginal cost of punishment h^f for the firm. The equilibrium condition implies that the optimal level of the firm's pollution emissions $e^\star$ is higher in the case of collusion. At this point, any increase in the probability of the central government detecting the collusion σ will increase the firm's marginal abatement cost, implying a decreasing level of optimal pollution emissions $e^\star$. The result is described in Proposition 2.

Proposition 2. *When the firm and the local regulator collude, the increase in the central government's probability of detection reduces the firm's optimal pollution emissions.*

As illustrated on the right side of Fig. 1, direct supervision over the firm will increase the central government's knowledge of the firm's real pollution emissions. This policy increases the probability of detecting the collusion, resulting in a higher corresponding marginal abatement cost of the firm's pollution emissions. As a result, the higher cost leads to a smaller optimal pollution emissions level.

Propositions 1 and 2 suggest that under a regime of decentralized environmental regulation, direct supervision by the central government is effective at reducing firms' excessive pollution emissions by mitigating the problem of imperfect information about polluting firms stemming from the local regulator's capacity constraints or collusion with firms. This theoretical analysis elucidates the mechanism of central supervision and provides guidance for the empirical analysis addressing reforms to China's environmental decentralization, to which we now turn.

4. Empirical strategy

In the empirical analysis, we focus on the NSMF program to evaluate the impact of central supervision. Since the program provides an exogenous cutoff, we adopt an RD design, pioneered by Thistlethwaite and Campbell (1960), for treatment evaluation. Economists have recently strengthened RD estimation methods, and it has been widely adopted in various fields including education, health and labor market (see Lee and Lemieux, 2010 for a review). In this section, we detail how we tailored the RD design to our setting to evaluate the NSMF program.

4.1. The identification strategy

We estimate the following relationship to determine the impact of the NSMF program:

$$Y_i^t = \alpha X_i^t + \beta NSM\_firm_i + \mu_i + \varepsilon_i, t = 0, 1, 2 \tag{11}$$

where Y is the firm's water pollution emissions as measured based on COD.[9] COD is a commonly used measurement of water pollution in environmental chemistry that measures the oxygen required to oxidize soluble and particulate organic matter in water. Compared to other indicators of water pollution, e.g., ammonia nitrogen (NH3-N) and biological oxygen demand (BOD), which are emitted by some industrial firms, COD is emitted by most industrial firms and has thus become a key indicator under routine monitoring in China. Moreover, COD is a key indicator of the Chinese local government's environmental performance because of its data integrity and because of the attention it has garnered. For these reasons, we follow the literature and use COD as our main dependent variable (Kahn et al., 2015).

X is a vector of firm characteristics, and *NSM_firm* denotes treatment by the NSMF program, taking a value of one if the firm is an NSM firm and zero otherwise. The subscript i indicates a firm. In our setting, it may take time for the effects of the NSMF program to be observed; thus t indicates the firm's outcome t years after the program's implementation. μ represents unobserved firm characteristics, and ε is a stochastic error term.

Given the non-random selection process of the NSM firms, the correlation between μ and *NSM_firm*, i.e., $cov(NSM\_firm, \mu) \neq 0$, may bias estimates of the treatment effect with the OLS estimator. However, as discussed in the previous section, water-polluting firms are designated as NSM firms based on the 65th percentile of firms ranked by their amount of water pollutants discharged in 2005. This rule provides two cutoff points, i.e., COD and NH3-N, for the identification of the treatment effect.[10] Provided that COD is a key indicator with a far superior data integrity, we rely on the COD cutoff for primary analysis and consider the NH3-N cutoff later in our robustness check. As shown in Fig. 2, the COD cutoff creates a highly nonlinear relationship between the firm's COD emissions in 2005 and its probability of being designated as an NSM water-polluting firm.

There is a clear jump in the probability of being an NSM firm at the COD cutoff point. Polluting firms above the threshold have an 80% chance of being designated as an NSM water-polluting firm, while those below it have less than a 10% chance. The rule provides exogenous variation in treatment, which can be used for identification. However, the critical exogeneity of treatment requires that a firm be unable to manipulate their data to avoid treatment (McCrary, 2008). We believe this type of intentional manipulation is implausible in our case because the cutoff point is predetermined using emissions data from two years prior to the program. Firms were unable to modify their data at the time of the program and had no access to national data from the CES, which was not publicly available at the time.

Intuitively, assuming that unobservable characteristics vary continuously within a small range of the cutoff point, NSM and non-NSM firms immediately above and immediately below the threshold can be assumed to be similar and their environmental outcomes comparable. If NSM firms directly above the cutoff emit less COD than non-NSM firms directly below it after the NSMF program's implementation, the NSMF program has a positive impact on water pollution reduction.

4.2. Estimation using a non-parametric method

If an NSM firm's status is perfectly determined by the 65% threshold, polluting firms immediately above the cutoff point may be assumed to be similar to those immediately below it aside from their exposure to treatment. This case is referred to as the Sharp RD, and a simple method of estimating the treatment effect β is to compare the mean outcome of polluting firms at the cutoff point. This method can be represented mathematically with the following expression:

$$\beta_{RD} = \lim_{Z_i \downarrow z_c} E\left[Y_i^t(1)|Z_i\right] - \lim_{Z_i \uparrow z_c} E\left[Y_i^t(0)|Z_i\right] \tag{12}$$

where Z denotes the running variable used to determine an NSM firm's status, and z_c is the cutoff point, i.e., 65% of total COD emissions in 2005. The treatment effect is then calculated as the difference in

[9] The NSM list is pollutant-specific and involves water-polluting firms, air-polluting firms and sewage treatment plants. Given the paper's length, we decided to focus primarily on the NSM list of water-polluting firms and thus its impact on water pollution reduction. Its impacts on air pollution and sewage treatment are left for future investigation.

[10] Using the environmental statistics dataset, we calculated that the thresholds of 199 metric tons COD per year and 60 metric tons Nh3-N per year approximately correspond to 65% of total emissions. In the following study, we use these calculated thresholds for our RD design.

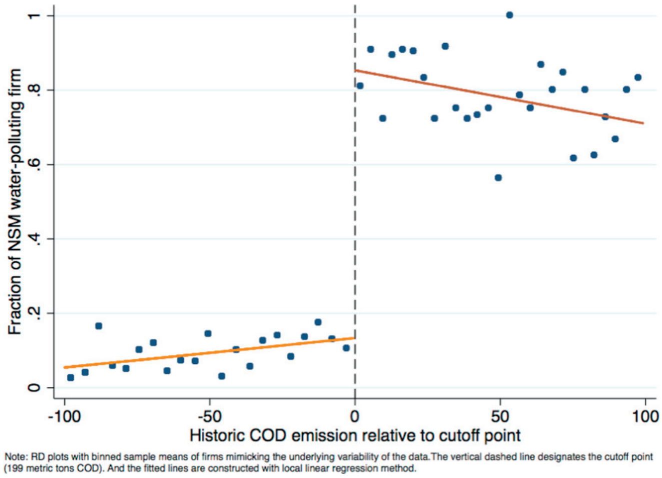

Fig. 2. Distribution of NSM water-polluting firms around the cutoff.

the limits of conditional expectations of future outcomes close to the threshold.

However, as illustrated in Fig. 2, the determination of an NSM firm's status is more complex in reality. Approximately 20% of firms across the cutoff received waivers, while 10% of firms below the cutoff were designated as NSM firms. This "fuzzy" discontinuity may stem from a number of causes such as administrative record errors and other related determinants of NH3-N. In a situation with "fuzzy" discontinuity, provided the probability of treatment changes discontinuously at the cutoff (i.e., as illustrated in Fig. 2), the implementation of an RD design remains feasible. The treatment effect β_{RD} merely needs to be scaled by the difference in the probability of treatment, which yields the following local Wald estimator:

$$\beta_{FRD} = \frac{\lim_{Z_i \downarrow z_c} E\left[Y_i^t(1)|Z_i\right] - \lim_{Z_i \uparrow z_c} E\left[Y_i^t(0)|Z_i\right]}{\lim_{Z_i \downarrow z_c} E\left[NSM\_firm_i(1)|Z_i\right] - \lim_{Z_i \uparrow z_c} E\left[NSM\_firm_i(0)|Z_i\right]} \quad (13)$$

Although polluting firms near the cutoff are similar, they are not identical. One potential concern is that differences in outcome between the two groups reflect differences in firm characteristics rather than the impact of the program. This assumption is indeed critical and needs to be made to ensure the internal validity of the RD design. Based on our available data, we can check differences in some observable characteristics such as output, water treatment equipment quantity and capacity, and NH3-N emissions at the cutoff.[11] Regarding unobservable characteristics, we can examine the change in outcome relative to the initial level to partially remove time-invariant characteristics. If we assume that trends in pollution reduction are the same for both NSM and non-NSM firms near the cutoff, we can identify the treatment effect using the following difference-in-differences likewise estimator:

$$\beta_{FRD\_DID} = \frac{\lim_{Z_i \downarrow z_c} E\left[Y_i^t(1) - Y_i^0(1)|Z_i\right] - \lim_{Z_i \uparrow z_c} E\left[Y_i^t(0) - Y_i^0(0)|Z_i\right]}{\lim_{Z_i \downarrow z_c} E[NSM\_firm_i(1)|Z_i] - \lim_{Z_i \uparrow z_c} E[NSM\_firm_i(0)|Z_i]} \quad (14)$$

A central issue surrounding the non-parametric estimation in RD design is choosing for analysis an appropriate bandwidth, i.e., the range of observations around the cutoff point. The advantage of selecting a narrow bandwidth is that the treatment effect is estimated with limited bias because the underlying relationship between the outcome and the running variable tends to be linear. Increasing the bandwidth may increase bias if the true relationship is nonlinear. However, a narrow bandwidth may also reduce the precision of impact estimates by shrinking the sample size. Furthermore, a narrow bandwidth complicates the extrapolation of estimates as the estimated treatment effect generalizes only to units in the range around the cutoff. The choice of bandwidth therefore involves a tradeoff between statistical power and estimation bias.

Various methods have been developed to determine the optimal bandwidth. For instance, the "cross-validation" and "plug-in" procedures have been proposed to select a bandwidth based on minimizing the mean squared errors (MSE) (Imbens and Kalyanaraman, 2012; Ludwig and Miller, 2007). However, these approaches lead to a first-order bias in the distributional approximation, so the MSE-optimal bandwidths are too large in practice. More recently, Calonico et al. (2014) have proposed a novel approach (referred to as the CCT approach) to address this concern based on bias correction. In our analysis, we rely primarily on the CCT approach to select the optimal bandwidth. We also test narrower and wider bandwidths to check the sensitivity of our results to different bandwidths.

4.3. Estimation in an instrumental variable framework

In addition to the non-parametric method, we can also implement the fuzzy RD design in an instrumental variable(IV) framework, i.e., a parametric two-stage least squares (2SLS) regression. We assume that the water-pollution equation takes the following form:

$$Y_i^t = \alpha X_i^t + \beta NSM\_firm_i + \gamma f\left(Z_i^{2005}\right) + \varepsilon_i, t = 0, 1, 2 \quad (15)$$

where Y is the firm's water pollution emissions measured based on the COD (in log) one or two years after the NSMF program's implementation. X is a vector of the firm's characteristics related to their COD emissions, including the firm's output, NH3-N emissions, wastewater treatment equipment and capacity; *NSM_firm* is a dichotomous indicator of whether a firm participated in the NSMF (yes or no); $f(Z)$ is a function of the running variable; and ε is the stochastic error term. We refer to this specification the future outcome model. In addition, we test another specification by examining the change in COD emissions relative to the base level (i.e., in 2006 before the NSMF program's implementation), which we call the outcome change model, as follows:

$$\triangle Y_i^t = \alpha X_i^t + \beta NSM\_firm_i + \gamma f\left(Z_i^{2005}\right) + \varepsilon_i, t = 0, 1, 2 \quad (16)$$

The first-stage model is given by:

$$NSM\_firm_i = \lambda X_i^t + \rho f\left(Z_i^{2005}\right) + \delta Eligibility_i^{2005} + \upsilon_i, t = 0, 1, 2 \quad (17)$$

where *Eligibility* is a dummy variable indicating whether the firm's COD emissions in 2005 were above the cutoff (i.e., the IV), and υ is the error term.

The main advantage of using this 2SLS estimator is its flexibility to include firm-level characteristics as well as industry- and region-fixed effects to check the robustness of our results. However, one pitfall is that it relies on an assumption about the functional form of the running variable $f(Z)$. For instance, if the true functional form is nonlinear around the cutoff, whereas we specify it as linear, then the estimated treatment effect may simply pick up any non-linearity in the outcome relationship. To address this concern, we follow the literature and allow the first- and second-order polynomials of the

[11] We did not include any covariates in the non-parametric estimation because the error in the coefficients of covariates can reduce efficiency, and any endogenous covariates can increase bias.

running variable to relax the functional form assumption (Gelman and Imbens, 2014)[12].

We note that no consensus has been reached in the literature on the use of the non-parametric versus the parametric approach in estimating RD design. On the one hand, the non-parametric approach involves selecting an optimal bandwidth within which the functional form can be approximated with a linear function effectively reducing estimation bias at the cost of reduced statistical power. On the other hand, the parametric approach seeks to specify the correct functional form using all available observations. If it improves the estimate's precision, the risk of bias due to misspecification is inevitable. The choice between the two approaches is thus a trade-off between estimation bias and efficiency (Jacob et al., 2012). To ensure the robustness of our results, we use both non-parametric and parametric approaches for the estimation.

5. Data

The data used for the analysis are derived from the large administrative CES dataset collected by the MEP, which has compiled administrative data on ambient environmental quality and pollution by industrial firms since the 1980s. The monitoring system covers the majority of polluting industrial firms, which contribute approximately 85% of China's total major pollutants (e.g., COD, ammonia nitrogen, sulfur dioxide, industrial smoke and dust as well as solid waste).

Polluting firms are required to report detailed environmental data-such as major pollutant emissions and treatment conditions, hazardous waste disposal, raw materials, and energy inputs-on an annual basis. This information is verified by local EPBs, combined at the MEP and finally used to construct the CES dataset and to produce the Chinese Environmental Yearbook. The CES dataset is thus regarded as the most comprehensive and reliable environmental microeconomic data in China. However, CES data were only recently made accessible to researchers.

We derive a sample of 20,607 polluting firms between 2005 and 2009 from the CES dataset for our analysis.[13] Our data comprise information on major water pollutant (COD and NH3-N) generation, treatment and emissions, as well as firm characteristics such as industry code, location, industrial output, and the quantity and capacity of wastewater treatment equipment. We use the 2007 NSM list published by the MEP and focus on the water-polluting firms in our dataset. Table A2 in the appendix contains data on the definition and sources of the variables, and Table A3 reports the summary statistics for the variables by NSM firm status.

From Table A3, we can derive a preliminary understanding of NSM firms with regard to their counterparts. The 2007 data, for example, clearly show that NSM firms were much more polluting than the non-NSM firms. On average, an NSM firm generated 4767 metric tons and emitted 1020 metric tons of COD; whereas, a non-NSM firm generated only 235 metric tons and emitted 37 metric tons of COD per year. This significant pattern persisted over time. Meanwhile, NSM firms are also distinct from non-NSM firms in a number of aspects. For instance, NSM firms had a higher industrial output (1.85 billion yuan/year versus 0.29 billion yuan/year), more water treatment equipment (3.4 versus 1.4) and a greater capacity for wastewater treatment (44,900 metric tons/day versus 3100 metric tons/day). Based on these descriptive statistics, a simple OLS regression analysis yields biased estimates of the treatment effect. We further plot the data in Fig. A1 to provide an overview of firms' COD emissions over year. As noted, massive firms gathered below a threshold (many small polluting firms in our data), which suggests a cutoff in firms' COD emissions. Although the massive data makes it difficult to eyeball test any significant discontinuity at the cutoff, one can still observe that a small break at the cutoff has been enlarging since 2008. This preliminary observation motivates us to zoom in on the data and supports our analysis using a small subset near the cutoff and implementation of an RD design.

6. Results

6.1. The graphical evidence

We first present the graphical evidence from the fuzzy RD design. The graphs provide a straightforward visualization of the program's impact. In particular, we focus on the observations in a small window around the cutoff point.[14] We adopt the method used in Lee and Lemieux (2010) to divide the window into evenly spaced "bins" and plot the mean value of each "bin" against the running variable.[15] The fitted line is constructed using a flexible local linear regression, enabling us to obtain a clear picture of the data distribution.

Fig. 3 shows the pattern in firms' COD emissions over time. A "jump" (discontinuity) at the cutoff point indicates the potential impact of the NSMF program on a firm's COD emissions. As the figure shows, the distribution of firm's COD emission is quasi-continuous at the cutoff point before the program's implementation in 2006. This continuous trend confirms the internal validity of the RD design. In 2007, as the program took effect, a slight "jump" begins to appear at the cutoff point, which suggests that the program had a limited effect at the time. The "jump" becomes clearer and larger in 2008 and 2009, suggesting that the program's impact was substantial and lasting over time.

However, the discontinuity of firms' characteristics at the cutoff may confound the policy impact; therefore, we check some of the critical characteristics (i.e., firms' output, NH3-N emissions, water treatment equipment, and capacity) at the program's start in 2007. Fig. 4 shows that while the NH3-N is slightly different, all other characteristics are continuous at the cutoff point, meaning polluting firms near the cutoff point are similar in terms of these characteristics.

To make sure that firms near the cutoff have the same pre-trends in their COD emissions, we then construct Fig. 5 for a check. Firms above the cutoff are defined as "eventually regulated" group and firms below the cutoff are defined as "eventually not regulated" group. We contrast the mean COD emissions of the two groups over time to show their trends before and after 2007 when the NSMF program took effect.

As shown in Fig. 5, before the program took effect (2005–2006), the regulated group emitted more COD emissions than the non-regulated group. The trends are quasi parallel for two groups, regardless of the regulatory status. After 2007, the COD emissions of

[12] According to Gelman and Imbens (2014), the use of high-order polynomials in RD analysis is a flawed approach. They demonstrate three major problems in which higher order polynomials lead to noisy estimates, sensitivity to the degree of the polynomial, and poor coverage of confidence intervals. Therefore, we include only the first- and second-order polynomials to relax the functional form assumption of the parametric RD approach.

[13] To achieve consistent results over time, we keep firms with observations during the entire period to construct a balanced panel dataset. Firms that were shut down due to the NSMF program are thus not included in our analysis. This sample selection may introduce a downward bias to our estimates, and the estimation in this analysis is therefore a lower bound of the NSMF program impact.

[14] The preliminary bandwidth here is 100 metric tons of COD, similar to that in the following parametric analysis. We also adopt a more rigorous approach to select the optimal bandwidth in the following econometric analysis.

[15] The approach is implemented using the "rdplot" user-written command in STATA version 13.

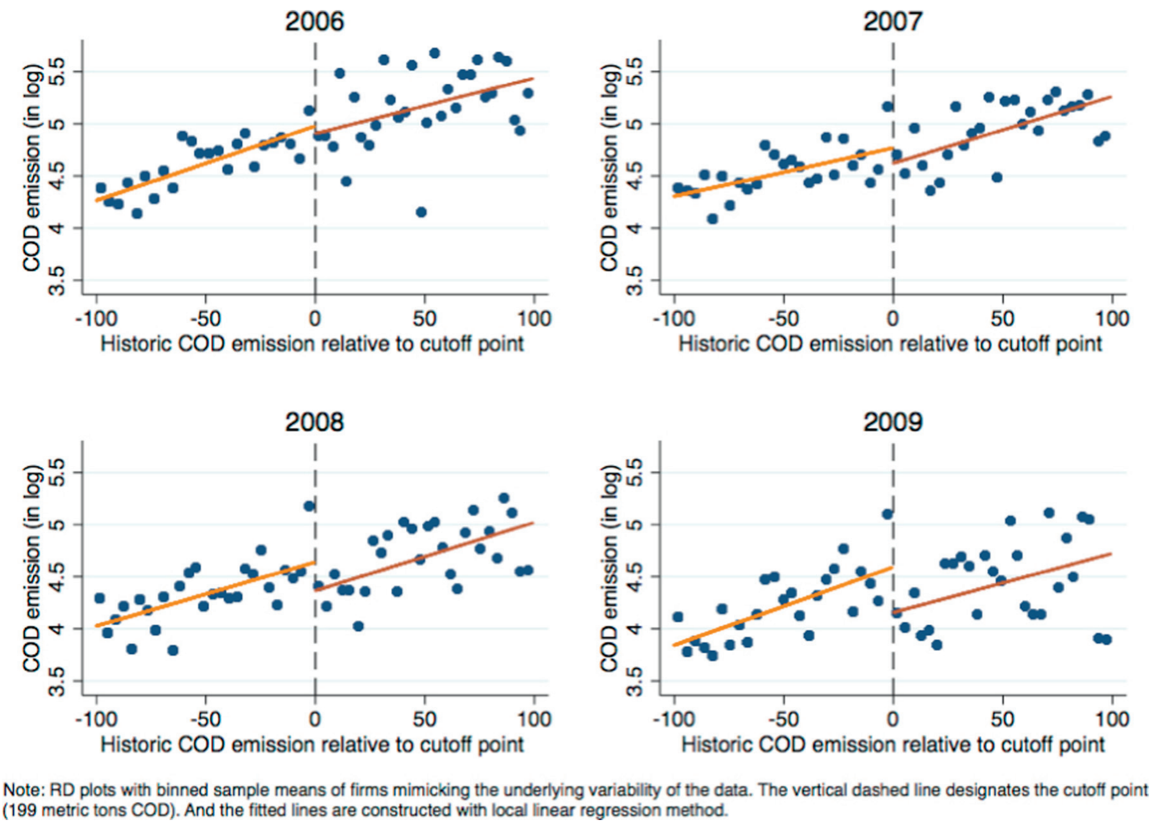

Fig. 3. The discontinuity of outcomes at the cutoff.

regulated group substantially decreased, while the decrease of their counterpart was slower. As a result, the COD emissions level of the regulated group were lower than that of the non-regulated group. This pre-trends graph sheds lights on the internal validity of the RD analysis. Although the graphs lend transparency to the program's impact, the magnitude and the statistical significance of the impact must be quantified. We continue our regression analysis by focusing on observations within a small subset around the cutoff to estimate the impact using both non-parametric and parametric estimation.

6.2. The impact of central supervision on industrial water pollution

In the non-parametric estimation, we rely primarily on the CCT approach to select the optimal bandwidth. In addition, we use half and double the CCT bandwidth to check the sensitivity of our results to the bandwidth selection. In the parametric estimation, we use all observations within the narrower bandwidth and test the model specification with first- and second-order polynomials of the running variable.[16] Both estimation methods are used to estimate two models, i.e., the future outcome model and the outcome change model (c.f., Eqs. (13) and (14)). The results are presented in chronological order to check the impact's delay.

Table 2 reports the results of the non-parametric estimation. The results are generally in line with our graphical analysis. For the future outcome model and the outcome change model, the impact estimated by the local linear regression is non-significant in 2006 and 2007; whereas, it is negative and significant in 2008 and 2009 (at the 95% and 99% levels, respectively). These estimates are robust to the choice of different bandwidths. The results indicate that before the NSMF program's implementation, the general time trends in COD emissions were similar for both NSM and non-NSM firms. After the program took effect in 2007, central supervision had enhanced the local regulation of water-polluting firms and reduced their COD emissions. However, the NSMF program's effect was not immediate but was subject to a delay: the official announcement was made in mid-2007, but local governments needed time to implement the program, as did designated NSM firms, who had to adapt to more stringent environmental regulations. With direct central supervision over their emissions, NSM firms made more substantial COD emissions reductions than non-NSM firms near the cutoff. Moreover, this impact appears to have been lasting over time: when a firm spent a longer amount of time under central supervision, the firm reduced its COD emissions by a greater amount.

The results of the parametric IV regression are presented in Table 3.[17] We first check the functional form of the model. Note that the coefficients of the quadratic term of the running variable are small and insignificant throughout all regressions, which suggests a linear relationship between the running variable and the outcome. Once we have controlled for this linear relationship, the coefficient of the treatment effect appears to be negative and significant in 2008 and 2009 but insignificant in 2006 and 2007. In addition, we control for firms' characteristics as well as for province- and industry-fixed effects to improve estimation precision. The standard errors are clustered at the province level. The results are consistent

[16] As in the previous graphical analyses, we use a bandwidth of ± 100 metric tons COD for the parametric regression analysis. We also test different bandwidths and report the results in Table A4 in the appendix.

[17] The first-stage results are presented in Table A5 in the appendix.

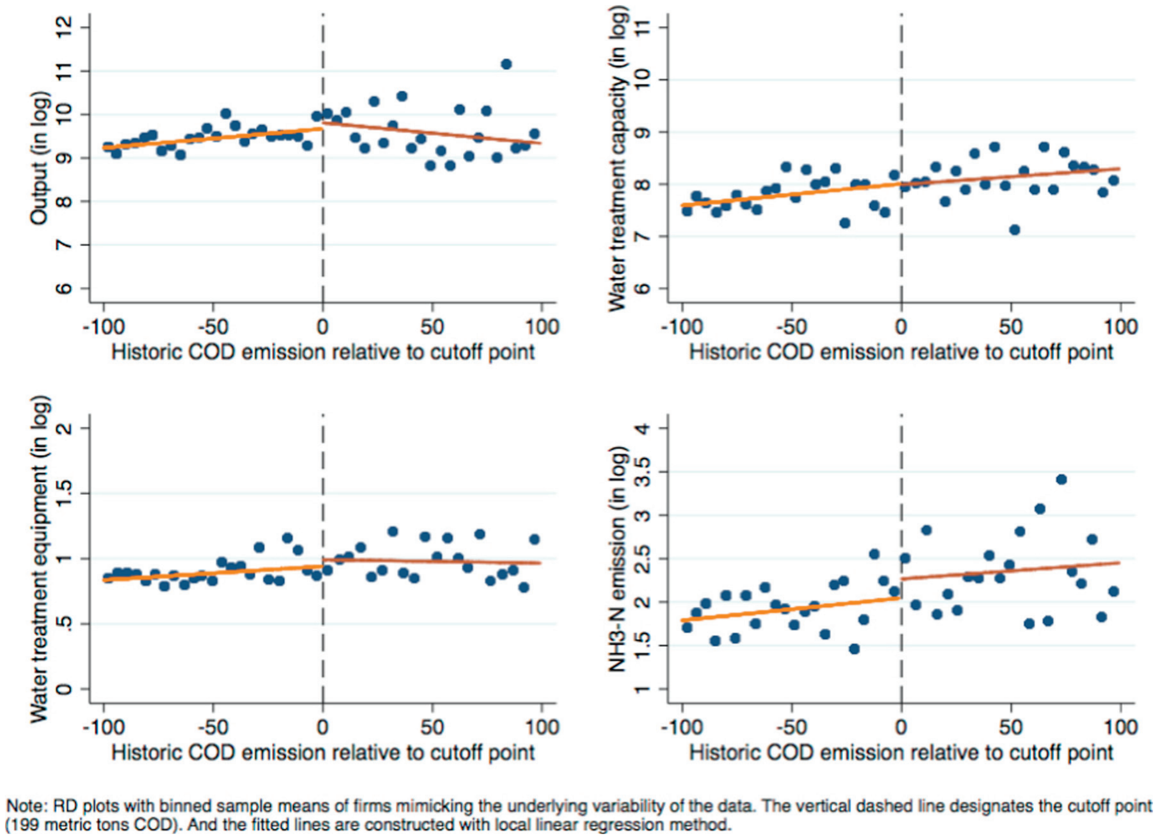

Fig. 4. Continuous firm characteristics at the cutoff.

with those from the non-parametric estimation and confirm the delayed but lasting impact of the NSMF program on water pollution reduction. To determine the impact's magnitude, we base the most conservative interpretation on the more precise parametric estimates. *Ceteris paribus*, the NSMF program reduced firms' COD emissions by at least 26.8%. This significant and lasting impact suggests substantial marginal benefits from central supervision over the pollution emissions of NSM firms and highlights the efficacy of direct central supervision in China's environmental decentralization.

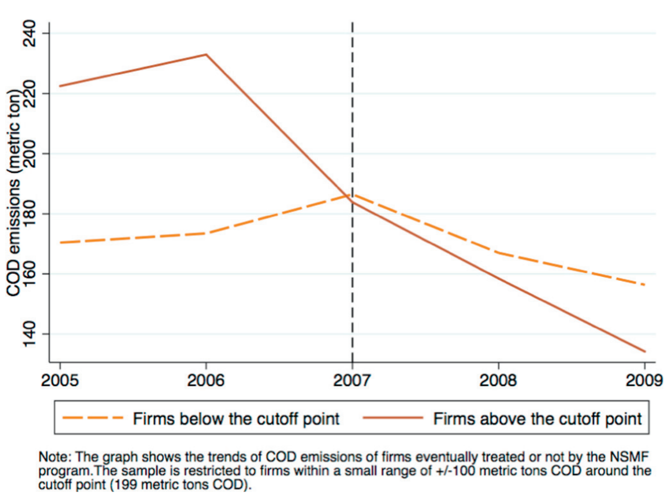

Fig. 5. Trends of COD emissions by firms near the cutoff.

We compare our results with findings on environmental federalism in the literature. As previously discussed, evidence on the efficacy of environmental federalism is mixed. Using country-level data, Sigman (2007) examines the effects of decentralization on water pollution. She finds that federal countries exhibit greater interjurisdictional variation in pollution, but pollution levels are not affected. From another angle, Bulte et al. (2007) explore US state-level data to provide evidence that air pollution converged during the early years of federal pollution control but not when local control was in effect. Sjöberg (2016) uses Swedish municipality data and finds that local politicians affected not only environmental policy but also outcomes through policy implementation. Our study on an alternative model of environmental decentralization in China complements the current literature and suggests that decentralized environmental regulation may have limited effectiveness in developing countries. In this case, direct central supervision complementing local environmental regulation can effectively enhance local environmental enforcement and thereby reduce water pollution.

6.3. Firm's behaviors and social welfare under central supervision

One important concern over central supervision is its potential detrimental effect on firms' economic output. As centrally supervised polluting firms are free to choose their own pollution abatement strategies, so reductions in COD emissions may be achieved by either increasing end-of-pipe abatement or reducing production. If most supervised firms elect the latter strategy, central supervision is likely to induce unexpected economic losses and reduce social welfare. To test this potential hypothesis, we use a firm's COD production (i.e., the amount of COD generated in a firm's production processes) and a firm's economic output as the dependent variables and repeat the RD analysis. The results are reported in Tables 4 and 5.

Table 2
Non-parametric RD estimates of the impact of the NSMF program on firm COD emissions by year.

	(1)	(2)	(3)	(4)	(5)	(6)
Dep var.	$Ln(COD)$			$\triangle Ln(COD)$		
Bandwidth	50%CCT	100%CCT	200%CCT	50%CCT	100%CCT	200%CCT
2006	−0.271	−0.0765	0.0251	−0.0879	0.00948	−0.0345
	(0.327)	(0.221)	(0.166)	(0.253)	(0.202)	(0.145)
Bandwidth	18.777	37.554	75.108	29	58	116
Obs.	183	429	953	314	695	1773
2007	−0.484	−0.239	−0.233	−0.155	−0.220	−0.189
	(0.477)	(0.236)	(0.171)	(0.217)	(0.165)	(0.149)
Bandwidth	21.2145	42.429	84.858	39.974	79.948	159.896
Obs.	227	499	1123	469	1045	3567
2008	−0.939**	−0.706**	−0.533**	−0.728***	−0.670***	−0.419**
	(0.448)	(0.301)	(0.226)	(0.272)	(0.218)	(0.173)
Bandwidth	22.215	44.43	88.86	27.087	54.174	108.348
Obs.	238	524	1174	297	644	1567
2009	−0.966**	−0.937***	−0.720***	−0.805***	−0.863***	−0.590***
	(0.465)	(0.306)	(0.235)	(0.303)	(0.301)	(0.215)
Bandwidth	26.1495	52.299	104.598	32.563	65.126	130.252
Obs.	284	620	1500	365	800	2205

Note: RD estimates with non-parametric local linear regressions. The results of the future outcome model are reported in columns (1)–(3), and the results of the outcome change model are reported in columns (4)–(6). The cutoff point is 199 metric tons COD. The optimal bandwidth is selected using the approach of Calonico et al. (2014). Half and double the CCT bandwidth are used to check the sensitivity of the results. Conventional local linear regression robust standard errors clustered at the province level are reported in parentheses.
*** Indicates p<0.01.
* Indicates p<0.05.
* Indicates p<0.1.

Table 3
Parametric RD estimates of the impact of the NSMF program on firms' COD emissions by year.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep var.	$Ln(COD)$				$\triangle Ln(COD)$			
Year	2006	2007	2008	2009	2006	2007	2008	2009
NSM_firm	−0.060	0.026	−0.268*	−0.520***	−0.032	−0.037	−0.461**	−0.468**
	(0.178)	(0.172)	(0.153)	(0.166)	(0.178)	(0.191)	(0.203)	(0.206)
COD2005	0.005***	0.003***	0.005***	0.006***	−0.001	−0.001	0.001	0.0004
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
COD2005_sq	−1.44e−05	−9.24e−06	−8.17e−06	−1.93e−05	7.97e−07	−1.46e−06	5.42e−06	−6.67e−06
	(9.95e−06)	(1.21e−05)	(1.36e−05)	(1.35e−05)	(9.97e−06)	(8.42e−06)	(1.27e−05)	(1.10e−05)
Constant	5.102***	1.715***	2.151***	2.991***	−0.215	−2.909***	−2.691***	−2.095***
	(0.360)	(0.395)	(0.335)	(0.438)	(0.359)	(0.317)	(0.320)	(0.340)
Province FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control of firm characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	947	915	985	1088	947	915	985	1088
F-test of IV	236.67	399.15	376.5	428.34	236.67	399.15	376.5	428.34
Prob > F	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
R-squared	0.228	0.200	0.240	0.186	0.154	0.099	0.110	0.095

Note: RD estimates with the parametric regression of 2SLS models. The results of the future outcome model are reported in columns (1)–(4), and the results of the outcome change model are reported in columns (5)–(8). The cutoff point is 199 metric tons COD and the bandwidth is 100 metric tons COD by default. An alternative bandwidth is used in Table A4. Conventional robust standard errors clustered at the province level are reported in parentheses.
*** Indicates p<0.01.
** Indicates p<0.05.
* Indicates p<0.1.

Table 4 reports the results of the non-parametric RD estimates of the impact on a firm's COD production.[18] As the table clearly shows, neither the future outcome model nor the outcome change model yields a significant impact of the NSMF program on a firm's COD production. The results suggest that NSM firms are more likely to reduce their COD emissions through end-of-pipe abatement than by reducing production. It is plausible that end-of-pipe abatement is less costly than fundamentally adjusting production to comply with the regulation. This evidence clearly implies that NSM firms were not affected by central supervision at the production stage.

Table 5 reports the results of the non-parametric RD estimates of the impact on a firm's output. The results of most of the regressions are insignificant except the 2009 change in output. Yet the significance is not robust to larger bandwidths, which suggests an insignificant and negligible detrimental impact on firms' economic output. The results are in line with previous findings on firms' COD production. Taken together, they confirm that central supervision is socially optimal in the sense that it reduces water pollution without substantially damaging the local economy.

[18] The parametric IV estimation yields similar results, which are available upon request.

Table 4
Non-parametric RD estimates of the impact of the NSMF program on firm COD production by year.

	(1)	(2)	(3)	(4)	(5)	(6)
Dep var.	*Ln(CODP)*			△*Ln(CODP)*		
Bandwidth	50%CCT	100%CCT	200%CCT	50%CCT	100%CCT	200%CCT
2006	−0.569	−0.699	−0.412	−0.410	−0.145	−0.178
	(0.846)	(0.666)	(0.359)	(0.420)	(0.333)	(0.211)
Bandwidth	20.294	40.588	81.176	32.647	65.294	130.588
Obs.	162	365	836	273	620	1703
2007	−0.593	−0.473	−0.451	0.0397	0.00436	0.00143
	(0.713)	(0.580)	(0.339)	(0.197)	(0.256)	(0.154)
Bandwidth	20.598	41.196	82.392	29.4605	58.921	117.842
Obs.	150	354	817	194	466	1223
2008	−0.619	−0.0137	−0.292	0.268	−0.118	−0.0342
	(0.538)	(0.388)	(0.350)	(0.269)	(0.291)	(0.162)
Bandwidth	14.006	28.012	56.024	29.759	59.518	119.036
Obs.	108	235	541	206	489	1298
2009	−0.255	−0.278	−0.217	−0.208	−0.0209	0.00397
	(0.619)	(0.356)	(0.182)	(0.410)	(0.334)	(0.241)
Bandwidth	18.662	37.324	74.648	36.1045	72.209	144.418
Obs.	135	330	754	253	603	1916

Note: RD estimates with the non-parametric local linear regressions. The results of the future outcome model are reported in columns (1)–(3), and the results of the outcome change model are reported in columns (4)–(6). The cutoff point is 199 metric tons COD. The optimal bandwidth is selected using the Calonico et al. (2014) approach. Half and double of the CCT bandwidth are employed to check the sensitivity of the results. Conventional local linear regression robust standard errors clustered at the province level are reported in parentheses.
*** Indicates $p<0.01$.
* Indicates $p<0.05$.
* Indicates $p<0.1$.

Table 5
Non-parametric RD estimates of the impact of the NSMF program on firm output by year.

	(1)	(2)	(3)	(4)	(5)	(6)
Dep var.	*Ln(Output)*			△*Ln(Output)*		
Bandwidth	50%CCT	100%CCT	200%CCT	50%CCT	100%CCT	200%CCT
2006	0.449	0.433	0.0686	0.422	0.163	−0.00713
	(0.690)	(0.524)	(0.327)	(0.610)	(0.410)	(0.267)
Bandwidth	37.2605	74.521	149.042	38.7415	77.483	154.966
Obs.	422	936	2955	445	979	3266
2007	0.544	0.619	0.302	0.186	0.180	0.0704
	(0.649)	(0.506)	(0.360)	(0.166)	(0.131)	(0.0964)
Bandwidth	28.601	57.202	114.404	69.913	139.826	279.652
Obs.	309	684	1723	859	2560	19,486
2008	0.209	0.257	0.0527	−0.275	−0.194	−0.0117
	(0.735)	(0.540)	(0.363)	(0.259)	(0.222)	(0.171)
Bandwidth	36.1135	72.227	144.454	36.261	72.522	145.044
Obs.	407	893	2758	407	895	2788
2009	0.135	0.230	0.0543	−0.461**	−0.213	−0.113
	(0.700)	(0.488)	(0.337)	(0.202)	(0.172)	(0.145)
Bandwidth	38.189	76.378	152.756	30.8705	61.741	123.482
Obs.	438	970	3135	337	748	1986

Note: RD estimates with the non-parametric local linear regressions. The results of the future outcome model are reported in columns (1)–(3), and the results of the outcome change model are reported in columns (4)–(6). The cutoff point is 199 metric tons COD. The optimal bandwidth is selected using the Calonico et al. (2014) approach. Half and double of the CCT bandwidth are employed to check the sensitivity of the results. Conventional local linear regression robust standard errors clustered at the province level are reported in parentheses.
*** Indicates $p<0.01$.
* Indicates $p<0.05$.
* Indicates $p<0.1$.

6.4. Potential pollution substitution and relocation across regions

Another concern is that firms may offset water pollution by increasing other forms of pollution such as air. If this is the case, central supervision is not an effective policy because it merely results in pollution substitution (Greenstone, 2003). To address this critical issue, we use sulfur dioxide (SO_2), one of the most common air pollutants, as the dependent variable to check whether the NSM firms systematically increased their air pollution. The results are reported in Table 6.

In Table 6, estimates from most of the regressions are insignificant. The results show that the SO_2 emitted by NSM firms was not systematically higher than that from non-NSMF firms. In contrast, NSM firms may even have produced lower SO_2 emissions when a large bandwidth is used.[19] This result is in line with our previous results on firms' behavior, as NSM firms essentially reduced their COD emissions through end-of-pipe wastewater treatment rather than by changing production. The substitution of COD by other forms of pollutants is therefore less likely.

Nevertheless, given the different levels of environmental enforcement between regions, NSM firms may redistribute their pollution across regions, especially firms with plants in multiple locations

[19] Large bandwidth involves more heterogeneous firms. The result should be interpreted with caution.

Table 6
Non-parametric RD estimates of the impact of the NSMF program on firm SO_2 emissions.

	(1)	(2)	(3)	(4)	(5)	(6)
Dep var.	$Ln(SO_2)$			$\triangle Ln(SO_2)$		
Bandwidth	50%CCT	100%CCT	200%CCT	50%CCT	100%CCT	200%CCT
2006	−0.378	−0.428	−0.784**	−0.527*	−0.262	−0.168
	(0.660)	(0.409)	(0.308)	(0.304)	(0.181)	(0.146)
Bandwidth	39.6075	79.215	158.43	33.5785	67.157	134.314
Obs.	324	716	2499	267	577	1673
2007	0.0501	−0.279	−0.504*	0.226	0.203	0.228
	(0.544)	(0.400)	(0.292)	(0.341)	(0.222)	(0.159)
Bandwidth	43.0525	86.105	172.21	35.332	70.664	141.328
Obs.	350	760	3235	265	557	1759
2008	−0.539	−0.456	−0.548*	−0.124	−0.0122	0.114
	(0.609)	(0.409)	(0.318)	(0.494)	(0.285)	(0.175)
Bandwidth	38.8955	77.791	155.582	34.4935	68.987	137.974
Obs.	304	667	2310	249	533	1601
2009	−0.0243	−0.342	−0.755***	−0.275	0.0415	−0.0544
	(0.578)	(0.343)	(0.267)	(0.520)	(0.371)	(0.201)
Bandwidth	39.457	78.914	157.828	28.9055	57.811	115.622
Obs.	338	721	2514	199	429	1127

Note: RD estimates with the non-parametric local linear regressions. The results of the future outcome model are reported in columns (1)–(3), and the results of the outcome change model are reported in columns (4)–(6). The cutoff point is 199 metric tons COD. The optimal bandwidth is selected using the Calonico et al. (2014) approach. Half and double of the CCT bandwidth are employed to check the sensitivity of the results. Conventional local linear regression robust standard errors clustered at the province level are reported in parentheses.

*** Indicates $p<0.01$.

** Indicates $p<0.05$.

* Indicates $p<0.1$.

(Gibson, 2016). To check potential pollution relocation, we divide the full sample into two subsamples, i.e., coastal and non-coastal regions, and repeat the analysis.[20] We report the graphical evidence in Fig. 6 and the non-parametric RD estimates in Table 7.

As shown in Fig. 6, the NSMF program had a similar impact on water pollution reduction in both coastal and non-coastal regions. Central supervision reduced firms' COD emissions across the country, and its impact increased over time. This graphical evidence sheds light on overall pollution reduction and rejects the hypothesis of pollution relocation across regions under uniform central supervision.

The results in Table 7 once again reject the hypothesis of pollution transfer across regions. The NSMF program had a negative and significant impact on firms' COD emissions in coastal and non-coastal regions. To interpret the results, one can read the bounded coefficients[21]. Specifically, the NSMF program reduced firms' COD emissions by approximately 73% in 2008 and 52% in 2009 in the coastal region. In the non-coastal region, the impact of the program has been more substantially delayed. It was only 2009 when the program appeared to reduce firms' COD emissions by approximately 59%. This fact is probably explained by more serious environmental problems in the former. According to our previous study, more than 60% of industrial polluting firms are located in China's coastal regions, which signifies that pollution levels are much higher there (Wu et al., 2017). These results imply a substantial impact in the Chinese context, where local environmental regulation was weakly enforced in general and firms' violations were severe. According to a technical report by China's MEP, room for firms' pollution reduction remains large. If all firms in major polluting industries were to meet the environmental standard, pollution would reduce by 55.1% for SO2 and 41.5% for COD in China[22]. In the case of the NSMF program, firms need to reduce their pollution emissions by half or even more to meet the environmental standard under enhanced central supervision. Our results thus confirm the critical situation of weak local environmental regulation in China and the substantial room for improvement via central supervision.

6.5. Additional robustness checks

We conduct further robustness checks to ensure the validity of the RD analysis and check the sensitivity of our results. First, as noted by Barreca et al. (2016), if attributes related to outcomes predict heaping in the running variable, the usual RD analysis may arrive at biased estimates. The "Donut RD" is thus proposed to correct for the bias. In our case, if a few firms heap at the cutoff and shape the linear estimate, our results may be driven by these firms. To rule out this concern, we adopt the "Donut RD" to drop 25 firms within small bandwidth (± 3 metric tons COD) around the cutoff point.[23] To remain consistent, the estimation method is non-parametric local linear regression and the bandwidths are chosen using CCT as in the previous analysis. The results are provided in Table 8.

As noted, after dropping these firms around the cutoff, our main results remain consistent and robust.The estimates using the smallest bandwidth (50% CCT) become non-significant for the future outcome model (column 1). The result is probably due to the overly small sample size after dropping the data heaps. The results confirm that the impact of the NSMF program is indeed substantial, which is not driven by only a few firms near the cutoff.

Second, another concern surrounding the cutoff point is that the estimated impact is simply due to the fact that heavy polluters can reduce pollution emissions much more easily.[24] If this is the case, one may observe the impact with alternative cutoff points. To address this concern, we use fake cutoff points created by halving and doubling the original value and repeat the RD analysis. The results are reported in Table 9.[25]

[20] Further divisions could not be created because the sample size would be too small.

[21] As our estimates are in log points, a value below −1 is meaningless in the economic sense. Therefore, we need to rescale the coefficients to be bounded above −1 using the formula $\exp(\alpha) - 1$.

[22] Readers can find the report in the Chinese official press: http://legal.people.com.cn/n/2015/0720/c188502-27327691.html (in Chinese).

[23] We also drop more suspicious firms within larger bandwidths (± 5 metric tons COD) around the cutoff point. The results are similar and are available upon request.

[24] We thank an anonymous reviewer for comments on this critical issue.

[25] To save space, we present only the results of non-parametric estimates of the future outcome model. Other regressions yield similar results, which are available upon request.

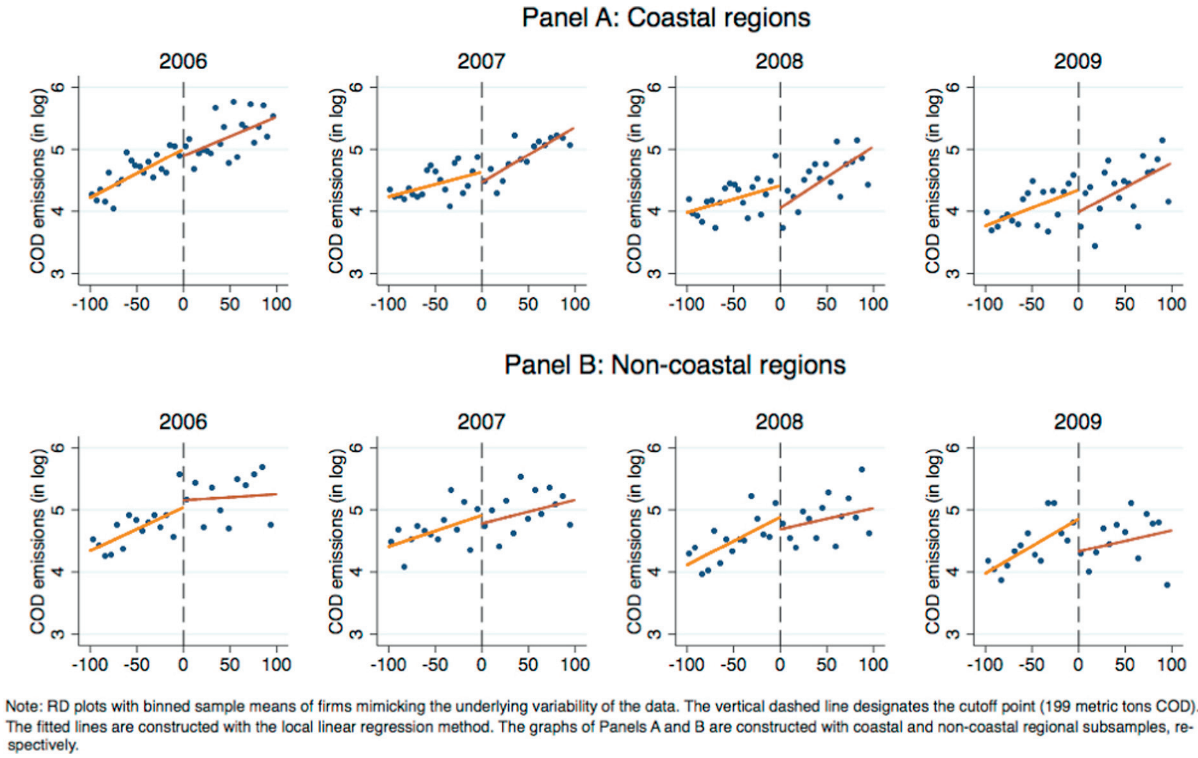

Fig. 6. The impact of central supervision across regions.

Table 7
The impact of the NSMF program across regions.

	(1)	(2)	(3)	(4)	(5)	(6)
Dep var.	*Ln*(*COD*)					
Region	Coastal region			Non-coastal region		
Bandwidth	50%CCT	100%CCT	200%CCT	50%CCT	100%CCT	200%CCT
2006	1.015	0.160	−0.507	−0.640	0.307	0.403
	(0.626)	(0.476)	(0.321)	(0.552)	(0.294)	(0.260)
Bandwidth	8.604	17.208	34.416	17.549	35.098	70.196
Obs.	40	79	202	93	194	402
2007	−0.553	−0.383	−0.262**	−0.529	0.0512	−0.148
	(0.650)	(0.255)	(0.115)	(0.833)	(0.378)	(0.310)
Bandwidth	28.189	56.378	112.756	19.1555	38.311	76.622
Obs.	159	363	957	102	208	450
2008	−1.503**	−1.310***	−0.693**	−0.504	−0.0784	−0.358
	(0.716)	(0.452)	(0.276)	(0.854)	(0.411)	(0.358)
Bandwidth	21.735	43.47	86.94	18.7465	37.493	74.986
Obs.	117	277	632	98	201	437
2009	−1.160**	−0.736**	−0.454**	−0.666	−0.880*	−0.963***
	(0.579)	(0.304)	(0.191)	(0.744)	(0.509)	(0.359)
Bandwidth	30.113	60.226	120.452	25.766	51.532	103.064
Obs.	169	394	1108	138	278	640

Note: RD estimates with the non-parametric local linear regressions. The results of the future outcome model are reported in columns (1)–(3), and the results of the outcome change model are reported in columns (4)–(6). The cutoff point is 199 metric tons COD. The optimal bandwidth is selected using the Calonico et al. (2014) approach. Half and double of the CCT bandwidth are employed to check the sensitivity of the results. Conventional local linear regression robust standard errors clustered at the province level are reported in parentheses.
*** Indicates p<0.01.
** Indicates p<0.05.
* Indicates p<0.1.

It can be seen that even though the coefficients are large in numbers, none of the alternative cutoff points yield significant results[26]. This result confirms the unique cutoff point in our RD design. Along with the rigorous RD design, this check rules out alternative explanations of water pollution reduction and reduces it to the impact of the NSMF program.

Third, to establish the internal validity of the RD design, one critical assumption is made about the continuity of firms' characteristics at the cutoff point: firms should be similar in all aspects except their exposure to treatment. Although we restricted our analysis to a small bandwidth around the cutoff, we nevertheless formally test the

[26] In the fuzzy RD model, the treatment effect is scaled by the difference in the probability of treatment. At the fake cutoff, the probability of treatment is continuous so that the difference is very small. That is the reason for the large coefficients in most of the cases.

Table 8
Robustness check with the Donut RD method.

	(1)	(2)	(3)	(4)	(5)	(6)
Dep var.	$Ln(COD)$			$\triangle Ln(COD)$		
Bandwidth	50%CCT	100%CCT	200%CCT	50%CCT	100%CCT	200%CCT
2006	−0.661	−0.181	−0.0196	−0.208	−0.0449	−0.0693
	(0.679)	(0.417)	(0.212)	(0.516)	(0.289)	(0.159)
Bandwidth	18.777	37.554	75.108	29	58	116
Obs.	158	404	928	289	670	1748
2007	−0.791	−0.300	−0.276	−0.110	−0.217	−0.183
	(0.959)	(0.381)	(0.212)	(0.443)	(0.252)	(0.177)
Bandwidth	21.2145	42.429	84.858	39.974	79.948	159.896
Obs.	202	474	1098	444	1020	3542
2008	−1.484	−0.883**	−0.598**	−0.929**	−0.756***	−0.428**
	(0.918)	(0.404)	(0.237)	(0.457)	(0.282)	(0.189)
Bandwidth	22.215	44.43	88.86	27.087	54.174	108.348
Obs.	213	499	1149	272	619	1542
2009	−1.346	−1.102***	−0.762***	−0.956**	−0.928**	−0.589**
	(0.823)	(0.369)	(0.244)	(0.487)	(0.367)	(0.241)
Bandwidth	26.1495	52.299	104.598	32.563	65.126	130.252
Obs.	259	595	1475	340	775	2180

Note: Donut RD estimates by dropping observations within ±3 metric tons COD around the cutoff point. The results of the future outcome model are reported in columns (1)–(3), and the results of the outcome change model are reported in columns (4)–(6). The cutoff point is 199 metric tons COD. The optimal bandwidth is selected using the approach of Calonico et al. (2014). Half and double the CCT bandwidth are used to check the sensitivity of the results. Conventional local linear regression robust standard errors clustered at the province level are reported in parentheses.
*** Indicates $p<0.01$.
* Indicates $p<0.05$.
* Indicates $p<0.1$.

assumption for some critical characteristics. To this end, we run the local linear regression using four characteristics of firms (i.e., output, NH3-N emissions, wastewater treatment equipment and capacity) as dependent variables. Any significant results may indicate a violation of the assumption.

Table 10 reports the non-parametric estimation results from the test. As expected, none of these characteristics are significantly different around the cutoff. These results are in accordance with our graphical analysis (c.f., Fig. 4) and support the internal validity of the RD design in our case.

A final check is conducted on the multiple running variables. As discussed in Section 2, NSM water-polluting firms are designated on the basis of two pollutants: COD and NH3-N. Although the NH3-N levels of NSM and non-NSM firms are almost the same at the COD cutoff point, as shown by our graphical analysis and regression results, the potential selection bias due to the NH3-N cutoff remains a concern. In the literature, RD designs with multiple running variables are estimated using methods such as the frontier, centering, univariate and IV approaches (see Wong et al., 2013 for a review). Since we are interested in the separate treatment effect on firms' COD

Table 9
Robustness check with alternative cutoff points.

	(1)	(2)	(3)	(4)	(5)	(6)
Dep var.	$Ln(COD)$					
Cutoff point	50%			200%		
Bandwidth	50%CCT	100%CCT	200%CCT	50%CCT	100%CCT	200%CCT
2006	3.437	13.65	2.533	−2.826	−43.28	−91.65
	(4.141)	(36.40)	(4.501)	(4.307)	(255.0)	(1038)
Bandwidth	9.4375	18.875	37.75	25.5665	51.133	102.266
Obs.	96	185	434	277	611	1445
2007	−2.887	−3.391	−7.717	−3.140	−33.39	−64.71
	(6.516)	(3.190)	(6.047)	(4.164)	(222.4)	(783.5)
Bandwidth	10.992	21.984	43.968	25.833	51.666	103.332
Obs.	110	235	522	278	615	1467
2008	−10.21	−42.10	−3.638	−0.303	−13.57	80.87
	(9.439)	(76.06)	(4.511)	(1.628)	(241.2)	(2358)
Bandwidth	9.307	18.614	37.228	29.211	58.422	116.844
Obs.	95	180	425	324	701	1797
2009	−6.875	−24.75	0.107	−0.801	−24.26	63.19
	(7.703)	(37.84)	(3.648)	(1.824)	(420.2)	(1336)
Bandwidth	8.851	17.702	35.404	28.875	57.75	115.5
Obs.	88	175	408	314	690	1761

Note: RD estimates with the non-parametric local linear regressions. The results of the future outcome model are reported in columns (1)–(3), and the results of the outcome change model are reported in columns (4)–(6). The cutoff point is 100 metric tons COD for 0.5 and 400 metric tons COD for 2. The optimal bandwidth is selected using the Calonico et al. (2014) approach. Half and double of the CCT bandwidth are employed to check the sensitivity of the results. Conventional local linear regression robust standard errors clustered at the province level are reported in parentheses.
*** Indicates $p<0.01$.
* Indicates $p<0.05$.
* Indicates $p<0.1$.

Table 10
Robustness check of similar firm characteristics near the cutoff point.

	(1)	(2)	(3)	(4)
Dep Var.	*Ln(Output)*	*Ln(Equipment)*	*Ln(Capacity)*	*Ln(NH3 − N)*
2006	0.433	0.0427	0.103	−0.240
	(0.430)	(0.108)	(0.653)	(0.398)
Bandwidth	74.521	64.483	43.893	82.877
No of Obs.	936	721	473	848
2007	0.619	−0.0126	0.324	0.192
	(0.472)	(0.103)	(0.580)	(0.363)
Bandwidth	57.202	66.021	47.546	80.749
No of Obs.	684	723	495	810

Note: Estimates from the non-parametric local linear regressions. The cutoff point is 199 metric tons COD. The optimal bandwidth is selected using the Calonico et al. (2014) approach. Standard errors in parentheses.
*** Indicates $p<0.01$.
* Indicates $p<0.05$.
* Indicates $p<0.1$.

emissions, we adopt a univariate approach to drop all observations of NSM water-polluting firms that cross the threshold of NH3-N. This approach simplifies the analysis and serves as a robustness check on our previous results.

Table 11 reports the results and clearly shows that the results support our main findings. The impact of the NSMF program was insignificant in 2006 and 2007 but become more significant in 2008 and 2009. In sum, these additional robustness checks lend confidence to our conclusion that direct central supervision enhances local environmental enforcement; hence it reduces industrial water pollution emissions and yields substantial marginal benefits for China's environmental decentralization.

7. Discussion and conclusion

By devolving environmental authority from the central government to local and regional governments, China has de facto decentralized its environmental regulation. However, weak local environmental enforcement has cast doubts on its environmental decentralization and led to calls for a reform of the current system. In this context, the NSMF program imposes direct central supervision on polluting firms to complement local environmental regulations and thus provides a unique combination of central supervision and environmental decentralization for study. This paper relies on the natural experiment emerging from the NSMF program and uses an RD design to evaluate the efficacy of central supervision for improving local environmental enforcement. The results show that only a year after its implementation, the NSMF program had significantly reduced industrial COD emissions by at least 26.8%, with even greater reductions since. The robust and lasting impact of the program suggests that substantial room remains to improve environmental decentralization via central supervision in China.

Based on the experiences of developed countries, we observe that the evolution of environmental federalism has oscillated between periods of more centralized and more decentralized control. In the US, most environmental responsibilities were left to state governments until 1969, when national standards were rapidly established for almost all forms of air and water pollution. With centralized regulation, many indicators of environmental quality improved significantly in the following 25 years. This centralized period lasted until the 1980s, when many state regulations became more stringent or comprehensive than those of the federal government, and environmental authority was once again delegated to the states. The evolution of environmental federalism clearly depends on a number of factors such as the country's level of development, the institutional context, and the specificity of policy areas (Fredriksson and Wollscheid, 2014; Vogel et al., 2010). In developing countries such as China, where local economic interests continue to take precedent over environmental protection, problems of collusion and interjurisdictional externalities persist (Cai et al., 2015; Kahn et al., 2015), and in such a context, environmental decentralization appears to be ineffective at addressing critical environmental problems.

Having acknowledged the problem of environmental decentralization, a reform could entail enhanced supervision by the central government. Based on the experience of the NSMF program, the MEP has announced the implementation of a new vertical administration system to supervise environmental regulation within the framework of China's 13th Five-Year Plan (2016–2020). In this new system, environmental monitoring and inspection will be strengthened from the hierarchy and EPBs will report directly to provincial instead of local governments.

Table 11
Robustness check by dropping key firms across the cutoff point of NH3-N.

	(1)	(2)	(3)	(4)	(5)	(6)
Dep Var.	*Ln(COD)*			*△Ln(COD)*		
Bandwidth	50%CCT	100%CCT	200%CCT	50%CCT	100%CCT	200%CCT
2006	−0.518	−0.111	0.0201	−0.152	−0.0241	−0.0431
	(0.413)	(0.219)	(0.172)	(0.232)	(0.192)	(0.144)
Bandwidth	16.47	32.94	65.88	28.1635	56.327	112.654
Obs.	146	342	751	284	626	1585
2007	−0.947	−0.420	−0.346*	−0.308	−0.345*	−0.277
	(0.650)	(0.274)	(0.191)	(0.269)	(0.195)	(0.169)
Bandwidth	19.464	38.928	77.856	38.318	76.636	153.272
Obs.	185	419	919	410	907	3053
2008	−1.312**	−0.898***	−0.687***	−0.932***	−0.776***	−0.531***
	(0.563)	(0.311)	(0.236)	(0.351)	(0.244)	(0.182)
Bandwidth	20.2255	40.451	80.902	26.6595	53.319	106.638
Obs.	195	443	990	267	592	1441
2009	−1.379**	−1.135***	−0.862***	−1.033***	−1.002***	−0.711***
	(0.557)	(0.301)	(0.237)	(0.364)	(0.324)	(0.226)
Bandwidth	23.8815	47.763	95.526	31.324	62.648	125.296
Obs.	242	514	1217	324	707	1954

Note: RD estimates with the non-parametric local linear regressions. The results of the future outcome model are reported in columns (1)-(3), and the results of the outcome change model are reported in columns (4)-(6). The cutoff point is 199 metric tons COD. The optimal bandwidth is selected using the Calonico et al. (2014) approach. Half and double of the CCT bandwidth are employed to check the sensitivity of the results. Conventional local linear regression robust standard errors clustered at the province level are reported in parentheses.
*** Indicates $p<0.01$.
** Indicates $p<0.05$.
* Indicates $p<0.1$.

To incentivize local governments to embrace reform, we recommend a more flexible regime of environmental decentralization whereby authority and responsibility are allocated on the basis of local environmental performance. For localities with serious environmental problems and a weak capacity for regulation, more public investment in the construction of in time pollution monitoring facilities and strengthened supervision from the upper hierarchy is appropriate. In extreme cases, the central government should overrule the local environmental authority and implement more stringent regulations in its place. In contrast, localities with strong regulatory capacity and superior environmental performance should be granted greater autonomy to determine local environmental standards and conduct relevant regulatory activities. In both cases, more reliable pollution information will need to be disclosed and made accessible to the public because a more accurate and transparent environmental information system is at the core of the successful environmental decentralization.

Acknowledgment

This research was supported by the National Science Foundation of China (Grant No. 71433007).

Appendix A

Table A1
The 2007 list of national key firms.

Province/municipality	NSM water-polluting firms	NSM air-polluting firms	NSM sewage treatment plants
Coastal regions			
Beijing	7	9	18
Tianjin	25	19	9
Hebei	313	205	29
Shanghai	22	17	32
Jiangsu	266	207	91
Zhejiang	198	100	43
Fujian	57	52	23
Hainan	18	4	2
Guangdong	61	86	60
Shandong	201	270	81
Non-coastal regions			
Liaoning	101	165	19
Jilin	59	81	6
Heilongjiang	77	109	6
Henan	218	109	26
Hubei	153	136	20
Hunan	184	315	18
Anhui	96	102	16
Jiangxi	81	138	3
Shanxi	168	403	14
Guangxi	275	203	7
Chongqing	30	30	22
Sichuan	133	217	21
Guizhou	16	72	5
Yunnan	43	31	25
Inner Mongolia	68	165	15
Shaanxi	106	139	9
Gansu	33	49	9
Qinghai	18	33	1
Ningxia	39	46	6
Xinjiang	49	80	22
Total	3115	3592	658

Source: MEP (2007).

Table A2
Variable definitions and sources.

Variable	Definition and measure	Year	Source
Key	The dummy variable indicating the status of the firm. "1" if the firm is a key state-monitored water polluting firm."0" otherwise.	2007–2009	MEP
NH3-N	The amount of NH_3-N discharged by firm in metric tons.	2005–2009	CES
COD	The amount of COD emitted by firm in metric tons.	2005–2009	CES
CODG	The amount of COD generated by firm in metric tons.	2005–2009	CES
SO2	The amount of SO_2 emitted by firm in metric tons.	2005–2009	CES
Output	The firm annual industrial output in multiples of 10,000 yuan.	2005–2009	CES
Equipment	The number of wastewater treatment facilities the firm owns.	2005–2009	CES
Capacity	The wastewater treatment capacity of the firm in metric tons/day.	2005–2009	CES
Industry	The two-digit SIC code of the industry to which the firm belongs.	2005–2009	CES
Province	The province in which the firm is registered.	2005–2009	CES

Table A3
Summary statistics by years and by status of NSM water-polluting firm.

	Non-NSM firm					NSM firm					
Variable	Obs	Mean	Std.dev.	Min	Max	Obs	Mean	Std.dev.	Min	Max	MeanDiff
2005											
COD2005	18,964	29.04	145.30	0	7247	1643	1203	2476	0.06	36,688	−1.20e+03***
CODG2005	18,964	148.40	1121	0	100,006	1643	4363	14,770	0.35	345,590	−4.20e+03***
Nh3-N2005	18,964	1.80	16.85	0	892.30	1643	133.30	553.90	0	17,162	−131.50***
SO22005	18,677	74.16	507.30	0	28,782	1603	1266	5159	0	102,924	−1.20e+03***
Output2005	18,964	19,415	111,192	0	7.64e+06	1643	128,356	504,180	0	1.00e+07	−1.10e+05***
Equipment2005	18,964	1.08	1.65	0	68	1643	3.30	6.35	0	106	−2.22***
Capacity2005	18,964	1839	19,020	0	1.48e+06	1643	31,132	150,858	0	2.96e+06	−2.90e+04***
2006											
COD2006	18,964	35.54	162.60	0	5652	1643	1127	2542	0.20	36939	−1.10e+03***
CODG2006	12,928	234.50	1148	0	50,219	1355	4797	11,750	0.48	172,440	−4.60e+03***
Nh3-N2006	13,282	3.29	22.63	0	1350	1316	120.1	310.80	0	3700	−116.82***
SO22006	12,949	94.42	563.20	0	30,318	1204	1310	4622	0	78,130	−1.20e+03***
Output2006	18,854	24,553	138,219	0	8.56e+06	1633	149,030	582,166	0	1.22e+07	−1.20e+05***
Equipment2006	15,732	1.34	1.87	0	104	1563	3.34	5.96	0	65	−2.00***
Capacity2006	15,756	2639	27,020	0	1.56e+06	1564	36,028	167,359	0	3.58e+06	−3.30e+04***
2007											
COD2007	18,964	36.74	145.80	0	5750	1643	1020	2448	0.07	38,620	−983.13***
CODG2007	12,364	235	1082	0	61,405	1327	4767	13,931	0.07	292,008	−4.50e+03***
Nh3-N2007	13,098	3.26	17.05	0	568	1301	92.70	236.30	0	2979	−89.43***
SO22007	12,476	96.18	606.60	0	39,735	1188	1332	4396	0	66,023	−1.20e+03***
Output2007	18,801	29,491	190,496	0	1.32e+07	1637	186,282	760,358	0	1.65e+07	−1.60e+05***
Equipment2007	15,539	1.45	8.91	0	950	1553	3.41	7.06	0	131	−1.96***
Capacity2007	15,573	3138	36,725	0	2.12e+06	1559	44,379	225,991	0	4.41e+06	−4.10e+04***
2008											
COD2008	18,964	34.10	134.20	0	5660	1643	878.10	2329	0.01	44,073	−844.00***
CODG2008	13,279	251.30	1083	0	42,820	1380	4739	14262	0.07	290,376	−4.50e+03***
Nh3-N2008	13,857	2.85	13.99	0	568	1357	70.86	179.80	0	2020	−68.01***
SO22008	12,056	94.06	575.50	0	33,740	1167	1259	4125	0	51,415	−1.20e+03***
Output2008	18,866	32,344	175,741	0	7.78e+06	1641	194,706	816,931	0.10	1.94e+07	−1.60e+05***
Equipment2008	15,660	1.45	16.05	0	2000	1566	3.19	5.82	0	67	−1.73***
Capacity2008	15,702	2755	26,949	0	1.32e+06	1570	44,363	274,934	0	7.72e+06	−4.20e+04***
2009											
COD2009	18,964	33.17	138.30	0	6872	1643	790.40	2060	0.05	31,732	−757.24***
CODG2009	14,048	230.40	1093	0	69,839	1388	4651	14,768	0.57	290,375	−4.40e+03***
Nh3-N2009	15,253	2.53	12.95	0	430.20	1475	53.32	148.10	0	2055	−50.79***
SO22009	12,500	87.31	523.90	0	24,544	1191	1259	4218	0	50,894	−1.20e+03***
Output2009	18,884	34,578	209,260	0	1.01e+07	1640	208,177	1.07e+06	1	2.69e+07	−1.70e+05***
Equipment2009	16,148	1.31	3.18	0	350	1590	3.09	6.33	0	87	−1.79***
Capacity2009	16,197	2986	33,894	0	2.20e+06	1592	46,507	269,416	0	6.17e+06	−4.40e+04***

Note: The mean differences of all variables are significant at 99%. See Table A2 in the appendix for the definition of variables.
*** Indicates $p<0.01$.
* Indicates $p<0.05$.
* Indicates $p<0.1$.

Table A4
Check of parametric RD estimates using different bandwidths.

Dep. var.	*Ln*(*COD*)							
Bandwidth	50%				200%			
Year	2006	2007	2008	2009	2006	2007	2008	2009
NSM_firm	−0.144	0.135	−0.594**	−0.765***	0.177	−0.878***	−0.855***	−0.979***
	(0.359)	(0.274)	(0.283)	(0.263)	(0.142)	(0.273)	(0.243)	(0.257)
COD2005	0.004	0.002	0.011**	0.011***	−0.001**	0.006***	0.005***	0.006***
	(0.005)	(0.005)	(0.005)	(0.004)	(0.001)	(0.001)	(0.001)	(0.001)
COD2005_sq	−2.12e−05	1.16e−05	6.08e−05	9.93e−05	2.03e−05***	−6.25e−05***	−5.60e−05***	−5.30e−05***
	(6.25e−05)	(0.00011)	(8.38e−05)	(8.55e−05)	(2.27e−06)	(4.25e−06)	(4.22e−06)	(3.58e−06)
Constant	4.183***	1.925***	2.426***	2.464***	0.176**	6.266***	5.455***	5.147***
	(0.369)	(0.349)	(0.256)	(0.292)	(0.076)	(0.594)	(0.482)	(0.312)
Province FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control of firm characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	416	401	440	491	11,674	11,674	12,036	13,668

(*continued on next page*)

Table A4 (*continued*)

Dep. var.	*Ln*(*COD*)							
Bandwidth	50%				200%			
Year	2006	2007	2008	2009	2006	2007	2008	2009
F-test of IV	73.95	73.47	139.07	129.69	284.79	201.86	224.34	300.22
Prob > F	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
R-squared	0.255	0.291	0.319	0.242	0.100	0.516	0.497	0.464

Note: RD estimates with the parametric regression of 2SLS models. The cutoff point is 199 metric tons COD. The results using half of original bandwidth (50 metric ton COD) are reported in columns (1)–(4), and the results using double of original bandwidth (200 metric tons COD) are reported in columns (5)–(8). Conventional robust standard errors clustered at the province level are reported in parentheses.

*** Indicates $p<0.01$.

* Indicates $p<0.05$.

* Indicates $p<0.1$.

Table A5

First-stage regression of the parametric estimation.

Dep var.	NSM_firm							
Year	2006	2007	2008	2009	2006	2007	2008	2009
Eligibility	0.729***	0.728***	0.751***	0.752***	0.729***	0.728***	0.751***	0.752***
	(0.047)	(0.036)	(0.039)	(0.036)	(0.047)	(0.036)	(0.039)	(0.036)
COD2005	−1.09e−04	−1.10e−04	−2.74 e−04	−1.62e−04	−1.09e−04	−1.10e−04	−2.74e−04	−1.62e−04
	(4.12e−04)	(3.03e−04)	(2.96e−04)	(2.70e−04)	(4.12e−04)	(3.03e−04)	(2.96e−04)	(2.70e−04)
COD2005_sq	−8.14e−06*	−8.21e−06**	−8.43e−06**	−6.74e−06**	−8.14e−06*	−8.21e−06**	−8.43e−06**	−6.74e−06**
	(4.56e−06)	(3.96e−06)	(3.54e−06)	(3.28e−06)	(4.56e−06)	(3.96e−06)	(3.54e−06)	(3.28e−06)
Constant	0.342***	0.396***	0.380***	0.353***	0.342***	0.396***	0.380***	0.353***
	(0.063)	(0.057)	(0.050)	(0.074)	(0.063)	(0.057)	(0.050)	(0.074)
Province FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control of firm								
Characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	947	915	985	1088	947	915	985	1088
R-squared	0.619	0.599	0.614	0.621	0.619	0.599	0.614	0.621

Note: RD estimates with the parametric regression of 2SLS models. The results of future outcome model are reported in columns (1)–(4), and the results of outcome change model are reported in columns (5)–(8). The cutoff point is 199 metric tons COD. Conventional robust standard errors clustered at province level are reported in parentheses.

*** Indicates $p<0.01$.

** Indicates $p<0.05$.

* Indicates $p<0.1$.

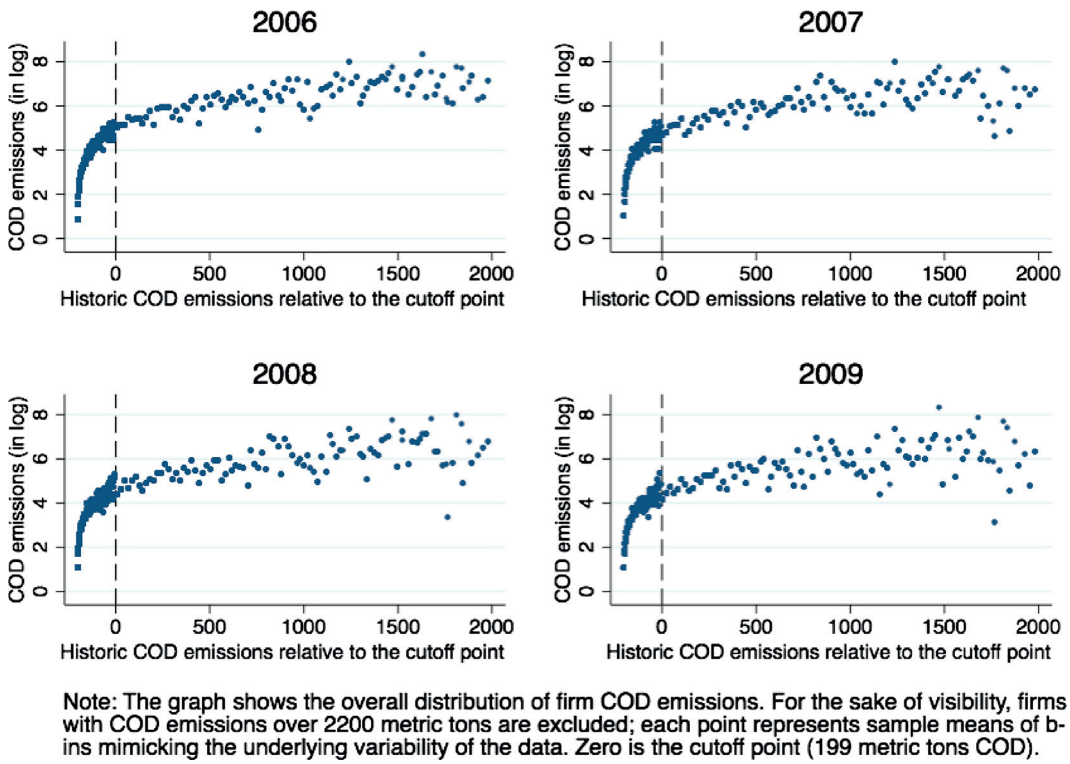

Fig. A1. Overview of firm COD emissions over year.

References

Anderson, T.L., Hill, P.J., 1997. Environmental Federalism. Rowman & Littlefield Publishers.

Barreca, A.I., Lindo, J.M., Waddell, G.R., 2016. Heaping-induced bias in regression-discontinuity designs. Econ. Inq. 54 (1), 268–293.

Becker, G.S., 1968. Crime and Punishment: An Economic Approach. Palgrave Macmillan, UK. , pp. 13–68. chap. The Economic Dimensions of Crime.

Bulte, E., List, J.A., Strazicich, M.C., 2007. Regulatory federalism and the distribution of air pollutant emission. J. Reg. Sci. 47 (1), 155–178.

Cai, H., Chen, Y., Gong, Q., 2015. Polluting thy neighbor: unintended consequences of China's pollution reduction mandates. J. Environ. Econ. Manag. 76, 86–104.

Calonico, S., Cattaneo, M.D., Titiunik, R., 2014. Robust nonparametric confidence intervals for regression discontinuity designs. Econometrica 82 (6), 2295–2326.

Dasgupta, S., Wang, H., Wheeler, D., 1997. Surviving Success: Policy Reform and the Future of Industrial Pollution in China. World Bank Publications.

Duflo, E., Greenstone, M., Pande, R., Ryan, N., 2013. Truth-telling by third-party auditors and the response of polluting firms: experimental evidence from India. Q. J. Econ. 1499 (1545), 1499.

Duflo, E., Greenstone, M., Pande, R., Ryan, N., 2014. The Value of Regulatory Discretion: Estimates From Environmental Inspections in India. National Bureau of Economic Research.

Engel, K.H., 1996. State environmental standard-setting: is there a race and is it to the bottom. Hast. Law J. 48, 271.

Fan, C.S., Lin, C., Treisman, D., 2009. Political decentralization and corruption: evidence from around the world. J. Public Econ. 93 (1-2), 14–34.

Fredriksson, P.G., Mani, M., Wollscheid, J., 2006. Environmental Federalism: A Panacea or Pandora's Box for Developing Countries? World Bank Publications.

Fredriksson, P.G., Millimet, D.L., 2002. Strategic interaction and the determination of environmental policy across US states. J. Urban Econ. 51 (1), 101–122.

Fredriksson, P.G., Wollscheid, J.R., 2014. Environmental decentralization and political centralization. Ecol. Econ. 107, 402–410.

Gelman, A., Imbens, G., 2014. Why High-order Polynomials Should Not Be Used in Regression Discontinuity Designs. National Bureau of Economic Research.

Gibson, M., 2016. Regulation-induced Pollution Substitution. University of California at San Diego Economics Working Paper.

Golding, W.F., 2011. Incentives for change: China's cadre system applied to water quality. Pac. Rim Law Policy J. 20, 399–428.

Gray, W.B., Shadbegian, R.J., 2004. "Optimal" pollution abatement-whose benefits matter, and how much? J. Environ. Econ. Manag. 47 (3), 510–534.

Greenstone, M., 2003. Estimating regulation-induced substitution: the effect of the clean air act on water and ground pollution. Am. Econ. Rev. 93 (2), 442–448.

Hanna, R.N., Oliva, P., 2010. The impact of inspections on plant-level air emissions. BE J. Econ. Anal. Policy 10 (1).

Huang, Y., 1996. Central-local relations in China during the reform era: the economic and institutional dimensions. World Dev. 24 (4), 655–672.

Imbens, G., Kalyanaraman, K., 2012. Optimal bandwidth choice for the regression discontinuity estimator. Rev. Econ. Stud. 79 (3), 933–959.

Jacob, R.T., Zhu, P., Somers, M.-A., Bloom, H.S., 2012. A Practical Guide to Regression Discontinuity. Citeseer.

Jia, R., Nie, H., 2017. Decentralization, collusion and coalmine deaths. Rev. Econ. Stat. 99 (1), 105–118.

Jin, H., Qian, Y., Weingast, B.R., 2005. Regional decentralization and fiscal incentives: federalism, Chinese style. J. Public Econ. 89 (9), 1719–1742.

Kahn, M., Li, P., Zhao, D., 2015. Pollution control effort at China's river borders: when does free riding cease? Am. Econ. J. Econ. Policy 7 (4), 223–242.

Konisky, D.M., 2007. Regulatory competition and environmental enforcement: is there a race to the bottom? Am. J. Polit. Sci. 51 (4), 853–872.

Konisky, D.M., Woods, N.D., 2012. Environmental free riding in state water pollution enforcement. State Polit. Policy Q. 12 (3), 227–251.

Kunce, M., Shogren, J.F., 2005. On interjurisdictional competition and environmental federalism. J. Environ. Econ. Manag. 50 (1), 212–224.

Lee, D.S., Lemieux, T., 2010. Regression discontinuity designs in economics. J. Econ. Lit. 48 (2), 281–355.

Levinson, A., 1997. A note on environmental federalism: interpreting some contradictory results. J. Environ. Econ. Manag. 33 (3), 359–366.

Levinson, A., 2003. Environmental regulatory competition: a status report and some new evidence. Natl. Tax J. 91–106.

List, J.A., Gerking, S., 2000. Regulatory federalism and environmental protection in the United States. J. Reg. Sci. 40 (3), 453–471.

Lo, C.W.-H., Fryxell, G.E., Wong, W.W.-H., 2006. Effective regulations with little effect? The antecedents of the perceptions of environmental officials on enforcement effectiveness in China. Environ. Manag. 38 (3), 388–410.

Ludwig, J., Miller, D.L., 2007. Does head start improve children's life chances? Evidence from a regression discontinuity design. Q. J. Econ. 122 (1), 159–208.

Ma, X., Ortolano, L., 2000. Environmental Regulation in China. Rowman & Littlefield Publishing Group, Lanham, MD.

Markusen, J.R., Morey, E.R., Olewiler, N., 1995. Competition in regional environmental policies when plant locations are endogenous. J. Public Econ. 56 (1), 55–77.

McAusland, C., 2003. Voting for pollution policy: the importance of income inequality and openness to trade. J. Int. Econ. 61 (2), 425–451.

McCrary, J., 2008. Manipulation of the running variable in the regression discontinuity design: a density test. J. Econ. 142 (2), 698–714.

MEP, 2007. Circular Regarding the Enhancement of the Environmental Monitoring and Statistics. Ministry of Environmental Protection, Beijing.

Millimet, D.L., 2014. Environmental federalism: a survey of the empirical literature. Case West. Res. Law Rev. 64 (4), 1669.

Oates, W.E., 1972. Fiscal Federalism. Harcourt Brace, New York.

Oates, W.E., 2001. A Reconsideration of Environmental Federalism. Resources for the Future, Washington, DC.

OECD, 2006. Environmental compliance and enforcement in China: an assessment of current practices and ways forward. Tech. rep. Organisation for Economic Co-operation and Development, Paris.

Revesz, R.L., 1997. Federalism and Environmental Regulation: A Normative Critique Hoover Institution, Stanford, California, pp. 13–68. chap. Federalism: Can the States be Trusted?.

Sigman, H., 2001. International Spillovers and Water Quality in Rivers: Do Countries Free Ride? National Bureau of Economic Research.

Sigman, H., 2005. Transboundary spillovers and decentralization of environmental policies. J. Environ. Econ. Manag. 50 (1), 82–101.

Sigman, H., 2007. Decentralization and Environmental Quality: An International Analysis of Water Pollution. National Bureau of Economic Research.

Sjöberg, E., 2016. An empirical study of federal law versus local environmental enforcement. J. Environ. Econ. Manag. 76, 14–31.

Thistlethwaite, D.L., Campbell, D.T., 1960. Regression-discontinuity analysis: an alternative to the ex post facto experiment. J. Educ. Psychol. 51 (6), 309.

Tiebout, C.M., 1956. A pure theory of local expenditures. J. Polit. Econ. 64 (5), 416–424.

Van Rooij, B., 2006. Implementation of Chinese environmental law: regular enforcement and political campaigns. Dev. Chang. 37 (1), 57–74.

Vogel, D., Toffel, M., Post, D., Uludere, N., 2010. Environmental Federalism in the European Union and the United States.

Wang, H., Mamingi, N., Laplante, B., Dasgupta, S., 2003. Incomplete enforcement of pollution regulation: bargaining power of Chinese factories. Environ. Resour. Econ. 24 (3), 245–262.

White, L.T., Landry, P.F., 2010. Decentralized authoritarianism in China: the communist party's control of local elites in the post-Mao era. J. Chin. Polit. Sci. 15 (3), 337–338.

Wong, V.C., Steiner, P.M., Cook, T.D., 2013. Analyzing regression-discontinuity designs with multiple assignment variables a comparative study of four estimation methods. J. Educ. Behav. Stat. 38 (2), 107–141.

WorldBank, 2000. World development report 1999-2000: entering the 21st century. Tech. rep. World Bank Group, Washington, DC.

Wu, H., Guo, H., Zhang, B., Bu, M., 2017. Westward movement of new polluting firms in China: pollution reduction mandates and location choice. J. Comp. Econ. 45 (1), 119–138.

Yang, D.L., 2017. China's illiberal regulatory state in comparative perspective. Chin. Polit. Sci. Rev. 2 (1), 114–133.

Zheng, Y., 2007. De Facto Federalism in China: Reforms and Dynamics of Central-Local Relations. World Scientific Publishing.