Indirect Value of Public Infrastructure Technology

Abstract

While prior research has primarily focused on the direct value of information technology (IT) and IT adoption by individuals and firms, this study explores the indirect value of IT in the form of public infrastructure technology. Exploiting a spatial discontinuity in water monitoring stations, we discover that firms located immediately upstream of water monitoring stations exhibit significantly lower levels of corruption than firms located immediately downstream. These findings are particularly noteworthy given that water monitoring stations have the potential to generate significant indirect value as they are not explicitly designed to mitigate corruption. Further analyses reveal that public infrastructure technology alone does not hold the key to mitigating corporate corruption. Instead, it is the synergistic interplay between public infrastructure technology and organizational change that drives ethical behaviors. These findings contribute to a deeper understanding of the broader IT value landscape, emphasizing the indirect value of technological advancements in public infrastructure that were not originally intended for such benefit.

Keywords: corruption; public infrastructure technology; IT value; organizational change; technology complementarity.

Jialu Liu, Siqi Pei, and Xiaoquan (Michael) Zhang, "Indirect Value of Public Infrastructure Technology," Management Science, forthcoming.

1. Introduction

In today's digitally interconnected world, the value of information technology (IT) has been extensively studied (Brynjolfsson 1996, Brynjolfsson and Hitt 1996, Melville et al. 2004), with a particular emphasis on its *direct* effects on domains like consumer welfare (Brynjolfsson 1996, Mithas et al. 2016, Babar and Burtch 2020, Liu et al. 2022), business value (Bharadwaj et al. 2007, Nevo and Wade 2010, Grover and Kohli 2012, Havakhor et al. 2019), or innovation (Gómez et al. 2017, Trantopoulos et al. 2017, Guo et al. 2022). However, the *indirect* value of IT, particularly in the realm of promoting ethical business practices, remains an underexplored area of research. The term "indirect value of IT" is used to describe situations where IT artifacts have unintended consequences or unanticipated effects, despite not being designed for such purposes. An illustrative example is that the ride-hailing platforms, while not designed to impact the number of rape incidents, have been found to have a dampening effect on such occurrences (Park et al. 2021). Similarly, Liu et al. (2023) find that online food delivery platforms have an indirect effect to increase female labor market participation, despite lacking explicit intentions to do so. Moreover, existing literature on IT (direct and indirect) value primarily focuses on the adoption of IT artifacts by individuals (e.g., social media (Allcott et al. 2020), sharing platforms (Li et al. 2022), ride-hailing platforms (Burtch et al. 2018), crowdfunding platforms (Wang and Overby 2022)) or organizations (e.g., IT investment (Gómez et al. 2017), enterprise resource planning (Jia et al. 2020), supply chain management (Dong et al. 2009), healthcare IT (Ganju et al. 2022)). However, our research stands apart by examining the indirect value of IT in the context of public infrastructure technology and discussing its impact on corporate corruption.

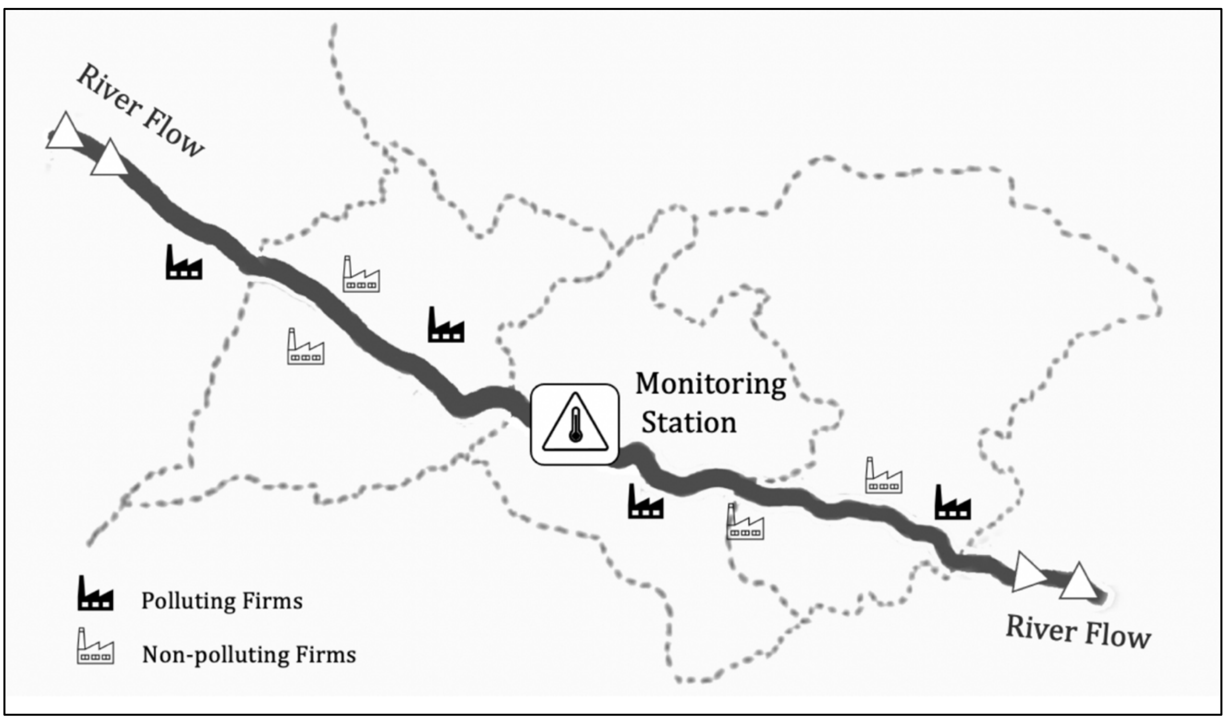
The issue of corporate corruption presents a major threat to the stability and economic growth worldwide. According to the Dow Jones State of Anti-Corruption Survey conducted in 2014, [1] approximately one-third of companies have reported losing business to their competitors who secured contracts through bribery. According to Pei (2007)'s estimation, the direct expenses incurred due to corruption in China amount to approximately three percent of the country's annual GDP. Corporate managers often corrupt to leverage their political connections to channel resources toward their firms or to evade regulations (Faccio 2006, Li et al. 2008). However, such actions can result in distortions that ultimately lead to suboptimal decision-making and underperformance (Faccio 2006, Fan et al. 2007, Fisman and Wang 2015).

Recognizing the potential of public infrastructure technology as a catalyst for change, our study seeks to examine its indirect impact on reducing corporate corruption. We go beyond the traditional focus on individuals' or firms' IT adoption and delve into the broader context of public infrastructure technology,

[1] https://pro.wsj.com/press-room/dow-jones-risk-compliance-issues-sixth-annual-anti-corruption-survey/

which encompasses technological advancements in public facilities. Specifically, the public infrastructure technology we examine is water monitoring stations. During the 1990s, the Chinese government constructed numerous state-controlled water quality monitoring stations, with the objective of gathering surface water quality information (e.g., the depth, speed, and width of surface water, water quality). This has provided a distinctive empirical context for identifying causality and assessing the impact of public infrastructure technology. As the water flows from upstream to downstream, the impact of technology (i.e., monitoring stations) on water quality is limited to upstream firms because downstream firms' pollution and emissions cannot be detected by the water quality monitoring stations (see Figure 1 for an illustration). In other words, companies situated in close proximity to water monitoring stations, either upstream or downstream, share many similarities, with the exception that upstream enterprises are subject to the influence of water monitoring stations. This allows us to utilize a regression discontinuity design to investigate the effects of public infrastructure technology.

Figure 1. Research Context



Leveraging the spatial discontinuity in monitoring water pollution, we discovered that polluting businesses located immediately upstream of water monitoring stations showed significantly lower corporate corruption costs compared to polluting firms located immediately downstream. Remind that the objective of the water monitoring stations was to assess the quality of water, rather than to address issues of corruption. The results demonstrate that the employment of public infrastructure technology, particularly the water monitoring stations, yields an unintended impact on the corrupted activities of polluting enterprises, despite lacking explicit design for such purposes.

Further analysis reveals, however, that the indirect impact of public facility technology cannot be solely attributed to technological resources. We demonstrated that the complementarity of public facility technology and organizational change is the key to mitigating corporate misconduct. In China, an organizational change has been implemented suddenly whereby water quality measurements obtained from water monitoring stations have been integrated as a fundamental metric for assessing the performance of the administrations in the regions where the companies are located. Moreover, our further investigation shows that the organizational change in isolation is inadequate for such effect on reducing corporate corruption. Several other robustness checks confirm the validity of our regression discontinuity design and the findings.

By examining the relationship between public infrastructure technology, organizational change, and corporate corruption, our research contributes to a nuanced understanding of the broader IT value landscape. The existing body of literature has traditionally concentrated on the direct effects of IT artifacts (Brynjolfsson 1996, Bharadwaj et al. 2007, Nevo and Wade 2010, Grover and Kohli 2012, Mithas et al. 2016, Gómez et al. 2017, Trantopoulos et al. 2017, Havakhor et al. 2019, Babar and Burtch 2020, Guo et al. 2022, Liu et al. 2022), and the indirect effects of IT adopted by individuals (Burtch et al. 2018, Allcott et al. 2020, Li et al. 2022, Wang and Overby 2022). Our study contributes to the field of IT value by drawing attention to the indirect effects of public infrastructure technology. Specifically, we highlight the unintended consequences of such technology in the fight against corporate corruption. By broadening our perspective on IT value and recognizing the complementary role of organizational change, we contribute to a deeper understanding of the strategies that can effectively promote ethical behavior and integrity.

Furthermore, our study holds implications for policymakers and practitioners involved in anti-corruption efforts. By recognizing the potential of public infrastructure technology as a catalyst for the positive change, we provide valuable insights into the design and implementation of effective interventions. We advocate for a holistic approach that integrates technological advancements in public infrastructure with organizational change initiatives.

Section 2 reviews related literature. In Section 3, we discuss the background of organizational change, water monitoring stations, and corruption in China. We then describe the data and present our empirical strategy in Section 4. In Section 5, we report our main results and discuss the underlying mechanisms. We then provide several robustness checks and additional analyses in Section 6. In Section 7, we conclude with a discussion.

2. Literature Review

Our study makes contribution to the literature on the value of IT. IT value research examines the benefits and strategic value of information technology (Hitt and Brynjolfsson 1996, Oh and Pinsonneault

2007). This topic has been extensively studied (Brynjolfsson 1996, Brynjolfsson and Hitt 1996, Melville et al. 2004), and researchers have adopted myriad approaches to assess the mechanisms through which IT value is generated and to estimate its magnitude (Mukhopadhyay et al. 1995, Kohli and Devaraj 2003). The literature on IT value can be broadly classified into two categories: IT direct value, which examines whether the anticipated benefits of IT have been realized (Hitt and Brynjolfsson 1996), and IT indirect value, which investigates unintended consequences resulting from IT adoption. Our paper falls within the scope of the second category.

Existing research on IT value has predominantly focused on its direct value, examining how IT enhances productivity, fosters competitive advantages, and generates consumer value (Hitt and Brynjolfsson 1996). Regarding productivity, studies have found that IT spending leads to a positive increase in gross marginal product (Brynjolfsson and Hitt 1996), and contributes significantly to GDP growth (Dewan and Kraemer 2000), with varying impacts across industries (Mittal and Nault 2009). In terms of competitive advantages, IT significantly improves supply chain efficiency (Rai et al. 2006, Dong et al. 2009, Yao et al. 2019) and firm performance (Bharadwaj et al. 2007, Chakravarty et al. 2013, Tafti et al. 2013). It also helps mitigate diminishing returns to R&D (Ravichandran et al. 2017), facilitates knowledge flows within alliances (Ravichandran and Giura 2019), improves quality and cost outcomes (Mishra et al. 2022), and promotes efficiency and innovation across multiple domains (Nevo and Wade 2010, Gao and Hitt 2012, Gómez et al. 2017, Trantopoulos et al. 2017, Wang and Overby 2022). In terms of consumer welfare, studies have shown positive associations between IT investments and consumer surplus (Brynjolfsson 1996) as well as customer satisfaction (Mithas et al. 2016). Furthermore, in the transportation sector, IT has assisted commuters and local governments in managing urban traffic (Babar and Burtch 2020, Cheng et al. 2020, Rhee et al. 2022). In contrast to this body of literature, our study diverges by focusing on the indirect value of IT. Rather than examining the anticipated benefits of IT, our primary objective is to explore and analyze the unintended consequences that arise from its use.

The indirect value of IT remains a relatively underexplored area of research, with only a limited number of studies addressing it, as shown in Table 1. For instance, research shows that the use of an internet-enabled matching platform, which aims to facilitate social interactions and meetings among couples and groups, unintendedly leads to an increase in HIV incidence (Greenwood and Agarwal 2016) and prostitution cases (Chan et al. 2019). Similarly, Chan et al. (2016) discover an unexpected correlation between greater broadband availability, intended to provide high-speed internet access, and a rise in racial hate crimes. The ride-hailing platform Uber, designed to facilitate the matching of passengers with drivers, unintentionally results in a reduction in homicide rates (Greenwood and Wattal 2017), a decrease in entrepreneurial activities (Burtch et al. 2018), and a decline in the occurrence of rape incidents (Park et al. 2021), despite not intending to bring about such outcomes. Moreover, Allcott et al. (2020) find that

Facebook, a social media and social networking platform, unintentionally contributes to political polarization among individuals, though it was not designed to impact political polarization. Liu et al. (2023) find that online food delivery platforms, which act as intermediaries facilitating the delivery of food orders from restaurants to consumers, inadvertently contribute to the rise in female labor market participation. Ozer et al. (2023) discover that a peer-to-peer lending platform, a micro-financing platform facilitating direct connections between borrowers and investors, is correlated with a rise in abortion rates among women, despite not being designed for such purpose. Our study adds to the limited yet growing body of literature on the unintended consequences of IT. Specifically, we demonstrate that water monitoring stations, originally established for scientific purposes, unintentionally help mitigate corporate misconduct, despite not being explicitly designed for such purposes. The significance of this field lies in the need for a comprehensive evaluation of the value and impact of IT, by considering both the anticipated benefits and unintended consequences associated with its adoption.

Table 1. Literature Review on the Indirect Value of IT

Paper	IT Artifact	Areas	Key Findings
Ozer et al. (2023)	LendingClub	Public Health	There is a strong and significant increase in the abortion rate following the entry of LendingClub.
Liu et al. (2023)	Online Food Delivery Platform	Labor Market	The entry of an online food delivery platform significantly increased the female employment rate.
Park et al. (2021)	Uber	Criminality	Uber's entry into a city is negatively associated with the number of rape incidents.
Allcott et al. (2020)	Facebook	Belief	Deactivating Facebook helps reduce political polarization.
Chan et al. (2019)	Craigslist	Criminality	Entry of Craigslist to a county lead to a 17.58% increase in prostitution cases.
Burtch et al. (2018)	Uber	Entrepreneurship	Uber platform entry has a negative and significant relationship with entrepreneurial activity.
Greenwood and Wattal (2017)	Uber	Criminality	The entry of Uber services led to a significant drop in the rate of homicides.
Chan et al. (2016)	Internet broadband	Criminality	Broadband availability increases racial hate crimes.

Greenwood and Agarwal (2016)	Craigslist	Public health	Internet enabled matching platform implementation led to a significant increase in HIV incidence.

Our research diverges from the current literature on the indirect value of IT in two distinct ways. First, the existing literature primarily addresses the adoption of IT artifacts by individuals (see Table 1), such as social media (Allcott et al. 2020), ride-hailing platforms (Greenwood and Wattal 2017, Burtch et al. 2018, Park et al. 2021), online matching platform (Greenwood and Agarwal 2016, Chan et al. 2019), food delivery platform (Liu et al. 2023), P2P lending platform (Ozer et al. 2023) and internet broadband (Chan et al. 2016). However, none of the aforementioned studies explore the unintended consequences of public infrastructure technology. Even when considering the broader scope of literature that investigates both the direct and indirect value of IT, the prevailing emphasis remains on investigating IT adoption by individuals and organizations in areas such as crowdfunding platforms (Wang and Overby 2022), IT investment (Gómez et al. 2017), enterprise resource planning systems (Jia et al. 2020), supply chain management (Dong et al. 2009), and healthcare IT (Ganju et al. 2022). The academic literature has predominantly overlooked the influence of public infrastructure technology.

Second, while the existing studies have identified the unexpected consequences of IT implementation in various domains (see Table 1), including public health (Greenwood and Agarwal 2016, Ozer et al. 2023), labor market (Liu et al. 2023), criminality (Chan et al. 2016, Greenwood and Wattal 2017, Chan et al. 2019, Park et al. 2021), entrepreneurship (Burtch et al. 2018), and individual belief (Allcott et al. 2020), we expand the literature on indirect value of IT by focusing on ethical business practices. Such indirect value is noteworthy, since the issue of corporate corruption presents a major threat to the stability and economic growth worldwide (Cheung et al. 2021, Griffin et al. 2022), leading to distortions that result in suboptimal decision-making and underperformance (Faccio 2006, Fan et al. 2007, Fisman and Wang 2015).

In summary, our research contributes to the literature by examining the indirect value of IT on ethical business practices in the context of public infrastructure technology, thereby extending the research on IT value.

3. Background

3.1. Water Monitoring Stations

In the 1990s, the Chinese central government installed state-controlled water monitoring stations along major rivers, lakes, and reservoirs, to scientifically examine hydrological features such as the depth, speed, and width of surface water. The water monitoring stations sample and analyze the water, and then report the water quality information to China's National Environmental Monitoring Center (CNEMC). The water quality information includes the temperature (°C), pH, Chemical Oxygen Demand (COD) (mg/L), conductivity (μS/cm), turbidity (NTU), permanganate (mg/L), NH_3-N (mg/L), total phosphorous (mg/L), total nitrogen (mg/L), chlorophyII α (mg/L) (for lakes), algal density (cells/L) (for lakes), as well as an overall assessment of the water. The overall assessment of the water is based on the Environmental Quality Standards for Surface Water issued by the Ministry of Ecology and Environment. Through the Internet and virtual private network technology, the water quality information of the automatic water monitoring stations is transmitted to the CNEMC in real-time every four hours (0:00, 4:00, 8:00, 12:00, 16:00, and 20:00).

The purpose of the water monitoring stations was to get a comprehensive understanding of the surface water quality in China. Guided by this motivation, the locations of the water monitoring stations were chosen to spatially monitor the neighboring water bodies. The stations cover all the major rivers, lakes, and reservoirs in China. According to the Ministry of Environment Protection, the factors that affect the location of the water monitoring stations include river flow, areas of the water surface, geographical features of the riverbed, etc. In summary, the water monitoring stations are established and sponsored by the central government, and the locations were chosen in accordance with a whole host of scientific factors. In other words, non-scientific factors such as the needs of local officials, firms, and the local economy were excluded from the consideration when picking the location.

Since the locations of the monitoring stations were determined by scientific reasons, we can eliminate the concern that polluting firms lobbied the government to locate water monitoring stations far away from their locations to avoid being monitored. Furthermore, we conduct a robustness check to test whether the firms strategically relocated downstream to avoid pollution detection. The results in Table A5 in the Appendix dismiss this possibility.

Our study also avoids the concern that the local officials might manipulate data in their favor because we focus on state-controlled water quality monitoring stations. The Technical Specifications for Automatic Monitoring of Surface Water, issued by the Chinese Ministry of Ecology and Environment (MEE), specifies that state-controlled stations directly report water quality readings to the central government.[2] Therefore, local officials did not have approaches to manipulate the readings.

These monitoring stations capture and measure emissions only within a few kilometers upstream, and they cannot detect the pollution from downstream firms and firms located far away. Therefore, only immediate upstream firms are influenced by the technology (i.e., water monitoring stations). This unique context provides us the opportunity to use regression discontinuity methods to study the impact of public

[2] http://english.mee.gov.cn/Resources/standards/SpecificationsTestProcedures/201801/t20180126_430442.shtml

infrastructure technology.

3.2. Organizational Change

In 2003, a significant organizational change in government has been implemented whereby water quality measurements obtained from water monitoring stations have been integrated as a fundamental metric for assessing the performance of the administrations in the respective regions where the companies are located. In this section, we briefly introduce the organizational change.

Chinese President Hu Jintao formally proposed the “Scientific Outlook of Development” (SOD) in 2003, which is composed of solutions for maintaining sustainable development, enhancing social welfare, and combatting environmental problems. Following the SOD, MEP released detailed plans to reduce environmental pollution. Specifically, in 2003, the central government decentralized authority for monitoring surface and wastewater to local governments, and established a five-year plan for pollution abatement among all major rivers, including the Huai, Liao, and Hai Rivers.[3] In its 10th Five-Year Plan,[4] “by 2005, the readings of chemical oxygen demand (COD) and ammonia-nitrogen (NH_3-N) of water monitoring stations in Huai River should be lower than 64.3 ton/annual and 9.2 ton/annual, respectively. The readings of COD and NH_3-N in Hai River should be lower than 106.5 ton/annual and 20.5 ton/annual respectively. The readings of COD and NH_3-N in Liao River should be lower than 32.58 ton/annual and 5.2 ton/annual respectively. The readings of COD and NH_3-N in the water conveyance line of the “South-to-North Water Diversion” project should be lower than 54.7 ton/annual and 7 ton/annual respectively. 87% of the quality readings of water monitoring stations along the “South-to-North Water Diversion” project should reach Grade III by 2005 and the readings for the remaining water monitoring stations should not be worse than Grade IV. All state-controlled water monitoring stations should have water quality readings better than Grade V by 2005”.

From then on, water quality readings, reported by the monitoring stations, turned out to be a factor of consideration for provincial leaders’ political promotions after 2003.[5] Specifically, the official document explicitly stated that “the achievement of pollution abatement target become one component of cadre evaluation metric (KPI of the local officials)”, which was the paramount political motivation to local officials.

Local government officials could enforce environmental regulations by interfering with the production process of polluting firms in various ways. Given the paramount political motivations to reduce emissions and to improve water quality, policies such as the Jiangsu Environmental Protection

[3] See http://www.mee.gov.cn/ywgz/fgbz/bz/bzwb/jcffbz/200301/t20030101_66890.htm and http://www.mee.gov.cn/ywgz/fgbz/bz/bzwb/jcffbz/200301/t20030101_66891.htm.
[4] http://www.gov.cn/gongbao/content/2002/content_61775.htm
[5] In China, junior officials are appointed by higher-ranked government officials rather than chosen by voters.

Enforcement Plan 2003 explicitly stated that local officials would be promoted only if the water monitoring stations reported good water quality.[6] As a result, this organizational change in government indeed reduced pollutant and emissions levels after 2003 (He et al. 2020).

3.3. Corruption in China

Corruption is an "economic and political evil" (Seligson 2002) in both developed and developing countries. It increases transaction costs, intensifies economic inequality, distorts resource allocation, undermines political trust, and fuels social unrest. Corruption has severely impacted nearly every Chinese sector, including law enforcement, military, medicine, and education. In 1995, Transparency International ranked China as the fortieth of forty-one countries in the worldwide Corruption Perceptions Index.[7]

In China, firms might escape investigation, improve sales, leverage financial advantages, and avoid regulations by bribing officials for "protective umbrellas" (World Bank 1997). For example, in the coal mining industry, the central government delegated the management of state coal mines to local governments in 1998, which created opportunities for local officials to take bribes in exchange for overlooking mismanagement and regulation violations. As a result, the death rate of coal mine workers increased dramatically during the delegation periods (Jia and Nie 2017). The Zijinshan mining area bribed Chen Junan, director of the Shanghang County Environmental Protection Bureau, to protect the company from regular supervision, which led to a severe sewage leakage in 2010. This leakage polluted the Tingjiang River, killed fish in the river, and cost 22.2 million RMB (about US$3 million) in damages.

Some Chinese firms often conduct corruptive transactions and build relationships (*guanxi*) through banqueting, drinking, gift exchange, etc. Specifically, businessmen usually build connections with government officials through "drinking at the banquet, singing in the karaoke clubs (KTVs), carousing in the nightclub, group visits to brothels and saunas, and playing mahjong in teahouses" and the bills are paid by the businessmen (Osburg 2018). Many business deals were thus sealed by the expensive dishes and fine liquors on the table, and by paying and receiving bribes, kickbacks, and luxury gifts under the table. Such expenses are usually recorded as "entertainment and travel cost" (ETC), a standard expenditure item in the accounting book of Chinese firms. As the name suggests, ETC is used to cover entertainment (including banqueting, drinking, gift, karaoke, sports club membership, etc.) and travel expenditure (Cai et al. 2011). Therefore, ETC is often used as a proxy to measure corruption between businesses and in the Chinese context (Cai et al. 2011, Huang and Li 2013).

---

[6] Original documents: http://www.js.gov.cn/xxgk/project/P0201606/P020160629/P020160629314508753226.pdf. More policy documents and translations can be found in He et. al (2020).

[7] https://www.transparency.org/en/cpi/1995/results/. The Corruption Perceptions Index is based on expert and business executive rankings of countries. The composite index combines 13 surveys and assessments collected by various reputable institutions.

During our research sample period (1996 – 2011), the Chinese government did not lay a strong emphasis on anti-corruption. Anti-corruption became a priority only after 2012 when President Xi Jinping launched the largest organized anti-corruption effort in Chinese history and successfully cracked down on corrupt officials. Among the measures, he banned alcoholic beverages at official banquets and required disclosure of government officials' personal assets. Thus, the corruption we observe in this study is not affected by anti-corruption policies.

4. Research Design and Data

4.1. Research Design and Empirical Model

The main purpose of our empirical analysis is to identify whether public infrastructure technology has an indirect effect on corruption. Given that water monitoring stations can only measure and report the aggregate emissions coming from immediate upstream manufacturers, we use this spatial discontinuity in monitoring stations to apply a regression discontinuity framework to identify whether the public facility technology could mitigate the corruption problem. The spatial distance is the running variable in which positive (negative) distance indicates that firms are located upstream (downstream) of their nearest water monitoring stations, and spend less (more) money to bribe governments.

We use both non-parametric and parametric approaches to estimate regression discontinuity and find quite consistent results. Our main results are based on the non-parametric approach. This approach is widely applied in quasi-experimental research, and it provides point estimates and inference procedures that are more robust to parametric misspecification bias (Calonico et al. 2018).[8]

To illustrate the non-parametric estimation, let Y_i be the outcome variable of firm i (entertainment and travel costs as a proxy of corruption). Let c_i be the location relative to water monitoring: positive (negative) c_i indicates upstream (downstream). The treatment effect of water monitoring stations on corruption is:

$$\hat{\beta} = \lim_{c \downarrow 0} E[Y_i | c_i] - \lim_{c \uparrow 0} E[Y_i | c_i] \tag{1}$$

The intuition of the regression discontinuity is that the treatment is "as good as" randomly assigned in a neighborhood of water monitoring stations. Since whether the firms belong to the treatment or control group is "as good as" random, treatment and control firms are all else similar except that the treatment firms are being treated (i.e., being affected by the water monitoring stations). Hence, the average treatment

[8] We choose the Cattaneo et al. (2020) estimator by following several papers on spatial regression discontinuity that leverage this estimator (Campante and Yanagizawa-Drott 2018, Ehrlich and Seidel 2018, Ito and Zhang 2020). The running variable for Campante and Yanagizawa-Drott (2018) is the distance of the regions to the airport. The running variable for Ehrlich and Seidel (2018) is the distance of the city to the border. The running variable for Ito and Zhang (2020) is the distance between the city and the river.

effect is equal to the difference between the limits of the treated and control average observed outcome. We will show more details about the assumption in Section 4.3. In our case, as the distance between the firms and water monitoring stations approaches zero, the treatment effect is equal to the differences in corruption between upstream and downstream firms.

Following Hahn et al. (2001) and Imbens and Lemieux (2008), the conventional non-parametric estimation of $\hat{\beta}$ is:

$$\min_{\beta,\gamma,\tau,\sigma} \sum_{i=1}^{N} K\left(\frac{c_i}{h}\right) [Y_i - \sigma - \gamma \times c_i - \beta D_i - \tau \times D_i \times c_i]^2 \quad (2)$$

D_i is 1 if the firm is located upstream, and 0 otherwise; h is the bandwidth; and $K(\cdot)$ is the rectangle kernel function. As Hahn et al. (2001) and Imbens and Lemieux (2008) suggest, discontinuity is estimated using local polynomial regression, and γ, τ, σ are the parameters for the polynomial regression. (Detailed derivations can be found in Imbens and Lemieux (2008).) We calculate optimal bandwidth using the MSE optimal bandwidth method proposed by Calonico et al. (2014), because it has a smaller coverage error and it is less sensitive to tuning parameter choices. We also use alternative optimal bandwidth selection methods to check the robustness and find consistent results (see Appendix Table A10). We implement triangular, Epanechnikov, and uniform kernel functions following Hahn et al. (2001). The kernel is used to weigh the observation around the water monitoring stations, which is commonly used in non-parametric estimation techniques. Uniform kernel means to compute the treatment effect using the unweighted observations. Triangular kernel and Epanechikov kernel put more weight on observations near the water monitoring stations than observations located far away from the water monitoring stations. The kernel density functions are different for the two kernels.

Conventional non-parametric regression discontinuity estimation is the most commonly used model. It requires a specified and balanced bandwidth, which is small enough to remove the smoothing bias yet large enough to ensure adequate precision. Its performance is quite sensitive to the bandwidth used; inappropriate bandwidths could bias the results. Thus, we also report the bias-corrected estimation proposed by Calonico et al. (2014), which removes the potentially large effect of unknown leading bias of the conventional estimator and justifies it by using the MSE-optimal bandwidth choices.

However, bias-corrected estimators have some unappealing properties such as potentially large-coverage distortions in applications and poor performance in finite samples. Therefore, we follow Calonico et al. (2014) and report the robust estimator. In addition to the benefit of correcting biases, this robust estimator offers excellent finite sample performance.

For each regression discontinuity model, we report estimators using conventional, bias-corrected, and robust methods with triangular, Epanechnikov, and uniform kernel types proposed by Cattaneo et al. (2020). This estimator is a nonparametric density estimator based on local polynomial techniques. The estimator is boundary adaptive and does not require pre-binning of the data. The estimator is constructed by smoothing out the empirical distribution function using local polynomial techniques. Importantly, the estimator is an asymptotic distributional approximation with precise leading bias and variance characterizations, and a consistent standard error estimator. These properties apply to both interior and boundary points in a fully automatic and data-driven way. In other words, it does not require prior knowledge of the shape of the sample.

Our results are highly robust to different estimation methods and kernel types. In addition, we present comprehensive robustness checks in the Appendix, including alternative estimators, difference-in-differences analysis, alternative bandwidth selectors, and covariates in the estimation. All the results from alternative specifications are highly consistent with our main findings. Furthermore, we conduct a placebo test by moving the water monitoring stations upstream or downstream by 2km, and then estimate using the same approach. As expected, we find that the corruptive behavior is similar for firms located upstream and downstream relative to the fictitious stations, proving that our main findings are robust. All standard errors in our analysis are clustered at the water monitoring station level to eliminate concerns about the spatial correlation of the error term.

4.2. Data

Water Quality Monitoring Stations

We use the China Environmental Yearbook to collect information about water quality monitoring stations. They were constructed in the 1990s to measure water quality in major rivers, lakes, and reservoirs in China for scientific purposes. We exclude stations located on lakes and reservoirs because we could not identify the relative upstream and downstream locations between these stations and their nearest firms. This leaves 478 water monitoring stations to be geocoded.

Firm-Government Corruption

The entertainment and travel costs (ETC) accounting category is commonly used to document both ordinary business expenses and costs for bribing government officials, clients, and suppliers in China (Cai et al. 2011, Huang and Li 2013). The association of ETC with firms' corrupt practices has been widely acknowledged in both news media and academic literature because of its frequent utilization by firms as a means of providing benefits to cultivate relationships with external entities (Huang and Li 2013, Zhu and Wu 2014, Cull et al. 2015, Lin et al. 2016, Huang et al. 2017, Kong et al. 2017, You and Nie 2017, Wei et al. 2020). As Huang and Li (2013) pointed out, firms do not have incentives to lie about ETC in the survey because the ETC is normally below the maximum allowed amount set by China's taxation authorities. In addition, during our sample period, the accounting practice of ETC is not changed. Thus, we adhere to the existing body of scholarly works in utilizing ETC as a measure of corporate corruption.

Furthermore, we also demonstrate the connection between the accounting variables ETC and firms' actual misconduct in Appendix Table A1. The Poisson regression analysis presented in Appendix Table A1 reveals a statistically significant positive correlation between the number of regulation breaches observed in firms and the variable ETC. This provides compelling evidence of the association between our measurement of ETC and firms' actual misbehaviors.[9]

We collect multiple officially released datasets to gather firm information, measure firm corruption, and conduct a thorough and robust analysis. The different sources can be cross-validated. Each dataset has advantages and disadvantages. We will discuss them in greater detail.

The first dataset we use is the Annual Survey of Industrial Firms (ASIF) 2004 dataset, conducted by the Chinese National Bureau of Statistics (NBS). ASIF provides basic information (e.g., firm name, address, industry classification, ownership), and major accounting items (e.g., income, sales, and tax). Although ASIF surveys are conducted yearly, we only use ASIF 2004 because that was the only year that NBS surveyed and reported ETC. This cross-sectional dataset covers all industrial firms with annual sales above 5 million RMB (about US$762,000) in China.

Although the ASIF dataset contains some data irregularities, many empirical studies have used it for analysis. We clean the data following the practice of the previous literature (Brandt et al. 2012, Rudai 2015, He et al. 2020), which dropped about 1.71% of firms in the original dataset.[10] First, we drop observations with missing key values, such as addresses and ETC figures. Second, we drop observations with negative values for variables that should be positive, such as employment and capital stock. Third, we drop observations that violate generally accepted accounting principles. For example, we drop the companies whose total assets were unequal to the sum of total liabilities and equity. Fourth, following He et al. (2020), we drop the firms that belong to parent multi-unit firms, because subsidiaries may have reallocated production to avoid tighter regulations after 2003. Last, we winsorize the data by keeping records in which the values of our dependent variable (ETC) ranged from the 0.5th to 99.5th percentile.

We first draw a 10-km radius circle around each water monitoring station and keep only those towns within the circle. Then, combined with geolocation information about the water monitoring stations, we can determine whether each firm was situated upstream or downstream relative to its nearest station, and then calculate the distance. If a firm was upstream (downstream) to its nearest station, and downstream (upstream) to its second nearest station, we drop the observation to avoid ambiguity.

In addition, ASIF identifies industry categories of firms. Following the official Ministry of

[9] However, due to the limited number of datapoints available during our sample periods, we were unable to utilize the data on regulation breaches to perform regression discontinuity analyses in our sample periods.

[10] In Appendix Table A19, we alleviate the concern of sample selection by conducting t-tests about the characteristics for the sample after the data cleaning process (column (1)) and before the data cleaning process (column (2)).

Environmental Protection (MEP) classification, we classify firms into two categories: polluting and non-polluting. Non-polluting firms could serve as a control group since neither technologies nor organizational changes aimed at reducing pollution would have had an impact on them.

The second dataset we use is Chinese Private Enterprise Survey (CPES) 1996 – 2011, one of China's longest-lasting large-scale nationwide surveys providing key micro firm-level information. CPES is collected and supervised by the Privately Owned Enterprises Research Project Team. The team members are from the United Front Work Department of CPC Central Committee, All-China Federation of Industry and Commerce, State Administration for Market Regulation, and the Chinese Academy of Social Sciences. CPES has been used in hundreds of empirical studies in the last two decades (Jia 2014, Zhao and Lu 2016, Ge et al. 2019). CPES data are released as multiyear cross-sectional data sets, allowing an analysis of dynamic corruptive activities across years. Note that CPES provides multiyear cross-sectional data rather than panel data, because they surveyed different samples of private firms in different years.

Unfortunately, the address information for CPES firms is limited to zipcode level only, so we can only roughly measure their relative distances from the nearest monitoring stations and upstream/downstream locations. Moreover, CPES is insufficient for classifying firms into polluting industries and non-polluting industries.

The summary statistics of the data can be found in Appendix Table A2 and Table A3.

Geo-data

The geocoding locations of ASIF firms are obtained through Baidu Map API (https://api.map.baidu.com/geocoding/v3/), which returns coordinates and outputs a confidence level parameter that measures errors of street addresses. We keep the coordinates with errors within 5km (confidence level > 30).[11] We obtain geocoded information for each zip code from the National Bureau of Statistics and use the data to match with our firm datasets. We also obtain GIS data on the water basin systems from the Ministry of Water Resources.[12] We then use the ArcGIS software to identify firm-station upstream and downstream relationships and firms' distances to water monitoring stations. The summary statistics are presented in Table A4. In addition, we collect the river length data of each administration in China from the National Geomatics Center of China.

4.3. Regression-Discontinuity Assumption Check

Our spatial discontinuity design basically assumes that firms located immediately upstream and

---

[11] The confidence level ranges between 20 (error within 10km) and 100 (error within 20m). We also conduct robustness checks using samples with different confidence level cutoffs and find consistent results (Table A17).

[12] http://mwr.gov.cn/english/aboutmwr.html.

downstream of water monitoring stations are similar. We thus check the manipulation of the assignment variable and examine the balances between firm-level characteristics of upstream and downstream firms (Lee and Lemieux 2010). If the assignment variable is correlated with the treatment, upstream and downstream firms would show significantly different characteristics. Appendix Table A2 shows the summary statistics and balance checks for variables including the firm's opening year, whether it is in a polluting industry, its profits, value-added tax, number of employees (men and women), capital stock, and intermediate input. Characteristics are all well-balanced between upstream and downstream firms. Appendix Table A3 further demonstrates that firms in various industries are equally distributed on both sides of the monitoring stations.[13]

In addition, sentiment toward corruption or pollution did not differentially affect firms located upstream and downstream during our sample periods. Anti-corruption became a priority only after 2012, which is after our sample periods. Therefore, there should be no sentiment change in corruption during our sample periods. To further check whether upstream and downstream firms have different sentiments toward pollution, we investigate whether upstream and downstream invest differently in combating pollution. T-test (*t=0.836, p=0.403*) suggests that upstream and downstream firms display similar sentiments toward pollution.

5. Results and Mechanisms

5.1. Main Results

We report the non-parametric estimates using conventional, bias-corrected, and robust estimation methods proposed by Calonico et al. (2014) with triangular, Epanechnikov, and uniform kernel types (Hahn et al. 2001) (Section 3.1). Each cell in Table 2 is a separate regression discontinuity estimate using one of the three estimation methods and one of the three kernel types. We use corruption and ETC interchangeably in the following sections.

Columns (1) – (3) of Table 2 report the regression discontinuity estimates of the ETC gap between the upstream and downstream private enterprises in the polluting industries in 2004. Columns (4) – (6) of Table 2 display results for private enterprises in the non-polluting industries. In the polluting industries, firms located immediately upstream showed significantly lower ETC than the downstream counterparts (90 ~ 125 thousand RMB or US$13,000 ~ 18,000 for each firm annually) (Table 2 columns (1) – (3)), which translates to 61.61% ~ 85.58% reduction compared to the average corruption costs.[14] The

[13] Our results are quite robust when we remove the unbalanced industries (Table A18).
[14] The average corruption costs for ASIF firms are 146,070 RMB.

regression discontinuity estimates are highly robust to different estimation methods and kernel types.

The results indicate that upstream firms in polluting industries engage in substantially less corruption than downstream firms, demonstrating the indirect effect of technology in public infrastructure on reducing bribes. Although the water monitoring stations are not designed for mitigating the incidence of corrupt practices of firms, our empirical analyses reveal that the deployment of this public infrastructure has unanticipated effects on corporate corruption. For non-polluting firms, the upstream-downstream corruption gap was much smaller, and is not significantly different from zero (Table 2 columns (4) – (6)), suggesting that the indirect effect of the technology in public infrastructure is not universal across all firms. Given that the water quality readings of water monitoring stations are only influenced by the emissions of polluting firms, it is logical to infer that non-polluting enterprises were not affected by water monitoring stations.

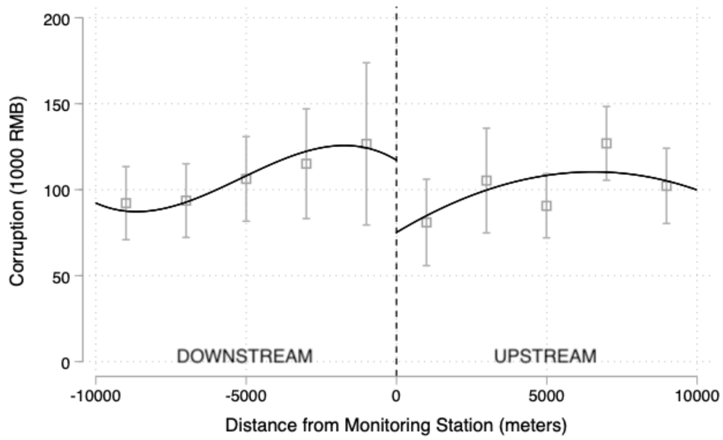
Table 2. Private Enterprises: Upstream-Downstream Corruption Gaps (ASIF)

	Polluting industries			Non-Polluting industries		
Method	(1)	(2)	(3)	(4)	(5)	(6)
Conventional	-93.11**	-95.93**	-107.34**	-14.91	-15.27	-21.29
	(33.75)	(32.83)	(35.41)	(23.18)	(24.02)	(27.66)
Bias-corrected	-106.41**	-109.50***	-124.71***	-21.69	-23.79	-28.98
	(33.75)	(32.83)	(35.41)	(23.18)	(24.02)	(27.66)
Robust	-106.41**	-109.50**	-124.71**	-21.69	-23.79	-28.98
	(37.09)	(35.91)	(39.16)	(26.02)	(26.53)	(30.88)
Observations	3,502	3,502	3,502	9,288	9,288	9,288
Kernel	Triangular	Epanechnikov	Uniform	Triangular	Epanechnikov	Uniform
Bandwidth	11100	10970	7834	11729	11650	9384

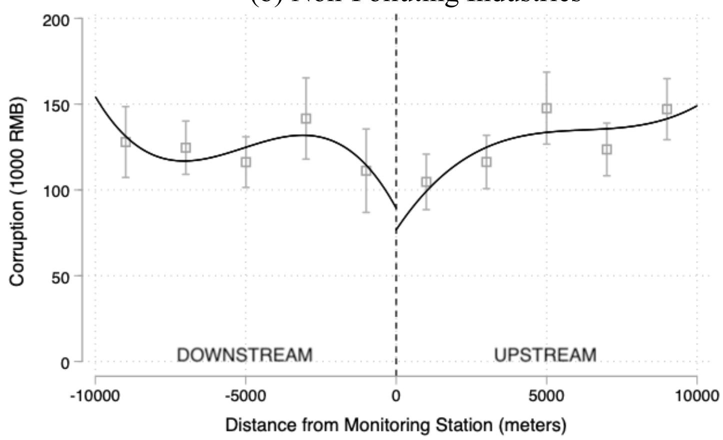
Note: Each cell represents a separate regression discontinuity regression. Data are from ASIF. The running variable is the distance between a firm and a monitoring station. A positive (negative) distance means the firm is located upstream (downstream). The negative coefficients indicate that upstream firms have lower ETC. The discontinuities at monitoring stations are estimated using methods proposed by Calonico et al. (2014) and the MSE optimal bandwidth proposed by Calonico et al. (2014) for different kernel weighting methods. Standard errors clustered at the monitoring station level are reported below the estimates. * significant at 5% ** significant at 1% *** significant at 0.1%.

Figure 2. Regression Discontinuity Plot from ASIF Data

(a) Polluting Industries



(b) Non-Polluting Industries



We then visualize our findings in Figure 2, which plots corruption (ETC) against the distance between a polluting firm (Figure 2(a)) or a non-polluting firm (Figure 2(b)) and its nearest monitoring station. Each dot represents the average ETC level within 2km. Similarly, we observe a sharp decrease in corruption around the monitoring stations. Upstream polluting firms showed a significantly lower corruption level than the downstream polluting firms.

5.2. Mechanism

Having established that public facility technology has a significant indirect effect on reducing corporate corruption, we explore the mechanisms driving this effect in this section. Specifically, we investigate the role of public facility technology and its complementarity with organizational change in reducing corporate corruption. We also examine the role of organizational change in isolation.

While previous studies have emphasized the value of technology alone, our analysis reveals that the technological resource alone does not hold the answer to the indirect effect of public facility technology. Instead, our findings suggest that the indirect effect of public facility technology in reducing corporate corruption is contingent on whether it is complemented by organizational change in government. Furthermore, we find that the implementation of organizational change in isolation may not lead to the reduction of firms' corruption.

We first investigate whether technology alone can lead to the reduction of firms' corruption. To investigate this, we compare the corruption gap between upstream firms and downstream firms before 2003 when there was no relevant organizational change at that time, which shows the effect of public facility technology alone on corruption. The results are displayed in Table 3 columns (1) – (3). We find that most of the coefficients do not deviate from zero, indicating that the corruption from upstream firms of water monitoring stations was not different from the downstream firms. Therefore, we do not find supporting evidence showing that technology alone is effective in reducing corruption.

Next, we examine whether public infrastructure technology, together with organizational change in government, has an effect on corruption. Subsequent to 2003, an organizational change was suddenly implemented whereby the water quality measurements obtained from water monitoring stations were incorporated as a fundamental metric for evaluating the performance of the administrations in the respective regions where the companies are situated. It should be noted that the organizational change was not implemented with the aim of addressing corporate corruption. Rather, the organizational change is primarily focused on environmental concerns. Results in Table 3 columns (4) – (6) demonstrate that the upstream firms significantly reduced their corruption costs (130,000 RMB ~ 170,000 RMB or US$20,000 ~ US$27,000 for each firm annually) compared with their downstream counterparts after 2003, or 88.98% ~ 116.36% decrease relative to average corruption costs.[15] The regression discontinuity estimates are highly robust to different estimation methods and kernel types. The evidence indicates that the co-presence of public infrastructure technology and organizational change leads to the reduction of firms' corruption.

Table 3. Private Enterprises in Different Years: Upstream-Downstream Corruption Gaps (CPES)

	Before 2003			After 2003		
Method	(1)	(2)	(3)	(4)	(5)	(6)
Conventional	-1.95	-2.99	-4.51	-13.06***	-14.51***	-14.01***
	(2.76)	(2.66)	(3.46)	(2.78)	(3.04)	(3.04)
Bias-corrected	-1.90	-3.40	-6.53	-15.53***	-17.20***	-15.71***
	(2.76)	(2.66)	(3.46)	(2.78)	(3.04)	(3.04)
Robust	-1.90	-3.40	-6.53	-15.53***	-17.20***	-15.71***
	(3.59)	(3.80)	(4.62)	(3.66)	(4.02)	(5.14)
Observations	2,195	2,195	2,195	3,960	3,960	3,960
Kernel	Triangular	Epanechnikov	Uniform	Triangular	Epanechnikov	Uniform

[15] The average corruption costs for CPES firms are 146,100 RMB.

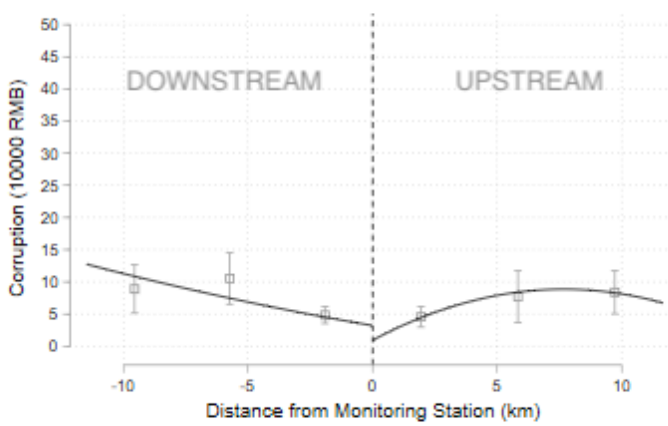
Bandwidth	5.682	5.199	3.456	4.660	4.174	4.297

Note: Each cell represents a separate regression discontinuity regression. Data are from CPES (1996 – 2009). The running variable is the distance between a firm and a monitoring station. A positive (negative) distance means the firm is located upstream (downstream). Negative coefficients indicate that upstream firms have lower ETC. Discontinuities at monitoring stations are estimated using methods proposed by Calonico et al. (2014) and the MSE optimal bandwidth proposed by Calonico et al. (2014) for different kernel weighting methods. Standard errors clustered at the monitoring station level are reported below the estimates. Year fixed effects are included in each regression. * significant at 5% ** significant at 1% *** significant at 0.1%.

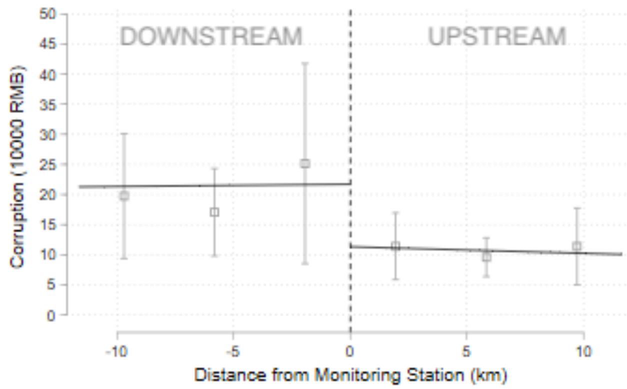
We visualize our findings in Figure 3, which plots corruption (ETC) against the distance between a firm and its nearest monitoring station. Each dot represents the average ETC level within 4km. We also plot their 95% confidence intervals and overlaid the plot with a fitted curve to display the discontinuity around stations. Figure 3(a) shows the discontinuity of corruption among private enterprises before 2003. Little corruption gap existed around monitoring locations before 2003, suggesting when only public infrastructure technology existed, the corruption costs of upstream and downstream private enterprises were rather similar. Figure 3(b) is the regression discontinuity plot for the corruption level of private enterprises in polluting industries after 2003. A sharp decrease in corruption appears around the monitoring stations. Upstream polluting industries showed a significantly lower corruption level than downstream firms after 2003.

Figure 3. Regression Discontinuity Plot for Firms Before and After 2003

(a) Before 2003



(b) After 2003



We further conduct an alternative difference-in-discontinuity estimation (Grembi et al. 2016), which takes the differences of the two upstream-downstream corruptive gaps before and after 2003 for polluting firms. The approach helps control effects around the cutoff point that are uncorrelated with the organizational change, such as changes in attitudes about corruptive behaviors. We report the difference-in-discontinuity estimates in Table 4. We find negative and statistically significant coefficients (100,000 to 200,000 RMB or US$15,000 to US$30,000 for each firm annually) for upstream and downstream polluting firms before and after 2003. It shows that in fact, the complementarity between the public facility technology and the organizational change is shown to play the strongest role in reducing corporate corruption.

Table 4. Difference-in-Discontinuities Estimates of Upstream-Downstream Corruptive Gap Before and After 2003

Method	(1)	(2)	(3)
Conventional	-10.94*	-10.71*	-18.68*
	(5.04)	(5.01)	(7.94)
Bias-corrected	-13.31**	-13.04**	-21.72**
	(5.04)	(5.01)	(7.94)
Robust	-13.31*	-13.04*	-21.72*
	(6.40)	(6.10)	(8.48)
Observations	6,155	6,155	6,155
Kernel	Triangular	Epanechnikov	Uniform
Bandwidth	3.754	3.920	2.572

Note: Each cell represents a separate difference-in-discontinuities estimate: the differences between corruption discontinuity before and after 2003. Data are from CPES. The running variable is the distance between the county center of a firm and a monitoring station. Positive (negative) distance means firms are located upstream (downstream). Discontinuities at monitoring stations are estimated using methods proposed by Calonico et al. (2014) and MSE optimal bandwidth proposed by Calonico et al. (2014) for different kernel weighting methods. Year-fixed effects are included in the estimation. Standard errors clustered at the monitoring station level are reported below the estimates. * significant at 5% ** significant at 1% *** significant at 0.1%.

We then test whether our findings simply capture the effect of organizational change alone, rather than the indirect effect of the water monitoring stations or the complementary effect. In other words, we check whether misconduct behaviors would be reduced even in the absence of public infrastructure technology, so long as the administrations where the firms are located place a strong emphasis on environmental protection. To determine whether our results are solely attributable to organizational change, we conduct a difference-in-differences analysis. We notice that the intensity of the organizational change may vary across regions. We could then compare the difference in corruption for firms located in regions where organizational change was more applicable before and after the implementation of organizational change with the difference in corruption for firms located in regions where the organizational change was less applicable. The estimates capture the effect of organizational change alone.

We construct a measurement to indicate whether the firms belong to the treatment group (i.e., organizational change is more applicable) or control group (i.e., organizational change is less applicable). We develop a variable, *Long River*, which functions as a proxy for the likelihood that the organizational change will be implemented. Whilst the implementation of organizational change is executed at a national level, the intensity of the policy may differ across regions. As an illustration, the total length of all rivers within a given jurisdiction correlates positively with the degree of attention paid to issues of water pollution. In extreme cases, if a province lacks any rivers, it is possible that the issue of water pollution may not garner significant attention or concern. On the contrary, in provinces possessing a substantial number of rivers, greater emphasis will be placed on concerns pertaining to water pollution. Consequently, regions with longer rivers are more likely to implement stringent organizational changes regarding water pollution. Column (1) of Table 5 shows the results where *Long River* is a dummy variable indicating whether the firms are located in places where the length of river is above the median level of river length. The interaction term *Long River* $\times$ *After* is not significantly different from zero. Therefore, we do not find supporting evidence that organizational change alone is sufficient to reduce firms' corruption.

To eliminate the concern that the length of the river may be correlated with the number of water monitoring stations, we restrict the sample in column (2) of Table 5 to firms located in regions without water monitoring stations. The interaction term *Long River* $\times$ *After* is not significantly different from zero. Hence, there is a lack of evidence to substantiate the claim that organizational change alone is sufficient to reduce firms' corruption.

Table 5. The Effect of Organizational Change alone on Corruption

	(1)	(2)
Long River × After	8.408	4.620
	(6.316)	(2.276)
Long River	-0.913	-0.789
	(2.255)	(2.105)
After	3.867	-1.068

	(3.346)	(1.138)
Log Likelihood	-49421.1	-3561.1
Observations	7,160	705

Note: The data are from CPES. The dependent variables are ETC. Long River is a dummy variable indicating whether the firms are located in places where the length of the river is above the median level. After is a dummy variable indicating whether the time is after 2003. Samples for column (2) exclude firms located in administrations that have water monitoring stations. * significant at 5% ** significant at 1% *** significant at 0.1%.

In summary, we discover that technological resources alone, and the organizational change in government alone, fail to explain the indirect effect of public facility technology on corruption. Rather, the complementarity between public facility technology and organizational change has been demonstrated to play the role in reducing corporate abuse/business-government collusion.

6. Robustness Checks and Additional Analyses

In this section, we conduct additional robustness checks including (1) a manipulation check, (2) alternative estimation methods, (3) alternative radius circles, (4) alternative bandwidth estimation, (5) placebo tests, (6) inclusion of covariates, (7) heterogeneity analysis, (8) alternative samples, and (9) city-level analysis. More details can be found in Online Appendix.

A potential concern about our main analysis is that polluting firms might avoid locating above monitoring stations and instead, they moved downstream to escape regulations and fines. We test the distribution of polluting firms across monitoring stations (Table A5 in the Appendix) following the procedures proposed in Cattaneo et al. (2020). We find no discontinuity in the distribution of polluting firms across monitoring stations, which suggests that the possibilities of polluting firms relocating to downstream areas do not confound our main results.

We also conduct our analysis using alternative estimation methods including parametric estimates (Gelman and Imbens 2019) in Table A6 in the Appendix, difference-in-differences approach in Table A7, and Imbens and Wager (2019) estimator in Table A8, and all the results are consistent with our main ones.

To remove the concerns that our results are sensitive to the choice of a 10-km cutoff, we conduct additional tests with the cutoffs of 20-km (Panel A), and 30-km (Panel B) in Table A9. The results are consistent.

The bandwidth chosen in our main result is a common MSE-optimal bandwidth selector (Calonico et al. 2014). To check whether our main findings are sensitive to optimal bandwidth selection methods, we use five alternative bandwidth selectors suggested by Calonico et al. (2018) and report the results in Table A10 in the Appendix, which are highly consistent with our main results in Table 4.

We conduct placebo tests by using fake water monitoring stations. We move the original stations upstream or downstream by 2km (Table A11, Panel A), 3km (Table A11, Panel B), and 4km (Table A11, Panel C), and re-estimate the regression discontinuity models. The results show that the fake relative distance between firms and the placebo stations does not cause discontinuity of corruption at the fabricated cutoff. This test (Table A11 in the Appendix) confirms that the discontinuity of corruption exists only in actual monitoring stations, not placebo stations, thus providing additional evidence supporting our main findings.

In addition, we follow Lee and Lemieux (2010)'s suggestions to include additional covariates as a robustness check. As additional covariates, we include firm sales, firm value-added tax, the logarithm of the number of employees, the logarithm of one plus firm age, and the logarithm of province per capita GDP. Table A12 in the Appendix shows the results, which confirms our main finding.

Given the large variance among different provinces in China in terms of their local economy, leadership, corruption level, and water quality, we conduct the difference-in-discontinuities analysis to investigate the heterogeneity effect (Table A13). We do not find evidence that corruption discontinuity between upstream and downstream polluting firms differs in terms of their social-economic condition (Panel A), regions' centralization level (Panel C), and water quality (Panel E). We observe that regions with politically motivated leaders experience larger corruption reduction gaps than regions with non-politically motivated leaders (Panel B). Moreover, high corruption regions show larger corruption reduction gaps than low corruption regions (Panel D). These findings offer additional corroborating evidence.

We further show the robustness of our results using alternative samples. We conduct our analysis using samples excluding the water monitoring stations located at the border of the provinces in Table A14, samples including ambiguous firms which are located upstream of one water monitoring station and at the same time also located downstream of another water monitoring station in Table A15, and samples including missing ETC in Table A16. All the findings are consistent with our main results.

7. Conclusion

This study shed light on the indirect value of public infrastructure technology in reducing corporate corruption. By investigating the interplay between public infrastructure technology, organizational change, and corruption, we provide insights into the indirect value of IT and its impact on ethical business practices.

In the early 1990s and early 2000s, the Chinese central government set up hundreds of water monitoring stations for scientific purposes. We leverage the fact that water monitoring stations can detect, measure, and report pollution only for upstream polluting firms. Exploiting this spatial discontinuity, we implement a regression discontinuity design to compare the corruptive activities of upstream and downstream polluting firms. We discover that upstream polluting firms have significantly less corruption costs than downstream polluting firms, though water monitoring stations were not explicitly designed for reducing corporate corruption. We conduct multiple robustness checks to confirm the validity of our regression discontinuity design and the results. These results highlight the unintended consequences of public infrastructure technology.

When analyzing the underlying mechanism, our empirical analysis shows that technology adoption alone is insufficient to reduce corporate corruption. Instead, we demonstrate that the effectiveness of public infrastructure technology in combating corporate corruption is contingent upon the extent to which it is complemented by organizational change. Moreover, it is noteworthy that organizational change in isolation is inadequate to achieve such impact as well. This highlights the need for a holistic approach that combines technological advancements in public infrastructure with organizational change.

Importantly, our study extends the literature on IT value, which has predominantly concentrated on the direct value of IT (Brynjolfsson 1996, Brynjolfsson and Hitt 1996, Melville et al. 2004). Instead, our research endeavors to shed light on the indirect effects of public infrastructure technology that have been adopted for specific purposes. Furthermore, our research moves beyond the traditional focus on individuals' IT adoption (Burtch et al. 2018, Allcott et al. 2020, Li et al. 2022, Wang and Overby 2022) or organizations' IT adoption (Dong et al. 2009, Gómez et al. 2017, Jia et al. 2020, Ganju et al. 2022) and explores the broader context of public infrastructure technology. By doing so, we contribute to a more comprehensive understanding of the potential of IT to drive ethical behavior and integrity.

The implications of our study are significant for policymakers and practitioners involved in anti-corruption efforts. By recognizing the indirect value of public infrastructure technology, policymakers can prioritize investments in technology-enabled public facilities and leverage their potential to create a conducive environment for ethical business practices. Additionally, practitioners can incorporate organizational change initiatives that align with and amplify the benefits of public infrastructure technology in reducing corruption.

Although our finding may not be directly generalized to other countries with different political systems, the impact of public infrastructure technology on corporation corruption is substantial and meaningful in China, where the direct costs of corruption are estimated to amount to approximately three percent of the country's GDP annually (Pei 2007). The present study establishes and quantifies the indirect benefits of public infrastructure technology in mitigating corporate corruption. Such indirect effects of technology are often poorly measured (Brynjolfsson et al. 2019). Policymakers may find it necessary to consider these indirect effects when estimating the economic and societal effects of these technologies.

Our work has some limitations. First, we use ETC, an aggregate measurement of all corruptive activities, as our proxy for corruption but we are unable to establish different corruption categories. Second, our time window was relatively short. Firms may need time to adjust their production, investment, and corruptive activities. It would be interesting for future research to examine if there are long-term effects or if the reduced corruption can spill over to influence downstream firms.

References

Allcott, H., Braghieri, L., Eichmeyer, S., Gentzkow, M. (2020). The Welfare Effects of Social Media. *American Economic Review*. 110(3) 629-676.

Babar, Y., Burtch, G. (2020). Examining the Heterogeneous Impact of Ride-Hailing Services on Public Transit Use. *Information Systems Research*. 31(3) 820-834.

Bharadwaj, S., Bharadwaj, A., Bendoly, E. (2007). The Performance Effects of Complementarities between Information Systems, Marketing, Manufacturing, and Supply Chain Processes. *Information systems research*. 18(4) 437-453.

Brandt, L., Van Biesebroeck, J., Zhang, Y. (2012). Creative Accounting or Creative Destruction? Firm-Level Productivity Growth in Chinese Manufacturing. *Journal of Development Economics*. 97(2) 339-351.

Brynjolfsson, E. (1996). The Contribution of Information Technology to Consumer Welfare. *Information Systems Research*. 7(3) 281-300.

Brynjolfsson, E., Collis, A., Eggers, F. (2019). Using Massive Online Choice Experiments to Measure Changes in Well-Being. *Proceedings of the National Academy of Sciences*. 116(15) 7250-7255.

Brynjolfsson, E., Hitt, L. (1996). Paradox Lost? Firm-Level Evidence on the Returns to Information Systems Spending. *Management science*. 42(4) 541-558.

Burtch, G., Carnahan, S., Greenwood, B.N. (2018). Can You Gig It? An Empirical Examination of the Gig Economy and Entrepreneurial Activity. *Management science*. 64(12) 5497-5520.

Cai, H., Fang, H., Xu, L.C. (2011). Eat, Drink, Firms, Government: An Investigation of Corruption from the Entertainment and Travel Costs of Chinese Firms. *The Journal of Law and Economics*. 54(1) 55-78.

Calonico, S., Cattaneo, M.D., Farrell, M.H. (2018). On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference. *Journal of the American Statistical Association*. 113(522) 767-779.

Calonico, S., Cattaneo, M.D., Titiunik, R. (2014). Robust Nonparametric Confidence Intervals for Regression‐Discontinuity Designs. *Econometrica*. 82(6) 2295-2326.

Campante, F., Yanagizawa-Drott, D. (2018). Long-Range Growth: Economic Development in the Global Network of Air Links. *The Quarterly Journal of Economics*. 133(3) 1395-1458.

Cattaneo, M.D., Jansson, M., Ma, X. (2020). Simple Local Polynomial Density Estimators. *Journal of the American Statistical Association*. 115(531) 1449-1455.

Chakravarty, A., Grewal, R., Sambamurthy, V. (2013). Information Technology Competencies, Organizational Agility, and Firm Performance: Enabling and Facilitating Roles. *Information systems research*. 24(4) 976-997.

Chan, J., Ghose, A., Seamans, R. (2016). The Internet and Racial Hate Crime. *Mis Quarterly*. 40(2) 381-404.

Chan, J., Mojumder, P., Ghose, A. (2019). The Digital Sin City: An Empirical Study of Craigslist's Impact on Prostitution Trends. *Information Systems Research*. 30(1) 219-238.

Cheng, Z., Pang, M.-S., Pavlou, P.A. (2020). Mitigating Traffic Congestion: The Role of Intelligent Transportation Systems. *Information Systems Research*. 31(3) 653-674.

Cheung, Y.-L., Rau, P.R., Stouraitis, A. (2021). What Determines the Return to Bribery? Evidence from Corruption Cases Worldwide. *Management Science*. 67(10) 6235-6265.

Cull, R., Li, W., Sun, B., Xu, L.C. (2015). Government Connections and Financial Constraints: Evidence from a Large Representative Sample of Chinese Firms. *Journal of Corporate Finance*. 32(2015) 271-294.

Dewan, S., Kraemer, K.L. (2000). Information Technology and Productivity: Evidence from Country-Level Data. *Management science*. 46(4) 548-562.

Dong, S., Xu, S.X., Zhu, K.X. (2009). Research Note—Information Technology in Supply Chains: The Value of It-Enabled Resources under Competition. *Information systems research*. 20(1) 18-32.
Ehrlich, M.v., Seidel, T. (2018). The Persistent Effects of Place-Based Policy: Evidence from the West-German Zonenrandgebiet. *American Economic Journal: Economic Policy*. 10(4) 344-374.
Faccio, M. (2006). Politically Connected Firms. *American economic review*. 96(1) 369-386.
Fan, J.P., Wong, T.J., Zhang, T. (2007). Politically Connected Ceos, Corporate Governance, and Post-Ipo Performance of China's Newly Partially Privatized Firms. *Journal of financial economics*. 84(2) 330-357.
Fisman, R., Wang, Y. (2015). The Mortality Cost of Political Connections. *The Review of Economic Studies*. 82(4) 1346-1382.
Ganju, K.K., Atasoy, H., Pavlou, P.A. (2022). Do Electronic Health Record Systems Increase Medicare Reimbursements? The Moderating Effect of the Recovery Audit Program. *Management Science*. 68(4) 2889-2913.
Gao, G., Hitt, L.M. (2012). Information Technology and Trademarks: Implications for Product Variety. *Management Science*. 58(6) 1211-1226.
Ge, J., Carney, M., Kellermanns, F. (2019). Who Fills Institutional Voids? Entrepreneurs' Utilization of Political and Family Ties in Emerging Markets. *Entrepreneurship Theory and Practice*. 43(6) 1124-1147.
Gelman, A., Imbens, G. (2019). Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs. *Journal of Business & Economic Statistics*. 37(3) 447-456.
Gómez, J., Salazar, I., Vargas, P. (2017). Does Information Technology Improve Open Innovation Performance? An Examination of Manufacturers in Spain. *Information Systems Research*. 28(3) 661-675.
Greenwood, B.N., Agarwal, R. (2016). Matching Platforms and Hiv Incidence: An Empirical Investigation of Race, Gender, and Socioeconomic Status. *Management Science*. 62(8) 2281-2303.
Greenwood, B.N., Wattal, S. (2017). Show Me the Way to Go Home. *MIS quarterly*. 41(1) 163-188.
Grembi, V., Nannicini, T., Troiano, U. (2016). Do Fiscal Rules Matter? *American Economic Journal: Applied Economics*. 8(3) 1-30.
Griffin, J.M., Liu, C., Shu, T. (2022). Is the Chinese Anticorruption Campaign Authentic? Evidence from Corporate Investigations. *Management Science*. 68(10) 7248-7273.
Grover, V., Kohli, R. (2012). Cocreating It Value: New Capabilities and Metrics for Multifirm Environments. *Mis Quarterly* 225-232.
Guo, F., Li, Y., Maruping, L.M., Masli, A. (2022). Complementarity between Investment in Information Technology (It) and It Human Resources: Implications for Different Types of Firm Innovation. *Information Systems Research*.
Hahn, J., Todd, P., Van der Klaauw, W. (2001). Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design. *Econometrica*. 69(1) 201-209.
Havakhor, T., Sabherwal, R., Steelman, Z.R., Sabherwal, S. (2019). Relationships between Information Technology and Other Investments: A Contingent Interaction Model. *Information Systems Research*. 30(1) 291-305.
He, G., Wang, S., Zhang, B. (2020). Watering Down Environmental Regulation in China. *The Quarterly Journal of Economics*. 135(4) 2135-2185.
Hitt, L.M., Brynjolfsson, E. (1996). Productivity, Business Profitability, and Consumer Surplus: Three Different Measures of Information Technology Value. *MIS quarterly* 121-142.
Huang, J., Li, K. (2013). Eat-Drink, Corruption and Firm's Purchase Order. *Economic Research Journal (Jing-Ji-Yan-Jiu)*. 48(6) 71-84.
Huang, Z., Li, L., Ma, G., Xu, L.C. (2017). Hayek, Local Information, and Commanding Heights: Decentralizing State-Owned Enterprises in China. *American Economic Review*. 107(8) 2455-2478.

Imbens, G., Wager, S. (2019). Optimized Regression Discontinuity Designs. *Review of Economics and Statistics*. 101(2) 264-278.
Imbens, G.W., Lemieux, T. (2008). Regression Discontinuity Designs: A Guide to Practice. *Journal of Econometrics*. 142(2) 615-635.
Ito, K., Zhang, S. (2020). Willingness to Pay for Clean Air: Evidence from Air Purifier Markets in China. *Journal of Political Economy*. 128(5) 1627-1672.
Jia, N. (2014). Are Collective Political Actions and Private Political Actions Substitutes or Complements? Empirical Evidence from China's Private Sector. *Strategic Management Journal*. 35(2) 292-315.
Jia, N., Rai, A., Xu, S.X. (2020). Reducing Capital Market Anomaly: The Role of Information Technology Using an Information Uncertainty Lens. *Management Science*. 66(2) 979-1001.
Jia, R., Nie, H. (2017). Decentralization, Collusion, and Coal Mine Deaths. *Review of Economics and Statistics*. 99(1) 105-118.
Kohli, R., Devaraj, S. (2003). Measuring Information Technology Payoff: A Meta-Analysis of Structural Variables in Firm-Level Empirical Research. *Information systems research*. 14(2) 127-145.
Kong, D., Wang, L., Wang, M. (2017). Effects of Anti-Corruption on Firm Performance: Evidence from a Quasi-Natural Experiment in China. *Finance Research Letters*. 23(2017) 190-195.
Lee, D.S., Lemieux, T. (2010). Regression Discontinuity Designs in Economics. *Journal of economic literature*. 48(2) 281-355.
Li, H., Kim, Y., Srinivasan, K. (2022). Market Shifts in the Sharing Economy: The Impact of Airbnb on Housing Rentals. *Management Science*. 68(11) 8015-8044.
Li, H., Meng, L., Wang, Q., Zhou, L.-A. (2008). Political Connections, Financing and Firm Performance: Evidence from Chinese Private Firms. *Journal of development economics*. 87(2) 283-299.
Lin, C., Morck, R., Yeung, B., Zhao, X. (2016). *Anti-Corruption Reforms and Shareholder Valuations: Event Study Evidence from China*. National Bureau of Economic Research.
Liu, B., Pavlou, P.A., Cheng, X. (2022). Achieving a Balance between Privacy Protection and Data Collection: A Field Experimental Examination of a Theory-Driven Information Technology Solution. *Information Systems Research*. 33(1) 203-223.
Liu, J., Pei, S., Zhang, M.X. (2023). Online Food Delivery Platforms and Female Labor Force Participation. *Information Systems Research*. forthcoming.
Melville, N., Kraemer, K., Gurbaxani, V. (2004). Information Technology and Organizational Performance: An Integrative Model of It Business Value. *MIS quarterly* 283-322.
Mishra, A.N., Tao, Y., Keil, M., Oh, J.-h. (2022). Functional It Complementarity and Hospital Performance in the United States: A Longitudinal Investigation. *Information Systems Research*. 33(1) 55-75.
Mithas, S., Krishnan, M.S., Fornell, C. (2016). Research Note—Information Technology, Customer Satisfaction, and Profit: Theory and Evidence. *Information Systems Research*. 27(1) 166-181.
Mittal, N., Nault, B.R. (2009). Research Note—Investments in Information Technology: Indirect Effects and Information Technology Intensity. *Information systems research*. 20(1) 140-154.
Mukhopadhyay, T., Kekre, S., Kalathur, S. (1995). Business Value of Information Technology: A Study of Electronic Data Interchange. *MIS quarterly* 137-156.
Nevo, S., Wade, M.R. (2010). The Formation and Value of It-Enabled Resources: Antecedents and Consequences of Synergistic Relationships. *MIS quarterly* 163-183.
Oh, W., Pinsonneault, A. (2007). On the Assessment of the Strategic Value of Information Technologies: Conceptual and Analytical Approaches. *MIS quarterly* 239-265.
Osburg, J. (2018). Making Business Personal: Corruption, Anti-Corruption, and Elite Networks in Post-Mao China. *Current Anthropology*. 59(S18) S149-S159.
Ozer, G.T., Greenwood, B.N., Gopal, A. (2023). Digital Multisided Platforms and Women's Health: An Empirical Analysis of Peer-to-Peer Lending and Abortion Rates. *Information Systems Research*. 34(1) 223-252.
Park, J., Pang, M.-S., Kim, J., Lee, B. (2021). The Deterrent Effect of Ride-Sharing on Sexual Assault and Investigation of Situational Contingencies. *Information Systems Research*. 32(2) 497-516.

Pei, M. (2007). Corruption Threatens China's Future.
Rai, A., Patnayakuni, R., Seth, N. (2006). Firm Performance Impacts of Digitally Enabled Supply Chain Integration Capabilities. *MIS quarterly* 225-246.
Ravichandran, T., Giura, S.I. (2019). Knowledge Transfers in Alliances: Exploring the Facilitating Role of Information Technology. *Information Systems Research*. 30(3) 726-744.
Ravichandran, T., Han, S., Mithas, S. (2017). Mitigating Diminishing Returns to R&D: The Role of Information Technology in Innovation. *Information Systems Research*. 28(4) 812-827.
Rhee, K.S., Zheng, J., Wang, Y., Tan, Y. (2022). Value of Information Sharing Via Ride-Hailing Apps: An Empirical Analysis. *Information Systems Research*.
Rudai, Y. (2015). Study on the Total Factor Productivity of Chinese Manufacturing Enterprises. *Economic Research Journal*. 2(2015) 61-74.
Seligson, M.A. (2002). The Impact of Corruption on Regime Legitimacy: A Comparative Study of Four Latin American Countries. *The Journal of Politics*. 64(2) 408-433.
Tafti, A., Mithas, S., Krishnan, M.S. (2013). The Effect of Information Technology–Enabled Flexibility on Formation and Market Value of Alliances. *Management science*. 59(1) 207-225.
Trantopoulos, K., von Krogh, G., Wallin, M.W., Woerter, M. (2017). External Knowledge and Information Technology. *MIS quarterly*. 41(1) 287-300.
Wang, H., Overby, E.M. (2022). How Does Online Lending Influence Bankruptcy Filings? *Management Science*. 68(5) 3309-3329.
Wei, C., Hu, S., Chen, F. (2020). Do Political Connection Disruptions Increase Labor Costs in a Government-Dominated Market? Evidence from Publicly Listed Companies in China. *Journal of Corporate Finance*. 62(2020) 101554.
World Bank. (1997). *Helping Countries Combat Corruption: The Role of the World Bank*. World Bank Washington, DC.
Yao, Y., Dresner, M., Zhu, K.X. (2019). "Monday Effect" on Performance Variations in Supply Chain Fulfillment: How Information Technology–Enabled Procurement May Help. *Information Systems Research*. 30(4) 1402-1423.
You, J., Nie, H. (2017). Who Determines Chinese Firms' Engagement in Corruption: Themselves or Neighbors? *China Economic Review*. 43(2017) 29-46.
Zhao, H., Lu, J. (2016). Contingent Value of Political Capital in Bank Loan Acquisition: Evidence from Founder-Controlled Private Enterprises in China. *Journal of Business Venturing*. 31(2) 153-174.
Zhu, J., Wu, Y. (2014). Who Pays More "Tributes" to the Government? Sectoral Corruption of China's Private Enterprises. *Crime, Law and Social Change*. 61(3) 309-333.

Online Appendix

Appendix A: Robustness Checks and Additional Analyses

In Section 4, we confirm the validity of our regression discontinuity design by showing that upstream and downstream firms were well-balanced in the CPES and the ASIF datasets. In this appendix, we conduct additional robustness checks: (1) a manipulation check, (2) alternative estimation methods, (3) alternative radius circles, (4) alternative bandwidth estimation, (5) placebo tests, (6) inclusion of covariates, (7) heterogeneity analysis, and (8) alternative samples.

Manipulation Check

A potential concern about our main analysis is that polluting firms might avoid locating above monitoring stations and instead, they moved downstream to escape regulations and fines. We test the distribution of polluting firms across monitoring stations (Table A5 in the Appendix) following the procedures proposed in Cattaneo et al. (2020). The manipulation check essentially compares the density of polluting firms around the water monitoring stations. We find that there is no discontinuity in the distribution of polluting firms around the monitoring stations. If firms were strategically located downstream to avoid detection by the monitoring stations, we would observe fewer polluting firms upstream. However, we find no discontinuity in the distribution of polluting firms across monitoring stations, suggesting that the possibilities of firms relocating to downstream areas do not confound our main results.

Alternative Estimation Methods

In Section 5, we report non-parametric estimation results. To check whether our main result is sensitive to the use of a non-parametric approach, we also report parametric estimates (Gelman and Imbens 2019) in Table A6 in the Appendix. We estimate regression discontinuity using linear, quadratic, and cubic functions to check whether the estimates are sensitive to the order of polynomial functions. The results are consistent with those in Table 2. For private firms in polluting industries after 2003, the regression discontinuity estimates are negative and statistically significant (Table A6 in the Appendix, column (1) – (3)), while for private firms in non-polluting industries, the estimates are statistically nonsignificant (Table A6 in the Appendix, column (4) – (6)).

In our main result, we use the regression discontinuity method to account for the fact that the closer a firm is to the monitoring station, the more sensitive that firm will be to the complementarities between technology and organizational change. We now consider a difference-in-differences approach to investigate the research question. In Table A7, the interaction term of the upstream dummy and polluting industries dummy remains negative and statistically significant.

We also present the results of an alternative regression discontinuity estimator – Imbens and Wager (2019) estimator. This estimator is fully data-driven and calculates "optimal" weights for each observation. In addition, the estimator is defined regardless of the shape of the treatment region and is not affected by the potential discreteness of the running variable. The results are presented in Table A8 and are consistent with our main results.

Alternative Radius Circles

Recall that in our main analysis, we draw a 10-km circle around the water monitoring stations to select the towns within the range. To remove the concerns that our results are sensitive to the choice of a 10-km cutoff, we conduct additional tests with the cutoffs of 20-km (Panel A), and 30-km (Panel B) in Table A9. The results are consistent. We observe statistically significant upstream and downstream corruption gaps for polluting industry (columns (1) – (3)) but not for non-polluting industries (columns (4) – (6)).

Alternative Bandwidth Estimation

The bandwidth chosen in our main result is a common MSE-optimal bandwidth selector (Calonico et al. 2014), which minimizes the asymptotic mean squared error (MSE) of the average effect of the treatment. To check whether our main findings are sensitive to optimal bandwidth selection methods, we use five alternative bandwidth selectors suggested by Calonico et al. (2018) and report the results in Table A10 in the Appendix. They are (1) MSE-two: This method allows for a different bandwidth below and above the cutoff. It applies MSE-optimal bandwidth selectors method on both sides of the cutoff. (2) MSE-sum: This method uses a common MSE-optimal bandwidth, but the objective function includes the mean squared-errors on both the left and the right of the cutoff point, whereas MSE and MSE-two focus on the mean-squared-error of the difference. (3) CER: This method employs a common bandwidth selector, and aims to minimize the coverage error probability (CER). It is also known as the CER-optimal bandwidth selector. Usually, CER-optimal bandwidth is smaller than the MSE-optimal bandwidth. (4) CER-two: This method allows for different bandwidths below and above the cutoff. It applies CER-optimal bandwidth selectors method on both sides of the cutoff. (5) CER-sum: This method utilizes a common CER-optimal bandwidth, but the objective function includes the sum of mean squared errors on both sides of the cutoff point, whereas in CER and CER-two, the objective function is mean-squared-error of the difference. Technical details are in Calonico et al. (2018). The results in Table A10 in the Appendix are highly consistent with our main results in Table 4, which demonstrates that our main findings are robust to various bandwidth selection methods.

Placebo Tests

We conduct placebo tests by using fake water monitoring stations. We move the original stations upstream or downstream by 2km (Table A11, Panel A), 3km (Table A11, Panel B), and 4km (Table A11, Panel C), and re-estimate the regression discontinuity models. The results show that the fake relative distance and location between firms and the placebo stations do not cause discontinuity of corruption at the fabricated cutoff. This test (Table A11 in the Appendix) confirms that the discontinuity of corruption exists only in actual monitoring stations, not placebo stations, providing additional evidence supporting our main findings.

Inclusion of covariates

Although we have checked balances between the upstream and downstream firms, we now follow Lee and Lemieux (2010)'s suggestions to include additional covariates. If our research design is valid, the additional covariates should have little effect on the estimation. As additional covariates, we include firm sales, firm value-added tax, the logarithm of the number of employees, the logarithm of one plus firm age, and the logarithm of province per capita GDP. Table A12 in the Appendix shows results confirming our main findings: polluting firms show negative and statistically significant upstream-downstream gaps (Table A12, columns (1) – (3)), but non-polluting firms show statistically nonsignificant gaps (Table A12, columns (4) – (6)).

Heterogeneity Analysis

Given the large variance among different provinces in China in terms of their local economy, leadership, corruption level, and water quality, we conduct the difference-in-discontinuities analysis to investigate the heterogeneity effect (Table A13). Specifically, we analyze the differences in corruption discontinuity between high GDP regions and low GDP regions (Panel A), the differences in corruption discontinuity between politically motivated leaders and non-politically motivated leaders (Panel B), the differences in corruption discontinuity between centralized regions and less centralized regions (Panel C), the differences in corruption discontinuity between high corruption regions and low corruption regions (Panel D), and the differences in corruption discontinuity between high water pollution regions and low water pollution regions (Panel E). We do not find evidence that corruption discontinuity between upstream and downstream polluting firms differs in terms of their social-economic condition (Panel A), regions' centralization level (Panel C), and water quality (Panel E).

We observe that regions with politically motivated leaders experience larger corruption reduction gaps than regions with non-politically motivated leaders (Panel B). Moreover, high corruption regions show larger corruption reduction gaps than low corruption regions (Panel D). These findings offer additional corroborating evidence.

Alternative Samples

It is possible that upstream firms and downstream firms are governed by different politicians if the water monitoring stations are located at the boundary of provinces. To remove the concern, we conduct the difference-in discontinuities analysis after excluding the water monitoring stations located at the border of the provinces. The results are shown in Table A14, and are consistent with our main results.

In the main analysis, we removed ambiguous firms which are located upstream of one water monitoring station and at the same time also located downstream of another water monitoring station. To alleviate the concern that our results are sensitive to this, we reconduct the analysis including these ambiguous firms. The results are shown in Table A15 and are consistent with our main results.

In our data cleaning process, we dropped firms with missing ETC. To alleviate concerns about the data cleaning process, we reconduct the analysis including the firms with missing ETC and treated their ETC as 0. The results (Table A16) remain consistent with our main results.

Appendix B: Tables

Table A1. The link between ETC and Firms' Actual Misconduct

	Number of Regulation Breaches
ETC	0.014***
	(0.002)
Log Likelihood	-24811.4
Observations	32,459

Note: The dependent variable is the number of regulation breaches for a firm in a year from China Stock Market & Accounting Research Database. The independent variable is ETC (million RMB) from Wind Database. Poisson regression is employed. * significant at 5% ** significant at 1% *** significant at 0.1%.

Table A2. Covariate Balance Between Upstream and Downstream Firms

	Mean		Mean Difference
	Downstream	Upstream	
	(1)	(2)	(3)
Panel A: ASIF			
Year of Opening	1994.12	1993.02	1.100
	(11.98)	(13.90)	(1.082)
Polluting industries	0.28	0.28	-0.001
(1=Yes, 0=Others)	(0.45)	(0.45)	(0.023)
Profit	5583.67	4451.76	1131.909
(1,000 RMB)	(144712.98)	(140953.98)	(2514.313)
Value-Added Tax	3685.50	3072.04	613.459
(1,000 RMB)	(42045.42)	(34363.86)	(702.502)
# of Employees	258.11	240.56	17.541
(Male)	(1042.80)	(1078.64)	(33.022)
# of Employees	102.83	95.38	7.454
(Female)	(332.37)	(360.07)	(12.175)
Capital Stock	40620.32	26047.07	14573.255
(1,000 RMB)	(658851.78)	(155018.10)	(8121.275)
Intermediate Input	75408.55	71480.62	3927.938
(1,000 RMB)	(874074.76)	(572838.81)	(14806.010)
Panel B: CPES			
Year of Opening	1998.99	1997.99	0.998
	(5.77)	(5.24)	(0.635)

Sales	7,796.03	6,394.30	1,401.732
(10000 RMB)	(102,832.91)	(67,007.83)	(2,510.164)
Tax	268.30	277.56	-9.267
(10000 RMB)	(1,740.17)	(3,570.98)	(84.882)
Profit	279.99	354.42	-74.426
(10000 RMB)	(1,831.14)	(3,349.80)	(87.264)

Note: Columns (1)–(2) report the means and standard deviations of firm characteristics. In column (3), we test the covariate balance between upstream and downstream firms. The difference coefficients are obtained by running OLS regressions of firm characteristics on an upstream dummy. Standard errors reported in the parentheses are clustered at the water monitoring station level. * significant at 5% ** significant at 1% *** significant at 0.1%.

Table A3. Industry Balance Between Upstream and Downstream Firms

	Mean		Mean Difference
	Downstream	Upstream	
	(1)	(2)	(3)
Agricultural and Sideline Food Processing	0.03	0.02	0.011
Ind. Code: 13	(0.18)	(0.14)	(0.007)
Food Manufacturing	0.02	0.02	0.001
Ind. Code: 14	(0.15)	(0.15)	(0.007)
Beverage Manufacturing	0.01	0.01	-0.003
Ind. Code: 15	(0.10)	(0.11)	(0.004)
Textile Mills	0.09	0.07	0.017
Ind. Code: 17	(0.29)	(0.26)	(0.036)
Apparel and Clothing Accessories Manufacturing	0.05	0.06	-0.014
Ind. Code: 18	(0.21)	(0.24)	(0.022)
Leather, Fur, and Related Product Manufacturing	0.04	0.01	0.031
Ind. Code: 19	(0.19)	(0.08)	(0.016)
Wood and Bamboo Products Manufacturing	0.01	0.01	0.005
Ind. Code: 20	(0.11)	(0.08)	(0.003)
Furniture Manufacturing	0.02	0.00	0.013*
Ind. Code: 21	(0.13)	(0.07)	(0.005)
Paper Products Manufacturing	0.02	0.02	0.006
Ind. Code: 22	(0.15)	(0.13)	(0.006)
Printing and Reproduction of Recorded Media	0.03	0.03	-0.002
Ind. Code: 23	(0.18)	(0.18)	(0.009)
Education and Entertainment Articles Manufacturing	0.01	0.01	-0.001
Ind. Code: 24	(0.08)	(0.08)	(0.004)
Petrochemicals Manufacturing	0.01	0.01	0.003

Ind. Code: 25	(0.11)	(0.09)	(0.005)
Chemical Products Manufacturing	0.06	0.08	-0.027
Ind. Code: 26	(0.23)	(0.28)	(0.018)
Medical Goods Manufacturing	0.03	0.03	-0.001
Ind. Code: 27	(0.16)	(0.16)	(0.006)
Rubber Products Manufacturing	0.03	0.01	0.020
Ind. Code: 29	(0.18)	(0.11)	(0.022)
Plastic Products Manufacturing	0.05	0.04	0.003
Ind. Code: 30	(0.21)	(0.20)	(0.009)
Non-Metallic Mineral Products Manufacturing	0.04	0.05	-0.005
Ind. Code: 31	(0.21)	(0.22)	(0.012)
Basic Metal Processing	0.01	0.01	-0.002
Ind. Code: 32	(0.11)	(0.12)	(0.006)
Non-Ferrous Metal Processing	0.02	0.02	-0.004
Ind. Code: 33	(0.13)	(0.14)	(0.006)
Fabricated Metal Products Manufacturing	0.04	0.06	-0.018
Ind. Code: 34	(0.21)	(0.24)	(0.011)
General Purpose Machinery Manufacturing	0.07	0.10	-0.025
Ind. Code: 35	(0.26)	(0.30)	(0.014)
Special Purpose Machinery Manufacturing	0.05	0.07	-0.022*
Ind. Code: 36	(0.21)	(0.25)	(0.011)
Transport Equipment Manufacturing	0.06	0.06	-0.004
Ind. Code: 37	(0.24)	(0.24)	(0.012)
Electrical Equipment Manufacturing	0.08	0.07	0.014
Ind. Code: 39	(0.27)	(0.25)	(0.015)
Computers and Electronic Products Manufacturing	0.03	0.04	-0.010
Ind. Code: 40	(0.18)	(0.20)	(0.011)
General Instruments and Other Equipment Manufacturing	0.02	0.03	-0.013
Ind. Code: 41	(0.14)	(0.18)	(0.012)
Craftworks Manufacturing	0.02	0.01	0.011
Ind. Code: 42	(0.13)	(0.08)	(0.008)
Electricity and Heat Supply	0.02	0.01	0.009
Ind. Code: 44	(0.14)	(0.11)	(0.005)
Water Production and Supply	0.01	0.00	0.007**
Ind. Code: 46	(0.10)	(0.06)	(0.002)

Note: Columns (1)–(2) report the means and standard deviations of firm characteristics. In columns (3), we test the covariate balance between upstream and downstream firms within 5km of water monitoring stations. The difference coefficients are obtained by running OLS regressions of firm characteristics on an upstream dummy. Standard errors reported in the parentheses are clustered at the water monitoring station level. * significant at 5% ** significant at 1% *** significant at 0.1%.

Table A4. Summary Statistics of Entertainment and Travel Costs and Company's Location

Variable	Mean	Standard Deviation	Min	Max
CPES (10000 RMB)				
ETC	14.61	240.76	0	19,450
ASIF (1000 RMB)				
ETC	146.07	1,153.20	0	249,405
Distance to nearest water monitoring station (meters)	27,444.2	32,315.6	13.65	3,519,410
Distance to the second nearest water monitoring station (meters)	46,846.9	38,270.6	13.65	3,520,434

Table A5. Regression Discontinuity Manipulation Tests

	(1)	(2)
T	0.04	-1.59
P>|T|	0.97	0.11
Bandwidth Left	1,845.18	2,462.04
Bandwidth Right	1,687.48	2,462.04
Observations	1,207	1,207
Bandwidth Selectors	Each	Diff

Note: This table reports regression discontinuity manipulating tests using the local polynomial density estimators proposed by Cattaneo et al. (2020). The sample consists of ASIF firms located within 5km of water monitoring stations. Two bandwidth selectors are used to test the density discontinuity. "Each" means we use two distinct bandwidths based on MSE of each density separately for upstream and downstream firms. "Diff" bandwidth selection is based on MSE of the difference of densities with one common bandwidth. Technical explanations are in Cattaneo et al. (2020).

Table A6. Parametric Regression Discontinuity Estimation for Private Enterprise

	Polluting industries			Non-Polluting industries		
	(1)	(2)	(3)	(4)	(5)	(6)
RD in Corruption	-72.372 *	-107.742 ***	-103.296 **	-18.099	4.851	-1.303
	(32.121)	(30.961)	(31.938)	(29.227)	(23.105)	(24.470)
Observations	2143	2143	2143	5717	5717	5717
Log Likelihood	-14731.7	-14729.5	-14729.2	-40465.7	-40464.3	-40462.5
Polynomial Function	Linear	Quadratic	Cubic	Linear	Quadratic	Cubic

Note: Each cell represents a separate regression discontinuity regression. The sample consists of ASIF private firms within 10km of water monitoring stations. The running variable is the distance between a firm and a monitoring station. Positive (negative) distance means firms are located upstream (downstream). Negative coefficients indicate that upstream firms have lower ETC. We report OLS estimates of the coefficient on an "upstream" dummy after controlling for polynomial functions in distance from the monitoring stations interacted with an upstream dummy. Standard errors clustered at the monitoring station level are reported below the estimates. * significant at 5% ** significant at 1% *** significant at 0.1%.

Table A7. Difference-in-Differences Result

	(1)
Upstream x Polluting Industries	-67.858 **
	(22.857)
Upstream	-2.138
	(12.001)
Polluting Industries	42.143 *
	(19.033)
Log Likelihood	-22307.2
Observations	3,183

Note: The data are from ASIF firms. The sample consists of private firms within 5km of water monitoring stations. Upstream indicates whether the firms are located upstream of water monitoring stations. Polluting Industries indicate whether the firm belongs to polluting industries. * significant at 5% ** significant at 1% *** significant at 0.1%.

Table A8. Imbens and Wager (2019) Estimators

	Estimator	Confidence Interval	Maximum Bias	Sample Error
Polluting industries	-313.56	-313.56±288.06	0.16	146.97
Non-polluting industries	-12.72	-12.72±280.56	0.12	143.14

Note: Each row represents a separate regression discontinuity regression. Data are private firms from ASIF. The running variable is the distance between a firm and a monitoring station. A positive (negative) distance means the firm is located upstream (downstream). The negative coefficients indicate that upstream firms have lower ETC. The discontinuities at monitoring stations are estimated using methods proposed by Imbens and Wager (2019). Reported are bias-adjusted 95% confidence intervals, a bound on the maximum bias, and an estimate of the sampling error.

Table A9. Alternative Radius Circles Around the Monitoring Stations

	Polluting Industries			Non-Polluting Industries		
Method	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Within 20-km Radius			
Robust	-92.04*	-99.28*	-115.30**	-19.77	-18.91	-29.27
	(41.15)	(39.47)	(38.02)	(26.01)	(26.77)	(28.97)
Observations	6,397	6,397	6,397	16,117	16,117	16,117
Kernel	Triangular	Epanechnikov	Uniform	Triangular	Epanechnikov	Uniform
Bandwidth	8528	8461	7548	11034	10600	9313
			Panel B: Within 30-km Radius			
Robust	-95.94**	-102.51**	-122.45***	-18.67	-19.54	-5.30
	(36.36)	(36.65)	(36.64)	(31.29)	(32.12)	(29.94)
Observations	10,978	10,978	10,978	26,074	26,074	26,074
Kernel	Triangular	Epanechnikov	Uniform	Triangular	Epanechnikov	Uniform
Bandwidth	9164	8749	8581	12001	11203	14224

Note: Each cell represents a separate regression. Data are private firms from ASIF. The running variable is the distance between a firm and a monitoring station. Positive (negative) distance means firms are located upstream (downstream). Negative coefficients indicate that upstream firms have lower ETC. Discontinuities at monitoring stations are estimated using methods proposed by Calonico et al. (2014) and MSE optimal bandwidth proposed by Calonico et al. (2014) for different kernel weighting methods.

Standard errors clustered at the monitoring station level are reported below the estimates. * significant at 5% ** significant at 1% *** significant at 0.1%.

Table A10. Alternative Bandwidths for Private Enterprise in Polluting Industries

Bandwidth Selection Method	(1)	(2)	(3)
MSE-Two	-96.65**	-96.07**	-91.62**
	(41.97)	(40.75)	(37.76)
MSE-Sum	-101.01**	-107.19***	-106.72**
	(40.86)	(39.17)	(46.38)
CER-RD	-101.01**	-107.19***	-106.72**
	(40.86)	(39.17)	(46.38)
CER-Two	-73.55	-73.15*	-84.06**
	(46.77)	(44.40)	(40.34)
CER-Sum	-85.08*	-91.12*	-126.63***
	(51.61)	(50.10)	(38.74)
Kernel	Triangular	Epanechnikov	Uniform
Observations	3,502	3,502	3,502

Note: Each cell represents a separate regression discontinuity regression. Data are from ASIF. The running variable is the distance between a firm and a monitoring station. Positive (negative) distance means firms are located upstream (downstream). Negative coefficients indicate that upstream firms have lower ETC. Discontinuities at monitoring stations are estimated using methods proposed by Calonico et al. (2014) for different kernel weighting methods. We use alternative bandwidth selectors proposed by Calonico et al. (2014). Technical details are in Calonico et al. (2018). Robust estimates are reported. Standard errors clustered at the station level are reported below the estimates. * significant at 10% ** significant at 5% *** significant at 1%.

Table A11. Placebo Test for Private Enterprises in Polluting Industries

	Move Monitoring Stations Downstream			Move Monitoring Stations Upstream		
Method	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Move by 2km						
Robust	-47.51	-51.99	-57.74	5.49	-2.63	-36.15
	(32.73)	(32.96)	(33.35)	(43.94)	(42.56)	(38.67)
Observations	3,502	3,502	3,502	3,502	3,502	3,502
Kernel	Triangular	Epanechnikov	Uniform	Triangular	Epanechnikov	Uniform
Bandwidth	10503	9753	8371	12477	12243	13845
Panel B: Move by 3km						
Robust	-26.92	-26.38	-29.27	12.57	12.87	-7.22
	(30.00)	(30.15)	(31.05)	(40.92)	(39.53)	(36.47)
Observations	3,502	3,502	3,502	3,502	3,502	3,502
Kernel	Triangular	Epanechnikov	Uniform	Triangular	Epanechnikov	Uniform
Bandwidth	10318	9427	8344	13053	12748	12929
Panel C: Move by 4km						
Robust	28.32	24.98	16.56	34.58	43.22	31.20

	(32.81)	(33.07)	(32.88)	(33.10)	(32.71)	(30.84)
Observations	3,502	3,502	3,502	3,502	3,502	3,502
Kernel	Triangular	Epanechnikov	Uniform	Triangular	Epanechnikov	Uniform
Bandwidth	9582	9633	9113	11523	11025	11488

Note: Each cell represents a separate regression. Data are private enterprises in polluting industries from ASIF. The running variable is the distance between a firm and a monitoring station. Positive (negative) distance means firms are located upstream (downstream). Negative coefficients indicate that upstream firms have lower ETC. Discontinuities at monitoring stations are estimated using methods proposed by Calonico et al. (2014) and MSE optimal bandwidth proposed by Calonico et al. (2014) for different kernel weighting methods. Standard errors clustered at the monitoring station level are reported below the estimates. * significant at 5% ** significant at 1% *** significant at 0.1%.

Table A12. Inclusion of Covariates for Private Enterprises

	Polluting Industries			Non-Polluting Industries		
Method	(1)	(2)	(3)	(4)	(5)	(6)
Conventional	-88.81**	-88.93**	-93.48**	-15.62	-3.22	-25.86
	(30.11)	(29.50)	(31.84)	(24.23)	(25.54)	(25.58)
Bias-corrected	-102.64***	-103.11***	-108.46***	-22.12	-3.22	-33.00
	(30.11)	(29.50)	(31.84)	(24.23)	(25.54)	(25.58)
Robust	-102.64**	-103.11**	-108.46**	-22.12	-3.22	-33.00
	(35.33)	(34.42)	(37.07)	(27.52)	(28.66)	(29.18)
Observations	3,499	3,499	3,499	9,276	9,276	9,276
Kernel	Triangular	Epanechnikov	Uniform	Triangular	Epanechnikov	Uniform
Bandwidth	9065	8924	7188	12946	10286	10957

Note: Each cell represents a separate regression discontinuity regression. Data are from ASIF. The running variable is the distance between a firm and a monitoring station. Positive (negative) distance means firms are located upstream (downstream). The negative coefficients indicate that upstream firms have lower ETC. Discontinuities at monitoring stations are estimated using methods proposed by Calonico et al. (2014) and MSE optimal bandwidth proposed by Calonico et al. (2014) for different kernel weighting methods. Covariates include sales, value added tax, log(# of employees), log(1+firm age) and log(per capita GDP). Standard errors clustered at the monitoring station level are reported below the estimates. * significant at 5% ** significant at 1% *** significant at 0.1%.

Table A13. Heterogeneity Effect

Method	(1)	(2)	(3)
	Panel A: Social Economy		
Conventional	-26.06	-17.12	3.60
	(72.21)	(69.59)	(66.53)
Bandwidth	11,837	11,480	10,533
	Panel B: Political Structure		
Conventional	-137.38*	-138.53*	-143.97**
	(73.86)	(72.77)	(59.15)
Bandwidth	13,290	13,387	15,878
	Panel C: Centralized vs Less Centralized		

Conventional	-96.94	-85.67	0.25
	(76.22)	(72.52)	(43.95)
Bandwidth	15,811	15,401	18,933
		Panel D: Corruption	
Conventional	-169.92**	-161.65**	-165.11**
	(73.02)	(73.22)	(72.31)
Bandwidth	11,980	11,014	9,996
		Panel E: Water Quality	
Conventional	-62.15	-53.54	-36.30
	(70.35)	(67.14)	(63.30)
Bandwidth	12285	11921	11174
Observations	3,113	3,113	3,113
Kernel	Triangular	Epanechnikov	Uniform
Bandwidth	3.686	3.987	2.576

Note: Each cell represents a separate difference-in-discontinuities estimate: the differences in corruption discontinuity between high GDP regions and low GDP regions (Panel A), the differences in corruption discontinuity between politically motivated leaders and non-politically motivated leaders (Panel B), the differences in corruption discontinuity between centralized regions and less centralized regions (Panel C), the differences in corruption discontinuity between high corruption regions and low corruption regions (Panel D), and the differences in corruption discontinuity between high water pollution regions and low water pollution regions (Panel E). We define regions with GDP higher than the median GDP as high GDP regions, city officials greater than 60 years old as politically motivated leaders, regions with distance to Capital Beijing less than the median distance as centralized regions (Huang et al. 2017), regions with corruption costs greater than median corruption costs as high corruption region, and regions with COD levels higher than median COD levels as high water pollution regions. Data are from CPES. The running variable is the distance between the county center of a firm and a monitoring station. Positive (negative) distance means firms are located upstream (downstream). Discontinuities at monitoring stations are estimated using methods proposed by Calonico et al. (2014) and MSE optimal bandwidth proposed by Calonico et al. (2014) for different kernel weighting methods. Year-fixed effects are included in the estimation. Standard errors clustered are reported below the estimates. * significant at 10% ** significant at 5% *** significant at 1%.

Table A14. Difference-in-Discontinuities Estimates Excluding Boundary Stations

Method	(1)	(2)	(3)
Conventional	-10.96**	-10.73**	-10.93*
	(5.10)	(5.01)	(5.96)
Bias-corrected	-13.09**	-13.30***	-12.98**
	(5.10)	(5.01)	(5.96)
Robust	-13.09**	-13.30**	-12.98*
	(6.52)	(6.21)	(7.18)
Observations	5,853	5,853	5,853
Kernel	Triangular	Epanechnikov	Uniform
Bandwidth	3.830	3.808	3.471

Note: Each cell represents a separate difference-in-discontinuities estimate: the differences between corruption discontinuity before and after 2003. Data are from CPES after removing water monitoring stations located at the boundary. The running variable is the distance between the county center of a firm and a monitoring station. Positive (negative) distance means firms are located upstream (downstream). Discontinuities at monitoring stations are estimated using methods proposed by Calonico et al. (2014) and MSE optimal bandwidth proposed by Calonico et al. (2014) for different kernel weighting methods. Year-fixed effects are included in the estimation. Standard errors clustered at the monitoring station level are reported below the estimates. * significant at 10% ** significant at 5% *** significant at 1%.

Table A15. Analysis Including Ambiguous Firms

	Polluting Industries			Non-Polluting Industries		
Method	(1)	(2)	(3)	(4)	(5)	(6)
Conventional	-73.76*	-76.92*	-88.07**	-16.54	-17.22	-15.40
	(34.14)	(34.28)	(33.68)	(20.40)	(21.16)	(23.80)
Bias-corrected	-85.48*	-89.98**	-100.83**	-22.30	-23.77	-24.55
	(34.14)	(34.28)	(33.68)	(20.40)	(21.16)	(23.80)
Robust	-85.48*	-89.98*	-100.83*	-22.30	-23.77	-24.55
	(39.94)	(39.90)	(39.95)	(23.21)	(23.82)	(25.50)
Observations	4,320	4,320	4,320	12,221	12,221	12,221
Kernel	Triangular	Epanechnikov	Uniform	Triangular	Epanechnikov	Uniform
Bandwidth	9434	9011	7644	10731	9934	8743

Note: Each cell represents a separate regression discontinuity regression. Data are private firms from ASIF. The running variable is the distance between a firm and a monitoring station. A positive (negative) distance means the firm is located upstream (downstream). The negative coefficients indicate that upstream firms have lower ETC. The discontinuities at monitoring stations are estimated using methods proposed by Calonico et al. (2014) and the MSE optimal bandwidth proposed by Calonico et al. (2014) for different kernel weighting methods. Standard errors clustered at the monitoring station level are reported below the estimates. * significant at 5% ** significant at 1% *** significant at 0.1%.

Table A16. Analysis Including Firms with Missing ETC

	Polluting Industries			Non-Polluting Industries		
Method	(1)	(2)	(3)	(4)	(5)	(6)
Conventional	-92.51**	-95.51**	-107.41**	232.31	237.43	170.32
	(33.56)	(32.80)	(35.12)	(164.79)	(172.99)	(178.67)
Bias-corrected	-105.63**	-109.01***	-125.40***	274.95	280.46	213.54
	(33.56)	(32.80)	(35.12)	(164.79)	(172.99)	(178.67)
Robust	-105.63**	-109.01**	-125.40**	274.95	280.46	213.54
	(36.83)	(35.90)	(38.73)	(193.43)	(204.30)	(207.43)
Observations	3,509	3,509	3,509	889	889	889
Kernel	Triangular	Epanechnikov	Uniform	Triangular	Epanechnikov	Uniform
Bandwidth	11246	10994	7950	7834	7441	6498

Note: Each cell represents a separate regression discontinuity regression. Data are private firms from ASIF. The running variable is the distance between a firm and a monitoring station. A positive (negative) distance means the firm is located upstream (downstream). The negative coefficients indicate that upstream firms have lower ETC. The discontinuities at monitoring stations are estimated using methods proposed by Calonico et al. (2014) and the MSE optimal bandwidth proposed by Calonico et al. (2014) for different kernel weighting methods. Standard errors clustered at the monitoring station level are reported below the estimates. * significant at 5% ** significant at 1% *** significant at 0.1%.

Table A17. The Upstream-Downstream Corruption Gap for Private Enterprises Using Different Confidence Level Cutoffs

	Polluting Industries			Non-Polluting Industries		
Method	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Confidence level greater than 20 (error within 10km)						
Conventional	-89.28**	-94.14**	-98.09**	168.55	157.47	113.10
	(32.85)	(32.51)	(33.06)	(169.85)	(177.90)	(191.16)
Bias-corrected	-101.70**	-106.85**	-112.42***	208.69	198.24	160.32

	(32.85)	(32.51)	(33.06)	(169.85)	(177.90)	(191.16)
Robust	-101.70**	-106.85**	-112.42**	208.69	198.24	160.32
	(36.27)	(35.75)	(36.56)	(196.16)	(206.60)	(217.84)
Observations	3,956	3,956	3,956	960	960	960
Kernel	Triangular	Epanechnikov	Uniform	Triangular	Epanechnikov	Uniform
Bandwidth	10272	10034	7892	8484	8121	6801
Panel B: Confidence level greater than 50 (error within 1km)						
Conventional	-95.84*	-103.57**	-115.07**	258.80	263.73	225.86
	(40.31)	(37.86)	(36.39)	(190.12)	(199.23)	(204.51)
Bias-corrected	-111.41**	-118.44**	-130.51***	302.80	306.04	270.35
	(40.31)	(37.86)	(36.39)	(190.12)	(199.23)	(204.51)
Robust	-111.41*	-118.44**	-130.51**	302.80	306.04	270.35
	(46.50)	(41.90)	(39.72)	(223.93)	(236.21)	(242.88)
Observations	3,178	3,178	3,178	786	786	786
Kernel	Triangular	Epanechnikov	Uniform	Triangular	Epanechnikov	Uniform
Bandwidth	8661	9229	9482	7570	7251	6439

Note: Each cell represents a separate regression discontinuity regression. Data are from ASIF. Confidence level is the output parameter from Baidu Map API, which indicates the error between street address and output coordinates. The running variable is the distance between a firm and a monitoring station. Positive (negative) distance means firms are located upstream (downstream). Negative coefficients indicate that upstream firms have lower ETC. Discontinuities at monitoring stations are estimated using methods proposed by Calonico et al. (2014) and MSE optimal bandwidth proposed by Calonico et al. (2014) for different kernel weighting methods. Standard errors clustered at the monitoring station level are reported below the estimates. * significant at 10% ** significant at 5% *** significant at 1%.

Table A18. Corruption Gap for Private Enterprises After Removing Unbalanced Industries

	Polluting Industries			Non-Polluting Industries		
Method	(1)	(2)	(3)	(4)	(5)	(6)
Conventional	-93.11**	-95.93**	-107.34**	-14.75	-15.22	-7.37
	(33.75)	(32.83)	(35.41)	(23.69)	(24.68)	(27.69)
Bias-corrected	-106.41**	-109.50***	-124.71***	-19.10	-20.36	-14.69
	(33.75)	(32.83)	(35.41)	(23.69)	(24.68)	(27.69)
Robust	-106.41**	-109.50**	-124.71**	-19.10	-20.36	-14.69
	(37.09)	(35.91)	(39.16)	(26.72)	(27.65)	(30.13)
Observations	3,502	3,502	3,502	8,449	8,449	8,449
Kernel	Triangular	Epanechnikov	Uniform	Triangular	Epanechnikov	Uniform
Bandwidth	11100	10970	7834	9987	9499	8174

Note: Each cell represents a separate regression. Data are from ASIF after removing two unbalanced industries in Table A2. The running variable is the distance between a firm and a monitoring station. Positive (negative) distance means firms are located upstream (downstream). Negative coefficients indicate that upstream firms have lower ETC. Discontinuities at monitoring stations are estimated using methods proposed by Calonico et al. (2014) and MSE optimal bandwidth proposed by Calonico et al. (2014) for different kernel weighting methods. Standard errors clustered at the monitoring station level are reported below the estimates. * significant at 10% ** significant at 5% *** significant at 1%.

Table A19. Covariate Balance of Data Cleaning Process

	Mean		Mean Difference
	After Data Cleaning	Before Data Cleaning	
	(1)	(2)	(3)
Panel A: ASIF			
Year of Opening	1995.21	1995.06	0.150***
	(10.83)	(11.00)	(0.029)
Polluting industries	0.33	0.33	0.000
(1=Yes, 0=Others)	(0.47)	(0.47)	(0.001)
Private Enterprise	0.87	0.86	0.010***
(1=Yes, 0=Others)	(0.33)	(0.34)	(0.001)
Profit	4,106.95	4,053.04	53.912
(1,000 RMB)	(160,262.77)	(158,948.18)	(429.102)
Value-Added Tax	2,463.55	2,435.87	27.684
(1,000 RMB)	(41,047.29)	(40,781.59)	(109.998)
# of Employees	244.14	240.63	3.510
(Male)	(1,097.35)	(1,100.13)	(2.954)
# of Employees	104.75	103.30	1.446
(Female)	(389.85)	(402.35)	(1.065)
Capital Stock	21,060.69	20,799.08	261.606
(1,000 RMB)	(467,307.80)	(463,420.54)	(1,251.140)
Intermediate Input	54,942.99	54,232.17	710.825
(1,000 RMB)	(527,172.64)	(523,939.60)	(1412.949)

Note: Columns (1)–(2) report the means and standard deviations of firm characteristics. In column (3), we conduct t-test about the sample after data cleaning process and before data cleaning process. Standard errors are reported in the parentheses. *** significant at 0.1%.