Optimal Emission Regulation under Market Uncertainty

Guokai Li, Pin Gao, Zizhuo Wang
School of Data Science, The Chinese University of Hong Kong, Shenzhen (CUHK-Shenzhen), Guangdong 518116, China
guokaili@link.cuhk.edu.cn, {gaopin, wangzizhuo}@cuhk.edu.cn

Government regulations on emission control can be broadly divided into two categories: price instruments and quantity instruments. In this paper, we develop a stylized model to compare the two instruments in the presence of market uncertainty. We find that when the emission intensity (i.e., emissions per unit of production) and the market uncertainty are both high or low, the expected social welfare under the price instruments will be higher; otherwise, the performance of the quantity instruments is comparatively better. The results are robust when incorporating firm competition and national/regional pollution damage. Afterward, we demonstrate that the government's quick-response capability or a hybrid of the price and quantity instruments can improve the expected social welfare, especially for high-emitting industries when the market uncertainty is intermediate. Lastly, for heterogeneous firms, we find that allowing permit trading in the quantity instrument may not be beneficial when pollution from each firm is more likely to have regional effects.

Key words: environmental regulation, sustainable operations, uncertainty, emission tax, emission cap

1. Introduction

With the highest temperature record being broken year by year, the adverse effects of the climate crisis continue to accumulate and intensify. Examples include the disappearance of island countries (Farbotko and Lazrus 2012), the melting of Greenland's ice sheet (Keegan et al. 2014), and the tragic wildfire in Australia.[1] In order to address such environmental issues, governments around the world have been cooperating with each other to reduce the global emission of greenhouse gases such as methane and carbon dioxide. One significant milestone in these efforts is the well-known Paris Agreement.[2] As a legally binding international treaty, the agreement was signed by 196 parties in December 2015, which sets a goal of limiting global warming to a range well below 2 degrees Celsius.

To achieve the goal set forth in the Paris Agreement, each signatory needs to impose regulations domestically to limit greenhouse gas emissions in each respective country. There are many types of regulations for reducing greenhouse gas emissions. Among them, the mainstream regulations can be divided into two kinds: price instruments and quantity instruments. In particular, under the price instrument, a regulating authority sets a tax rate, and each regulated firm has the obligation to pay the emission tax for every tonne of gas emitted. Under the quantity instrument, a regulating authority imposes a cap (or some permits) for each regulated firm, while the latter's net emission amount cannot exceed its holding amount of permits. According to the World Bank,[3] as of April 2023, there are 27 national jurisdictions that have implemented a tax system, and those countries account for about 6% of global greenhouse gas emission. In the meantime, the quantity instruments are adopted by 12 national jurisdictions, which account for about 18% of global greenhouse gas emissions. Discussion on which of the two instruments is better is still in full swing (Tang et al. 2019). As pointed out by Montgomery (1972) and Anand and Giraud-Carrier (2020), in absence of market uncertainty, the two instruments can essentially achieve the same performance. However, in the presence of market uncertainty, the equivalence does not hold in general. In this paper, we aim to shed light on the trade-off between the two instruments when the market uncertainty factors in.

To answer the above question, in the base model, we consider an ecosystem with one regulator and one monopolistic firm. The firm produces one type of product with a linear inverse demand function and its production process generates pollution emissions that can be mitigated through its abatement efforts. The consequent pollution damage is convex in the net emission amount, defined as the emission amount minus the abatement amount. The market size is random, which is either a high demand or a low demand. Before the market size is realized, in order to maximize the expected social welfare (i.e., the firm's profit plus consumer surplus minus pollution damage), the regulator needs to announce and commit to a regulatory instrument, including the choice of instruments and its specification, i.e., either a price instrument with a specified tax rate or a quantity instrument with a specified cap. Subject to the regulatory policy, the firm decides its production and abatement amounts to maximize its own profit upon observing the realization of the market size. We are interested in finding out under what circumstance, one instrument will lead to a higher expected social welfare than the other.

We find that if the emission intensity (i.e., emissions per unit of production) is smaller than a threshold, then neither regulation should be implemented. When the emission intensity is above the threshold, a regulation is beneficial to social welfare. However, which type of regulation is better depends on both the size of market uncertainty and the emission intensity. Specifically, when the market uncertainty is low, the performance of the price instrument is better (worse, resp.) if the emission intensity is relatively small (large, resp.). However, the ranking is reversed when the market uncertainty is large. That is, the performance of the price instrument is worse (better, resp.) if the emission intensity is relatively small (large, resp.).

To illustrate the robustness of our results, we further incorporate various practical factors into our base model, including firm competition and regional/national pollution damage. We show that the structure of the comparison results in the base model still qualitatively holds, but the thresholds will be influenced by the incorporated factors. We further investigate the importance of the government's quick-response capability, and explore a hybrid instrument that combines the advantages of both the price and the quantity instruments. Not considering the additional administrative cost, we show that when the emission intensity is large and the market uncertainty is intermediate, the quick-response capability and the hybrid instrument both can significantly improve social welfare, compared to the pure instruments in the base model. This result has three implications. First, it emphasizes the importance of government's quick-response capability. Second, it supports regulatory agencies in implementing price regulations on the cap-and-trade market. Third, for jurisdictions that use emission tax tools, providing firms with some tax-free emission allowances can be beneficial to the entire society. Lastly, we investigate firms with different abatement coefficients and damage coefficients. Our results suggest that permit trading among heterogeneous firms will be detrimental when the pollution from each firm is more likely to have regional impacts.

1.1. Literature Review

Our paper is broadly related to two streams of literature: environmental economics and sustainable operations.

Environmental Economics. In the face of a gradually deteriorating environment, regulatory instruments attracted increasing attention from governments. Two main forms of environmental regulations are the price instruments and the quantity instruments. The price instrument was first proposed by Pigou (1920), of which the basic idea is to price the pollution and to charge firms for their emission. Following this idea, researchers explore the price instrument in various scenarios to test its robustness. For example, Barnett (1980) studies the imperfect competition among polluters and shows that the equilibrium tax rate may be below the first-best level. Bovenberg and De Mooij (1994) find that the pollution tax will exacerbate the distortion induced by the preexisting labor income tax. The quantity instrument was discussed in Dales (1968), of which the idea is to set a global emission reduction target and utilize the market to allocate quotas efficiently. Various follow-up works investigate this instrument by incorporating various practical factors. For example, Montgomery (1972) shows that the unique cost-efficient equilibrium can be achieved regardless of the initial allocation of emission permits. Nagurney and Dhanda (2000) study the impact of transition costs in an oligopolistic market for the quantity instrument. Carmona and Hinz (2011) study how to price options related to emission permits in the presence of the emission permit market.

In addition to the analysis of each instrument alone, the debate on which instrument is better has a long history. Without market uncertainty, Montgomery (1972) proves the equivalence of these two instruments. That is, the regulator can achieve efficient emission amounts under the quantity instrument by imposing a tax rate equal to the corresponding shadow price of permits. Afterward, researchers test the equivalence in other deterministic situations. For example, Requate (1993) considers a model with multiple local monopolies and finds that the two instruments are equivalent if only aggregate pollution matters. Anand and Giraud-Carrier (2020) show that under Cournot competition, the two instruments can still get the same equilibrium by appropriately tuning the specifications. However, in the seminal work "Prices *vs.* Quantities" by Weitzman (1974), the author finds that the equivalence may break down if market uncertainty exists. Specifically, in terms of social welfare, the price instrument is better if and only if the emission intensity is small. Following this framework, Laffont (1977) considers more general randomness and includes a new timing structure where the target price (or tax) is first announced to consumers, and the cap value is determined by consumers; Ireland (1977) analyzes the case when prices are contingent on the realized information; Chen (1990) analyzes the case when the firm has more information on its cost and more market power to affect the market price.

In this work, we consider a setting similar to that in Weitzman (1974), however, with two key differences: First, negative production allowed in Weitzman's model is prohibited in our model. Second, partially exercising the cap prohibited in Weitzman's model is allowed in ours. As a result, we demonstrate that, in addition to the emission intensity, the comparison between the price and the quantity instruments also depends on the market uncertainty. It is worth noting that Subramanian et al. (2007) and Anand and Giraud-Carrier (2020) also use similar assumptions as ours, but the focuses of the two works are on deterministic situations. As for the uncertain case, Fu et al. (2021) focus on the value of the intertemporal permit transfer under the quantity instrument when firms can partially exercise the emission permits.

Sustainable Operations. The study of sustainable operations management is a growing area in the operations management society. For a comprehensive review of recent developments in this area, see Corbett and Klassen (2006); Sunar (2016) and Atasu et al. (2020). Our paper is mainly related to studies on firms' strategic decisions under different environmental regulations. Empirical studies, including Klassen and McLaughlin (1996) and King and Lenox (2002), show that better environmental performance can make firms more attractive to shareholders and induce better financial performance accordingly. Drake and Just (2016) study firms' compliance responses to environmental regulations and provide realistic examples for each case. Sunar and Plambeck (2016) study firms' emission allocation plan and the impact of the price instrument when co-products exist. Sunar and Birge (2019) study the strategic behaviors of both renewable and conventional firms in a day-ahead electricity market with an undersupply penalty regulation. Gopalakrishnan et al. (2021) propose a footprint-balanced scheme based on the Shapley value for allocating the emission responsibilities along the supply chain. Sunar and Swaminathan (2021) analyze the impact of distributed renewable energy on energy firms' profit and social welfare under the net metering regulation.

Environmental regulations also influence firms' behavior on other dimensions, including the adoption of green technologies and product design. As Porter (1991) argues, environmental regulation does not conflict with economic competitiveness because encouraging technology innovation will not only reduce pollution but also lower production costs. Krass et al. (2013) find that the firm's response to the tax is non-monotone: the increase of the tax rate will first motivate the firm to switch to a greener technology but will result in a reverse switch afterward. Gong and Zhou (2013) study the firm's optimal production planning policy under the quantity instrument when it has both green and regular technologies. Drake (2018) finds that under the carbon tariff and the border adjustment, the carbon leakage may prompt foreign firms to adopt clean technology, resulting in decreased global emissions. Peng et al. (2021) examine the impact of the government's regulation on firms' green claims in the presence of consumers who bring their own reusable containers. Regarding competition, Subramanian et al. (2007) study both the monopoly and the Cournot competition cases under the quantity instrument and find that excessive permits can be detrimental. Drake et al. (2016) study the impacts of the price and the quantity instruments on firms' profit and investment in dirty and clean capacities. Anand and Giraud-Carrier (2020) find that there always exists a range of moderate cap-and-trade regulation such that the resulting industry profit is greater than that of laissez-faire even when the abatement cost is concerned. Compared with these works, we also incorporate firm competition and the adoption of emission reduction technologies, but our focus is on the impact of market randomness on the regulator's choice of environmental regulatory tools.

1.2. Organization

In Section 2, the base model is introduced. In Section 3, we present our main results. In Section 4, we incorporate various practical factors into the base model: In Section 4.1, we incorporate firm competition and national/regional environmental damage; In Section 4.2, we explore the government's quick-response capability and the hybrid instrument; In Section 4.3, we examine the value of the trading market in the cap-and-trade scheme. The paper is concluded in Section 5. The proofs of all results are relegated to Appendix D.

2. Base Model

In the base model, we consider a market with one regulator and one monopolistic firm, where the latter produces a single type of product and can make pollution abatement. Only knowing the prior distribution of market demand, the regulator's decision is to issue a regulatory instrument to influence the firm's production behavior. The objective of the regulator is to maximize the expected social welfare, considering the firm's profit, the consumer surplus and the pollution damage caused by the production and the abatement effort. In Section 4.1, we extend the analysis to the cases with firm competition.

The Consumer. Adopting the well-established model in Singh and Vives (1984), we study a representative consumer with the consumption utility function as follows:

$$U(x) = \theta x - \frac{1}{2}x^2,$$

where θ is a random variable that measures the size of the market and reflects the consumer's preference. Given the total production in the market as x, we assume that the equilibrium market price $p(\theta, x)$ satisfies the following formula,

$$x = \underset{z:0\leq z\leq x}{\arg\max}\{U(z) - p(\theta, x)\cdot z\}.$$

In other words, under the market price $p(\theta, x)$, to maximize the consumption utility minus the payment, the consumer's optimal order quantity coincides with that in the market. After some algebra, we have the following linear inverse demand function,

$$p(\theta, x) := \theta - x.$$

Subsequently, given the production quantity x, consumer surplus can be computed as $U(x) - p(\theta, x)x = x^2/2$.

The Firm. The firm can produce the product at a negligible cost. Production is always associated with harmful gas emissions, such as carbon dioxide, etc. To avoid severe penalties for excessive emissions, the firm may innovate (e.g., new production process and cleaner inputs) with an effort (see, e.g., Anand and Giraud-Carrier 2020). Following the classical literature in environmental economics (see, e.g., Hartman et al. 1997, Subramanian et al. 2007), the emission abatement cost is assumed to be quadratic in the abatement amount to reflect the diminishing return of the abatement investment. That is,

$$C(w) := \frac{\gamma}{2}w^2,$$

where $\gamma \in \mathbb{R}^+$ is the abatement cost coefficient and w denotes the production quantity at which emissions are entirely mitigated. Given the market size θ, the firm's production profit, when choosing production quantity x and abatement amount w, will be

$$R(\theta, x, w) := p(\theta, x)\cdot x - C(w) = \theta x - x^2 - \frac{\gamma}{2}w^2.$$

For analytical tractability, we assume that θ follows a two-point distribution on $\{\mu - \sigma, \mu + \sigma\}$ with equal probabilities, where $\mu \in \mathbb{R}^+$ can be regarded as the average market size and $\sigma \in [0, \mu)$ measures the uncertainty. The binary state can be interpreted as that the market size will be affected by some unknown event with two likely outcomes, or the economic market is either prosperous or depressed. For example, due to the epidemic, the global market is affected by much more unpredictable factors, resulting in a sharp increase in market uncertainty.[4] For ease of exposition, in the subsequent discussions, we will call the market state high-demand and low-demand if $\theta = \mu + \sigma$ and $\theta = \mu - \sigma$, respectively. Moreover, for analytical tractability, we assume the marginal production cost to be zero. We note that the main results still qualitatively hold if the firm has an additional production cost (e.g., $\propto x^a$ with $a \geq 0$), and when the distribution of θ is generalized to some continuous functions (e.g., a superposition of two distinct normal distributions).

The Regulator. Different from the profit-maximizing firm, the regulator is also concerned with the consumer surplus and the pollution damage caused by the firm's production. Following the classical literature in environmental economics (see, e.g., Weitzman 1974, Tietenberg and Lewis 2011, Nordhaus 2019), we assume that the pollution damage is quadratic in the net pollution quantity $(x - w)$, namely

$$D(x, w) := d \cdot [(x - w)^+]^2,$$

where $(\cdot)^+ = \max\{\cdot, 0\}$. The level of the pollution damage coefficient d can be attributed to several factors. First, industries such as cement and chemicals are generally classified as high-emitting industries, but others such as glass and paper are considered as low-emitting industries.[5] Then, with the same production volume, the emission intensity levels of different industries will lead to different values of d. Second, different emission sources, like CO_2, NO_X and SO_2, will also cause different pollution severity (see, e.g., Muller and Mendelsohn 2009). Third, as commented in Xue et al. (2021), regional pollution will lead to losses of talented and skilled workers, thereby reducing the economic performance. Then, d can be understood as the level of the social cost of the regulator. For consistency, throughout the paper, we will mainly interpret d as the *emission intensity*, and hence call it the emission intensity coefficient.

Though the regulator can anticipate the distribution of market size, prior to the realization of uncertainty, it has to choose one regulatory instrument. We note that the assumption that the regulator's decision needs to be made ex-ante is quite reasonable. Since deliberating and enacting a regulatory instrument is often a lengthy process, government agencies need to start the process early and hence can only use the limited information available at that time. In addition, after the announcement of regulatory instruments, it is not easy for governments to make further changes until the announced policy expires. For example, as required by the reformed Climate Change Response Act 2002, by the end of September of every year, the Ministry of Environment in New Zealand needs to set the overall emission cap, which takes effect several months later.[6] Moreover, in most existing regulatory instruments, for convenience of implementation, governments normally commit to a simple regulation (a constant tax rate or cap) rather than an adaptive regulation depending on the realized market information.[7,8] Most research papers, including Yu et al. (2018), Arifoğlu and Tang (2022) and Fan et al. (2023), have also adopted similar setups. Since there is a one-to-one mapping between emissions and net pollution quantity, we will simply consider direct regulations of the net pollution quantity, i.e., production amount net abatement amount (see, e.g., Krass et al. 2013, Anand and Giraud-Carrier 2020).

Particularly, in our base model, the regulator ex-ante chooses one of the following two types of instruments:

(i) *Price Instrument.* Under this instrument, the regulator imposes an emission tax for each unit of pollution by the firm. Given a tax rate $t \geq 0$, the firm has to pay emission tax $t \cdot (x-w)^+$ if its production amount is x and abatement amount is w.

(ii) *Quantity Instrument.* Under this instrument, the regulator sets a cap for the net pollution quantity of the firm. Specifically, when the regulator chooses a production cap $k \geq 0$, the net pollution quantity of the firm $(x-w)^+$ is restricted to be no more than k.

We note that due to the simplicity of implementation, the adoption of one of the two pure instruments above is commonly observed in reality.[9] We extend the analysis to a hybrid instrument in Section 4.2.

After choosing a regulatory instrument and its specification (the tax rate under the price instrument or the cap under the quantity instrument), the market size uncertainty unfolds and the firm makes its production and abatement decisions based on the imposed regulatory instrument and the realized market size. The objective of the regulator is to maximize the expected social welfare:

$$\mathbb{E}_{\theta}^{\pi(\alpha)}\Bigg[\underbrace{(\theta-x)x-\frac{\gamma}{2}w^2}_{\text{Production Profit}} + \underbrace{\frac{x^2}{2}}_{\text{Consumer Surplus}} - \underbrace{d[(x-w)^+]^2}_{\text{Pollution Damage}}\Bigg]. \tag{1}$$

In (1), we use $\pi(\alpha)$ to denote the type of the regulatory instrument chosen by the regulator, with α indicating its specification. More precisely, $\pi(\alpha)=p(t)$ means that the regulator chooses a price instrument with tax rate t, while $\pi(\alpha)=q(k)$ means that the regulator chooses a quantity instrument with cap k. We also note that (x,w) in (1) is the firm's optimal decisions in response to $\pi(\alpha)$ as well as the realization of θ. The timing of events in our model is summarized as follows:

a) Based on the prior distribution of market size θ, the regulator ex-ante commits to a regulatory instrument, namely $\pi(\alpha)$;
b) The value of market size θ is realized and observed by the firm;
c) The firm chooses its production quantity x and abatement quantity w;
d) The firm's production profit, the pollution damage and the social welfare are calculated.

In the next section, we will compare different regulatory instruments in terms of expected social welfare. In particular, we will study the conditions under which one regulatory instrument will lead to a greater expected social welfare than the other.

3. Main Results

In this section, we first analyze the price instrument and the quantity instrument respectively. After that, we compare the two instruments. The superscripts 'P' and 'Q' are provided as mnemonics for 'Price' and 'Quantity', respectively.

3.1. Price Instrument

Under the price instrument, the regulator sets a tax rate t for each unit of pollution. Given a tax rate t, the firm chooses the production amount x and abatement amount w to maximize its profit:

$$\max_{x,w:0\le w\le x}\left\{(\theta-x)x-\frac{\gamma}{2}w^2-t(x-w)^+\right\}. \tag{2}$$

From (2), we can deduce that the firm's optimal production amount and abatement amount are

$$x^P(\theta,t)=\begin{cases}\frac{\theta-t}{2} & t\in[0,\frac{\gamma\theta}{\gamma+2})\\ \frac{\theta}{\gamma+2} & t\in[\frac{\gamma\theta}{\gamma+2},\infty),\end{cases}\qquad w^P(\theta,t)=\begin{cases}\frac{t}{\gamma}, & t\in[0,\frac{\gamma\theta}{\gamma+2})\\ \frac{\theta}{\gamma+2}, & t\in[\frac{\gamma\theta}{\gamma+2},\infty).\end{cases}$$

The above result indicates that when the tax rate is relatively high, the firm will make net-zero-emission production ($x^P(\theta,t)=w^P(\theta,t)$), i.e., the abatement fully offsets the emission (see Fankhauser et al. 2022). Then the regulator determines the optimal tax rate t by solving the following problem:

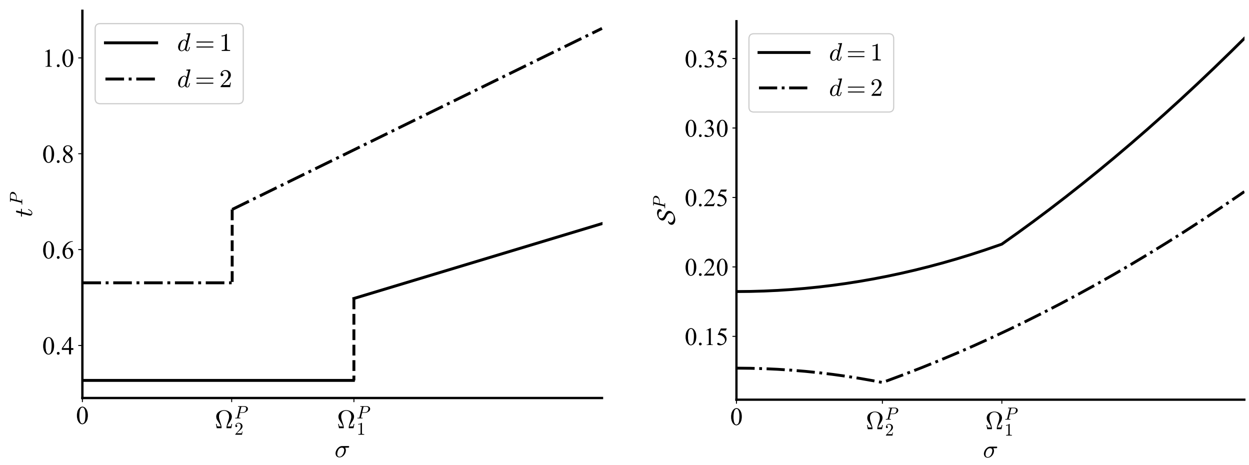
$$\mathcal{S}^P=\max_t \mathbb{E}_\theta^{p(t)}\left[\left(\theta-x^P(\theta,t)\right)\cdot x^P(\theta,t)-\frac{\gamma}{2}\left(w^P(\theta,t)\right)^2+\frac{1}{2}\left(x^P(\theta,t)\right)^2-d\left[(x^P(\theta,t)-w^P(\theta,t))^+\right]^2\right]. \tag{3}$$

The result is summarized in the following proposition.

Proposition 1 **(Optimal Price Instrument)**. *Define* $\delta^P=\frac{\gamma}{2(\gamma+2)}$ *and* $\Omega^P=\frac{2\gamma(\gamma+3)}{2\sqrt{2}d(\gamma+2)^2+(6-2\sqrt{2})\gamma+(2-\sqrt{2})\gamma^2}$. *The optimal price instruments are as follows.*

- *When $d\le\delta^P$, the optimal tax rate and the corresponding expected social welfare are $t^P=0$ and $\mathcal{S}^P=\frac{3-2d}{8}(\mu^2+\sigma^2)$.*
- *When $d>\delta^P$ and $\sigma/\mu\le\Omega^P$, the optimal tax rate and the corresponding expected social welfare are $t^P=\mathcal{C}_1^P\mu$ and $\mathcal{S}^P=\mathcal{C}_2^P\mu^2+\mathcal{C}_3^P\sigma^2$. As σ increases, t^P remains constant, and $\mathcal{S}^P$ increases (decreases, resp.) when $d<3/2$ ($d>3/2$, resp.).*
- *When $d>\delta^P$ and $\sigma/\mu>\Omega^P$, the optimal tax rate and the corresponding expected social welfare are $t^P=\mathcal{C}_1^P(\mu+\sigma)$ and $\mathcal{S}^P=\mathcal{C}_4^P(\mu^2+\sigma^2)+\mathcal{C}_5^P\mu\sigma$. Both t^P and $\mathcal{S}^P$ increase in σ.*

Here, $\mathcal{C}_1^P$, $\mathcal{C}_2^P$, $\mathcal{C}_3^P$, $\mathcal{C}_4^P$ and $\mathcal{C}_5^P$ are constants depending only on d and γ, and their detailed forms are provided in the proof.



(a) Optimal tax rate

(b) Expected social welfare

Figure 1 Optimal tax rate and expected social welfare, where $\mu=1$, $\gamma=10$ and $d=1$ or 2.

Figure 1 illustrates Proposition 1. According to the proposition, when the emission intensity is sufficiently small ($d \le \delta^P$), the regulator should not impose an emission tax (i.e., $t^P=0$). The reason is that in this case, consumer surplus outweighs pollution damage, and the abatement cost is relatively high (compared to the pollution damage). Therefore, the firm's production is always beneficial to social welfare and the abatement effort is not worthwhile to be made, suggesting that the regulator should not impose an emission tax to restrict the firm's production or urge more abatement efforts.

When the emission intensity is relatively large ($d > \delta^P$), pollution damage outweighs consumer surplus and abatement becomes cost-efficient, suggesting that the emission tax should be imposed. According to Proposition 1, the optimal solution to (3) has two phases. In the first phase, the market uncertainty is low ($\sigma/\mu < \Omega^P$), indicating that the market size under the low-demand case is still sizable. As a result, there will still be a significant contribution to social welfare from the low-demand case, even with consideration of the pollution damage. Therefore, the optimal tax rate is one such that it is still profitable for the firm to keep a positive net pollution quantity under the low-demand case. We call this phase the "inclusive" phase. (For ease of memory, we note that whether it is inclusive or not indicates whether the regulator cares about both demand cases or only the high-demand case when determining regulations.) In this phase, the optimal tax rate t^P depends on the expected market size μ, but not on the uncertainty of the market size σ. However, the optimal social welfare depends on both the expected market size and the uncertainty. Particularly, when the emission intensity is small ($d < 1.5$), the expected social welfare increases in σ because the increment of production profit and consumer surplus outweighs the additional pollution damage. In contrast, with large emission intensity ($d > 1.5$), the expected social welfare decreases in σ as increased pollution damage outweighs the increment of other components.

In the second phase, the market uncertainty is high ($\sigma/\mu \geq \Omega^P$), and the market size under the low-demand case is minuscule. Then, the optimal tax rate is one such that under the low-demand case, the firm will choose net-zero-emission production and its production decisions are independent of the tax rate, leaving the high-demand market the only case under which the firm's decisions depend on the tax rate. We call this phase the "exclusive" phase. In this phase, as σ further increases, the regulator will increase the tax rate to moderate the net pollution amount under the high-demand case. Therefore, the tax rate increases in σ. Moreover, with a proper tax rate, a higher market size is always more beneficial to the entire social welfare. Therefore, the expected social welfare (which is mainly contributed by the high-demand case) increases in the market uncertainty σ.

Lastly, we discuss the two crucial thresholds (δ^P and Ω^P) introduced in Proposition 1. The former specifies the minimum emission intensity d at which the price instrument should be imposed. The latter specifies the relative market size σ/μ at which the price instrument switches from the inclusive phase to the exclusive phase. Both thresholds decrease in the firm's abatement capability $1/\gamma$. The reason is as follows. As the firm's abatement ability increases, the regulator will naturally prompt the firm to adopt more abatement by narrowing the no-regulation region and the inclusive phase (which implements a relatively small tax rate). Moreover, the threshold Ω^P decreases in the emission intensity coefficient d. This is intuitively reasonable. When the emission intensity becomes larger, the region where the low-demand case can still contribute significant social welfare will be smaller. Therefore, the region where the regulator only focuses on the high-demand case (i.e., the exclusive phase) will expand, resulting in the earlier switch to the exclusive phase when σ increases.

3.2. Quantity Instrument

Under the quantity instrument, the regulator sets a cap k for the firm's production amount. Given a cap k, the firm chooses the production amount x and the abatement amount w to maximize its profit:

$$\max_{x,w:0\leq x-w\leq k; w\geq 0} \left\{(\theta - x)x - \frac{\gamma}{2}w^2\right\}. \tag{4}$$

From (4), we can deduce that the firm's optimal production amount and abatement amount are

$$x^Q(\theta,k) = \begin{cases} \frac{\theta+k\gamma}{\gamma+2} & k\in[0,\frac{\theta}{2}) \\ \frac{\theta}{2} & k\in[\frac{\theta}{2},\infty), \end{cases} \qquad w^Q(\theta,k) = \begin{cases} \frac{\theta-2k}{\gamma+2} & k\in[0,\frac{\theta}{2}) \\ 0 & k\in[\frac{\theta}{2},\infty). \end{cases}$$

The above result indicates that when the given cap is relatively high, the firm will partially exercise the given cap ($x^Q(\theta,k) - w^Q(\theta,k) < k$). For example, in Europe ETS, firms could partially exercise the caps during recession[10] or pandemics.[11] Then the regulator determines the optimal cap k by solving the following problem:

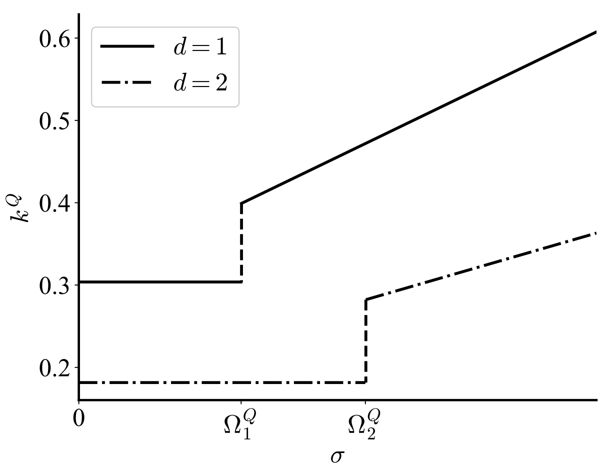
$$\mathcal{S}^Q = \max_{k\geq 0} \mathbb{E}_\theta^{q(k)} \left[\big(\theta - x^Q(\theta,k)\big)\cdot x^Q(\theta,k) - \frac{\gamma}{2}\big(w^Q(\theta,k)\big)^2 + \frac{1}{2}\big(x^Q(\theta,k)\big)^2 - d\big[(x^Q(\theta,k) - w^Q(\theta,k))^+\big]^2 \right]. \tag{5}$$

The result is summarized in the following proposition.

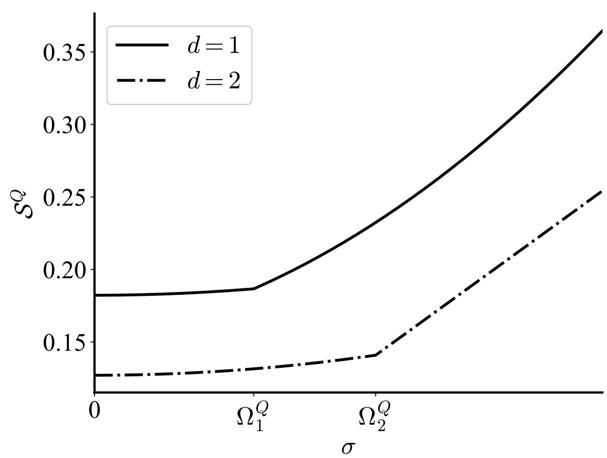
Proposition 2 **(Optimal Quantity Instrument)**. *Define* $\delta^Q = \frac{\gamma}{2(\gamma+2)}$ *and* $\Omega^Q = \frac{(\gamma+2)(4d-\gamma+2d\gamma)}{2d(\gamma+2)^2+\gamma(6\sqrt{2}-2+(2\sqrt{2}-1)\gamma)}$ *The optimal quantity instruments are as follows.*

- *When* $d \leq \delta^Q$*, the optimal cap and the corresponding expected social welfare are* $k^Q \geq \frac{\mu+\sigma}{2}$ *and* $\mathcal{S}^Q = \frac{3-2d}{8}(\mu^2+\sigma^2)$.
- *When* $d > \delta^Q$ *and* $\sigma/\mu \leq \Omega^Q$*, the optimal cap and the corresponding expected social welfare are* $k^Q = \mathcal{C}_1^Q\mu$ *and* $\mathcal{S}^Q = \mathcal{C}_2^Q\mu^2 + \mathcal{C}_3^Q\sigma^2$*. As* σ *increases,* k^Q *remains constant and* $\mathcal{S}^Q$ *increases.*
- *When* $d > \delta^Q$ *and* $\sigma/\mu > \Omega^Q$*, the optimal cap and the corresponding expected social welfare are* $k^Q = \mathcal{C}_1^Q(\mu+\sigma)$ *and* $\mathcal{S}^Q = \mathcal{C}_4^Q(\mu^2+\sigma^2) + \mathcal{C}_5^Q\mu\sigma$*. Both* k^Q *and* $\mathcal{S}^Q$ *increase in* σ.

Here, $\mathcal{C}_1^Q$, $\mathcal{C}_2^Q$, $\mathcal{C}_3^Q$, $\mathcal{C}_4^Q$ *and* $\mathcal{C}_5^Q$ *are constants depending only on* d *and* γ*, and their detailed forms are provided in the proof.*



(a) Optimal cap



(b) Expected social welfare

Figure 2 Optimal cap and expected social welfare, where $\mu=1$, $\gamma=10$ and $d=1$ or 2.

Figure 2 illustrates Proposition 2. According to the proposition, similar to the price instrument, when the emission intensity is sufficiently small ($d<\delta^Q$), the regulator should not restrict the firm's production or urge more abatement (i.e., $k^Q \geq \frac{\mu+\sigma}{2}$). When the emission intensity is relatively large ($d>\delta^Q$), pollution damage becomes significantly serious, and hence the regulator should impose a cap. Similar to the price instrument, there are also two phases here. In the first phase, i.e., the inclusive phase, the market uncertainty is low. As a result, even in the low-demand case, the pollution damage is sizable. Therefore, the regulator will limit the production quantity even in the low-demand case. In particular, under the optimal quantity instrument, the firm's net pollution amount is the same as the cap under either market size, such that the optimal cap depends on the expected market size. Since the pollution damage is fixed, the expected social welfare increases in the market uncertainty σ due to the convexity of the production profit and consumer surplus.

In the second phase, i.e., the exclusive phase, the market uncertainty is high, and the market size under the low-demand case is sufficiently small, under which the net pollution amount determined by the firm will naturally be below the optimal cap set by the regulator. Thus, when determining the cap, the regulator only focuses on the high-demand case. Similar to the price instrument, in the exclusive case, as σ increases, the market size under the high-demand case increases, and so will the optimal cap and the corresponding expected social welfare. Lastly, similar to the price instrument, the threshold $\delta^Q = \delta^P$ guides whether the quantity instrument should be imposed and Ω^Q guides the switch of these two phases. As the firm's abatement ability $1/\gamma$ is enhanced, the regulator shall consider a smaller δ^Q and a larger Ω^Q to narrow the no-regulation region and the exclusive phase (which implements a larger cap). Moreover, as the emission intensity increases, the region without a restriction on the low-demand case will shrink under the optimal decision, resulting in a wider range of the inclusive phase (i.e., larger Ω^Q).

3.3. Comparison between the Price and Quantity Instruments

Now we compare the two instruments and investigate which yields a higher expected social welfare. Our main result is summarized in the following theorem.

Theorem 1 **(Instrument Comparison)**. *Define* $\delta = \frac{\gamma}{2(\gamma+2)}$, $\Delta = \frac{\gamma(3\gamma+8)}{2(\gamma+2)^2}$ *and* $\Omega = \sqrt{2}-1$. *The comparison result is as follows.*

- *When* $d \in [0,\delta]$, *neither instruments should be implemented.*
- *When* $d \in (\delta,\Delta)$, *the expected social welfare under the optimal price instrument is higher (lower, resp.) than that under the optimal quantity instrument when* $\sigma/\mu < \Omega$ ($\sigma/\mu > \Omega$, *resp.).*
- *When* $d \in (\Delta,\infty)$, *the expected social welfare under the optimal quantity instrument is higher (lower, resp.) than that under the optimal price instrument when* $\sigma/\mu < \Omega$ ($\sigma/\mu > \Omega$, *resp.).*

Figure 3 illustrates Theorem 1. According to the theorem, the comparison between the two instruments depends on both the emission intensity coefficient d and the relative market uncertainty σ/μ. In particular, when the emission intensity is smaller than a threshold ($d < \delta$), consumer surplus outweighs pollution damage and abatement is relatively expensive, such that the firm's production is always beneficial to social welfare and abatement is not cost-efficient. Therefore, neither instrument should be implemented to restrict the firm's production or urge more abatement.

When the emission intensity is above the threshold ($d > \delta$), using a proper regulatory instrument would benefit the whole society. In particular, when the emission intensity and the relative market uncertainty are both high ($d > \Delta$ and $\sigma/\mu > \Omega$) or low ($\delta < d < \Delta$ and $\sigma/\mu < \Omega$), the optimal expected social welfare under the price instrument is higher than that under the quantity instrument; otherwise, the ranking is reversed. To interpret, we plot the expected social welfare under different instruments for different emission intensity coefficients in Figure 4 (the curves in Figure 4 are plotted based on the results in Propositions 1 and 2).

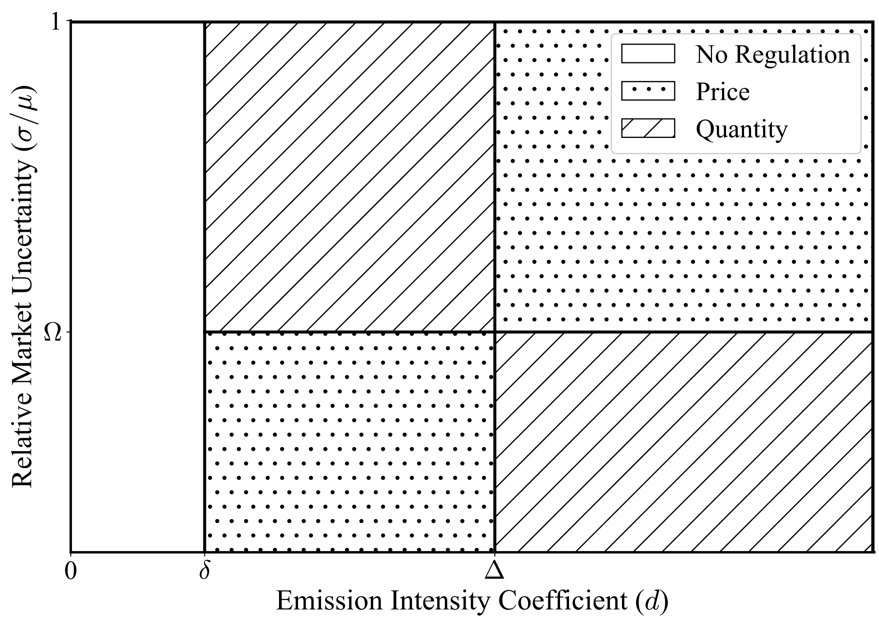


Figure 3 **The preferred regulatory instrument under different market conditions when** $\gamma = 10$.

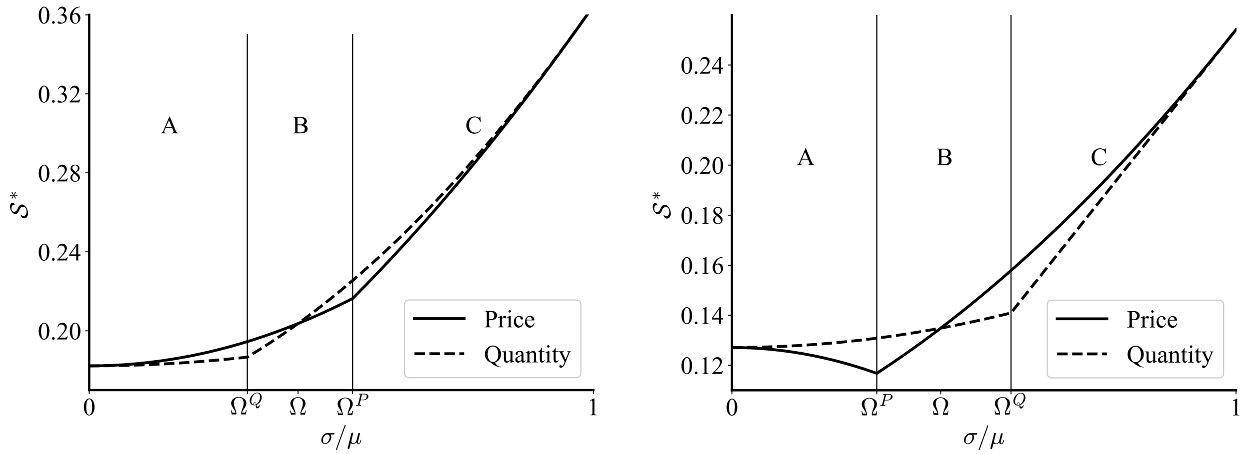
Now we provide intuitions for Theorem 1. For ease of illustration, we first introduce two forms of flexibility.

- **Emission Flexibility.** Consider the case when both instruments are in the inclusive phase. As discussed in Sections 3.1 and 3.2, under the quantity instrument, the firm's net emission amount under either market size is the same as the imposed cap; while under the price instrument, the firm's net emission amount increases in the realized market size. Therefore, the firm is provided with greater "emission flexibility" under the price instrument.
- **Abatement Flexibility.** Consider the case when both instruments are in the exclusive phase and the market is in the low-demand case. As argued in Sections 3.1 and 3.2, under the price instrument, the firm will choose an abatement amount equal to production volume to achieve the net-zero-emission target; while under the quantity instrument, it does not need to make abatement and operates without any constraints. Therefore, when facing a small market size, the firm is provided with greater "abatement flexibility" under the quantity instrument.

Either flexibility is a double-edged sword for the whole society: On the positive side, the firm can exploit the flexibility to achieve a higher production profit; on the negative side, its flexible emission behaviors may induce amplified pollution damage. In the following, we discuss in detail how these two forms of flexibility affect the comparison between the instruments.

Recall that for both instruments, there are two phases (i.e., the inclusive and the exclusive phases) in their optimal solutions. We start from the case when both instruments are in the inclusive phase, i.e., the market uncertainty is sufficiently low ($\sigma/\mu \le \min\{\Omega^P, \Omega^Q\}$). See region A in both Figures 4a and 4b for an illustration. In this case, according to the above discussion, the firm is provided with greater *emission flexibility* under the price instrument. On the one hand, such emission flexibility allows the firm to adjust its net pollution amount according to the realized market size, which can increase the firm's profit. On the other hand, due to the convexity of the pollution damage function, the expected pollution damage under the price instrument is amplified by the firm's uneven pollution amounts across different cases. In other words, the emission flexibility is a double-edged sword. Specifically, when the emission intensity is not significant ($d < \Delta$), the additional profit induced by the emission flexibility outweighs the additional pollution damage, resulting in an overall positive effect for the emission flexibility. Therefore, the expected social welfare under the optimal price instrument increases in the market uncertainty with a larger slope and hence is greater than that under the optimal quantity instrument in this phase (see region A in Figure 4a). When the emission intensity is significant ($d > \Delta$), the additional pollution damage induced by the emission flexibility outweighs the additional profit, resulting in an overall negative effect for the emission flexibility. Therefore, the expected social welfare under the optimal price instrument decreases in the market uncertainty and hence is less than that under the optimal quantity instrument in this phase (see region A in Figure 4b). Moreover, as the market uncertainty increases, the overall positive or negative effect is strengthened, and thus the absolute difference in social welfare between the two instruments always increases. Similar results are also found in Weitzman (1974), though the latter considers a different setting without abatement efforts, and assumes that a negative production amount is allowed and partially exercising the cap is prohibited. As elaborated in the following paragraph, our amendments in the assumptions lead to a *reversed* ranking of the two instruments as the market uncertainty increases.

When the market uncertainty is sufficiently high, both instruments are in the exclusive phase ($\sigma/\mu \ge \max\{\Omega^P, \Omega^Q\}$). See region C in both Figures 4a and 4b for an illustration. In this case, as argued in Sections 3.1 and 3.2, since the firm's decisions in the low-demand case are independent of regulation, the regulator is only concerned about the high-demand case when determining the specification of each instrument. Thus, the regulation for the high-demand case will be the same as if the market size is certainly high. Notice that Anand and Giraud-Carrier (2020) prove that



(a) $d = 1$ (b) $d = 2$

Figure 4 Expected social welfare of the two instruments, where $\mu = 1$ and $\gamma = 10$.

if the market is deterministic, the optimal price instrument and the optimal quantity instrument will essentially lead to the same social welfare. Therefore, in our model, the optimal social welfare generated from the high-demand case is the same under both instruments. Accordingly, the difference between the two instruments only comes from the low-demand case. As discussed, in the low-demand case, the quantity instrument provides the firm with greater *abatement flexibility*. Similar to the emission flexibility, the abatement flexibility is also a double-edged sword. On the positive side, the firm chooses the production amount that yields more profit; on the negative side, the emission without abatement efforts will cause more pollution damage. Hence, the overall effect of the abatement flexibility depends on the emission intensity. Specifically, when the emission intensity is not significant ($d < \Delta$), the production profit outweighs the pollution damage, resulting in an overall positive effect of the abatement flexibility. Therefore, the expected social welfare under the quantity instrument is greater than that under the price instrument (see region C in Figure 4a). When the emission intensity is significant ($d > \Delta$), the pollution damage outweighs the production profit, resulting in an overall negative effect of the abatement flexibility. Therefore, the expected social welfare under the price instrument is greater than that under the quantity instrument (see region C in Figure 4b). Moreover, as the market uncertainty increases, the contribution to social welfare by the low-demand market decreases so that the difference between the two instruments is diminishing.

Then, we discuss the case when the market uncertainty is intermediate. See region B in both Figures 4a and 4b for an illustration. As argued in the previous two paragraphs, the better instrument under sufficiently high and low market uncertainty is different. Thus, the curves of the expected social welfare under different instruments intersect when the market uncertainty is intermediate.

One important remark is that here the two instruments are now in different phases. Specifically, when the emission intensity is not significant, the price instrument is in the inclusive phase while the quantity instrument is in the exclusive phase (see region B in Figure 4a). This is because as emission intensity is small, providing the firm with flexibility either in emission or abatement is beneficial to the entire society. The opposite becomes true when the emission intensity is significant, implying different phases of the two instruments in region B of Figure 4b.

Next, we discuss the crucial thresholds (i.e., δ, Δ and Ω) in Theorem 1. Similar to δ^P and δ^Q, the first threshold δ represents the minimal level of emission intensity at which regulation should be imposed; the second threshold Δ represents the level of emission intensity at which the ranking of the two instruments reverses; the last threshold Ω represents the size of market uncertainty at which the preferred instrument changes. Then, we analyze the impact of the firm's abatement capability $1/\gamma$ on the first two thresholds. The first threshold δ decreases in $1/\gamma$. As the firm's abatement ability is enhanced, the regulator can urge the firm to take more abatement responsibility. Therefore, the region where neither instrument should be implemented narrows. The second threshold Δ also decreases in $1/\gamma$. When $d < \Delta$, the regulator will provide either the emission or abatement flexibility to firms. Thus, when the firm's abatement ability is enhanced, the regulator should provide less flexibility to the firm such that the firm can take more abatement responsibility. Therefore, the region with flexibility shrinks.

Lastly, we compare the expected abatement and production amounts under these two instruments.

Proposition 3. *Compared to the quantity instrument, the price instrument always induces a (weakly) higher expected abatement amount but a (weakly) lower expected production amount.*

According to Proposition 3, compared to the quantity instrument, the price instrument always induces more abatement but less production. Notice that, when the emission intensity is low and the market uncertainty is high, or the emission intensity is high and the market uncertainty is low, the price instrument leads to lower social welfare but a higher abatement amount. Therefore, more abatement does not necessarily mean higher social welfare, and the balance between abatement and production is intriguing.

3.4. Managerial Implications of Main Results

Our analysis suggests that the selection of a better regulatory tool should be based on the comprehensive information of both industry and market, particularly the industry emission intensity and the market uncertainty. Based on our above results, we provide answers to the following two questions.

(1) *Should regulators provide flexibility to the industry (firm)?* The trade-off is to compare the additional industry profit and the additional pollution damage due to flexibility. Clearly, flexibility always increases industry profit, but the level of the resulting environmental damage will depend on the industry emission intensity. In particular, for high-emitting (low-emitting) industries, the flexibility will significantly (not significantly) increase emissions, which in turn will significantly (not significantly) affect the environment. Therefore, the answer is that *regulators should provide (not provide) flexibility to low-emitting (high-emitting) industries.*

(2) *Which regulatory tool is more flexible for firms under different market conditions?* The trade-off is as follows. With high market uncertainty, a firm may gain abatement flexibility from the quantity instrument, because it will benefit from sufficient permits rather than suffer from a high tax in the depressed market; while with low market uncertainty, a firm may prefer the price instrument, because it can flexibly adjust its net emission amount contingent on the market condition rather than always emit exactly the licensed cap. Therefore, the answer is that *with high (low) market uncertainty, the quantity (price) instrument is more flexible for firms.*

The above results provide profound guidance for practices. For example, in recent years, due to the epidemic and the Russo-Ukrainian War, the market uncertainty has increased significantly.[12,13] Our results suggest that regulators may need to change their regulatory approach to deal with such market unpredictability. At the same time, each regulator is constantly incorporating more corporate entities into the regulated market. Our analysis implies that habitually assimilating them into one specific regulatory system will result in unexpected welfare losses as these regulated industries often differ in their emission intensity. Based on our model setup, we numerically calculate $\frac{\max\{\mathcal{S}^P,\mathcal{S}^Q\}}{\min\{\mathcal{S}^P,\mathcal{S}^Q\}}-1$ and find that this value can reach around 30%. Moreover, the numerical study also suggests that choosing the better instrument is particularly important for regulating high-emitting industries.

Lastly, we provide some insights related to the abatement. Nowadays, more and more companies are realizing the importance of green investments,[14] and several world leaders have planned to join forces at COP26 to boost green technologies.[15] Our research shows that as emission reduction technologies continue to improve, there will be fewer unregulated industries, implying that more entities should be included in the regulatory market. Furthermore, for industries that are already regulated, we find that in addition to strengthening the specific regulation (such as increasing emission tax in the price instrument or reducing emission cap in the quantity instrument), regulators can also change the regulatory instrument to adapt to a more progressive and advanced market.

4. Extensions

In this section, we incorporate more practical factors to illustrate the robustness of our main results and provide more insightful implications. To be specific, we introduce firm competition and national/regional environmental damage in Section 4.1, explore the government's quick-response capability and the hybrid instrument in Section 4.2, and examine the value of the permit trading market in the cap-and-trade scheme in Section 4.3.

4.1. Firm Competition

Competition exists in most industries and may affect the effectiveness of regulatory instruments (Requate 2006). In this subsection, we modify our base model into a Cournot competition setting. In the following, we investigate the case when there are multiple homogeneous firms, aiming to figure out the impact of competition. (In Appendix A, we further study the case with imperfect market information.) The subscript 'c' is provided as a mnemonic for 'competition'.

Consider a market with one regulator and N symmetric firms producing the same type of product. The firms face a linear inverse demand function: $p(\theta, x_1, x_2, \ldots, x_N) := \theta - x_i - x_{-i}$, where x_i is the production quantity of firm i, $x_{-i} := \sum_{j \neq i} x_j$ and p is the equilibrium product price. For ease of analysis, we focus on the symmetric equilibrium. In practice, firms may be located in different cities or states, and their pollution (e.g., water pollution, solid waste pollution) may not fully spread to the overall environment (see, e.g., Eichner and Runkel 2012, Holland et al. 2016, Deryugina et al. 2019). In this case, in addition to the national pollution damage (see, e.g., Anand and Giraud-Carrier 2020), there may also be regional pollution damage. For example, sulfur dioxide induces not only national pollution damage, e.g., global warming, but also regional pollution damage, e.g., acid rain. Thus, we assume that the total pollution damage is

$$D(x_1, x_2, \ldots, x_N; w_1, w_2, \ldots, w_N) = \underbrace{\beta \cdot d \sum_{i=1}^{N} \left[(x_i - w_i)^+\right]^2}_{\text{Regional Pollution Damage}} + \underbrace{(1-\beta) \cdot d \Big[\sum_{i=1}^{N} \big(x_i - w_i\big)^+\Big]^2}_{\text{National Pollution Damage}},$$

where $\beta \in (0,1)$ is the proportion of regional pollution damage and w_i is firm i's abatement amount. In this case, the regional pollution damage in each region i is $d[(x_i - w_i)^+]^2$ and the national pollution damage is convex in the total emission amount $\sum_i (x_i - w_i)^+$. Then, similar to (1), the objective of the regulator is to maximize the expected social welfare:

$$\mathbb{E}_\theta^{\pi(\alpha)}\Bigg[\underbrace{\sum_{i=1}^{N} \big(\theta - \sum_{i=1}^{N} x_i\big) \cdot x_i - \sum_{i=1}^{N} \frac{\gamma}{2} w_i^2}_{\text{Production Profit}} + \underbrace{\frac{1}{2}\Big(\sum_{i=1}^{N} x_i\Big)^2}_{\text{Consumer Surplus}} - \underbrace{\Big(\beta d \sum_{i=1}^{N} \left[(x_i - w_i)^+\right]^2 + (1-\beta) d \Big[\sum_{i=1}^{N} \big(x_i - w_i\big)^+\Big]^2\Big)}_{\text{Pollution Damage}}\Bigg]. \tag{6}$$

We first analyze the results for each instrument.

Price Instrument. We start with the price instrument. Given a tax rate t, firm i chooses the production amount x_i and the abatement amount w_i to maximize its profit:

$$\max_{x_i, w_i: 0 \le w_i \le x_i} \left\{ (\theta - x_{-i} - x_i)x_i - \frac{\gamma}{2} w_i^2 - t(x_i - w_i)^+ \right\}. \tag{7}$$

The optimal solution in this case is summarized in the following lemma.

Lemma 1 **(Optimal Price Instrument with Firm Competition)**. *Given N identical competing firms, the optimal price instruments are as follows.*

- *When $d \le \delta_c^P$, the optimal tax rate and the corresponding expected social welfare are $t_c^P = 0$ and $\mathcal{S}_c^P = \frac{N(2\beta d(N-1) - 2dN + N + 2)}{2(N+1)^2}(\mu^2 + \sigma^2)$.*
- *When $d > \delta_c^P$ and $\sigma/\mu \le \Omega_c^P$, the optimal tax rate and the corresponding expected social welfare are $t_c^P = \mathcal{C}_{c,1}^P \mu$ and $\mathcal{S}_c^P = \mathcal{C}_{c,2}^P \mu^2 + \mathcal{C}_{c,3}^P \sigma^2$. As σ increases, t_c^P remains constant, and $\mathcal{S}_c^P$ increases (decreases, resp.) when $d < \frac{N+2}{2(N+\beta-\beta N)}$ ($d < \frac{N+2}{2(N+\beta-\beta N)}$, resp.).*
- *When $d > \delta_c^P$ and $\sigma/\mu > \Omega_c^P$, the optimal tax rate and the corresponding expected social welfare are $t_c^P = \mathcal{C}_{c,1}^P(\mu + \sigma)$ and $\mathcal{S}_c^P = \mathcal{C}_{c,4}^P(\mu^2 + \sigma^2) + \mathcal{C}_{c,5}^P \mu\sigma$. Both t_c^P and $\mathcal{S}_c^P$ increase in σ.*

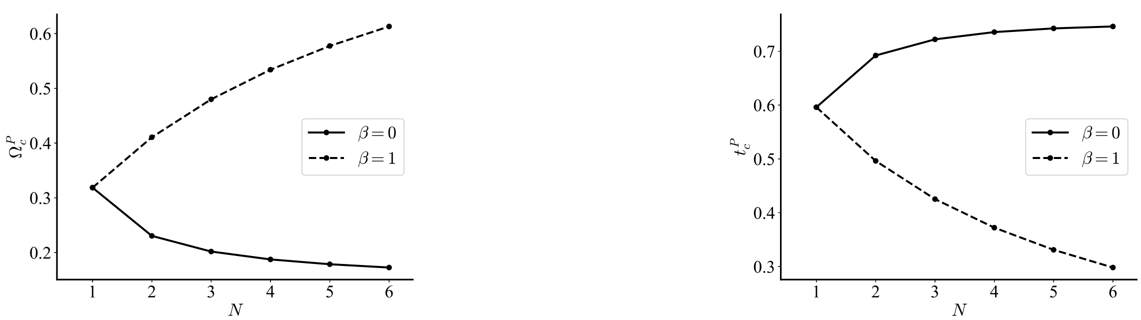
Here, δ_c^P, Ω_c^P, $\mathcal{C}_{c,1}^P$, $\mathcal{C}_{c,2}^P$, $\mathcal{C}_{c,3}^P$, $\mathcal{C}_{c,4}^P$ and $\mathcal{C}_{c,5}^P$ are constants depending only on γ, d, β and N, and their detailed forms are provided in the proof. Moreover, as $\gamma \to \infty$, we have the following comparative statics:

- *δ_c^P always decreases in N.*
- *Ω_c^P increases in N when $\beta > (2d+1)/4d$ and decreases otherwise.*
- *t_c^P decreases in N when $\beta > (2d+1)/4d$ and increases otherwise.*
- *When $d > \delta_c^P$ and $\sigma/\mu \le \Omega_c^P$, $\mathcal{S}_c^P$ increases in N when $\beta > \hat{\beta}^P$ and decreases otherwise. When $d > \delta_c^P$ and $\sigma/\mu > \Omega_c^P$, $\mathcal{S}_c^P$ always increases in N.*

The results in Lemma 1 are similar to the base model (see Proposition 1). In the following, we will explain the comparative statics in Proposition 1, illustrated in Figure 5. For ease of exposition, we focus on the case when the abatement cost is relatively high ($\gamma \to \infty$), and hence the pollution damage is the main impact factor. To interpret the comparative statics, we need to understand the impact of firm competition on pollution damage. Given a fixed tax rate t, when the number of firms increases, the total pollution amount $\sum_i \left(x_{c,i}^P(\theta,t) - \sum_i w_{c,i}^P(\theta,t)\right)^+$ will increase, while the pollution amount of each single firm $\left(x_{c,i}^P(\theta,t) - w_{c,i}^P(\theta,t)\right)^+$ will decrease. Accordingly, there are two opposite impacts of firm competition on pollution damage.

- **Accumulation Effect:** The increased total pollution amount will result in higher national pollution damage. That is, the national pollution damage $\left[\sum_i \left(x_{c,i}^P(\theta,t) - w_{c,i}^P(\theta,t)\right)^+\right]^2$ increases in N.
- **Dispersion Effect:** The decreased production amount of each single firm will lead to lower regional pollution damage. That is, the regional pollution damage $\sum_i \left[\left(x_{c,i}^P(\theta,t) - w_{c,i}^P(\theta,t)\right)^+\right]^2$ decreases in N.

These two effects together play a key role in understanding Lemma 1. According to Lemma 1, the unregulated region (i.e., δ_c^P) shrinks as the number of firms increases. The reason is that additional firms induce excessive production, such that regulation will be expected to be imposed to mitigate competition and pollution damage. Similar to the base model (Proposition 1), there are also two phases under the regulated region. We can show that the phase boundary Ω_c^P increases in N when $\beta > (2d+1)/4d$, but decreases otherwise (see Figure 5a). This is because when the regional (national, resp.) damage is the main concern, the dispersion (accumulation, resp.) effect dominates such that pollution damage becomes less (more, resp.) significant as N increases, implying that the regulator should (should not, resp.) offer flexibility to firms. As discussed in Section 3.3, the flexibility of the price instrument only exists in the inclusive phase, namely $\sigma/\mu \le \Omega_c^P$. Thus, Ω_c^P has different monotonic property in N for different β. As for t_c^P, illustrated in Figure 5b, in either the inclusive or exclusive phase, it always decreases in N when $\beta > (2d+1)/4d$, and increases otherwise. This is also because when the dispersion (accumulation, resp.) effect dominates, the regulator should adopt a looser (stricter, resp.) regulation due to the decreased (increased, resp.) pollution damage.



(a) Ω_c^P (b) t_c^P

Figure 5 **The threshold and the optimal tax rate as functions of N under the price instrument, where $\mu=1$, $\sigma=0.1$, $d=2$, $\beta\in\{0,1\}$ and $\gamma=200$.**

Now we discuss the property of the optimal expected social welfare $\mathcal{S}_c^P$ with respect to N. We can show that when $\sigma/\mu > \Omega_c^P$, $\mathcal{S}_c^P$ always increases in N. This is because under this condition, the regulation is in the exclusive phase. Then whether the accumulation or dispersion effect dominates, the regulator will always use a strategy that maximizes social welfare for the high-demand case and induces the firm to make minuscule production in the low-demand case. Thus, the social welfare in this phase is mainly contributed by the high-demand case, where the accumulation effect is well regulated, and hence social welfare $\mathcal{S}_c^P$ rises due to the dispersion effect. When $\sigma/\mu < \Omega_c^P$, the regulator cares about both demand cases, and thus the regulation will not maximize social welfare for either demand case. It implies that the dispersion and accumulation effects will factor in. In fact, we can show that there is a threshold $\hat{\beta}^P$ such that when $\beta > \hat{\beta}^P$ ($\beta < \hat{\beta}^P$, resp.), the dispersion (accumulation, resp.) effect dominates, then $\mathcal{S}_c^P$ increases (decreases, resp.) in N.

Quantity Instrument. Next, we study the quantity instrument. Since firms are symmetric, we consider a uniform cap k and do not consider permit trading among firms. Given a cap k, each firm i chooses the production amount x_i and the abatement amount w_i to maximize its profit:

$$\max_{x_i, w_i: 0 \le x_i - w_i \le k; w_i \ge 0} \left\{ (\theta - x_{-i} - x_i) x_i - \frac{\gamma}{2} w_i^2 \right\}. \tag{8}$$

The optimal solution in this case is summarized in the following lemma.

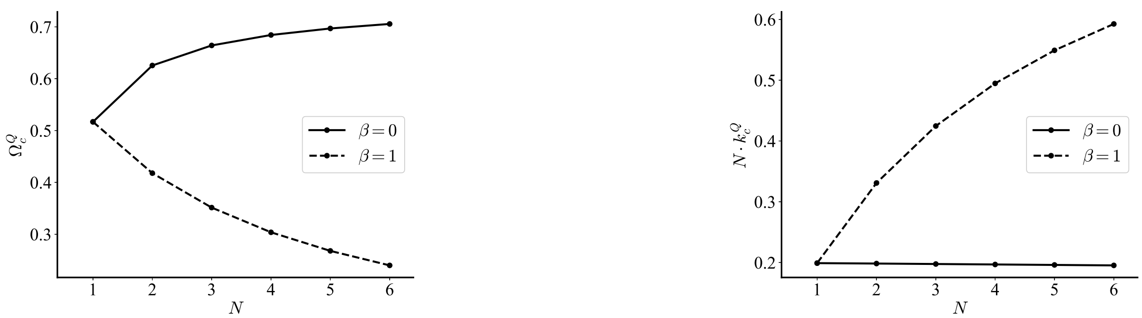
Lemma 2 **(Optimal Quantity Instrument with Firm Competition)**. *Given N identical competing firms, the optimal quantity instruments are as follows.*

- *When $d \le \delta_c^Q$, the optimal tax rate and the corresponding expected social welfare are $k_c^Q \ge \frac{\mu+\sigma}{N+1}$ and $\mathcal{S}_c^Q = \frac{N(2\beta d(N-1) - 2dN + N + 2)}{2(N+1)^2}(\mu^2 + \sigma^2)$.*
- *When $d > \delta_c^Q$ and $\sigma/\mu \le \Omega_c^Q$, the optimal tax rate and the corresponding expected social welfare are $k_c^Q = \mathcal{C}_{c,1}^Q \mu$ and $\mathcal{S}_c^Q = \mathcal{C}_{c,2}^Q \mu^2 + \mathcal{C}_{c,3}^Q \sigma^2$. As σ increases, k_c^Q remains constant and $\mathcal{S}_c^Q$ increases.*
- *When $d > \delta_c^Q$ and $\sigma/\mu > \Omega_c^Q$, the optimal tax rate and the corresponding expected social welfare are $k_c^Q = \mathcal{C}_{c,1}^Q(\mu + \sigma)$ and $\mathcal{S}_c^Q = \mathcal{C}_{c,4}^Q(\mu^2 + \sigma^2) + \mathcal{C}_{c,5}^Q \mu\sigma$. Both k_c^Q and $\mathcal{S}_c^Q$ increase in σ.*

Here, δ_c^Q, Ω_c^Q, $\mathcal{C}_{c,1}^Q$, $\mathcal{C}_{c,2}^Q$, $\mathcal{C}_{c,3}^Q$, $\mathcal{C}_{c,4}^Q$ and $\mathcal{C}_{c,5}^Q$ are constants depending only on γ, d, β and N, and their detailed forms are provided in the proof. Moreover, as $\gamma \to \infty$, we have the following comparative statics:

- *δ_c^Q always decreases in N.*
- *Ω_c^Q decreases in N when $\beta > (2d+1)/4d$ and increases otherwise.*
- *Nk_c^Q increases in N.*
- *When $d > \delta_c^Q$ and $\sigma/\mu \le \Omega_c^Q$, $\mathcal{S}_c^Q$ increases in N. When $d > \delta_c^Q$ and $\sigma/\mu > \Omega_c^Q$, $\mathcal{S}_c^Q$ increases in N when $\beta \to 1$ and decreases when $\beta \to 0$.*

In this case, the unregulated region under the quantity instrument will also shrink as N increases, which is because of firms' excessive production. According to Lemma 2, similar to the base model (Proposition 2), there are also two phases under the regulated region. Similar to the price instrument, we explain the comparative statics in Lemma 2 when the abatement cost is relatively high, illustrated in Figure 6. In this case, the threshold Ω_c^Q decreases in N when $\beta > (2d+1)/4d$, but increases otherwise (see Figure 6a). The explanation is similar to that for the price instrument by arguing whether to offer flexibility to firms. Differently, the flexibility of the quantity instrument only exists in the exclusive phase, namely $\sigma/\mu > \Omega_c^Q$. As for k_c^Q, it always decreases in N in either the inclusive or exclusive phase. This is because when N increases, *ceteris paribus*, the firm's self-protection incentive will always let it reduce its production amount, thus so will the imposed cap. We can also show that the optimal total cap $N \cdot k_c^Q$ always increases in N (see Figure 6b). This is because the accumulation effect is well regulated by the total cap and the dispersion effect induces a loose regulation, i.e., a larger total cap.



(a) Ω_c^Q

(b) $N \cdot k_c^Q$

Figure 6 **The threshold and the optimal total cap as functions of N under the quantity instrument, where $\mu = 1$, $\sigma = 0.1$, $d = 2$, $\beta \in \{0, 1\}$ and $\gamma = 200$.**

Now we analyze the impact of N on the optimal expected social welfare $\mathcal{S}_c^Q$. The expected social welfare $\mathcal{S}_c^Q$ increases in N when $\sigma/\mu \le \Omega_c^Q$. This is because under this condition, with either demand market size, the firm will always emit exactly the imposed cap, such that the accumulation effect can be well regulated. At the same time, due to the amplified dispersion effect by a larger N, the pollution damage is less significant, and hence the optimal social welfare increases. When $\sigma/\mu > \Omega_c^Q$, we find that $\mathcal{S}_c^Q$ increases in N when β is sufficiently large, but decreases when β is sufficiently small. This is because the regulation is under the exclusive phase, for which the regulator designs the policy that maximizes social welfare only for the high-demand case. As the firms still make unlimited production under the low-demand case, their contribution to the expected social welfare is non-negligible. When the national pollution damage is the main concern ($\beta \to 0$), the accumulation effect in the low-demand case is significant, and hence the social welfare $\mathcal{S}_c^Q$ decreases in N. When the regional damage is the main concern ($\beta \to 1$), the social welfare $\mathcal{S}_c^Q$ increases in N due to the dispersion effects from both cases.

Finally we compare the two instruments under competition. The result is summarized in the following theorem.

THEOREM 2 **(Instrument Comparison with Firm Competition)**. *Define* $\delta_c = \frac{\gamma}{2(N+\beta-\beta N)(N+\gamma+1)}$, $\Delta_c = \frac{\gamma(N^2+4N+3+(N+2)\gamma)}{2(N+\gamma+1)^2(N+\beta-\beta N)}$ *and* $\Omega_c = \sqrt{2} - 1$. *If there are N identical firms competing in the market, then the comparison of the optimal price and quantity instruments is as follows.*

- *When* $d \in [0, \delta_c]$*, neither instruments should be implemented.*
- *When* $d \in (\delta_c, \Delta_c)$*, the expected social welfare under the optimal price instrument is higher (lower, resp.) than that under the optimal quantity instrument when* $\sigma/\mu < \Omega_c$ *(*$\sigma/\mu > \Omega_c$*, resp.).*
- *When* $d \in (\Delta_c, \infty)$*, the expected social welfare under the optimal quantity instrument is higher (lower, resp.) than that under the optimal price instrument when* $\sigma/\mu < \Omega_c$ *(*$\sigma/\mu > \Omega_c$*, resp.).*

From Theorem 2, we can see that the comparison between the two instruments qualitatively shares the same structure as that in the base model (Theorem 1). However, the thresholds, δ_c and Δ_c, are affected by the number of firms N and the regional effect ratio β. We start with the sensitivity analysis with respect to the regional effect ratio β. Both thresholds increase in β because the regional damage $\sum_i [(x_i - w_i)^+]^2$ is relatively smaller than the national damage $[\sum_i (x_i - w_i)^+]^2$. Then a larger β has the same effect of reducing the emission intensity coefficient d. Therefore the thresholds increase in the regional effect ratio β.

Then we conduct the sensitivity analysis of the two thresholds δ_c and Δ_c with respect to the number of firms N. The first threshold δ_c decreases in the number of firms N. This is consistent with the analysis of δ_c^P and δ_c^Q. That is, as N increases, the excessive production is more significant, so the unregulated region shrinks. We note that the second threshold Δ_c is also monotone in N, but its direction depends on the regional effect ratio β. For ease of exposition, we focus on the case when the abatement cost is sufficiently high ($\gamma \to \infty$). In this case, when $\beta < 2/3$ ($\beta > 2/3$, resp.), this threshold decreases (increases, resp.) in N. To understand this property, recall the discussion in Section 3.4: When $d \le \Delta_c$, the regulator will provide either the emission or abatement flexibility to firms. Thus, when β is large (small, resp.), the dispersion (accumulation, resp.) effect dominates, and hence the pollution damage decreases (increases, resp.) in N, implying a larger (smaller, resp.) region of providing flexibility to firms.

Managerial Implications of Competitive Markets. The study in this subsection complements the analysis of the base model by incorporating (1) firm competition and (2) the dispersion effect and the accumulation effect of pollution. The results provide insights into a regulated market that encompasses multiple firms, especially those that are located in different regions. Our analysis suggests that the proper regulation should be based on the number of competing firms in the market. In fact, the number of regulated entities in the regulatory markets is constantly increasing, either because of voluntary or mandatory inclusion of new territories. For example, California and Quebec share a market, with Manitoba, Ontario, and provinces in Mexico and Brazil all planning to join.[16] China has also begun to activate a national cap-and-trade system consisting of most cities.[17] We find that the level of competition will particularly affect the regulator's design of the regulatory flexibility for firms. To specify the level of flexibility, regulators need to investigate whether pollution from each firm, industry, or region is more likely to have regional or national impacts. For example, SO_2 and NO_X will cause regional acid rains, whereas CO_2 and CH_4 will result in global warming. The basic trade-off is that when pollution is more likely to be regional, the dispersion effect predominates, and then when competition becomes more intense, the regulator should deregulate.

4.2. Quick-Response Capability and Hybrid Instrument

In this subsection, we explore two potential ways the government can further improve social welfare. First, we study the quick-response capability that enables the government to adjust regulatory instruments according to the realized market size θ. Second, we investigate a hybrid instrument that appropriately combines the strengths of the two pure instruments studied.

Quick-Response Capability. We start with the quick-response case where the government can set different tax rates or caps for different demand cases. The superscript 'R' is provided as a mnemonic for 'Quick-Response'. Specifically, with the price instrument, the government determines tax rates t_h^R and t_ℓ^R for the high-demand case ($\theta_h = \mu + \sigma$) and the low-demand case ($\theta_\ell = \mu - \sigma$), respectively, as follows:

$$\max_{t_i^R \geq 0} \left\{ \big(\theta_i - x^P(\theta_i, t_i^R)\big) x^P(\theta_i, t_i^R) - \frac{\gamma}{2}\big(w^P(\theta_i, t_i^R)\big)^2 + \frac{1}{2}\big(x^P(\theta_i, t_i^R)\big)^2 - d\big[(x^P(\theta_i, t_i^R) - w^P(\theta_i, t_i^R))^+\big]^2 \right\}, \; i = h, \ell.$$

With the quantity instrument, the government determines caps k_h^B and k_ℓ^B for the two demand cases as follows:

$$\max_{k_i^R \geq 0} \left\{ \big(\theta_i - x^Q(\theta_i, k_i^R)\big) x^Q(\theta_i, k_i^R) - \frac{\gamma}{2}\big(w^Q(\theta_i, k_i^R)\big)^2 + \frac{1}{2}\big(x^Q(\theta_i, k_i^R)\big)^2 - d\big[(x^Q(\theta_i, k_i^R) - w^Q(\theta_i, k_i^R))^+\big]^2 \right\}, \; i = h, \ell.$$

The results are summarized in the following proposition.

PROPOSITION 4 **(Quick-Response Capability)**. *Define $\delta^R = \frac{\gamma}{2(\gamma+2)}$. Given the government's quick-response capability, these two regulatory instruments achieve the same optimal expected social welfare. Specifically, we have*

- *When $d \leq \delta^R$, the optimal tax rates are $t_h^R = t_\ell^R = 0$, and the optimal caps are $k_h^R \geq \frac{\mu+\sigma}{2}$ and $k_\ell^R \geq \frac{\mu-\sigma}{2}$. The corresponding expected social welfare is $\mathcal{S}^R = \frac{3-2d}{8}(\mu^2 + \sigma^2)$.*
- *When $d > \delta^R$, the optimal tax rates are $t_h^R = \mathcal{C}_1^R(\mu+\sigma)$ and $t_\ell^R = \mathcal{C}_1^R(\mu-\sigma)$, and the optimal caps are $k_h^R = \mathcal{C}_2^R(\mu+\sigma)$ and $k_\ell^R = \mathcal{C}_2^R(\mu-\sigma)$. The corresponding expected social welfare is $\mathcal{S}^R = \mathcal{C}_3^R(\mu^2+\sigma^2)$.*

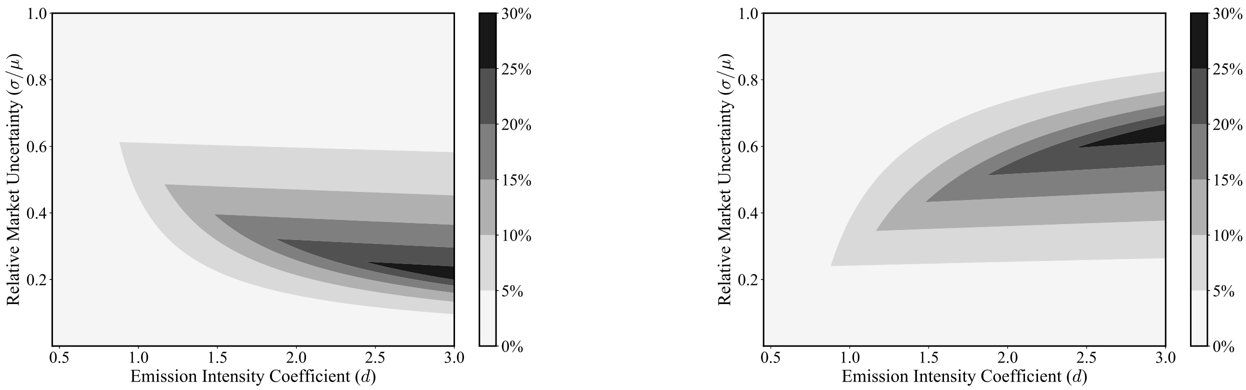
Here, $\mathcal{C}_1^R$, $\mathcal{C}_2^R$ and $\mathcal{C}_3^R$ are non-negative constants depending only on d and γ, and their detailed forms are provided in the proof.

According to Proposition 4, given the government's quick-response capability, the optimal expected social welfare under these instruments is the same. Such a result is similar to Anand and Giraud-Carrier (2020) which show that without uncertainty these two instruments achieve the same expected social welfare at optimality. This is because, for each demand case, the regulator maximizes social welfare when the market size is realized (we refer to such maximal value of either demand case as the "*ex-post optimality*"). Therefore, the expected social welfare of the quick-response case is always greater than that under the setup of the base model, i.e., $\mathcal{S}^R \geq \max\{\mathcal{S}^P, \mathcal{S}^Q\}$. In the following, we will compare the quick-response case with the base model to facilitate understanding the importance of the government's quick-response capability.

Comparison with Base Model. It is foreseeable that pursuing the quick-response capability will incur additional administrative costs (e.g., more staff and advanced information systems) to the government. Therefore, the government may not have enough incentive to make such an effort if the improved social welfare is not significant. In order to identify the situations where the quick-response capability is worth implementing, in the following, we compute the improvement ratio of the quick-response case compared to the base case. Since no regulation shall be implemented under both cases when $d \leq \delta^R$, we will focus on the cases when $d > \delta^R$.

Comparison Regarding Price Instruments. We first compare the price instruments under these two cases. According to Propositions 1 and 4, in terms of expected social welfare, the improvement ratio of the quick-response case compared to the base case is $\mathcal{R}^P := (\mathcal{S}^R - \mathcal{S}^P)/\mathcal{S}^P$. Then we analyze the property of $\mathcal{R}^P$. In the inclusive (exclusive, resp.) phase, the improvement ratio $\mathcal{R}^P$ increases (decreases, resp.) in the market uncertainty σ. This is because the quick-response case always achieves the ex-post optimality for either demand case. Whereas, in the inclusive phase, the price instrument in the base model cannot attain the ex-post optimality for either demand case, and the optimality gap increases in the difference between the demand cases. Therefore, as the market uncertainty increases, the importance of the quick-response capability is enhanced. In contrast, in the exclusive phase, the price instrument in the base model achieves the ex-post optimality for the high-demand case. As the market uncertainty increases, the contribution of the low-demand case to social welfare becomes increasingly negligible, resulting in a diminishing importance of the quick-response capability. Moreover, according to Figure 7a, the quick-response capability improves the social welfare the most when the market uncertainty is small ($\sigma = \Omega^P \mu$) and the emission intensity is large ($d \to \infty$). This is because the overall negative effect of the emission flexibility is amplified by the large emission intensity.

Comparison Regarding Quantity Instruments. Next, we compare the quantity instrument under these two cases. According to Propositions 2 and 5, in terms of expected social welfare, the improvement ratio of the quick-response case compared to the base case is $\mathcal{R}^Q := (\mathcal{S}^R - \mathcal{S}^Q)/\mathcal{S}^Q$. Similar to



(a) $\frac{\mathcal{S}^R - \mathcal{S}^P}{\mathcal{S}^P}$ (b) $\frac{\mathcal{S}^R - \mathcal{S}^Q}{\mathcal{S}^Q}$

Figure 7 **Social welfare improvement ratios of the quick-response case compared to the base case as functions of $d \in [\delta^R, 3]$ and $\sigma \in [0, 1)$ with $\mu = 1$ and $\gamma = 20$.**

the comparison regarding price instruments, in the inclusive (exclusive, resp.) phase, the improvement ratio $\mathcal{R}^Q$ increases (decreases, resp.) in the market uncertainty σ. In the inclusive phase, the optimal quantity instrument cannot achieve the ex-post optimality for either demand case, and hence the importance of the quick-response capability increases in the market uncertainty. In the exclusive phase, as the market uncertainty increases, the contribution by the low-demand case becomes increasingly negligible, resulting in the diminishing importance of the quick-response capability. Moreover, according to Figure 7b, the overall negative effect of the abatement flexibility is amplified by the large emission intensity, and hence the quick-response capability improves social welfare significantly when the market uncertainty is large ($\sigma = \Omega^Q \mu$) and the emission intensity is large ($d \to \infty$).

Hybrid Instrument. In addition to the quick-response capability, we provide an alternative way for the government to improve the expected social welfare. In the following, we investigate a hybrid regulatory instrument that combines the price and the quantity instruments. In the hybrid instrument, the regulator provides baseline emission permits κ to the firm, and the firm can buy extra permits at a price τ from the regulator. Notice that the price instrument with a tax rate t can be represented by $(\kappa, \tau) = (0, t)$, while the quantity instrument with a cap k can be represented by $(\kappa, \tau) = (k, \infty)$. Therefore, the hybrid instrument performs no worse than either pure instrument if not considering the additional implementation cost. We note that the hybrid instrument differs from the cap-and-trade system studied in Section 4.3, and the difference between the two is who the firm buys extra permits from, the regulator or other firms. In addition, unlike the hybrid instrument in this subsection, a cap-and-trade system requires the participation of multiple firms. Therefore, we can interpret the two studies as exploring the additional compulsory regulation of the regulator or the additional self-regulating effect of the market, respectively. The superscript 'H' is provided as a mnemonic for 'Hybrid'.

Given κ and τ, the firm will choose its production amount x and abatement amount w by solving the following problem:

$$\max_{x\geq w\geq 0}\left\{\theta x - x^2 - \frac{\gamma}{2}w^2 - \tau\cdot\max\{x-w-\kappa,0\}\right\}. \tag{9}$$

PROPOSITION 5 **(Hybrid Instrument)**. *Define $\delta^H=\frac{\gamma}{2(\gamma+2)}$. The optimal hybrid instruments are as follows.*

- *When $d\leq\delta^H$, no regulation shall be imposed and the corresponding expected social welfare is $\mathcal{S}^H=\frac{3-2d}{8}(\mu^2+\sigma^2)$.*
- *When $d>\delta^H$, the optimal hybrid instrument and the corresponding expected social welfare are $(\kappa^H,\tau^H)=\big(\mathcal{C}_1^H(\mu-\sigma),\,\mathcal{C}_2^H(\mu+\sigma)\big)$ and $\mathcal{S}^H=\mathcal{C}_3^H(\mu^2+\sigma^2)$.*

Here $\mathcal{C}_1^H$, $\mathcal{C}_2^H$ and $\mathcal{C}_3^H$ are non-negative constants depending only on d and γ, and their detailed forms are provided in the proof.

According to Proposition 5, similar to the previous section, when the emission intensity is sufficiently small ($d<\delta^H$), the regulator should not restrict the firm's production. When the emission intensity is relatively large ($d>\delta^H$), under the optimal hybrid instrument, the firm will choose the production and the abatement amounts the same as those in the quick-response case. Subsequently, the hybrid instrument achieves the same expected social welfare as the quick-response case, $\mathcal{S}^H=\mathcal{S}^R$. Moreover, as the market uncertainty σ increases, the optimal baseline cap κ^H decreases, and the optimal permit price τ^H increases. The result contradicts that of adopting a pure instrument where both the optimal tax rate t^P and cap k^Q of pure instruments increase in the market uncertainty σ. This is because, under the hybrid instrument, the regulator mainly uses κ^H (τ^H, resp.) to regulate the low-demand (high-demand, resp.) case. As the market uncertainty increases, the low-demand market becomes less profitable, and the regulator can adjust the optimal production of the low-demand case by reducing the baseline cap κ^H. In the meantime, the firm under the high-demand case becomes more profitable, and hence the regulator increases the tax rate τ^H to prevent the firm from excessively increasing production.

Similar to the discussion in Section 4.2, coordinating both the price and quantity tools at the same time will incur additional administrative costs. Therefore, the regulator will implement the hybrid instrument only if the improved social welfare is sufficiently significant. Since we have $\mathcal{S}^H=\mathcal{S}^R$, the improvement ratios of the hybrid instrument compared to the pure price and the pure quantity instruments are the same as those of the quick-response case, i.e., $\mathcal{R}^P$ and $\mathcal{R}^Q$. As Figure 7 illustrated, the hybrid instrument improves over the price (quantity, resp.) instrument significantly when the market uncertainty is large (small, resp.) and the emission intensity is large. Moreover, such results qualitatively hold in the simulation study where the distribution of the market size is a superposition of two truncated normal distributions (see Appendix C).

Managerial Implications. Despite the administrative costs (e.g., hiring staff, information system, and coordinating the permit baseline and permit price in reality), our analysis suggests that it is still worthwhile for the government to pursue the quick-response capability or implement a hybrid instrument, especially for industries with high emission intensity. These high-emitting industries include metallurgy, building materials, and petrochemicals, which often account for a large proportion of national emissions and are heavily regulated in different jurisdictions. As shown in Figure 7, through a quick response or a hybrid instrument, the improvement ratio in social welfare can reach around 30%, compared to any fixed pure instrument. Moreover, compared to the best pure instrument, our numerical study shows that the improvement ratio can still be about 14%. As the simulation study illustrates (see Figures C4 and C5), for a different distribution of the market uncertainty, the hybrid instrument can still significantly improve social welfare and hence serves as a substitution for the quick-response capability. Such a result also implies that the hybrid instrument can better cope with market uncertainty than the pure instruments.

Our results also shed light on ways to intervene in each of the established regulatory instrument markets. More precisely, in order to improve the expected social welfare, regulators who have already implemented the price instrument should distribute some free emission permits to firms so that they can raise the carbon tax to better punish non-compliant emissions under a prosperous market. For example, with such a purpose, South Africa is currently implementing the emission tax along with basic tax-free allowances.[18] Conversely, when mainly using the quantity instrument, the regulator can set a permit price at which firms can buy extra permits. For example, in January 2021, California state added a price ceiling to its existing regulatory system.[19] By doing so, the regulators can reduce the supplied permits and consequently reduce the excessive emissions under a depressed market.

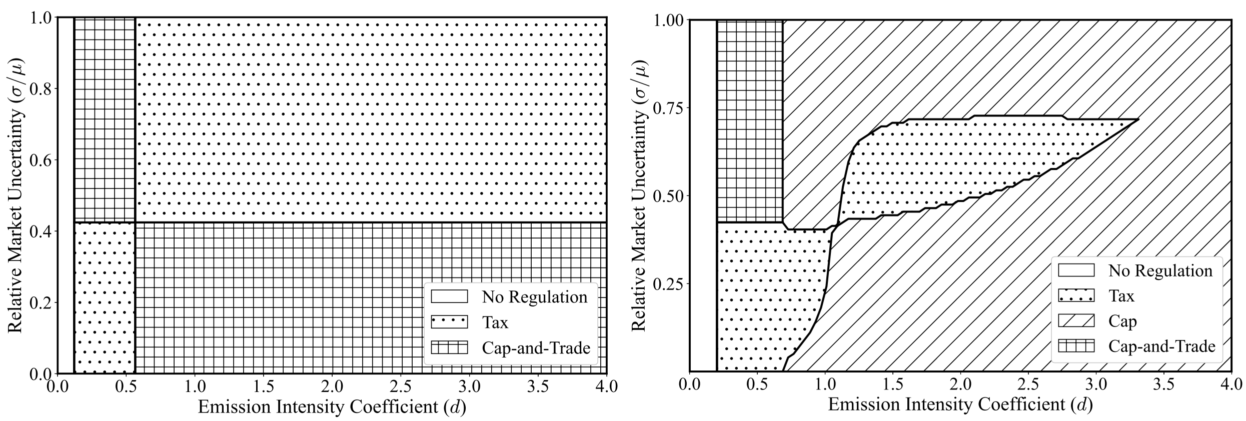
4.3. The Impact of Permit Trading

In this section, we study the case when there are two firms in different regions with heterogeneous abatement coefficients ($\gamma_1 \geq \gamma_2$). As commented in Requate (2006), the regulator generally knows the available technologies and their prevalence, but not exactly which firm has which technology. Therefore, without specific statements, we assume that the regulator cannot distinguish these two firms and hence sets the same tax rate or cap for the firms. (In Appendix B, we further explore the value of differentiated regulations and an alternative regulatory instrument.) Moreover, as Ricke et al. (2018) illustrated, different regions may have different damage severity, and hence we assume different damage coefficients, i.e., $d_1 \geq d_2$ for regional pollution damage and the damage coefficient for national pollution damage is still d.

In practice, the quantity instrument is sometimes equipped with a trading market where firms can trade their permits with each other. Such a mechanism is often called the cap-and-trade mechanism or Emission Trading System.[20] In this subsection, we explore the regulatory value of the permit trading market. The regulatory policies are now divided into three categories: (1) *The price instrument.* (2) *The quantity instrument without trading.* (3) *The quantity instrument with trading.* For the quantity instrument with trading, in more detail, given a cap k, firm i chooses its production amount x_i, abatement amount w_i and trading amount s_i to maximize its profit:

$$\max_{x_i \geq w_i \geq 0, x_i - w_i + s_i \leq k} \left\{ (\theta - x_1 - x_2)x_i - \frac{\gamma_i}{2} w_i^2 + p_c s_i \right\}, \tag{10}$$

where p_c is the market-clearing price of the trading market (such that $s_1 + s_2 = 0$). In Figure 8, we numerically show the best policy under different parameters when the pollution is either purely national or regional, namely $\beta = 0$ or $\beta = 1$.



(a) National pollution ($\beta = 0$). (b) Regional pollution ($\beta = 1$).

Figure 8 **The preferred instrument under different conditions when $\mu = 1$, $\sigma \in [0,1]$, $d_1 = d_2 = d$, $\gamma_1 = 4$ and $\gamma_2 = 2$.**

The results in Figure 8a are similar to that of the base model. Notice that according to Figure 8a, if $\beta = 0$, then the optimal expected social welfare under cap-and-trade is always higher than that under cap (without trading). The reason is that in this case, through trading, a higher total abatement amount may be induced as the firm with a smaller abatement cost coefficient will exert more abatement efforts to make profits in the trading market. However, we note that under the regional pollution setting, trading will become detrimental as d increases. See Figure 8b. The reason is that trading is a double-edged sword. On the positive side, it can induce more efficient abatement from the firm with a smaller abatement cost coefficient. On the negative side, it will induce uneven emissions, resulting in amplified regional pollution damage due to the convexity of the damage function. Moreover, according to Figure 8b, when the emission intensity is sufficiently high, the preferred instrument is always the quantity instrument without trading, contradicting the base model and the case with purely national pollution effect. The reason is that due to the heterogeneous abatement cost coefficients, the net pollution amounts across firms, $x_i - w_i$, will be uneven under the price or cap-and-trade instrument while uniform under the quantity instrument without trading (namely the cap-without-trade), resulting in additional regional pollution damage in the first two instruments.

Managerial Implications of Heterogeneous Firms. The study in this section complements the analysis of the base model by incorporating (1) competition between heterogeneous firms and (2) permit trading. Our work provides some guidance for market linkages. As stated in Green (2017), many policymakers believe that different regions should be included in the same "cap-and-trade" market. However, our study demonstrates that for cities with regional pollution, linking them with other markets may actually lead to aggravation of regional pollution. By a numerical study in Figure 9, we can observe that for cities with high emissions and regional pollution, the relative difference in social welfare between cap-without-trade and cap-and-trade can be around 20%. Therefore, we suggest that the regulators should manage these cities separately or by region, and restrict the permit trading between different regions.

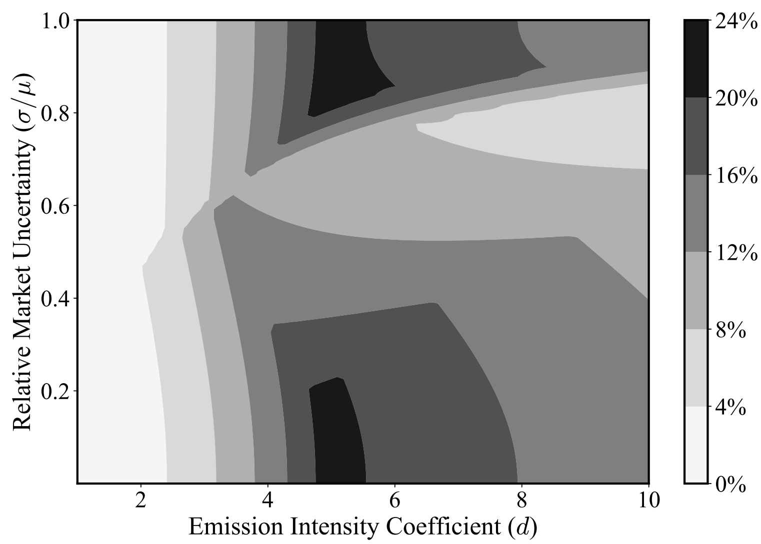


Figure 9 **The relative difference in expected social welfare between cap-without-trade and cap-and-trade when $\beta = 1$, $d_1 = d_2 = d$, $\mu = 1$, $\gamma_1 = 50$ and $\gamma_2 = 5$. The numerical model is based on the results in Section 4.3.**

5. Concluding Remarks

Nowadays, environmental issues are becoming increasingly critical and have attracted much attention of governments. When governments plan to implement an environmental regulation, there is always a debate on the choice of regulatory instruments. Among many others, the mainstream regulations include two types of instruments: price instruments and quantity instruments. In this paper, we compare the two instruments in terms of the expected social welfare. Our results demonstrate that the ranking of the two instruments flips as, *ceteris paribus*, either the emission intensity or the market uncertainty exceeds some threshold. More precisely, when the emission intensity and the relative market uncertainty are both high or both low, the expected social welfare under the price instrument is higher than that under the quantity instrument; otherwise, the expected social welfare under the quantity instrument is higher than that under the price instrument. The emission flexibility and the abatement flexibility are introduced to explain these intriguing results. Our results motivate governments to collect more information when determining the regulatory instrument. Specifically, in addition to the expected profit and pollution damage, the size of market uncertainty is also crucial in decision-making.

In Section 4.1 and Appendix A, we further incorporate some other practical factors (i.e., firm competition, regional pollution damage, and imperfect information) to test the robustness of our results. We find that when these factors are incorporated, although the thresholds are influenced, the comparison between the two instruments shares a similar structure as that in the base model (Theorem 1). We also note that more firms can be beneficial to the entire society if the regional pollution damage is the main concern. In Section 4.2, we investigate the importance of the government's quick-response capability under different market conditions. Then, a hybrid mode that combines both the price and the quantity instruments is proposed and can also significantly improve social welfare over the base model. This result has two implications. First, it supports regulatory agencies in implementing price regulations on the cap-and-trade market. Second, for jurisdictions that use carbon tax tools, providing firms with some free emission allowances can increase social welfare. Lastly, we study firms with heterogeneous abatement cost coefficients and damage coefficients through numerical experiments in Section 4.3 and Appendix B. We find that allowing permit trading under the quantity instrument may not benefit social welfare, especially when the pollution is more likely to have regional impacts. Furthermore, we find that both differentiated instruments and an alternative uniform instrument on pollution damage can significantly improve social welfare over uniform price and quantity instruments, especially when the firms' heterogeneity is large and the market uncertainty is intermediate.

To conclude, we acknowledge the limitations of our work and outline potential avenues for future research. First, we implicitly assume that all firms are regulated by the same regulator and hence overlook the impact of foreign firms and their corresponding regulators. One future direction is to investigate how regulatory instruments of foreign regulators affect the choice of regulatory instruments by the domestic regulator. Second, we mainly focus on the case with homogeneous firms and only numerically find some interesting results of the heterogeneous case. It would be intriguing to analyze the equilibrium of heterogeneous firms in the presence of market uncertainty and compare different regulatory instruments theoretically. Moreover, our work inspires some empirical studies to compare different regulatory instruments and provide detailed guidance to governments.

Notes

[1]'This is not normal': Climate change and escalating bushfire risk: https://www.climatecouncil.org.au/resources/bushfire-briefing-paper/.

[2]The Paris Agreement: https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement.

[3]The World Bank: https://carbonpricingdashboard.worldbank.org/map_data.

[4]Predicting Consumer Demand in an Unpredictable World: https://hbr.org/2020/11/predicting-consumer-demand-in-an-unpredictable-world.

[5]Federal Climate Policy 105: The Industrial Sector: https://www.rff.org/publications/explainers/federal-climate-policy-105-the-industrial-sector/.

[6]New Zealand updates cap and price control settings: https://icapcarbonaction.com/en/news/new-zealand-updates-cap-and-price-control-settings.

[7]British Columbia's Carbon Tax: https://www2.gov.bc.ca/gov/content/environment/climate-change/clean-economy/carbon-tax.

[8]EU-ETS Emissions cap and allowances: https://climate.ec.europa.eu/eu-action/eu-emissions-trading-system-eu-ets/emissions-cap-and-allowances_en.

[9]ETS & Carbon Taxes: https://carbonpricingdashboard.worldbank.org/map_data.

[10]Carbon permits said to be in excess in Europe: https://www.nytimes.com/2010/04/01/business/energy-environment/01carbon.html.

[11]EU confirms carbon market permit surplus grew in 2020: https://www.reuters.com/business/energy/eu-confirms-carbon-market-permit-surplus-grew-2020-2021-05-12/.

[12]The impact of the recent spike in uncertainty on economic activity in the euro area: https://www.ecb.europa.eu/pub/economic-bulletin/html/eb202006.en.html.

[13]Ukraine brings significant economic uncertainty, Rishi Sunak warns: https://www.bbc.com/news/business-60699815.

[14]Britain strikes green investment partnership with Bill Gates: https://money.usnews.com/investing/news/articles/2021-10-19/britain-strikes-green-investment-partnership-with-bill-gates.

[15]COP26: Leaders agree global plan to boost green technology: https://www.bbc.com/news/science-environment-59138622.

[16]Western U.S., Canada go own way on carbon trading: https://www.reuters.com/article/climate-california-idUSN1120328820100312.

[17]China's national carbon market is about to launch: https://chinadialogue.net/en/climate/chinas-national-carbon-market-is-about-to-launch/.

[18]Carbon tax in South Africa: https://www.sars.gov.za/customs-and-excise/excise/environmental-levy-products/carbon-tax/.

[19]The California Air Resources Board: https://ww2.arb.ca.gov/our-work/programs/cap-and-trade-program/cost-containment-information/price-ceiling-information.

[20]The ICAP ETS Map: https://icapcarbonaction.com/en/ets-map.

References

Anand KS, Giraud-Carrier FC (2020) Pollution regulation of competitive markets. *Management Science* 66(9):4193–4206.

Arifoğlu K, Tang CS (2022) A two-sided incentive program for coordinating the influenza vaccine supply chain. *Manufacturing & Service Operations Management* 24(1):235–255.

Atasu A, Corbett CJ, Huang X, Toktay LB (2020) Sustainable operations management through the perspective of manufacturing & service operations management. *Manufacturing & Service Operations Management* 22(1):146–157.

Barnett AH (1980) The Pigouvian tax rule under monopoly. *American Economic Review* 70(5):1037–1041.

Bovenberg AL, De Mooij RA (1994) Environmental levies and distortionary taxation. *American Economic Review* 84(4):1085–1089.

Burkardt J (2014) The truncated normal distribution. *Department of Scientific Computing Website, Florida State University* 1:35.

Carmona R, Hinz J (2011) Risk-neutral models for emission allowance prices and option valuation. *Management Science* 57(8):1453–1468.

Chen P (1990) Prices vs. quantities and delegating price authority to a monopolist. *The Review of Economic Studies* 57(3):521–529.

Corbett CJ, Klassen RD (2006) Extending the horizons: Environmental excellence as key to improving operations. *Manufacturing & Service Operations Management* 8(1):5–22.

Dales JH (1968) *Pollution, Property & Prices* (University of Toronto Press, Toronto).

Deryugina T, Heutel G, Miller NH, Molitor D, Reif J (2019) The mortality and medical costs of air pollution: Evidence from changes in wind direction. *American Economic Review* 109(12):4178–4219.

Drake DF (2018) Carbon tariffs: Effects in settings with technology choice and foreign production cost advantage. *Manufacturing & Service Operations Management* 20(4):667–686.

Drake DF, Just RL (2016) Ignore, avoid, abandon, and embrace: What drives firm responses to environmental regulation? *Environmentally Responsible Supply Chains*, 199–222 (Springer).

Drake DF, Kleindorfer PR, Van Wassenhove LN (2016) Technology choice and capacity portfolios under emissions regulation. *Production and Operations Management* 25(6):1006–1025.

Eichner T, Runkel M (2012) Interjurisdictional spillovers, decentralized policy making, and the elasticity of capital supply. *American Economic Review* 102(5):2349–57.

Fan X, Chen K, Chen YJ (2023) Is price commitment a better solution to control carbon emissions and promote technology investment? *Management Science* 69(1):325–341.

Fankhauser S, Smith SM, Allen M, Axelsson K, Hale T, Hepburn C, Kendall JM, Khosla R, Lezaun J, Mitchell-Larson E, Obersteiner M, Rajamani L, Rickaby R, Seddon N, Wetzer T (2022) The meaning of net zero and how to get it right. *Nature Climate Change* 12(1):15–21.

Farbotko C, Lazrus H (2012) The first climate refugees? Contesting global narratives of climate change in Tuvalu. *Global Environmental Change* 22(2):382–390.

Fisher ML, Hammond JH, Obermeyer WR, Raman A (1994) Making supply meet demand in an uncertain world. *Harvard Business Review* 72:83–83.

Fu X, Chen YJ, Gallego G, Gao P, Lu M (2021) Temporal flexibility in emission permits regulation. *Available at SSRN 3900094* .

Gong X, Zhou SX (2013) Optimal production planning with emissions trading. *Operations Research* 61(4):908–924.

Gopalakrishnan S, Granot D, Granot F, Sošić G, Cui H (2021) Incentives and emission responsibility allocation in supply chains. *Management Science* 67(7):4172–4190.

Green JF (2017) Don't link carbon markets. *Nature* 543(7646):484–486.

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

Holland SP, Mansur ET, Muller NZ, Yates AJ (2016) Are there environmental benefits from driving electric vehicles? The importance of local factors. *American Economic Review* 106(12):3700–3729.

Ireland NJ (1977) Ideal prices vs. prices vs. quantities. *The Review of Economic Studies* 44(1):183–186.

Keegan KM, Albert MR, McConnell JR, Baker I (2014) Climate change and forest fires synergistically drive widespread melt events of the greenland ice sheet. *Proceedings of the National Academy of Sciences* 111(22):7964–7967.

King A, Lenox M (2002) Exploring the locus of profitable pollution reduction. *Management Science* 48(2):289–299.

Klassen RD, McLaughlin CP (1996) The impact of environmental management on firm performance. *Management Science* 42(8):1199–1214.

Krass D, Nedorezov T, Ovchinnikov A (2013) Environmental taxes and the choice of green technology. *Production and Operations Management* 22(5):1035–1055.

Laffont JJ (1977) More on prices vs. quantities. *The Review of Economic Studies* 44(1):177–182.

Montgomery WD (1972) Markets in licenses and efficient pollution control programs. *Journal of Economic Theory* 5(3):395–418.

Muller NZ, Mendelsohn R (2009) Efficient pollution regulation: Getting the prices right. *American Economic Review* 99(5):1714–39.

Nagurney A, Dhanda KK (2000) Marketable pollution permits in oligopolistic markets with transaction costs. *Operations Research* 48(3):424–435.

Nordhaus W (2019) Climate change: The ultimate challenge for economics. *American Economic Review* 109(6):1991–2014.

Peng Y, Gao F, Chen J (2021) Green packaging or greenwashing? Implications of bring-your-own-container. *Available at SSRN 3888378* .

Pigou AC (1920) *The Economics of Welfare* (Macmillan, London).

Porter ME (1991) America's Green Strategy. *Scientific American* 264(4):168.

Requate T (1993) Equivalence of effluent taxes and permits for environmental regulation of several local monopolies. *Economics Letters* 42(1):91–95.

Requate T (2006) Environmental policy under imperfect competition: A survey. Tietenberg T, Folmer H, eds., *The International Yearbook of Environmental and Resource Economics 2006/2007* (Edward Elgar, Cheltenham, UK), 120-207.

Ricke K, Drouet L, Caldeira K, Tavoni M (2018) Country-level social cost of carbon. *Nature Climate Change* 8(10):895–900.

Singh N, Vives X (1984) Price and quantity competition in a differentiated duopoly. *The RAND Journal of Economics* 546–554.

Subramanian R, Gupta S, Talbot B (2007) Compliance strategies under permits for emissions. *Production and Operations Management* 16(6):763–779.

Sunar N (2016) Emissions allocation problems in climate change policies. *Environmentally Responsible Supply Chains*, 261–282 (Springer).

Sunar N, Birge JR (2019) Strategic commitment to a production schedule with uncertain supply and demand: Renewable energy in day-ahead electricity markets. *Management Science* 65(2):714–734.

Sunar N, Plambeck E (2016) Allocating emissions among co-products: Implications for procurement and climate policy. *Manufacturing & Service Operations Management* 18(3):414–428.

Sunar N, Swaminathan JM (2021) Net-metered distributed renewable energy: A peril for utilities? *Management Science* 67(11):6716–6733.

Tang BJ, Wang XY, Wei YM (2019) Quantities versus prices for best social welfare in carbon reduction: A literature review. *Applied Energy* 233:554–564.

Tietenberg T, Lewis L (2011) *Environmental and Natural Resource Economics* (Pearson Addison Wesley, New York).

Weitzman ML (1974) Prices vs. quantities. *The Review of Economic Studies* 41(4):477–491.

Xue S, Zhang B, Zhao X (2021) Brain drain: The impact of air pollution on firm performance. *Journal of Environmental Economics and Management* 110:102546, ISSN 0095-0696.

Yu JJ, Tang CS, Shen ZJM (2018) Improving consumer welfare and manufacturer profit via government subsidy programs: Subsidizing consumers or manufacturers? *Manufacturing & Service Operations Management* 20(4):752–766.

Appendix A: Extension: Imperfect Market Information and Firm Competition

In this section, we extend the model in Section 4.1 to the case where firms only have access to partial information on the market size θ rather than the exact value. Specifically, firms will perceive a signal with probability ϕ_H (ϕ_L, resp.) in the high-demand (low-demand, resp.) case and $\phi_H > \phi_L$. Given the signal (no signal, resp.), the firm will infer the posterior probability of the high-demand case as $\frac{\phi_H}{\phi_H+\phi_L}$ ($\frac{1-\phi_H}{2-\phi_H-\phi_L}$, resp.), and the random demand is denoted by $\tilde{\theta}_1$ ($\tilde{\theta}_0$, resp.). Moreover, we have $\mathbb{E}[\tilde{\theta}_0] = \mu - \frac{\phi_H-\phi_L}{2-\phi_H-\phi_L}\sigma$ and $\mathbb{E}[\tilde{\theta}_1] = \mu + \frac{\phi_H-\phi_L}{\phi_H+\phi_L}\sigma$. The subscript '$mc$' is provided as a mnemonic for 'imperfect market information & firm competition'.

Price Instrument. We start with the price instrument. Similar to Section 4.1, given a tax rate t and a signal s, firm i will choose its production amount x_i and its abatement amount w_i to maximize its expected profit:

$$\max_{x_i,w_i:0\le w_i\le x_i} \mathbb{E}_{\tilde{\theta}_s}\left[(\tilde{\theta}_s - x_i - x_{-i})x_i - \frac{\gamma}{2}w_i^2 - t(x_i - w_i)^+\right].$$

Given the firm's optimal responses $x^P_{mc,i}(s,t)$ and $w^P_{mc,i}(s,t)$, the regulator determines the optimal tax rate t to maximize the expected social welfare:

$$\mathcal{S}^P_{mc} = \max_{t\ge 0} \mathbb{E}^{p(t)}_{\theta,s}\Bigg[\big(\theta - \sum_{i=1}^N x^P_{mc,i}(s,t)\big)\big(\sum_{i=1}^N x^P_{mc,i}(s,t)\big) - \sum_{i=1}^N \frac{\gamma}{2}\big(w^P_{mc,i}(s,t)\big)^2 + \frac{1}{2}\big(\sum_{i=1}^N x^P_{mc,i}(s,t)\big)^2 - d\big[\sum_{i=1}^N \big(x^P_{mc,i}(s,t) - w^P_{mc,i}(s,t)\big)^+\big]^2\Bigg].$$

Then, we summarize the results in the following lemma.

Lemma A.1 **(Optimal Price Instrument with Imperfect Information and Firm Competition)**. *The optimal price instruments with imperfect information and firm competition are as follows.*

- *When $d \le \delta^P_{mc}$, the optimal tax rate is $t^P_{mc} = 0$.*
- *When $d > \delta^P_{mc}$ and $\sigma/\mu \le \Omega^P_{mc}$, the optimal tax rate is $t^P_{mc} = \mathcal{C}^P_{mc,1}\mu$.*
- *When $d > \delta^P_{mc}$ and $\sigma/\mu > \Omega^P_{mc}$, the optimal tax rate is $t^P_{mc} = \mathcal{C}^P_{mc,1}\big(\mu + \frac{\phi_H-\phi_L}{\phi_H+\phi_L}\sigma\big)$.*

Here, δ^P_{mc}, Ω^P_{mc}, $\mathcal{C}^P_{mc,1}$ are non-negative constants depending only on d, γ, N, ϕ_H and ϕ_L, and their detailed forms are provided in the proof.

Quantity Instrument. Next, we discuss the quantity instrument. Similar to Section 4.1, given a cap k and a signal s, firm i will choose its production amount x_i and its abatement amount w_i to maximize its expected profit:

$$\max_{x_i,w_i:0\le x_i-w_i\le k, w_i\ge 0} \mathbb{E}_{\tilde{\theta}_s}\left[(\tilde{\theta}_s - x_i - x_{-i})x_i - \frac{\gamma}{2}w_i^2\right].$$

Given the firm's optimal responses $x^Q_{mc,i}(s,k)$ and $w^Q_{mc,i}(s,k)$, the regulator determines the optimal cap k to maximize the expected social welfare:

$$\mathcal{S}^Q_{mc} = \max_{k\ge 0} \mathbb{E}^{q(k)}_{\theta,s}\Bigg[\big(\theta - \sum_{i=1}^N x^Q_{mc,i}(s,t)\big)\big(\sum_{i=1}^N x^Q_{mc,i}(s,t)\big) - \sum_{i=1}^N \frac{\gamma}{2}\big(w^Q_{mc,i}(s,t)\big)^2 + \frac{1}{2}\big(\sum_{i=1}^N x^Q_{mc,i}(s,t)\big)^2 - d\big[\sum_{i=1}^N \big(x^Q_{mc,i}(s,t) - w^Q_{mc,i}(s,t)\big)^+\big]^2\Bigg].$$

Then, we summarize the results in the following lemma.

LEMMA A.2 **(Optimal Quantity Instrument with Imperfect Information and Firm Competition)**. *The optimal quantity instruments with imperfect information and firm competition are as follows.*

- *When* $d \le \delta^Q_{mc}$*, the optimal cap is* $k^Q_{mc} \ge \frac{\mu}{2} + \frac{\phi_H - \phi_L}{(N+1)(\phi_H+\phi_L)}\sigma$.
- *When* $d > \delta^Q_{mc}$ *and* $\sigma/\mu \le \Omega^Q_{mc}$*, the optimal cap is* $k^Q_{mc} = \mathcal{C}^Q_{mc,1}\mu$.
- *When* $d > \delta^Q_{mc}$ *and* $\sigma/\mu > \Omega^Q_{mc}$*, the optimal cap is* $k^Q_{mc} = \mathcal{C}^Q_{mc,1}\big(\mu + \frac{\phi_H-\phi_L}{\phi_H+\phi_L}\sigma\big)$.

Here, δ^Q_{mc}*,* Ω^Q_{mc} *and* $\mathcal{C}^Q_{mc,1}$ *are non-negative constants depending only on* d*,* γ*,* N*,* ϕ_H *and* ϕ_L*, and their detailed forms are provided in the proof.*

Instrument Comparison. Lastly, we compare these two instruments with imperfect market information and firm competition. The result is summarized in the following theorem.

PROPOSITION A.1 **(Instrument Comparison with Imperfect Information and Firm Competition)**. *Define* $\delta_{mc} = \frac{\gamma}{2N(N+\gamma+1)}$, $\Delta_{mc} = \frac{(N+2)\gamma^2+(N+1)(N+3)\gamma}{2N(N+\gamma+1)^2}$ *and* $\Omega_{mc} = \frac{\sqrt{2(\phi_H+\phi_L)}-(\phi_H+\phi_L)}{\phi_H-\phi_L}$*. If the firm will receive a signal with probability* ϕ_H *(*ϕ_L*, resp.) in the high-demand (low-demand, resp.) case, then the comparison of the optimal price and quantity instruments is as follows.*

- *When* $d \in [0, \delta_{mc}]$*, neither instruments should be implemented.*
- *When* $d \in (\delta_{mc}, \Delta_{mc})$*, the expected social welfare under the optimal price instrument is higher (lower, resp.) than that under the optimal quantity instrument when* $\sigma/\mu < \Omega_{mc}$ *(*$\sigma/\mu > \Omega_{mc}$*, resp.).*
- *When* $d \in (\Delta_{mc}, \infty)$*, the expected social welfare under the optimal quantity instrument is higher (lower, resp.) than that under the optimal price instrument when* $\sigma/\mu < \Omega_{mc}$ *(*$\sigma/\mu > \Omega_{mc}$*, resp.).*

According to Lemmas A.1 and A.2 and Proposition A.1, the results in the base model still qualitatively hold. The threshold of the market uncertainty, Ω_{mc}, decreases in the probability ϕ_H but increases in the probability ϕ_L. As we discuss in Section 3.4, when the gap between the demand cases increases, the abatement flexibility gradually becomes the mainly concerned flexibility. Given a more informational signal (i.e., ϕ_H increases or ϕ_L decreases), the difference between the posterior expectations of the two demand cases increases, which amplifies the gap of the demand cases and hence results in an earlier switch of the concerned flexibility.

Managerial Implications. The study in this subsection complements the analysis of the base model by incorporating imperfect market information. The results provide insights into a regulated market when firms cannot perfectly perceive the market condition. Our analysis suggests that firms' ability to forecast market conditions plays an important role in the government's choice of regulatory instrument. Since demand forecasting plays a crucial role in each firm's operations, managers invest substantial effort and embrace cutting-edge information technologies to advance their forecast systems.[21,22] Our results imply that for the low-emitting (high-emitting, resp.) industry, the preferred region of the price (quantity, resp.) instrument shrinks as the forecast precision increases (i.e., ϕ_H increases and ϕ_L decreases).

Appendix B: Exploration on Heterogeneous Firms

In the following, we follow the settings in Section 4.3 and further investigate two important issues. First, we study when it is worthwhile for the regulator to pay the administrative cost to impose differentiated tax rates or caps. Second, we explore an alternative instrument that sets a cap on the pollution damage, i.e., $d_i(x_i - w_i)^2 \le \Gamma$. The subscript '$h$' is provided as a mnemonic for 'Heterogeneous'.

B.1. Value of Differentiated Regulation

In the following, we investigate when differentiated tax rates or caps can significantly improve social welfare than a uniform instrument. As mentioned in Section 4.3, the regulator in practice knows the abatement coefficients γ_1 and γ_2 but cannot distinguish these two firms, and hence impose a uniform tax rate or cap. Such a case is called the *indistinguishable* case and is denoted by the subscript 'i'. We also study an alternative case when the government pays the administrative cost to investigate firms' specific characteristics and hence can distinguish these two firms and impose differentiated tax rates or caps. Such a case is called the *distinguishable* case and is denoted by the subscript 'd'. In the following, we start with the price instrument.

Price Instrument. Given a tax rate t_i, firm i determines its production amount x_i and abatement amount w_i to maximize its profit:

$$\max_{x_i, w_i: 0 \le w_i \le x_i} \left\{ (\theta - x_i - x_{-i}) x_i - \frac{\gamma_i}{2} w_i^2 - t_i (x_i - w_i)^+ \right\},$$

from which we can deduce the firms' optimal responses as follows:

$$x^P_{v,i}(\theta, t_1, t_2) = \begin{cases} \frac{\theta - 2t_i + t_{-i}}{3} & t_i < \frac{\gamma_i(\theta + t_{-i})}{2\gamma_i + 3} \,\&\, t_{-i} < \frac{\gamma_{-i}(\theta + t_i)}{2\gamma_{-i} + 3} \\ \frac{(\gamma_{-i}+1)\theta - (\gamma_{-i}+2)t_i}{2\gamma_{-i}+3} & t_i < \frac{\gamma_i(\gamma_{-i}+1)\theta}{\gamma_i\gamma_{-i} + 2\gamma_i + 2\gamma_{-i} + 3} \,\&\, t_{-i} \ge \frac{\gamma_{-i}(\theta + t_i)}{2\gamma_{-i}+3} \\ \frac{\theta + t_{-i}}{2\gamma_i + 3} & t_i \ge \frac{\gamma_i(\theta + t_{-i})}{2\gamma_i + 3} \,\&\, t_{-i} < \frac{(\gamma_i + 1)\gamma_{-i}\theta}{\gamma_i\gamma_{-i} + 2\gamma_i + 2\gamma_{-i} + 3} \\ \frac{(\gamma_{-i}+1)\theta}{\gamma_i\gamma_{-i} + 2\gamma_i + 2\gamma_{-i} + 3} & t_i \ge \frac{\gamma_i(\gamma_{-i}+1)\theta}{\gamma_i\gamma_{-i} + 2\gamma_i + 2\gamma_{-i} + 3} \,\&\, t_{-i} \ge \frac{(\gamma_i+1)\gamma_{-i}\theta}{\gamma_i\gamma_{-i} + 2\gamma_i + 2\gamma_{-i} + 3} \end{cases}$$

$$w^P_{v,i}(\theta, t_1, t_2) = \begin{cases} \frac{t_i}{\gamma_i} & t_i < \frac{\gamma_i(\theta + t_{-i})}{2\gamma_i + 3} \,\&\, t_{-i} < \frac{\gamma_{-i}(\theta + t_i)}{2\gamma_{-i} + 3} \\ \frac{t_i}{\gamma_i} & t_i < \frac{\gamma_i(\gamma_{-i}+1)\theta}{\gamma_i\gamma_{-i} + 2\gamma_i + 2\gamma_{-i} + 3} \,\&\, t_{-i} \ge \frac{\gamma_{-i}(\theta + t_i)}{2\gamma_{-i}+3} \\ \frac{\theta + t_{-i}}{2\gamma_i + 3} & t_i \ge \frac{\gamma_i(\theta + t_{-i})}{2\gamma_i + 3} \,\&\, t_{-i} < \frac{(\gamma_i + 1)\gamma_{-i}\theta}{\gamma_i\gamma_{-i} + 2\gamma_i + 2\gamma_{-i} + 3} \\ \frac{(\gamma_{-i}+1)\theta}{\gamma_i\gamma_{-i} + 2\gamma_i + 2\gamma_{-i} + 3} & t_i \ge \frac{\gamma_i(\gamma_{-i}+1)\theta}{\gamma_i\gamma_{-i} + 2\gamma_i + 2\gamma_{-i} + 3} \,\&\, t_{-i} \ge \frac{(\gamma_i+1)\gamma_{-i}\theta}{\gamma_i\gamma_{-i} + 2\gamma_i + 2\gamma_{-i} + 3} \end{cases}.$$

In the *distinguishable case*, the regulator determines different tax rates according to the firms' abatement coefficients to maximize the expected social welfare

$$\begin{aligned} \mathcal{S}^P_{hd} = \max_{t_1, t_2 \ge 0} \mathbb{E}\Bigg[& \sum_{i=1}^2 \big(\theta - \sum_{i=1}^2 x^P_{v,i}(\theta, t_1, t_2)\big) x^P_{v,i}(\theta, t_1, t_2) - \sum_{i=1}^2 \frac{\gamma_i}{2} \big(w^P_{v,i}(\theta, t_1, t_2)\big)^2 + \frac{1}{2}\big(\sum_{i=1}^2 x^P_{v,i}(\theta, t_1, t_2)\big)^2 \\ & - \beta \sum_{i=1}^2 d_i \big[(x^P_{v,i}(\theta, t_1, t_2) - w^P_{v,i}(\theta, t_1, t_2))^+\big]^2 - (1-\beta) d \big[\sum_{i=1}^2 \big(x^P_{v,i}(\theta, t_1, t_2) - w^P_{v,i}(\theta, t_1, t_2)\big)^+\big]^2 \Bigg]. \end{aligned}$$

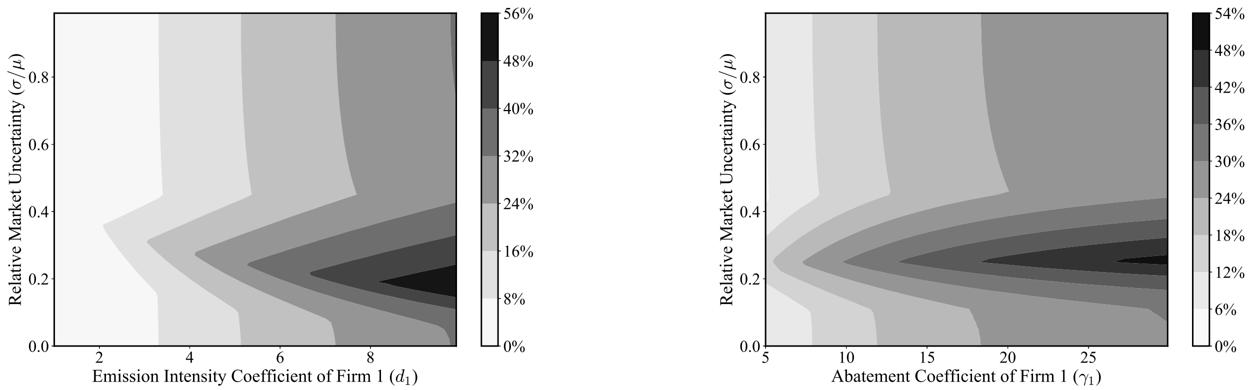
In contrast, in the *indistinguishable case*, the regulator sets the same tax rate for the firms, i.e., $t_1 = t_2$, and the expected social welfare is denoted by $\mathcal{S}^P_{hi}$. Then, we compare the price instrument in the distinguishable case with that in the indistinguishable case, and illustrate the numerical improvement ratio in Figure 1.

Quantity Instrument. Given a cap k_i, firm i determines its production amount x_i and abatement amount w_i to maximize its profit:

$$\max_{x_i, w_i: 0 \le x_i - w_i \le k_i; w_i \ge 0} \left\{ (\theta - x_i - x_{-i}) x_i - \frac{\gamma_i}{2} w_i^2 \right\},$$

from which we can deduce the firms' optimal responses as follows:

$$x^Q_{v,i}(\theta, k_1, k_2) = \begin{cases} \frac{(\gamma_{-i}+1)\theta + \gamma_i(\gamma_{-i}+2)k_i - \gamma_{-i}k_{-i}}{\gamma_i\gamma_{-i} + 2\gamma_i + 2\gamma_{-i} + 3} & k_i < \frac{(\gamma_{-i}+1)\theta - \gamma_{-i}k_{-i}}{2\gamma_{-i}+3} \,\&\, k_{-i} < \frac{(\gamma_i+1)\theta - \gamma_i k_i}{2\gamma_i + 3} \\ \frac{\theta + 2\gamma_i k_i}{2\gamma_i + 3} & k_i < \frac{\theta}{3} \,\&\, k_{-i} \ge \frac{(\gamma_i+1)\theta - \gamma_i k_i}{2\gamma_i+3} \\ \frac{(\gamma_{-i}+1)\theta - \gamma_{-i}k_{-i}}{2\gamma_{-i}+3} & k_i \ge \frac{(\gamma_{-i}+1)\theta - \gamma_{-i}k_{-i}}{2\gamma_{-i}+3} \,\&\, k_{-i} < \frac{\theta}{3} \\ \frac{\theta}{3} & k_i \ge \frac{\theta}{3} \,\&\, k_{-i} \ge \frac{\theta}{3} \end{cases}$$



(a) d_1 (b) γ_1

Figure 1 **Social welfare improvement ratio of the price instrument under the distinguishable case compared to that under the indistinguishable case** $\frac{\mathcal{S}^P_{hd}-\mathcal{S}^P_{hi}}{\mathcal{S}^P_{hi}}$ **with** $d_1=5$, $d_2=1$, $d=2$, $\beta=0.8$, $\mu=1$, $\gamma_1=10$ **and** $\gamma_2=15$.

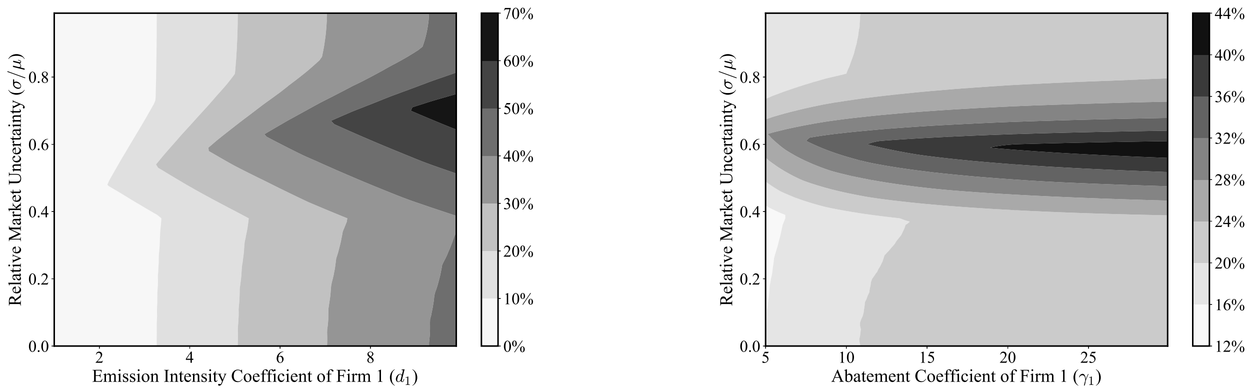
$$w^Q_{v,i}(\theta,k_1,k_2)=\begin{cases}\frac{(\gamma_{-i}+1)\theta-(2\gamma_{-i}+3)k_i-\gamma_{-i}k_{-i}}{\gamma_i\gamma_{-i}+2\gamma_i+2\gamma_{-i}+3} & k_i<\frac{(\gamma_{-i}+1)\theta-\gamma_{-i}k_{-i}}{2\gamma_{-i}+3}\ \&\ k_{-i}<\frac{(\gamma_i+1)\theta-\gamma_ik_i}{2\gamma_i+3}\\ \frac{\theta-3k_i}{2\gamma_i+3} & k_i<\frac{\theta}{3}\ \&\ k_{-i}\geq\frac{(\gamma_i+1)\theta-\gamma_ik_i}{2\gamma_i+3}\\ 0 & k_i\geq\frac{(\gamma_{-i}+1)\theta-\gamma_{-i}k_{-i}}{2\gamma_{-i}+3}\ \&\ k_{-i}<\frac{\theta}{3}\\ 0 & k_i\geq\frac{\theta}{3}\ \&\ k_{-i}\geq\frac{\theta}{3}\end{cases}.$$

Similarly, in the *distinguishable case*, the regulator determines different caps according to the firms' abatement coefficients to maximize the expected social welfare

$$\begin{aligned}\mathcal{S}^Q_{hd}=\max_{k_1,k_2\geq 0}\mathbb{E}\Bigg[&\sum_{i=1}^2\Big(\theta-\sum_{i=1}^2 x^Q_{v,i}(\theta,k_1,k_2)\Big)x^Q_{v,i}(\theta,k_1,k_2)-\sum_{i=1}^2\frac{\gamma_i}{2}\big(w^Q_{v,i}(\theta,k_1,k_2)\big)^2+\frac{1}{2}\Big(\sum_{i=1}^2 x^Q_{v,i}(\theta,k_1,k_2)\Big)^2\\ &-\beta\sum_{i=1}^2 d_i\big[(x^Q_{v,i}(\theta,k_1,k_2)-w^Q_{v,i}(\theta,k_1,k_2))^+\big]^2-(1-\beta)d\Big[\sum_{i=1}^2\big(x^Q_{v,i}(\theta,k_1,k_2)-w^Q_{v,i}(\theta,k_1,k_2)\big)^+\Big]^2\Bigg].\end{aligned}$$

In the *indistinguishable case*, the regulator sets the same cap for the firms, i.e., $k_1=k_2$, and the expected social welfare is denoted by $\mathcal{S}^Q_{hi}$. Then, we compare the quantity instrument in the distinguishable case with that in the indistinguishable case, and illustrate the numerical improvement ratio in Figure 2.

According to Figures 1 and 2, under both instruments, the value of differentiation is significant when the firms' heterogeneity ($|d_1-d_2|$ or $|\gamma_1-\gamma_2|$) is large, which is intuitive. Moreover, the value of differentiation under the price (quantity, resp.) instrument is significant when the market uncertainty σ is relatively small (large, resp.), which is similar to results in Section 4.2. Intuitively, as the problem of the indistinguishable case becomes more challenging, the value of differentiation is amplified. Similar to the discussion in Section 3.3, as the market uncertainty increases, the regulation problem in the inclusive phase becomes more challenging because the gap between demand cases increases. In contrast, as the market uncertainty increases, the regulation problem in the exclusive phase becomes less challenging because the low-demand case becomes more minuscule. Therefore, the value of differentiation is significant at the switch point: low uncertainty for the price instrument and high uncertainty for the quantity instrument.



(a) d_1 (b) γ_1

Figure 2 Social welfare improvement ratio of the quantity instrument under the distinguishable case compared to that under the indistinguishable case $\frac{\mathcal{S}^Q_{hd}-\mathcal{S}^Q_{hi}}{\mathcal{S}^Q_{hi}}$ with $d_1=5$, $d_2=1$, $d=2$, $\beta=0.8$, $\mu=1$, $\gamma_1=10$ **and** $\gamma_2=15$.

B.2. Regulation on Pollution Damage

In the following, we explore an alternative regulatory instrument restricting the pollution damage. Specifically, under such an instrument, the regulator sets a cap on the pollution damage, $d_i\cdot(x_i-w_i)^2\le\Gamma$. The superscript '$D$' is provided as a mnemonic for 'Pollution Damage'. For the case with one monopolistic firm, such an instrument ($d_i(x_i-w_i)^2\le\Gamma$) is equivalent to a quantity instrument on pollution amount ($x_i-w_i\le\sqrt{\Gamma/d}$). In this case, given a pollution damage cap Γ, each firm i chooses its production amount x_i and abatement amount w_i to maximize its profit:

$$\max_{x_i,w_i:x_i,w_i\ge 0;0\le x_i-w_i\le\sqrt{\Gamma/d_i}}\Big\{(\theta-x_i-x_{-i})x_i-\frac{\gamma_i}{2}w_i^2\Big\}.$$

Then, we can deduce the equilibrium production and abatement amounts as follows:

$$x_1^D(\theta,\Gamma)=\begin{cases}\frac{(\gamma_2+1)\theta+\gamma_1(\gamma_2+2)\sqrt{\Gamma/d_1}-\gamma_2\sqrt{\Gamma/d_2}}{\gamma_1\gamma_2+2\gamma_1+2\gamma_2+3} & \Gamma\in[0,\tilde\Gamma)\\ \frac{\theta+2\gamma_1\sqrt{\Gamma/d_1}}{2\gamma_1+3} & \Gamma\in[\tilde\Gamma,\frac{d_1\theta^2}{9})\\ \frac{\theta}{3} & \Gamma\in[\frac{d_1\theta^2}{9},\infty)\end{cases}\qquad w_1^D(\theta,\Gamma)=\begin{cases}\frac{(\gamma_2+1)\theta-(2\gamma_2+3)\sqrt{\Gamma/d_1}-\gamma_2\sqrt{\Gamma/d_2}}{\gamma_1\gamma_2+2\gamma_1+2\gamma_2+3} & \Gamma\in[0,\tilde\Gamma)\\ \frac{\theta-3\sqrt{\Gamma/d_1}}{2\gamma_1+3} & \Gamma\in[\tilde\Gamma,\frac{d_1\theta^2}{9})\\ 0 & \Gamma\in[\frac{d_1\theta^2}{9},\infty)\end{cases}$$

$$x_2^D(\theta,\Gamma)=\begin{cases}\frac{(\gamma_1+1)\theta+\gamma_2(\gamma_1+2)\sqrt{\Gamma/d_2}-\gamma_1\sqrt{\Gamma/d_1}}{\gamma_1\gamma_2+2\gamma_1+2\gamma_2+3} & \Gamma\in[0,\tilde\Gamma)\\ \frac{\theta-\gamma_1\sqrt{\Gamma/d_1}+\gamma_1\theta}{2\gamma_1+3} & \Gamma\in[\tilde\Gamma,\frac{d_1\theta^2}{9})\\ \frac{\theta}{3} & \Gamma\in[\frac{d_1\theta^2}{9},\infty)\end{cases}\qquad w_2^D(\theta,\Gamma)=\begin{cases}\frac{(\gamma_1+1)\theta-(2\gamma_1+3)\sqrt{\Gamma/d_2}-\gamma_1\sqrt{\Gamma/d_1}}{\gamma_1\gamma_2+2\gamma_1+2\gamma_2+3} & \Gamma\in[0,\tilde\Gamma)\\ 0 & \Gamma\in[\tilde\Gamma,\frac{d_1\theta^2}{9})\\ 0 & \Gamma\in[\frac{d_1\theta^2}{9},\infty)\end{cases},$$

where $\tilde\Gamma=\frac{d_1d_2(\gamma_1+1)^2\theta^2}{(\sqrt{d_1}(2\gamma_1+3)+\sqrt{d_2}\gamma_1)^2}$. The regulator determines the optimal pollution damage cap Γ to maximize the expected social welfare:

$$\mathcal{S}^D_{hi}=\max_{\Gamma\ge0}\mathbb{E}_\theta\Bigg[\sum_{i=1}^2\big(\theta-\sum_{i=1}^2x_i^D(\theta,\Gamma)\big)x_i^D(\theta,\Gamma)+\frac{1}{2}\big(\sum_{i=1}^2x_i^D(\theta,\Gamma)\big)^2-\sum_{i=1}^2\frac{\gamma_i}{2}w_i^D(\theta,\Gamma)^2$$
$$-\beta\sum_{i=1}^2d_i\big[(x_i^D(\theta,\Gamma)-w_i^D(\theta,\Gamma))^+\big]^2-(1-\beta)d\big[\sum_{i=1}^2\big(x_i^D(\theta,\Gamma)-w_i^D(\theta,\Gamma)\big)^+\big]^2\Bigg].$$

Then, we compare the quantity instruments by numerical experiments and illustrate the improvement ratio in Figure 3. According to Figure 3, the quantity instrument on pollution damage can significantly improve social welfare compared to the quantity instrument on pollution amount, especially when the firms'

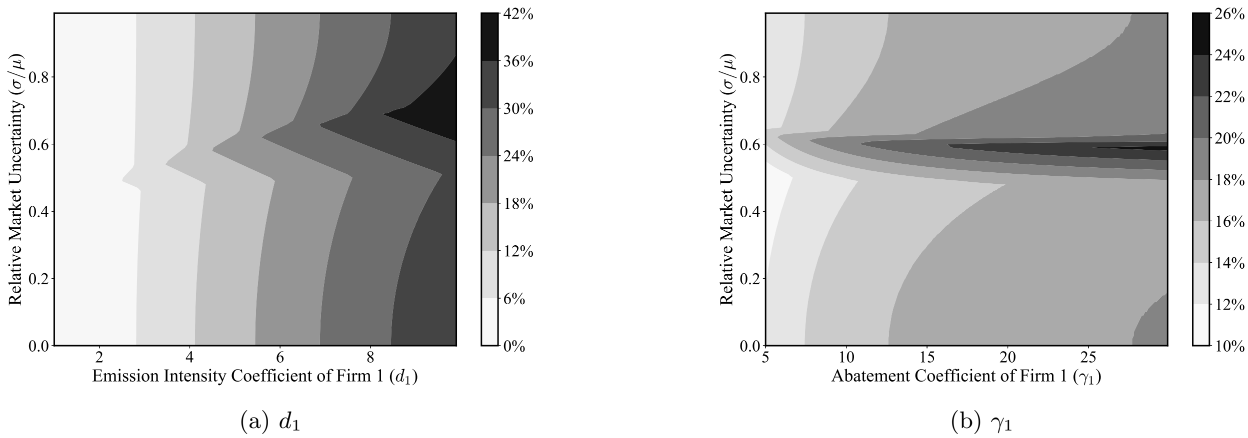


Figure 3 **Social welfare improvement ratio of the instrument on pollution damage compared to quantity instrument** $\frac{\mathcal{S}_{hi}^{D}-\mathcal{S}_{hi}^{Q}}{\mathcal{S}_{hi}^{Q}}$ **with** $d_1=5$**,** $d_2=1$**,** $d=2$**,** $\beta=0.8$**,** $\mu=1$**,** $\gamma_1=10$ **and** $\gamma_2=15$**.**

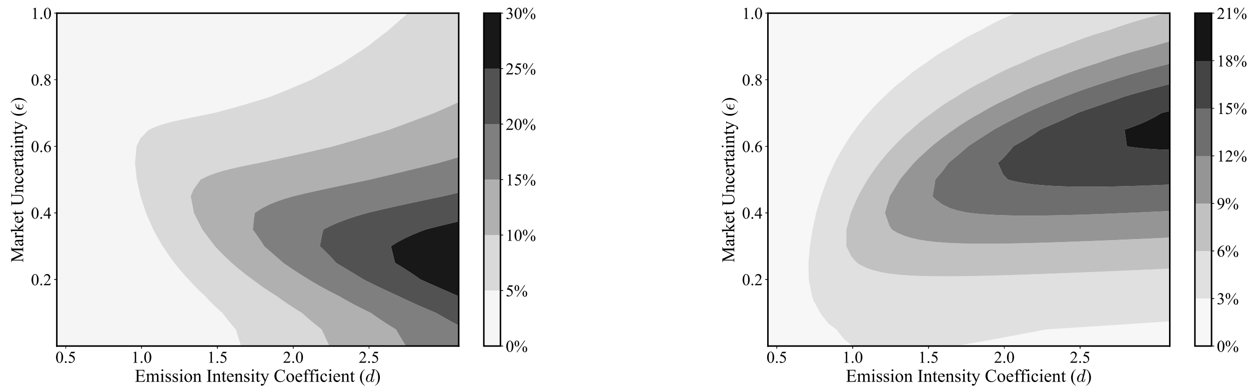
heterogeneity ($|d_1-d_2|$ or $|\gamma_1-\gamma_2|$) is large and the market uncertainty σ is large. Compared with the regulatory instrument on pollution amount, such an instrument imposes a large cap for low-emitting firms and a small cap for low-emitting firms, which is similar to the distinguishable case.

Managerial Insights. First, our results suggest the government investigates each firm's characteristics (e.g., abatement technology) such that it can implement a differentiated instrument, especially when the firms' heterogeneity is large. Moreover, under the price (quantity, resp.) instrument, the value of a differentiated instrument is significant when the market uncertainty is small (large, resp.). Second, if the administrative cost of investigation and implementation is unacceptable or a differentiated instrument is illegal, an alternative uniform regulatory instrument restricting the pollution damage can still improve social welfare significantly.

Appendix C: Simulation Study

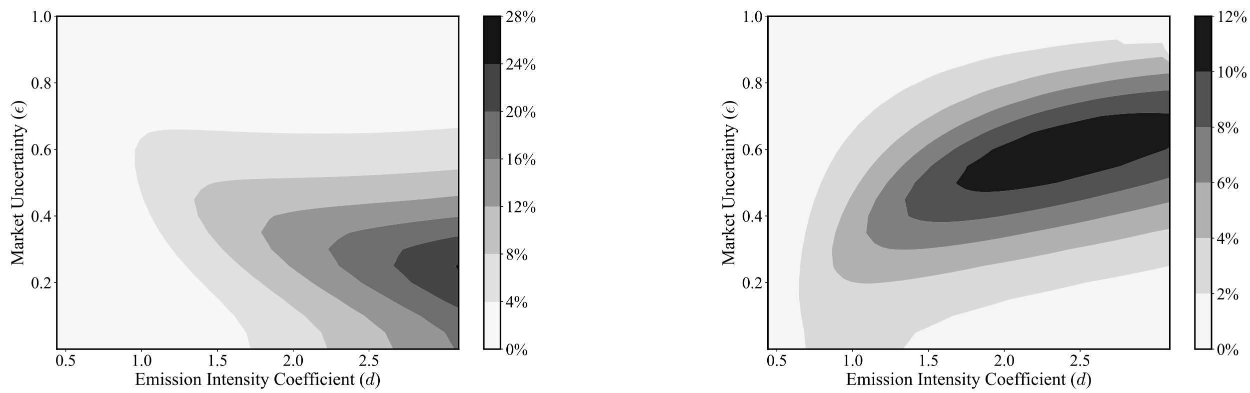
In the following, we will use a simulation to further investigate the performance of the quick-response case and the hybrid instrument in Section 4.2. We consider the case when the market size is drawn from a bimodal distribution.[23] Let $\Phi(\mu,\sigma,a,b;\cdot)$ denote the PDF of a truncated normal distribution specified by the "parent" mean μ and the "parent" standard deviation σ and the truncation range (a,b) (see Burkardt 2014). The distribution of the market size θ is a superposition of two truncated normal distributions. Specifically, the random market size is $D=\frac{1}{2}D_1+\frac{1}{2}D_2$, where $D_1\sim\Phi(\mu-\epsilon,\sigma,0,\infty;\cdot)$ and $D_2\sim\Phi(\mu+\epsilon,\sigma,0,\infty;\cdot)$. Similar to Fisher et al. (1994) and Fan et al. (2023), we choose appropriate parameters $\mu=1$, $\sigma=0.2$ and $\gamma=15$. Then, we illustrate the improvement ratios of the quick-response case and the hybrid instrument over the base model in Figures 4 and 5, respectively.

As Figures 4 and 5 illustrate, although the hybrid instrument cannot achieve the same expected social welfare as the quick-response case, the hybrid instrument can still significantly improve social welfare over pure regulatory instruments. Moreover, the results in Figure 7 still qualitatively hold: for the price (quantity, resp.) instrument, the improvement is significant when the emission intensity is large and the market uncertainty is small (large, resp.).



(a) $\frac{\mathcal{S}^R - \mathcal{S}^P}{\mathcal{S}^P}$ (b) $\frac{\mathcal{S}^R - \mathcal{S}^Q}{\mathcal{S}^Q}$

Figure 4 **Social welfare improvement ratio of the quick-response case compared to the base case by simulation as a function of** $d \in [\frac{\gamma}{2(\gamma+2)}, 3]$ **and** $\epsilon \in [0,1)$.



(a) $\frac{\mathcal{S}^R - \mathcal{S}^P}{\mathcal{S}^P}$ (b) $\frac{\mathcal{S}^R - \mathcal{S}^Q}{\mathcal{S}^Q}$

Figure 5 **Social welfare improvement ratio of the hybrid instrument compared to the base case by simulation as a function of** $d \in [\frac{\gamma}{2(\gamma+2)}, 3]$ **and** $\epsilon \in [0,1)$.

Appendix D: Omitted Proofs

For the sake of clarity, we omit both subscripts and superscripts in proofs.

D.1. Solution to Problem (2)

Given a tax rate t, the firm chooses its production amount x and abatement amount w to maximize its profit:

$$\max_{x,w:x\geq w\geq 0}\left\{\theta x - x^2 - \frac{\gamma}{2}w^2 - t(x-w)\right\}.$$

We first relax the constraint $w \geq 0$ and consider a larger feasible set. Then we will check the feasibility of the optimal solution. Therefore, we can write the Lagrangian as

$$\mathcal{L} = (\theta - t)x - x^2 - \frac{\gamma}{2}w^2 + tw + \lambda(x-w).$$

According to the KKT conditions, we have $\lambda \geq 0$ and

$$\frac{\partial \mathcal{L}}{\partial x} = \theta - t - 2x + \lambda = 0; \quad \frac{\partial \mathcal{L}}{\partial w} = -\gamma w + t - \lambda + \eta = 0; \quad \lambda(x-w) = 0. \tag{D.1}$$

Then we consider two candidate cases:

1. When $\lambda > 0$, since $\lambda(x-w)=0$, we have $x=w$. Solving (D.1) with $x=w$, we have $x_1^* = w_1^* = \frac{\theta}{\gamma+2}$ and $\lambda = t - \frac{\lambda\theta}{\gamma+2} > 0$. The corresponding Lagrangian is $\mathcal{L}_1^* = \frac{\theta^2}{2(\gamma+2)}$.
2. When $\lambda = 0$, solving (D.1), we have $x_2^* = \frac{\theta-t}{2}$ and $w_2^* = \frac{t}{\gamma}$. The corresponding Lagrangian is $\mathcal{L}_2^* = \frac{(\theta-t)^2}{4} + \frac{t^2}{2\gamma}$.

When $t \le \frac{\lambda\theta}{\gamma+2}$, we only have one possible candidate solution (x_2^*, w_2^*) and hence it is optimal. When $t > \frac{\lambda\theta}{\gamma+2}$, both solutions above are possible candidates. Since $\mathcal{L}_1^* \ge \mathcal{L}_2^*$ when $t > \frac{\lambda\theta}{\gamma+2}$, the optimal solution should be (x_1^*, w_1^*). The optimal solutions of the larger feasible set satisfy the constraint $w \ge 0$ and hence are optimal solutions for this problem. To summarize, we have the firm's optimal decision as follows:

$$x^*(\theta,t) = \begin{cases} \frac{\theta-t}{2} & t \in [0, \frac{\gamma\theta}{\gamma+2}) \\ \frac{\theta}{\gamma+2} & t \in [\frac{\gamma\theta}{\gamma+2}, \infty). \end{cases} \qquad w^*(\theta,t) = \begin{cases} \frac{t}{\gamma} & t \in [0, \frac{\gamma\theta}{\gamma+2}) \\ \frac{\theta}{\gamma+2} & t \in [\frac{\gamma\theta}{\gamma+2}, \infty). \end{cases} \qquad \square$$

D.2. Proof of Proposition 1

Having the firm's response in Section D.1, the regulator's optimization problem is as follows:

$$\mathcal{S}^* = \max_{t\ge 0} \mathbb{E}_\theta \left[\theta x^*(\theta,t) - (x^*(\theta,t))^2 - \frac{\gamma}{2}\big(w^*(\theta,t)\big)^2 + \frac{1}{2}(x^*(\theta,t))^2 - d\big(x^*(\theta,t) - w^*(\theta,t)\big)^2 \right]. \tag{D.2}$$

We first divide the feasible space into three subspaces and compute the optimal solution in each subspace. Then, we compare the optimal solutions of these subspaces and derive the global optimal solution. We start with the computation of the optimal solution in each subspace.

1. When $t \in [0, \frac{\gamma}{\gamma+2}(\mu-\sigma))$, the firm's production and abatement amounts of the low-demand case and the high-demand case are $(x_L, w_L) = (\frac{\mu-\sigma-t}{2}, \frac{t}{\gamma})$ and $(x_H, w_H) = (\frac{\mu+\sigma-t}{2}, \frac{t}{\gamma})$, respectively. Then the expected social welfare in (D.2) becomes

$$\mathcal{S}_1 = -\frac{2d(\gamma+2)^2 + \gamma(\gamma+4)}{8\gamma^2} t^2 + \frac{4d + (2d-1)\gamma}{4\gamma}\mu t + \frac{3-2d}{8}(\mu^2+\sigma^2).$$

The unconstrained optimal solution that maximizes $\mathcal{S}_1$ is $t^* = \frac{2d\gamma(\gamma+2)-\gamma^2}{2d(\gamma+2)^2+\gamma(\gamma+4)}\mu$. Since it is a quadratic function and may not be monotonic in this subspace, we discuss different situations. If $d < \frac{\gamma}{2(\gamma+2)}$, then we have $t^* < 0$, and the optimal solution and the corresponding objective are

$$t_1^* = 0; \qquad \mathcal{S}_1^* = \frac{3-2d}{8}(\mu^2+\sigma^2).$$

If $d \ge \frac{\gamma}{2(\gamma+2)}$ and $\sigma < \frac{2\gamma(\gamma+3)}{2d(\gamma+2)^2+\gamma(\gamma+4)}\mu$, then we have $0 \le t^* \le \frac{\gamma}{\gamma+2}(\mu-\sigma)$, and the optimal solution and the corresponding objective are

$$t_1^* = t^* = \frac{2d\gamma(\gamma+2)-\gamma^2}{2d(\gamma+2)^2+\gamma(\gamma+4)}\mu; \qquad \mathcal{S}_1^* = \frac{(\gamma+3)(\gamma+2d)}{4d(\gamma+2)^2+2\gamma(\gamma+4)}\mu^2 + \frac{3-2d}{8}\sigma^2.$$

If $d \ge \frac{\gamma}{2(\gamma+2)}$ and $\sigma \ge \frac{2\gamma(\gamma+3)}{2d(\gamma+2)^2+\gamma(\gamma+4)}\mu$, then we have $t^* \ge \frac{\gamma}{\gamma+2}(\mu-\sigma)$, and the optimal solution and the corresponding objective are

$$t_1^* = \frac{\gamma}{\gamma+2}(\mu-\sigma); \qquad \mathcal{S}_1^* = \frac{(2\gamma+6)\mu^2 + 2\gamma(\gamma+3)\mu\sigma + (6-2d(\gamma+2)^2+\gamma(\gamma+4))\sigma^2}{4(\gamma+2)^2}.$$

2. When $t \in [\frac{\gamma}{\gamma+2}(\mu-\sigma), \frac{\gamma}{\gamma+2}(\mu+\sigma))$, the firm's production and abatement amounts of the low-demand case and the high-demand case are $(x_L, w_L) = (\frac{\mu-\sigma}{\gamma+2}, \frac{\mu-\sigma}{\gamma+2})$ and $(x_H, w_H) = (\frac{\mu+\sigma-t}{2}, \frac{t}{\gamma})$, respectively. Then the expected social welfare in (D.2) becomes
$$\mathcal{S}_2 = -\frac{2d(\gamma+2)^2+\gamma(\gamma+4)}{16\gamma^2}t^2 + \frac{4d+(2d-1)\gamma}{8\gamma}(\mu+\sigma)t + \frac{1}{16(\gamma+2)^2}\Big(\big(24-2d(\gamma+2)^2+\gamma(3\gamma+16)\big)\mu^2 + \big(2\gamma(3\gamma+8)-4d(\gamma+2)^2\big)\mu\sigma + \big(24-2d(\gamma+2)^2+\gamma(3\gamma+16)\big)\sigma^2\Big).$$
The unconstrained optimal solution that maximizes $\mathcal{S}_2$ is $t^* = \frac{2d\gamma(\gamma+2)-\gamma^2}{2d(\gamma+2)^2+\gamma(\gamma+4)}(\mu+\sigma)$, which is less than $\frac{\gamma}{\gamma+2}(\mu+\sigma)$. Similar to the last subspace, we discuss different situations. If $d < \frac{\gamma}{2(\gamma+2)}$, then we have $t^* < 0$, and the optimal solution and the corresponding objective are
$$t_2^* = \frac{\gamma}{\gamma+2}(\mu-\sigma); \qquad \mathcal{S}_2^* = \frac{(2\gamma+6)\mu^2+2\gamma(\gamma+3)\mu\sigma+(6-2d(\gamma+2)^2+\gamma(\gamma+4))\sigma^2}{4(\gamma+2)^2}.$$
If $d \ge \frac{\gamma}{2(\gamma+2)}$ and $\sigma \ge \frac{\gamma(\gamma+3)}{\gamma+2d(\gamma+2)^2}\mu$, then we have $\frac{\gamma}{\gamma+2}(\mu-\sigma) \le t^* \le \frac{\gamma}{\gamma+2}(\mu+\sigma)$, and the optimal solution and the corresponding objective are
$$t_2^* = t^* = \frac{2d\gamma(\gamma+2)-\gamma^2}{2d(\gamma+2)^2+\gamma(\gamma+4)}(\mu+\sigma);$$
$$\mathcal{S}_2^* = \frac{(\gamma+3)\Big(4d(\gamma+2)^2(\mu^2+\sigma^2)+\gamma\big((\gamma^2+5\gamma+8)\mu^2+2\gamma(\gamma+3)\mu\sigma+(\gamma^2+5\gamma+8)\sigma^2\big)\Big)}{8d(\gamma+2)^4+4\gamma(\gamma+2)^2(\gamma+4)}.$$
If $d \ge \frac{\gamma}{2(\gamma+2)}$ and $\sigma < \frac{\gamma(\gamma+3)}{\gamma+2d(\gamma+2)^2}\mu$, then we have $t^* < \frac{\gamma}{\gamma+2}(\mu-\sigma)$, and the optimal solution and the corresponding objective are
$$t_2^* = \frac{\gamma}{\gamma+2}(\mu-\sigma); \qquad \mathcal{S}_2^* = \frac{(2\gamma+6)\mu^2+2\gamma(\gamma+3)\mu\sigma+(6-2d(\gamma+2)^2+\gamma(\gamma+4))\sigma^2}{4(\gamma+2)^2}.$$
3. When $t \in [\frac{\gamma}{\gamma+2}(\mu+\sigma), \infty)$, the firm's production and abatement amounts of the low-demand case and the high-demand case are $(x_L, w_L) = (\frac{\mu-\sigma}{\gamma+2}, \frac{\mu-\sigma}{\gamma+2})$ and $(x_H, w_H) = (\frac{\mu+\sigma}{\gamma+2}, \frac{\mu+\sigma}{\gamma+2})$, respectively. Then the expected social welfare in (D.2) becomes
$$\mathcal{S}_3 = \frac{(\gamma+3)(\mu^2+\sigma^2)}{2(\gamma+2)^2}.$$
The optimal solution t_3^* can be any value larger than $\frac{\gamma}{\gamma+2}(\mu+\sigma)$, and the corresponding objective is always $\mathcal{S}_3^* = \frac{(\gamma+3)(\mu^2+\sigma^2)}{2(\gamma+2)^2}$.

Then, we compare the optimal values of the three subspaces in different situations.

1. When $d \in [0, \frac{\gamma}{2(\gamma+2)})$, the optimal values of these subspaces are
$$\mathcal{S}_1^* = \frac{3-2d}{8}(\mu^2+\sigma^2); \qquad \mathcal{S}_2^* = \frac{(2\gamma+6)\mu^2+2\gamma(\gamma+3)\mu\sigma+(6-2d(\gamma+2)^2+\gamma(\gamma+4))\sigma^2}{4(\gamma+2)^2};$$
$$\mathcal{S}_3^* = \frac{(\gamma+3)(\mu^2+\sigma^2)}{2(\gamma+2)^2}.$$
In this case, $\mathcal{S}_1^* \ge \mathcal{S}_2^* \ge \mathcal{S}_3^*$ and the optimal solution is $t^* = 0$.

2. When $d \in [\frac{\gamma}{2(\gamma+2)}, \infty)$ and $\sigma \in [0, \frac{\gamma(\gamma+3)}{\gamma+2d(\gamma+2)^2}\mu)$, the optimal values of these subspaces are
$$\mathcal{S}_1^* = \frac{(\gamma+3)(\gamma+2d)}{4d(\gamma+2)^2+2\gamma(\gamma+4)}\mu^2 + \frac{3-2d}{8}\sigma^2;$$
$$\mathcal{S}_2^* = \frac{(2\gamma+6)\mu^2+2\gamma(\gamma+3)\mu\sigma+(6-2d(\gamma+2)^2+\gamma(\gamma+4))\sigma^2}{4(\gamma+2)^2};$$
$$\mathcal{S}_3^* = \frac{(\gamma+3)(\mu^2+\sigma^2)}{2(\gamma+2)^2}.$$
In this case, $\mathcal{S}_1^* \ge \mathcal{S}_2^* \ge \mathcal{S}_3^*$ and the optimal solution is $t^* = \frac{2d\gamma(\gamma+2)-\gamma^2}{2d(\gamma+2)^2+\gamma(\gamma+4)}\mu$.

3. When $d\in[\frac{\gamma}{2(\gamma+2)},\infty)$ and $\sigma\in[\frac{\gamma(\gamma+3)}{\gamma+2d(\gamma+2)^2}\mu, \frac{2\gamma(\gamma+3)}{2d(\gamma+2)^2+\gamma(\gamma+4)})$, the optimal values of these subspaces are

$$\mathcal{S}_1^* = \frac{(\gamma+3)(\gamma+2d)}{4d(\gamma+2)^2+2\gamma(\gamma+4)}\mu^2 + \frac{3-2d}{8}\sigma^2;$$
$$\mathcal{S}_2^* = \frac{(\gamma+3)\Big(4d(\gamma+2)^2(\mu^2+\sigma^2)+\gamma\big((\gamma^2+5\gamma+8\big)\mu^2+2\gamma(\gamma+3)\mu\sigma+(\gamma^2+5\gamma+8)\sigma^2\Big)}{8d(\gamma+2)^4+4\gamma(\gamma+2)^2(\gamma+4)};$$
$$\mathcal{S}_3^* = \frac{(\gamma+3)(\mu^2+\sigma^2)}{2(\gamma+2)^2}.$$

If $\sigma \le \frac{2\gamma(\gamma+3)\mu}{2\sqrt{2}d(\gamma+2)^2+(6-2\sqrt{2})\gamma+(2-\sqrt{2})\gamma^2}$, then $\mathcal{S}_1^* \ge \mathcal{S}_2^* \ge \mathcal{S}_3^*$ and hence the optimal solution is $t^* = \frac{2d\gamma(\gamma+2)-\gamma^2}{2d(\gamma+2)^2+\gamma(\gamma+4)}\mu$. Otherwise, the optimal solution is $t^*=\frac{2d\gamma(\gamma+2)-\gamma^2}{2d(\gamma+2)^2+\gamma(\gamma+4)}(\mu+\sigma)$.

4. When $d\in[\frac{\gamma}{2(\gamma+2)},\infty)$ and $\sigma\in[\frac{2\gamma(\gamma+3)}{2d(\gamma+2)^2+\gamma(\gamma+4)},\mu)$, the optimal values of these subspaces are

$$\mathcal{S}_1^* = \frac{(2\gamma+6)\mu^2+2\gamma(\gamma+3)\mu\sigma+(6-2d(\gamma+2)^2+\gamma(\gamma+4))\sigma^2}{4(\gamma+2)^2};$$
$$\mathcal{S}_2^* = \frac{(\gamma+3)\Big(4d(\gamma+2)^2(\mu^2+\sigma^2)+\gamma\big((\gamma^2+5\gamma+8\big)\mu^2+2\gamma(\gamma+3)\mu\sigma+(\gamma^2+5\gamma+8)\sigma^2\Big)}{8d(\gamma+2)^4+4\gamma(\gamma+2)^2(\gamma+4)};$$
$$\mathcal{S}_3^* = \frac{(\gamma+3)(\mu^2+\sigma^2)}{2(\gamma+2)^2}.$$

In this case, $\mathcal{S}_2^*\ge\mathcal{S}_1^*\ge\mathcal{S}_3^*$ and the optimal solution is $t^*=\frac{2d\gamma(\gamma+2)-\gamma^2}{2d(\gamma+2)^2+\gamma(\gamma+4)}(\mu+\sigma)$.

Lastly, we define $\delta^P := \frac{\gamma}{2(\gamma+2)}$ and $\Omega^P := \frac{2\gamma(\gamma+3)}{2\sqrt{2}d(\gamma+2)^2+(6-2\sqrt{2})\gamma+(2-\sqrt{2})\gamma^2}$, and summarize the above results as follows:

1. When $d\in[0,\delta^P)$, the optimal tax rate and the corresponding objective are

$$t^*=0 \quad\text{and}\quad \mathcal{S}^*=\frac{3-2d}{8}(\mu^2+\sigma^2).$$

2. When $d\in[\delta^P,\infty)$ and $\sigma/\mu\in[0,\Omega^P)$, the optimal solution and the corresponding objective are

$$t^*=\mathcal{C}_1^P\mu \quad\text{and}\quad \mathcal{S}^*=\mathcal{C}_2^P\mu^2+\mathcal{C}_3^P\sigma^2,$$

where $\mathcal{C}_1^P := \frac{2d\gamma(\gamma+2)-\gamma^2}{2d(\gamma+2)^2+\gamma(\gamma+4)}$, $\mathcal{C}_2^P := \frac{(\gamma+3)(\gamma+2d)}{4d(\gamma+2)^2+2\gamma(\gamma+4)}$ and $\mathcal{C}_3^P := \frac{3-2d}{8}$. Subsequently, the corresponding expected production and abatement amounts are

$$\mathcal{P}=\frac{(\gamma+2)(2d+\gamma)\mu}{2d(\gamma+2)^2+\gamma(\gamma+4)} \quad\text{and}\quad \mathcal{B}=\frac{2d(\gamma+2)-\gamma}{2d(\gamma+2)^2+\gamma(\gamma+4)}\mu.$$

3. When $d\in[\delta^P,\infty)$ and $\sigma/\mu\in[\Omega^P,1)$, the optimal solution and the corresponding objective are

$$t^*=\mathcal{C}_1^P(\mu+\sigma) \quad\text{and}\quad \mathcal{S}^*=\mathcal{C}_4^P(\mu^2+\sigma^2)+\mathcal{C}_P^5\mu\sigma,$$

where $\mathcal{C}_1^P = \frac{2d\gamma(\gamma+2)-\gamma^2}{2d(\gamma+2)^2+\gamma(\gamma+4)}$, $\mathcal{C}_4^P := \frac{(\gamma+3)\big(4d(\gamma+2)^2+\gamma(\gamma^2+5\gamma+8)\big)}{8d(\gamma+2)^4+4\gamma(\gamma+2)^2(\gamma+4)}$ and $\mathcal{C}_5^P := \frac{\gamma^2(\gamma+3)^2}{4d(\gamma+2)^4+2\gamma(\gamma+2)^2(\gamma+4)}$. Subsequently, the corresponding expected production and abatement amounts are

$$\mathcal{P}=\frac{\big(4d(\gamma+2)^2+\gamma(\gamma^2+5\gamma+8)\big)\mu+\gamma^2(\gamma+3)\sigma}{2(\gamma+2)\big(2d(\gamma+2)^2+\gamma(\gamma+4)\big)} \quad\text{and}\quad \mathcal{B}=\frac{2d(\gamma+2)^2\mu+\gamma(\mu-(\gamma+3)\sigma)}{(\gamma+2)\left(2d(\gamma+2)^2+\gamma(\gamma+4)\right)}.$$

□

D.3. Solution to Problem (4)

Given a cap k, the firm maximizes its profit as follows:

$$\max_{x,w:0\leq x-w\leq k;w\geq 0}\left\{\theta x-x^2-\frac{\gamma}{2}w^2\right\}.$$

When $k=0$, we have $x-w=0$. We can deduce that $x^*=\frac{\theta}{\gamma+2}$ and $w^*=\frac{\theta}{\gamma+2}$. Then we only consider the case when $k>0$. Similar to Section D.1, we first ignore the constraint $x-w\geq 0$. Then we will check the feasibility of the solutions. Therefore, we can write the Lagrangian as

$$\mathcal{L}=\theta x-x^2-\frac{\gamma}{2}w^2-\mu(x-w-k)+\eta w.$$

According to the KKT conditions, we have $\mu,\eta\geq 0$ and

$$\frac{\partial\mathcal{L}}{\partial x}=\theta-2x-\mu=0;\qquad \frac{\partial\mathcal{L}}{\partial w}=-\gamma w+\mu+\eta=0;\qquad \mu(x-w-k)=0;\qquad \eta w=0. \tag{D.3}$$

Then we consider candidate cases:

1. When $\mu>0$ and $\eta>0$, according to complementary conditions, we have $x=w+k$ and $w=0$. Then $0=\frac{\partial\mathcal{L}}{\partial w}=\mu+\eta>0$. Thus we reach a contradiction in this case.
2. When $\mu>0$ and $\eta=0$, according to the complementary condition $\mu(x-w-k)=0$, we have $x=w+k$. Solving (D.3), we have $x_2^*=\frac{\theta+k\gamma}{\gamma+2}$, $w_2^*=\frac{\theta-2k}{\gamma+2}$ and $\mu=\frac{\gamma(\theta-2k)}{\gamma+2}>0$. The corresponding Lagrangian is $\mathcal{L}_2^*=\frac{(\theta-2k)^2}{2(\gamma+2)}+\theta k-k^2$.
3. When $\mu=0$ and $\eta>0$, according to the complementary condition $\eta w=0$, we have $w=0$. Then $0=\frac{\partial\mathcal{L}}{\partial w}=\eta>0$. Thus we reach a contradiction in this case.
4. When $\mu=0$ and $\eta=0$, solving (D.3), we have $x_4^*=\frac{\theta}{2}$ and $w_4^*=0$. The corresponding Lagrangian is $\mathcal{L}_4^*=\frac{\theta^2}{4}$.

When $k\geq\frac{\theta}{2}$, (x_4^*,w_4^*) is the unique possible solution and hence is the optimal solution. When $k<\frac{\theta}{2}$, there exist two possible solutions (x_2^*,w_2^*) and (x_4^*,w_4^*). Since $\mathcal{L}_2^*\geq\mathcal{L}_4^*$, the optimal solution is (x_2^*,w_2^*). Then we can check that these optimal solutions satisfy the constraint $x-w\geq 0$. To summarize, we have the firm's optimal decision as follows:

$$x^*(\theta,k)=\begin{cases}\frac{\theta+k\gamma}{\gamma+2} & k\in[0,\frac{\theta}{2})\\ \frac{\theta}{2} & k\in[\frac{\theta}{2},\infty),\end{cases}\qquad w^*(\theta,k)=\begin{cases}\frac{\theta-2k}{\gamma+2} & k\in[0,\frac{\theta}{2})\\ 0 & k\in[\frac{\theta}{2},\infty).\end{cases}\qquad\square$$

D.4. Proof of Proposition 2

Having the firm's response in Section D.3, the regulator's optimization problem is as follows:

$$\mathcal{S}^*=\max_{k\geq 0}\mathbb{E}_\theta\left[\theta x^*(\theta,k)-(x^*(\theta,k))^2-\frac{\gamma}{2}\big(w^*(\theta,k)\big)^2+\frac{1}{2}(x^*(\theta,k))^2-d\big(x^*(\theta,k)-w^*(\theta,k)\big)^2\right]. \tag{D.4}$$

Again we divide the feasible space into three subspaces and compute the firm's solution in each subspace. Then, we compare the optimal solutions of these subspaces and derive the global optimal solution. We first compute the optimal solution in each subspace.

1. When $k \in [0, \frac{\mu-\sigma}{2})$, the optimal production and abatement amounts of the low-demand case and the high-demand case are $(x_L, w_L) = (\frac{\mu-\sigma+k\gamma}{\gamma+2}, \frac{\mu-\sigma-2k}{\gamma+2})$ and $(x_H, w_H) = (\frac{\mu+\sigma+k\gamma}{\gamma+2}, \frac{\mu+\sigma-2k}{\gamma+2})$, respectively. Then the expected social welfare in (D.4) becomes
$$\mathcal{S}_1 = -\frac{\gamma^2 + 2d(\gamma+2)^2 + 4\gamma}{2(\gamma+2)^2}k^2 + \frac{\gamma(\gamma+3)}{(\gamma+2)^2}\mu k + \frac{\gamma+3}{2(\gamma+2)^2}(\mu^2+\sigma^2).$$
The unconstrained optimal solution that maximizes $\mathcal{S}_1$ is $k^* = \frac{\gamma(\gamma+3)\mu}{2d(\gamma+2)^2+\gamma(\gamma+4)}$. Since it is a quadratic function and may not be monotonic in this subspace, we discuss different situations. If $d < \frac{\gamma}{2(\gamma+2)}$, then we have $k^* > \frac{\mu-\sigma}{2}$, and the optimal solution and the corresponding objective are
$$k_1^* = \frac{\mu-\sigma}{2}; \qquad \mathcal{S}_1^* = \frac{3-2d}{8}\mu^2 + \frac{4d+(2d-1)\gamma}{4\gamma+8}\mu\sigma - \frac{\gamma^2+2d(\gamma+2)^2-12}{8(\gamma+2)^2}\sigma^2.$$
If $d \geq \frac{\gamma}{2(\gamma+2)}$ and $\sigma < \frac{(\gamma+2)(4d+(2d-1)\gamma)\mu}{2d(\gamma+2)^2+\gamma(\gamma+4)}$, then we have $0 \leq k^* < \frac{\mu-\sigma}{2}$, and the optimal solution and the corresponding objective are
$$k_1^* = k^* = \frac{\gamma(\gamma+3)\mu}{2d(\gamma+2)^2+\gamma(\gamma+4)}; \qquad \mathcal{S}_1^* = \frac{(\gamma+2d)(\gamma+3)}{4d(\gamma+2)^2+2\gamma(\gamma+4)}\mu^2 + \frac{\gamma+3}{2(\gamma+2)^2}\sigma^2.$$
If $d \geq \frac{\gamma}{2(\gamma+2)}$ and $\sigma \geq \frac{(\gamma+2)(4d+(2d-1)\gamma)\mu}{2d(\gamma+2)^2+\gamma(\gamma+4)}$, then we have $k^* \geq \frac{\mu-\sigma}{2}$, and the optimal solution and the corresponding objective are
$$k_1^* = \frac{\mu-\sigma}{2}; \qquad \mathcal{S}_1^* = \frac{3-2d}{8}\mu^2 + \frac{4d+(2d-1)\gamma}{4\gamma+8}\mu\sigma - \frac{\gamma^2+2d(\gamma+2)^2-12}{8(\gamma+2)^2}\sigma^2.$$

2. When $k \in [\frac{\mu-\sigma}{2}, \frac{\mu+\sigma}{2})$, the firm's production and abatement amounts of the low-demand case and the high-demand case are $(x_L, w_L) = (\frac{\mu-\sigma}{2}, 0)$ and $(x_H, w_H) = (\frac{\mu+\sigma+k\gamma}{\gamma+2}, \frac{\mu+\sigma-2k}{\gamma+2})$, respectively. Then the expected social welfare in (D.4) becomes
$$\begin{aligned}\mathcal{S}_2 = &-\frac{2d(\gamma+2)^2+\gamma(\gamma+4)}{4(\gamma+2)^2}k^2 + \frac{\gamma(\gamma+3)}{2(\gamma+2)^2}(\mu+\sigma)k \\ &+ \frac{3\gamma^2(\mu-\sigma)^2 - 2d(\gamma+2)^2(\mu-\sigma)^2 + 24(\mu^2+\sigma^2) + 16\gamma(\mu-\mu\sigma+\sigma^2)}{16(\gamma+2)^2}.\end{aligned}$$
The unconstrained optimal solution that maximizes $\mathcal{S}_2$ is $k^* = \frac{\gamma(\gamma+3)}{2d(\gamma+2)^2+\gamma(\gamma+4)}(\mu+\sigma)$. Similarly, we discuss different situations. If $d < \frac{\gamma}{2(\gamma+2)}$, then we have $k^* > \frac{\mu+\sigma}{2}$, and the optimal solution and the corresponding objective are
$$k_2^* = \frac{\mu+\sigma}{2}; \qquad \mathcal{S}_2^* = \frac{3-2d}{8}(\mu^2+\sigma^2).$$
If $d \geq \frac{\gamma}{2(\gamma+2)}$ and $\sigma \geq \frac{(\gamma+2)(4d+(2d-1)\gamma)}{2d(\gamma+2)^2+\gamma(3\gamma+10)}\mu$, then we have $\frac{\mu-\sigma}{2} \leq k^* \leq \frac{\mu+\sigma}{2}$, and the optimal solution and the corresponding objective are
$$\begin{aligned}k_2^* &= k^* = \frac{\gamma(\gamma+3)}{2d(\gamma+2)^2+\gamma(\gamma+4)}(\mu+\sigma); \\ \mathcal{S}_2^* &= \frac{1}{16\big(2d(\gamma+2)^2+\gamma(\gamma+4)\big)}\Big(4d\big((\gamma^2+6\gamma+12)\mu^2 - 2\gamma(\gamma+2)\mu\sigma + (\gamma^2+6\gamma+12)\sigma^2\big) \\ &\qquad + \gamma\big((24+7\gamma)\mu^2 + 2\gamma\mu\sigma + (24+7\gamma)\sigma^2\big) - 4d^2(\gamma+2)^2(\mu-\sigma)^2\Big).\end{aligned}$$
If $d \geq \frac{\gamma}{2(\gamma+2)}$ and $\sigma < \frac{(\gamma+2)(4d+(2d-1)\gamma)}{2d(\gamma+2)^2+\gamma(3\gamma+10)}\mu$, then we have $k^* < \frac{\mu-\sigma}{2}$, and the optimal solution and the corresponding objective are
$$k^* = \frac{\mu-\sigma}{2}; \qquad \mathcal{S}_2^* = \frac{3-2d}{8}\mu^2 + \frac{4d+(2d-1)\gamma}{4\gamma+8}\mu\sigma - \frac{\gamma^2+2d(\gamma+2)^2-12}{8(\gamma+2)^2}\sigma^2.$$

3. When $k \in [\frac{\mu+\sigma}{2}, \infty)$, the firm's production and abatement amounts of the low-demand case and the high-demand case are $(x_L, w_L) = (\frac{\mu-\sigma}{2}, 0)$ and $(x_H, w_H) = (\frac{\mu+\sigma}{2}, 0)$, respectively. Then the expected social welfare in (D.4) becomes
$$\mathcal{S}_3 = \frac{3-2d}{8}(\mu^2+\sigma^2).$$
The optimal solution k_3^* can be any value greater than $\frac{\mu+\sigma}{2}$, and the corresponding objective is always $\mathcal{S}_3^* = \frac{3-2d}{8}(\mu^2+\sigma^2)$.

Then we compare the optimal values of the three subspaces in different situations.

1. When $d \in [0, \frac{\gamma}{2(\gamma+2)})$, the optimal values of these subspaces are
$$\begin{aligned}\mathcal{S}_1^* &= \frac{3-2d}{8}\mu^2 + \frac{4d+(2d-1)\gamma}{4\gamma+8}\mu\sigma - \frac{\gamma^2+2d(\gamma+2)^2-12}{8(\gamma+2)^2}\sigma^2;\\ \mathcal{S}_2^* &= \frac{3-2d}{8}(\mu^2+\sigma^2);\\ \mathcal{S}_3^* &= \frac{3-2d}{8}(\mu^2+\sigma^2).\end{aligned}$$
In this case, $\mathcal{S}_2^* = \mathcal{S}_3^* \geq \mathcal{S}_1^*$ and the optimal solution is $k^* = c$, $\forall c \geq \frac{\mu+\sigma}{2}$.

2. When $d \in [\frac{\gamma}{2(\gamma+2)}, \infty)$ and $\sigma \in [0, \frac{(\gamma+2)(4d+(2d-1)\gamma)}{2d(\gamma+2)^2+\gamma(3\gamma+10)}\mu)$, the optimal values in the first and second subspaces are
$$\begin{aligned}\mathcal{S}_1^* &= \frac{(\gamma+2d)(\gamma+3)}{4d(\gamma+2)^2+2\gamma(\gamma+4)}\mu^2 + \frac{\gamma+3}{2(\gamma+2)^2}\sigma^2;\\ \mathcal{S}_2^* &= \frac{3-2d}{8}\mu^2 + \frac{4d+(2d-1)\gamma}{4\gamma+8}\mu\sigma - \frac{\gamma^2+2d(\gamma+2)^2-12}{8(\gamma+2)^2}\sigma^2;\\ \mathcal{S}_3^* &= \frac{3-2d}{8}(\mu^2+\sigma^2).\end{aligned}$$
In this case, $\mathcal{S}_1^* \geq \mathcal{S}_2^* \geq \mathcal{S}_3^*$ and the optimal solution is $k^* = \frac{\gamma(\gamma+3)}{2d(\gamma+2)^2+\gamma(\gamma+4)}\mu$.

3. When $d \in [\frac{\gamma}{2(\gamma+2)}, \infty)$ and $\sigma \in [\frac{(\gamma+2)(4d+(2d-1)\gamma)}{2d(\gamma+2)^2+\gamma(3\gamma+10)}\mu, \frac{(\gamma+2)(4d+(2d-1)\gamma)\mu}{2d(\gamma+2)^2+\gamma(\gamma+4)})$, the optimal values of these subspaces are
$$\begin{aligned}\mathcal{S}_1^* &= \frac{(\gamma+2d)(\gamma+3)}{4d(\gamma+2)^2+2\gamma(\gamma+4)}\mu^2 + \frac{\gamma+3}{2(\gamma+2)^2}\sigma^2;\\ \mathcal{S}_2^* &= \frac{1}{16\big(2d(\gamma+2)^2+\gamma(\gamma+4)\big)}\Big(4d\big((\gamma^2+6\gamma+12)\mu^2 - 2\gamma(\gamma+2)\mu\sigma + (\gamma^2+6\gamma+12)\sigma^2\big)\\ &\qquad + \gamma\big((24+7\gamma)\mu^2 + 2\gamma\mu\sigma + (\gamma^2+6\gamma+12)\sigma^2\big) - 4d^2(\gamma+2)^2(\mu-\sigma)^2\Big);\\ \mathcal{S}_3^* &= \frac{3-2d}{8}(\mu^2+\sigma^2).\end{aligned}$$
If $\sigma \leq \frac{(\gamma+2)(4d-\gamma+2d\gamma)\mu}{2d(\gamma+2)^2+\gamma(6\sqrt{2}+(2\sqrt{2}-1)\gamma)}$, then $\mathcal{S}_1^* \geq \mathcal{S}_2^* \geq \mathcal{S}_3^*$ and hence the optimal solution is $k^* = \frac{\gamma(\gamma+3)}{2d(\gamma+2)^2+\gamma(\gamma+4)}\mu$. Otherwise, $\mathcal{S}_2^* \geq \mathcal{S}_1^* \geq \mathcal{S}_3^*$ and hence the optimal solution is $k^* = \frac{\gamma(\gamma+3)}{2d(\gamma+2)^2+\gamma(\gamma+4)}(\mu+\sigma)$.

4. When $d \in [\frac{\gamma}{2(\gamma+2)}, \infty)$ and $\sigma \in [\frac{(\gamma+2)(4d+(2d-1)\gamma)\mu}{2d(\gamma+2)^2+\gamma(\gamma+4)}, \mu)$, the optimal values of these subspaces are
$$\begin{aligned}\mathcal{S}_1^* &= \frac{3-2d}{8}\mu^2 + \frac{4d+(2d-1)\gamma}{4\gamma+8}\mu\sigma - \frac{\gamma^2+2d(\gamma+2)^2-12}{8(\gamma+2)^2}\sigma^2;\\ \mathcal{S}_2^* &= \frac{1}{16\big(2d(\gamma+2)^2+\gamma(\gamma+4)\big)}\Big(4d\big((\gamma^2+6\gamma+12)\mu^2 - 2\gamma(\gamma+2)\mu\sigma + (\gamma^2+6\gamma+12)\sigma^2\big)\\ &\qquad + \gamma\big((24+7\gamma)\mu^2 + 2\gamma\mu\sigma + (\gamma^2+6\gamma+12)\sigma^2\big) - 4d^2(\gamma+2)^2(\mu-\sigma)^2\Big);\\ \mathcal{S}_3^* &= \frac{3-2d}{8}(\mu^2+\sigma^2).\end{aligned}$$

In this case, $\mathcal{S}_2^* \geq \mathcal{S}_1^* \geq \mathcal{S}_3^*$ and the optimal solution is $k^* = \frac{\gamma(\gamma+3)}{2d(\gamma+2)^2+\gamma(\gamma+4)}(\mu+\sigma)$.

Lastly, we define $\delta^Q := \frac{\gamma}{2(\gamma+2)}$ and $\Omega^Q := \frac{(\gamma+2)(4d-\gamma+2d\gamma)}{2d(\gamma+2)^2+\gamma(6\sqrt{2}-2+(2\sqrt{2}-1)\gamma)}$, and summarize the above results as follows:

1. When $d \in [0, \delta^Q)$, the optimal solution and the corresponding objective are

$$k^* = c, \ \forall c \geq \frac{\mu+\sigma}{2}; \quad \text{and} \quad \mathcal{S}^* = \frac{3-2d}{8}(\mu^2+\sigma^2).$$

2. When $d \in [\delta^Q, \infty)$ and $\sigma/\mu \in [0, \Omega^Q)$, the optimal solution and the corresponding objective are

$$k^* = \mathcal{C}_1^Q \mu \quad \text{and} \quad \mathcal{S}^* = \mathcal{C}_2^Q \mu^2 + \mathcal{C}_3^Q \sigma^2,$$

   where $\mathcal{C}_1^Q := \frac{\gamma(\gamma+3)}{2d(\gamma+2)^2+\gamma(\gamma+4)}$, $\mathcal{C}_2^Q := \frac{(\gamma+2d)(\gamma+3)}{4d(\gamma+2)^2+2\gamma(\gamma+4)}$ and $\mathcal{C}_3^Q := \frac{\gamma+3}{2(\gamma+2)^2}$. Subsequently, the corresponding expected production and abatement amounts are

$$\mathcal{P} = \frac{(\gamma+2)(2d+\gamma)\mu}{2d(\gamma+2)^2+\gamma(\gamma+4)} \quad \text{and} \quad \mathcal{B} = \frac{2d(\gamma+2)-\gamma}{2d(\gamma+2)^2+\gamma(\gamma+4)}\mu.$$

3. When $d \in [\delta^Q, \infty)$ and $\sigma \in [\Omega^Q, 1)$, the optimal solution and the corresponding objective are

$$k^* = \mathcal{C}_1^Q(\mu+\sigma) \quad \text{and} \quad \mathcal{S}^* = \mathcal{C}_4^Q(\mu^2+\sigma^2) + \mathcal{C}_5^Q \mu\sigma,$$

   where $\mathcal{C}_1^Q = \frac{\gamma(\gamma+3)}{2d(\gamma+2)^2+\gamma(\gamma+4)}$, $\mathcal{C}_4^Q := \frac{4d(\gamma^2+6\gamma+12)+\gamma(7\gamma+24)-4d^2(\gamma+2)^2}{16\left(2d(\gamma+2)^2+\gamma(\gamma+4)\right)}$ and $\mathcal{C}_5^Q := \frac{(\gamma-4d-2d\gamma)^2}{8\left(2d(\gamma+2)^2+\gamma(\gamma+4)\right)}$. Subsequently, the corresponding expected production and abatement amounts are

$$\mathcal{P} = \frac{\left(2d(\gamma+2)(\gamma+4)+\gamma(3\gamma+8)\right)\mu + \left(\gamma^2-2d\gamma(\gamma+2)\right)\sigma}{8d(\gamma+2)^2+4\gamma(\gamma+4)} \quad \text{and} \quad \mathcal{B} = \frac{(2d(\gamma+2)-\gamma)(\mu+\sigma)}{4d(\gamma+2)^2+2\gamma(\gamma+4)}.$$

□

D.5. Proof of Theorem 1

We will discuss the expected social welfare under these instruments under different market conditions. We use the “price curve” and the “quantity curve” to represent the expected social welfare curves under the price instrument and the quantity instrument, respectively. According to Propositions 1 and 2, we analyze these two curves under different market conditions (Define $\delta := \frac{\gamma}{2(\gamma+2)}$, $\Delta := \frac{\gamma(3\gamma+8)}{2(\gamma+2)^2}$ and $\Omega := \sqrt{2}-1$).

1. $d \in [0, \delta)$. In this case, since $\delta^P = \delta^Q = \delta$, the two curves coincide with the expected social welfare without regulation.

2. $d \in [\delta, \Delta)$. In this case, we have $\Omega^Q \leq \Omega^P$. The price curve is over the quantity curve when $\sigma/\mu \in [0, \Omega^Q)$, and the quantity curve is over the price curve when $\sigma/\mu \in (\Omega^P, 1)$. When $\sigma/\mu \in [\Omega^Q, \Omega^P]$, the two curves have one unique cross point $\mu/\sigma = \Omega$.

3. $d \in (\Delta, \infty)$. In this case, we have $\Omega^Q \geq \Omega^P$, and the two curves have one unique cross point $\mu/\sigma = \Omega$.
   □

D.6. Proof of Proposition 3

Let $\mathcal{B}^P$ and $\mathcal{B}^Q$ denote the expected abatement amount under the price and the quantity instruments, respectively. According to Lemmas 1 and 2, we have compare the expected abatement amount as follows:

1. When $d \in [0, \delta)$, we have $\mathcal{B}^P = \mathcal{B}^Q = 0$.
2. When $d \in [\delta, \infty)$ and $\sigma/\mu \in [0, \min\{\Omega^P, \Omega^Q\})$, we have $\mathcal{S}^P = \mathcal{S}^Q = \frac{2d(\gamma+2)-\gamma}{2d(\gamma+2)^2+\gamma(\gamma+4)}\mu$.
3. When $d \in [\delta, \Delta)$ and $\sigma/\mu \in [\Omega^Q, \Omega^P)$, we have
$$\mathcal{B}^P = \frac{2d(\gamma+2)-\gamma}{2d(\gamma+2)^2+\gamma(\gamma+4)}\mu \quad \text{and} \quad \mathcal{B}^Q = \frac{(2d(\gamma+2)-\gamma)(\mu+\sigma)}{4d(\gamma+2)^2+2\gamma(\gamma+4)},$$
and $\mathcal{B}^P \geq \mathcal{B}^Q$.
4. When $d \in [\Delta, \infty)$ and $\sigma/\mu \in [\Omega^P, \Omega^Q)$, we have
$$\mathcal{B}^P = \frac{2d(\gamma+2)^2\mu + \gamma(\mu-(\gamma+3)\sigma)}{(\gamma+2)\,(2d(\gamma+2)^2+\gamma(\gamma+4))} \quad \text{and} \quad \mathcal{B}^Q = \frac{(2d(\gamma+2)-\gamma)\mu}{4d(\gamma+2)^2+2\gamma(\gamma+4)},$$
and $\mathcal{B}^P \geq \mathcal{B}^Q$.
5. When $d \in [\delta, \infty)$ and $\sigma/\mu \in [\max\{\Omega^P, \Omega^Q\}, \infty)$, we have
$$\mathcal{B}^P = \frac{2d(\gamma+2)^2\mu + \gamma(\mu-(\gamma+3)\sigma)}{(\gamma+2)\,(2d(\gamma+2)^2+\gamma(\gamma+4))} \quad \text{and} \quad \mathcal{B}^Q = \frac{(2d(\gamma+2)-\gamma)(\mu+\sigma)}{4d(\gamma+2)^2+2\gamma(\gamma+4)},$$
and $\mathcal{B}^P \geq \mathcal{B}^Q$.

Similarly, we can prove that $\mathcal{P}^P \leq \mathcal{P}^Q$. To summarize, compared to the quantity instrument, the price instrument always induces more abatement but less production. □

D.7. Solution to Problem (7)

Given a tax rate t, firm i will optimize its production and abatement according to the following problem:

$$\max_{x_i \geq w_i \geq 0} \left\{ (\theta - x_i - x_{-i})x_i - t(x_i - w_i) - \frac{\gamma}{2}w_i^2 \right\}.$$

Similar to Proof D.1, we can deduce that firm i's optimal decision is

$$x_i^*(\theta,t) = \begin{cases} \frac{\theta-t-x_{-i}}{2} & \frac{\theta-t-x_{-i}}{2} \geq \frac{t}{\gamma} \\ \frac{\theta-x_{-i}}{\gamma+2} & \frac{\theta-t-x_{-i}}{2} < \frac{t}{\gamma}, \end{cases} \qquad w_i^*(\theta,t) = \begin{cases} \frac{t}{\gamma} & \frac{\theta-t-x_{-i}}{2} \geq \frac{t}{\gamma} \\ \frac{\theta-x_{-i}}{\gamma+2} & \frac{\theta-t-x_{-i}}{2} < \frac{t}{\gamma}. \end{cases}$$

In symmetric Nash equilibrium, each firm would produce the same amount of products (i.e., $x_{-i} = (N-1)x_i$) and this symmetric condition leads to the following fixed point problem:

$$x_i^*(\theta,t) = \begin{cases} \frac{\theta-t-(N-1)x_i^*}{2} & \text{when } \frac{\theta-t-(N-1)x_i^*}{2} \geq \frac{t}{\gamma} \\ \frac{\theta-(N-1)x_i^*}{\gamma+2} & \frac{\theta-t-(N-1)x_i^*}{2} < \frac{t}{\gamma}, \end{cases} \qquad w_i^*(\theta,t) = \begin{cases} \frac{t}{\gamma} & \frac{\theta-t-(N-1)x_i^*}{2} \geq \frac{t}{\gamma} \\ \frac{\theta-(N-1)x_i^*}{\gamma+2} & \frac{\theta-t-(N-1)x_i^*}{2} < \frac{t}{\gamma}. \end{cases}$$

Then, we can deduce that the equilibrium solution is:

$$x_i^*(\theta,t) = \begin{cases} \frac{\theta-t}{N+1} & t \leq \frac{\gamma\theta}{N+\gamma+1} \\ \frac{\theta}{N+\gamma+1} & t > \frac{\gamma\theta}{N+\gamma+1}, \end{cases} \qquad w_i^*(\theta,t) = \begin{cases} \frac{t}{\gamma} & t \leq \frac{\gamma\theta}{N+\gamma+1} \\ \frac{\theta}{N+\gamma+1} & t > \frac{\gamma\theta}{N+\gamma+1}. \end{cases}$$

Due to symmetric Nash equilibrium, each firm will have the same decision. □

D.8. Proof of Lemma 1

Having each firm's response in Section D.7, the regulator's optimization problem is as follows (Since we study the symmetric Nash equilibrium, we can write $\sum_{i=1}^{N} x_i^*(\theta,t)$ as $Nx^*(\theta,t)$ and $\sum_{i=1}^{N} w_i^*(\theta,t)$ as $Nw^*(\theta,t)$):

$$\mathcal{S}^* = \max_{t\geq 0} \mathbb{E}_\theta \left[\theta N x^*(\theta,t) - \frac{1}{2}\big(Nx^*(\theta,t)\big)^2 - \frac{\gamma N}{2} w^*(\theta,t)^2 - d\big(\beta N + (1-\beta)N^2\big)(x^*(\theta,t) - w^*(\theta,t))^2\right]. \quad \text{(D.5)}$$

Similar to previous proofs, we divide the decision space into three subspaces and find the optimal solution of each subspace. Then we compare the optimal solutions of the three subspaces to get the global optimal solution. First, we compute the optimal solution in each subspace.

1. When $t \in [0, \frac{\gamma(\mu-\sigma)}{N+\gamma+1})$, each firm's response in the low-demand and high-demand cases are $(x_{i,L}, w_{i,L}) = (\frac{\mu-\sigma-t}{N+1}, \frac{t}{\gamma})$ and $(x_{i,H}, w_{i,H}) = (\frac{\mu+\sigma-t}{N+1}, \frac{t}{\gamma})$, respectively. Then the expected social welfare in (D.5) becomes

$$\begin{aligned}\mathcal{S}_1 = \frac{N}{2\gamma^2(N+1)^2}\Big(&\gamma^2\big(\left(\mu^2+\sigma^2\right)(2\beta d(N-1) - 2dN + N + 2) + 2\mu t(2d(\beta(-N)+\beta+N) - 1)\\ &+ t^2(2\beta d(N-1) - (2d+1)N)\big) - \gamma(N+1)t(4\beta d(N-1)(\mu-t) + N(-4d\mu+4dt+t)+t)\\ &+ 2d(N+1)^2t^2(\beta(N-1)-N)\Big).\end{aligned}$$

The unconstrained optimal solution that maximizes $\mathcal{S}_1$ is $t^* = \frac{\gamma\left(2d(N-\beta(N-1))(\gamma+N+1)-\gamma\right)}{-2\beta d(N-1)(\gamma+N+1)^2+2dN(\gamma+N+1)^2+\gamma N(\gamma+N+2)+\gamma}\mu$. Since $\mathcal{S}_1$ is a quadratic function of t and may not be monotonic in this subspace, we discuss different situations. If $d < \frac{\gamma}{2(N+\beta-\beta N)(N+\gamma+1)}$, then we have $t^* < 0$, and the optimal solution and the corresponding objective are

$$t_1^* = 0 \qquad \text{and} \qquad \mathcal{S}_1^* = \frac{N\left(\mu^2+\sigma^2\right)(2\beta d(N-1) - 2dN + N + 2)}{2(N+1)^2}.$$

If $d \geq \frac{\gamma}{2(N+\beta-\beta N)(N+\gamma+1)}$ and $\sigma < \frac{\gamma\mu(N+1)(\gamma+N+2)}{-2\beta d(N-1)(\gamma+N+1)^2+2dN(\gamma+N+1)^2+\gamma N(\gamma+N+2)+\gamma}$, then we have $0 \leq t^* < \mu-\sigma$, and the optimal solution and the corresponding objective are

$$t_1^* = t^* = \frac{\gamma\left(2d(N-\beta(N-1))(\gamma+N+1)-\gamma\right)}{-2\beta d(N-1)(\gamma+N+1)^2+2dN(\gamma+N+1)^2+\gamma N(\gamma+N+2)+\gamma}\mu \qquad \text{and}$$

$$\begin{aligned}\mathcal{S}_1^* = -&\frac{N}{2(N+1)^2\left(-2\beta d(N-1)(\gamma+N+1)^2+2dN(\gamma+N+1)^2+\gamma N(\gamma+N+2)+\gamma\right)}\Big(4\beta^2d^2(N-1)^2\sigma^2(\gamma+N+1)^2\\ &+ 2\beta d(N-1)\left(\sigma^2\left(\gamma^2(2-4dN) - \gamma(N+1)(8dN-N-3) - (N+1)^2(4dN-N-2)\right) + \mu^2(N+1)^2(\gamma+N+2)\right)\\ &+ \sigma^2((2d-1)N-2)\left(\gamma^2(2dN+N) + \gamma(N+1)(4dN+N+1) + 2dN(N+1)^2\right)\\ &- \mu^2(N+1)^2(\gamma+N+2)(2dN+\gamma)\Big).\end{aligned}$$

If $d \geq \frac{\gamma}{2(N+\beta-\beta N)(N+\gamma+1)}$ and $\sigma \geq \frac{\gamma\mu(N+1)(\gamma+N+2)}{-2\beta d(N-1)(\gamma+N+1)^2+2dN(\gamma+N+1)^2+\gamma N(\gamma+N+2)+\gamma}$, then we have $t^* \geq \frac{\gamma(\mu-\sigma)}{N+\gamma+1}$, and the optimal solution and the corresponding objective are

$$t^* = \frac{\gamma(\mu-\sigma)}{N+\gamma+1} \qquad \text{and}$$

$$\begin{aligned}\mathcal{S}_1^* = \frac{N}{2(N+1)^2(\gamma+N+1)^2}\Big(&\sigma^2\big(4\beta d(N-1)(\gamma+N+1)^2 + \gamma^2(2-4dN) - \gamma(N+1)(8dN-N-3)\\ &- (N+1)^2(4dN-N-2)\big) + \mu^2(N+1)^2(\gamma+N+2) + 2\gamma\mu(N+1)\sigma(\gamma+N+2)\Big).\end{aligned}$$

2. When $t \in [\frac{\gamma(\mu-\sigma)}{N+\gamma+1}, \frac{\gamma(\mu+\sigma)}{N+\gamma+1})$, each firm's optimal production amounts in the low-demand and high-demand cases are $(x_{i,L}, w_{i,L}) = (\frac{\mu-\sigma}{N+\gamma+1}, \frac{\mu-\sigma}{N+\gamma+1})$ and $(x_{i,H}, w_{i,H}) = (\frac{\mu+\sigma-t}{N+1}, \frac{t}{\gamma})$, respectively. Then the expected social welfare in (D.5) becomes

$$
\begin{aligned}
\mathcal{S}_2 = \frac{N}{4\gamma^2(N+1)^2(\gamma+N+1)^2}\Big(&\gamma^2(N+1)^2(2\mu^2(\beta d(N-1)-dN+N+2)-2t(\mu+\sigma)(6\beta d(N-1)-6dN+1) \\
&+4d\mu\sigma(\beta(N-1)-N)+2\sigma^2(\beta d(N-1)-dN+N+2)+t^2(12\beta d(N-1)-3(4d+1)N-2)) \\
&+\gamma^3(N+1)(\mu^2(4\beta d(N-1)-4dN+3N+5)+2\mu\sigma(4\beta d(N-1)-4dN+N+3)-4\mu t(3\beta d(N-1)-3dN+1) \\
&+(\sigma-t)(\sigma(N(4(\beta-1)d+3)-4\beta d+5)+t(8d(\beta(-N)+\beta+N)+3N+1))) \\
&+\gamma^4(\mu+\sigma-t)((\mu+\sigma)(2\beta d(N-1)-2dN+N+2)+t(2d(\beta(-N)+\beta+N)+N)) \\
&+\gamma(N+1)^3t(-4\beta d(N-1)(\mu+\sigma-2t)+4dN(\mu+\sigma)-t(8dN+N+1))+2d(N+1)^4t^2(\beta(N-1)-N)\Big).
\end{aligned}
$$

The unconstrained optimal solution that maximize $\mathcal{S}_2$ is $t^* = \frac{\gamma\left(2d(N-\beta(N-1))(\gamma+N+1)-\gamma\right)}{-2\beta d(N-1)(\gamma+N+1)^2+2dN(\gamma+N+1)^2+\gamma N(\gamma+N+2)+\gamma}(\mu+\sigma)$, which is less than $\mu+\sigma$. If $d<\frac{\gamma}{2(N+\beta-\beta N)(N+\gamma+1)}$, then we have $t^*<0$, and the optimal solution and the corresponding objective are

$$
t_2^* = \mu-\sigma \qquad \text{and} \qquad \mathcal{S}_2^* = \frac{N\sigma\left((N+1)\mu+\sigma+2\beta d(N-1)\sigma-2dN\sigma\right)}{(N+1)^2}.
$$

If $d\geq\frac{\gamma}{2(N+\beta-\beta N)(N+\gamma+1)}$ and $\sigma\geq\frac{\gamma\mu(N+1)(\gamma+N+2)}{-4\beta d(N-1)(\gamma+N+1)^2+\gamma^2(4dN+N-1)+(8d+1)\gamma N(N+1)+4dN(N+1)^2}$, then we have $t^*\geq\mu-\sigma$, and the optimal solution and the corresponding objective are

$$
\begin{aligned}
&t_2^* = t^* = \frac{\gamma\left(2d(N-\beta(N-1))(\gamma+N+1)-\gamma\right)}{-2\beta d(N-1)(\gamma+N+1)^2+2dN(\gamma+N+1)^2+\gamma N(\gamma+N+2)+\gamma}(\mu+\sigma) \qquad \text{and} \\
&\mathcal{S}_2^* = \frac{N(\gamma+N+2)}{4(\gamma+N+1)^2\left(-2\beta d(N-1)(\gamma+N+1)^2+2dN(\gamma+N+1)^2+\gamma N(\gamma+N+2)+\gamma\right)}\times \\
&\quad\Big(-4\beta d(N-1)(\gamma+N+1)^2\left(\mu^2+\sigma^2\right)+\gamma^2\left(\mu^2(4dN+3N+2)+\sigma^2(4dN+3N+2)+2\mu(N+2)\sigma\right) \\
&\quad+2\gamma(N+1)(4dN+N+1)\left(\mu^2+\sigma^2\right)+4dN(N+1)^2\left(\mu^2+\sigma^2\right)+\gamma^3(\mu+\sigma)^2\Big).
\end{aligned}
$$

If $d\geq\frac{\gamma}{2(N+\beta-\beta N)(N+\gamma+1)}$ and $\sigma<\frac{\gamma\mu(N+1)(\gamma+N+2)}{-4\beta d(N-1)(\gamma+N+1)^2+\gamma^2(4dN+N-1)+(8d+1)\gamma N(N+1)+4dN(N+1)^2}$, then we have $t^*<\frac{\gamma(\mu-\sigma)}{N+\gamma+1}$, the optimal solution and the corresponding objective are

$$
\begin{aligned}
&t^* = \frac{\gamma(\mu-\sigma)}{N+\gamma+1} \qquad \text{and} \\
&\mathcal{S}_2^* = \frac{N}{2(N+1)^2(\gamma+N+1)^2}\Big(\sigma^2\left(4\beta d(N-1)(\gamma+N+1)^2+\gamma^2(2-4dN)-\gamma(N+1)(8dN-N-3)\right. \\
&\quad\left.-(N+1)^2(4dN-N-2)\right)+\mu^2(N+1)^2(\gamma+N+2)+2\gamma\mu(N+1)\sigma(\gamma+N+2)\Big).
\end{aligned}
$$

3. When $t\in[\frac{\gamma(\mu+\sigma)}{N+\gamma+1},\infty)$, each firm's optimal production amounts in the low-demand and high-demand cases are $(x_{i,L},w_{i,L})=(\frac{\mu-\sigma}{N+\gamma+1},\frac{\mu-\sigma}{N+\gamma+1})$ and $(x_{i,H},w_{i,H})=(\frac{\mu+\sigma}{N+\gamma+1},\frac{\mu+\sigma}{N+\gamma+1})$, respectively. Then the expected social welfare in (D.5) becomes

$$
\mathcal{S}_3 = \frac{N(\gamma+N+2)\left(\mu^2+\sigma^2\right)}{2(\gamma+N+1)^2}.
$$

The optimal solution t_3^* can be any value larger than $\frac{\gamma(\mu+\sigma)}{N+\gamma+1}$ and the corresponding objective is $\mathcal{S}_3^*=\frac{N(\gamma+N+2)\left(\mu^2+\sigma^2\right)}{2(\gamma+N+1)^2}$.

Then, similar to Section D.2, we compare the optimal values in these subspaces in different situations and get the following results. Lastly, we define $\delta_c^P:=\frac{\gamma}{2(N+\beta-\beta N)(N+\gamma+1)}$ and $\Omega_c^P:=\frac{\gamma(N+1)(\gamma+N+2)}{2\sqrt{2}d(N-\beta(N-1))(\gamma+N+1)^2+\gamma\left(\gamma(N-\sqrt{2}+1)+(N+1)(N-\sqrt{2}+2)\right)}$, and summarize the results as follows:

1. When $d\in[0,\delta_c^P)$, the optimal solution and the corresponding objective are

$$
t^*=0 \quad \text{and} \quad \mathcal{S}^* = \frac{N(2\beta d(N-1)-2dN+N+2)}{2(N+1)^2}(\mu^2+\sigma^2).
$$

2. When $d \in [\delta_c^P, \infty)$ and $\sigma/\mu \in [0, \Omega_c^P)$, the optimal solution and the corresponding objective are

$$t^* = \mathcal{C}_{c,1}^P \mu \qquad \text{and} \qquad \mathcal{S}^* = \mathcal{C}_{c,2}^P \mu^2 + \mathcal{C}_{c,3}^P \sigma^2,$$

where $\mathcal{C}_{c,1}^P := \frac{\gamma\left(2d(N-\beta(N-1))(\gamma+N+1)-\gamma\right)}{2dN(\gamma+N+1)^2-2\beta d(N-1)(\gamma+N+1)^2+\gamma N(\gamma+N+2)+\gamma}$, $\mathcal{C}_{c,2}^P := \frac{N(N+\gamma+2)(2dN+\gamma-2\beta d(N-1))}{2\left(2dN(N+\gamma+1)^2+\gamma N(N+\gamma+2)-2\beta d(N-1)(N+\gamma+1)^2+\gamma\right)}$ and $\mathcal{C}_{c,3}^P := \frac{N\left(N+2+2\beta d(N-1)-2dN\right)}{2(N+1)^2}$.

3. When $d \in [\delta_c^P, \infty)$ and $\sigma/\mu \in [\Omega_c^P, 1)$, the optimal solution and the corresponding objective are

$$t^* = \mathcal{C}_{c,1}^P(\mu+\sigma) \qquad \text{and} \qquad \mathcal{S}^* = \mathcal{C}_{c,4}^P(\mu^2+\sigma^2) + \mathcal{C}_{c,5}^P \mu\sigma,$$

where $\mathcal{C}_{c,1}^P = \frac{\gamma\left(2d(N-\beta(N-1))(\gamma+N+1)-\gamma\right)}{2dN(\gamma+N+1)^2-2\beta d(N-1)(\gamma+N+1)^2+\gamma N(\gamma+N+2)+\gamma}$, $\mathcal{C}_{c,4}^P := \frac{N(N+\gamma+2)\left(2\gamma+4dN-4\beta d(N-1)(N+\gamma+1)^2+(N+\gamma+2)(4dN^2+\gamma^2+2\gamma N+4d\gamma N)\right)}{4(N+\gamma+1)^2\left(2dN(N+\gamma+1)^2+\gamma N(N+\gamma+2)-2\beta d(N-1)(N+\gamma+1)^2+\gamma\right)}$ and $\mathcal{C}_{c,5}^P := \frac{\gamma^2 N(N\gamma+2)^2}{2(N+\gamma+1)^2\left(2dN(N+\gamma+1)^2+\gamma N(N+\gamma+2)-2\beta d(N-1)(N+\gamma+1)^2+\gamma\right)}$.

Lastly, we analyze the comparative statics when $\gamma \to \infty$.

1. The threshold of d. Obviously, δ_c^P decreases in N.

2. The threshold Ω_c^P. We have

$$\lim_{\gamma\to\infty} \frac{\partial \Omega_c^P}{\partial N} = \frac{\sqrt{2}(4d\beta - 2d - 1)}{(1-\sqrt{2}+N+2\sqrt{2}d(\beta+N-\beta N))^2},$$

which implies that Ω_c^P increases in N when $\beta \geq \frac{2d+1}{4d}$ and decreases in N when $\beta < \frac{2d+1}{4d}$.

3. The optimal tax rate t_c^P. When $\sigma/\mu < \Omega_c^P$, we have

$$\lim_{\gamma\to\infty} \frac{\partial t_c^P}{\partial N} = \frac{(2d+1-4d\beta)}{(N+2d(\beta+N-\beta N))^2}\mu.$$

When $\sigma/\mu > \Omega_c^P$, we have

$$\lim_{\gamma\to\infty} \frac{\partial t_c^P}{\partial N} = \frac{(2d+1-4d\beta)}{(N+2d(\beta+N-\beta N))^2}(\mu+\sigma).$$

Therefore, we have t_c^P increases in N when $\beta < \frac{2d+1}{4d}$ and decreases in N when $\beta > \frac{2d+1}{4d}$.

4. The expected social welfare $\mathcal{S}_c^P$. When $\sigma/\mu < \Omega_c^P$ and $d \geq \delta_c^P$, we have

$$h(\beta) := \lim_{\gamma\to\infty} \frac{\partial \mathcal{S}_c^P}{\partial N} = \frac{\beta d\mu^2}{(2d(\beta+N-\beta N)+N)^2} + \frac{\sigma^2(\beta d(3N-1)-2dN+1)}{(N+1)^3}$$
$$\frac{\partial h(\beta)}{\partial \beta} = \frac{d\mu^2(2d(\beta(N-1)+N)+N)}{(2d(\beta(-N)+\beta+N)+N)^3} + \frac{d(3N-1)\sigma^2}{(N+1)^3} \geq 0$$
$$h(0) = \frac{\sigma^2(1-2dN)}{(N+1)^3} \leq 0$$
$$h(1) = \frac{d\mu^2}{(2d+N)^2} + \frac{\sigma^2(d(N-1)+1)}{(N+1)^3} \geq 0.$$

Therefore, there exists a threshold $\hat{\beta}^P$ such that $\mathcal{S}_c^P$ increases in N when $\beta > \hat{\beta}^P$ and decreases in N when $\beta < \hat{\beta}^P$. When $\sigma/\mu > \Omega_c^P$, we have

$$\lim_{\gamma\to\infty} \frac{\partial \mathcal{S}_c^P}{\partial N} = \frac{\beta d(\mu+\sigma)^2}{2(N+2d(\beta+N-\beta N))^2} \geq 0.$$

Therefore, we have $\mathcal{S}_c^P$ increases in N when $\sigma/\mu > \Omega_c^P$.

□

D.9. Solution to Problem (8)

Given a cap k, firm i will optimize its production and abatement according to the following problem.

$$\max_{0\le x_i - w_i \le k;\ w_i\ge 0} \Big\{(\theta - x_i - x_{-i})x_i - \frac{1}{2}\gamma w_i^2\Big\}.$$

Similar to Proof D.3, solving the problem, we have

$$x_i^*(\theta,k) = \begin{cases} \frac{\theta - x_{-i}}{2} & \frac{\theta-x_{-i}}{2} \le k \\ \frac{\theta - x_{-i}+\gamma k}{\gamma+2}, & \frac{\theta-x_{-i}}{2} > k, \end{cases} \qquad w_i^*(\theta,k) = \begin{cases} 0 & \frac{\theta-x_{-i}}{2} \le k \\ \frac{\theta - x_{-i}-2k}{\gamma+2} & \frac{\theta-x_{-i}}{2} > k. \end{cases}$$

In symmetric Nash equilibrium, each firm would produce the same amount of products (i.e., $x_{-i} = (N-1)x_i$) and this symmetric condition leads to the following fixed point problem:

$$x_i^*(\theta,k) = \begin{cases} \frac{\theta-(N-1)x_i^*}{2} & \frac{\theta-(N-1)x_i^*}{2} \le k \\ \frac{\theta-(N-1)x_i^*+\gamma k}{\gamma+2} & \frac{\theta-(N-1)x_i^*}{2} > k, \end{cases} \qquad w_i^*(\theta,k) = \begin{cases} 0 & \frac{\theta-(N-1)x_i^*}{2} \le k \\ \frac{\theta-(N-1)x_i^*-2k}{\gamma+2} & \frac{\theta-(N-1)x_i^*}{2} > k. \end{cases}$$

Then, we can deduce that the equilibrium solution is:

$$x_i^*(\theta,k) = \begin{cases} \frac{\theta+\gamma k}{N+\gamma+1} & k \le \frac{\theta}{N+1} \\ \frac{\theta}{N+1} & k > \frac{\theta}{N+1}, \end{cases} \qquad w_i^*(\theta,k) = \begin{cases} \frac{\theta-(N+1)k}{N+\gamma+1} & k \le \frac{\theta}{N+1} \\ 0 & k > \frac{\theta}{N+1}. \end{cases}$$

□

D.10. Proof of Lemma 2

Having each firm's response in Section D.9, the regulator's optimization problem is as follows (Since we study the symmetric Nash equilibrium, we can write $\sum_{i=1}^N x_i^*(\theta,k)$ as $Nx^*(\theta,k)$ and $\sum_{i=1}^N w_i^*(\theta,t)$ as $Nw^*(\theta,t)$):

$$\mathcal{S}^* = \max_{k\ge 0} \mathbb{E}_\theta\left[\theta N x^*(\theta,k) - \frac{1}{2}\big(Nx^*(\theta,k)\big)^2 - \frac{\gamma N}{2} w^*(\theta,k)^2 - d\big(\beta N + (1-\beta)N^2\big)(x^*(\theta,k) - w^*(\theta,k))^2\right]. \tag{D.6}$$

Similar to previous proofs, we divide the decision space into three subspaces and find the optimal solution of each subspace. Then we compare the optimal solutions of the three subspaces to get the global optimal solution. First, we compute the optimal solution in each subspace.

1. When $k \in [0, \frac{\mu-\sigma}{N+1}]$, each firm's optimal production amounts in the low-demand and high-demand cases are $(x_{i,L}, w_{i,L}) = (\frac{\mu-\sigma+\gamma k}{N+\gamma+1}, \frac{\mu-\sigma-(N+1)k}{N+\gamma+1})$ and $(x_{i,H}, w_{i,H}) = (\frac{\mu+\sigma+\gamma k}{N+\gamma+1}, \frac{\mu+\sigma-(N+1)k}{N+\gamma+1})$, respectively. Then the social welfare in (D.6) becomes

$$\begin{aligned}\mathcal{S}_1 = \frac{N}{2(\gamma+N+1)^2}\Big(&2\beta dk^2(N-1)(\gamma+N+1)^2 - \gamma^2((2d+1)kN - 2\mu)k - 2dk^2N + N\sigma^2 + 2\sigma^2 + \\ &+\gamma\left(2k\mu(N+2) - \left(k^2(N+1)(4dN+N+1)\right) + \mu^2 + \sigma^2\right) + (N+2)\left(\mu^2 - 2dk^2N^2\right)\Big).\end{aligned}$$

The unconstrained optimal solution that maximizes $\mathcal{S}_1$ is $k^* = \frac{\gamma\mu(\gamma+N+2)}{-2\beta d(N-1)(\gamma+N+1)^2 + 2dN(\gamma+N+1)^2 + \gamma N(\gamma+N+2)+\gamma}$. Since $\mathcal{S}_1$ is a quadratic function of k and may not be monotonic in this subspace, we discuss different situations. If $d < \frac{\gamma}{2(N+\beta-\beta N)(N+\gamma+1)}$, then we have $k^* > \frac{\mu-\sigma}{N+1}$, and the optimal solution and the corresponding objective are

$$k_1^* = \frac{\mu-\sigma}{N+1} \qquad \text{and}$$

$$\begin{aligned}\mathcal{S}_1^* = \frac{N}{2(N+1)^2(\gamma+N+1)^2}\Big(&\sigma^2\left(2\beta d(N-1)(\gamma+N+1)^2 - N\left(2d(\gamma+N+1)^2 + \gamma^2 - N(N+4) - 5\right) + 2\right) \\ &+ \mu^2(\gamma+N+1)^2(2\beta d(N-1) - 2dN + N + 2) - 2\mu\sigma(\gamma+N+1)(2d(\beta(N-1) - N)(\gamma+N+1) + \gamma)\Big).\end{aligned}$$

If $d \geq \frac{\gamma}{2(N+\beta-\beta N)(N+\gamma+1)}$ and $\sigma < \frac{\mu(\gamma+N+1)(2d(N-\beta(N-1))(\gamma+N+1)-\gamma)}{-2\beta d(N-1)(\gamma+N+1)^2+2dN(\gamma+N+1)^2+\gamma N(\gamma+N+2)+\gamma}$, then we have $k^* < \frac{\mu-\sigma}{N+1}$, then the optimal solution and the corresponding objective are

$$k_1^* = k^* = \frac{\gamma\mu(\gamma+N+2)}{-2\beta d(N-1)(\gamma+N+1)^2+2dN(\gamma+N+1)^2+\gamma N(\gamma+N+2)+\gamma} \qquad \text{and}$$

$$\begin{aligned}\mathcal{S}_1^* = {} & \frac{N(\gamma+N+2)}{2(\gamma+N+1)^2\left(-2\beta d(N-1)(\gamma+N+1)^2+2dN(\gamma+N+1)^2+\gamma N(\gamma+N+2)+\gamma\right)} \times \\ & \Big(-2\beta d(N-1)(\gamma+N+1)^2\left(\mu^2+\sigma^2\right)+\sigma^2\left(\gamma^2(2dN+N)+\gamma(N+1)(4dN+N+1)+2dN(N+1)^2\right) \\ & +\mu^2(\gamma+N+1)^2(2dN+\gamma)\Big).\end{aligned}$$

If $d \geq \frac{\gamma}{2(N+\beta-\beta N)(N+\gamma+1)}$ and $\sigma \geq \frac{\mu(\gamma+N+1)(2d(N-\beta(N-1))(\gamma+N+1)-\gamma)}{-2\beta d(N-1)(\gamma+N+1)^2+2dN(\gamma+N+1)^2+\gamma N(\gamma+N+2)+\gamma}$, then we have $k^* \geq \frac{\mu-\sigma}{N+1}$, and the optimal solution and the corresponding objective are

$$k_1^* = \frac{\mu-\sigma}{N+1} \qquad \text{and}$$

$$\begin{aligned}\mathcal{S}_1^* = {} & \frac{N}{2(N+1)^2(\gamma+N+1)^2}\Big(\sigma^2\left(2\beta d(N-1)(\gamma+N+1)^2-N\left(2d(\gamma+N+1)^2+\gamma^2-N(N+4)-5\right)+2\right) \\ & +\mu^2(\gamma+N+1)^2(2\beta d(N-1)-2dN+N+2)-2\mu\sigma(\gamma+N+1)(2d(\beta(N-1)-N)(\gamma+N+1)+\gamma)\Big).\end{aligned}$$

2. When $k \in (\frac{\mu-\sigma}{N+1}, \frac{\mu+\sigma}{N+1}]$, each firm's optimal production amounts in the low-demand and high-demand cases are $(x_{i,L}, w_{i,L}) = (\frac{\mu-\sigma}{N+1}, 0)$ and $(x_{i,H}, w_{i,H}) = (\frac{\mu+\sigma+\gamma k}{N+\gamma+1}, \frac{\mu+\sigma-(N+1)k}{N+\gamma+1})$, respectively. Then the social welfare in (D.6) becomes

$$\begin{aligned}\mathcal{S}_2 = {} & \frac{N}{4(N+1)^2(\gamma+N+1)^2}\Big(2\beta d(N-1)(\gamma+N+1)^2\left(k^2(N+1)^2+(\mu-\sigma)^2\right)+\gamma^2\big(-(2d+1)k^2N(N+1)^2 \\ & -(2dN-N-2)(\mu-\sigma)^2+2k(N+1)^2(\mu+\sigma)\big)-\gamma(N+1)\big(k^2(N+1)^2(4dN+N+1) \\ & +\mu^2(4dN-3N-5)+2\mu\sigma(-4dN+N+3)+\sigma^2(4dN-3N-5)-2k(N+1)(N+2)(\mu+\sigma)\big) \\ & -2(N+1)^2\left(dk^2N(N+1)^2+\mu^2((d-1)N-2)-2d\mu N\sigma+\sigma^2((d-1)N-2)\right)\Big).\end{aligned}$$

The unconstrained optimal solution that maximizes $\mathcal{S}_2$ is $k^* = \frac{\gamma(\gamma+N+2)(\mu+\sigma)}{-2\beta d(N-1)(\gamma+N+1)^2+2dN(\gamma+N+1)^2+\gamma N(\gamma+N+2)+\gamma}$. If $d < \frac{\gamma}{2(N+\beta-\beta N)(N+\gamma+1)}$, then we have $k^* > \frac{\mu+\sigma}{N+1}$, and the optimal solution and the corresponding objective are

$$k_2^* = \frac{\mu+\sigma}{N+1} \qquad \text{and} \qquad \mathcal{S}_2^* = \frac{N\left(\mu^2+\sigma^2\right)(2\beta d(N-1)-2dN+N+2)}{2(N+1)^2}.$$

If $d \geq \frac{\gamma}{2(N+\beta-\beta N)(N+\gamma+1)}$ and $\sigma \geq -\frac{\mu(\gamma+N+1)(2d(\beta(N-1)-N)(\gamma+N+1)+\gamma)}{-2\beta d(N-1)(\gamma+N+1)^2+\gamma^2(2(d+1)N+1)+\gamma(N+1)(4dN+2N+3)+2dN(N+1)^2}$, then we have $\frac{\mu+\sigma}{N+1} \geq k^* \geq \frac{\mu-\sigma}{N+1}$, and the optimal solution and the corresponding objective are

$$k_2^* = k^* = \frac{\gamma(\gamma+N+2)(\mu+\sigma)}{-2\beta d(N-1)(\gamma+N+1)^2+2dN(\gamma+N+1)^2+\gamma N(\gamma+N+2)+\gamma} \qquad \text{and}$$

$$\begin{aligned}\mathcal{S}_2^* = {} & \frac{N}{4(N+1)^2\left(2\beta d(N-1)(\gamma+N+1)^2-2dN(\gamma+N+1)^2-\gamma N(\gamma+N+2)-\gamma\right)}\Big(4\beta^2 d^2(N-1)^2(\gamma+N+1)^2(\mu-\sigma)^2 \\ & -4\beta d(N-1)\big(\mu^2\left(\gamma^2(2dN-1)+\gamma(N+1)(4dN-N-2)+(N+1)^2(2dN-N-2)\right) \\ & +\sigma^2\left(\gamma^2(2dN-1)+\gamma(N+1)(4dN-N-2)+(N+1)^2(2dN-N-2)\right)-2\mu\sigma(\gamma+N+1)(\gamma(2dN-1)+2dN(N+1))) \\ & +2\gamma(N+1)\left(\mu^2\left(N\left(4d^2N-2d(N+2)-N-3\right)-2\right)+\sigma^2\left(N\left(4d^2N-2d(N+2)-N-3\right)-2\right)+4d\mu N\sigma(1-2dN)\right) \\ & +\gamma^2\left(\mu^2(2N(2d(dN-1)-N-2)-1)-2\mu\sigma(1-2dN)^2+\sigma^2(2N(2d(dN-1)-N-2)-1)\right) \\ & +4dN(N+1)^2\left(\mu^2((d-1)N-2)-2d\mu N\sigma+\sigma^2((d-1)N-2)\right)\Big).\end{aligned}$$

If $d \geq \frac{\gamma}{2(N+\beta-\beta N)(N+\gamma+1)}$ and $\sigma < -\frac{\mu(\gamma+N+1)(2d(\beta(N-1)-N)(\gamma+N+1)+\gamma)}{-2\beta d(N-1)(\gamma+N+1)^2+\gamma^2(2(d+1)N+1)+\gamma(N+1)(4dN+2N+3)+2dN(N+1)^2}$, then we have $k^* < \frac{\mu-\sigma}{N+1}$, and the optimal solution and the corresponding objective are

$$k_2^* = \frac{\mu-\sigma}{N+1} \qquad \text{and}$$
$$\mathcal{S}_2^* = \frac{N}{4(N+1)^2(\gamma+N+1)^2}\Big(2\sigma^2\left(2\beta d(N-1)(\gamma+N+1)^2 - N\left(2d(\gamma+N+1)^2+\gamma^2-N(N+4)-5\right)+2\right) + 2\mu^2(\gamma+N+1)^2(2\beta d(N-1)-2dN+N+2) - 4\mu\sigma(\gamma+N+1)(2d(\beta(N-1)-N)(\gamma+N+1)+\gamma)\Big).$$

3. When $k \in (\frac{\mu+\sigma}{N+1}, \infty)$, each firm's optimal production amounts in the low-demand and high-demand cases are $(x_{i,L}, w_{i,L}) = (\frac{\mu-\sigma}{N+1}, 0)$ and $(x_{i,H}, w_{i,H}) = (\frac{\mu+\sigma}{N+1}, 0)$, respectively. Then the social welfare in (D.6) becomes
$$\mathcal{S}_3 = \frac{N\left(\mu^2+\sigma^2\right)\left(2\beta d(N-1)-2dN+N+2\right)}{2(N+1)^2}.$$
The optimal solution k_3^* can be any value larger than $\frac{\mu+\sigma}{N+1}$ and the corresponding objective is $\mathcal{S}_3^* = \frac{N\left(\mu^2+\sigma^2\right)(2\beta d(N-1)-2dN+N+2)}{2(N+1)^2}$.

Then, similar to Section D.4, we compare the optimal values in the three subspaces in different situations and get the following results. Lastly, we define $\delta_c^Q := \frac{\gamma}{2(N+\beta-\beta N)(N+\gamma+1)}$ and $\Omega_c^Q := \frac{(\gamma+N+1)(2d(N-\beta N+\beta)(\gamma+N+1)-\gamma)}{2dN(N+1)^2-2\beta d(N-1)(\gamma+N+1)^2+\gamma^2(2dN+\sqrt{2}N+\sqrt{2}-1)+\gamma(N+1)(4dN+\sqrt{2}N+2\sqrt{2}-1)}$, and summarize the results as follows:

1. When $d \in [0, \delta_c^Q)$, the optimal solution and the corresponding objective are
$$k^* = c,\ \forall c \geq \frac{\mu+\sigma}{N+1} \quad \text{and} \quad \mathcal{S}^* = \frac{N(2\beta d(N-1)-2dN+N+2)}{2(N+1)^2}\left(\mu^2+\sigma^2\right).$$

2. When $d \in [\delta_c^Q, \infty)$ and $\sigma/\mu \in [0, \Omega_c^Q)$, the optimal solution and the corresponding objective are
$$k^* = \mathcal{C}_{c,1}^Q \mu \qquad \text{and} \qquad \mathcal{S}^* = \mathcal{C}_{c,2}^Q\mu^2 + \mathcal{C}_{c,3}^Q\sigma^2,$$
where $\mathcal{C}_{c,1}^Q := \frac{\gamma(N+\gamma+2)}{2dN(\gamma+N+1)^2-2\beta d(N-1)(\gamma+N+1)^2+\gamma N(\gamma+N+2)+\gamma}$, $\mathcal{C}_{c,2}^Q := \frac{N(N+\gamma+2)\left(\gamma+2d(N+\beta-\beta N)\right)}{4dN(N+\gamma+1)^2+2\gamma N(N+\gamma+2)-4\beta d(N-1)(N+\gamma+1)^2+2\gamma}$ and $\mathcal{C}_{c,3}^Q := \frac{N(N+\gamma+2)}{2(N+\gamma+1)^2}$.

3. When $d \in [\delta_c^Q, \infty)$ and $\sigma/\mu \in [\Omega_c^Q, 1)$, the optimal solution and the corresponding objective are
$$k^* = \mathcal{C}_{c,1}^Q(\mu+\sigma) \qquad \text{and} \qquad \mathcal{S}^* = \mathcal{C}_{c,4}^Q(\mu^2+\sigma^2) + \mathcal{C}_{c,5}^Q\mu\sigma.$$
where $\mathcal{C}_{c,1}^Q = \frac{\gamma(N+\gamma+2)}{2dN(\gamma+N+1)^2-2\beta d(N-1)(\gamma+N+1)^2+\gamma N(\gamma+N+2)+\gamma}$, $\mathcal{C}_{c,4}^Q := \big[N\big(4\beta d(N-1)(\gamma^2(2dN-1) + (N+1)^2(2dN-N-2) + \gamma(N+1)(4dN-N-2)) - 4\beta^2 d^2(N-1)^2(N+\gamma+1)^2 - 4dN(N+1)^2(dN-N-2) + \gamma^2(1+2N(2+2d+N-2d^2N)) + 2\gamma(N+1)(2+3N+N^2-4d^2N^2+4dN+2dN^2)\big)\big]/\big[4(N+1)^2\big(2dN(N+\gamma+1)^2 + \gamma N(N+\gamma+2) - 2\beta d(N-1)(N+\gamma+1)^2+\gamma\big)\big]$ and $\mathcal{C}_{c,5}^Q := \frac{N\left(\gamma+2d(\beta N-\beta-N)(N+\gamma+1)\right)^2}{2(N+1)^2\left(2dN(N+\gamma+1)^2+\gamma N(N+\gamma+2)-2\beta d(N-1)(N+\gamma+1)^2+\gamma\right)}$.

Lastly, we analyze the comparative statics when $\gamma \to \infty$.

1. The threshold of d. Obviously, δ_c^Q decreases in N.

2. The threshold Ω_c^Q. We have
$$\lim_{\gamma\to\infty}\frac{\partial\Omega_c^Q}{\partial N} = \frac{\sqrt{2}(2d+1-4d\beta)}{(\sqrt{2}-1+\sqrt{2}N+2d(\beta+N-\beta N))^2},$$
which implies that Ω_c^Q decreases in N when $\beta \geq \frac{2d+1}{4d}$ and increases in N when $\beta < \frac{2d+1}{4d}$.

3. The optimal total cap Nk_c^Q. When $\sigma/\mu < \Omega_c^Q$, we have
$$\lim_{\gamma\to\infty}\frac{\partial(Nk_c^Q)}{\partial N}=\frac{2\beta d}{(N+2d(\beta+N-\beta N))^2}\mu.$$
When $\sigma/\mu > \Omega_c^Q$, we have
$$\lim_{\gamma\to\infty}\frac{\partial(Nk_c^Q)}{\partial N}=\frac{2\beta d}{(N+2d(\beta+N-\beta N))^2}(\mu+\sigma).$$
Therefore, we have Nk_c^Q increases in N.

4. The expected social welfare $\mathcal{S}_c^Q$. When $\sigma/\mu < \Omega_c^Q$ and $d \geq \delta_c^Q$, we have
$$\lim_{\gamma\to\infty}\frac{\partial\mathcal{S}_c^Q}{\partial N}=\frac{\beta d\mu^2}{(N+2d(\beta+N-\beta N))^2},$$
which implies that $\mathcal{S}_c^Q$ increases in N when $\sigma/\mu < \Omega_c^Q$. When $\sigma/\mu > \Omega_c^Q$, we have
$$\begin{aligned}g(\beta) := \lim_{\gamma\to\infty}\frac{\partial\mathcal{S}_c^Q}{\partial N} = &\frac{1}{2(N+1)^3(2d(\beta(-N)+\beta+N)+N)^2}\Big(4\beta^3d^3(N-1)^2(3N-1)(\mu-\sigma)^2\\ &-4\beta^2d^2(N-1)(N(d(8N-4)+3N-2)+1)(\mu-\sigma)^2+\beta d\big(\mu^2(N(4(7d^2+5d+1)N^2\\ &-2(10d(d+1)+1)N+8d+7)+1)-2\mu\sigma\big((4d(7d+5)+2)N^3-4(5d(d+1)+2)N^2\\ &+8dN+N-1\big)+\sigma^2\left(N\left(4(d(7d+5)+1)N^2-2(10d(d+1)+1)N+8d+7\right)+1\right)\big)\\ &-(2dN-1)(2dN+N)^2(\mu-\sigma)^2\Big).\end{aligned}$$
$$g(0)=\frac{(1-2dN)(\mu-\sigma)^2}{2(N+1)^3}\leq 0$$
$$\begin{aligned}g(1)=&\frac{1}{2(N+1)^3(2d+N)^2}\Big(4d^3(N-1)(\mu-\sigma)^2+4d^2(N^2-N+1)(\mu-\sigma)^2+d\big(\mu^2(N(2N(N+1)+7)+1)\\ &+2\mu(4N^2-N+1)\sigma+(N(2N^2+2N+7)+1)\sigma^2\big)+N^2(\mu-\sigma)^2\Big)\geq 0,\end{aligned}$$
Therefore, we have $\mathcal{S}_c^Q$ increases in N when $\beta\to 1$ and decreases in N when $\beta\to 0$.

□

D.11. Proof of Theorem 2

Similar to Section D.5, we compare these two curves in different situations and find the cross points (Define $\delta_c := \frac{\gamma}{2(N+\beta-\beta N)(N+\gamma+1)}$, $\Delta_c := \frac{\gamma(N^2+4N+3+(N+2)\gamma}{2(N+\gamma+1)^2(N-\beta(N-1))}$ and $\Omega_c := \sqrt{2}-1$).

1. $d\in[0,\delta_c)$. In this case, the two curves overlap.
2. $d\in[\delta_c,\Delta_c)$. In this case, $\Omega_c^Q \leq \Omega_c^P$. The two curves cross in the interval $[\Omega_c^Q,\Omega_c^P]$. Similarly, we can get the cross point Ω_c.
3. $d\in(\Delta_c,\infty)$. In this case, we have $\Omega_c^Q \geq \Omega_c^P$. The two curves cross in the interval $[\Omega_c^P,\Omega_c^Q]$. Similarly, we can get the cross point Ω_c. □

D.12. Proof of Proposition 4

We start with the price instrument. Since the social welfare keeps constant when $t\in[\frac{\gamma\theta}{\gamma+2},\infty)$, we can focus on the interval $[0,\frac{\gamma\theta}{\gamma+2}]$ to find the optimal tax rate. Given the realized market size θ, the social welfare can be written as
$$\mathcal{S}=-\frac{2d(\gamma+2)^2+\gamma(\gamma+4)}{8\gamma^2}t^2+\frac{(2d(\gamma+2)-\gamma)\theta}{4\gamma}t+\frac{3-2d}{8}\theta^2,$$
for $t\in[0,\frac{\gamma\theta}{\gamma+2}]$. When $d\in[0,\frac{\gamma}{2(\gamma+2)})$, the optimal tax rate is $t^*=0$. When $d\in[\frac{\gamma}{2(\gamma+2)},\infty)$, the optimal tax rate is $t^*=\frac{(2d(\gamma+2)-\gamma)\gamma\theta}{2d(\gamma+2)^2+\gamma(\gamma+4)}$. Therefore, the optimal price instrument for the quick-response case is as follows:

1. When $d \in [0, \frac{\gamma}{2(\gamma+2)})$, the optimal tax rates are $t_\ell^* = t_h^* = 0$ and the corresponding expected social welfare is $\mathcal{S}^* = \frac{3-2d}{8}(\mu^2+\sigma^2)$.
2. When $d \in [\frac{\gamma}{2(\gamma+2)}, \infty)$, the optimal tax rates are $t_\ell^* = \mathcal{C}_1^R(\mu-\sigma)$ and $t_h^* = \mathcal{C}_1^R(\mu+\sigma)$, and the corresponding expected social welfare is $\mathcal{S}^* = \mathcal{C}_3^R(\mu^2+\sigma^2)$. Here, we have $\mathcal{C}_1^R := \frac{(2d(\gamma+2)-\gamma)\gamma}{2d(\gamma+2)^2+\gamma(\gamma+4)}$ and $\mathcal{C}_3^R := \frac{(\gamma+3)(2d+\gamma)}{4d(\gamma+2)^2+2\gamma(\gamma+4)}$.

Then, we study the quantity instrument. Similarly, given the realized market size θ, we can focus on the interval $k \in [0, \frac{\theta}{2}]$ and the social welfare is

$$\mathcal{S} = -\frac{2d(\gamma+2)^2+\gamma(\gamma+4)}{2(\gamma+2)^2}k^2 + \frac{\gamma(\gamma+3)\theta}{(\gamma+2)^2}k + \frac{(\gamma+3)\theta^2}{2(\gamma+2)^2}.$$

When $d \in [0, \frac{\gamma}{2(\gamma+2)})$, the optimal cap is $k^* \geq \frac{\theta}{2}$. When $d \in [\frac{\gamma}{2(\gamma+2)}, \infty)$, the optimal cap is $k^* = \frac{\gamma(\gamma+3)\theta}{2d(\gamma+2)^2+\gamma(\gamma+4)}$. Therefore, the optimal quantity instrument for the quick-response case is as follows (Define $\delta^R := \frac{\gamma}{2(\gamma+2)}$):

1. When $d \in [0, \delta^R)$, the optimal caps are $k_\ell^* \geq \frac{\mu-\sigma}{2}$ and $k_h^* \geq \frac{\mu+\sigma}{2}$, and the corresponding expected social welfare is $\mathcal{S}^* = \frac{3-2d}{8}(\mu^2+\sigma^2)$.
2. When $d \in [\delta^R, \infty)$, the optimal caps are $k_\ell^* = \mathcal{C}_2^R(\mu-\sigma)$ and $k_h^* = \mathcal{C}_2^R(\mu+\sigma)$, and the corresponding expected social welfare is $\mathcal{S}^* = \mathcal{C}_3^R(\mu^2+\sigma^2)$. Here, we have $\mathcal{C}_2^R := \frac{\gamma(\gamma+3)}{2d(\gamma+2)^2+\gamma(\gamma+4)}$ and $\mathcal{C}_3^R = \frac{(\gamma+3)(2d+\gamma)}{4d(\gamma+2)^2+2\gamma(\gamma+4)}$.

□

D.13. Solution to Problem (9)

Under a hybrid instrument (κ, τ), the firm maximizes its profit:

$$\max_{x,w:x\geq w\geq 0}\left\{\theta x - x^2 - \frac{\gamma}{2}w^2 - \tau\max\{x-w-\kappa, 0\}\right\}.$$

We first divide the decision space into two subspaces according to the sign of $x-w-\kappa$ and find the optimal solution of each subspace. Then we compare the optimal solutions of the two subspaces to deduce the global optimal solution.

1. When $x-w-\kappa<0$, the firm maximizes its profit:
$$\max_{x,w:x\geq w\geq 0; x-w-\kappa<0}\left\{\theta x - x^2 - \frac{\gamma}{2}w^2\right\}.$$
Similar to Section D.3, we can derive the optimal solution as $(x^*, w^*) = (\frac{\theta}{2}, 0)$ when $\kappa \geq \frac{\theta}{2}$ and $(x^*, w^*) = (\frac{\theta+\kappa\gamma}{\gamma+2}, \frac{\theta-2\kappa}{\gamma+2})$ when $\kappa < \frac{\theta}{2}$.
2. When $x-w-\kappa \geq 0$, firm i maximizes its profit:
$$\max_{x,w:x\geq w\geq 0; x-w-\kappa\geq 0}\left\{\theta x - x^2 - \frac{\gamma}{2}w^2 - \tau(x-w-\kappa)\right\}.$$
Similar to Section D.1, we can derive the optimal solution as $(x^*, w^*) = (\frac{\theta-\tau}{2}, \frac{\tau}{\gamma})$ when $2\kappa + \frac{\gamma+2}{\gamma}\tau \leq \theta$ and $(x^*, w^*) = (\frac{\theta+\gamma\kappa}{\gamma+2}, \frac{\theta-2\kappa}{\gamma+2})$ when $2\kappa + \frac{\gamma+2}{\gamma}\tau > \theta$.

Then, we compare the maximal values of these subspaces in different cases, and derive the firm's optimal decisions as follows:

$$x^*(\theta,\kappa,\tau) = \begin{cases} \frac{\theta-\tau}{2} & 2\kappa + \frac{\gamma+2}{\gamma}\tau < \theta \\ \frac{\theta+\gamma\kappa}{\gamma+2} & 2\kappa + \frac{\gamma+2}{\gamma}\tau \geq \theta, \kappa < \frac{\theta}{2} \\ \frac{\theta}{2} & \kappa \geq \frac{\theta}{2}. \end{cases} \qquad w^*(\theta,\kappa,\tau) = \begin{cases} \frac{\tau}{\gamma} & 2\kappa + \frac{\gamma+2}{\gamma}\tau < \theta \\ \frac{\theta-2\kappa}{\gamma+2} & 2\kappa + \frac{\gamma+2}{\gamma}\tau \geq \theta, \kappa < \frac{\theta}{2} \\ 0 & \kappa \geq \frac{\theta}{2}. \end{cases}$$

□

D.14. Proof of Proposition 5

Having the firm's decision in Section D.13, the regulator's optimization problem is as follows:

$$\mathcal{S}^* = \max_{\kappa\geq 0;\tau\geq 0} \mathbb{E}_\theta\left[\theta\cdot x^*(\theta,\kappa,\tau) - x^*(\theta,\kappa,\tau)^2 - \frac{\gamma}{2}w^*(\theta,\kappa,\tau)^2 - d\big(x^*(\theta,\kappa,\tau)-w^*(\theta,\kappa,\tau)\big)^2 + \frac{1}{2}x^*(\theta,\kappa,\tau)^2\right]. \tag{D.7}$$

Similar to previous proofs, we first divide the feasible space into subspaces and find the optimal solution for each subspace. Then we compare the optimal values of the subspaces to get the global optimal solution. We start with the case when $d\in(0,\frac{\gamma}{2(\gamma+2)})$.

1. When $2\kappa\in[\mu+\sigma,\infty)$, the firm's optimal production amounts in the low-demand case and the high-demand case are $(x_L,w_L)=(\frac{\mu-\sigma}{2},0)$ and $(x_H,w_H)=(\frac{\mu+\sigma}{2},0)$, respectively. Then the social welfare in (D.7) becomes
$$\mathcal{S}_1 = \frac{3-2d}{8}(\mu^2+\sigma^2).$$
The optimal solution and the corresponding objective are
$$(\kappa_1^*,\tau_1^*)=(c_1,c_2),\ \forall c_1\geq\frac{\mu+\sigma}{2},\ c_2\geq 0;\qquad \mathcal{S}_1^*=\frac{3-2d}{8}(\mu^2+\sigma^2).$$

2. When $2\kappa\in[\mu-\sigma,\mu+\sigma)$ and $2\kappa+\frac{\gamma+2}{\gamma}\tau\in[\mu+\sigma,\infty)$, the firm's optimal production amounts in the low-demand case and the high-demand case are $(x_L,w_L)=(\frac{\mu-\sigma}{2},0)$ and $(x_H,w_H)=(\frac{\mu+\sigma+\gamma\kappa}{\gamma+2},\frac{\mu+\sigma-2\kappa}{\gamma+2})$, respectively. Then the social welfare in (D.7) becomes
$$\begin{aligned}\mathcal{S}_2 = \frac{1}{16(\gamma+2)^2}\Big(&-2d(\gamma+2)^2\left(4\kappa^2+(\mu-\sigma)^2\right)+\gamma^2\left(-4\kappa^2+8\kappa(\mu+\sigma)+3(\mu-\sigma)^2\right)\\ &-8\gamma\left(2\kappa^2-3\kappa(\mu+\sigma)-2\left(\mu^2-\mu\sigma+\sigma^2\right)\right)+24\left(\mu^2+\sigma^2\right)\Big)<\mathcal{S}_1^*.\end{aligned}$$

3. When $2\kappa\in[\mu-\sigma,\mu+\sigma)$ and $2\kappa+\frac{\gamma+2}{\gamma}\tau\in[\mu-\sigma,\mu+\sigma)$, the firm's optimal production amounts in the low-demand case and the high-demand case are $(x_L,w_L)=(\frac{\mu-\sigma}{2},0)$ and $(x_H,w_H)=(\frac{\mu+\sigma-\tau}{2},\frac{\tau}{\gamma})$, respectively. Then the social welfare in (D.7) becomes
$$\mathcal{S}_3 = -\frac{2(2d-3)\gamma^2\left(\mu^2+\sigma^2\right)-2\gamma\tau(2d(\gamma+2)-\gamma)(\mu+\sigma)+\tau^2\left(2d(\gamma+2)^2+\gamma(\gamma+4)\right)}{16\gamma^2}\leq\mathcal{S}_1^*,$$
where the equality holds when $\kappa\in[\frac{\mu-\sigma}{2},\frac{\mu+\sigma}{2})$ and $\tau=0$.

4. When $2\kappa\in[0,\mu-\sigma)$ and $2\kappa+\frac{\gamma+2}{\gamma}\tau\in[\mu+\sigma,\infty)$, the firm's optimal production amounts in the low-demand case and the high-demand case are $(x_L,w_L)=(\frac{\mu-\sigma+\gamma\kappa}{\gamma+2},\frac{\mu-\sigma-2\kappa}{\gamma+2})$ and $(x_H,w_H)=(\frac{\mu+\sigma+\gamma\kappa}{\gamma+2},\frac{\mu+\sigma-2\kappa}{\gamma+2})$, respectively. Then the social welfare in (D.7) becomes
$$\mathcal{S}_4 = \frac{-2d(\gamma+2)^2\kappa^2+\gamma^2(-\kappa)(\kappa-2\mu)+\gamma\left(-4\kappa^2+6\kappa\mu+\mu^2+\sigma^2\right)+3\left(\mu^2+\sigma^2\right)}{2(\gamma+2)^2}<\mathcal{S}_1^*.$$

5. When $2\kappa\in[0,\mu-\sigma)$ and $2\kappa+\frac{\gamma+2}{\gamma}\tau\in[\mu-\sigma,\mu+\sigma)$, the firm's optimal production amounts in the low-demand case and the high-demand case are $(x_L,w_L)=(\frac{\mu-\sigma+\gamma\kappa}{\gamma+2},\frac{\mu-\sigma-2\kappa}{\gamma+2})$ and $(x_H,w_H)=(\frac{\mu+\sigma-\tau}{2},\frac{\tau}{\gamma})$, respectively. Then the social welfare in (D.7) becomes
$$\begin{aligned}\mathcal{S}_5 = &\frac{-2d(\gamma+2)^2\left(4\kappa^2+(\mu+\sigma)^2\right)+\gamma^2\left(-4\kappa^2+8\kappa(\mu-\sigma)+3(\mu+\sigma)^2\right)-8\gamma\left(2\kappa^2+3\kappa(\sigma-\mu)-2\left(\mu^2+\mu\sigma+\sigma^2\right)\right)+24\left(\mu^2+\sigma^2\right)}{16(\gamma+2)^2}\\ &+\frac{(2d(\gamma+2)-\gamma)(\mu+\sigma)}{8\gamma}\tau-\frac{2d(\gamma+2)^2+\gamma(\gamma+4)}{16\gamma^2}\tau^2<\mathcal{S}_1^*.\end{aligned}$$

6. When $2\kappa \in [0, \mu-\sigma)$ and $2\kappa + \frac{\gamma+2}{\gamma}\tau \in [0, \mu-\sigma)$, the firm's optimal production amounts in the low-demand case and the high-demand case are $(x_L, w_L) = (\frac{\mu-\sigma-\tau}{2}, \frac{\tau}{\gamma})$ and $(x_H, w_H) = (\frac{\mu+\sigma-\tau}{2}, \frac{\tau}{\gamma})$, respectively. Then the social welfare in (D.7) becomes
$$\mathcal{S}_6 = -\frac{(2d-3)\gamma^2(\mu^2+\sigma^2) + 2\gamma\mu\tau(\gamma - 2d(\gamma+2)) + \tau^2(2d(\gamma+2)^2 + \gamma(\gamma+4))}{8\gamma^2} \le \mathcal{S}_1^*,$$
where the equality holds when $\kappa \in [0, \frac{\mu-\sigma}{2})$ and $\tau = 0$.

Therefore the optimal solution and the corresponding objective are
$$(\kappa^*, \tau^*) = (c_1, c_2)\, \forall c_1 \ge \frac{\mu+\sigma}{2},\, c_2 \ge 0 \quad \text{or} \quad (\kappa^*, \tau^*) = (c, 0)\, \forall c \in \Big[0, \frac{\mu+\sigma}{2}\Big);$$
$$\mathcal{S}^* = \frac{3-2d}{8}(\mu^2+\sigma^2).$$

Then we consider the case when $d \in (\frac{\gamma}{2(\gamma+2)}, \infty)$. When $2\kappa \in [0, \mu-\sigma)$ and $2\kappa + \frac{\gamma+2}{\gamma}\tau \in [\mu-\sigma, \mu+\sigma)$, the firm's optimal production amounts in the low-demand case and the high-demand case are $(x_L, w_L) = (\frac{\mu-\sigma+\gamma\kappa}{\gamma+2}, \frac{\mu-\sigma-2\kappa}{\gamma+2})$ and $(x_H, w_H) = (\frac{\mu+\sigma-\tau}{2}, \frac{\tau}{\gamma})$, respectively. Then the social welfare in (D.7) becomes
$$\mathcal{S}_1 = \frac{-2d(\gamma+2)^2\left(4\kappa^2 + (\mu+\sigma)^2\right) + \gamma^2\left(-4\kappa^2 + 8\kappa(\mu-\sigma) + 3(\mu+\sigma)^2\right) - 8\gamma\left(2\kappa^2 + 3\kappa(\sigma-\mu) - 2\left(\mu^2+\mu\sigma+\sigma^2\right)\right) + 24\left(\mu^2+\sigma^2\right)}{16(\gamma+2)^2}$$
$$+ \frac{(2d(\gamma+2)-\gamma)(\mu+\sigma)}{8\gamma}\tau - \frac{2d(\gamma+2)^2 + \gamma(\gamma+4)}{16\gamma^2}\tau^2.$$

The optimal solution and the corresponding objective are
$$(\kappa_1^*, \tau_1^*) = \Big(\frac{\gamma(\gamma+3)(\mu-\sigma)}{2d(\gamma+2)^2 + \gamma(\gamma+4)}, \frac{\gamma(2d(\gamma+2)-\gamma)(\mu+\sigma)}{2d(\gamma+2)^2+\gamma(\gamma+4)}\Big); \qquad \mathcal{S}_1^* = \frac{(\gamma+3)(2d+\gamma)(\mu^2+\sigma^2)}{4d(\gamma+2)^2 + 2\gamma(\gamma+4)}.$$

Similarly, we can find that the objective is no greater than $\mathcal{S}_1^*$. Therefore, the optimal solution and the corresponding objective are
$$(\kappa^*, \tau^*) = \Big(\mathcal{C}_1^H(\mu-\sigma), \mathcal{C}_2^H(\mu+\sigma)\Big); \qquad \mathcal{S}^* = \mathcal{C}_3^H(\mu^2+\sigma^2),$$
where $\mathcal{C}_1^H := \frac{\gamma(\gamma+3)}{2d(\gamma+2)^2+\gamma(\gamma+4)}$, $\mathcal{C}_2^H := \frac{\gamma(2d(\gamma+2)-\gamma)}{2d(\gamma+2)^2+\gamma(\gamma+4)}$ and $\mathcal{C}_3^H := \frac{(\gamma+3)(2d+\gamma)}{4d(\gamma+2)^2+2\gamma(\gamma+4)}$. □

D.15. Proof of Lemma A.1

Similar to Section D.2, we divide the feasible space into subspaces and compute the optimal solution in each subspace. After that, we compare the optimal objective values in these subspaces and derive the global optimal solution.

1. When $t \in [0, \frac{\gamma}{N+\gamma+1}\mathbb{E}[\tilde{\theta}_0])$, the firm's production and abatement amounts of the no-signal case and the with-signal case are $(x_0, w_0) = (\frac{\mathbb{E}[\tilde{\theta}_0]-t}{N+1}, \frac{t}{\gamma})$ and $(x_1, w_1) = (\frac{\mathbb{E}[\tilde{\theta}_1]-t}{N+1}, \frac{t}{\gamma})$, respectively. Then, the expected social welfare becomes
$$\mathcal{S}_1 = \frac{N}{2(N+1)^2}\Big(-\frac{N(2d(\gamma+N+1)^2 + \gamma(\gamma+N+2))}{\gamma^2}t^2 + \frac{2\mu(\gamma(2dN-1) + 2dN(N+1))}{\gamma}t$$
$$+ \mu^2(-2dN+N+2) + \frac{\sigma^2((2d-1)N-2)(\phi_H-\phi_L)^2}{(\phi_H+\phi_L-2)(\phi_H+\phi_L)} - \frac{t^2}{\gamma}\Big).$$

2. When $t \in [\frac{\gamma}{N+\gamma+1}\mathbb{E}[\tilde{\theta}_0], \frac{\gamma}{N+\gamma+1}\mathbb{E}[\tilde{\theta}_1])$, the firm's production and abatement amounts of the no-signal case and the with-signal case are $(x_0, w_0) = (\frac{\mathbb{E}[\tilde{\theta}_0]}{N+\gamma+1}, \frac{\mathbb{E}[\tilde{\theta}_0]}{N+\gamma+1})$ and $(x_1, w_1) = (\frac{\mathbb{E}[\tilde{\theta}_1]-t}{N+1}, \frac{t}{\gamma})$, respectively. Then, the expected social welfare becomes

$$
\begin{aligned}
\mathcal{S}_2 = &-\frac{N}{4\gamma^2(N+1)^2(\gamma+N+1)^2(\phi_H+\phi_L-2)(\phi_H+\phi_L)}\Big(\gamma^2(N+1)^2\Big(2\mu^2(\phi_H+\phi_L-2)(\phi_H+\phi_L)(N(d(\phi_H+\phi_L)-1)-2) \\
&+2\mu(\phi_H+\phi_L-2)(\phi_H+\phi_L)(2dN(\sigma(\phi_H-\phi_L)-3t(\phi_H+\phi_L))+t(\phi_H+\phi_L))+2\sigma^2(\phi_H-\phi_L)^2(dN(\phi_H+\phi_L-2)+N+2) \\
&-2\sigma t(6dN-1)(\phi_H-\phi_L)(\phi_H+\phi_L-2)(\phi_H+\phi_L)+t^2(3(4d+1)N+2)(\phi_H+\phi_L-2)(\phi_H+\phi_L)^2\Big) \\
&+\gamma^3(N+1)\Big(\mu^2(\phi_H+\phi_L-2)(\phi_H+\phi_L)(N(4d\phi_H+4d\phi_L-\phi_H-\phi_L-2)-3\phi_H-3\phi_L-2) \\
&+2\mu(\phi_H+\phi_L-2)(\phi_H+\phi_L)(\sigma((4d-1)N-3)(\phi_H-\phi_L)-2t(3dN-1)(\phi_H+\phi_L)) \\
&+\sigma^2(\phi_H-\phi_L)^2(N(4d(\phi_H+\phi_L-2)-\phi_H-\phi_L+4)-3\phi_H-3\phi_L+8)-4\sigma t(3dN-1)(\phi_H-\phi_L)(\phi_H+\phi_L-2)(\phi_H+\phi_L) \\
&+t^2((8d+3)N+1)(\phi_H+\phi_L-2)(\phi_H+\phi_L)^2\Big) \\
&+\gamma^4(\phi_H+\phi_L-2)(\mu(\phi_H+\phi_L)+\phi_H(\sigma-t)-\phi_L(\sigma+t))(((2d-1)N-2)(\mu(\phi_H+\phi_L)+\sigma(\phi_H-\phi_L))-(2d+1)Nt(\phi_H+\phi_L)) \\
&-\gamma(N+1)^3t(\phi_H+\phi_L-2)(\phi_H+\phi_L)(4dN(\mu(\phi_H+\phi_L)+\sigma(\phi_H-\phi_L))-t(8dN+N+1)(\phi_H+\phi_L)) \\
&+2dN(N+1)^4t^2(\phi_H+\phi_L-2)(\phi_H+\phi_L)^2\Big).
\end{aligned}
$$

3. When $t \in [\frac{\gamma}{N+\gamma+1}\mathbb{E}[\tilde{\theta}_1], \infty)$, the firm's production and abatement amounts of the no-signal case and the with-signal case are $(x_0, w_0) = (\frac{\mathbb{E}[\tilde{\theta}_0]}{N+\gamma+1}, \frac{\mathbb{E}[\tilde{\theta}_0]}{N+\gamma+1})$ and $(x_1, w_1) = (\frac{\mathbb{E}[\tilde{\theta}_1]}{N+\gamma+1}, \frac{\mathbb{E}[\tilde{\theta}_1]}{N+\gamma+1})$, respectively. Then, the expected social welfare becomes
$$\mathcal{S}_3 = \frac{N(\gamma+N+2)\left(\mu^2(\phi_H+\phi_L-2)(\phi_H+\phi_L)-\sigma^2(\phi_H-\phi_L)^2\right)}{2(\gamma+N+1)^2(\phi_H+\phi_L-2)(\phi_H+\phi_L)}.$$

We define $\delta^P_{mc} := \frac{\gamma}{2N(N+\gamma+1)}$ and $\Omega^P_{mc} := \frac{\gamma(N+1)(\gamma+N+2)(2-\phi_H-\phi_L)\sqrt{\phi_H+\phi_L}}{(\phi_H-\phi_L)\left((\gamma+N+1)(2dN(N+1)+\gamma(2dN-1))\sqrt{2}+\gamma(N+1)(\gamma+2+N)\sqrt{\phi_H+\phi_L}\right)}$, and summarize the results as follows:

1. When $d \in [0, \delta^P_{mc})$, the optimal tax and the corresponding expected social welfare are
$$t^* = 0 \qquad \text{and} \qquad \mathcal{S}^* = \frac{N((2d-1)N-2)\left(\sigma^2(\phi_H-\phi_L)^2-\mu^2(\phi_H+\phi_L-2)(\phi_H+\phi_L)\right)}{2(N+1)^2(\phi_H+\phi_L-2)(\phi_H+\phi_L)}.$$

2. When $d \in [\delta^P_{mc}, \infty)$ and $\sigma/\mu \in [0, \Omega^P_{mc})$, the optimal tax and the corresponding expected social welfare are
$$t^* = \mathcal{C}^P_{mc,1}\mu$$
and
$$
\begin{aligned}
\mathcal{S}^* = &\frac{N}{2(N+1)^2(\phi_H+\phi_L-2)(\phi_H+\phi_L)\left(\gamma^2(2dN+N)+\gamma(N+1)(4dN+N+1)+2dN(N+1)^2\right)}\times \\
&\Big(\gamma^2\left((2d+1)N\sigma^2(2dN-N-2)(\phi_H-\phi_L)^2+\mu^2(N+1)^2(\phi_H+\phi_L-2)(\phi_H+\phi_L)\right) \\
&+\gamma(N+1)\left(\mu^2(N+1)(2dN+N+2)(\phi_H+\phi_L-2)(\phi_H+\phi_L)+\sigma^2(2dN-N-2)(4dN+N+1)(\phi_H-\phi_L)^2\right) \\
&+2dN(N+1)^2\left(\sigma^2(2dN-N-2)(\phi_H-\phi_L)^2+\mu^2(N+2)(\phi_H+\phi_L-2)(\phi_H+\phi_L)\right)\Big),
\end{aligned}
$$
where $\mathcal{C}^P_{mc,1} := \frac{\gamma(\gamma(2dN-1)+2dN(N+1))}{\gamma^2(2dN+N)+\gamma(N+1)(4dN+N+1)+2dN(N+1)^2}$.

3. When $d \in [\delta^P_{mc}, \infty)$ and $\sigma/\mu \in [\Omega^P_{mc}, 1)$, the optimal tax and the corresponding expected social welfare are
$$t^* = \mathcal{C}^P_{mc,1}\left(\mu + \frac{\phi_H-\phi_L}{\phi_H+\phi_L}\sigma\right),$$
and
$$
\begin{aligned}
\mathcal{S}^* = &\frac{N(\gamma+N+2)}{4(\gamma+N+1)^2(\phi_H+\phi_L-2)(\phi_H+\phi_L)\left(\gamma^2(2dN+N)+\gamma(N+1)(4dN+N+1)+2dN(N+1)^2\right)}\times \\
&\Big(\gamma^2(\mu^2(\phi_H+\phi_L-2)(\phi_H+\phi_L)(N(4d+\phi_H+\phi_L+2)+2(\phi_H+\phi_L)) \\
&+\sigma^2(\phi_H-\phi_L)^2(N(-4d+\phi_H+\phi_L-4)+2(\phi_H+\phi_L-2))+2\mu(N+2)\sigma(\phi_H-\phi_L)(\phi_H+\phi_L-2)(\phi_H+\phi_L)) \\
&+2\gamma(N+1)(4dN+N+1)\left(\mu^2(\phi_H+\phi_L-2)(\phi_H+\phi_L)-\sigma^2(\phi_H-\phi_L)^2\right) \\
&+4dN(N+1)^2\left(\mu^2(\phi_H+\phi_L-2)(\phi_H+\phi_L)-\sigma^2(\phi_H-\phi_L)^2\right)+\gamma^3(\phi_H+\phi_L-2)(\mu(\phi_H+\phi_L)+\sigma(\phi_H-\phi_L))^2\Big).
\end{aligned}
$$

$\square$

D.16. Proof of Lemma A.2

Similar to Section D.4, we divide the feasible space into subspaces and compute the optimal solution in each subspace. After that, we compare the optimal objective values in these subspaces and derive the global optimal solution.

1. When $k \in [0, \frac{\mathbb{E}[\tilde{\theta}_0]}{2})$, the firm's production and abatement amounts of the no-signal case and the with-signal case are $(x_0, w_0) = (\frac{\mathbb{E}[\tilde{\theta}_0]+k\gamma}{N+\gamma+1}, \frac{\mathbb{E}[\tilde{\theta}_0]-(N+1)k}{N+\gamma+1})$ and $(x_1, w_1) = (\frac{\mathbb{E}[\tilde{\theta}_1]+k\gamma}{N+\gamma+1}, \frac{\mathbb{E}[\tilde{\theta}_1]-(N+1)k}{N+\gamma+1})$, respectively. Then, the expected social welfare becomes

$$\begin{aligned}\mathcal{S}_1 = &-\frac{N}{2(\gamma+N+1)^2(\phi_H+\phi_L-2)(\phi_H+\phi_L)}\Big(\gamma^2 k(\phi_H+\phi_L-2)(\phi_H+\phi_L)((2d+1)kN-2\mu)\\ &+\gamma\big(k^2(N+1)(4dN+N+1)(\phi_H+\phi_L-2)(\phi_H+\phi_L)-2k\mu(N+2)(\phi_H+\phi_L-2)(\phi_H+\phi_L)-\mu^2(\phi_H+\phi_L-2)(\phi_H+\phi_L)\\ &+\sigma^2(\phi_H-\phi_L)^2\big)+2dk^2N(N+1)^2(\phi_H+\phi_L-2)(\phi_H+\phi_L)-\mu^2(N+2)(\phi_H+\phi_L-2)(\phi_H+\phi_L)+(N+2)\sigma^2(\phi_H-\phi_L)^2\Big).\end{aligned}$$

2. When $k \in [\frac{\mathbb{E}[\tilde{\theta}_0]}{2}, \frac{\mathbb{E}[\tilde{\theta}_1]}{2})$, the firm's production and abatement amounts of the no-signal case and the with-signal case are $(x_0, w_0) = (\frac{\mathbb{E}[\tilde{\theta}_0]}{N+1}, 0)$ and $(x_1, w_1) = (\frac{\mathbb{E}[\tilde{\theta}_1]+k\gamma}{N+\gamma+1}, \frac{\mathbb{E}[\tilde{\theta}_1]-(N+1)k}{N+\gamma+1})$, respectively. Then, the expected social welfare becomes

$$\begin{aligned}\mathcal{S}_2 = &-\frac{N}{4(N+1)^2(\gamma+N+1)^2(\phi_H+\phi_L-2)(\phi_H+\phi_L)}\Big(\gamma^2(\phi_H+\phi_L)((2d+1)k^2N(N+1)^2(\phi_H+\phi_L-2)(\phi_H+\phi_L)\\ &-\sigma^2(2dN-N-2)(\phi_H-\phi_L)^2-2k(N+1)^2\sigma(\phi_H-\phi_L)(\phi_H+\phi_L-2))\\ &-2\mu(\phi_H+\phi_L-2)(\phi_H+\phi_L)(\sigma(\phi_H-\phi_L)\left(\gamma^2(2dN-N-2)+\gamma(N+1)(4dN-N-3)+2dN(N+1)^2\right)\\ &+\gamma k(N+1)^2(\gamma+N+2)(\phi_H+\phi_L))-\mu^2(\phi_H+\phi_L-2)(\phi_H+\phi_L)(\gamma^2(2dN-N-2)(\phi_H+\phi_L-2)\\ &+\gamma(N+1)(N(4d(\phi_H+\phi_L-2)-\phi_H-\phi_L+4)-3\phi_H-3\phi_L+8)+2(N+1)^2(dN(\phi_H+\phi_L-2)+N+2))\\ &+\gamma(N+1)\big(k^2(N+1)^2(4dN+N+1)(\phi_H+\phi_L-2)(\phi_H+\phi_L)^2-\sigma^2(\phi_H-\phi_L)^2(N(4d\phi_H+4d\phi_L-\phi_H-\phi_L-2)-3\phi_H-3\phi_L-2)\\ &-2k(N+1)(N+2)\sigma(\phi_H-\phi_L)(\phi_H+\phi_L-2)(\phi_H+\phi_L)\big)+2dk^2N(N+1)^4(\phi_H+\phi_L-2)(\phi_H+\phi_L)^2\\ &-2(N+1)^2\sigma^2(\phi_H-\phi_L)^2(N(d(\phi_H+\phi_L)-1)-2)\Big).\end{aligned}$$

3. When $k \in [\frac{\mathbb{E}[\tilde{\theta}_1]}{2}, \infty)$, the firm's production and abatement amounts of the no-signal case and the with-signal case are $(x_0, w_0) = (\frac{\mathbb{E}[\tilde{\theta}_0]}{N+1}, 0)$ and $(x_1, w_1) = (\frac{\mathbb{E}[\tilde{\theta}_1]}{N+1}, 0)$, respectively. Then, the expected social welfare becomes

$$\mathcal{S}_3 = \frac{N((2d-1)N-2)\,(\sigma^2(\phi_H-\phi_L)^2-\mu^2(\phi_H+\phi_L-2)(\phi_H+\phi_L))}{2(N+1)^2(\phi_H+\phi_L-2)(\phi_H+\phi_L)}.$$

We define $\delta_{mc}^P := \frac{\gamma}{2N(N+\gamma+1)}$ and $\Omega_{mc}^Q := \frac{(\gamma+N+1)\left(2dN(N+1)+\gamma(2dN-1)\gamma\right)(2-\phi_H-\phi_L)(\phi_H+\phi_L)}{(\phi_H-\phi_L)\left(((\gamma+N+1)(2dN(N+1)+\gamma(2dN-1))(\phi_H+\phi_L)+\gamma(N+1)(\gamma+N+2)\sqrt{2(\phi_H+\phi_L)}\right)}$ and summarize the results as follows:

1. When $d \in [0, \delta_{mc}^Q)$, the optimal cap and the corresponding expected social welfare are

$$k^* \geq \frac{\mathbb{E}[\tilde{\theta}_1]}{N+1} \qquad \text{and} \qquad \mathcal{S}^* = \frac{2d-3}{8}\Big(\mu^2 + \frac{(\phi_H-\phi_L)^2}{(2-\phi_H-\phi_L)(\phi_H+\phi_L)}\sigma^2\Big).$$

2. When $d \in [\delta_{mc}^Q, \infty)$ and $\sigma/\mu \in [0, \Omega_{mc}^Q)$, the optimal cap and the corresponding expected social welfare are

$$k^* = \mathcal{C}_{mc,1}^Q \mu,$$

and

$$\mathcal{S}^* = \frac{1}{2}N(\gamma+N+2)\left(\frac{\mu^2(2dN+\gamma)}{\gamma^2(2dN+N)+\gamma(N+1)(4dN+N+1)+2dN(N+1)^2} - \frac{\sigma^2(\phi_H-\phi_L)^2}{(\gamma+N+1)^2(\phi_H+\phi_L-2)(\phi_H+\phi_L)}\right),$$

where $\mathcal{C}_{mc,1}^Q := \frac{\gamma(\gamma+N+2)}{\gamma^2(2dN+N)+\gamma(N+1)(4dN+N+1)+2dN(N+1)^2}$.

3. When $d \in [\delta^Q_{mc}, \infty)$ and $\sigma/\mu \in [\Omega^Q_{mc}, 1)$, the optimal cap and the corresponding expected social welfare are

$$k^* = \mathcal{C}^Q_{mc,1}\big(\mu + \frac{\phi_H - \phi_L}{\phi_H + \phi_L}\sigma\big),$$

and

$$\begin{aligned}
\mathcal{S}^* = {} & \frac{N}{4(N+1)^2(\phi_H+\phi_L-2)(\phi_H+\phi_L)\left(\gamma^2(2dN+N)+\gamma(N+1)(4dN+N+1)+2dN(N+1)^2\right)} \times \\
& \Big(\gamma^2(\sigma^2(\phi_H-\phi_L)^2\left(N^2\left(4d^2(\phi_H+\phi_L)-2\right)-4N(d(\phi_H+\phi_L)+1)+\phi_H+\phi_L-2\right) \\
& +\mu^2(\phi_H+\phi_L-2)(\phi_H+\phi_L)(2N(2d(dN-1)(\phi_H+\phi_L-2)+N+2)+\phi_H+\phi_L) \\
& +2\mu\sigma(1-2dN)^2(\phi_H-\phi_L)(\phi_H+\phi_L-2)(\phi_H+\phi_L)) \\
& +2\gamma(N+1)(\sigma^2(\phi_H-\phi_L)^2\left(N\left(4d^2N(\phi_H+\phi_L)-2d(N+\phi_H+\phi_L+1)-N-3\right)-2\right) \\
& +\mu^2(\phi_H+\phi_L-2)(\phi_H+\phi_L)(N(2d(2dN(\phi_H+\phi_L-2)+N-\phi_H-\phi_L+3)+N+3)+2) \\
& +4d\mu N\sigma(2dN-1)(\phi_H-\phi_L)(\phi_H+\phi_L-2)(\phi_H+\phi_L))+4dN(N+1)^2(\mu^2(\phi_H+\phi_L-2)(\phi_H+\phi_L)(dN(\phi_H+\phi_L-2)+N+2) \\
& +2d\mu N\sigma(\phi_H-\phi_L)(\phi_H+\phi_L-2)(\phi_H+\phi_L)+\sigma^2(\phi_H-\phi_L)^2(N(d(\phi_H+\phi_L)-1)-2))\Big).
\end{aligned}$$

□

D.17. Proof of Proposition A.1

Similar to Appendix D.5, we compare the price and the quantity instruments. Define $\delta_{mc} := \frac{\gamma}{2N(N+\gamma+1)}$, $\Delta_{mc} := \frac{(N+2)\gamma^2+(N+1)(N+3)\gamma}{2N(N+\gamma+1)^2}$ and $\Omega_{mc} := \frac{\sqrt{2(\phi_H+\phi_L)}-(\phi_H+\phi_L)}{\phi_H-\phi_L}$.

1. $d \in [0, \delta_{mc})$. In this case, the two curves coincide with the expected social welfare without regulation.
2. $d \in [\delta_{mc}, \Delta_{mc})$. In this case, we have $\Omega^Q_{mc} \le \Omega^P_{mc}$. The price curve is over the quantity curve when $\sigma/\mu \in [0, \Omega^Q_{mc})$, and the quantity curve is over the price curve when $\sigma/\mu \in (\Omega^P_{mc}, 1)$. When $\sigma/\mu \in [\Omega^Q_{mc}, \Omega^P_{mc}]$, the two curves have one unique cross point Ω_{mc}.
3. $d \in (\Delta_{mc}, \infty)$. In this case, we have $\Omega^Q_{mc} \ge \Omega^P_{mc}$, and the two curves have one unique cross point Ω_{mc}. □